

High-Intensified Resemblance and Statistic-Restructured Alignment in Few-Shot Domain Adaptation for Industrial-Specialized Employment

Jirayu Petchhan[✉] and Shun-Feng Su[✉], *Fellow, IEEE*

Abstract—The industrial status quo has been changed hastily. Facilities require new technological factors and are less data-driven to make precise inferences at most. Thus, the use of virtual knowledge can be adapted to practical applications to limit human intervention in explanation, namely, comprehensive digital transformation. Deep transfer learning has the role of enabling transferable knowledge and awareness practice of annotating endpoints. However, practical transfer tasks are burdened while considering scarce or non-annotated instances accumulated from external sampling, and these may inevitably degenerate knowledge during training and further degrade performance. In this study, a novel framework composed of the spectral correlation alignment is proposed to enable spectral patterned structure for statistical criterion to diminish the domain disparity under a few-shot alignment and multiple attention orchestration to accelerate a few-shot adaptation. This technique also increases high-concentrated recognition in cross-domain similarity verification. The demonstrations are conducted under public visual adaptation bench-marks and realistic deployment. The empirical experiment illustrates that our approach is better in efficacy and adaptability under data-limited conditions. Besides, in similar realistic applications, it is evident that the proposed scheme is deployable and can be re-practicable to less time as well as data consumption, yet is better in performance.

Index Terms—Digital transformation, spectral correlation alignment, few-shot unsupervised domain adaptation, multiple attention orchestration.

I. INTRODUCTION

THE ERA of intelligent systems has emerged in recent years. Most industries require innovation in order to greatly increase productivity, to optimize maintenance, and to integrate with emerging technologies. Besides, the outgrowth of intelligent computation has contributed to providing more assessment and promptly decision making, such as the adapting to Human-Machine Interaction [1], [44], Internet of Things [2], systematic security [3], and industrial employment [4], [5]. Furthermore,

Manuscript received 6 May 2022; revised 24 July 2022 and 11 December 2022; accepted 11 February 2023. Date of publication 16 February 2023; date of current version 18 August 2023. This work was supported in part by the “Center for Cyber-Physical System Innovation” from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan. (*Corresponding author: Jirayu Petchhan.*)

The authors are with the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei City 106, Taiwan (e-mail: petchhan.jirayu@gmail.com; fsu@mail.ntust.edu.tw).

Digital Object Identifier 10.1109/TCE.2023.3245821

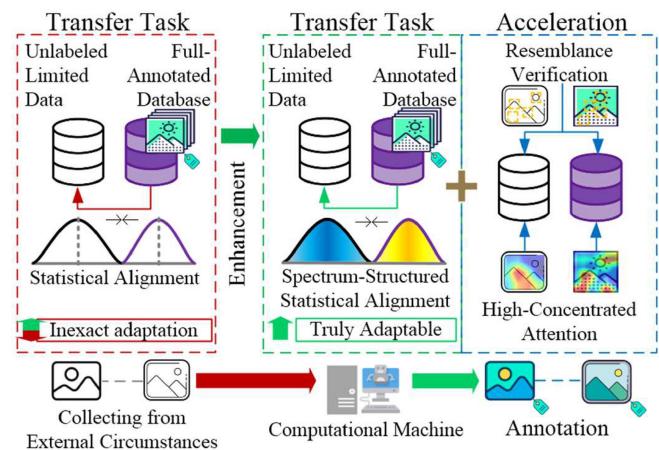


Fig. 1. The idea of the proposed schematic representation.

at this time, to get involved in a revolution from industrial 4.0 to 5.0, which requires agility, durability, and system flow perfection without human interference, all these lead to the desire for the adoption of full digital transformation so as to parse and determine useful approaches attributed to the engaged knowledge of physical and virtual entities. Recently, computational intelligence studies are continuously advanced so as to increase the effectiveness and efficiency of productive processes, to reduce evaluation time, and to increase the possibility of self-decision. In particular, deep learning-based collective studies play a role in consumer technologies considerably, such as access to useful manufacturing resources in investigation, like, object recognition, or detection [7], [10], and segmentation [8], [9]. In fact, transfer learning [22], [31] is the trending issue from many studies that brings us to concentrate this mentioned study.

The purpose of this study is to meet customers’ needs for increasing productivity through computational technology. Our challenges are to deal with dilemma issues on an adaptation task that needs to fitly adapt the interval domain-different circumstances as shown in Fig. 1. Hence, the proposed scheme consists of preparing a new structure for statistical metrics and accelerating recognition with high element-wise attention under the limitation of both information disparity and quantity. Initially, domain-invariant feature learning is incapable of or frequently degrades from being heavily attempted in a transfer task under

TABLE I
THE RECENT STUDIES AND OUR APPROACH

Studies	Feature Enhancement	Attention Strengthening	Similarity Learning	Deep Transfer Learning
Mixup [46], Cutmix [57]	✓ (Image-level mix)	-	-	-
HCDN [59]	-	-	✓ ¹	-
HardMix [39]	✓ (Feature-level mix)	-	✓ ¹ (Their approach can be utilized as a few-shot DA approach)	-
SoftTriple [45]	-	-	✓ ² (Sort & Sim)	-
Xmixup [58]	✓ (Image-level mix)	-	-	✓ (Fine-Tuning)
MDD [15], MCC [40], Robust & High-order CORAL [50], etc.	-	-	-	✓ (UDA)
FADA [37], MME [42]	-	-	-	✓ (Few-shot SDA)
PCS [43]	-	-	-	✓ (Few-shot Self-DA)
TAN [47]	-	✓ (Self-attention only)	-	✓ (UDA)
STOS (Ours)	✓ (Image-level mix & Feature-level orchestration)	✓ (Multiple attention orchestration)	✓ (Sort, Sim, & Shorten domain disparity as a few-shot UDA approach)	

Where SDA, UDA, and Self-DA stand for supervised- or semi-supervised-, unsupervised-, and self-supervised domain adaptation, respectively.

*¹ means similarity-based regularization and the decision module are absolutely independent.

*² Sort & Sim means the scheme, which can both categorize instance classes and discriminate inter-domain instance similarities simultaneously.

the considered constraints. We present high concentration that aids similarity-based learning to accelerate feature awareness and to benefit before problematic transfer.

We have further studied how to deal with a considerable issue to alleviate the domain disparity. It can be discovered that most existing techniques align general correlation matrices without considering the messy correlation inside before regularization. This makes domain adjustment quite ineffective. Although some studies have focused on and addressed this issue, unfortunately, most valuable correlations are eliminated. We therefore propose a technique to tackle the issue and this technique transforms the prepared correlation matrix into a spectral pattern for better alignment. This adoption can have a new structure for gradient computation. Despite having to deal with a few-shot transfer, new intrinsically valuable correlations in the knowledge are neither degraded nor interfered with.

II. RELATED WORK

We have summarized briefly for study comparisons in recent existing studies and our novel approach as tabulated in Table I.

A. Deep Similarity Learning

Due to the difficulty of not realizing the characteristic features of an object, it is not effective as an expectation for learning to improve the model. Data enhancement is a way of accelerating initial realization. For instance, Mixup [46], CutMix [57], and Xmixup [58]. In the network-level adjustment, HCDN [59] implemented triplet sampling to compare an exemplary instance between the most similar and different instances from two samples that are updated as the new triplet cost. HardMix [39] learns the feature-space blending and adjustment with sampling triplet cost instead of general transfer practice to adapt to other domains. SoftTriple [45] was further developed into a structure-level model that learn to classify and verify object similarity from predicted multiple vectors, each of which has a high-confidence predicted class. These are formulated as the multi-dimensional SoftMax operator, which is embedded and calculated for cost into only a single module.

B. Deep Transfer Learning

With a phenomenon known as negative transfer. This might cause sub-optimal outcomes and have a detrimental influence on performance from transferring all parameters that are irrelevant for the target, yet are still problematic for endpoint deployment in the consumer market, leading to the utilization of the transfer learning technique. Recently, the domain-adaptive learning has been arguably the most widely used technique that can learn and adapt simultaneously. DANN [54] was presented for domain adversarial-based learning by updating via a Gradient Reversal Layer (GRL) as domain regularization loss to predict domain category correctness. FADA [37] was presented as a few-shot supervised adaptation task. DDC [55] and MK-MMD [63] is learning domain invariance at low-level latent feature mapping in Replicative Kernel Hilbert Space (RKHS). For statistic-based diminishing domain disparity, D-CORAL [40] presented to form two-order covariance matrices for statistical metrics. MCC [26] is to re-form the class confusion minimization of the correlation. MDD [15] presented adapting theoretical and algorithmic bridging in order to achieve domain-adaptive learning, KHOMM [50], [56] reform features into high-order moments and is considered to shorten distances in RKHS as well. Robust-CORAL [50] reformulates iterative matrix inversion to correlation matrices with theoretical Newton iteration. MME [42] presented semi-supervised adaptation with minimax entropy regularization via both two few-labeled domains. PCS [43] has been introduced as a self-supervised adaptation technique, over which the memory bank builds a clustering of present and previous vectors for self- and cross-domain adaptation practice.

C. Implementing Computational Intelligence to Consumer Technology and Digital Transformation

The breakthrough has dramatically affected consumer technology, whether it be in the household or industrial sectors. Nowadays, the intelligence of consumer devices is measured by how much they can increase intelligence or how well

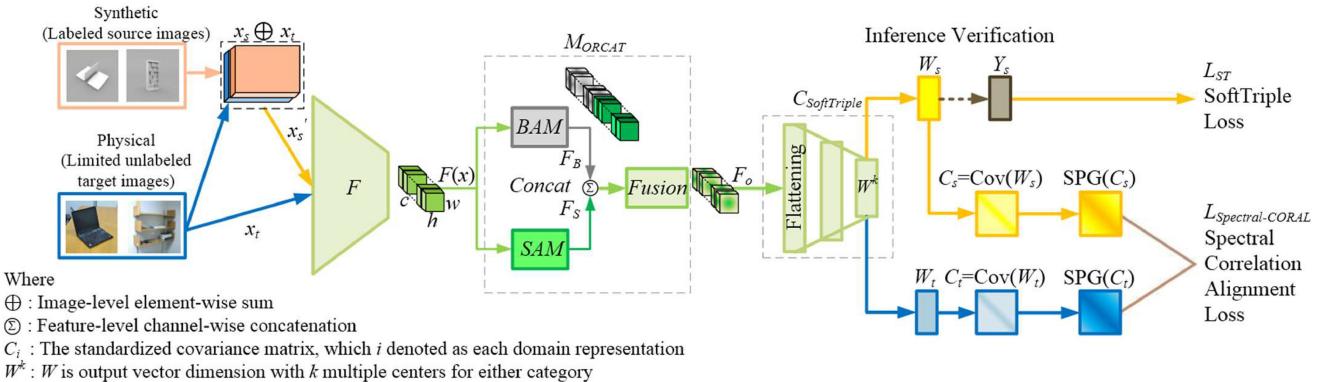


Fig. 2. The illustration of STOS framework.

they respond to human requirements. For explanation, a study described how intelligent computation relates to consumer technology and end users [12]. Concretely, in Mobile-DA [16], they presented a framework appropriate for household tasks where knowledge is transferred to a computational board in the TX2 mobile edge device for mobile analysis, similar to that of a robot. The lightweight Efficient-Fi [35] is compressible to Wi-Fi sensing and deploys in consumer devices to support Wi-Fi large-scale system compression to reduce massive transmission costs. Moreover, they performed full-supervision learning before deploying them in consumer devices [36] and attaching them to facility robots for exploring virtual as well as practical 3D physical objects. In order to protect the customers' data privacy, A study of [34] used intelligent computation with a blockchain-based structure to detect computerized attacks in IoE scenarios using smart things sensing data in order to protect customers' data privacy. The fog-enabled robotics [17] are surrounded by deep robot learning, which readjusts to the private real instances for many of edge devices to increase the potential of accessing low-level machine intelligence. By virtue of the emergency condition, the building-interior positioning on a consumer mobile device is calculable from sensors, as the computed network inference states [32]. The use of fast-adaptive learning to speed up deployment on consumer edges [19]. With the inaccessibility of complex object structures, point cloud generation [25] thoroughly explores local features and aids in the creation of new instances of interactive content and consumer products via their new edge convolution. These advancements have been further applied across the digital transformation to integrate with consumer facilities. In the absence of a thorough assessment, digitization of shop-floor manufacturing [14], [21] was thus established until smart health prognosis [28] followed. In spite of the tricky equipment health investigation in the past, to simplify prognosis and properly fit into facility-based typical hardware, DTL framework [31] is adaptable to analytical schemes in the diagnostic task via well-invested knowledge to understand unlabeled states of the motor bearing-failed vibration signal as well as for DFDD framework [11] of a two-phase digital-twin-assisted fault diagnosis to identify machine states that learn from virtual entities to adapt in the direction of the physical prognostic maintenance phase. And many studies, e.g., DATLN [30], DAFD [27], and others [22], [24], similarly

presented a framework for the aforementioned fault diagnosis on machines. Hereupon, we propose the approach for our contributions to two new manifolds in this article as follows:

- Meet multiple concentrated features that lead to multiple attention orchestration. The competency strengthens more intensive attention in feature-wise recognition and accelerates the validation of similarity instances under SoftTriplet.
- We present the spectral correlation alignment under optimizing factor adjustment of the newly nearest correlation structure. Specifically, the proposed approach is to enable more consistent computing in the projected spectral gradient pattern and include mitigation of the domain disparity under the new-patterned statistical criterion. Our two manifolds are established as STOS as illustrated in Fig. 2.
- Comprehensive demonstrations are performed by several strategies following official visual adaptation benchmarks and our circumstance preparations. In these results, our proposed approaches mostly obtained cutting-edge performances and retained their robustness, even though they performed in pragmatic transfers under the data-scarce condition.

III. OUR PROPOSED APPROACHES

To begin with the proposed system, we would start to describe the fundamental notation of many variants. We determine this by the fact that the labeled source domain $D_s = \{x_s^i, y_s^i\}_{(i=1)}^{N_s}$ and the unlabeled target domain $D_t = \{x_t^i\}_{(i=1)}^{N_t}$. They come from various locations, but have similar distributions. The goal of unsupervised domain adaptation is to train a deep neural network to alleviate classification errors and realize diminishing discrepancies in both domains. Ideally, the output labels are defined as the model function. We defined them as $F(\cdot)$, which is the feature extractor, and $C(\cdot)$, which is the classifier. Generally, the classification result is inference output $W = C_\theta(F_\theta(X))$. θ denotes the model parameters to be learned and adapted over time. H_s and H_t are denoted as outputs of latent feature representations over the source and target domains. Thus, output features from extracting process, $H_{s(t)} = F_\theta(X_{s(t)}) = [h_{s(t)}^1, \dots, h_{s(t)}^b] \in \mathbb{R}^{b \times L_f}$ and classification output dimension, $W_{s(t)} = C_\theta(F_\theta(X_{s(t)})) = [w_{s(t)}^1, \dots, w_{s(t)}^b] \in \mathbb{R}^{b \times L_o}$, $h_{s(t)}^i$ denotes the feature vector of the i -th instance, $w_{s(t)}^i$

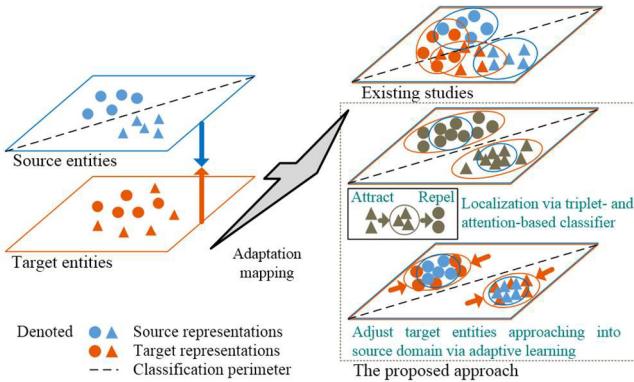


Fig. 3. The illustration of the final bound of labeling adapted by several strategies, each of which learns under a few-shot transfer condition. Existing studies still assign many target instances out-of-correct similar source, while our method realizes cross-domain homogeneity and disparity mutually.

denotes the output vector of the i -th instance, b is the batch size of the input images, L_f is the post-flattening latent feature output dimension, and L_o is the output vector dimension before mapping to label vectors. Although the target domain is only available through output dimensional vectors due to unlabeled data collection, this can be mitigated via source data to shorten domain shifting through hands-on adaptation. Existing studies and our system are contrasted in Fig. 3 to illustrate the final results of the inter-domain bound mapping for both domains.

A. Image-Level Mixup: Varying-Styled Auxiliary Data

The Mixup [46] image preprocessing is augmented data preparation for blending both images at the image-level features. Intuitively, in adaptation practices, the target protocol is inaccessible to labeled data. Instead, we take advantage of label data through source features. Thus, it is useful to combine two image-level features completely in a blending style. The instance style would change depending on the trade-off balance, which controls the intensity of the mixing.

$$x' = \alpha \cdot x_s + (1 - \alpha) \cdot x_t \quad (1)$$

where x_s and x_t are respectively raw input vectors, and the final instance is drawn from blending a few samples. $\alpha \in [0,1]$ is a weight of trade-off balance to control the level of instance-blending concentration between both image representations turning into new augmented data, as formulated simply in (1).

B. Cross-Domain Similarity Learning for Deep Few-Shot Unsupervised Domain Adaptation Preparation

In accordance with an illustration in Fig. 4, we explored the SoftTriplet [45] (a.k.a. SoftTriple or ST), which gains the ability to investigate empirical object similarity and regularize a model under categorization learning jointly without sampling any isolated triplet cost. With multiple k centers in either class that can measure many distances, each of them can be distinguished into maximally different and identical exemplars, similar to general triplet loss that is shaped as multi-kernel output computing, which is formulated as the Softmax operator. Based on likeness-based metrics from (2), this figures

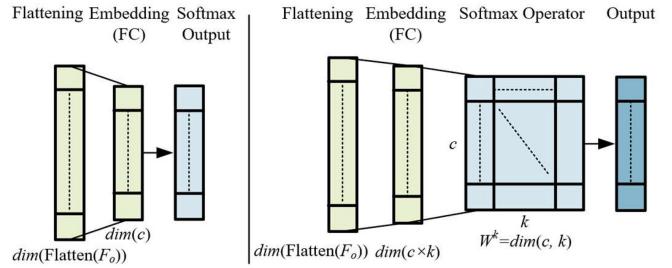


Fig. 4. The illustration of vanilla classifier and SoftTriplet [45] collaborated for few-shot unsupervised domain adaptation, where c and k are denoted as the number of classes and multiple centers for either class, respectively.

out a smoothed similarity output to measure SoftTriplet loss at (3).

$$S'_{i,c} = \sum_k \frac{\exp\left(\frac{1}{\gamma} x_t^T w_c^k\right)}{\sum_k \exp\left(\frac{1}{\gamma} x_t^T w_c^k\right)} \cdot x_t^T w_c^k \quad (2)$$

$$L_{\text{SoftTriplet}}(x_s^i) = -\log \frac{\exp\left(\lambda(S'_{i,y_s^i} - \delta)\right)}{\exp\left(\lambda(S'_{i,y_s^i} - \delta)\right) + \sum_{j \neq y_s^i} \exp\left(\lambda S'_{i,j}\right)} \quad (3)$$

where λ is a predefined scaling factor for smoothing output, γ is a predefined scaling factor for similarity between the sample x_i and the class c , in this case, it can be either ground truth y_i or the class probability prediction j , which can be processed into the source domain only. Although inferred target data is unable to use class vectors over smoothed similarity calculations due to the unlabeled data in hand for the target domain, the massive cross-domain knowledge gained from similarity mapping is therefore crucially significant for annotating the labels from the fully labeled source into the target domain. Other essential variables, W_c^k denoted $[w_1^k, \dots, w_C^k] \in \mathbb{R}^{b \times C}$ is the last fully connected (fc) layer output dimension along with similar to shaped output vectors in row rank. b is the batch size of the input images, and C is the number of classes that have shapes proportional to the output vector dimension length. These are defined to figure out the finite smoothed similarity S' as in (2) that assumes output probabilities along with each class having k multiple centers. The distributions over different classes are computed with the smoothed similarity obtained from either class C , and δ is a predefined margin for Triplet loss.

C. Emphasizing Feature: Multiple Attention Orchestration

To increase element-specific attention, we have considered a new style under an attention mechanism. The purpose is to discriminate between feature-characteristic styles and strengthen models' recognizability before adjusting domain-specific knowledge. The self-attention module (SAM), depicted in the upper-left corner of Fig. 5, encourages highlighting local region understanding before minimizing domain differences, as demonstrated in a study [47] that has implemented SAM to learn with batch spectral penalization into the conditional adversarial-based learning; however, extracted feature maps with the majority do not emphasize as they should on feature-specific spaces under our considered constraints.

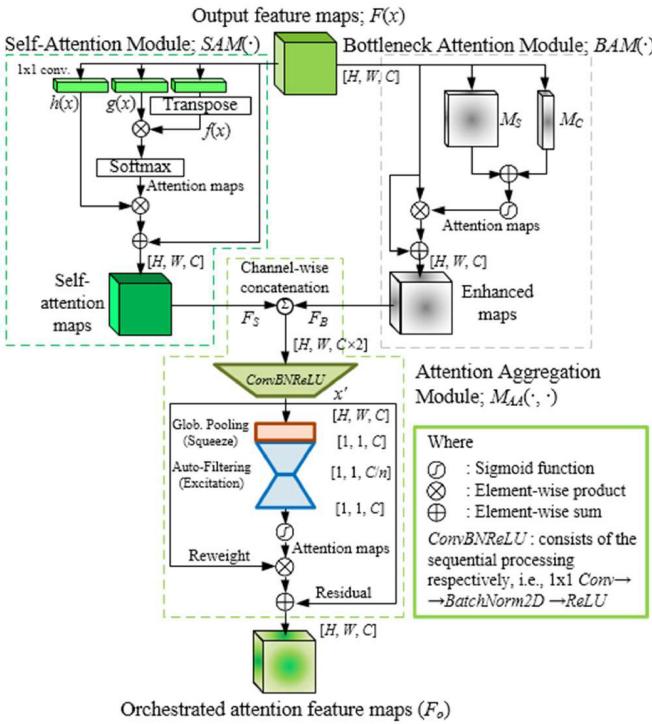


Fig. 5. The illustration of self-attention module (SAM) (upper-left), bottleneck attention module (BAM) (upper-right), and the orchestrated attention module (shown in full in the figure). This prepared module filters and excites over concatenated attentions obtained after BAM and SAM refined features.

The proposed multiple attention orchestration (M_{ORCAT}) implements refined features based on blending computation that enable more intense perception, as illustrated in Fig. 5, in which the aggregation module is generated by [62]. This module ensures that the positioning is coordinated with respect to some enhanced features. We decided on the bottleneck attention module (BAM) [33] at the upper-right corner of Fig. 5. This module uses the combination of spartial- and channel-based elemental attention to intensify a map to a certain extent. Hereupon, BAM- and SAM-based calculations are parallelly computed alongside each other. Following that, refined maps are sent to the attention aggregation module (M_{AA}) which is comparable to auto-filtering for high concentration on features. In general, the global-and-local exploration seeks to discover what is the remarkable region for BAM and where the informational element for both attentions is located; however, there are a few emphases on which our approach capitalizes to enable re-interpretability via filtration to emphasize whether multiple attention-wise bounds are of similar attentiveness. Furthermore, since unfolded vectors are typically prone to information loss, the influence of non-transferable features can be to clamp down on the trivial weight and keep up the crucial information more competently rather than with our highlighting descriptor. Besides, prior to alleviating inconvenient negative transfers in low-shot adaptive learning, mitigating the occurrence of important data loss from embedding expansion is also beneficial.

$$F_B = BAM(x) = \sigma(M_C(x) + M_S(x)) \times x + x \quad (4)$$

$$F_S = SAM(x) = \mu(\text{softmax}(f(x) \times g(x)) \times h(x)) + x \quad (5)$$

where x represents the encoded feature block from the backbone network, M_C and M_S in BAM processing in (4) are channel- and spartial-wise attention, respectively, whereas $f(x)$, $g(x)$, and $h(x)$ in SAM processing in (5) are equivalent to query, key, and value [49]. Instead, the image recognition is processed as feature outputs from a 1×1 convolution operation. Likewise, a control value μ is also thought of as an initialized value ranging from zero to the highest-boundary constraints ($\mu \geq 0$), which is progressively assigned many parameters as the training process progresses. Hereinafter, the initial state of SAM is then set according to [47], [48], i.e., $\mu = 0$, and σ represents the sigmoid activation function.

$$x' = ConvBNReLU(cat(F_B, F_S)) \quad (6)$$

$$F_o = \sigma(SE(x') \times x') + x' \quad (7)$$

where x' represents the post-normalized feature obtained by concatenating two enhanced features; $cat(\cdot, \cdot)$, followed by 1×1 batch-normalized convolution via the $ConvBNReLU$ block in (6). Finally, as formulated in (7), F_o is computed to high-intensive attention feature maps from feature aggregation enhancing over $SE(\cdot)$ layers [29]. These two modules from (6) to (7) are analogous to the sequential extract-squeeze-and-excitation processing, i.e., ESE-layer from incoming channel-concatenated features to encourage greater concentration. The entire module can be expressed as $F_o = M_{ORCAT}(F(x))$.

D. New Adaptation Paradigm: Spectral Correlation Alignment

The purpose of this study is to mitigate the domain disparity. This is a benefit for the learning network in preventing excessive overfitting from the source classification process, whereas this learning algorithm has been optimized to learn only to reduce domain shifting without regard to distinguishing image classes from labeled data. It is discouraging for the network that is unable to learn better representations, i.e., the feature instances have degenerated as the rich features have worsened on account of the fact that they are no longer arranged as instance classes. Therefore, training in categorizing classes and mitigating domain-specific issues is essential for more efficient network enhancement.

Previously, we have observed that the Deep CORAL (a.k.a. D-CORAL or CORAL) [40] increases the network's transferability by integrating a cost function in (8) that integrates the classification loss in the source domain with a second appropriately weighted loss, which penalizes the covariance difference. Despite its outstanding performance, D-CORAL is somewhat regardless of essential internal features for covariance matrices. We work based on the deep correlation alignment, which is formulated as:

$$L_{CORAL} = \frac{1}{4d^2} \|Cov(W_s) - Cov(W_t)\|_F^2 \quad (8)$$

where $W_{s(t)} \in \mathbb{R}^{b \times C}$ is the final fc layer output dimension. $Cov(\cdot)$ is computed to a standardized covariance matrix, shaped by $Cov(\cdot) \in \mathbb{R}^{C \times C}$, d is equal to the length of a correlation matrix C of any dimension. With the Frobenius norm $\|\cdot\|_F^2$,

a domain disparity can be minimized by the nearness cost function.

More recently, we discovered that Robust-CORAL (a.k.a. sqrt-CORAL) [50] uses practical transferability through the style-transformed statistical array, which is employed by square-root normalization to compute covariance that enables numerical stability. This method ensures the computation of covariance in an iterative matrix inversion layout by employing Denman-Beavers iterations. Nonetheless, it requires a calculational pair in every iteration in which the latter elements change a little, but the earliest elements have a large computing cost since the remainder can be calculated from earlier elements with only a few batches processing a variant of Newton's iteration background crucial for alignment and metrification afterward. Some techniques with adaptation properties in adversarial learning, such as DANN [54], transfer knowledge across low-level latent features as in an adversarial domain. MK-MMD (a.k.a. DAN) is a method for reducing domain confusion [55], [63]. Whereas numerous studies have implemented the mentioned abilities to mitigate domain gaps, data-limited constraints prevent the enhancement of features. Unfortunately, this problem still has a direct impact on the mentioned approach, which relies on multi-level layers for calculating transfer costs. Latterly, higher-order statistical metrics are expected to tackle similar issues, i.e., High-order CORAL [50] (a.k.a. KHoMM) generated high-order moment measures to achieve fine-grained correlation alignment. Despite the fact that this method uses higher-order moments that more precisely reflect the new form of features, as the number of orders and features contributes to the complexity of the features, it becomes more difficult to provide particulars regarding domain-specific information since they must still confront similar challenges to few-shot transfer to eliminate domain dislocation hardship.

Through our new insight, we have empowered the background to guarantee numerical consistency by applying the spectral projected gradient method (a.k.a. SPGM) [51]. This well-known technique for transforming normalized correlation matrices to the maximizes diagonal variance to the maximal boundary while retaining the off-diagonal correlation values, viz., this method rearranges and distributes relationships so as to form a new relevant pattern. Our method establishes a minor readjustment, but gains literally more productivity via factor tuning for the purpose of structuring newly effective correlation matrices instead. The proposed technique will be referred to as the spectral correlation alignment (Spectral-CORAL). To concretely establish the novel-fashioned patterned spectral gradient structure into standardized covariance matrices, formalize the closeness still under the Frobenius norm, and enable a single computation in an iteration. Initially, we adhere the nearness issue challenge, in which factor structure is defined on correlation matrices. The motivating structure is estimable for the factor model from their randomly generated expectation vector [51]. We have explicit directions to apply to predicted high-confidence output vectors. Accordingly,

$$\text{Cov}(\omega) = E(\omega \cdot \omega^T) = \sum_{j=1}^n w_j \cdot w_j^T + D \quad (9)$$

$$\begin{aligned} \text{Norm}_{i=1}^{\sqrt{2}}(\text{Cov}(W)) &= \text{Norm}_{i=1}^{\sqrt{2}} \left(\begin{bmatrix} 0.2361 & -0.1058 & -0.1302 \\ -0.1058 & 0.2617 & -0.1558 \\ -0.1302 & -0.1558 & 0.2861 \end{bmatrix} \right) \rightarrow \begin{bmatrix} 0.4064 & 0.0 & 0.0 \\ 0.0 & 0.4450 & 0.0 \\ 0.0 & 0.0 & 0.4809 \end{bmatrix} \\ C_{b=1}(\text{Cov}(W)) &= C_{b=1} \left(\begin{bmatrix} 0.2361 & -0.1058 & -0.1302 \\ -0.1058 & 0.2617 & -0.1558 \\ -0.1302 & -0.1558 & 0.2861 \end{bmatrix} \right) \rightarrow \begin{bmatrix} 1.00 & 0.6581 & 0.5581 \\ 0.6337 & 1.00 & 0.5825 \\ 0.5837 & 0.6325 & 1.00 \end{bmatrix} \end{aligned}$$

Fig. 6. The toy covariance in accordance with the first iteration.

The element-wise vector $\omega = \mathbb{R}^C$ is regarded according to the calculational theorem [51], which is comparable to $W \in \mathbb{R}^{b \times C}$, i.e., the predicted vector in multiple batches, while $\text{Cov}(\cdot)$ and $E(\cdot)$ are correspondingly denoted as the standardized covariance matrix and the expectation estimator to establish a two-order statistical association. As it follows, $E(\omega) = 0$. Thus, this enables them to assume in the simplest form, as in (9). By expecting that a variance of ω_i is the maximum boundary at 1 as for every element i for $\text{Cov}(\omega)$, i.e., the correlation matrix of ω and (9) denotes satisfying the b factor of the nearest structure expected to be an array, in which off-diagonal elements have a convergence criterion, expressively, $\sum_{j=1}^m w_{ij}^2 \leq 1$, where $i = 1 : n$.

As a result, the final form is rewriteable to $C_b(W)$, which has a b factor to obtain the optimal nearest correlation matrix, which $C_b(W)$ can be computed as (10).

$$C_b(W) = W_{b-1} \cdot W_{b-1}^T + \text{diag}(I - W_{b-1} \cdot W_{b-1}^T), \quad b \in \mathbb{N} \quad (10)$$

In its initial state, I is an identity matrix. We can express that $C_0 = \text{Cov}(W) = W_0 \cdot W_0^T$. The $\text{diag}(\cdot)$ function creates a diagonal-style structure. The matrix argument can be either a numeric square matrix or a vector. In this case, our matrix is a real-square symmetric correlation matrix. So, $\text{diag}(\cdot)$ establishes a column vector with its elements w_i equal to the corresponding diagonal elements $w_{i,i}$ of covariance. Therefore, the b factor is alterable to the range of a countable number \mathbb{N} , which the final b guarantees for approaching the statistic-gradient one.

$$L_{\text{Spectral-CORAL}} = \frac{1}{4d^2} \|C_b(w_s) - C_b(w_t)\|_F^2, \quad b \in \mathbb{N} \quad (11)$$

The final by-products of C_b for the spectral gradient correlation structures, computed from (10), are aligned with the Frobenius norm to shorten the domain discrepancy.

$$L = L_{\text{SoftTriple}} + \beta \cdot L_{\text{Spectral-CORAL}} \quad (12)$$

In summary, we define composite objective functions as follows in (12) for optimization under our considered problems. With the combination of SoftTriple loss from (3) and Spectral-CORAL loss from (11), where variables were assigned as the weight parameter β to appreciably adjust transfer loss. We have summarized the end-to-end processing of the scheme as clarified in Algorithm 1.

IV. EXPERIMENTS AND ANALYSIS

A. Experimentation Resources

We set up experiments on hardware utilized to train and verify the system. Our machine is powered by an Intel Core i5-10400 CPU, 32 GB of dynamic memory, and a single

Algorithm 1: The End-to-End Processing of STOS Framework

Require: A few-shot exemplary images in both the source and target domains (x_s, x_t) with the same and different classes; Mixup data-augmentation as auxiliary source x'_s according to Eq. (1); Concentration constant for image-level mix adjustment α ; The number of factors to optimize the nearest correlation structure b ; **k-etc** multiple centers including other hyper-parameters for determining SoftTriple constraints;

- 1 Compute latent features $f_{s(t)} = F(x_{s(t)})$
- 2 Compute high-concentrated attention via Multiple Attention Orchestration $F_o = M_{ORCAT}(BAM(f), SAM(f))$ according to Eq. (4)-(7)
- 3 Compute predicted high-confidence outputs through SoftTriplet descriptor $W_{s(t)} = C_{SoftTriple}(F_o^{s(t)}, k\text{-etc})$
- 4 Calculate source SoftTriplet loss $L_{SoftTriplet}$ according to Eq. (2)-(3)
- 5 **Initialize:** $C_0^{s(t)}$ from the standardized covariance $Cov(W_s)$ and $Cov(W_t)$
- 6 **Repeat:**
 - | Update $C_b^{s(t)}$ according to Eq. (10)
 - | $b = b + 1$;
- 6 **until:** Convergence to the spectral projected gradient correlation $C_b^{s(t)}$;
- 7 Calculate the spectral correlation alignment loss $L_{Spectral-CORAL}$ in Eq. (11)

Outcome: Composited objective function according to Eq. (12) for backward updating to model Φ

GeForce RTX 2060 Super graphics card with 8 GB of dedicated memory to use with the deep learning frameworks. PyTorch ver. 1.9.0, Torchvision ver. 0.10.0, NumPy ver. 1.18.1, the GPU-acceleration library (CU-DA) ver. 11.1, the deep neural network-acceleration library (cu-DNN) ver. 8.1, and kmeans_pytorch ver. 0.3 for the PCS approach.

B. Datasets

The first public dataset is Office-31 [52], which is composed of office supplies, has an abundance of 4,110 instances across 31 different classes. We conduct six transfer ways: A→W, A→D, W→A, W→D, D→A, and D→W. In addition, we use the Modern Office-31 [41] that has the recent Modern domain (S) with 100 images in a category. We have partly determined three adaptation ways: S→A, S→W, S→D. The second dataset is the ImageCLEF-DA 2014 domain adaptation [16]. Instances are randomly picked from a variety of public data, such as Caltech-256 (C), ILSVRC 2012 (I), Pascal VOC 2012 dataset (P), and Bing (B), gathered into four domains, each of which has 12 categories. This dataset is arranged in each category and whole domain for 150 and 600 images, respectively. We conduct twelve practical transfer tasks: C→I, C→P, C→B, I→C, I→P, I→B, P→C, P→I, P→B, B→C, B→I, and B→P.

In order to achieve a realistic implementation of virtual-to-physical adaptation, we generated a crystallization growth dataset. The crystallization has four main stages, the first of which is an unsaturated stage with no crystal nuclei. The metastable zone is marginally above the saturation point. There will be an oversupply of crystals being formed, leading to varying sizes. The labile zone creates numerous free nuclei for crystallization spontaneously from highly viscous solutions, but crystals will not fully mature. The intermediate zone is

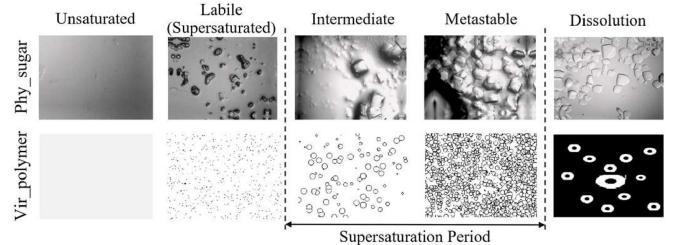


Fig. 7. Exemplary instances of the crystallization growth dataset.

a gap range, allowing solids to build spontaneous nuclei and progressively become bigger. We use growth simulations of new nuclei [60], which consider three positive stages from nucleation to full-grown crystals. In addition, the reverse dissolution simulation described in [61] and the unsaturated stage were also established. All virtual entities correspond to the actual sugar crystallization and dissolution [53], which contains microscopic illustrations of the crystalline sugar. We prepare data for two domains, each of which has five categories as shown in Fig. 7.

C. System Implementation Details

All verification is carried out in unsupervised adaptations according to the official protocol, with the source domain labeled while the target domain is completely unlabeled. We build up an exemplary pair by using scarce target exemplars and fully annotated source samples. Those are generated as numbers that are equal to the number of shots, viz., a few source data are turned incessantly for each iteration, and the target data uses a few instances in all trials. We used a varying-way five-shot with alternated two pairs, or five source/target exemplars from the same and different classes in the training phase while the evaluation phase uses a full range, for hyper-parameters, the learning rate at 1×10^{-3} , the weight decaying at 5×10^{-4} , and adopted SGD optimizer with a momentum of 0.9. For batch size, the batch size is the number of classes in each dataset. For the model preparation, the resolution images are minimized into 224x224 small images and transformed into the Mixup pre-processing as auxiliary features for the source data, in which the concentration-level parameter is set at 0.2 for every assessment. We decided on ResNet-18 for Office-31 and ResNet-50 for ImageCLEF-DA and event-realistic datasets. The model pre-processing utilized the ImageNet pre-trained parameters as the encoder initialization equipped in all visual benchmarks. The descriptor has mostly set up the SoftTriplet constraint as default, like most margins, but without their weight regularization term, which is provided in [45] since their term is used to minimize increasing multiple centers. We used $k = 1$, which is equivalent to the general classification, to test transferability under our constraints. We set $k = n$ (n is the best parameter for all runs), equivalent to a SoftTriplet descriptor with multiple n centers in a class. The width of the pre-last embedding layer is figured out by the size of the number of classes multiplied by the defined multiple centers.

We provide benchmarks in their performances with those of the state-of-the-art unsupervised adaptation performing

TABLE II
MEAN ACCURACY (%) OF FEW-SHOT (VARYING-WAY 5-SHOT) ADAPTATION ON IMAGECLEF-DA THROUGH FOUR INDEPENDENT PROCESSING BASED ON RESNET-50 (BEST IN BOLD)

Studies	C→I	C→P	C→B	I→C	I→P	I→B	P→C	P→I	P→B	B→C	B→I	B→P	Avg.
Few-Shot Unsupervised Adaptation Task													
Source only	88.83	75.00	63.83	94.00	79.67	64.33	94.17	90.83	64.17	94.00	89.33	75.17	81.11
DANN [54]	76.83	72.17	62.83	93.33	79.67	63.83	94.00	90.67	63.17	92.83	88.33	74.50	79.35
DAN [63]	85.83	75.00	60.67	93.50	77.33	61.83	93.33	88.64	60.17	65.50	63.50	29.83	71.26
D-CORAL [40]	89.50	76.33	63.50	93.83	79.67	64.67	94.50	91.17	63.17	94.67	89.33	77.50	81.49
Robust-CORAL [50]	88.50	75.00	62.17	93.50	79.67	64.33	93.33	90.83	63.17	94.50	88.33	75.33	80.72
High-order CORAL [50]	89.17	75.00	62.33	93.00	79.67	65.00	94.00	91.00	63.50	95.00	90.67	76.33	81.22
MDD [15]	88.83	74.17	63.17	93.83	79.67	64.50	94.50	91.17	63.83	94.00	89.17	76.67	81.13
MCC [26]	88.33	75.33	62.83	93.67	79.83	64.33	94.00	90.83	63.17	93.67	88.33	75.67	80.83
MME [42]	87.67	74.50	64.50	93.17	80.00	63.00	93.50	91.33	64.00	94.00	89.33	76.17	80.93
PCS [43]	88.00	74.67	63.00	93.50	80.83	63.83	94.33	91.33	63.83	94.17	88.83	76.00	81.03
Spectral-CORAL (Ours)	89.67	77.00	65.50	94.17	80.00	65.67	94.50	91.33	65.33	95.00	91.00	76.33	82.13
Replenishment													
SoftTriplet (best k) [45]	89.83	77.17	66.00	95.00	80.00	66.33	95.33	91.17	65.00	96.33	91.00	76.17	82.44
STOS (Ours)	93.33	80.67	69.00	95.83	82.83	69.67	96.00	94.33	68.67	97.33	94.33	80.83	85.24

in few-shot transfer tasks: MK-MMD [63], DANN [54], CORAL [40], MDD [15], MCC [26], MME [42], PCS [43], Robust-CORAL [50], High-order CORAL [50], [56]. Our approach and Robust-CORAL set the sixth step of the number of iterations to gain the highest performance in all evaluation metrics. We choose the best of two results since the higher third and fourth order dimensions are optimal order [50], [56]. The higher tensor dimension that must be identical on every facet, this requires a lot of computational memory. Thus, we use a ResNet-18 model for fairness of all tasks in Office-31. The weight is set to vary linearly over time in DANN to control domain adversarial loss directly. DDC, D-CORAL, Robust-CORAL, High-order CORAL, and our technique implement the optimal weight β to calculate domain loss. We discovered the use of those parameters for all tasks, each of which gains an optimization inefficiency at most. DANN and MDD use encoded latent features to account for the GRL block. DAN uses all the fc layers to compute the Gaussian kernel and to measure pairwise squared distances. PCS store the features for cluster calculations based on the number of classes while other parameters are set as in [43]. At the inferred outputs, MME is used in adversarial entropy. High-order CORAL is drawn by the high-order tensor dimension instead such that measures a distance similar to DAN, some outputs are calculated to covariance matrix and measured by D-CORAL, MCC, Robust-CORAL and our Spectral-CORAL. Our attention strategy is determined of 4096 to 2048 channels following the fourth block of ResNet-50 output that generates 2048 output channels, whereas dimensions in ResNet-18 with the reduction of 4 times. Our resource is also available on <https://github.com/pjirayu/STOS>.

D. Results and Benchmarks

Hereinafter, every measurement indicates the highest execution for each comparative task in the bold text over every table in Part D onward. To investigate ImageCLEF-DA following Table II, we discovered that our results are remarkable without dropping or disrupting knowledge even if domain regularization is only investigated at some practices, i.e., P→B and C→I are actually able to promote in both ways up to about 1.2% and 0.8%. Conversely, other backgrounds have

TABLE III
MEAN ACCURACY (%) OF FEW-SHOT (VARYING-WAY 5-SHOT)
ADAPTATION ON OFFICE-31 THROUGH THREE INDEPENDENT
PROCESSING BASED ON RESNET-18 (BEST IN BOLD)

Studies	A	A	W	W	D	D	Avg.
	↓ W	↓ D	↓ A	↓ D	↓ A	↓ W	
Few-Shot Unsupervised Adaptation Task							
Source only	70.38	74.57	56.37	90.70	58.76	91.43	73.70
DANN [54]	68.60	75.90	54.41	91.08	58.58	89.45	73.00
DAN [63]	71.45	74.38	57.20	91.84	59.00	92.86	74.46
D-CORAL [40]	70.81	75.14	56.03	90.89	58.58	92.24	73.95
Robust-CR [50]	70.54	75.33	57.15	89.75	58.76	91.04	73.76
High-CR [50]	70.50	74.95	56.42	91.08	58.76	92.36	74.01
MDD [15]	70.85	74.38	55.52	90.70	58.12	91.81	73.56
MCC [26]	70.53	73.81	57.29	90.51	58.58	91.94	73.84
MME [42]	70.75	74.38	56.43	90.32	57.73	91.94	73.59
PCS [43]	70.19	74.19	57.31	92.03	59.00	92.56	74.21
Spec-CR (Ours)	71.33	76.09	57.31	91.84	59.05	92.80	74.74
Replenishment							
ST ($k=5$) [45]	80.60	82.16	67.72	99.70	68.81	97.27	82.71
STOS (Ours)	84.70	88.91	70.65	99.81	71.72	98.64	85.74

uncertain results on domain improvement, such as High-order CORAL, which degrades by 0.67% in P→B and promotes a bit by 0.34% in C→I. At the average efficiency, it is obvious that the newly developed practice boost improvement from source-only up to 1.0% and overtake the second place on D-CORAL by 0.6%. Furthermore, with an aggregated scheme, our findings demonstrated greater adaptability than ST, the second-place transfer task, and source-only by 2.8%, 3.7%, and 4.1%, respectively.

Table III for Office-31, We explored that our system is not only improved by this new-adaptive operator, but also confidently outperforms without devaluating the weight presentation, like some approaches. As examined in W→A and A→D tasks, our method is advanced by 0.94% and 1.52%. On the contrary, some domain-adaptable styles have instability in practice. DAN suitably realize to adapt in the small latent fc layer in ResNet-18 [63], is promoted by 0.76%, but Robust-CORAL has a minor improvement of 0.06%. Besides, the recent PCS technique is a rather notable adjustment, but its performing is moderate. Our proposed method is superior to the second place over DAN and source-only at around 0.2% and 1.0%. Including, STOS surpasses SoftTriplet, the second place, and source-only by 3.0%, 11.2%, and 12.1%,

TABLE IV
MEAN ACCURACY (%) OF FEW-SHOT (VARYING-WAY 5-SHOT)
ADAPTATION ON MODERN OFFICE-31 THROUGH SYNTHETIC
ADAPTING TO PHYSICAL DATA IN THREE PERSPECTIVE
PROCESSING BASED ON RESNET-18 (BEST IN BOLD)

Studies	S→A	S→W	S→D	Avg.
Few-Shot Unsupervised Adaptation Task				
Source only	54.33	65.84	73.81	64.66
DANN [54]	51.41	64.42	69.45	61.76
DAN [63]	54.62	66.52	75.52	65.55
D-CORAL [40]	53.08	66.28	73.24	64.20
Robust-CORAL [50]	54.13	66.34	73.06	64.51
High-order CORAL [50]	54.01	65.86	73.43	64.43
MDD [15]	54.54	66.84	73.43	64.94
MCC [26]	53.40	66.92	73.24	64.52
MME [42]	53.15	66.23	74.19	64.52
PCS [43]	54.35	66.28	75.14	65.26
Spectral-CORAL (Ours)	55.07	66.60	75.52	65.73
Replenishment				
SoftTriplet ($k=5$) [45]	64.42	74.36	83.68	74.15
STOS (Ours)	67.58	76.84	86.34	76.92

TABLE V
MEAN ACCURACY (%) OF FEW-SHOT (VARYING-WAY 5-SHOT)
ADAPTATION ON CRYSTALLIZATION GROWTH THROUGH
Vir_Polymer (Vir) → Phy_Sugar (Phy) PROCESSING
BASED ON RESNET-50 (BEST IN BOLD)

Studies	Vir_polymer → Phy sugar	Studies	Vir_polymer → Phy sugar
Few-Shot Unsupervised Adaptation Task			
Source only	76.70 ± 0.1	MDD [15]	78.71 ± 3.0
DAN [63]	63.80 ± 20.9	MCC [26]	82.41 ± 1.3
D-CORAL [40]	87.21 ± 3.1	MME [42]	83.69 ± 1.8
Robust-CR [50]	71.33 ± 3.5	PCS [43]	84.54 ± 1.8
High-or CR [50]	79.49 ± 2.6	Spec-CR (Ours)	89.61 ± 0.4
Replenishment			
ST ($k=10$) [45]	82.44 ± 6.2	STOS (Ours)	95.38 ± 1.6

respectively. Moreover, the measures at Modern Office-31 are tabulated in Table IV. The proposed method surpasses the second place, similar as over DAN, and source-only by 0.2% and 1.1%, while system improvement outperforms vanilla ST and source-only by 2.8% and 12.3%.

Lastly, for practical performance in crystallization activity in Table V, our proposed Spectral-CORAL can eliminate the uncertainty in transfer practice and outperforms the second place via D-CORAL by 2.4% and source-only by 12.9%, whereas our STOS scheme outpaces Soft-Triplet, the second place, and source-only by 12.94%, 8.17%, and 18.68%, respectively. The comparative demonstrations manifest how cross-domain training formalizes traditional transfer practices. This is vulnerable to potentially alleviate domain differences, even though the outcomes are unsettled while playing on low-shot transfer practices. Our proposed scheme, which intensifies more concentration, aids in object sameness validation, and reduces the Gaussian discrepancy through re-patterning the statistical representation. These are confidently preferable with network improvement.

E. Ablation Study

Network Enrichment & Attention Strengthening: According to Table VI, we explored our high-optimal STOS scheme determined between a general ResNet block and the best number of centers, for instance [45] integrated with the ResNet

TABLE VI
ABLATION STUDIES (MEAN ACCURACY) ON OFFICE-31,
IMAGCLEF-DA, AND CRYSTALLIZATION GROWTH
DATASETS (BEST IN BOLD)

Background	A ↓ D	D ↓ W	W ↓ A	B ↓ C	C ↓ P	P ↓ I	Vir ↓ Phy
Network Enrichment (simple encoder & classifier)							
Vanilla ResNet	74.6	91.4	56.4	94.2	75.0	90.1	76.7
SoftTriplet-ResNet (best k)	82.4	97.3	67.7	96.2	77.2	91.2	82.4
Attention Strengthening							
STOS-SAM only	83.1	98.4	65.3	96.5	77.3	91.5	89.3
STOS-BAM only	85.0	96.8	66.6	97.0	77.7	91.7	90.0
STOS-ORCAT	88.9	98.7	70.7	97.3	80.7	94.3	95.4
Dwindling Domain Disparity							
STOS (Adversarial)	86.2	98.1	68.9	96.3	79.0	91.7	86.3
STOS (MK-MMD)	86.0	98.1	69.8	95.7	79.3	91.3	42.5
STOS (D-CORAL)	87.7	98.1	69.6	96.2	80.2	93.0	90.2
STOS (High-or CR)	87.3	97.8	69.6	96.0	79.8	92.5	92.0
STOS (Robust-CR)	87.7	97.6	70.0	95.7	79.2	93.0	57.5
STOS (MDD)	87.1	97.6	69.3	96.3	80.5	94.2	92.8
STOS (MCC)	86.5	98.1	70.4	97.0	80.5	93.5	90.4
STOS (MME)	86.9	97.5	69.1	96.8	80.7	94.0	88.6
STOS (PCS)	86.3	97.9	69.2	96.2	80.0	94.2	92.6
STOS (Spectral-CR)	88.9	98.6	70.7	97.3	80.7	94.3	95.4

TABLE VII
MEAN ACCURACY OF TRANSFER TASKS w.r.t. k MULTIPLE
CENTERS IN A CLASS

k	1	2	3	4	5	10	15
A→W	80.26	82.13	83.24	83.67	84.70	83.15	82.81
C→P	93.17	93.50	94.00	94.00	94.33	93.67	93.17
Vir→Phy	82.15	91.47	87.03	91.47	88.03	95.38	88.75

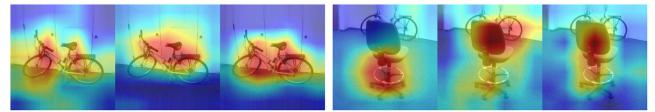


Fig. 8. The attention visualization with two perspectives refined by SAM, BAM, and the proposed Orchestrated Attention in order from left to right.

family, as well as the importance of determining the number of centers to raise the highest efficacy in accuracy. As demonstrated in Table VII, those were assessed at the number ranging from two to fifteen points for either class center in the A→W, C→P, and Vir_polymer→Phy_sugar transfer practices in order to appraise varying scheme capabilities for each configuration. Those findings are highly satisfactory at five multiple centers for Office-31 and Image-CLEF-DA datasets and up to ten points for the crystallization dataset. Further, to accelerate spatial awareness and help before reducing domain dislocation, we have examined whether the identity of our attention orchestration is comparably better than the existing improvement rooms. Hence, this promotes greater explicit attentiveness, for instance, up to 4% for A→D and 2.5% for P→I, than using either one of the attention-based blocks. Including, three attention mechanisms are illustrated to two extents in Fig. 8: single- and multi-class instances. BAM illustrations focus on an out-of-the-box region and another class instance field. SAM is defined in terms of increasing weight μ in order to identify the simpler tasks first and then incrementally raise the complicated tasks; however, when faced with a real multi-class vision, it still creates a too narrow attention bound for a focused task

TABLE VIII
MEAN ACCURACY OF TRANSFER TASKS w.r.t. β CONTROL
WEIGHT OF ALIGNMENT LOSS

β	0	1	1×10^1	1×10^2	1×10^3	1×10^4
D→A	69.12	68.76	69.46	70.61	71.72	68.90

since two existing methods are insufficiently comprehensible in the feature outputs in comparison to the quantity of target data on hand, even though our orchestrated attention is capable of greatly concentrating on the boundaries of specific entities.

Adaptation Measurement: We investigate the impact of different adaptation efforts for tackling domain dislocation, such as an adversarial loss, MK-MMD, and High-order CORAL-based kernel matching losses, Robust-CORAL-based statistical metrics, and others. Our Spectral-CORAL alignment loss is confronted with adversarial-based techniques such as DANN [54] and MDD [15] are the most popular approaches in deep adaptation task. These are trained under an end-to-end adversarial training loss. We also examine approaches for MME [42] and PCS [43]. Most of them are recent studies, which are more appropriate for few-shot unsupervised adaptation practice. Many methods are presented to work around the low-level domain-specific representations. MK-MMD [63] and High-order CORAL [50] are respectively presented as the two-sample tests of low-order and high-order moments in the RKHS, which are performed in the house of kernels. The first two original matrices and the reformatted matrices for the two-order statistical measures are CORAL [40], MCC [26], and Robust-CORAL [50]. With the exception of MCC, a domain alignment loss is shortened by the distinctness in the Frobenius norm. All cost functions are computed under the influence of the ResNet encoder, SoftTriplet, and orchestrated attention. In seven exemplary adaptation ways, the majority of their findings are still inadequate even when running on our module provided, whereas MME is as adaptable as to our method in C→P. Nevertheless, all transfer tasks mostly have minor improvement as indicated in D→W task because of instance differences in resolution only. Yet overall, our full preparation is superior to many transfer tasks that can reach about 1-4 percent. Table VIII also conducts that the weight β adjustment actually improves adaptability from the SoftTriplet descriptor.

F. Analysis & Discussion

Transferability: To analyze the factor optimization of the matrix transformation in accordance with iterative matrix inversion and spectral gradient structure in factorization at A→D, I→C, and Vir_polymer→Phy_sugar tasks were performed by Fig. 9. Our scheme is capable of transferring, namely, a greater improvement starting at the second to the sixth factor than an existing study in A→D run. In I→C execution, we outclass at the first factor using Spectral-CORAL-based statistical trick that can achieve the highest since the eighth factor without raising too many orders. To also ascertain the consistency of efficiency in crystal growth, Robust-CORAL would rather have instability in transfer since

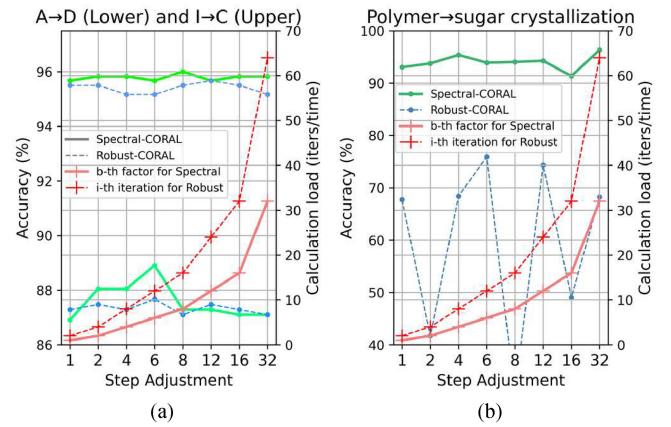


Fig. 9. The adaptability efficiency and calculation burden are obtained by considered STOS-based Robust- and Spectral-CORAL in transfer tasks of (a) A→D, (b) I→C, and (c) Vir_polymer→Phy_sugar under factor adjustment.

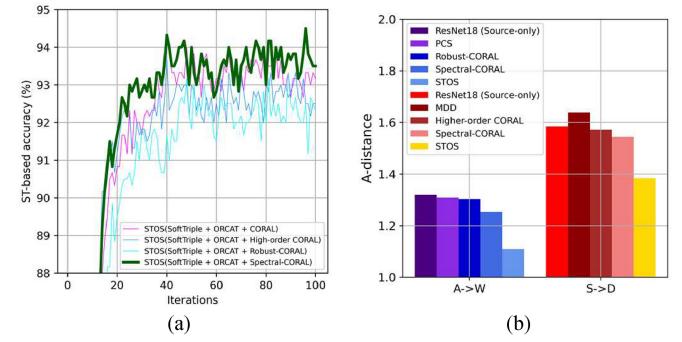


Fig. 10. (a) Performances in ImageCLEF-DA through the fully proposed scheme regularizing with many domain alignment adjustments over P→I transfer task and (b) A-distance with A→W and S→D transfer tasks.

second-order iterations, leading to performance fluctuations, but ours is highly stable every run. For domain gaps measured by A -distance following Fig. 10(b), studies appear to have quite distances between distributions. Spectral-CORAL is more adaptable at Gaussian differences, so the new attention has a greater influence in helping to reduce domain gaps to the lowest level in measures.

Improvability: In the convergence metrics of both exemplary executions in Fig. 10(a), our bench tests can stay on top of convergences when compared to the latest statistical extents and some related methods. Most of them are unstable, unadaptable, till to the point of deterioration. To study further in the generalization and examine the intrinsic results of our aggregated scheme without pre-processing images, we draw the accuracies and losses during the training process in Fig. 11. Since ST-based loss is only understandable in terms of object identicalness with a view to achieving categorization, they are inability to comprehend domain-individual parameters to indicate enough. In Fig. 11(a), source and target gains are at 80% and only 60% in one-third of the training amounts. Whether the domain alignment applies to an ST-based network or not, this is still effective with a single regularization of

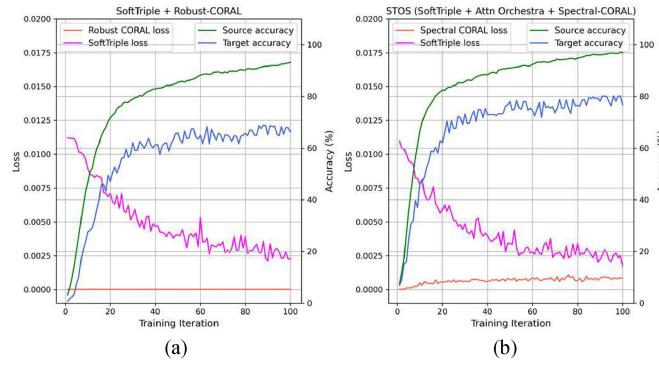


Fig. 11. The convergence of the intrinsic performance without any auxiliary data. (a) ST-based Robust-CORAL and (b) the proposed STOS scheme of A→D transfer task on Office-31.

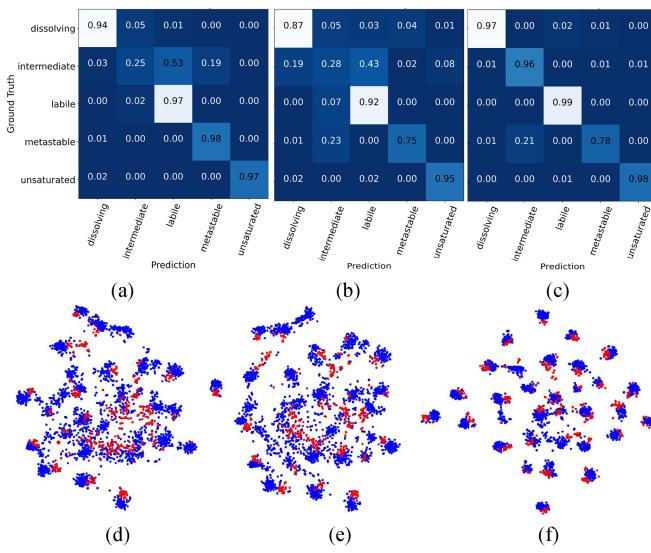


Fig. 12. Confusion matrix w.r.t. Vir_polymer→Phy_sugar transfer task, (a) CORAL only (88.39%), (b) ST-based High-order CORAL (82.01%), (c) STOS (96.27%), and (d) t-SNE of embeddings of (d) PCS, (e) MDD, and (f) STOS on A→W transfer task. Note that the blue and red dots are denoted as the source and target instances and the diagonal color for confusion matrices are not all white shade because a dataset is rather imbalanced, comparing in each category.

classification. Our approach yielded a marginal improvement of 87% in source, though it has been shown to not converge speedily, but it does not decrease processing knowledge that is able to reach target recognition at 74.0% within the similar one-third iteration amounts according to Fig. 11(b). Even the Spectral-CORAL loss starts from scratch and rises consistently to deal with the falling classification cost in order to increase adaptability and attentiveness to realize domain-object instances continually.

Confusion Matrix: Aside from comparing obvious concrete realizations, our method is noticeable in terms of competence. We validate for precise prediction on Vir_polymer→Phy_sugar transfer tasks as depicted in Fig. 12. Because the intermediate phase contains some large-scale crystals that have already built up in the crystal-growing process, while all crystals have grown to their greatest size in the final metastable, that is why some solids are similar in size between these two processes. Our approach is superior to conventional

TABLE IX
THE ASSESSMENTS FROM VARIOUS STUDIES IN RESPECT OF THE IMAGE RECOGNITION OF CRYSTALLIZATION AND/OR PARTICLE SEGMENTATION

Ref	Particle Type	Module and/or Contribution	Technique	Training			Infer. Perf.
				C	Data per iteration	T (hr)	Mean Acc.
[38]	Sugar crystalline image	Incep-ResNet V2	Fine-tuning TL (Sup'd)	4	800	9	90.1
[13]				5			
[23]	Sugar crystalline image (Seg)			2	1,500	10	93.3
[20]	Oil-in-water particle image	AlexNet	Single-domain multi-scale CNN (Sup'd)	3	270	0.5	91.6
[6]	Protein DNA sequence	Deep-Crystal (Plain-3CNNs)		2	23k	2.7	82.8
[10]		ATTCry (Custom -6ATT)		2	23k	0.3	86.6
Ours	Sugar crystalline image	ResNet-50 + STOS	Few-shot DA (Unsup'd)	5	≤ 100 (5-10img/C)	0.8	94.6

methods by being results that are close to the whole process.

Latent Feature Visualization: To know the domain classification from encoder activation of A→W task can be illustrated by t-SNE in Fig. 12(d). We can observe several post-adaptive measures. Most target instances, which are attempting to reach into the source clusters, have a low likelihood on account of the insufficient adaptive transfer from existing techniques. Compromise under new statistics and attention styles can make more clearly to instance boundaries, by which many target instances are allocated to each corresponding source entity.

Toward Consumer Technology: This phenomenon has led consumers that embraced the technological advancement that has the potential to reduce data adjustment in recognition, which can be contained without the need for strong hardware as instantiated by Section IV-A. Once the compression technology is in place, it can insert intelligent things into the edge, making it easier to meet consumer needs. Although the limitation to the responses and capacities of edge appliances is technically not confirmed, our findings across all benchmarks in Section IV are indeed performed to the highest predictability and durability, which is enough to be applied to the accessible device market.

G. Scenario: Syrup Crystallizer in Sugar Production

In spite of the fact that there is not much proposed for working on sugar growth with the use of modern computing as seen in Table IX. We therefore compared it to many similar tasks in particle recognition and related areas. All existing methods rely on full supervision and a huge data appetite for classification to testify to their greatest feasibility. Regardless of the precise measurement of other particle-based applications, our approach is apparently equivalent to the end-to-end calculation rapidity, with fewer data consumption than the current studies.

Besides, the improbability of the transfer task, the practice of fine-tuning transfer learning typically requires a well-trained

encoder. Those method still take up to half a day to calibrate the parameters to the optimal weights. Some studies were able to reduce the training time, but the performance has declined. With our prepared system that takes no upwards of an hour, provides an outstanding 94.6% better assessment, and can cut down on data usage. These make our approach uppermost for threefold aspects, until the system is prone to degeneration. The model is retrainable and can be replaced in a rush.

V. CONCLUSION

In this article, our considered the issue to diminish domain shift due to the burdensome bottleneck of the target protocol that led to an adaptation into unsteadiness and worsening, which should be alleviated by beneficial knowledge transferability. To end issues, the framework is made up of attention-wise orchestration that aids in more similarity validation and statistic-regenerated alignment. These are refined not only to eliminate the knowledge degeneration, but also to expedite the understanding of domain-object disparate entities. Our outcomes achieve cutting-edge performance from extensive experiments across public datasets and a realistic application. Witness that the novelty of our prepared technique is certainly adaptable in a few-shot unsupervised adaptation, into which the backing of the very similar feature element-wise perception is greatly strengthened. For upcoming work, we intend to enable the realization to adapt from a fully virtual world energized to the real-physical perception leading to the digital twin-based domain-adaptive learning. A new proposal is preferred to our prospective challenge.

REFERENCES

- [1] F. Basile, P. Chiacchio, and D. Gerbasio, "On the implementation of industrial automation systems based on PLC," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 4, pp. 990–1003, Oct. 2013.
- [2] W. Booyse, D. N. Wilke, and S. Heyns, "Deep digital twins for detection, diagnostics and prognostics," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106612.
- [3] C. Gehrmann and M. Gunnarsson, "A digital twin based industrial automation and control system security architecture," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 669–680, Jan. 2020.
- [4] K. Y. H. Lim, P. Zheng, and C.-H. Chen, "A state-of-the-art survey of digital twin: Techniques, engineering product lifecycle management and business innovation perspectives," *J. Intell. Manuf.*, vol. 31, no. 6, pp. 1313–1337, 2020.
- [5] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.
- [6] A. Elbasir et al., "DeepCrystal: A deep learning framework for sequence-based protein crystallization prediction," *J. Bioinformat.*, vol. 35, no. 13, pp. 2216–2225, 2019.
- [7] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.
- [8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.
- [10] C. Jin, J. Gao, Z. Shi, and H. Zhang, "ATTCry: Attention-based neural network model for protein crystallization prediction," *Neurocomputing*, vol. 463, pp. 265–274, Nov. 2021.
- [11] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, "A digital-twin-assisted fault diagnosis using deep transfer learning," *IEEE Access*, vol. 7, pp. 19990–19999, 2019.
- [12] A. Fong and M. Usman, "Machine learning for end consumers," *IEEE Consum. Electron. Mag.*, vol. 9, no. 5, pp. 77–78, Sep. 2020.
- [13] S. Chayatummagoon and P. Chongstitvattana, "Image classification of sugar crystal with deep learning," in *Proc. 13th Int. Conf. Knowl. Smart Technol. (KST)*, 2021, pp. 118–122.
- [14] F. Tao and M. Zhang, "Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing," *IEEE Access*, vol. 5, pp. 20418–20427, 2017.
- [15] Y. Zhang, T. Liu, M. Long, and M. I. Jordan, "Bridging theory and algorithms for domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1–10.
- [16] J. Yang, H. Zou, S. Cao, Z. Chen, and L. Xie, "MobileDA: Toward edge-domain adaptation," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6909–6918, Aug. 2020.
- [17] A. K. Tanwani, N. Mor, J. Kubiatowicz, J. E. Gonzalez, and K. Goldberg, "A fog robotics approach to deep robot learning: Application to object recognition and grasp planning in surface decluttering," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, 2019, pp. 1–8.
- [18] C. Jin, Z. Shi, C. Kang, K. Lin, and H. Zhang, "TLCrys: Transfer learning based method for protein crystallization prediction," *Int. J. Mol. Sci.*, vol. 23, no. 2, p. 972, 2022.
- [19] J. Li, M. Jing, H. Su, K. Lu, L. Zhu, and H. T. Shen, "Faster domain adaptation networks," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5770–5783, Dec. 2022.
- [20] J. Jeevanandam, D. Agyei, M. K. Danquah, and C. Udenigwe, "Food quality monitoring through bioinformatics and big data," in *Future Foods*. London, U.K.: Academic, 2022, pp. 733–744.
- [21] Y. Fang, C. Peng, P. Lou, Z. Zhou, J. Hu, and J. Yan, "Digital-twin-based job shop scheduling toward smart manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6425–6435, Dec. 2019.
- [22] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.
- [23] Z. Hayali and A. Gholamreza, "Transfer learning on semantic segmentation for sugar crystal analysis," in *Proc. Int. Conf. Machine Vision Image Process. (MVIP)*, 2022, pp. 1–6.
- [24] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.
- [25] S. Lim, M. Shin, and J. Paik, "Point cloud generation using deep adversarial local features for augmented and mixed reality contents," *IEEE Trans. Consum. Electron.*, vol. 68, no. 1, pp. 69–76, Feb. 2022.
- [26] Y. Jin, X. Wang, M. Long, and J. Wang, "Minimum class confusion for versatile domain adaptation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 1–17.
- [27] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, "Deep model-based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017.
- [28] S. Sundaram and A. Zeid, "Smart prognostics and health management (SPHM) in smart manufacturing: An interoperable framework," *Sensors*, vol. 21, no. 18, p. 5994, 2021.
- [29] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.
- [30] J. Li, R. Huang, G. He, S. Wang, G. Li, and W. Li, "A deep adversarial transfer learning network for machinery emerging fault detection," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8413–8422, Aug. 2020.
- [31] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019.
- [32] H.-G. Kim and G.-Y. Kim, "Deep neural network-based indoor emergency awareness using contextual information from sound, human activity, and indoor position on mobile device," *IEEE Trans. Consum. Electron.*, vol. 66, no. 4, pp. 271–278, Nov. 2020.
- [33] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "BAM: Bottleneck attention module," 2018, *arXiv:1807.06514*.
- [34] B. Bera, A. K. Das, M. S. Obaidat, P. Vijayakumar, K.-F. Hsiao, and Y. Park, "AI-enabled blockchain-based access control for malicious attacks detection and mitigation in IoE," *IEEE Consum. Electron. Mag.*, vol. 10, no. 5, pp. 82–92, Sep. 2021.
- [35] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "EfficientFi: Towards large-scale lightweight WiFi sensing via CSI compression," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13086–13095, Aug. 2022.

- [36] K. Alexopoulos, N. Nikolakis, and G. Chryssolouris, "Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing," *Int. J. Comput. Integr.*, vol. 33, no. 5, pp. 429–439, 2020.
- [37] S. Motljan, Q. Jones, S. M. Iranmanesh, and G. Doretto, "Few-shot adversarial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [38] J. Zhang et al., "Monitoring sugar crystallization with deep neural networks," *J. Food Eng.*, vol. 280, Sep. 2020, Art. no. 109965.
- [39] Z. Liang, Y. Gu, and J. Yang, "Hardmix: A regularization method to mitigate the large shift in few-shot domain adaptation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2021, pp. 454–458.
- [40] B. Sun and S. Kate, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 443–450.
- [41] T. Ringwald and R. Stiefelhagen, "Adaptiope: A modern benchmark for unsupervised domain adaptation," in *Proc. IEEE Winter Conf. Appl. Comput. (WACV)*, 2021, pp. 1–10.
- [42] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised domain adaptation via minimax entropy," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 8050–8058.
- [43] X. Yue et al., "Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1–16.
- [44] J. Gao et al., "Omni SCADA intrusion detection using deep learning algorithms," *IEEE Internet Things J.*, vol. 8, no. 2, pp. 951–961, Jan. 2021.
- [45] Q. Qian, L. Shang, B. Sun, J. Hu, H. Li, and R. Jin, "SoftTriple loss: Deep metric learning without triplet sampling," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1–10.
- [46] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–13.
- [47] C. Zhang, Q. Zhao, and Y. Wang, "Transferable attention networks for adversarial domain adaptation," *Inf. Sci.*, vol. 539, pp. 422–433, Oct. 2020.
- [48] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1–10.
- [49] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
- [50] Z. Cheng et al., "Robust and high-order correlation alignment for unsupervised domain adaptation," *Neural. Comput. Appl.*, vol. 33, no. 12, pp. 6891–6903, 2021.
- [51] R. Borsdorff, N. J. Higham, and M. Raydan, "Computing a nearest correlation matrix with factor structure," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 5, pp. 2603–2622, 2010.
- [52] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 213–226.
- [53] I. Usama, "Crystal-microscopics-images-and-annotations." Kaggle. 2021. [Online]. Available: <https://www.kaggle.com/datasets/usamaimdad/crystalmicroscopicsimagesandannotations>
- [54] Y. Ganin et al., "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.
- [55] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*.
- [56] C. Chen et al., "HoMM: Higher-order moment matching for unsupervised domain adaptation," in *Proc. Conf. AAAI Artif. Intell.*, vol. 34, 2020, pp. 3422–3429.
- [57] S. Yun et al., "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 6023–6032.
- [58] X. Li, H. Xiong, H. An, C. Xu, and D. Dou, "XMixup: Efficient transfer learning with auxiliary samples by cross-domain mixup," 2020, *arXiv:2007.10252*.
- [59] Z. Zhang and D. Li, "Hybrid cross deep network for domain adaptation and energy saving in visual Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6026–6033, Aug. 2019.
- [60] Z. J. Liu, J. Ouyang, W. Zhou, and X. D. Wang, "Numerical simulation of the polymer crystallization during cooling stage by using level set method," *Comput. Mater. Sci.*, vol. 97, pp. 245–253, Feb. 2015.
- [61] S. Yang, N. Ukrainczyk, A. Caggiano, and E. Koenders, "Numerical phase-field model validation for dissolution of minerals," *Appl. Sci.*, vol. 11, no. 6, p. 2464, 2021.
- [62] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiseNET: Bilateral segmentation network for real-time semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 334–349.
- [63] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 97–105.



Jirayu Petcharan received the B.E. (first class Hons.) degree in control engineering and the M.E. degree in instrumentation engineering from the School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with the College of Electrical Engineering and Computer Science, National Taiwan University of Science and Technology.

He was a System Integration Engineer with Contrologic Company Ltd., Thailand, for two years. His research interests include deep transfer learning, machine learning, and Internet of Things.



Shun-Feng Su (Fellow, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei City, Taiwan, in 1983, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1991, respectively.

He is currently the Chair Professor with the Department of Electrical Engineering, National Taiwan University of Science and Technology. He has published more than 300 refereed journal and conference papers in the areas of robotics, intelligent control, fuzzy systems, neural networks, and non-derivative optimization. His current research interests include computational intelligence, machine learning, virtual reality, intelligent transportation systems, smart home, robotics, and intelligent control.

Dr. Su is very active in various international/domestic professional societies. He also serves as a board member of various academic societies. He also acted as the General Chair, the Program Chair, or various positions for many international and domestic conferences. He currently serves as an Associate Editor for *IEEE TRANSACTIONS ON CYBERNETICS* and *Information Science*, a Senior Editor and an Associate Editor of *IEEE ACCESS*, the Executive Editor of the *Journal of the Chinese Institute of Engineers*, and an Area Editor and an Associate Editor of *International Journal of Fuzzy Systems*. He is currently the IEEE SMC Society Vice President for Publication and a member of board of government in the same society and also a Fellow of IFSA, CACS, and RST.