Unsupervised Domain Adaptation Using Full-Feature Whitening and Colouring

Subhankar Roy[1,2], Aliaksandr Siarohin[1(✉)], and Nicu Sebe[1]

[1] Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
{subhankar.roy,aliaksandr.siarohin,niculae.sebe}@unitn.it
[2] Fondazione Bruno Kessler (FBK), Trento, Italy

Abstract. It is a very well known fact in computer vision that classifiers trained on source datasets do not perform well when tested on other datasets acquired under different conditions. To this end, Unsupervised Domain adaptation (UDA) methods address the shift between the source and target domain by adapting the classifier to work well in the target domain despite having no access to the target labels. A handful of UDA methods bridge domain shift by aligning the source and target feature distributions through embedded domain alignment layers that are based on batch normalization (BN) or grouped whitening. Contrarily, in this work we propose to align feature distributions with domain specific full-feature whitening and domain agnostic colouring transforms, abbreviated as F^2WCT. The proposed F^2WCT optimally aligns the feature distributions by ensuring that the source and target features have identical covariance matrices. Our claim is also substantiated by the experimental results on Digits datasets for both single source and multi source unsupervised adaptation settings.

Keywords: Feature whitening · Colouring · Unsupervised Domain Adaptation · Multi source domain adaptation

1 Introduction

In the recent years deep learning has been exceptionally successful in supervised object recognition tasks [1,2]. Despite its effectiveness in *supervised* regime, object recognition in *unsupervised* regime is still an open ended problem because the lack of labels makes the training complicated. Off-the-shelf networks pretrained on some domain do not work well when transferred to a novel but related domain due to a problem called *domain-shift* [3]. To mitigate domain shift among datasets numerous Unsupervised Domain Adaptation (UDA) methods [4–10] have been proposed which leverage *unlabeled* target data together with *labeled* source data to learn a predictor for the target samples.

UDA methods can be roughly categorized under two broad categories. The first category includes Generative Adversarial Network (GAN) based methods [10–12] that learn a cross-domain mapping to emulate *target*-like source

E. Ricci et al. (Eds.): ICIAP 2019, LNCS 11752, pp. 225–236, 2019.
https://doi.org/10.1007/978-3-030-30645-8_21

images, which are then leveraged for training a target classifier. The second category of methods aims to reduce the discrepancy between source and target domains by leveraging the first order statistics [13,14] or second order statistics [15,25]. Some of the methods from this category achieve alignment of feature distributions by directly embedding batch normalization (BN) based [9,16,17] *domain alignment* (DA) layers into the network.

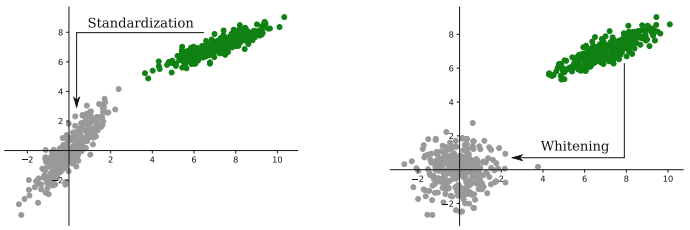
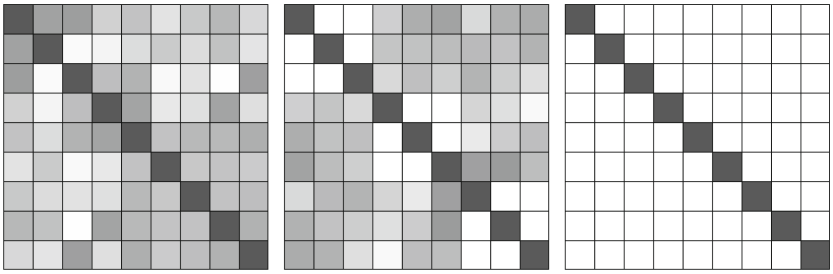


Fig. 1. Visualization of 2D features with different normalization transformations: (a) Feature standardisation; and (b) Feature whitening.

While BN based methods align feature distributions by setting variance of features to 1 and mean to 0, yet they leave the feature correlations intact (see Fig. 1a), leading to sub-optimal alignment. Conversely, we argue that to completely eliminate discrepancy between domains the source and target features should have the same covariance matrix. This can be ensured by projecting the feature distributions onto a canonical unit hyper-sphere through *full-feature whitening* (see Fig. 1b), such that both source and target domain features have identity covariance matrix. While Roy *et al.* [8] proposed to align feature distributions with domain-specific *grouped*-feature whitening (DWT), it suffers from imperfect alignment due to partial feature whitening (see Sect. 2).

To overcome the drawbacks of previous DA layers we propose to first *whiten* the feature representations and then apply *colouring*. Our *whiten* operation use domain specific whitening, while *colouring* operation is domain agnostic and is used to re-project the whitened features to a distribution having an arbitrary covariance matrix. Inspired by [18], we realize these transformations through **F**ull-**F**eature **W**hitening and **C**olouring **T**ransform (**F^{2}WCT**) blocks, embedded inside the network, replacing the BN-based and DWT-based DA layers. However, different from [18], which uses these operations for conditional image generation, we propose this technique for UDA. We also extend this to multi-source unsupervised DA (MSDA) setting where multiple source domains are available during training. Finally, we evaluate our proposed method on the *digits* datasets for both single source UDA and MSDA settings and set new state-of-the-art results.

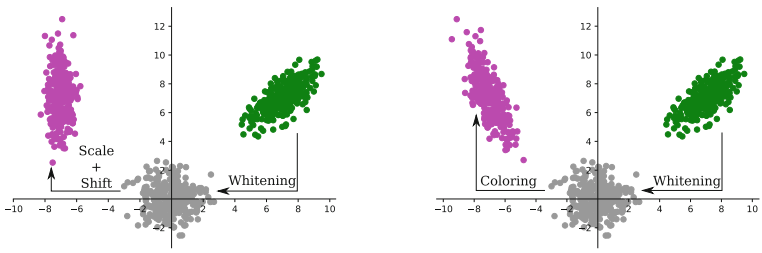


(a) (b) (c)

Fig. 2. Covariance matrices of features undergoing different normalization transformations: (a) BN [9]; (b) DWT [8]; and (c) Full-Feature Whitening. Black pixels denote value 1, white pixels denote value 0 and gray denotes intermediate values.

2 Related Works

Single Source UDA. Several UDA methods have been proposed in the recent years that operate under the assumption that there is only a single source domain. A multitude of UDA methods have utilized GAN [10–12] to learn a mapping between the source and target domains in order to generate synthetic data in the target domain. SBADA-GAN [10] and CyCADA [11] are trained with adversarial and cycle-consistent losses to generate labeled target-like source samples which are used for training a classifier for the target domain. Although very effective, GAN based methods require large amount of data from each domain to capture the inherent data distributions.



(a) (b)

Fig. 3. Visualization of 2D features after whitening and different feature re-projection techniques: (a) whitening with scale and shift as in DWT [8]; and (b) proposed whitening with colouring for aligning feature distributions. (Color figure online)

Another genre of UDA methods aim to reduce the discrepancy between source and target domains by leveraging the first and second order statistics. Minimum

Mean Discrepancy based methods [13,14] minimize the discrepancy between domains by minimizing the difference of the mean (i.e., first order statistics) of their respective feature representations. Correlation alignment methods [5, 15,25] leverage second order statistics by minimizing the loss derived from the covariance matrices of source and target feature representations. Carlucci *et al.* [9] and Roy *et al.* [8] showed that discrepancy between domains can be reduced efficiently by directly embedding BN-based and DWT based DA layers into the network, respectively. Albeit effective, BN-based and DWT based DA layers result in features which are correlated and therefore imperfectly aligned. As can be observed from the covariance matrices in Fig. 2a and b, the variance of the features are 1 but the features are still correlated due to non-zero off-diagonal elements. Ideally, we would like to have *identity* covariance matrix, $\Sigma = I$, (see Fig. 2c) to achieve complete alignment of features. This is achieved with our proposed F^{2}WCT. Moreover, DWT [8] re-projects partially whitened features with scale and shift transforms of [9] which is sub-optimal because it reduces the capacity of the network [18]. Hence, we propose to re-project whitened features with colouring operation as shown in Fig. 3b. Different from scale and shift operation of DWT (see Fig. 3a), which can only have axis-aligned re-projection of features, our colouring operation can re-project the whitened features to any arbitrary orientation and the network is flexible to choose through training.

Multi Source UDA. In practical scenarios source data can possess different underlying marginal distributions and therefore multiple domain shifts need to be addressed coherently while adapting to the target domain. MSDA was first addressed in [19] which showed the necessity to borrow knowledge from nearest source domains to avoid *negative transfer*. Xu *et al.* [20] adapted to the *distribution-weighted* combining rule in [21] with an adversarial framework. More recently, Peng *et al.* [22] proposed a Moment Matching Network for reducing domain shift from multiple sources to the target domain. Departing from the above methods, we easily extend F^{2}WCT to simultaneously align feature distributions of multiple source and target domains to a reference distribution.

3 Method

In this section we present our proposed method for UDA and MSDA. Specifically, first we will discuss some preliminaries and then introduce the proposed F^{2}WCT.

3.1 Preliminaries

Let us assume that $\mathcal{S} = \{(I_j^s, y_j^s)\}_{j=1}^{N_s}$ be the labeled source dataset, where I_j^s is the j^{th} source image and $y_j^s \in \mathcal{Y} = \{1, 2 \ldots, C\}$ be its associated label. Also, let $\mathcal{T} = \{I_i^t\}_{i=1}^{N_t}$ be the unlabeled target dataset where I_i^t is the i^{th} target image *without* any associated label. The aim of UDA is to train a target domain predictor by jointly utilizing samples from $\mathcal{S}$ and $\mathcal{T}$.

A fairly common technique to bridge domain shift is to use DA layers, which can either be BN-based [9] or DWT-based [8], that project source and target feature distributions onto a canonical distribution through per feature standardisation and grouped feature whitening, respectively. As mentioned in Sect. 1, we propose to replace these feature alignment techniques with domain specific full-feature whitening and domain agnostic colouring. Before introducing the proposed F^{2}WCT we will briefly recap BN [23] below.

A BN layer takes as input a mini-batch $B = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ of m samples, where $\mathbf{x}_i$ is the i^{th} element in the batch B and $\mathbf{x}_i \in \mathbb{R}^d$. As the name suggests, given a batch B the BN layer transforms each $\mathbf{x}_i \in B$ in the following way:

$$BN(x_{i,k}) = \gamma_k \frac{x_{i,k} - \mu_{B,k}}{\sqrt{\sigma^2_{B,k} + \epsilon}} + \beta_k, \tag{1}$$

where k ($1 \leq k \leq d$) signifies the k-th dimension of input data, $\mu_{B,k}$ and $\sigma_{B,k}$ are, respectively, the mean and the standard deviation corresponding to the k-th dimension of the samples in B and ϵ is used to prevent division by zero. Finally, γ_k and β_k are learnable scaling and shifting parameters. In essence, BN transforms a batch of features into having zero mean and unit variance and then re-projects the features with γ and β.

In Sect. 3.2 we present our proposed F^{2}WCT for UDA, while in Sect. 3.3 we extend the proposed F^{2}WCT for MSDA.

3.2 Full-Feature Whitening and Colouring Transform for UDA

As stated in Sect. 2 that BN based per-dimension feature *standardization* and DWT based *grouped feature whitening* is sub-optimal for marginal source and target distribution alignment due to the presence of correlated features. To alleviate domain shift we argue to replace BN and DWT with F^{2}WCT, derived from [18], and is defined as follows:

$$F^2WCT(\mathbf{x}_i; \Omega) = \mathbf{\Gamma}\hat{\mathbf{x}}_i + \boldsymbol{\beta}, \tag{2}$$

$$\hat{\mathbf{x}}_i = W_B(\mathbf{x}_i - \boldsymbol{\mu}_B). \tag{3}$$

In Eq. (3), $\boldsymbol{\mu}_B$ is the mean of B while W_B is the whitening matrix such that: $W_B^\top W_B = \Sigma_B^{-1}$, where Σ_B is the covariance matrix derived from B. $\Omega = (\boldsymbol{\mu}_B, \Sigma_B)$ indicates the batch-specific first and second-order statistics. Equation (3) performs the *whitening* of $\mathbf{x}_i \in B$ and the resulting elements of $\hat{B} = \{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_m\}$ lie in a hyper-spherical distribution, i.e., with a covariance matrix equal to the identity matrix (see Fig. 2c). Additionally, and differently from [8], in Eq. (2), with the help of learnable d dimensional vector $\boldsymbol{\beta}$ and $d \times d$ dimensional matrix $\mathbf{\Gamma}$ the *whitened* $\hat{B}$ is projected back to a multivariate Gaussian distribution having an arbitrary covariance matrix through the *colouring* operation. Implementation wise Eq. (2) can be realized with a convolutional layer having kernel size 1×1.

Our network, at any intermediate layer, takes as input two batches of input samples, $B^s = \{\mathbf{x}_1^s, \ldots, \mathbf{x}_m^s\}$ and $B^t = \{\mathbf{x}_1^t, \ldots, \mathbf{x}_m^t\}$ from the source and target domain, respectively. Every $\mathbf{x}_i^s \in B^s$ and $\mathbf{x}_i^t \in B^t$ is transformed through the F^{2}WCT block, where the whitening operation is domain specific but the colouring operation is domain agnostic. In details, using Eqs. (2)–(3) the output of F^{2}WCT blocks for the source and target samples are given respectively by:

$$F^2WCT(\mathbf{x}_i^s; \Omega^s) = \mathbf{\Gamma} W_{B^s}(\mathbf{x}_i^s - \boldsymbol{\mu}_{B^s}) + \boldsymbol{\beta}, \tag{4}$$

$$F^2WCT(\mathbf{x}_i^t; \Omega^t) = \mathbf{\Gamma} W_{B^t}(\mathbf{x}_i^t - \boldsymbol{\mu}_{B^t}) + \boldsymbol{\beta}. \tag{5}$$

Separate statistics ($\Omega^s = (\boldsymbol{\mu}_B^s, \Sigma_B^s)$ and $\Omega^t = (\boldsymbol{\mu}_B^t, \Sigma_B^t)$) are estimated for B^s and B^t which are then used for whitening the corresponding activations and then followed by colouring the spherical distribution to an arbitrary one (see Fig. 3b). Details about the computation of W_B can be found in [18]. In addition, the F^{2}WCT blocks maintain a moving average of the statistics Ω_{avg}^t of the target domain which is used during inference.

3.3 Full-Feature Whitening and Colouring Transform for MSDA

In the MSDA scenario we have access to P labeled source datasets $\{\mathcal{S}_j\}_{j=1}^P$, where $\mathcal{S}_j = \{(I_i, y_i)\}_{i=1}^{N_j}$, and a target unlabeled dataset $\mathcal{T} = \{I_i\}_{i=1}^{N_t}$. Since, we are addressing closed-set DA all the datasets share the same categories and each of them is associated to a domain $\mathbf{D}_1^s, \ldots, \mathbf{D}_P^s, \mathbf{D}^t$, respectively. Our end goal is to learn a predictor for the target domain $\mathbf{D}_t$ exploiting the data in $\{\mathcal{S}_j\}_{j=1}^P \cup \mathcal{T}$.

Unlike many UDA methods [10,11], the proposed F^{2}WCT can be extended to the MSDA setting in a very straightforward way by having dedicated F^{2}WCT blocks for every domain $\mathbf{D}$, where the colouring parameters are shared amongst $P + 1$ domains. In details:

$$F^2WCT(\mathbf{x}_i^{\mathbf{D}_1^s}; \Omega^{\mathbf{D}_1^s}) = \mathbf{\Gamma} W_{B^{\mathbf{D}_1^s}}(\mathbf{x}_i^{\mathbf{D}_1^s} - \boldsymbol{\mu}_{B^{\mathbf{D}_1^s}}) + \boldsymbol{\beta}, \tag{6}$$

$$\vdots$$

$$F^2WCT(\mathbf{x}_i^{\mathbf{D}_P^s}; \Omega^{\mathbf{D}_P^s}) = \mathbf{\Gamma} W_{B^{\mathbf{D}_P^s}}(\mathbf{x}_i^{\mathbf{D}_P^s} - \boldsymbol{\mu}_{B^{\mathbf{D}_P^s}}) + \boldsymbol{\beta}, \tag{7}$$

$$F^2WCT(\mathbf{x}_i^{\mathbf{D}^t}; \Omega^{\mathbf{D}^t}) = \mathbf{\Gamma} W_{B^{\mathbf{D}^t}}(\mathbf{x}_i^{\mathbf{D}^t} - \boldsymbol{\mu}_{B^{\mathbf{D}^t}}) + \boldsymbol{\beta}. \tag{8}$$

The whitening operation of F^{2}WCT projects the marginal feature distributions of all $P + 1$ domains onto a hyper-spherical reference distribution, thereby minimizing the multiple domain discrepancies in a coherent fashion. As in Sect. 3.2, the moving average of target statistics $\Omega_{avg}^{\mathbf{D}^t}$ is maintained during training and is used during inference.

3.4 Training

Let $B^s = \{\mathbf{x}_1^s, \ldots, \mathbf{x}_m^s\}$ and $B^t = \{\mathbf{x}_1^t, \ldots, \mathbf{x}_m^t\}$ be two batches of the network's last-layer activations, from the source and target domain, respectively. Since, the

source samples are associated with labels, the standard cross-entropy loss (L^s) can be used for B^s:

$$L^s(B^s) = -\frac{1}{m}\sum_{i=1}^{m}\log p(y_i^s|\mathbf{x}_i^s), \tag{9}$$

However, for the target samples entropy loss is calculated as in [9], which acts as a regularizer. The entropy loss forces the network to be more confident in its predictions by producing peaked probability distribution at the output.

$$L^t(B^t) = -\frac{1}{m}\sum_{i=1}^{m}p(\mathbf{x}_i^t)\log p(\mathbf{x}_i^t), \tag{10}$$

Finally, the network is trained with a weighted sum of L^s and L^t:

$$L(B^s, B^t) = L^s(B^s) + \lambda L^t(B^t) \tag{11}$$

4 Experimental Results

In this section we describe the datasets and provide details about the experimental protocols adopted. We also report our experimental evaluation on the considered datasets and compare our proposed method with the state-of-the-art methods in UDA and MSDA, respectively.

4.1 Datasets

We conduct all our experiments on the *Digits-Five* dataset, built for recognizing digits, consists of five unique domains having numerical digits ranging between 0 and 9. It includes the USPS, MNIST, MNIST-M, SVHN and *Synthetic numbers* (SYN) datasets. SVHN contains images of real-world house numbers acquired from Google Street View. SYN includes about 500K computer generated digits having varying orientation, position, color, etc. USPS and MNIST are datasets of digits scanned from U.S. envelopes but having different resolutions. Finally, MNIST-M is the colored counterpart of MNIST.

4.2 Experimental Setup

To ensure fair comparison with other UDA and MSDA methods we adopt base networks from [8] and [22] for UDA and MSDA experiments, respectively. In the network we have plugged F^2WCT blocks right after each of the first two convolutional layers. We reason that strong alignment of low level features (e.g., colour and texture) is very important to bridge the domain gap. As a consequence, we act in the early convolutional layers of the network, which deal with low level features, by fully aligning intermediate feature distributions with F^2WCT blocks. A typical block in the network is given by (Conv Layer $\rightarrow$ F^2WCT $\rightarrow$ ReLU). For the remainder layers we have used BN based DA layers as in [9].

We trained the networks with Adam for 150 epochs with an initial learning rate of `1e−3` and we dropped the learning rate by a factor of 10 after 50 and 90 epochs. To ensure well-conditioned covariance matrices we have used a mini-batch size of 128 and 512 for the UDA and MSDA settings, respectively. The source and target samples are drawn randomly such that each domain is well represented in a mini-batch. The value of λ in Eq. 11 is set to 0.1 as in [9].

4.3 Results and Discussion

In this section we analyze the impact of the proposed components on the final classification accuracy and compare F^{2}WCT with the state-of-the-art methods.

Ablation Study. We conduct ablation studies on the digits dataset for single source UDA to demonstrate the benefits of performing full-whitening followed by a *colouring* transformation. We consider the following models: (i) F^{2}WCT, our full model, is composed of full-feature whitening and colouring; (ii) F^{2}WT where the colouring operation is replaced by *scale-shift* operation. This will validate the importance of *colouring* transform over scaling and shifting; and (iii) DWT [8] which considers *grouped* whitening. This comparison allows us to determine the necessity of full-feature whitening as opposed to grouped whitening.

Table 1. Ablation study of full-feature whitening and colouring transform versus relevant normalization techniques on Digits-Five. The target domain is shown in *italics*. The best numbers are highlighted in bold and the second best numbers are underlined.

Methods	MNIST → *USPS*	USPS → *MNIST*	SVHN → *MNIST*	MNIST → *MNIST-M*	Avg
Source only	78.9	57.1	60.1	63.6	64.92
F^{2}WCT (Ours)	**99.13** ± 0.05	**98.81** ± 0.07	<u>97.37</u> ± 0.10	**96.33** ± 0.09	**97.91**
F^{2}WT	99.03 ± 0.04	98.30 ± 0.07	78.96 ± 0.64	<u>81.41</u> ± 0.98	<u>89.42</u>
DWT [8]	<u>99.09</u> ± 0.09	<u>98.79</u> ± 0.05	**97.75** ± 0.10	45.46 ± 0.05	85.27
Target only	96.5	99.2	99.5	96.4	97.9

As can be observed from Table 1 our proposed F^{2}WCT outperforms all other baselines. F^{2}WT demonstrates that the need of colouring is particularly evident for more complicated adaptation settings as in SVHN → MNIST and MNIST → MNIST-M. While in simpler MNIST ↔ USPS settings the network has enough capacity already. DWT [8] is especially worse than F^{2}WCT in the MNIST → MNIST-M setting because grouped feature whitening can not align the source and target feature distributions optimally (see Sect. 2). Conversely, F^{2}WCT enables strong alignment of low level features through full whitening.

Comparison with State-of-the-Art Results. We compare our proposed F^{2}WCT with state-of-the-art methods, in both single source UDA and MSDA settings.

Table 2. Classification accuracy (%) on the Digits-Five for single source UDA settings in comparison with the state-of-the-art methods. The target domain is shown in *italics*. The best numbers are highlighted in bold and the second best numbers are underlined.

Methods	MNIST → *USPS*	USPS → *MNIST*	SVHN → *MNIST*	MNIST → *MNIST-M*	Avg
Source only	78.9	57.1	60.1	63.6	64.9
CORAL [5]	81.7	–	63.1	57.7	–
DANN [30]	85.1	73.0 ± 2.0	73.9	77.4	77.3
DSN [29]	91.3	–	82.7	83.2	–
CoGAN [12]	91.2	89.1 ± 0.8	–	62.0	–
ADDA [7]	89.4 ± 0.2	90.1 ± 0.8	76.0 ± 1.8	–	–
DRCN [28]	91.8 ± 0.1	73.7 ± 0.1	82.0 ± 0.2	–	–
ATT [27]	–	–	86.20	94.2	–
AutoDIAL [9]	97.96	97.51	89.12	36.86	80.36
SBADA-GAN [10]	97.6	95.0	76.1	**99.4**	<u>92.02</u>
GAM [26]	95.7 ± 0.5	98.0 ± 0.5	74.6 ± 1.1	–	–
MECA [25]	–	–	95.2	–	–
SE [24]	88.14 ± 0.34	92.35 ± 8.61	93.33 ± 5.88	–	–
DWT [8]	99.09 ± 0.09	98.79 ± 0.05	**97.75** ± 0.10	45.46 ± 0.05	85.27
F^{2}WCT (Ours)	**99.13** ± 0.05	**98.81** ± 0.07	<u>97.37</u> ± 0.10	<u>96.33</u> ± 0.09	**97.91**
Target only	96.5	99.2	99.5	96.4	97.9

Single-Source Unsupervised Domain Adaptation. In Table 2 we consider single-source adaptation settings where we adapt from a single source domain to a target domain. We consider four adaptation settings: MNIST → USPS, USPS → MNIST, SVHN → MNIST and MNIST → MNIST-M. The entire *labeled* train set of the source domain and *unlabeled* train set of the target domain is used for training a network whereas the dedicated test set of the target domain is used for evaluating the performance. We have considered the baselines reported in [8]. It is to be noted that we have chosen the baselines that do not utilize data augmentation. The variant of SE [24] which does not make use of data augmentation is therefore reported for fair comparison with other methods. However, for some baselines we could not report all the numbers due to the lack of availability in the corresponding adaptation settings.

From Table 2 we observe that on average our proposed F^{2}WCT outperforms all considered state-of-the-art methods by a considerable margin. Individually, our F^{2}WCT has the best accuracy in MNIST ↔ USPS settings and is the second best in SVHN → MNIST and MNIST → MNIST-M settings. Particularly, SBADA-GAN performs the best in the MNIST → MNIST-M setting due to the implicit data-augmentation through generation of synthetic data. Surprisingly, in overall F^{2}WCT achieves at par performance with the *target only* setting

Table 3. Classification accuracy (%) on Digits-Five for multi-source domain adaptation settings. The target domain is shown in *italics*. Best number is in bold and second best is underlined.

Models	MNIST, USPS, SVHN, SYN → *MNIST-M*	MNIST-M, USPS, SVHN, SYN → *MNIST*	MNIST, MNIST-M, SVHN, SYN → *USPS*	MNIST, USPS, MNIST-M, SYN → *SVHN*	MNIST, USPS, SVHN, MNIST-M → *SYN*	Avg
Source combine						
Source only	63.70 ± 0.83	92.30 ± 0.91	90.71 ± 0.54	71.51 ± 0.75	83.44 ± 0.79	80.33
DAN [13]	67.87 ± 0.75	97.50 ± 0.62	93.49 ± 0.85	67.80 ± 0.84	86.93 ± 0.93	82.72
DANN [30]	70.81 ± 0.94	97.90 ± 0.83	93.47 ± 0.79	68.50 ± 0.85	87.37 ± 0.68	83.61
Multi-source						
Source only	63.37 ± 0.74	90.50 ± 0.83	88.71 ± 0.89	63.54 ± 0.93	82.44 ± 0.65	77.71
DAN [13]	63.78 ± 0.71	96.31 ± 0.54	94.24 ± 0.87	62.45 ± 0.72	85.43 ± 0.77	80.44
CORAL [5]	62.53 ± 0.69	97.21 ± 0.83	93.45 ± 0.82	64.40 ± 0.72	82.77 ± 0.69	80.07
DANN [30]	71.30 ± 0.56	97.60 ± 0.75	92.33 ± 0.85	63.48 ± 0.79	85.34 ± 0.84	82.01
ADDA [7]	71.57 ± 0.52	97.89 ± 0.84	92.83 ± 0.74	75.48 ± 0.48	86.45 ± 0.62	84.84
DCTN [20]	70.53 ± 1.24	96.23 ± 0.82	92.81 ± 0.27	77.61 ± 0.41	86.77 ± 0.78	84.79
M^3SDA [22]	72.82 ± 1.13	98.43 ± 0.68	96.14 ± 0.81	81.32 ± 0.86	89.58 ± 0.56	87.65
AutoDIAL [9]	80.15 ± 1.32	<u>99.30</u> ± 0.04	98.60 ± 0.09	80.87 ± 0.68	95.28 ± 0.13	90.84
DWT [8]	80.68 ± 1.35	99.26 ± 0.05	<u>98.81</u> ± 0.08	**86.11** ± 0.25	**95.94** ± 0.10	<u>92.16</u>
F^2WCT (Ours)	**93.47** ± 0.41	**99.41** ± 0.04	**98.97** ± 0.06	<u>82.46</u> ± 0.81	<u>95.92</u> ± 0.12	**94.04**

without having access to any target label, demonstrating the effectiveness of our method.

Multi-source Unsupervised Domain Adaptation. In Table 3 we report results for MSDA setting where we adapt from multiple source domains to a single target domain. We consider all possible combinations of the 5 domains in Digits-Five for the experiments. For fairness in comparison with the baseline methods we follow the training protocol used in [22]. According to this protocol we randomly sample 25000 training images from each domain and 9000 images for evaluation. For the USPS, entire train and test set is used instead. We compare our method with DWT [8], Autodial: Automatic domain alignment layers [9] (AutoDIAL) and other baselines taken from [22]. We observe similar behaviour in the MSDA setting as our proposed F^2WCT also out-performs all the baselines on average accuracy, thereby obtaining state-of-the-art results. Notably, for the adaptation setting where MNIST-M is the target domain, the proposed full-feature whitening and colouring provides a boost of 12.79% over grouped whitening and scale-shifting in [8]. This validates our hypothesis that complete alignment of source and target feature distributions with full-feature whitening followed by colouring of the whitened features is more beneficial for tackling domain shift.

5 Conclusions

In this work we address UDA and MSDA by proposing domain alignment layers based on domain specific full-feature whitening and domain agnostic colouring

with F^{2}WCT blocks. On the one hand, full-feature whitening of intermediate features allows optimal alignment of source and target feature distributions by guaranteeing same covariance matrices for both source and target features. On the other, the colouring transform helps in restoring the capacity of the network. The proposed F^{2}WCT blocks can be easily incorporated in any standard CNN. Our experiments on digits dataset show consistent improved performances over other state-of-the-art methods in both UDA and MSDA settings. As future work, we plan to adapt the proposed feature alignment technique for large scale benchmarks with deeper networks.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS (2012)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
3. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR (2011)
4. Zen, G., Sangineto, E., Ricci, E., Sebe, N.: Unsupervised domain adaptation for personalized facial emotion recognition. In: ICMI (2014)
5. Sun, B., Feng, J., Saenko, K.: Return of frustratingly easy domain adaptation. In: AAAI (2016)
6. Saha, S., Banerjee, B., Merchant, S.N.: Unsupervised domain adaptation without source domain training samples: a maximum margin clustering based approach. In: ICVGIP (2016)
7. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR (2017)
8. Roy, S., Siarohin, A., Sangineto, E., Bulò, S.R., Sebe, N., Ricci, E.: Unsupervised domain adaptation using feature-whitening and consensus loss. In: CVPR (2019)
9. Cariucci, F.M., Porzi, L., Caputo, B., Ricci, E., Bulò, S.R.: AutoDIAL: automatic domain alignment layers. In: ICCV (2017)
10. Russo, P., Carlucci, F.M., Tommasi, T., Caputo, B.: From source to target and back: symmetric bi-directional adaptive GAN. In: CVPR (2018)
11. Hoffman, J., et al.: CyCADA: cycle-consistent adversarial domain adaptation. In: ICML (2018)
12. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: NIPS (2016)
13. Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. In: ICML (2015)
14. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: maximizing for domain invariance. arXiv preprint arXiv:1412.3474 (2014)
15. Sun, B., Saenko, K.: Deep CORAL: correlation alignment for deep domain adaptation. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9915, pp. 443–450. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49409-8_35
16. Mancini, M., Bulò, S.R., Caputo, B., Ricci, E.: AdaGraph: unifying predictive and continuous domain adaptation through graphs. In: CVPR (2019)
17. Mancini, M., Porzi, L., Rota Bulò, S., Caputo, B., Ricci, E.: Boosting domain adaptation by discovering latent domains. In: CVPR (2018)
18. Siarohin, A., Sangineto, E., Sebe, N.: Whitening and coloring batch transform for GANs. In: ICLR (2019)

19. Yao, Y., Doretto, G.: Boosting for transfer learning with multiple sources. In: CVPR (2010)
20. Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L.: Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In: CVPR (2018)
21. Mansour, Y., Mohri, M., Rostamizadeh, A.: Domain adaptation with multiple sources. In: NIPS (2009)
22. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. arXiv preprint arXiv:1812.01754 (2018)
23. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: ICML (2015)
24. French, G., Mackiewicz, M., Fisher, M.: Self-ensembling for visual domain adaptation. In: ICLR (2018)
25. Morerio, P., Cavazza, J., Murino, V.: Minimal-entropy correlation alignment for unsupervised deep domain adaptation. In: ICLR (2017, 2018)
26. Huang, H., Huang, Q., Krähenbühl, P.: Domain transfer through deep activation matching. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11220, pp. 611–626. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01270-0_36
27. Saito, K., Ushiku, Y., Harada, T.: Asymmetric tri-training for unsupervised domain adaptation. In: ICML (2017)
28. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D., Li, W.: Deep reconstruction-classification networks for unsupervised domain adaptation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 597–613. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_36
29. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.: Domain separation networks. In: NIPS (2016)
30. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**(1), 1–35 (2016). 2096-2030