

Adaptação de Domínio usando Camadas de Alinhamento de Domínio com Várias Fontes

Lucas Fernando Alvarenga e Silva

Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo – UNIFESP
12247-014, São José dos Campos, SP – Brazil
Email: e.lucas@unifesp.br

Jurandy Almeida

Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo – UNIFESP
12247-014, São José dos Campos, SP – Brazil
Email: jurandy.almeida@unifesp.br

Resumo—Modelos profundos, frente a mudança de domínio que pode existir entre conjuntos de dados anotados de treinamento e conjuntos de dados não-anotados de teste, podem produzir resultados insatisfatórios. A mudança de domínio é objetivo das áreas de pesquisa denominadas adaptação de domínio não-supervisionada, quando adapta-se um domínio-fonte para um domínio-alvo; e a adaptação de domínio de várias fontes, a qual adapta múltiplos domínios-fonte para um único domínio-alvo. Este trabalho apresenta as camadas de alinhamento de domínio entre várias fontes (do inglês, *Multi-Source Domain Alignment Layers* – MS-DIAL), que reduzem a discrepância entre as distribuições dos domínios-fonte e do domínio-alvo a partir da inserção de camadas de alinhamento de domínio em diversos níveis da rede. Foram realizados experimentos em tarefas de reconhecimento de dígitos e de objetos, os quais obtiveram resultados promissores que superam métodos do estado da arte.

Abstract—Deep models, in front of the domain shift that may exist between supervised training datasets and unsupervised test datasets, may yield poor results. The domain shift problem has been the goal of the research areas known as unsupervised domain adaptation, when adapting one source domain to one target domain, and multi-source domain adaptation, which adapts multiple source domains to a single target domain. This work presents the Multi-Source Domain Alignment Layers (MS-DIAL), which reduce the discrepancies between the distributions of the source and target domains by inserting domain alignment layers at different levels of the network. Experiments were performed on digit and object recognition tasks, which have obtained promising results that outperform state-of-the-art methods.

I. INTRODUÇÃO

Normalmente, é esperado que redes neurais profundas treinadas em um conjunto de dados rotulados produzam resultados equivalentes quando aplicadas a um conjunto de dados não-rotulados. Entretanto, essa hipótese assume que os dados usados no treinamento e no teste pertencem a um mesmo domínio, ou seja, apresentam uma mesma distribuição de probabilidade. Isso muitas vezes não é verdade, pois os dados podem ser coletados por dispositivos diferentes ou sob condições distintas e, portanto, não fazer parte do mesmo domínio, produzindo resultados insatisfatórios. Esse problema acontece devido à mudança de domínio (do inglês, *domain shift*) que pode existir entre as distribuições de dados e é objetivo da área de pesquisa conhecida como adaptação de domínio não-supervisionada (do inglês, *unsupervised domain adaptation* – UDA).

A maioria dos trabalhos de UDA considera a transferência de conhecimento entre um domínio fonte (do inglês, *source*) e um alvo (do inglês, *target*). Em geral, essas abordagens podem ser agrupadas em duas categorias principais: (i) as que reduzem a discrepância entre as distribuições de dados ou (ii) as que exploram características invariantes entre domínios (do inglês, *domain invariant features*) [1].

Todavia, os dados podem ser obtidos de diferentes fontes e pertencer a diferentes distribuições. Isso caracteriza a adaptação de domínio de várias fontes (do inglês, *multi-source domain adaptation* – MSDA), que é considerado um problema mais complexo e desafiador devido aos seguintes fatores: (i) a possível existência de um deslocamento entre as distribuições dos domínios-fonte, (ii) as classes que categorizam os domínios-fontes podem ser diferentes (do inglês, *category shift*) e (iii) as informações que cada domínio-fonte pode fornecer para o processo de transferência de conhecimento para o domínio-alvo podem ser complementares.

Este trabalho contribui com uma nova proposta para o problema de MSDA denominada MS-DIAL (do inglês, *Multi-Source Domain Alignment Layers*). Tratam-se de camadas de alinhamento de domínio que podem ser incorporadas em diversos níveis de qualquer rede a fim de reduzir a discrepância entre as distribuições de vários domínios-fonte e do domínio-alvo. Dessa forma, os parâmetros da rede são compartilhados entre todos os domínios, exceto nas camadas MS-DIAL, o que reduz o tempo de processamento e uso de memória.

Foram realizados experimentos em cinco conjuntos de dados públicos, a saber: MNIST [2], MNIST-M [3], SVHN [4] e *SynthDigits* [5], que foram propostos para tarefas de reconhecimento de dígitos; e Office-Home [6], que aborda a tarefa de reconhecimento de objetos. Os resultados obtidos com as camadas MS-DIAL demonstram que elas não só são eficazes como, em média, superam abordagens do estado da arte.

O restante deste trabalho está organizado da seguinte maneira. A Seção II discute trabalhos relacionados. A Seção III introduz as camadas MS-DIAL e mostra como elas são usadas para lidar com tarefas de MSDA. A Seção IV apresenta o protocolo experimental, os resultados e uma comparação do MS-DIAL com outros métodos. E, por fim, conclusões e perspectivas futuras são oferecidas na Seção V.

II. TRABALHOS RELACIONADOS

A adaptação de domínio não-supervisionada de único domínio-fonte anotado para um único domínio-alvo não-anotado (do inglês, *single-source to single-target domain adaptation*) tem por objetivo adaptar um modelo treinado em dados do domínio-fonte para reconhecer instâncias de dados do domínio-alvo, sem perda de desempenho. É um problema difícil e que tem diversas propostas de solução, algumas voltadas para modelos rasos, como TCA [7] e DME [8], mas que vem se voltando ao uso de redes profundas, como DANN [5], WMMD [9]. Outras vertentes, como DIAL [1], [10], foram tomadas como base para a proposta deste trabalho.

Já a adaptação de domínio não-supervisionada de várias fontes para um único alvo (do inglês, *multi-source to single-target domain adaptation*) é ainda mais desafiadora. É um problema em evolução nos últimos anos e que atualmente existem algumas propostas de solução, como M3SDA [11], DCTN [12], DARN [13], MDMN [14], MDAN [15] e MDDA [16]. De modo geral, todas as abordagens supracitadas adotam redes neurais de múltiplos fluxos, na qual a quantidade de fluxos é ajustada proporcionalmente à quantidade de domínios.

III. CAMADAS DE ALINHAMENTO DE DOMÍNIOS COM VÁRIAS FONTES

Sejam os conjuntos de dados anotados $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$ referentes a M domínios-fonte que compartilham o mesmo conjunto de rótulos \mathcal{Y} com um conjunto de dados não-anotados \mathcal{T} referente ao único domínio-alvo. Suponha que cada domínio-fonte $\mathcal{S}_i = \{(\mathbf{x}_i^j, \mathbf{y}_i^j)\}_{j=1}^{N_i}$ corresponde a um conjunto de tuplas que contém os dados observados $X_i = \{\mathbf{x}_i^j\}_{j=1}^{N_i}$ e seus respectivos rótulos $Y_i = \{\mathbf{y}_i^j\}_{j=1}^{N_i}$, em que N_i é o número de amostras em \mathcal{S}_i . Os rótulos Y_i foram extraídos da distribuição-fonte $p_i(\mathbf{x}, \mathbf{y})$ com uso dos dados observados X_i . Como os rótulos do domínio-alvo \mathcal{T} não são conhecidos, assuma que ele seja formado por dados $X_T = \{\mathbf{x}_T^j\}_{j=1}^{N_T}$ extraídos da distribuição-alvo $p_T(\mathbf{x}, \mathbf{y}), \mathbf{y} \in \mathcal{Y}$, em que N_T é o número de amostras em \mathcal{T} . Assim, o problema de MSDA consiste em encontrar um conjunto de parâmetros $\theta \in \Theta$ para uma rede neural de forma que suas previsões para o conjunto de rótulos $Y_T = \{\mathbf{y}_T^j\}_{j=1}^{N_T}$ ainda não conhecidos do domínio-alvo \mathcal{T} sejam as melhores possíveis.

Trabalhos recentes de MSDA usam topologias de redes com múltiplos fluxos, com a quantidade de fluxos proporcional à quantidade de domínios. A maioria delas adota um conjunto de parâmetros θ diferente para cada fluxo, com um extrator de características e um classificador por domínio; ou parcialmente compartilhado entre os fluxos, geralmente pelos extractores de características de todos os domínios, mas cada um tendo o seu próprio classificador. Diferente desses métodos, na topologia de rede desenvolvida neste trabalho, o conjunto de parâmetros θ é compartilhado entre os fluxos de todos os domínios, tanto pelos extractores de características quanto pelos classificadores, com exceção das camadas de alinhamento de domínio com várias fontes (do inglês, *multi-source domain alignment layers* – MS-DIAL), nas quais há um fluxo independente para cada

domínio. Dessa forma, é necessário apenas uma instância da topologia de rede e, portanto, apenas um conjunto de parâmetros θ , reduzindo assim o custo computacional e a utilização de memória.

A. Preditores de Fonte e Alvo

Este trabalho tem como base as camadas de alinhamento de domínio (do inglês, *domain alignment layers* – DIAL) [1], que foram projetadas para problemas de UDA de única fonte para único alvo. Elas reduzem a discrepância entre os domínios fonte e alvo ao alinhar suas distribuições a uma mesma distribuição de referência. Para isso, as amostras do domínio-fonte $x_S \subseteq X_S$ e do domínio-alvo $x_T \subseteq X_T$ e, em seguida, encaminhadas às camadas de normalização de lote [17] BN_S e BN_T , respectivamente, as quais ajustam suas distribuições a uma distribuição de referência, porém, sem realizar transformações afins, como mostrado na Equação 1. No final, os parâmetros α e β aprendíveis em tempo de treinamento, são usados para controlar o nível de sobreposição das distribuições fonte e alvo sobre a distribuição de referência, como indicado na Equação 2, em que \oplus é a concatenação das saídas de BN_S e BN_T .

$$BN(x) = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} \quad (1)$$

$$DIAL(x) = \{BN_S(x_S) \oplus BN_T(x_T)\} \cdot \alpha + \beta \quad (2)$$

As camadas MS-DIAL podem ser vistas como uma generalização das camadas DIAL, ou seja, elas estendem as transformações realizadas na Equação 2, aplicando-as a todos os domínios-fonte de um problema de MSDA, como mostrado na Equação 3. Similar ao DIAL, as amostras $x \subseteq \{X_1 \cup X_2 \cup \dots \cup X_M \cup X_T\}$ dos mini-lotes são agrupadas em sub-lotes $x_1 \subseteq X_1, x_2 \subseteq X_2, \dots, x_M \subseteq X_M, x_T \subseteq X_T$ de acordo com o domínio ao qual pertencem e, em seguida, encaminhadas às camadas de normalização de lote $BN_1, BN_2, \dots, BN_M, BN_T$, respectivamente. Por fim, as saídas de todas as camadas de normalização de lote são concatenadas e transformadas linearmente por meio dos parâmetros α e β , aprendíveis em tempo de treinamento, que maximizam o acerto nas previsões das amostras do domínio-alvo.

$$MS-DIAL(x) = \left\{ \left\{ \bigoplus_{i=1}^M BN_i(x_i) \right\} \oplus BN_T(x_T) \right\} \cdot \alpha + \beta \quad (3)$$

A Figura 1 ilustra o fluxo de dados realizado nas camadas MS-DIAL durante as fases de treinamento e inferência. Na fase de treinamento, o mini-lote de entrada é submetido aos processos de separação em sub-lotes, encaminhamento para as camadas de normalização de lote e, por fim, transformação linear com uso dos parâmetros α e β . O fluxo de dados na fase de inferência é indicado pelo caminho destacado em vermelho, no qual o mini-lote é preservado e as camadas MS-DIAL atuam como camadas de normalização de lote padrão.

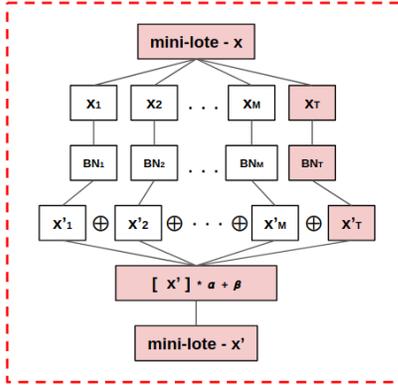


Figura 1. Fluxo de dados de mini-lotes x nas camadas MS-DIAL, em que BN_1, BN_2, \dots, BN_M são as camadas de normalização de lote aplicadas às amostras x_1, x_2, \dots, x_M dos domínios-fonte $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_M$, respectivamente; BN_T é a camada de normalização de lote aplicada às amostras x_T do domínio-alvo \mathcal{T} ; e α e β são parâmetros aprendíveis pela rede.

B. Treinamento e Inferência

Na fase de treinamento, os mini-lotes $x = x_1 \oplus x_2 \oplus \dots \oplus x_M \oplus x_T$ devem conter amostras provenientes de todos os domínios fontes, as quais são acompanhadas de seus rótulos; e também do domínio-alvo, para as quais os rótulos não são conhecidos. Ao término do encaminhamento de um mini-lote através da rede, são obtidas previsões $\{f_i^\theta(\mathbf{y}_i^k; \mathbf{x}_i^k)\}_{k=1}^{|\mathbf{x}_i|}$ para as amostras x_1, x_2, \dots, x_M dos domínios-fonte; e previsões $\{f_T^\theta(\mathbf{y}; \mathbf{x}_T^k)\}_{k=1}^{|\mathbf{x}_T|}$ para as amostras x_T do domínio-alvo.

Seguindo Carlucci *et al.* [1], o valor do erro entregue ao otimizador é dado por uma função de perda $\mathcal{L}(\theta)$ composta uma componente supervisionada $\mathcal{L}_S(\theta)$ e uma componente não-supervisionada $\mathcal{L}_T(\theta)$, as quais referem-se, respectivamente, a entropia cruzada das amostras dos domínios-fonte (Equação 4) e a entropia das amostras do domínio-alvo (Equação 5), a qual força o modelo a decidir com mais confiança. A função de perda $\mathcal{L}(\theta)$ é definida pela soma ponderada das componentes supracitadas, ou seja, $\mathcal{L}(\theta) = \mathcal{L}_S(\theta) + \lambda \mathcal{L}_T(\theta)$, em que λ é um hiperparâmetro associado ao peso da contribuição de $\mathcal{L}_T(\theta)$, o qual foi fixado em 0,1 nos experimentos.

$$\mathcal{L}_S(\theta) = - \sum_{i=1}^M \frac{1}{|\mathbf{x}_i|} \sum_{k=1}^{|\mathbf{x}_i|} \log f_i^\theta(\mathbf{y}_i^k; \mathbf{x}_i^k) \quad (4)$$

$$\mathcal{L}_T(\theta) = - \frac{1}{|\mathbf{x}_T|} \sum_{k=1}^{|\mathbf{x}_T|} \sum_{\mathbf{y} \in \mathcal{Y}} f_T^\theta(\mathbf{y}; \mathbf{x}_T^k) \log f_T^\theta(\mathbf{y}; \mathbf{x}_T^k) \quad (5)$$

Uma vez que os parâmetros θ estejam ajustados, os fluxos de dados associados aos domínios-fonte não são mais necessários. Em tempo de inferência, as camadas MS-DIAL operam como camadas de normalização de lote padrão, encaminhando mini-lotes apenas através do caminho associado ao domínio-alvo.

IV. EXPERIMENTOS E RESULTADOS

A abordagem proposta foi avaliada na tarefa de reconhecimento de dígitos, com uso dos conjuntos de dados MNIST [2], MNIST-M [3], SVHN [4] e *Synthetic Digits* (Synth) [5], e na

tarefa de reconhecimento de objetos, com uso do conjunto de dados Office-Home [6]. Esses cinco conjuntos de dados são públicos e têm sido amplamente usados na avaliação de métodos de UDA. Ambas as tarefas são bastante desafiadoras, pois há uma grande diferença entre as amostras dos domínios, como pode ser visto na Figura 2.



Figura 2. Exemplos de imagens do MNIST, MNIST-M, SVHN e Synth em (a) e dos domínios arte, clipart, produto e mundo real do Office-Home (b).

Os resultados do MS-DIAL foram comparados com os reportados por Wen *et al.* [13] para cinco abordagens de referência: **DANN** [5], **M3SDA** [11], **MDAN** [15], **MDMN** [14], **DARN** [13]. Também foram considerados os resultados **SRC** e **TAR** relatados por Wen *et al.* [13]. O primeiro refere-se ao treinamento do modelo em um conjunto composto pela união dos domínios-fonte e teste no domínio-alvo, o que caracteriza um limite inferior por não usar nenhuma adaptação de domínio. Já o segundo consiste no treinamento do modelo em dados do domínio-alvo com rótulos conhecidos e, portanto, refere-se a uma situação ideal e limita superiormente os resultados. Para ter uma comparação justa, foi adotado o mesmo protocolo experimental usado por Wen *et al.* [13]: em cada experimento, um domínio foi tomado como alvo e os demais foram usados como fonte, repetindo esse processo por várias vezes, sempre com um domínio diferente como alvo.

Para o reconhecimento de dígitos, os conjuntos de dados MNIST, MNIST-M, SVHN e Synth foram tratados como quatro domínios distintos. Para cada domínio, foram sub-amostrados aleatoriamente 20000 imagens para treinamento e 9000 para teste. Cada experimento foi repetido por 20 vezes, sendo reportado a média e o erro padrão dos melhores resultados alcançados. Adotou-se a mesma topologia de rede usada por Peng *et al.* [11], a qual foi inicializada com pesos aleatórios e suas camadas de normalização de lote foram substituídas por camadas MS-DIAL. Seguindo Roy *et al.* [18], a rede foi treinada do zero por 120 épocas usando o algoritmo de otimização Adam com tamanho de mini-lote de 64 (*i.e.*, 16 amostras por domínio), decaimento de peso de 0.0005 e taxa de aprendizado inicial de 0.001 com decaimento programado por um fator de 10 nas épocas 50 e 90.

Já para o reconhecimento de objetos, foram sub-amostrados aleatoriamente 2000 imagens de cada domínio para treinamento e as imagens restantes foram usadas para teste. Foram realizadas 20 repetições de cada experimento, sendo reportados a média e o erro padrão dos últimos resultados obtidos. Seguindo Roy *et al.* [18], foi adotada a rede ResNet-50 [19], substituindo-se camadas de normalização de lote padrão por camadas MS-DIAL. Primeiro, a rede foi inicializada com pesos pré-treinados na ImageNet [20] e a camada de saída

foi substituída por uma camada totalmente conectada com 65 neurônios de saída e pesos inicializados aleatoriamente. Em seguida, a rede foi treinada por 60 épocas usando o algoritmo de otimização SGD (do inglês, *Stochastic Gradient Descent*) com tamanho de mini-lote de 80 (*i.e.*, 20 por domínio), fator de momentum de 0,9, decaimento de peso de 0,0005, taxa de aprendizado inicial de 0,01 para os parâmetros da camada de saída e de 0,001 para os demais parâmetros da rede. Decaimento programado foi usado para reduzir as taxas de aprendizado iniciais por um fator de 10 na época 54.

As Tabelas I e II comparam os resultados das camadas MS-DIAL com os relatados por Wen *et al.* [13] para tarefas de MSDA com os conjuntos de dados de dígitos e Office-Home, respectivamente. Em negrito estão destacados os melhores resultados obtidos para cada domínio alvo. Como se pode observar, em todas as tarefas de MSDA, exceto pelo DARN, todas as demais abordagens foram superadas pelo MS-DIAL. Entretanto, apesar do DARN ser ligeiramente melhor para alguns domínios-alvo, o MS-DIAL tem um desempenho, em média, melhor que o DARN em ambas tarefas.

Tabela I
ACURÁCIA (%) NOS CONJUNTOS DE DADOS DE DÍGITOS.

Métodos	Domínios				Média
	MNIST	MNIST-M	SVHN	Synth	
SRC	96.78 ± 0.08	60.80 ± 0.21	68.99 ± 0.69	84.09 ± 0.27	77.66 ± 0.14
DANN	96.41 ± 0.13	60.10 ± 0.27	70.19 ± 1.30	83.83 ± 0.25	77.63 ± 0.35
M3SDA	96.95 ± 0.06	65.03 ± 0.80	71.66 ± 1.16	80.12 ± 0.56	78.44 ± 0.36
MDAN	97.10 ± 0.10	64.09 ± 0.31	77.72 ± 0.60	85.52 ± 0.19	81.11 ± 0.21
MDMN	97.15 ± 0.09	64.34 ± 0.27	76.43 ± 0.48	85.80 ± 0.21	80.93 ± 0.16
DARN	98.09 ± 0.03	67.06 ± 0.14	81.58 ± 0.14	86.79 ± 0.09	83.38 ± 0.06
MS-DIAL	94.33 ± 0.06	61.24 ± 1.27	85.61 ± 0.47	92.86 ± 0.20	83.51 ± 1.53
TAR	99.02 ± 0.02	94.66 ± 0.10	87.40 ± 0.17	96.90 ± 0.09	94.49 ± 0.07

Tabela II
ACURÁCIA (%) NO CONJUNTO DE DADOS OFFICE-HOME.

Métodos	Domínios				Média
	Arte	Clipart	Produto	Mundo Real	
SRC	58.02 ± 0.47	57.29 ± 0.30	74.26 ± 0.22	77.98 ± 0.25	66.89 ± 0.16
DANN	57.39 ± 0.69	57.35 ± 0.35	73.78 ± 0.27	78.12 ± 0.21	66.66 ± 0.19
M3SDA	64.05 ± 0.61	62.79 ± 0.37	76.21 ± 0.30	78.63 ± 0.22	70.42 ± 0.18
MDAN	68.14 ± 0.58	67.04 ± 0.21	81.03 ± 0.22	82.79 ± 0.15	74.75 ± 0.18
MDMN	68.67 ± 0.55	67.75 ± 0.20	81.37 ± 0.18	83.32 ± 0.14	75.28 ± 0.15
DARN	70.00 ± 0.38	68.42 ± 0.14	82.75 ± 0.21	83.88 ± 0.16	76.26 ± 0.13
MS-DIAL	82.85 ± 0.10	76.71 ± 0.10	80.74 ± 0.09	82.70 ± 0.09	80.75 ± 0.28
TAR	71.19 ± 0.38	79.16 ± 0.16	90.66 ± 0.15	85.60 ± 0.14	81.65 ± 0.12

V. CONCLUSÃO

Este trabalho apresentou uma nova abordagem para o problema de MSDA, denominada MS-DIAL, na qual a discrepância entre distribuições de vários domínios-fonte e um dado domínio-alvo é reduzida de forma automática a partir de parâmetros aprendíveis em tempo de treinamento por camadas de alinhamento de domínio inseridas em diversos níveis de uma rede neural. Isso permite compartilhar os parâmetros das demais camadas da rede entre todos os domínios, otimizando tempo de processamento e uso de memória.

A abordagem proposta foi validada em cinco conjuntos de dados públicos amplamente usados para avaliação de métodos de UDA, envolvendo as tarefas de reconhecimento de dígitos e de objetos. Os resultados obtidos foram promissores e superam, em média, os métodos do estado de arte.

A avaliação do MS-DIAL em outros conjuntos de dados, bem como a investigação de seu uso para abordar outras tarefas desafiadoras, como a adaptação de domínio de conjunto aberto (do inglês, *open set domain adaptation* – OSDA) e generalização de domínio (do inglês, *domain generalization* – DG), são possíveis direções para trabalhos futuros.

AGRADECIMENTOS

Este trabalho foi apoiado pela FAPESP (processos 2017/25908-6 e 2019/10998-5) e pelo CNPq (processos 423228/2016-1, 313122/2017-2 e 167857/2019-3).

REFERÊNCIAS

- [1] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. R. Bulò, “Just DIAL: domain alignment layers for unsupervised domain adaptation,” in *Int. Conf. Image Analysis and Processing*, 2017, pp. 357–369.
- [2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [3] Y. Ganin and V. S. Lempitsky, “Unsupervised domain adaptation by backpropagation,” in *ICML*, 2015, pp. 1180–1189.
- [4] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, “Reading digits in natural images with unsupervised feature learning,” in *NIPS Work. Deep Learning and Unsupervised Feature Learning*, 2011.
- [5] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. S. Lempitsky, “Domain-adversarial training of neural networks,” *JMLR*, vol. 17, pp. 59:1–59:35, 2016.
- [6] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, “Deep hashing network for unsupervised domain adaptation,” in *CVPR*, 2017, pp. 5385–5394.
- [7] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” *IEEE Trans. Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.
- [8] M. Baktashmotlagh, M. T. Harandi, and M. Salzmann, “Distribution-matching embedding for visual domain adaptation,” *JMLR*, vol. 17, pp. 108:1–108:30, 2016.
- [9] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, “Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation,” in *CVPR*, 2017, pp. 945–954.
- [10] F. M. Carlucci, L. Porzi, B. Caputo, E. Ricci, and S. R. Bulò, “Autodial: Automatic domain alignment layers,” in *ICCV*, 2017, pp. 5077–5085.
- [11] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, “Moment matching for multi-source domain adaptation,” in *ICCV*, 2019, pp. 1406–1415.
- [12] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin, “Deep cocktail network: Multi-source unsupervised domain adaptation with category shift,” in *CVPR*, 2018, pp. 3964–3973.
- [13] J. Wen, R. Greiner, and D. Schuurmans, “Domain aggregation networks for multi-source domain adaptation,” in *ICML*, 2020, pp. 10927–10937.
- [14] Y. Li, M. Murias, G. Dawson, and D. E. Carlson, “Extracting relationships by multi-domain matching,” in *NeurIPS*, 2018, pp. 6799–6810.
- [15] H. Zhao, S. Zhang, G. Wu, J. M. F. Moura, J. P. Costeira, and G. J. Gordon, “Adversarial multiple source domain adaptation,” in *NeurIPS*, 2018, pp. 8568–8579.
- [16] S. Zhao, G. Wang, S. Zhang, Y. Gu, Y. Li, Z. Song, P. Xu, R. Hu, H. Chai, and K. Keutzer, “Multi-source distilling domain adaptation,” in *AAAI Conf. Artificial Intelligence*, 2020, pp. 12975–12983.
- [17] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *ICML*, F. R. Bach and D. M. Blei, Eds., 2015, pp. 448–456.
- [18] S. Roy, A. Siarohin, E. Sangineto, S. R. Bulò, N. Sebe, and E. Ricci, “Unsupervised domain adaptation using feature-whitening and consensus loss,” in *CVPR*, 2019, pp. 9471–9480.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *CVPR*, 2016, pp. 770–778.
- [20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F.-F. Li, “Imagenet large scale visual recognition challenge,” *IJCV*, vol. 115, no. 3, pp. 211–252, 2015.