Weighted Pseudo Labeling Refinement for Plant Identification

Youshan Zhang, Brian D. Davison

Lehigh University, 113 Research Drive, Bethlehem, PA, 18015

Abstract

Unsupervised domain adaptation (UDA) focuses on transferring knowledge from a labeled source domain to an unlabeled target domain. However, existing domain adaptation methods try to handle various DA scenarios that are subject to imbalanced labels or large domain discrepancy datasets. In this paper, we propose a weighted pseudo labeling refinement model (WPLR) to balance the dataset using a weighted cross-entropy loss. We also utilize the CORAL loss to further reduce the domain difference. To improve the generalizability of the model, we develop an easy-to-hard pseudo labeling refinement process by probabilistic soft selection to suppress noisy predicted target labels. Experimental results demonstrate our WPLR model yields promising results on the PlantCLEF 2021 Challenge.

Keywords

Unsupervised domain adaptation, Pseudo labeling refinement, Plant identification

1. Introduction

Automatic plant identification is helpful for the general audience in recognizing plant species without the expertise of botanists. Deep neural networks can improve recognition performance when a large number of labeled data are used for training but suffer from significant performance degradation when deployed in a new domain due to the problem of domain shift. However, the domain shift or domain mismatch problem exists for the plant identification problem in PlantCLEF. Due to the significant difference between herbarium and real photos, classification models often do not generalize well to the novel field photo domain.

To circumvent the domain shift issue, the unsupervised domain adaptation (UDA) method has been proposed, which can transfer the model trained on the labeled source domain to an unlabeled target domain. Existing deep learning methods can be categorized into two major tracks: discrepancy-based methods [1, 2, 3] and adversarial learning methods [4, 5, 6]. The former aligns the distributions of source and target domains by directly minimizing the difference metric between feature distributions of the two domains, such as Maximum Mean Discrepancy (MMD) [1], CORrelation ALignment [2], Kullback-Leibler divergence [3], Jensen–Shannon divergence [7], and Wasserstein distance [8]. The latter category methods are inspired by GANs [9], and adversarial learning has shown its power in learning domain invariant

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania
yoz217@lehigh.edu (Y. Zhang); bdd3@lehigh.edu (B. D. Davison)
https://sites.google.com/view/youshanzhang (Y. Zhang); http://www.cse.lehigh.edu/~brian/ (B. D. Davison)
0000-0002-0074-0979 (Y. Zhang); 0000-0002-9326-3648 (B. D. Davison)

representations. It consists of a domain discriminator and a feature extractor. The domain discriminator aims to distinguish the source domain from the target domain, while the feature extractor aims to learn domain-invariant representations to fool the domain discriminator [4, 5, 6]. There is also much exploration of adversarial learning methods, such as DANN [10], MCD [11], TADA [12], SymNets [13], and ACDA [14].

Although many methods are proposed for domain adaptation, most of them are tested on small domain divergence datasets, which may have lower transferability to large-divergence datasets, and the data imbalance problem is not well addressed. To address these challenges, we offer two contributions:

1. We propose a weighted cross-entropy loss to balance the categorical data. To minimize the domain divergence, we utilize the existing CORAL loss.
2. To remove noisy pseudo labels in the target domain, we also employ an easy-to-hard pseudo labeling refinement process by probabilistic soft selection. We then form a high-quality pseudo-labeled target domain to improve the generalizability of the model.

2. Dataset

PlantCLEF 2021 is a large-scale dataset of the PlantCLEF 2021 task[15, 16], organized in the context of the LifeCLEF 2021 challenge. Fig. 1 shows some challenging images in this dataset. Tab. 1 lists the statistics on PlantCLEF 2021 dataset. Due to the significant difference between herbarium and real photos, it is extremely difficult to identify the correct class. All images are the same as PlantCLEF 2020 dataset [17], but it also introduces five "traits" covering exhaustively all species of the challenge.

Figure 1: Example images of the herbarium domain and photo domain. The large discrepancy between the two domains causes difficulty in improving the performance of the model.

Table 1
Statistics of the PlantCLEF 2021 dataset

Domain	Number of Samples	Number of Classes
Herbarium (H)	320,750	997
Herbarium_photo_associations (A)	1,816	244
Photo (P)	4,482	375
Test (T)	3,186	-

3. Methods

In this section, we will first introduce the problem and notation for UDA, and then introduce the different components of our Weighted Pseudo Labeling Refinement (WPLR) model.

3.1. Problem and notation

In this work, we consider the unsupervised domain adaptation (UDA) classification problem in the following setting. There exists a labeled source domain $\mathcal{D}_\mathcal{S} = \{\mathcal{X}^i_\mathcal{S}, \mathcal{Y}^i_\mathcal{S}\}^{\mathcal{N}_\mathcal{S}}_{i=1}$ of $\mathcal{N}_\mathcal{S}$ labeled samples in C categories and a target domain $\mathcal{D}_\mathcal{T} = \{\mathcal{X}^j_\mathcal{T}\}^{\mathcal{N}_\mathcal{T}}_{j=1}$ of $\mathcal{N}_\mathcal{T}$ samples without any labels (*i.e.*, $\mathcal{Y}_\mathcal{T}$ is unknown). The samples $\mathcal{X}_\mathcal{S}$ and $\mathcal{X}_\mathcal{T}$ obey the marginal distributions of $P_\mathcal{S}$ and $P_\mathcal{T}$. The conditional distributions of the two domains are denoted as $Q_\mathcal{S}$ and $Q_\mathcal{T}$. Due to the discrepancy between the two domains, the distributions are assumed to be different, i.e., $P_\mathcal{S} \neq P_\mathcal{T}$ and $Q_\mathcal{S} \neq Q_\mathcal{T}$. Our ultimate goal is to learn a classifier F under a feature extractor G, which reduces domain discrepancy and improves the generalization ability of the classifier to the target domain.

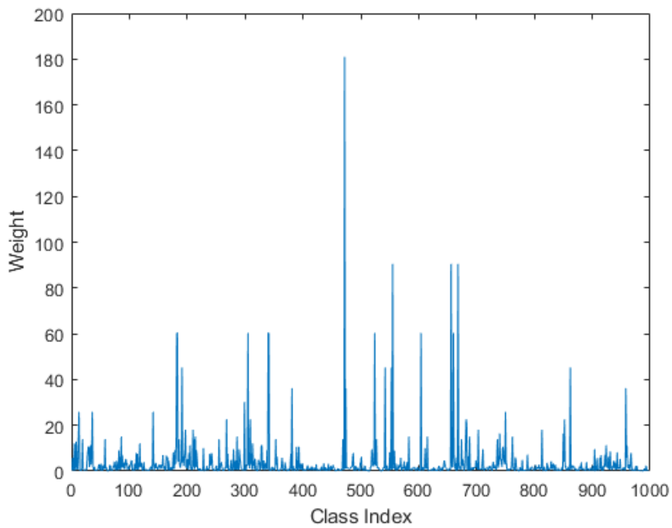

Figure 2: The weight of each class.

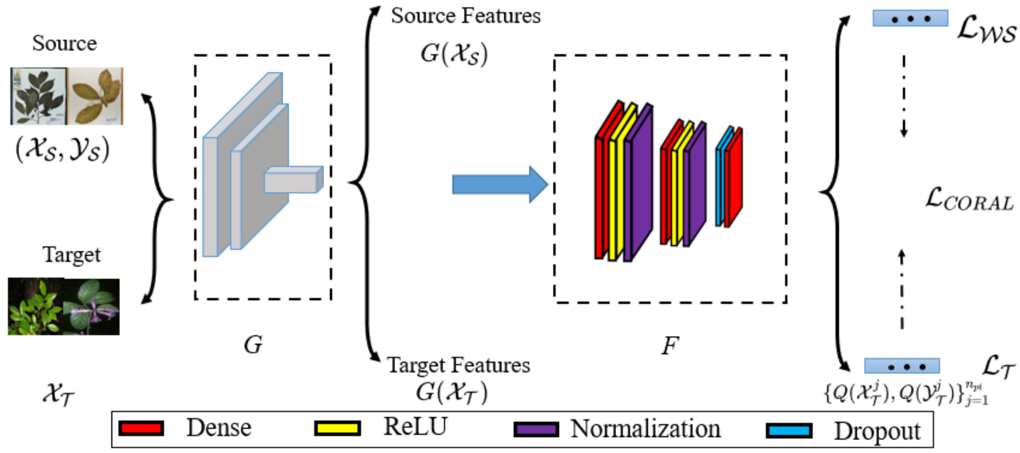

Figure 3: Architecture of the WPLR model. We first utilize NASNetLarge as the feature extractor G to extract features from the two domains ($G(\mathcal{X}_{\mathcal{S}})$ and $G(\mathcal{X}_{\mathcal{T}})$). The shared classifier F is then trained using the extracted features. $\mathcal{L}_{\mathcal{WS}}$ is the weighted source classification loss, $\mathcal{L}_{CORAL}$ is the CORAL loss, and $\mathcal{L}_{\mathcal{T}}$ is the pseudo-labeled target domain classification loss. $\{Q(\mathcal{X}_{\mathcal{T}}^j), Q(\mathcal{Y}_{\mathcal{T}}^j)\}_{j=1}^{n_{pt}}$ is the pseudo-labeled target domain after T times pseudo labeling refinement processes. Best viewed in color.

3.2. Weighted source classifier

The task in the source domain is trained using the typical cross-entropy loss. However, there are imbalanced numbers of samples of each category. To handle this issue, we develop a weighted source classifier to balance the weight of each category based on the source samples. We define the weight of each class in the following equation.

$$W = \frac{median(\{\frac{\mathcal{N}_{\mathcal{S}}^c}{\mathcal{N}_{\mathcal{S}}}\}_{c=1}^{C})}{\{\frac{\mathcal{N}_{\mathcal{S}}^c}{\mathcal{N}_{\mathcal{S}}}\}_{c=1}^{C}}, \tag{1}$$

where $\mathcal{N}_{\mathcal{S}}^c$ is the number of samples in each class, $\{\frac{\mathcal{N}_{\mathcal{S}}^c}{\mathcal{N}_{\mathcal{S}}}\}_{c=1}^{C} \in \mathbb{R}^{997\times 1}$ is the frequency of images in each class, $median(\cdot)$ takes the median value of the frequency. The frequency value varies; the median represents the middle frequency better than mean would. Fig. 2 shows the weight of each class (997 classes in total). Therefore, we develop the weighted cross-entropy loss for the labeled source domain in Eq. 2.

$$\mathcal{L}_{\mathcal{WS}} = \frac{1}{\mathcal{N}_{\mathcal{S}}} \sum_{i=1}^{\mathcal{N}_{\mathcal{S}}} W_i \times \mathcal{L}_{ce}(F(G(\mathcal{X}_{\mathcal{S}}^i)), \mathcal{Y}_{\mathcal{S}}^i), \tag{2}$$

where $\mathcal{L}_{ce}$ is the typical cross-entropy loss, F is the classifier in Fig. 3, and $F(G(\mathcal{X}_{\mathcal{S}}^i))$ is the predicted label.

3.3. CORAL loss

CORrelation ALignment loss (CORAL) [2] is one frequently used distance-based loss function to minimize the difference between source and target domain. We also integrate CORAL loss during the training as follows,

$$\mathcal{L}_{CORAL} = \frac{1}{4d^2}||COV(F(G(\mathcal{X}_\mathcal{S}))) - COV(F(G(\mathcal{X}_\mathcal{T})))||_F^2, \quad (3)$$

where d is the feature dimensionality, $COV(\cdot)$ is the covariance matrices of the source and target features, and $||\cdot||_F^2$ denotes the squared matrix Frobenius norm. Therefore, our model is able to minimize the domain divergence between the source domain and the target domain during the training.

3.4. Pseudo labeling refinement

To further reduce the domain difference, we also generate pseudo labels for the target domain. However, the detrimental effects of bad pseudo-labels are still significant. To mitigate this issue, we employ a T times recurrent easy-to-hard pseudo-label refinement process to improve the quality of the pseudo-labels in the target domain via imposing a probabilistic soft selection [18, 19].

The initial shared classifier F is optimized by $\mathcal{L}_{\mathcal{WS}}$. For the inference, we can directly get predicted results for one target domain sample $F(G(\mathcal{X}_\mathcal{T}^j))$. Let Softmax$(F(G(\mathcal{X}_\mathcal{T}^j)))$ be the predicted probability for each class, and $\mathcal{Y}_{\mathcal{PT}}^j = max(\text{Softmax}(F(G(\mathcal{X}_\mathcal{T}^j))))_{index}$ be its dominant class label, where $max(\cdot)_{index}$ return the index of the maximum probability value. Therefore, for the probabilistic soft selection, a higher quality pseudo label is defined as $max(\text{Softmax}(F(G(\mathcal{X}_\mathcal{T}^j)))) > p_t$, where p_t is a threshold probability in number of t training. For T times recurrent easy-to-hard pseudo-label refinement, for easy examples, p_t has a higher value and for hard examples, p_t has a lower value, hence $p_1 > p_2 > \cdots > p_T$.

In pseudo labeling refinement, we form a robust new pseudo-labeled domain in the following equation,

$$\{Q(\mathcal{X}_\mathcal{T}^j), Q(\mathcal{Y}_\mathcal{T}^j)\}_{j=1}^{n_{pt}} \quad \text{if and only if} \quad max(\text{Softmax}(F(G(\mathcal{X}_\mathcal{T}^j)))) > p_t \quad (4)$$

where $Q(\cdot)$ represents the high quality, n_{pt} is the number of higher quality pseudo labels for the target domain. We hence can mitigate detrimental effects of bad pseudo-labels using Eq. 4. Similar to Eq. 2, we define the pseudo-labeled target domain loss as:

$$\mathcal{L}_\mathcal{T} = \frac{1}{n_{pt}} \sum_{j=1}^{n_{pt}} W_j \times \mathcal{L}_{ce}(F(G(Q(\mathcal{X}_\mathcal{T}^j))), Q(\mathcal{Y}_\mathcal{T}^j)), \quad (5)$$

where W is the weight of each class and $\mathcal{L}_{ce}$ is the cross-entropy loss.

3.5. WPLR model

Fig. 3 depicts the overall framework of our proposed WPLR model. Taken together, our model minimizes the following objective function:

$$\arg\min\ (\mathcal{L}_{\mathcal{WS}} + \mathcal{L}_{CORAL} + \sum_{t=1}^{T} \mathcal{L}_\mathcal{T}^t) \quad (6)$$

where $\mathcal{L}_{\mathcal{WS}}$ is the weighted source classification loss, $\mathcal{L}_{CORAL}$ is the CORAL loss, and $\mathcal{L}_\mathcal{T}$ is the pseudo-labeled target domain classification loss.

4. Experiments

4.1. Implementation details

We first extract features from the last fully connected layer [20, 21, 22] of a retrained NASNet-Large [23] model. One image can be denoted by a feature vector with the size of 1×1000. Therefore, the feature representation of domain herbarium (H) has the size of $320,750 \times 1000$, domain herbarium_photo_associations (A) has the size of $1,816 \times 1000$, domain photo (P) has the size of $4,482 \times 1000$, and domain test (T) has the size of $3,186 \times 1000$. Domain H + A has the size of $322,566 \times 1000$. In Tab. 2, H → P represents learning knowledge from domain H, which is applied to domain P [24].

We implement our approach using PyTorch. The outputs of the three Linear layers are 1000, 1000 and $|C|$, respectively. Parameters in recurrent pseudo labeling are $T = 5$ and $\{p_t\}_{t=1}^{5} = [0.9, 0.8, 0.7, 0.6, 0.5]$. Learning rate (0.001), batch size (64), optimizer (Adam) and number of epochs ($\mathcal{N}_{\mathcal{S}}/64$) are determined by performance on the source domain. Experiments are performed with a GeForce 1080 Ti. We also compare our results with four domain adaptation methods: DANN [10], ADDA [5], NASNetLarge-ACL [24] and BA3US [25].

4.2. Results

Table 2
Accuracy (%) on PlantCLEF 2021 dataset for photo domain

Task	A → P	H → P	H+A → P
DANN [10]	1.07	1.85	2.01
ADDA [5]	2.95	3.05	3.43
BA3US [25]	3.56	4.65	5.31
NASNetLarge-ACL [24]	5.98	8.64	9.67
WPLR- $\mathcal{L}_{CORAL} - \mathcal{L}_{\mathcal{T}}$	6.03	9.12	10.03
WPLR- $\mathcal{L}_{\mathcal{T}}$	6.12	9.23	11.46
WPLR- $\mathcal{L}_{CORAL}$	6.22	9.47	12.51
WPLR	6.38	9.645	13.44

Tab. 2 shows the results of our WPLR model of the photo domain. We report the accuracy of the whole photo domain ($Acc = \sum_{j=1}^{\mathcal{N}_{\mathcal{T}}}(\hat{\mathcal{Y}_{\mathcal{T}}}_j == \mathcal{Y}_{\mathcal{T}j})/\mathcal{N}_{\mathcal{T}} \times 100$), where $\hat{\mathcal{Y}}_{\mathcal{T}}$ is the predicted label for the target domain. Compared with all other four methods, our WPLR model achieves the highest accuracy in all three tasks, and especially in H+A → P task.

We also carefully conduct an ablation study to demonstrate the effects of different loss functions on final classification accuracy. Notice that weighted source classification loss $\mathcal{L}_{\mathcal{WS}}$ is required for UDA. "**WPLR**- $\mathcal{L}_{CORAL} - \mathcal{L}_{\mathcal{T}}$" is implemented without $\mathcal{L}_{CORAL}$ and $\mathcal{L}_{\mathcal{T}}$. It is a simple model, which only reduces the source risk without minimizing the domain discrepancy using $\mathcal{L}_{\mathcal{WS}}$. "**WPLR**- $\mathcal{L}_{CORAL}$" reports results without performing CORAL loss. "**WPLR**- $\mathcal{L}_{\mathcal{T}}$" reports results without performing the T time pseudo labeling refinement process. We can find that with the increasing number of loss functions, the accuracy of our model keeps improving.

Table 3
MRR on PlantCLEF 2021 challenge for test domain

Team	Full test set	Sub-set of the test set
Organizer's submission [15]	0.198	0.093
Neuon AI	0.181	0.158
LU (ours)	0.065	0.037
Domain_run	0.065	0.037
To_be	0.056	0.038

The effectiveness of loss functions on classification accuracy is ordered as $\mathcal{L}_{\mathcal{T}} > \mathcal{L}_{CORAL}$. Therefore, the proposed weighted classification loss, CORAL loss, and easy-to-hard target domain pseudo labeling refinement approaches are effective in minimizing target domain risk and improving the accuracy.

We also list the final performance of our model in the test domain in Tab. 3. Our model earns the second position in the PlantCLEF 2021 challenge. We provided a total of nine submissions; the MRR of the full test set ranged from 0.034 to 0.065, as a result of varying hyperparameters (different number of iterations, T and p_t).

5. Discussion

There are two compelling advantages of our WPLR model. First, we propose a weighted cross-entropy loss to mitigate the imbalanced data issue in the source domain. Secondly, we develop an easy-to-hard refinement process to improve the quality of pseudo labels in the target domain. This strategy considers probabilistic soft selection, and it hence can push the shared classifier F towards the target domain. Compared with other baselines in Tab. 2, the T times easy-to-hard refinement process is effective in improving the classification accuracy and further reduces the domain discrepancy. However, our model only earns the second position in the challenge, and the results are a little bit lower than the Organizer's submission. One underlying reason is that our model cannot extract very robust invariant features. Therefore, we will consider designing a better feature extractor method and distill the domain invariant features across the two domains for future work. In addition, we would like to include more external data during the training (e.g., GBIF [26]).

6. Conclusion

In this paper, we propose a novel weighted pseudo labeling refinement (WPLR) method for domain adaptation to solve the plant identification problem. We develop a weighted cross-entropy loss to balance the categorical data and utilize the CORAL loss to minimize the domain divergence. We also employ an easy-to-hard pseudo labeling refinement process by probabilistic soft selection. It can improve the quality of pseudo labels and remove the detrimental effects of bad labels. Experimental results demonstrate our proposed WPLR model is better than several baselines.

References

[1] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance, arXiv preprint arXiv:1412.3474 (2014).

[2] B. Sun, K. Saenko, Deep CORAL: Correlation alignment for deep domain adaptation, in: European Conference on Computer Vision, Springer, 2016, pp. 443–450.

[3] Z. Meng, J. Li, Y. Gong, B. Juang, Adversarial teacher-student learning for unsupervised domain adaptation, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 5949–5953.

[4] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, The Journal of Machine Learning Research 17 (2016) 2096–2030.

[5] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.

[6] Y. Zhang, H. Ye, B. D. Davison, Adversarial reinforcement learning for unsupervised domain adaptation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 635–644.

[7] J. Jiang, X. Wang, M. Long, J. Wang, Resource efficient domain adaptation, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 2220–2228.

[8] B. Bhushan Damodaran, B. Kellenberger, R. Flamary, D. Tuia, N. Courty, DeepJDOT: Deep joint distribution optimal transport for unsupervised domain adaptation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 447–463.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.

[10] M. Ghifary, W. B. Kleijn, M. Zhang, Domain adaptive neural networks for object recognition, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2014, pp. 898–904.

[11] K. Saito, K. Watanabe, Y. Ushiku, T. Harada, Maximum classifier discrepancy for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3723–3732.

[12] X. Wang, L. Li, W. Ye, M. Long, J. Wang, Transferable attention for domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 5345–5352.

[13] Y. Zhang, H. Tang, K. Jia, M. Tan, Domain-symmetric networks for adversarial domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5031–5040.

[14] Y. Zhang, B. D. Davison, Adversarial continuous learning in unsupervised domain adaptation., in: ICPR Workshops (2), 2020, pp. 672–687.

[15] H. Goëau, P. Bonnet, A. Joly, Overview of PlantCLEF 2021: cross-domain plant identification, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.

[16] A. Joly, H. Goëau, S. Kahl, L. Picek, T. Lorieul, E. Cole, B. Deneu, M. Servajean, R. Ruiz De

Castañeda, I. Bolon, H. Glotin, R. Planqué, W.-P. Vellinga, A. Dorso, H. Klinck, T. Denton, I. Eggel, P. Bonnet, H. Müller, Overview of LifeCLEF 2021: a system-oriented evaluation of automated species identification and species distribution prediction, in: Proceedings of the Twelfth International Conference of the CLEF Association (CLEF 2021), 2021.
[17] H. Goëau, P. Bonnet, A. Joly, Overview of lifeclef plant identification task 2020, in: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece., 2020.
[18] Y. Zhang, B. D. Davison, Deep spherical manifold gaussian kernel for unsupervised domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 4443–4452.
[19] Y. Zhang, B. D. Davison, Efficient pre-trained features and recurrent pseudo-labeling in unsupervised domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 2719–2728.
[20] Y. Zhang, J. P. Allem, J. B. Unger, T. B. Cruz, Automated identification of hookahs (waterpipes) on Instagram: an application in feature extraction using convolutional neural network and support vector machine classification, Journal of Medical Internet Research 20 (2018) e10513.
[21] Y. Zhang, B. D. Davison, Modified distribution alignment for domain adaptation with pre-trained Inception ResNet, arXiv preprint arXiv:1904.02322 (2019).
[22] Y. Zhang, B. D. Davison, Impact of ImageNet model selection on domain adaptation, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops, 2020, pp. 173–182.
[23] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, Learning transferable architectures for scalable image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8697–8710.
[24] Y. Zhang, B. D. Davison, Adversarial consistent learning on partial domain adaptation of PlantCLEF 2020 challenge, in: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, 2020.
[25] J. Liang, Y. Wang, D. Hu, R. He, J. Feng, A balanced and uncertainty-aware approach for partial domain adaptation, arXiv preprint arXiv:2003.02541 (2020).
[26] L. Picek, M. Sulc, J. Matas, Recognition of the amazonian flora by inception networks with test-time class prior estimation, in: Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, 2019.