Improved Herbarium-Field Triplet Network for Cross-Domain Plant Identification: NEUON Submission to LifeCLEF 2021 Plant

Sophia Chulif, Yang Loong Chang

Department of Artificial Intelligence, NEUON AI, 94300, Sarwak, Malayisa

Abstract

This paper presents the submissions made by our team to PlantCLEF 2021. The challenge's goal was to identify plant species based on the test set made from only plant images in the field, given a training dataset consisting of primarily herbarium images. We implemented a two-streamed Herbarium-Field Triplet Loss Network to evaluate the similarity between herbarium and field pairs, thereby matching species from both herbarium and field domains. The network is made from two convolutional neural networks taking herbarium and field images as input, respectively. The network employed is a similar but improved version of our submission to the previous year's challenge [1]. In addition, we trained a one-streamed network taking both herbarium and field images as input to enable the learning of the features of each species irrespective of their domains. We found that an ensemble of these networks performed better than the Herbarium-Field Triplet Loss Network alone. We achieved a Mean Reciprocal Rank (MRR) of 0.181 for the primary metric, which focused on the whole test set. Comparably, we achieved an MRR of 0.158 for the secondary metric, which focused on the subset of species with fewer field training images.

Keywords

Cross-domain plant identification, herbarium, computer vision, triplet loss, convolutional neural networks

1. Introduction

The LifeCLEF evaluation campaign aims at boosting and evaluating the advances of plant and animal identification since 2011 [2]. The 2021 edition proposed four different challenges namely, PlantCLEF 2021 [3], BirdCLEF 2021 [4], GeoLifeCLEF 2021 [5], and SnakeCLEF 2021 [6]. The LifeCLEF 2021 plant identification challenge (PlantCLEF 2021) was evaluated as a cross-domain classification task. Likewise, in PlantCLEF 2020 [7], the objective was to identify plants in the field based on a training dataset composed primarily of herbarium images with little or no plant field images at all. The same training and test data from PlantCLEF 2020 were provided, however, 5 traits of the species were introduced.

The results obtained in PlantCLEF 2020 demonstrated that the challenge was particularly difficult as compared to the previous editions of PlantCLEF. Generally, herbarium and their respective plant field images vary in terms of their attributes like color, plant organs, captured

viewpoints, and illumination settings. Consequently, the difference in their input distribution makes it difficult to carry out conventional automated plant species identification in which the classification problem is straightforward, whereby the source and target domains are the same. In addition, it was shown that transfer learning from herbarium to field data based on conventional automated classification did not perform well [8, 9].

To tackle this problem, similarly in our approach from PlantCLEF 2020 [1], we adopt the triplet network architecture [10] from the face recognition domain. The core concept of this architecture is to feed the network with a triplet sample: two samples sharing the same label and one with a different label. Then, the network is trained to minimize the feature distance between the same labels and maximize the feature distance between different labels. The triplet loss aims to separate identical pairs from different pairs by a distance margin.

Likewise, we implemented a Herbarium-Field triplet loss network to minimize the feature distance between the same herbarium-field pairs, while maximizing the feature distance between different herbarium-field pairs. This network achieved impressive genericity by obtaining equivalent results regardless of whether the species has many or little field training images in PlantCLEF 2020.

As an improvement to our previous approach, we employed additional input augmentation, different network choices, finetune hyperparameters, and longer training duration in different stages. Furthermore, a one-streamed convolutional neural network (CNN) taking both herbarium and field images was trained to compare their behavior and performance to the triplet network. In contrast, the Herbarium-Field triplet loss network only takes into account existing herbarium-field pairs, meanwhile, the species with no herbarium-field pairs were neglected. Therefore, in terms of input data, triplet learning has less data used compared to our one-stream network. Moreover, we revised the extraction method of our herbarium dictionary to enhance the representation for the herbarium-field feature similarity comparison.

This paper presents our team's submission to PlantCLEF 2021. We discuss our implemented networks and methods in detail, our results obtained, and analyses made from the results.

2. Methodology

2.1. Networks and Architecture

We propose a Herbarium-Field Triplet Loss Network (HFTL Network) to model common features between herbarium-field pairs. Its main concept is to minimize the feature distance between the same species and maximize the feature distance between different species. A large feature distance difference denotes that a herbarium-field pair is of different species, while a small distance denotes the same species. In addition to HFTL Network, we construct a one-streamed mixed network (OSM Network) whereby herbarium and field images are trained together in a single network without distinction of herbarium and field classes. These networks are implemented based on the Inception-v4 and Inception-ResNet-v2 architectures [11] and are detailed in Section 3.3. The two core networks constructed in our submissions are as follows:
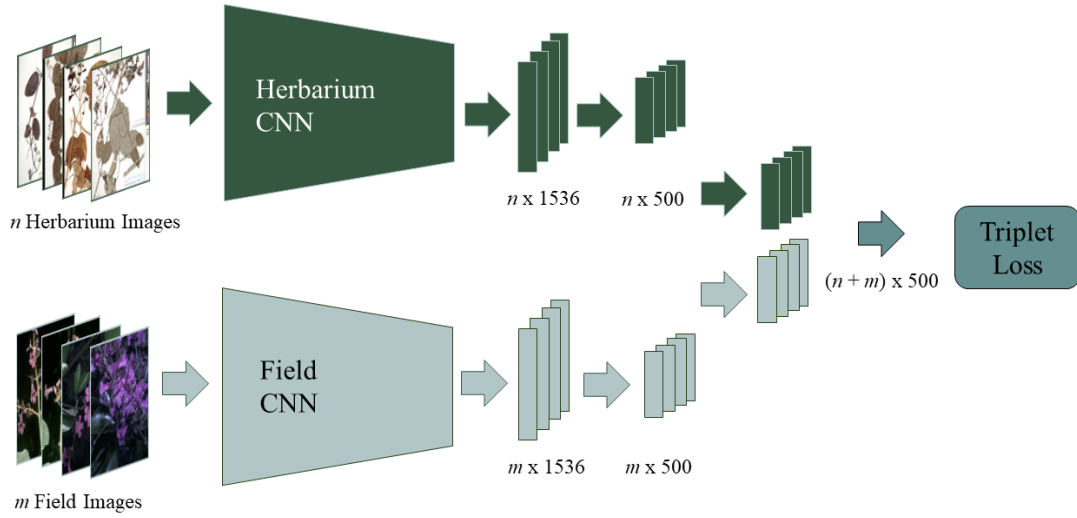
Figure 1: Network Architecture of the Herbarium-Field Triplet Loss Network.

2.1.1. Network 1: Herbarium-Field Triplet Loss Network

This network is made from two CNNs. One being the Herbarium Network, and the other being the Field Network. A batch normalization layer is added at their final embedding layers, and its feature vector is reduced from 1536 to 500. The output is then L2-normalized and concatenated to the output size of $(n + m) \times 500$. n denotes the batch size of the Herbarium Network, while m denotes the batch size of the Field Network. For the ease of implementation, we set n and m to be the same values. The concatenated feature embedding is then passed to the network's triplet loss layer[1] whereby the network optimizes the feature embeddings of herbarium and field to their species. It is trained to minimize the embedding distance of the same herbarium-field pair species while maximizing the embedding distance of different herbarium-field pair species. The network is illustrated in Figure 1.

2.1.2. Network 2: One-streamed Mixed Network

This network on the other hand is based on a single stream CNN approach. However, unlike the objective of a conventional CNN whereby its goal is to map the test data with its learned features, we do not directly map them but utilize the learned features since the training data (herbarium images) and test data (field images) share different feature distributions. Therefore, the learned features of the OSM Network are make used as a means to measure the feature

---

[1]The triplet loss is computed using triplet_semihard_loss function provided in Tensorflow 1.12 [12]
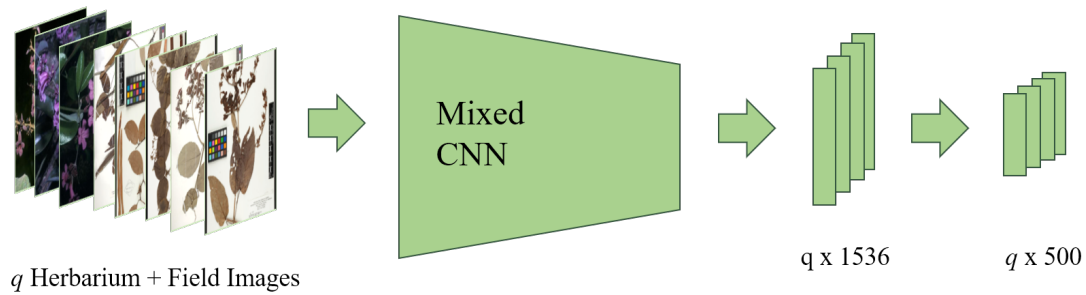
Figure 2: Network Architecture of the One-streamed Mixed Network.

similarity between herbarium-field pairs instead of directly classifying them. Implementing this mechanism allows us to predict classes with missing field images in a similar way the triplet network does. Likewise in the HFTL Network, its feature vector is reduced from 1536 to 500. This network is illustrated in Figure 2.

2.2. Data

The datasets used in our submissions are from PlantCLEF 2021 and PlantCLEF 2017. In PlantCLEF 2021, 997 species have herbarium images, and a subset of 435 species have both herbarium and field images as training data. Since the number of field images in this dataset is significantly less than that of herbarium images, we utilize the field images from PlantCLEF 2017 to allow the network to generalize features of field images better. In addition, PlantCLEF 2021 introduces new data related to five traits that include traits of the plants' growth form, habitat, lifeform, trophic guild, and woodiness. However, we did not apply these traits in the training of our networks.

2.2.1. Data in HFTL Network

This network has two separate streams that take in two different input domains: herbarium and field. In the initial stage of constructing the first stream (Herbarium Network), only PlantCLEF 2021 herbarium dataset is used. In the second stage, PlantCLEF 2017 field dataset is used to train the second stream (Field Network). We have also trained the Field Network with PlantCLEF 2021 field dataset instead, and this comparison is tabulated in Table 4. Finally, in the third stage, where the HFTL Network is established, only herbarium and field images from PlantCLEF 2021 are used.

2.2.2. Data in OSM Network

This network utilizes the training data solely from PlantCLEF 2021. It takes in both herbarium and field images as input without distinction between herbarium and field domains.

Table 1

Dataset Used in Training Networks.

Network	Number of images		Number of classes	
	Herbarium	Field	Herbarium	Field
Herbarium	306,005	-	997	-
Field (2017)	-	1,187,484	-	10,000
Field (2021)	-	4,685	-	435
HFTL	197,985	5,824	435	435
OSM	306,005	4,685	997	435

Table 2

Network Training Parameters.

Parameter	Herbarium, Field, OSM Network	HFTL Network
Batch Size	256	16
Input Image Size	$299 \times 299 \times 3$	$299 \times 299 \times 3$
Optimizer	Adam Optimizer [13]	Adam Optimizer [13]
Initial Learning Rate	0.0001	0.0001
Weight Decay	0.00004	0.00004
Loss Function	Softmax Cross Entropy	Triplet Loss

The overall datatset distribution is summarized in Table 1.

2.3. Training Setup

The networks trained are set up using Tensorflow 1.12 [12] alongside slim packages with hyperparameters as described in Table 2. The codes are available at https://github.com/NeuonAI/plantclef2021_challenge.

3. Experiments

3.1. Dataset

To evaluate the performance of our networks, we segregated a subset of species from the PlantCLEF 2021 dataset. This subset of species was catered for two categories of test sets: (1) with field images in the training data and (2) without field images in the training data. For the species without field training data, we obtained its field images from various resources via Google Images queries to create the test set. These experimented test sets for HFTL Network and OSM Network are detailed in Table 3.

Table 3

Test Set 1 (With Field Training Data) and Test Set 2 (Without Field Training Data).

| Dataset	Number of images	Number of classes
Test Set 1	1,219	345
Test Set 2	197	100

3.2. Inference Procedure

To evaluate the test set, we first generate a herbarium dictionary to store the reference embeddings of all 997 species. Then, the field embedding from the test set is compared with the herbarium dictionary formed to map the field embedding to their herbarium pair. The difference between our method of constructing the herbarium dictionary and our previous method is that instead of 5 corner crops, the extraction of herbarium embeddings was extended to 10 different corner crops. Furthermore, field images were also used to form the herbarium dictionary. The networks that utilized field images in the herbarium dictionary are tabulated in Table 4. The process of herbarium dictionary generation is illustrated in Figure 3 while the comparison of feature similarity is illustrated in Figure 4. The following steps describe the inference procedure:

1. **Generate herbarium dictionary**
    a) Using a predefined herbarium dataset* that contains the herbarium of all 997 species, extract the feature embeddings of each species from the network trained (*Note that this herbarium dataset is later added with field images to compare the effects. The results are seen in Table 4).
    b) Upon extracting the feature embedding of each test sample, apply *Center and Corner Crops* to the image to obtain 5 different images (center, top-left, bottom-left, top-right and bottom-right) from the original sample.
    c) Subsequently, flip those 5 images to obtain a total variety of 10 images.
    d) Hence for each test sample, 10 images are obtained resulting in 10 feature embeddings.
    e) Average the 10 feature embeddings for each sample.
    f) Group the averaged feature embeddings of each sample belonging to the same species.
    g) Average the embeddings to obtain a single feature embedding for each species.
    h) Store the averaged feature embeddings of each species in a herbarium dictionary to be used as reference embeddings.
    i) A herbarium dictionary of 997 feature embeddings is formed.

2. **Compare feature similarity**
    a) Extract the feature embedding of each test image using the same method as the extraction of herbarium dictionary embeddings:
        i. Apply *Center and Corner Crops* on the images before extraction to obtain 5 differnet images (center, top-left, bottom-left, top-right and bottom-right).
        ii. Flip aforementioned 5 images to obtain 10 images (feature embeddings).
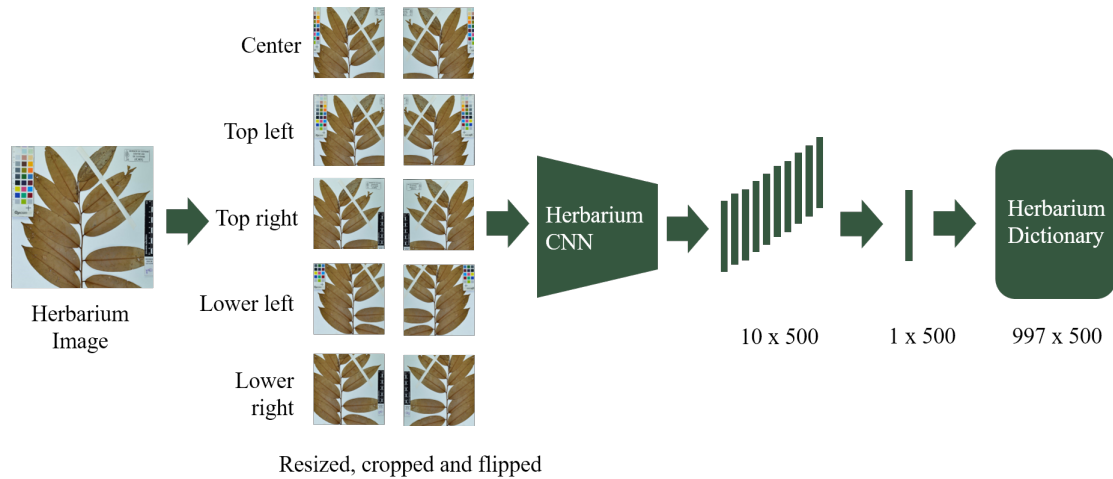
Figure 3: Process of Generating Herbarium Dictionary.

      iii. Average the 10 feature embeddings to obtain a single embedding for each test image.

b) Compute the cosine similarity between the feature embedding of each test image with the reference herbarium dictionary.

c) Subtract the computed cosine similarity from the value of 1 to obtain the cosine distance.

d) Employ inverse distance weighting on the cosine distance.

e) Acquire the probabilities of the test image mapped to the reference herbarium embeddings.

f) The species mapped with the highest probability denotes the class of the species.

3.3. Networks and results

We tested our networks on our test sets described in Table 3. The results are tabulated in Table 4 and the experimented networks are explained as follows:

3.3.1. Network 1: HFTL-I

An HFTL Network based on Inception-v4 in which its Field Network is pretrained from Plant-CLEF 2017.

3.3.2. Network 2: HFTL-I-AUG

An HFTL Network based on Inception-v4 in which its Field Network is pretrained from Plant-CLEF 2017 but with increased augmented training images.
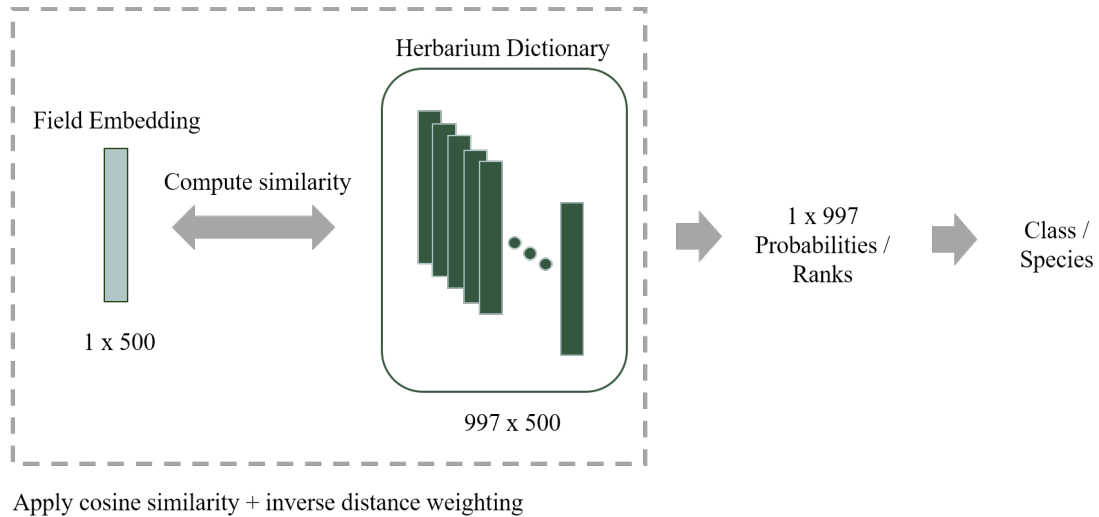
Figure 4: Process of Comparing Feature Similarity.

3.3.3. Network 3: HFTL-I-21

An HFTL Network based on Inception-v4 in which its Field Network is pretrained from Plant-CLEF 2021.

3.3.4. Network 4: HFTL-IR

An HFTL Network based on Inception-ResNet-v2 in which its Field Network is pretrained from PlantCLEF 2017.

3.3.5. Network 5: HFTL-IR-AUG

An HFTL Network based on Inception-ResNet-v2 in which its Field Network is pretrained from PlantCLEF 2017 dataset but with increased augmented training images.

3.3.6. Network 6: OSM-I

An OSM Network based on Inception-v4.

3.3.7. Network 7: OSM-IR

An OSM Network based on Inception-ResNet-v2.

3.4. Discussion

Two different test sets were used in the experiments: one on the species with field images present in training data (Test Set 1) and the other on the species with no field images in training

Table 4
MRR Scores of Experimental Test Sets.

Network	Field in Herbarium Dictionary	Test Set 1	Test Set 2
HFTL-I	No	**0.561**	0.137
HFTL-I (Field)	Yes	0.083	**0.154**
HFTL-I-AUG	No	**0.602**	0.141
HFTL-I-AUG (Field)	Yes	0.096	**0.167**
HFTL-I-21	No	**0.856**	0.071
HFTL-I-21 (Field)	Yes	**0.126**	0.09
HFTL-IR	No	**0.771**	0.129
HFTL-IR (Field)	Yes	0.119	**0.153**
HFTL-IR-AUG	No	**0.523**	0.071
HFTL-IR-AUG (Field)	Yes	0.077	**0.179**
OSM-I	No	**0.56**	0.14
OSM-I (Field)	Yes	0.657	0.052
OSM-IR	No	0.586	0.129
OSM-IR (Field)	Yes	**0.692**	0.055

data (Test Set 2). The extraction of herbarium features (embedding) for the herbarium dictionary was done in two ways: one with solely herbarium images, and the other including field images.

Comparing the results from both test sets, the networks performed better in Test Set 1 as compared to Test Set 2. This is naturally the case as the networks have learned features from seen classes better than unseen classes.

It can be observed that without field images in the herbarium dictionary, HFTL Networks performed better. It is as expected since they were trained without field data in their Herbarium stream. Introducing field images in the herbarium stream potentially caused the learned pairs to break down. In contrast, HFTL Networks performed better with field images used in the herbarium extraction in Test Set 2. Since Test Set 2 only contains unseen classes, it would not suffer from feature breakdown as they have in Test Set 1, instead, due to this exclusion, the overall predicted rank for the unseen class will move up. Although this happened involuntarily in our design, it is worth noting that it could be further improved if we could rework the embedding generation (for future work) to have both advantages encapsulate in one type of embedding generation that suits both Test Set 1 and Test Set 2. Aside from the effect of mixing the field in embedding generation, it can be observed that the herbarium-only embedding generation also gave promising results that support the triplet learning mechanism in general.

On the other hand, OSM Networks achieved a higher MRR score with field images used in the herbarium extraction for Test Set 1. As OSM Networks have learned features from both herbarium and field together, they perform better with field data in seen classes. However, field data in the herbarium dictionary does not help in unseen classes since the unseen herbarium-field pairs do not match the learned herbarium-field pairs.

Among the HFTL Networks, HFTL-I-21 performed the best in Test Set 1 but the poorest in Test Set 2. This is likely because its Field Network was pretrained from PlantCLEF 2021 dataset when the rest was pretrained from PlantCLEF 2017. Since its Field Network was pretrained from PlantCLEF 2021, it is able to perform well in its seen classes rather than its unseen classes.

Meanwhile, for OSM Networks, OSM-IR (Field) performed better in Test Set 1 while OSM-I performed better in Test Set 2.

4. Submission

4.1. Submitted Runs

The team submitted a total of 10 runs based on the networks mentioned in Section 3.3. The submitted runs are described as follows:

4.1.1. Run 1: HFTL-I

This model was based on HFTL-I Network.

4.1.2. Run 2: OSM-ENS

This model was based on an ensemble of OSM-I and OSM-IR Networks.

4.1.3. Run 3: HFTL-I-21

This model was based on HFTL-I-21 Network.

4.1.4. Run 4: HFTL-I-21 + OSM-ENS

This model was an ensemble of Run 2 and Run 3.

4.1.5. Run 5: HFTL-I + OSM-ENS

This model was an ensemble of HFTL-I Network (with 10 corner crops) and Run 2.

4.1.6. Run 6: HFTL-I-AUG + OSM-ENS

This model was an ensemble of HFTL-I-AUG Network (with 10 corner crops) and Run 2.

4.1.7. Run 7: HFTL-I (Field) + HFTL-I-AUG (Field) + HFTL-IR (Field) + OSM-ENS

This model was an ensemble of HFTL-I (with 10 corners + field in dictionary), HFTL-I-AUG (with 10 corner crops + field in herbarium dictionary), HFTL-IR (with 10 corner crops + field in herbarium dictionary), OSM-I (with 10 corner crops), and OSM-IR.

4.1.8. Run 8: HFTL-I (Field) + HFTL-I-AUG (Field) + HFTL-IR (Field)

This model was an ensemble of HFTL-I (with 10 corner crops + field in herbarium dictionary), HFTL-I-AUG (with 10 corner crops + field in herbarium dictionary), and HFTL-IR (with 10 corner crops + field in herbarium dictionary).

4.1.9. Run 9: HFTL-I + HFTL-I-AUG + HFTL-IR + OSM-ENS

This model was an ensemble of HFTL-I (with 10 corner crops), HFTL-I-AUG (with 10 corner crops), HFTL-IR (with 10 corner crops), OSM-I (with 10 corner crops), and OSM-IR.

Table 5
MRR Scores of the Submitted Runs

Run	Network	MRR Whole	MRR Subset
7	**HFTL-ENS + OSM-ENS**	**0.181**	**0.158**
10	HFTL-ENS + OSM-ENS	0.176	0.153
8	HFTL-ENS	0.169	0.150
2	OSM-ENS	0.152	0.117
9	HFTL-ENS + OSM-ENS	0.147	0.129
6	HFTL + OSM-ENS	0.143	0.126
5	HFTL + OSM-ENS	0.137	0.116
4	HFTL + OSM-ENS	0.088	0.073
1	HFTL	0.071	0.066
3	HFTL	0.060	0.056

4.1.10. Run 10: HFTL-I (Field) + HFTL-IR (Field) + HFTL-IR-AUG (Field) + OSM-ENS

This model was an ensemble of HFTL-I (with 10 corner crops + field in herbarium dictionary), HFTL-IR (with 10 corner crops + field in herbarium dictionary), HFTL-IR-AUG (with 10 corner crops + field in herbarium dictionary), OSM-I (with 10 corner crops), and OSM-IR.

4.2. Official Results

Our best-submitted run (Run 7) built from an HFTL-OSM ensembled network achieved an MRR score of 0.181 on the whole test set and an MRR score of 0.158 on the test set with few field training data. The results of the total runs submitted are tabularized in Table 5. The results of the overall participants are summarized in Figure 5 and Figure 6.

4.3. Discussion

Our best model achieved an MRR score of 0.181 on the primary metric, and 0.158 on the secondary metric. Our results show that our methods have improved from our last submission in PlantCLEF 2020 which was 0.121 on the primary metric and 0.108 on the secondary metric. It is also worth noting that for Triplets Learning-only network (HFTL-ENS) - performs better than the one-streamed-only network (OSM-ENS) thus proves that triplet learning is better in handling unknown classes, which further implies that triplet learning produces more generalized features compared to conventional classification network. Last but not least, our triplet network was trained with fewer data compared to its one-stream counterpart as it is only trained using valid herbarium-field pairs. Nevertheless, our triplet network still achieved comparable results or even better in some cases.
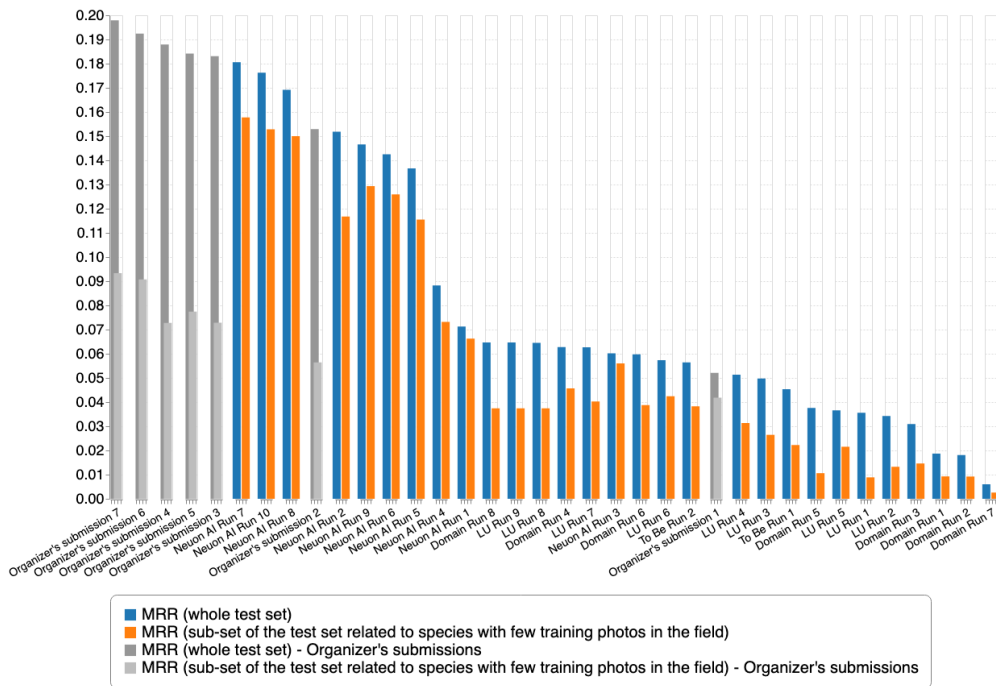
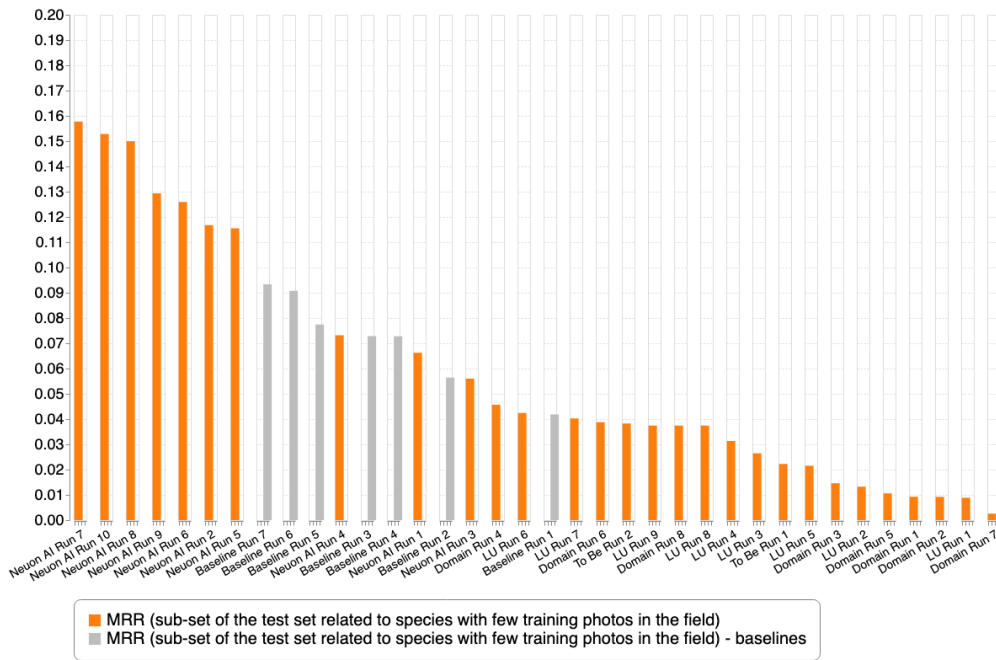**Figure 5:** Official Results of PlantCLEF 2021.

Figure 6: Official Results of PlantCLEF 2021 - Difficult Subset of Test Set.

5. Conclusion

In this paper, we presented our improved version of the Herbarium-Field Triplet Loss Network which aims to tackle cross-domain adaptation in plant identification between herbarium specimens and real-world plant images. Although we have gained improvements over the previous challenge, our MRR score for the primary metric did not surpass the organizer's submission. However, we can observe that the MRR score for the second metric is significantly higher than the organizer's submission. This indicates that our method would be more suitable to identify plants when their field samples are limited but herbarium specimens are available. This mechanism is better at predicting observations of species with missing field images in the training set than traditional CNNs.

As for future works, we would like to take into account the newly provided meta - traits - into our consideration during the learning process. Moreover, we would like to utilize the taxonomy data to further improve predictions.

Acknowledgments

References

[1] S. Chulif, Y. L. Chang, Herbarium-field triplets network for cross-domain plant identification-neuon submission to lifeclef 2020 plant, CLEF working notes (2020).

[2] 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. Durso, 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.

[3] 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.

[4] S. Kahl, T. Denton, H. Klinck, H. Glotin, H. Goëau, W.-P. Vellinga, R. Planqué, A. Joly, Overview of birdclef 2021: Bird call identification in soundscape recordings, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.

[5] T. Lorieul, E. Cole, B. Deneu, M. Servajean, A. Joly, Overview of geolifeclef 2021: Predicting species distribution from 2 million remote sensing images, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.

[6] L. Picek, A. M. Durso, R. Ruiz De Castañeda, I. Bolon, Overview of snakeclef 2021: Automatic snake species identification with country-level focus, in: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, 2021.

[7] A. Joly, B. Deneu, S. Kahl, H. Goëau, R. Ruiz De Castaneda, J. Champ, I. Eggel, E. Cole, P. Bonnet, C. Botella, A. Dorso, H. Glotin, T. Lorieul, M. Servajean, F.-R. Stöter, W.-P. Vellinga, H. Müller, Lifeclef 2020: Biodiversity identification and prediction challenges, in:

Proceedings of CLEF 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece., 2020.

[8] N. H. Krishna, M. Rakesh, R. Ram Kaushik, Plant species identification using transfer learning-plantclef 2020, CLEF working notes (2020).

[9] J. Villacis, H. Goëau, P. Bonnet, E. Mata-Montero, A. Joly, Domain adaptation in the context of herbarium collections: a submission to plantclef 2020, CLEF working notes (2020).

[10] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 815–823.

[11] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[12] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.

[13] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).