

# Herbarium-Field Triplet Network for Cross-Domain Plant Identification NEUON Submission to LifeCLEF 2020 Plant

Sophia Chulif and Yang Loong Chang

Department of Artificial Intelligence, NEUON AI, 94300 Sarawak, Malaysia

<https://neuon.ai/>

{sophiadouglas,yangloong}@neuon.ai

**Abstract.** This paper presents the implementation and performance of a Herbarium-Field triplet loss network to evaluate the herbarium-field similarity of plants which corresponds to the cross-domain plant identification challenge in PlantCLEF 2020. A two-streamed triplet loss network is trained to maximize the embedding distance of different plant species and at the same time minimize the embedding distance of the same plant species given herbarium-field pairs. The team submitted seven runs which achieved a Mean Reciprocal Rank score of 0.121 and 0.111 for the whole test set and the sub-set of the test set respectively.

**Keywords:** Cross-domain plant identification, computer vision, triplet loss, convolutional neural networks

## 1 Introduction

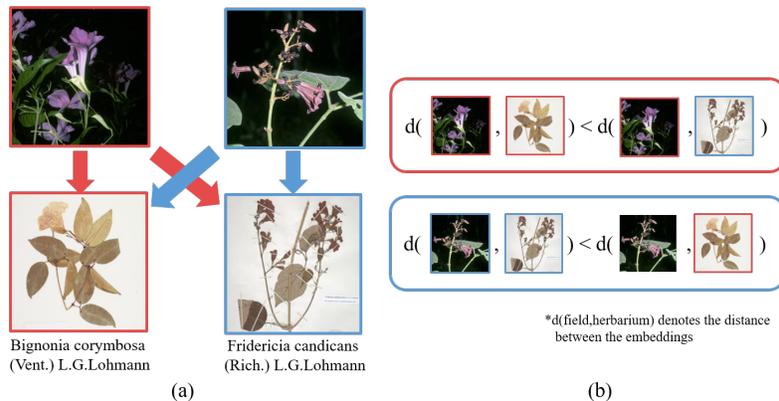
Plant specimens in herbaria have been used by novices and experts alike to study and confirm plant species as well as many other useful applications as described in [4]. Many works are being carried out to improve the access and preservation of these specimens as they would be considerably less expensive to obtain rather than field images. Despite its large collection, the application of herbaria specimens on the identification of real-world plants require more research [14].

The objective in PlantCLEF 2020 [5,6] involves a task of cross-domain plant classification between herbarium specimens and field (real-world plant) images. In this paper, we present our approach using a two-streamed network, namely Herbarium-Field triplet loss network to evaluate the similarity of herbarium-field pairs corresponding to the aforementioned task.

We adopt triplet loss function to optimize the plant embeddings which regulates the measure of plant similarity. The implemented network is trained to maximize the embeddings of different herbarium-field species pairs and minimize

---

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessaloniki, Greece.



**Fig. 1.** The triplet loss concept mainly revolves around minimizing the distances between same class and maximizing the distances between different classes. (a) shows two classes with its herbarium counterpart, the image embedding is compared with its own herbarium and the herbarium from another class (as indicated by the arrows). (b) The distances between herbarium-field pairs of the same species has to be less than the herbarium-pairs of different species (red and blue box denotes the class label).

the embeddings of same species pairs. It learns the similarity between herbarium sheets and field images instead of directly classifying plant species as conventional convolutional neural networks (CNN) [9].

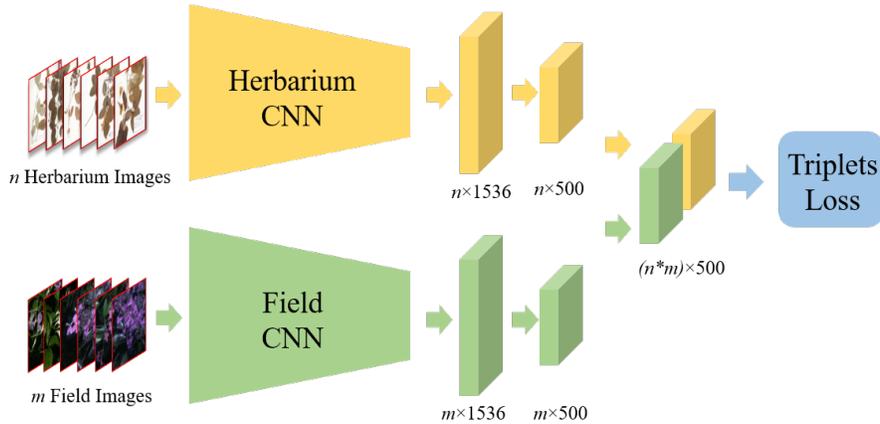
## 2 Related Works

### FaceNet: A Unified Embedding for Face Recognition and Clustering

The authors in [12] introduce triplet loss function that uses a CNN to optimize face embeddings which corresponds to a measure of face similarity. Instead of training an intermediate layer, the embeddings are directly optimized in an Euclidean space for face verification. Likewise, this triplet loss function is adopted in our networks to learn the optimized plant embeddings.

### Plant Disease Recognition with Siamese Network

The authors in [2] introduce Few-Shot Learning algorithms that classify leaf images with deep learning. They employ Siamese Network with triplet loss that shows the possibility of achieving high accuracy with small datasets. In addition, the authors in [3] address the classification problem using real-world images. They also show that the image embeddings extracted from the employed Siamese Network are better than using transfer learning. In the same way, we employed a two-streamed triplet loss network which works similarly to classify plants utilising the herbarium and field embeddings.



**Fig. 2.** Network Architecture of the Herbarium-Field Triplet Loss Network.

### 3 Methodology

This section describes our approach in PlantCLEF 2020, the implemented network architecture and training stages involved. The training process is split into three stages: pre-trained herbarium network, pre-trained field network and two-stream triplet loss network. The Herbarium and Field networks are trained individually to construct networks that could model generalized herbarium and field features. A triplet network is then employed to model the triplets distance between herbarium and field features. The objective is to train the network to behave: (i) herbarium features (or embeddings) of a species should be closer to the field features of the same class (ii) herbarium features of a species should be further from field features of a different class. Fig. 1 illustrates the concept of triplets learning for herbarium-field pairs.

#### 3.1 Network Architecture

The network architecture implemented in our approach is illustrated in Figure 2. This Herbarium-Field triplet loss network is constructed with two Inception-v4 CNNs [13], namely Herbarium CNN and Field CNN which were initialized with weights pre-trained on PlantCLEF 2020 [5] and PlantCLEF 2017 [7] respectively. Both networks are formed to cater for the generalization of herbarium and field features. At the final embedding layer of each network, a batch normalization layer is added and the output is fed into a fully-connected layer. The output size of the fully-connected layer is then reduced from 1536 to 500. Subsequently, these outputs are L2 normalized in the L2 layer and concatenated to give an output size of  $(n * m) \times 500$  whereby  $n$  and  $m$  is the batch size of the Herbarium and Field networks respectively. This concatenated embedding is later passed

into the triplet loss layer<sup>1</sup> through which the network learns to compute the herbarium and field embeddings with respect to their optimum embedding space. The network is trained to maximize the embedding distance of different species in herbarium-field pairs and minimize the embedding distance of the same species. The classification of species is dependent on the computed embedding space by which a large embedding distance denotes different species and a small embedding distance indicates same species. There are two types of training methods investigated i.e., frozen front layers and non-frozen front layers.

**Frozen Front Layers** In this method, the front layers of the pre-trained Herbarium and Field network, or simply, the extractor layer of the network is frozen. This allows only the weights in the newly added layer (triplet loss layer) to be updated.

**Non-Frozen Layers** This method on the other hand trains all layers in the network. It allows the network to relearn and recompute the embeddings of herbarium and field images with respect to their optimized embedding space from the triplet loss. The new layers are set to have a higher learning rate than the migrated layers.

### 3.2 Training stages

**Herbarium Network** As mentioned in 3.1, a Herbarium network based on the Inception-v4 model [13] is set up to make up the Herbarium-Field triplet loss network. The Herbarium network is initialized on weights pre-trained from ImageNet [11] and trained with PlantCLEF 2020 dataset (herbarium images) [5].

**Field Network** Likewise, the Field network adopts the Inception-v4 [13] network architecture. It is also initialized with weights pre-trained from ImageNet [11] but trained with PlantCLEF 2017 dataset (field images) [7] instead.

**Herbarium-Field Triplet Loss Network** Once the Herbarium and Field networks are trained, the Herbarium-Field Triplet Loss network is set up. The network is trained with PlantCLEF 2020 dataset [5] consisting of both herbarium and field images. The network trained in the Non-Frozen Layers setup is set with a learning rate of 0.00001 in the migrated layers and 0.0001 in the newly added layers, whereas the Frozen Front Layers setup is set with a learning rate of zero in the migrated layers.

---

<sup>1</sup> The triplet loss is computed using `triplet_semihard_loss` function provided in Tensorflow 1.13 [1]

**Table 1.** Training dataset distribution for different networks

Network	Number of images		Number of classes	
	Herbarium	Field	Herbarium	Field
Herbarium	305,531	-	997	-
Field	-	1,187,484	-	10,000
Herbarium-Field Triplet Loss	197,552	6,257	435	435

## 4 Training Setup

### 4.1 Data Preparation

As mentioned in the task description, only a subset of species for field images were provided to allow learning a mapping between the herbarium and field domain. We separated the species which possess both herbarium and field images to be used for mapping. Out of 997 classes, 435 classes were identified having both herbarium and field images. These classes were then used for training. Although the total number of classes was reduced from 997 to 435 species, the network was still trained to map the embedding space of 997 classes.

During the training of the Herbarium-Field triplet loss network, the images used for each batch were picked to be balanced for each class. For instance, in a batch of size 16, each class may not comprise more than 4 images, meanwhile the minimum number of images in each class is 2. This allows a balanced selection of anchors for the triplet loss.

### 4.2 Data Augmentation

In order to increase the network generalization and increase training sample size, data augmentation was applied on the training images. Random cropping, horizontal flipping and colour distortion (brightness, saturation, hue, and contrast) of images were performed on the training dataset. As a result, features and various transforms that are invariant to their original locations can be learned by the network, consequently reducing the chance of overfitting [10].

### 4.3 Training Dataset and Hyperparameters

The training dataset distributions and network setup parameters are summarized in Table 1 and Table 2 respectively.

## 5 Experiments

The experiments were conducted using Tensorflow 1.13 [1] alongside slim packages. The codes are available at [https://github.com/NeuronAI/plantclef2020\\_challenge](https://github.com/NeuronAI/plantclef2020_challenge)

**Table 2.** Network training parameters

Parameter	Herbarium and Field Network	Herbarium-Field Triplet Loss Network
	Value	Value
Batch Size	256	16
Input Image Size	$299 \times 299 \times 3$	$299 \times 299 \times 3$
Optimizer	Adam Optimizer[8]	Adam Optimizer[8]
Initial Learning Rate	0.0001	0.0001
Weight Decay	0.00004	0.00004
Loss Function	Softmax Cross Entropy	Triplet Loss

### 5.1 Dataset

Due to the limited field training samples, prior to training, a sample of images from each of the “herbarium\_photo\_associations” and “photo” folders were randomly segregated for validation purposes. 1,219 field images were separated from the test set leaving 5,038 field images for training instead of 6,257 as stated in Table 1. The number of images and classes present in the experimented training and testing dataset are summarized in Table 3. Nevertheless, the class number for the Herbarium-Field triplet loss network remains 997 and 10,000 in the Herbarium and Field network stream respectively.

**Table 3.** Dataset of experimented Herbarium-Field Triplet Loss Network.

Network	Herbarium		Field	
	Train	Test	Train	Test
Dataset				
Number of images	153,867	43,685	5,038	1,219
Number of classes present	435	434	435	345

### 5.2 Inference Procedure

#### Herbarium dictionary

For inference, the embeddings from 997 herbarium classes were first extracted using the trained Herbarium-Field triplet loss network to form the reference embeddings served as a herbarium dictionary. Random samples from each class were picked and fed into the network to obtain the embeddings. The extracted embeddings were then averaged to get a single embedding representation for each class. The embedding for each class was subsequently saved as a dictionary.

Note that the extraction was done with two different types of image cropping, namely, *Center Crop* and *Center and Corner Crop*. The *Center Crop* approach crops the centre region of the herbarium sample. Meanwhile, the *Corner Crop* approach on the other hand crops the top left, top right, bottom left, and bottom right region of the herbarium sample. Each region was cropped and resized then

**Table 4.** Validation Accuracy with *Center Crop* Herbarium Extraction.

Networks	Top 1 Center Crop	Top 1 Center Crop + Corner Crop	Top 5 Center Crop	Top 5 Center Crop + Corner Crop
FL	27.48 %	28.63 %	50.78 %	52.42 %
NFL	32.65 %	32.73 %	59.97 %	58.98 %
<b>NFL-ENS</b>	<b>36.42 %</b>	<b>37.33 %</b>	<b>65.14 %</b>	<b>67.51 %</b>
NFL-AUG	18.05 %	18.46 %	42.49 %	42.49 %
<b>NFL-AUG-ENS</b>	<b>36.42 %</b>	<b>37.33 %</b>	<b>65.14 %</b>	<b>67.51 %</b>

**Table 5.** Validation Accuracy with *Center and Corner Crop* Herbarium Extraction.

Networks	Top 1 Center Crop	Top 1 Center Crop + Corner Crop	Top 5 Center Crop	Top 5 Center Crop + Corner Crop
FL	27.40 %	29.20 %	50.78 %	52.17 %
NFL	33.06 %	34.29 %	59.80 %	58.98 %
<b>NFL-ENS</b>	<b>36.10 %</b>	<b>37.57 %</b>	<b>63.82 %</b>	<b>66.45 %</b>
NFL-AUG	18.29 %	18.79 %	41.84 %	42.74 %
<b>NFL-AUG-ENS</b>	<b>36.10 %</b>	<b>37.57 %</b>	<b>63.82 %</b>	<b>66.45 %</b>

passed into the network for the extraction of herbarium embeddings.

### Feature similarity

After obtaining the single embedding representation of each class, the saved dictionary is then used to compare the embedding distance between the 997 herbarium representation and the test image. During validation, *Center and Corner Crop* were also applied together with horizontal flip in obtaining the test images' embeddings. This resulted in 10 different variations for each image which was then averaged to obtain their similarity probability. Cosine similarity was used as the distance metric in measuring the embedding similarity. Then, the cosine distance was obtained by subtracting the cosine similarity from 1. Finally, inverse distance weighting was performed on the cosine distance to obtain the probabilities of each class.

### 5.3 Network and Results

The experimented results are tabulated in Table 4 and Table 5 for *Center Crop* and *Center Crop and Corner Crop* herbarium extraction methods respectively. The networks were tested on the same validation set of 1,219 images in which the Top 1 and Top 5 predictions were evaluated. *Center Crop* and *Corner Crop* were also applied on the field test set before validation. **5 different Herbarium-Field triplet loss networks were experimented, i.e.:**

**Network 1: Frozen Front Layers (FL)** A network trained with frozen front layers.

**Network 2: Non-Frozen Layers (NFL)** A network trained with non-frozen layers, or to put simply, trained with all layers.

**Network 3: Non-Frozen Layers Ensemble Model (NFL-ENS)** A ensemble of 3 different models trained on all layers.

**Network 4: Non-Frozen Layers Increased Augmentation (NFL-AUG)** A network trained with all layers whereby the training images were pre-processed with more transformations and augmentation.

**Network 5: Non-Frozen Layers Increased Augmentation Model Ensemble (NFL-AUG-ENS)** An ensemble of Network 3 and Network 4.

## 5.4 Discussion

From the experiments, it can be seen that the NFL ensemble models performed the best among the networks. The ensemble of these networks increased the robustness of the system and returned better predictions. On the other hand, the FL network performed the worst among the networks. It can be suggested that the training of all layers does help the prediction model instead of freezing the front layers or extractor layers of the network. It can be seen that the ensemble models with increased augmentation performed equally as to the ensemble model without increased augmentation. It can be suggested that the increased augmentation may have not produced enough new significant information for the network to learn. Since a portion of field images were separated from the training set to serve as test set, some of the classes may miss some field information. In addition, the trained model does not represent the entire classes as some classes miss field images. Consequently, the networks did not performed as well as it was not fed with sufficient images to represent the field domain. An approach to increasing the prediction accuracy would be increasing the training samples of the field images that are not present in the training set.

## 6 Submission

### 6.1 Inference Procedure

The procedure adopted to produce the submitted results are as follow:

- (i) Construct herbarium dictionary by extracting samples of herbarium embeddings for all 997 plant species using the trained Herbarium-Field triplet loss network.
  - (a) Apply *Center and Corner Crops* on the images before extraction.
  - (b) Average the cropped herbarium embeddings for each species and save them.

- (ii) Group the test images belonging to the same observation ID.
- (iii) For each image under the same observation ID, apply *Center and Corner Crops* which result in 5 images each.
- (iv) Subsequently flip the images horizontally resulting in 10 images each.
- (v) Average the 10 images and pass them to the Herbarium-Field triplet loss network.
- (vi) Obtain the image embeddings.
- (vii) Compute cosine similarity between each of the extracted embeddings with the saved 997 herbarium embeddings.
- (viii) Obtain cosine distance by subtracting the cosine similarity from the value of 1.
- (ix) Apply inverse distance weighting on the cosine distance.
- (x) Obtain the probabilities of the embedding distance.
- (xi) Average the probabilities over the total number of images for each observation ID.
- (xii) Repeat steps (iii) to (xii) for the remaining observation IDs.
- (xiii) Collect the predictions, probabilities and ranks for each observation ID.

## 6.2 Submitted Runs

The team submitted a total of seven runs based off the networks mentioned in Section 5.3.

**Run 1** This model was based off (FL). Unlike the rest of the runs, this network was trained with frozen front layers and does not apply image flipping during validation. Moreover, the embedding distances were normalized, inversed then applied with softmax to obtain the probabilities. In addition, the probabilities were based off the averaged embedding instead of all embeddings for each observation ID.

**Run 2** This model was based off (NFL). Similar to Run 1 however it was trained with all layers of the network, the embeddings of each observation IDs were averaged and then applied with Cosine Similarity and Inverse Distance Weighting to obtain the probabilities.

**Run 3** This model was based off (NFL). Similar to Run 2 however by using Cosine Similarity and Inverse Weighting, the probabilities of each embeddings were first computed then averaged for each observation IDs .

**Run 4** This model was based off (NFL). Similar to Run 3 however the probabilities take into account the total embeddings of each observation IDs multiplied by their croppings which consist of 10 variations.

**Table 6.** MRR Score of the Submitted Runs

Run	MRR Whole	MRR Sub-Set
<b>7</b>	<b>0.121</b>	0.107
<b>5</b>	0.111	<b>0.108</b>
3	0.103	0.094
2	0.099	0.076
6	0.093	0.066
4	0.088	0.073
1	0.081	0.061

**Run 5** This model was based off (NFL-ENS). Unlike Run 1 to 4, the network was trained together with the full dataset as stated in Table 1. It is also an ensemble of the predictions from 3 models of the same network.

**Run 6** This model was based off (NFL-AUG). Similar to Run 5 which was trained with the full dataset however it is not an ensemble of models and trained with increased image processing transformations and augmentations.

**Run 7** This model was based off (NFL-AUG-ENS). This run is the ensemble of the predictions from Run 5 and Run 6.

### 6.3 Submission Results

Our best submitted runs scored a Mean Reciprocal Rank (MRR) of 0.121 and 0.108 for the first and second metric respectively. Our results are tabulated in Table 6. The results by all the participating teams are summarised in Fig. 3 and Fig. 4.

### 6.4 Discussion

Similar to the experiment results, the ensemble models performed the best among the networks. The ensemble model with increased augmentation on the other hand performed best in the whole test set. In addition, the MRR score of the networks for the first and second metric are relatively close despite the few training photos in the sub-set species. It can be suggested that the number of training samples for each class does not directly influence the performance of the model. Other than filling the missing training samples of the field classes, the methods in obtaining the herbarium embedding representation can also be looked into to increase prediction accuracy. Such methods involve finding the best herbarium dictionary representation. Various image processing methods like flipping can be performed before extracting the herbarium embeddings. Meanwhile, finding the best model of the Herbarium-Field Triplet Loss Network and using it for the extraction of the herbarium embeddings would be significant as well.

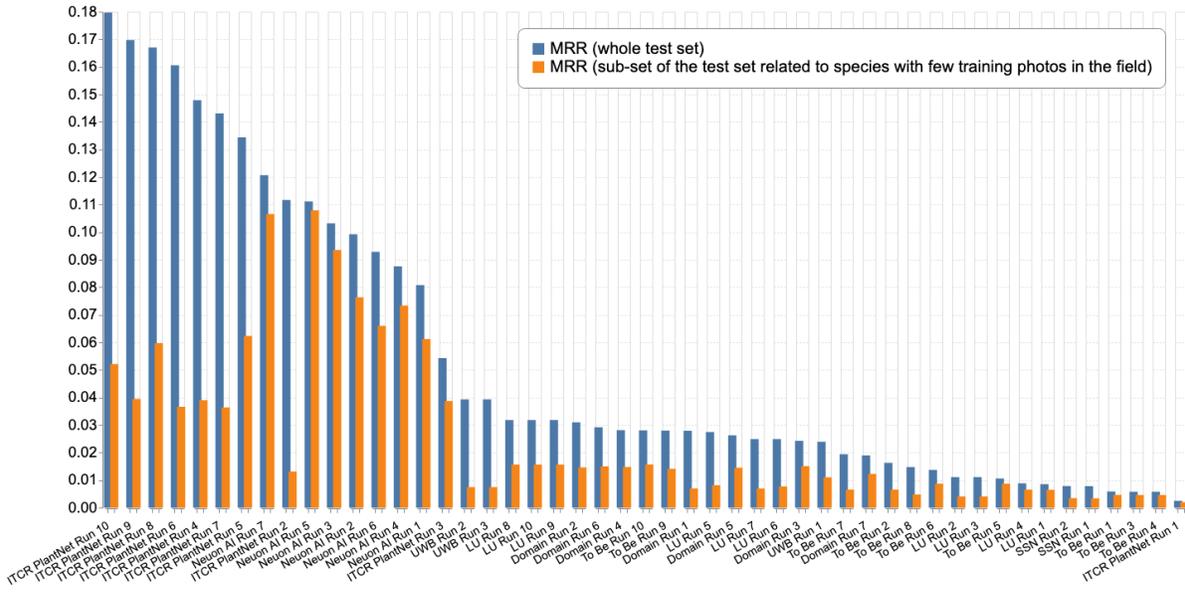


Fig. 3. Official Results of PlantCLEF 2020.

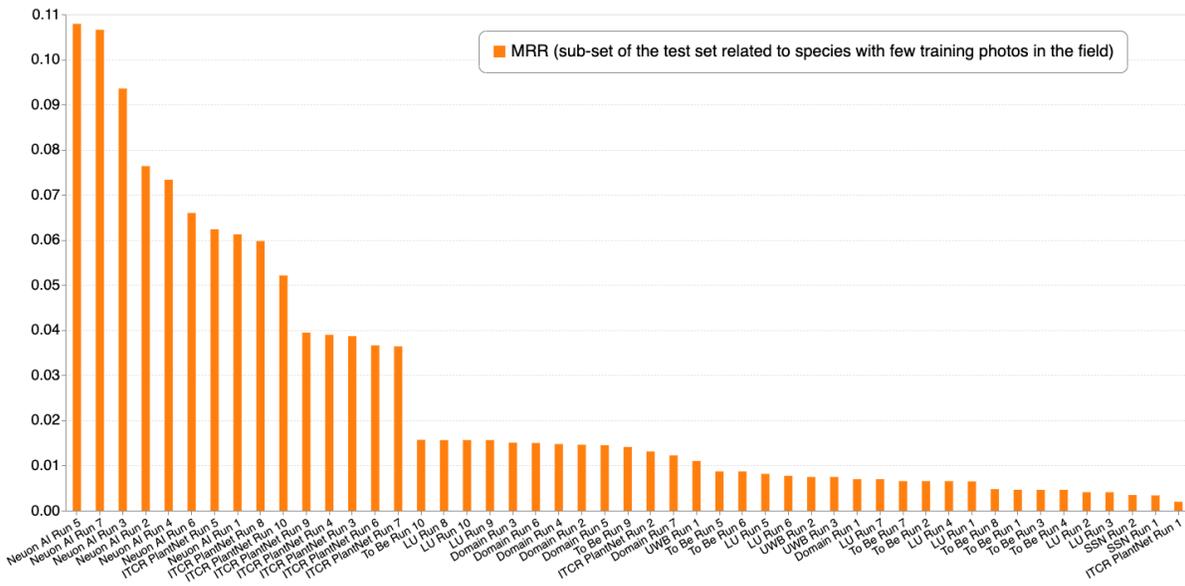


Fig. 4. Official Results of PlantCLEF 2020 (Second Metric Evaluation).

**Table 7.** MRR Score of Post-challenge Runs

Run	MRR Whole	MRR Sub-Set
8	0.101	0.094
<b>9</b>	<b>0.114</b>	0.105
<b>10</b>	0.110	<b>0.107</b>

**Table 8.** Post-challenge Validation Accuracy with *Center Crop* Herbarium Dictionary.

Run	Top 1 Center Crop	Top 1 Center Crop + Corner Crop	Top 5 Center Crop	Top 5 Center Crop + Corner Crop
<b>8</b>	<b>44.71%</b>	<b>45.94%</b>	<b>75.80%</b>	<b>77.19%</b>
9	36.42%	37.33%	65.14%	67.51%
10	36.42%	37.33%	65.14%	67.51%

**Table 9.** Post-challenge Validation Accuracy with *Center and Corner Crop* Herbarium Dictionary.

Run	Top 1 Center Crop	Top 1 Center Crop + Corner Crop	Top 5 Center Crop	Top 5 Center Crop + Corner Crop
<b>8</b>	<b>46.02%</b>	<b>48.32%</b>	<b>74.98%</b>	<b>76.95%</b>
9	36.10%	37.57%	63.82%	66.45%
10	36.10%	37.57%	63.82%	66.45%

## 7 Post-challenge Runs

In addition to the submitted results, the team trained another 3 runs which was based off the continuation of Run 6. However, the results did not performed better than the submitted runs. Since the runs were trained with the whole dataset, we believe the drop in performance is due to overfitting as there was no baseline to determine when to stop training the model. The MRR score of the runs are tabulated in Table 7.

**Run 8** This model was based off (NFL-AUG). This run was a continuation of the training from Run 6 which was trained with increased iterations.

**Run 9** This model was based of (NFL-AUG-ENS). This run was an ensemble of Run 8 and Run 5 predictions.

**Run 10** This model was based off (NFL-ENS). This run was an ensemble of 3 different models from Run 8.

We tested the post-challenge runs on our segregated test set as well and the results are tabulated in Table 8 and Table 9 for *Center Crop* and *Center and Corner Crop* herbarium dictionary construction methods respectively. In contrast

with its MRR score, Run 8 shows the best performance in the experimental validation setup when in fact it performed the worst among the post-challenge runs. This is likely due to overfitting as mentioned.

## 8 Conclusion

In this paper we have presented our approach in PlantCLEF 2020 which focused on the cross-domain plant identification between herbarium sheets and in-field photos. We adopted a two-streamed Herbarium-Field triplet loss network which performed relatively equal regardless if few field training images were given. Based on the similar score between MRR metric 1 and 2, it is proven that the proposed network feature is not directly affected by the plant class but it learns to perceive the similarity between a given field image with herbarium images. It is shown that even with a minimal amount of field images for each species, cross-domain plant identification can be performed. The identification of real-world plants based on herbarium sheets alone is indeed a challenging task. Although our machines did not performed as well with missing field classes which is the case in real-world, it shows that with sufficient data, it offers a step in alleviating the tedious task of herbarium-field classification which requires high level expertise. For future work, the field images that are not present among the training dataset can be added to improve the predictions. This would allow the model to learn the whole representation of plant species with respect to their herbarium and field domain. Furthermore, the extraction of herbarium embeddings to form a more powerful dictionary can be investigated to find the best representation of herbarium embeddings for the herbarium-field similarity comparison.

## Acknowledgment

The resources of this project is supported by NEUON AI SDN. BHD., Malaysia.

## References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <https://www.tensorflow.org/>, software available from tensorflow.org
2. Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M.G., Bereciartua, A., Alvarez-Gila, A.: Few-shot learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture* 175, 105542 (2020)

3. Chandra, M., Patil, P.S., Roy, S., Redkar, S.S.: Classification of various plant diseases using deep siamese network (2020)
4. Funk, V.A.: 100 uses for an herbarium: well at least 72. *American Society of Plant Taxonomists Newsletter* (2003)
5. Goëau, H., Bonnet, P., Joly, A.: Overview of the lifeclef 2020 plant identification task. In: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
6. Joly, A., Deneu, B., Kahl, S., Goëau, H., Ruiz De Castaneda, R., Champ, J., Eggel, I., Cole, E., Bonnet, P., Botella, C., Dorso, A., Glotin, H., Lorieul, T., Servajean, M., Stöter, F.R., Vellinga, W.P., Müller, H.: 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)
7. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Lombardo, J.C., Planque, R., Palazzo, S., Müller, H.: Lifeclef 2017 lab overview: multi-media species identification challenges. In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. pp. 255–274. Springer (2017)
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105 (2012)
10. Mikołajczyk, A., Grochowski, M.: Data augmentation for improving deep learning in image classification problem. In: *2018 international interdisciplinary PhD workshop (IIPhDW)*. pp. 117–122. IEEE (2018)
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: Imagenet large scale visual recognition challenge. *International journal of computer vision* 115(3), 211–252 (2015)
12. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 815–823 (2015)
13. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence* (2017)
14. Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P.: Automated plant species identification—trends and future directions. *PLoS computational biology* 14(4), e1005993 (2018)