Domain Adaptation in the context of herbarium collections

A submission to PlantCLEF 2020

Juan Villacis[1], Hervé Goëau[2,3], Pierre Bonnet[2,3], Alexis Joly[4], and Erick Mata-Montero[1]

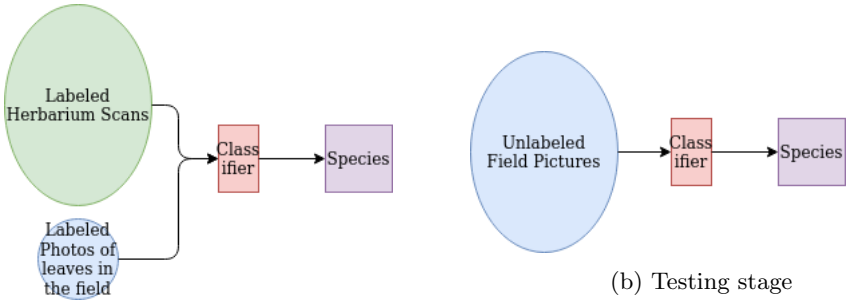
[1] Costa Rica Institute of Technology, Cartago, Costa Rica
jvillacis@ic-itcr.ac.cr, emata@itcr.ac.cr
[2] CIRAD, UMR AMAP, France, herve.goeau@cirad.fr, pierre.bonnet@cirad.fr
[3] AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France
[4] INRIA, Zenith Team, UMR LIRMM, Montpellier, France,alexis.joly@inria.fr

Abstract. This paper describes a submission to the PlantCLEF 2020 challenge, whose topic was the classification of plant images in the field, based on a dataset composed mainly of herbaria.. This work proposes the usage of domain adaptation techniques to tackle the problem. In particular, it makes use of the Few-Shot Adversarial Domain Adaptation method proposed by Motiian et al. (9). Additionally, a modification of this architecture is proposed to take advantage of upper taxa relations between species in the dataset. Experiments performed show that domain adaptation can provide very significant increases in accuracy when compared with traditional CNN-based approaches.

1 Introduction

Recent approaches to automated plant identification have relied on deep learning-based techniques (1). These techniques can be very effective and compete with human experts if a large amount of labeled data is available, even if it is partially noisy (2; 6). However, in the path towards achieving the goal of universal plant species identification, a significant obstacle is posed by the large number of species for which there are none or very few samples of their appearance in their natural state, thus making it very difficult to use this kind of methods. Carrying out missions to collect more data, typically in tropical regions, is not a feasible solution due to the elevated cost, difficulty to access the areas where the species are located and vast amount of data still not labeled. Nonetheless, vast amounts of data about these species exist in the form of herbarium sheets, collected over centuries by botanists and which has been recently massively digitized and published online. This is the topic of the PlantCLEF 2020 challenge

[5]. Given a large dataset of digitized herbarium sheets and very few photos in the field, the objective is to develop a classifier that can perform well on a test set consisting only of field photos after being trained primarily on herbariums. This article describes in detail the methods used for our submissions to the challenge (identified by the acronym *aabab* on the challenge web page[6]).



(a) Training stage

(b) Testing stage

Fig. 1: Description of the problem

2 Methodology

2.1 Data

Dataset The main dataset to be used is PlantCLEF 2020 (3), (8). This dataset has 320,752 herbarium images from 997 species, 4,482 field images from 375 species and 1,816 images from 244 species where for each specimen there are images both in its natural state and in herbarium sheets. In addition to this dataset, some experiments will include additional data from sources like GBIF[7] and PlantCLEF 2019 (7). These images come from the dataset used by (11). Figure 2 shows examples of images in the PlantCLEF 2020 dataset. As can be observed in these examples, the herbarium and field pictures differ greatly, even if they come from the same species (and even the same specimen). This aspect makes the task at hand a particularly challenging one.

2.2 Architecture and Models

Convolutional Neural Networks A common way to perform classification is to take a pretrained CNN and to re-train it on the new target classes. As mentioned earlier, this approach usually requires vast amounts of data. Given

[5] https://www.imageclef.org/PlantCLEF2020
[6] https://www.aicrowd.com/challenges/lifeclef-2020-plant/submissions
[7] https://www.gbif.org

Fig. 2: Herbarium and Field images from the *Asystasia gangetica (L.) T.Anderson* specie. Significant visual variations between them in color, background and shape can be observed even though they are from the same species

that the training set is comprised mostly of herbariums and the test set of field photos, we expect the performance of such an approach to be low. This reason motivates us to look for alternate solutions (which will be described in the next sections), but it is still necessary to measure the performance of the baseline CNN approach. Therefore, the submitted runs also include experiments with a CNN-based approach. The architecture chosen is Resnet50 (4) to maintain the same conditions as those used in the other experiments. Experiments will be performed using the PlantCLEF2020 dataset solely, or the union of all the datasets described in section 2.1. To take advantage of the data available, training will be performed in three stages. First the model will be trained on the ImageNet dataset, in a second stage only the herbariums will be used and in the last stage only the photos will be used.

Domain Adaptation Architectures To tackle the problem of having very few photos in the field domain we base our solution on the architecture presented in (9). This architecture which was devised to tackle the problem of few-shot domain adaptation has the following elements (see Figure 3):

- a CNN-based feature extractor E that maps from the source dataset (herbaria) and target dataset (field images) into a common space, in which it is expected for the features represented to be independent from the original domain.
- a classifier F that performs species classification on the common space
- a discriminator D that determines to which of the following categories a pair of samples from the common space belongs to

1. Samples from different domains and different classes
2. Samples from different domains but the same class
3. Samples from the same domain but different classes
4. Samples from the same domain and the same class

The division into these four categories instead of just into two categories determined by the domain is done to take advantage of label information in the target domain (9).

The feature extractor and the classifier are trained in an adversarial approach with the discriminator in order to guarantee a domain agnostic common space and a robust classifier. In addition to this strategy, data augmentation is done in the target domain in order to complete the feature space with more training samples from this domain.

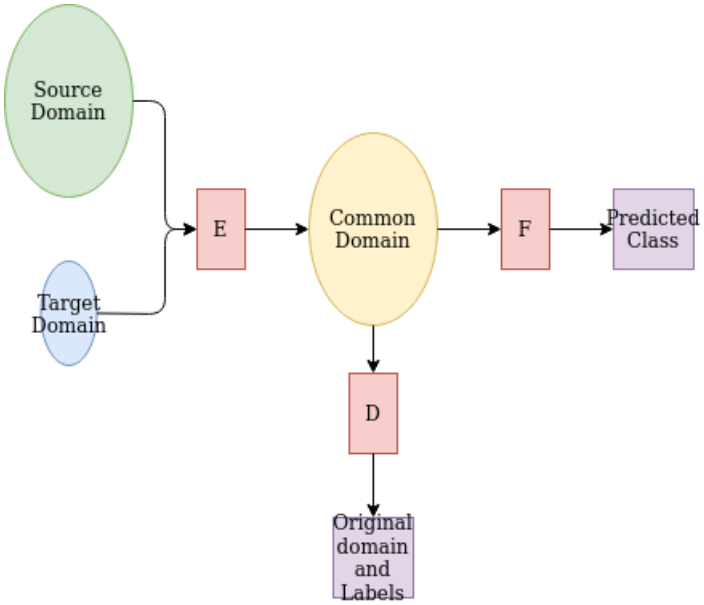


Fig. 3: Approach by (9) to tackle the problem of few-shot domain adaptation. It has an encoder E, classifier F and discriminator D

The training is completed in 3 stages. During the first stage the encoder E and the classifier F are trained in a standard way with samples from only the source domain. In the second stage the discriminator D is trained to distinguish between samples from the 4 categories mentioned before. The objective of the first two stages is to initialize the weights of E F and D. Finally, during the third stage they are all trained together with the objective of performing domain adaptation. It can be said that domain adaptation has been achieved once the discriminator is not able to distinguish samples from categories 1 and 2 and

categories 3 and 4. This means that once the samples have been encoded into the common domain, it is difficult for the discriminator to tell which was the original domain of such sample.

The architecture used has a ResNet50 (4) based encoder, which provides a good compromise between performance, memory use and training time. This is done by removing the last fully connected layer from the ResNet50 architecture. After applying these changes, the dimensionality of the common domain becomes 2048 features. This decision affects the architecture of the classifier and the discriminator. The first one is composed of a single fully-connected layer with 2048 inputs and 997 outputs. The discriminator is a multilayer perceptron, the input is composed of two feature vectors of 2048 features stacked together and it has 6 fully-connected layers that reduce the input size from 4096 features to just 4 outputs. Figure 4 portrays these components.

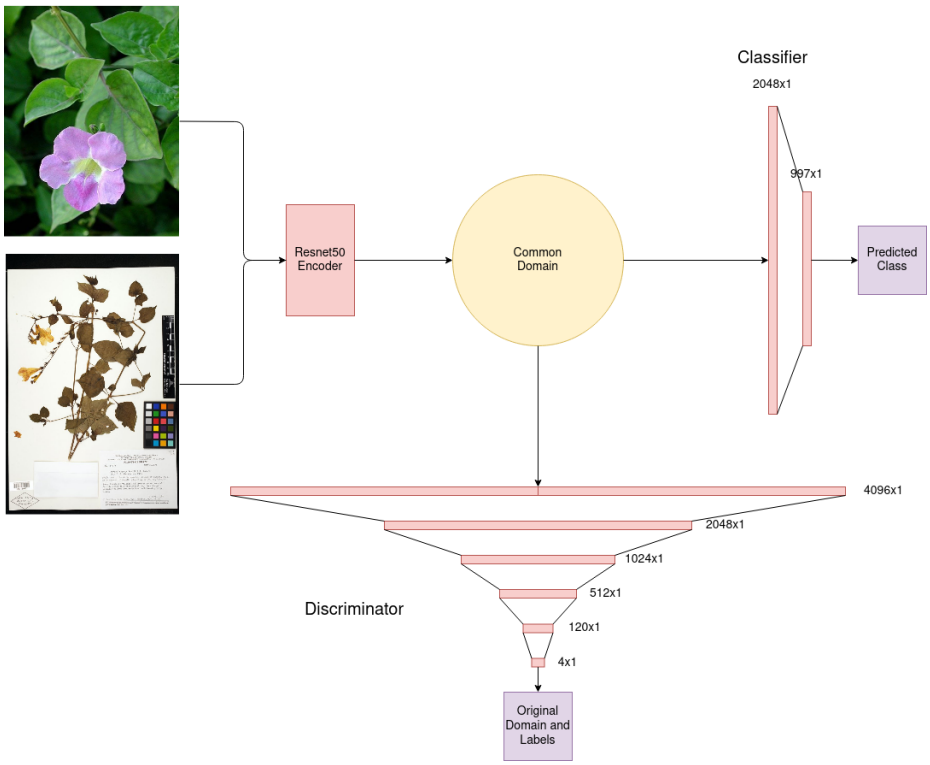


Fig. 4: Details of the FADA architecture used in the experiments

2.3 Additions to the main FADA architecture

Data Augmentation To obtain better results, data augmentation is used to increase the performance of the model. The traditional data augmentation operations used are: random rotations of $\pm$ 15 degrees, color jittering and random horizontal flips. Additionally, special transformations are added for each domain. In the herbariums, a special tilling around the center is added. This operation creates a crop of the original herbarium that is centered around the center and can has a zoom level randomly set between 0.9 and 1.3. This is done because in herbariums the plant samples are commonly placed around the center of the sheet. Examples of this crops can be observed in figure 5. In the field domain, the transformation used is a center crop as large as the original picture permits it to be.

Self Supervision Self supervision is a technique derived from unsupervised learning that tries to address situations found in supervised learning in which there might not be enough labeled data in order to train an efficient model. The objective of these tasks is to extract robust visual information from the pictures which can be useful either as initial weights or to help the main model during training. In our experiments self supervision is used following the ideas presented in (12), where it is used in a multi-task learning approach to help the main classifier. Figure 6 depicts how it is performed in the context of the FADA architecture. Self-supervision is only used during the third stage in the training of the encoder and the classifier. The self-supervision task is applied to the image after it has undergone the data augmentation process. This modification is also used in the traditional CNN approach. In this case, the main model is joined by an additional classifier in a multi-task learning approach. The new classifier tries to predict the correct self-supervision label for the data. In this case, the extra classifier loss is combined with the loss from the main classifier

From the several self supervision tasks in existence (5), we used the jigsaw puzzle solving (10). This decision was taken based on the findings of (12) that when incorporating self-supervision into domain adaptation it is important to choose tasks that do not reinforce domain-dependent features and the fact that the spatial information learned from this showed to be useful when compared to other tasks like recolorization. This task consists in dividing the original image into tiles, rearranging them randomly into one of the 64 possible orderings with the largest distance between them and then having the network try to determine which of the rearrangements is used. Figure 6 shows this process.

Upper taxa Given the nature of the dataset, it is possible to obtain taxonomical information from each species, like the genus or family name (= upper taxa). Because of the lack of data, we try to incorporate this information to the architecture on a multi-task learning approach, so that the features from the common domain are not only used to predict the species name, but also the genus or family of the specimen. This is done in two different approaches. For

Fig. 5: Special tilling used for herbariums to take maximum advantage of the information they have

the FADA architecture, both the classifier and the discriminator are extended with two additional sub-tasks, one for the genus level and one for the family level. These components have the same function as the original species classifier and discriminator, but they have to perform the discrimination and classification tasks with the genus or family instead. It is expected that specimens from the same group share partially similar visual content, and as such this training taking into account upper taxa can be indirectly used to increase performance on specimens that are poorly represented in the dataset but which might have related species in the dataset. In the flowers in figure 7 it is possible to observe the visual similarities between plants from a different specie and the same genus. This is the kind of information we hope to take advantage of.

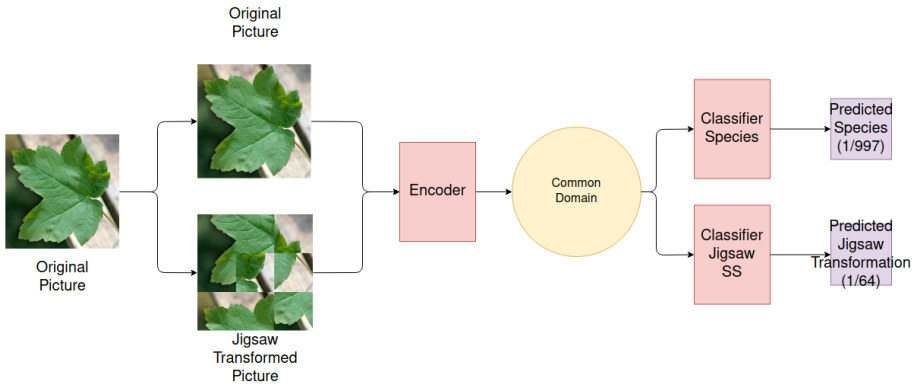


Fig. 6: Addition of jigsaw self-supervision complementary task (10) to the FADA architecture.

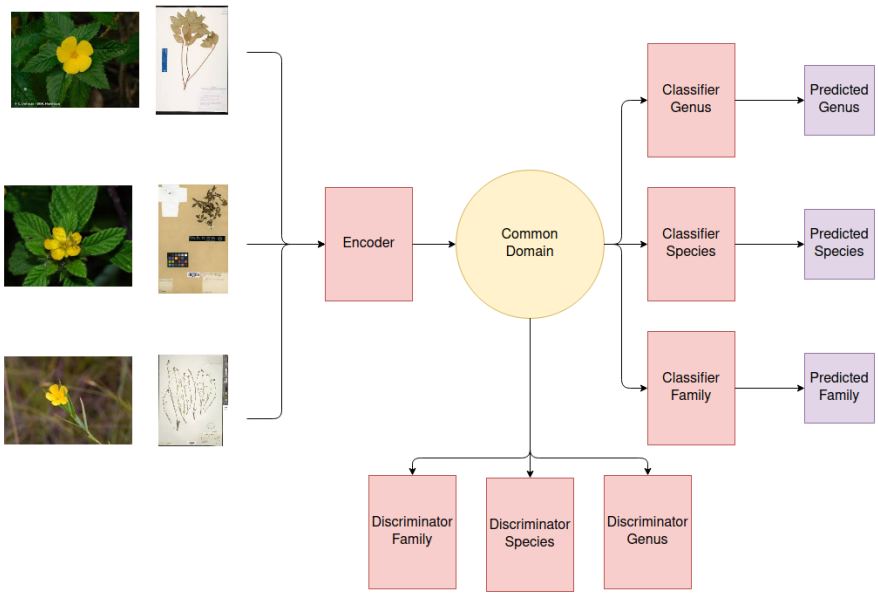


Fig. 7: Multi-task learning with upper taxons

2.4 Training procedure details

Several hyperparameters had to be tuned in order to obtain the best results possible. These are detailed in table 1

Hyperparameter	Value
Framework	Pytorch
Learning rate E and F in FADA	0.001
Learning rate D in FADA	0.001
Learning rate in CNN	0.03
Batch size in FADA	15
Batch size in CNN	64
Number of epochs, stage 1 FADA	60
Number of epochs, stage 2 FADA	60
Number of epochs, stage 3 FADA	30
Number of epochs CNN	60
Learning rate scheduler criterion	Step
Learning rate scheduler gamma	0.1
Learning rate scheduler step in FADA stage 3	15
Learning rate scheduler step in CNN and FADA stage 1	50

Table 1: Hyperparameters used in the training

3 Results

Results from the runs submitted to the challenge can be seen in table 2 and figure 8.

The metric used to present the results of the challenge is the Mean Reciprocal Rank (MRR), which measures the average rank of the correct answer in a series of predictions. It is described by the following formula

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$$

Two distinct MRRs are computed. A first MRR is computed on the full test set. Then, a second one is computed on a subset of the whole test set whose classes have particularly few field (or none) images in the training set.

As can be expected from the inherent difficulty of the challenge, the overall results obtained are low compared to previous editions of PlantCLEF. The method described here obtains the best result on the whole test set, and the second best result on the difficult subset of the test set.

In the runs submitted to the challenge, domain adaptation had a very significant impact on the results. Between the runs with a CNN and those that used this technique there is a 2600% and 1850% increase in the MRR All and the MRR Few. These results can be observed in figure 9.

Additional training data also seems to be a significant factor in obtaining higher values for the evaluation metric. In the MRR All there is a 5500% and a 165% increase in the values when comparing the results from a CNN and FADA

approaches with the same techniques but adding the complementary dataset into the training process. This high increase can be observed in figure 10

The last improvement that this work highlights is the benefit of the proposed extensions of FADA. The usage of self supervision and upper taxa information actually leads to slight but consistent increases of performance. Among this, the most useful to improve the MRR All turns out to be the combination of self supervision and upper taxa, with a 12% increase in the metric. Looking at the MRR on the difficult species, the use of upper taxa alone leads to an increase of 59% on the obtained value. This results can be observed in figure 11 and shows that the visual similarities between species of the same genus or family are particularly useful for species with very few training samples.

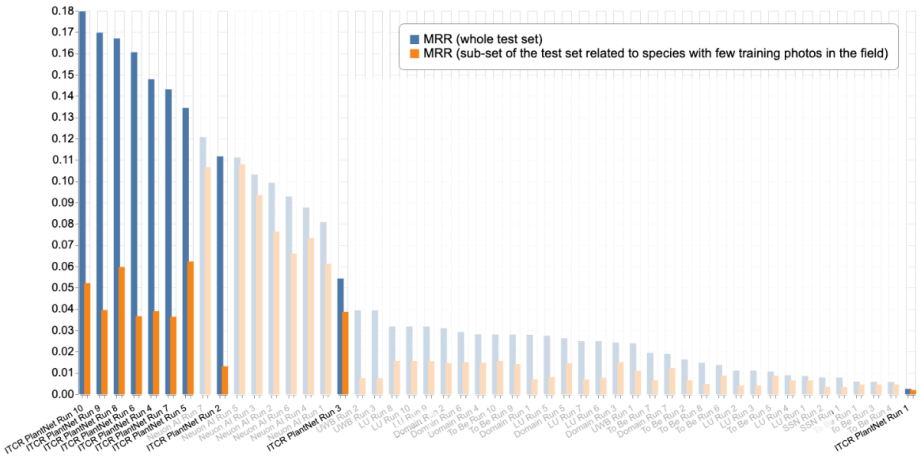


Fig. 8: Results of the PlantCLEF 2020 Challenge, the submissions by the *ITCR Pl@ntNet* team (AIcrowd username *aabab*) are highlighted

4 Conclusions and Future Work

The main conclusions from the experiments are the following

- Domain Adaptation can have a very significant impact in obtaining better results in scenarios where there is a very limited availability of data in one domain but a large dataset on the other
- The addition of extra data showed to be a very significant factor in achieving higher MRRs on the complete test set. On the subset of the most difficult species, however, this conclusion is not as clear-cut. The additional data provided an consistent gain when using the CNN approach, but on the other hand, the gain when using FADA was very small. This is expected to occur due to the fact that FADA is very sensitive to data (even one additional

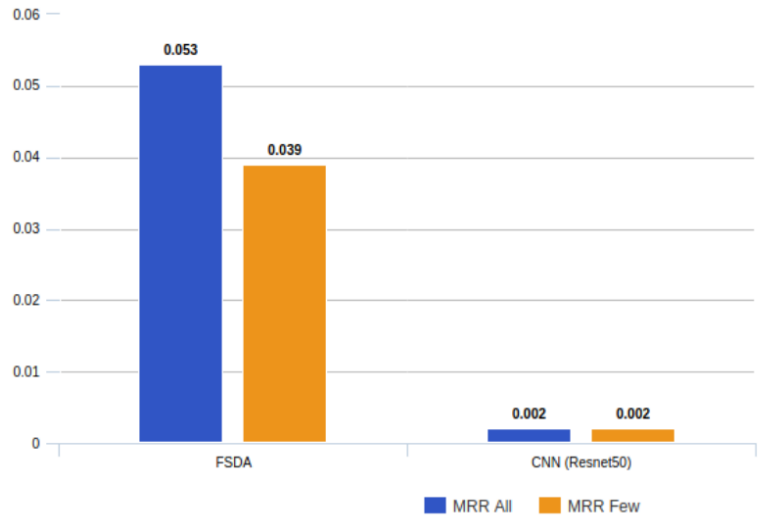


Fig. 9: Impact of domain adaptation on the results

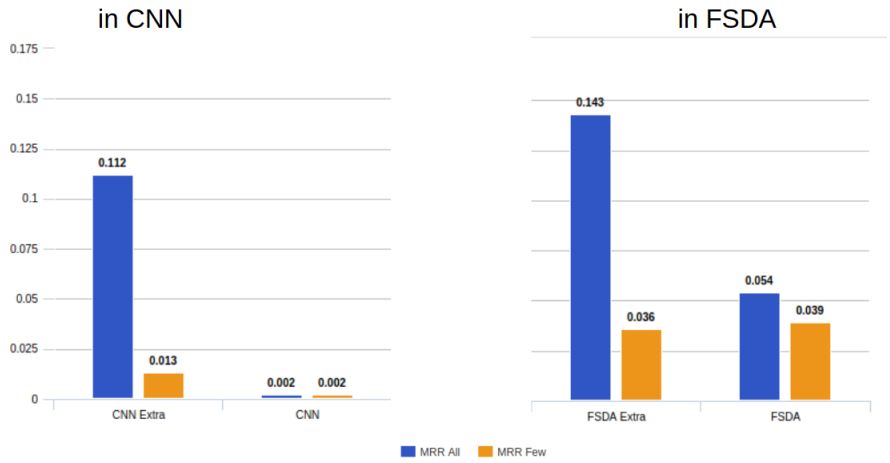


Fig. 10: Impact of extra data on the results

picture in the target domain showed to have a significant effect in modifying the results (9)) and the noisiness in the extra dataset.

– The main modification performed to the FADA architecture, i.e. the introduction of multi-task learning, proved to be important in obtaining better results on both metrics. Extending the classifier and discriminator to extra tasks at upper taxa level was successful for boosting results, in particular on the few shot classes.

Run	MRR	MRR Difficult Species
1. Resnet50 trained on PlantCLEF20	0,002	0,002
2. Resnet50 trained on PlantCLEF20 + extra datasets	0,112	0,013
3. FADA trained on PlantCLEF20	0,054	0,039
4. FADA trained on PlantCLEF20 + extra datasets	0,143	0,036
5. FADA trained on PlantCLEF20 + extra datasets with Self Supervision	0,148	0,039
6. FADA trained on PlantCLEF20 + extra datasets with MTL from genus and family	0,161	0,037
7. FADA trained on PlantCLEF20 + extra datasets with Self Supervision and MTL from genus and family	0,134	0,062
8. Ensemble or runs 6 and 7	0,167	0,06
9. Ensemble or runs 5 and 6	0,17	0,039
10. Ensemble or runs 4, 5, 6 and 7	0,18	0,052

Table 2: Results from the runs submitted to the challenge

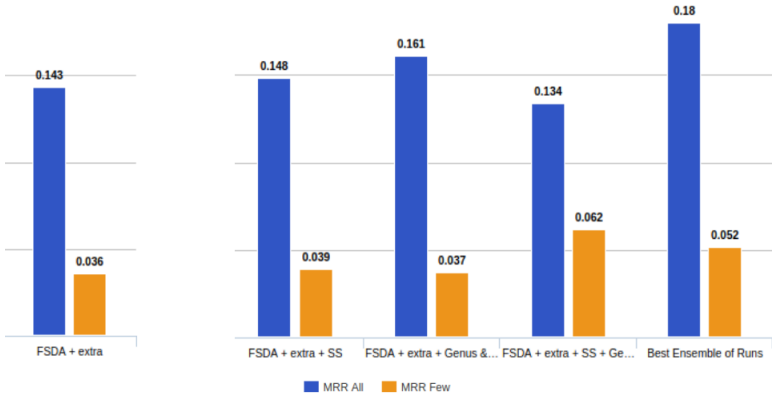


Fig. 11: Impact of performance improving techniques on the results

– As well as upper taxa information, self supervision was an important factor in obtaining increases in performance. Although the increases were smaller, they were nonetheless consistent in particular when combined with multi-task at upper taxa and on the few shot classes.

As future work, some other paths that can be explored are

– Test different botanical or morphological information in addition to taxonomy. For instance, whether a species is Woody/Non-Woody may be an additional task to be solved. The usage of this information is expected to help even more general features that can boost even more the results.

– Exploit additional metadata contained in the dataset like geolocation information or individual pairs. This might require modifications to the architectures used.

Bibliography

[1] Carranza-Rojas, J., Goëau, H., Bonnet, P., Mata-Montero, E., Joly, A.: Going deeper in the automated identification of herbarium specimens. BMC evolutionary biology **17**(1), 181 (2017)

[2] Goëau, H., Bonnet, P., Joly, A.: Plant identification based on noisy web data: the amazing performance of deep learning (lifeclef 2017). In: CLEF task overview 2017, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2017, Dublin, Ireland. (2017)

[3] Goëau, H., Bonnet, P., Joly, A.: Overview of the lifeclef 2020 plant identification task. In: CLEF task overview, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)

[4] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[5] Jing, L., Tian, Y.: Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)

[6] Joly, A., Goëau, H., Botella, C., Glotin, H., Bonnet, P., Vellinga, W.P., Planqué, R., Müller, H.: Overview of lifeclef 2018: a large-scale evaluation of species identification and recommendation algorithms in the era of ai. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 247–266. Springer (2018)

[7] Joly, A., Goëau, H., Botella, C., Kahl, S., Servajean, M., Glotin, H., Bonnet, P., Planqué, R., Robert-Stöter, F., Vellinga, W.P., et al.: Overview of lifeclef 2019: Identification of amazonian plants, south & north american birds, and niche prediction. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 387–401. Springer (2019)

[8] Joly, A., Goëau, H., Kahl, S., Deneu, B., Servajean, M., Cole, E., Picek, L., Ruiz De Castañeda, R., é, Lorieul, T., Botella, C., Glotin, H., Champ, J., Vellinga, W.P., Stöter, F.R., Dorso, A., Bonnet, P., Eggel, I., Müller, H.: Overview of lifeclef 2020: a system-oriented evaluation of automated species identification and species distribution prediction. In: Proceedings of CLEF 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)

[9] Motiian, S., Jones, Q., Iranmanesh, S., Doretto, G.: Few-shot adversarial domain adaptation. In: Advances in Neural Information Processing Systems. pp. 6670–6680 (2017)

[10] Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: European Conference on Computer Vision. pp. 69–84. Springer (2016)

[11] Picek, L., Sulc, M., Matas, J.: Recognition of the amazonian flora by inception networks with test-time class prior estimation. In: CLEF working

notes 2019, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2019, Lugano, Switzerland. (2019)
[12] Sun, Y., Tzeng, E., Darrell, T., Efros, A.A.: Unsupervised domain adaptation through self-supervision. arXiv preprint arXiv:1909.11825 (2019)