Hybridization of deep and prototypical neural network for rare defect classification on aircraft fuselage images acquired by an unmanned aerial vehicle

Julien Miranda, Jannic Veith, Stanislas Larnier, Ariane Herbulot, Michel Devy

Hybridization of deep and prototypical neural network for rare defect classification on aircraft fuselage images acquired by an UAV

Julien Miranda[a,b,c,*], **Jannic Veith**[c,d], **Stanislas Larnier**[c], **Ariane Herbulot**[a,b], **Michel Devy**[a,b]
[a]LAAS, CNRS, 7 avenue du colonel Roche, F-31400 Toulouse, France
[b]Univ de Toulouse, UPS, LAAS, F-31400 Toulouse, France
[c]Donecle, 201 Rue Pierre et Marie Curie, F-31670 Labège, France
[d]Swiss Federal Institute of Technology-ETHZ, CH-8092 Zürich, Switzerland

Abstract. In order to ease visual inspections of exterior aircraft fuselage, new technical approaches have been recently deployed. Automated UAVs are now acquiring high quality images of the aircraft in order to perform offline analysis. At first, some acquisitions are annotated by human operators in order to provide a large dataset required to train machine learning methods, especially for critical defects detection. An intrinsic problem of this dataset is its extreme imbalance (i.e there is an unequal distribution between classes). The rarest and most valuable samples represent few elements among thousands of annotated objects. Deep Learning-only based approaches have proven to be very effective when a sufficient amount of data is available for each desired class, whereas less complex systems such as Support Vector Machine theoretically need less data, and Few-Shot Learning dedicated methods (Matching Network, Prototypical Network, etc.) can learn from only few examples. Those approaches are compared on our applicative case. Preliminary results show the existence of empirical frontiers in term of training dataset volume that indicate which approach might be promoted. We propose a method to combine different approaches in order to achieve best performances on defect classification, that is an extension of previous work.[1]

Keywords: Deep Learning, Few-Shot Learning, Hybrid Model, Defect Detection, Support Vector Machine, Visual Inspection.

***Julien Miranda:** jmiranda@laas.fr

1 Introduction

Visual inspections are one of the most common operations for aircraft maintenance. A major inspection task, performed by maintenance operators, consists in detecting defects on an aircraft fuselage. To do this, they must use mobile elevating platforms to reach positions from where they can properly observe the aircraft skin, looking for those defects. To make those inspections faster, more effective and less painful for human experts, mobile platforms can be used,[2,3] especially automated UAV deployed by the French start-up Donecle, for which a localization with respect to the aircraft can be accurate to a few centimeters, using 3D models as shown in Figure 1.

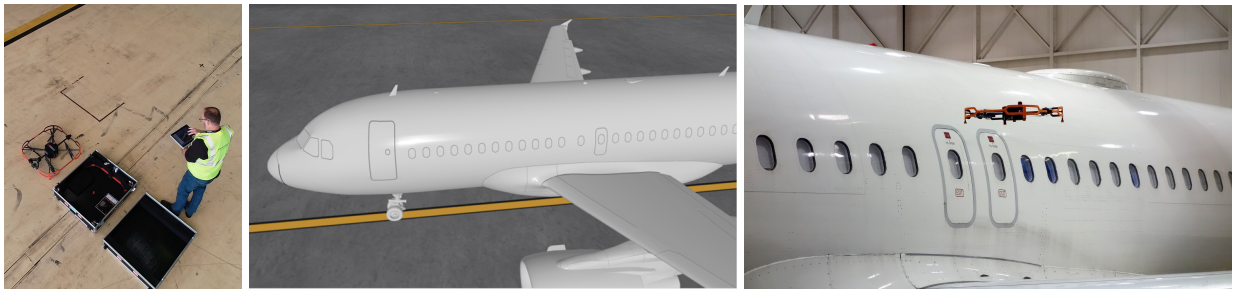

Fig 1 Automated drone inspection from left to right: drone with a tablet running the analysis software application, 3D model used for autonomous localization, drone inspecting aircraft.

Thus using automated drone to acquire images on the whole surface opens new perspectives for aircraft maintenance traceability and automation.[1] Visual inspection from images can be organized in two main tasks: object detection (in our context, an "object" corresponds to a salient image region on the fuselage), and object classification. In this paper we focus on the second task, using of state-of-the art object detection methods: automatic visual inspection relies on the use of a Deep Neural Network (DNN) as object detector[4] that performs well enough for our application (some detections results are displayed in Figure 3). Figure 2 contains images acquired by the drone under various conditions representative of the variability of the inputs and the difficulties encountered: **top views** (top, left) show that images often contain several specular areas due to the external lightning of the hangar. This difficulty is minimized by the image overlapping (about 30%) which allows to see the areas under reflection on the next acquisitions. **Bottom views** (top, right) show images that are much less bright. These variations in lightness illumination are mitigated by the presence of an on-board lightning device (LED rings). In addition, as shown in 3 some scenes contain almost no objects (bottom, left), while others contain dozens of them (bottom, right).

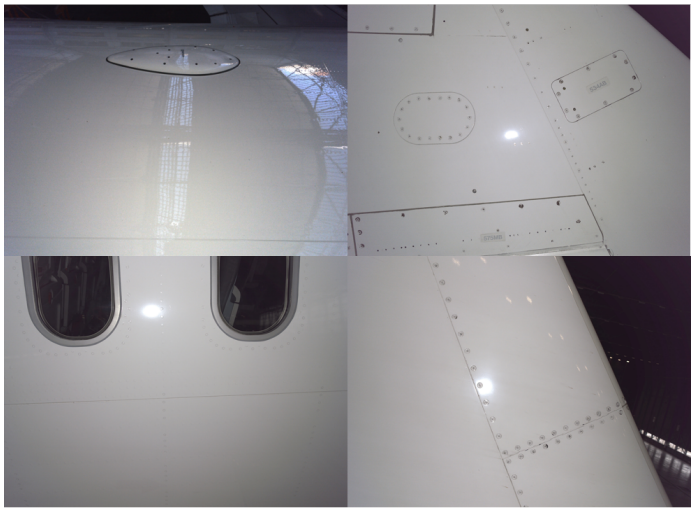

Fig 2 Images acquired by drone.

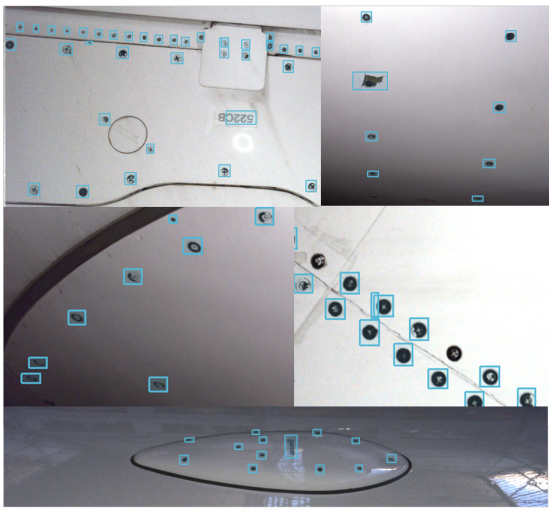

Fig 3 Detected objects.

Growing popularity of Deep Learning (DL) methods has led to great advances in Computer Vision during past years: more specifically, image classification has become a relatively simple problem provided there are enough available data to train deep models (they overpass human performances for this task since 2015). ImageNet[5] and CIFAR[6] challenges give to the research community ways to compare and improve their algorithms on public datasets. However in real world, accessing those data is often an impassable barrier. For the target application, defects such as lightning burns are not frequent enough to envision a classic DL approach, while a lot of other objects that have to be discriminated are very common. Thus, this paper concerns object classification, considering very high imbalance ratio (1:5000) between classes.

We first establish that with extreme class imbalance ratio, state-of-the-art methods are not sufficient and that there is a need to take advantage from both the power of big data algorithms and from more specific methods dealing with few data for some classes. To do so, we demonstrate that with our industrial data, different machine learning approaches are relevant for different volumes of balanced training set. Then we evaluate those models on extremely imbalanced datasets and

propose a method to combine models into an hybrid classifier able to deal with common objects as well as very rare ones. Finally, this hybrid strategy is also validated and characterized on public datasets, modified to create different imbalance ratios.

Section 2 describes our context: how data are acquired, what are the objects and the classes. In Section 3 we describe the machine learning approaches that could be evaluated for our application and justify the choice of Prototypical Network (see Table 2), before showing their limitations for high imbalance ratio between classes in Section 4. Finally we propose an hybrid method in Section 5 and compare results with others approaches on an imbalanced dataset. Those sections refer to our preliminary work published in QCAV conference proceedings.[7] The next ones are new material and results: based on the previously justified notion of hybridization, we propose to deepen these methods through general heuristics in Section 6, then by considering multiple views of the same object in Section 6.1.3. Section 7 is dedicated to the experimentations and analysis of the results obtained by adding these new material. Finally, Section 8 will discuss possible improvements for the classification of rare defects and new possibilities offered by the proposed new system, which go beyond the classification framework.

2 Acquisition and dataset

During inspection, images are taken by an autonomous UAV in order to cover the entire aircraft surface. For a typical aircraft, about $1,200$ high definition (16 MP) images are required, with an overlap between two successive acquisitions ensuring that each zone is acquired at least twice. This reduces the negative impact of specular components and allows to fuse classification results on the same real object seen on several views. The matching of objects from different points of view is made possible by using the location data of the drone thus on the on-board camera, but will

not be covered in this paper.

Acquisitions are sent to a separate laptop or tablet that can process automated analysis using GPUs. To create machine learning data sets, maintenance company experts annotate some batches of these acquisitions.

All objects that can occur on the fuselage have to be registered for various applications composing a general visual inspection (paint state evaluation, markings analysis, etc.). However using such a network with high definition images do not allow very small objects classification (less than $1\,\mathrm{mm}^2$) in reasonable time whereas those objects are crucial as they can be critical defects (lightning burns). Thus potential defects (small or ambiguous objects) are gathered for a supplementary classification step which is the object of this work.

Addressing the defect recognition as an image classification problem allows the use of advanced techniques that are much more complex to be introduce into a one-step object recognition algorithm in which detection and classification are inseparable,[4,8] that perform the best result in the state of the art on object recognition task (using mean average precision as metric).

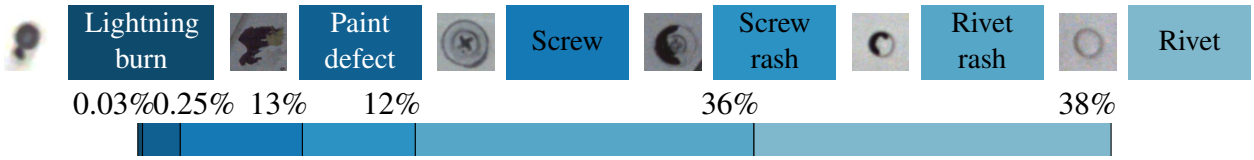

Fig 4 Unbalanced dataset description.

Class	Lightning burn	Paint defect	Screw	Screw rash	Rivet rash	Rivet	**Total**
Samples	11	91	4,758	4,496	13,228	13,959	**36,542**

Table 1 Unbalanced dataset composition with number of samples by class.

The extremely unbalanced number of samples between classes is a specificity of our data distribution compared to reference datasets. In this paper, we will consider a dataset with classes given

in Table 2 illustrated in Figure 4.

This database is not exhaustive and the image quality might not be representative of Donecle quality of acquisition but still gives an overview of the problem solved. In particular, several other types of objects are found on the fuselage of an aircraft and will not be mentioned here, for the sake of clarity and conciseness. Thus, while some classes are numerous enough to envision taking advantage of DL methods (screws or rivets), others are only represented by few samples (lightning burns) and may need dedicated Few-Shot Learning (FSL) methods. We splitted the original dataset into a train set (80%), a validation set (10%) used to tune hyper-parameters, and a test set (10%) used to obtain the presented results.

2.1 Model complexity and required amount of training data

How much data is needed for a given trainable model to perform well on real world data is a crucial question in Machine Learning. Finding an easy and accurate method to determine the required amount of data to reach a target generalization performance is the topic of many researches: statistical learning theory has given some clues, introducing capacity measures for such algorithm, e.g. the Vapnik-Chervonenkis dimension,[9] from which generalization bounds can be applied to learning algorithms like SVM. However those bounds are vacuous for complex models such as DNNs.[10] Thus, empirical tests have been also performed to observe performances on classification tasks versus volume of training data: a logarithmic relationship seems to exist,[11] but might be subject to a potential diminishing return on log-scale.[12] Based on those empirical observations the required number of samples to reach good accuracy is below the number of parameters of a Deep Network, but it still needs a lot of images to be accurate. Transfer learning is a highly popular way to train models using representations learned from another task.[13]

2.2 *Learning from imbalanced data*

Several approaches can be used to cope with imbalanced datasets. **Data-level** method modifies the data by oversampling, under sampling, transforming or generating training samples. **Algorithm-level** approach tunes existing learning algorithms to adapt them to data with skewed distributions. **Hybrid methods** combine those two with the possible add of handcrafted rules or another algorithm.

While existing works in class imbalance focus on imbalance ratios ranging from 1:4 up to 1:100, classification with extreme imbalance ratio that we are facing, and that can be found as well in other applications (fraud detection, detection of dangerous behavior, etc.) remains a challenge.[14] Precisely, the rarest data are often the most valuable ones, like in the present work.

3 Existing Machine Learning approaches

3.1 *Deep Neural Network*

DNN are very efficient for the classification task on reference datasets, assuming a sufficient amount of data. Deep Convolutional Neural Networks (CNNs) have indeed reached high accuracy rate (exceeding 90%), among the most effective approaches: Wide Residual Network,[15] Fractional Max Pooling,[16] Dual Path Network[17] or other advanced methods.[18–20]

We trained those models and fine-tuned some popular networks (ResNet, Inception, etc.) with available pre-trained weights. The best accuracy on our validation set was achieved by a fine-tuned ResNet50 architecture so we used this model, described by Figure 5, as CNN baseline for our problem. It is composed of residual blocks with skip connections that have proven to be very efficient.[21]

convolution 1	7×7, stride 2	
max pooling	3×3, stride 2	
convolutions 2.X	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix}$	$\times 3$
convolutions 3.X	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix}$	$\times 4$
convolutions 4.X	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix}$	$\times 6$
convolutions 5.X	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix}$	$\times 3$
average pooling		
fine-tuned dense layer		
softmax		

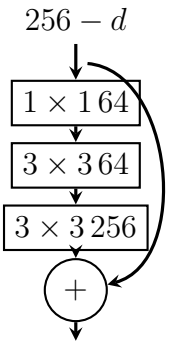

Fig 5 ResNet50 architecture (left) and residual block example (right).

All those DNNs are sensitive to data imbalance.[22] Common methods to tackle this issue are data-based approaches, and consist in creating new samples using transformations on real images, or generating random realistic samples with Generative Adversarial Networks (GANs).[23] Those methods have proven to be efficient, but they usually do not apply in case of extreme imbalanced dataset like ours.

3.2 Support Vector Machine

Another widely used learning approach, popular until the rise of DL hegemony in Computer Vision is Support Vector Machine (SVM). It is a statistical learning approach that needs image representation as inputs. It was first used with hand-crafted descriptors, such as Histogram Of Gradient (HOG).

We also tested SVM with representations learned from unsupervised learning, using a GAN trained to generate realistic images. Recent works have shown that replacing the softmax layer of a DNN by a SVM can give significant gain on classification datasets.[24] Moreover, data imbalance can be integrated into the algorithm using class weights or cost-sensitive learning. SVMs are sensitive to lack of data and to imbalanced classes as well as CNNs and can benefit from the same

data augmentation techniques.

On our test set SVM with HOG performs poorly (maximum accuracy is 0.76 while using the full training dataset), and provides good results on medium datasets ($100 - 1000$) when combined with learned representations (from pre-trained models and GAN). However it never outperformed our CNN baseline, so we did not include SVM on further comparisons.

3.3 Few-shot learning: algorithm-level approaches to face the lack of data

FSL algorithms are usually performing n shot, $k-way$ learning, with n being the number of needed learning samples and k is the number of possible classes for a new sample during inference. Two main public datasets dedicated to FSL algorithm training and test have been created. Omniglot[25] is a dataset composed of handwritten characters from different alphabets and MiniImageNet is a subset of ImageNet. Samples are shown in Figures 6 and 7. Next, we describe some of most popular FSL algorithms.

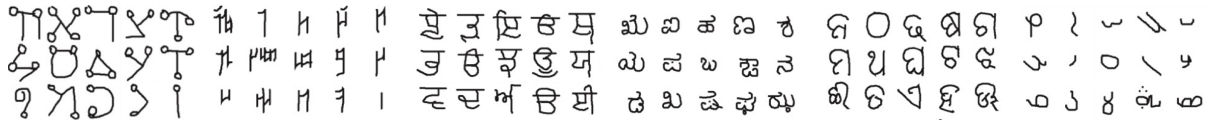

Fig 6 Omniglot samples from 6 alphabets.

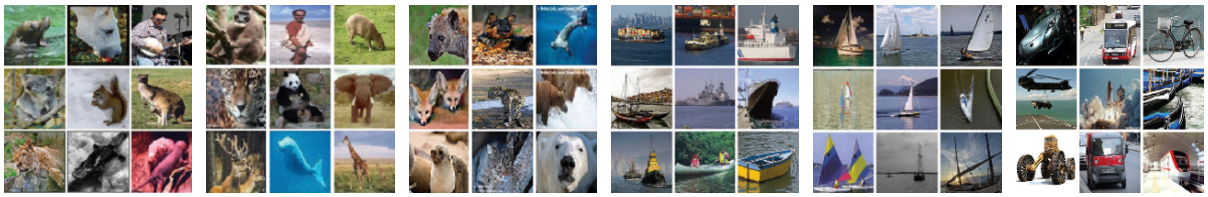

Fig 7 MiniImageNet samples from 6 classes.

To learn from few data, algorithm-level solutions have been proposed. Some of them are considering meta-learning processes[26–28] while others focus on the ability to learn metric.[29]

Siamese Networks. This approach combines multiple networks.[30] To be called 'siamese', two networks have to share the same architecture (same layers with the same parameters) and to share learning loss and weights of the junction layer. The choice of the loss function is of crucial importance. Some known examples are:

$$\textbf{Usual cross-entropy: } L = -y\log(p) + (1-y)\log(1-p)$$

$$\textbf{Triplet-loss}:[31]\ L = \max(d(a,p) - d(a,n) + m, 0)$$

With the triplet-loss formulation, d is the $L2$ loss (or another distance function), a is the input sample from the dataset, p is a sample from the target class (randomly picked), n is a sample from another class. m is an hyper-parameter (margin). With this method, it is possible to perform metric learning, then to use a classical nearest-neighbor classification algorithm to separate classes. This last operation is not part of an end-to-end training process and so cannot be called optimized for the task.

Matching Networks[32] introduced few-shot networks based on the idea of making nearest neighbor algorithm learnable during training process by using it in a differentiable form. This allows to perform end-to-end fully optimized learning and is achieved by embedding a sample into a representation space and then performs a nearest-neighbor-like algorithm with the equation:

$$\hat{y} = \sum_{i=1}^{k} a(\hat{x}, x_i) y_i$$

where $\hat{y}$ is the model prediction, x_i the support features, y_i the support labels, $\hat{x}$ the query sample features and a is a similarity function. This approach is the first end-to-end few-shot dedicated model, and outperforms Siamese network for this task, see Table 2.

Prototypical Network[29] are built within the assumption that a single prototype per class can be

used to compute distance in the representation space. If the used distance is the Euclidean distance, then the best prototype (in the sense that it minimizes the distance between a prototype and supports points) is the mean of the support set representations.

Model Agnostic Meta Learning[28] (MAML) is a very different approach for FSL that uses a clever initialization for deep models.[18,33] MAML is a way to learn how to initialize weights by optimizing the generalization of the model. We took Prototypical Networks as few-shot algorithm baseline, as they obtain the best performances on reference datasets and can dynamically perform k-way classification.

	5-way Accuracy		20-way Accuracy	
Omniglot	1-shot	5-shot	1-shot	5-shot
Siamese	97.3%	98.4%	88.2%	97.0%
Matching	98.1%	98.9%	93.8%	98.5%
Prototypical	**98.8**%	99.7%	**96.0**%	**98.9**%
MAML	98.7%	**99.9**%	95.8%	98.6%

	5-way Accuracy	
MiniImageNet	1-shot	5-shot
Matching	44.2%	57.0%
Prototypical	**49.4**%	**68.2**%
MAML	48.07%	63.15%

Table 2 Recent methods on FSL task.

We took the results from[28] and original papers, gathered them in Table 2.

4 State of the art methods evaluation

Based on the state of the art, we tested the described approaches on our datasets : original training dataset is sub-sampled to obtain training datasets of increasing size from 1 sample per class to $6,000$ samples per class, first to evaluate which one performs the best for a given number of training samples when there is no extreme imbalance. Results displayed on Figure 8 show that these methods have good performances when the classes are limited (< 4). With a large number ($\geqslant 4$), we found DL methods trained with large dataset ($6,000$ per classes) are more accurate. We have also implemented classifiers built on a SVM fed by HOG descriptor as well as on a SVM fed by ResNet features, that performed poorly on our test. We used prototypical networks, which

regarding the current state of the art metrics, see Table 2 appears to be the best alternative in terms of precision and recall for both the simple (Omniglot) and the hardest (minImageNet) datasets.

4.1 Experiment

To evaluate model relevancy regions in term of real training dataset size, we used object classes that we have in sufficient number to train deep models, then create decreasing subsets from the original dataset to observe the effect on top-1, top-2 and top-3 accuracies, with top-k accuracy being the proportion of samples for which the ground truth label is one of the k most probable predicted classes. We used common data augmentation techniques (horizontal and vertical flip, rotation, crop) to obtain presented results. For more general application of our work we also tested our method on public available datasets that where truncated in order to artificially create extremely imbalance datasets that are subsets of well known sets. For those dataset, we observe the performances on the randomly under-sampled class, which were randomly chosen.

4.2 Results

Considering ResNet model, we found that the assertion made in Sun *et al.*[11] that "performance increases logarithmically based onvolume of training data" fits our results in accuracy for top-k accuracy score in classification ($k \in \mathbb{N}, k < N_c$) with N_c the number of classes, however we did not test this assumption for datasets larger than $6,000$ elements per class. In Figure 8, dotted-lines represent log-regression on the obtained results.

We also observed the same relationship for FSL baseline for 2-way to N_c-way task. As expected it appears than under a certain amount of available data, DL methods are less accurate than prototypical network: top-1 accuracy of ResNet is lower than 7-way prototypical classifier: for

those classes it would be preferable to use a FSL dedicated approach. Results are illustrated in Figure 8: a frontier between two approaches is empirically set around 100 samples per class. The problem that stands next is simple: we cannot know in advance which model will be the most relevant for a given sample.

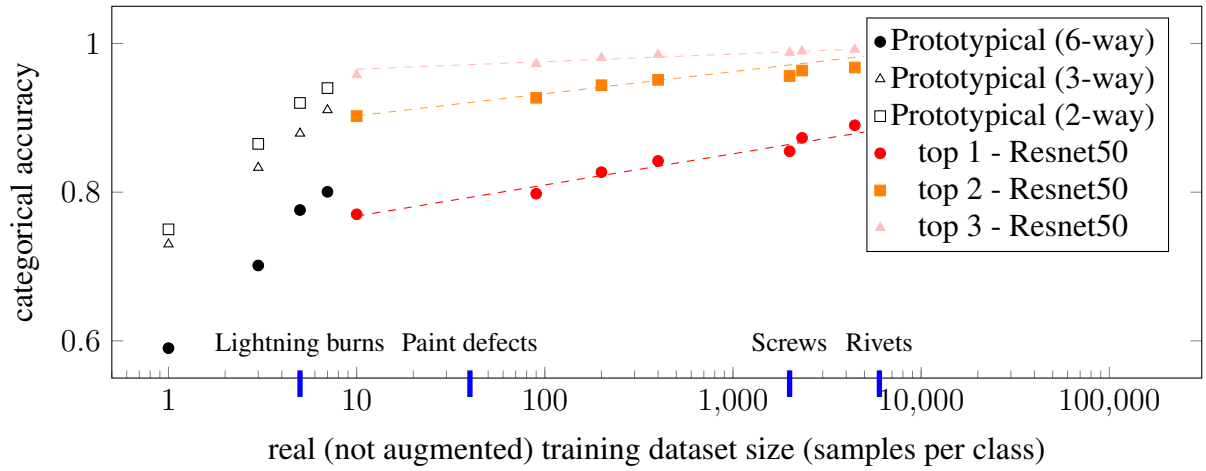

Fig 8 Different approaches performances regarding training dataset size.

5 Hybridization with Random Support Sampling (HRSS)

5.1 Description

Since CNN baseline top-3 accuracy giving good results, and 3-way Prototypical Network performing significantly better than 6-way (see Figure 8), we propose to combine classic DL model and FSL approaches in an hybrid system to combine their advantages.

The idea of our approach is to first estimate which classes are the most likely ones for a given input image, using a DL model with data augmentation. Then, in case one of the possible classes is known to be highly under-represented, few-shot dedicated model is applied, and a combination of the results from those two model gives the final output. If all the possible classes are known to be well represented, the output is the CNN baseline output. We tested our approach and compared

it with other methods on our test dataset. To do so, we took a defect class with 10 training samples while other classes are trained with $3,000$ samples before any augmentation.We compared ResNet50 with data augmentation, Prototypical Network with (10-shot, 6-way), and our hybrid method with top-3-way linking and two hybridization rules (few-shot wins and averaging outputs of deep model and few-shot model).

5.2 Results

We observed global categorical accuracy, the precision, recall and average precision on the imbalanced defect class (varying the algorithm defect output probability), results are shown in Table 3. We can see that hybrid method improves precision, recall and average precision (AP) for rare defect class. The best performances were obtained using average output hybridization method.

	Rare defect classification performances			
Truncated dataset	Categorical accuracy	Precision	Recall	AP
Hybrid Method (combined)	**0.877**	**0.97**	**0.77**	**0.79**
Hybrid Method (FS)	0.867	0.94	0.71	0.71
Prototypical network	0.821	0.84	0.75	0.75
ResNet50	0.863	0.90	0.75	0.77

Table 3 Classification results on imbalanced dataset.

The biggest gain is visible on precision score, which is very important for the addressed industrial application with big and imbalanced dataset, because operators can only handle a reasonable amount of false positive alerts. Nevertheless, the recall score is not high enough to constitute a truly reliable aid to the operator. Nevertheless, the recall score is not high enough to constitute a truly reliable aid to the operator. This is why, despite these encouraging initial results with regard to the hybridization of methods, this naïve approach seems to have considerable room for improvement. Indeed, the choice of the support element vectors of the highly represented classes was not questioned, and the interfacing of the two models is naive.

6 Heuristics for class subsampling

In order to overcome the imbalance of classes in the context of a classification problem by machine learning, some methods have been proposed in the literature[34] for oversampling as well as undersampling operations. The oversampling corresponds to the data augmentation process already mentioned in the previous sections. We describe here some undersampling heuristics that can be applied to the most well represented class sets.

6.0.1 Tomek's links

A sub-sampling approach is to try to remove unrepresentative or too noisy samples. The notion of Tomek's link[35] can be used for this purpose. Given a metric space, two examples E_i and E_j of two different are a Tomek link if there is not an example E_k, such that $d(E_i, E_j) > d(E_i, E_k)$ or $d(E_i, E_j) > d(E_j, E_k)$. This concept can be used via two purposes:

- Undersampling: the elements of the majority classes that belong to a Tomek link are deleted.
- Cleaning: all elements that belong to a Tomek link are deleted.

The use of this technique is possible in our case by considering the representation space learned by the prototypical network provided with the Euclidean distance. Thus we can use this method to sub-sample over-represented classes.

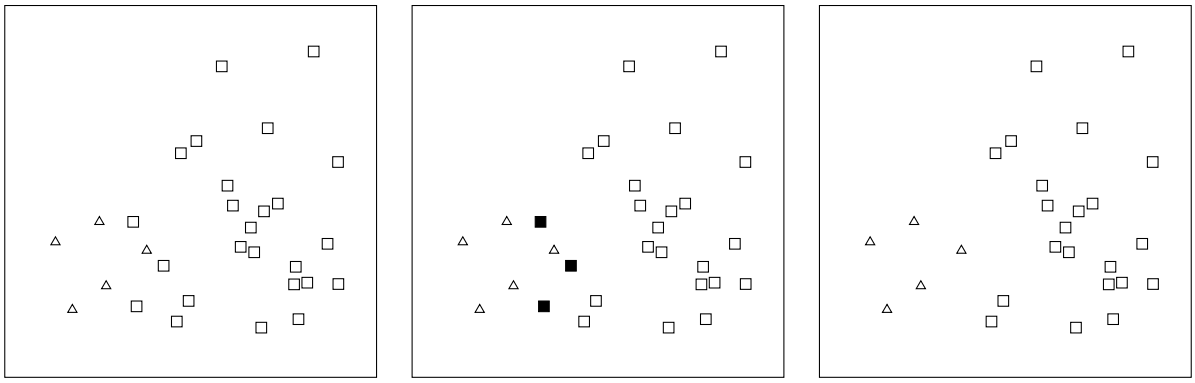

Fig 9 Undersampling iteration using Tomek's link: Two classes (left), sample of biggest class belonging to Tomek's link (center) and resulting datasets (right).

An iteration of sub-sampling over represented classes dataset using Tomek's link is illustrated in Figure 9: The elements of over-represented classes are highlighted (in black) and then deleted.

6.0.2 Condensed Nearest Neighbour Rule

Another iterative approach is based on the progressive construction of sets, selecting only those examples that cannot be explained by a simple classification based on elements already known. The idea is that the information provided by the elements that are correctly classified by this simple approach is already largely contained in the previously selected examples. Condensed Nearest Neighbour Rule proposed by Hart[36] and described in 1 is an algorithm reflecting this reasoning using neighbour algorithm as the naive classifier. It does not guarantee to find the smallest consistent subset.

6.0.3 One-sided selection (OSS)

It is possible to combine the two previous approaches by first applying the Tomek rule (under sampling) to eliminate noise and edges, then use the Condensed Nearest Neighbour algorithm.[37]

Algorithm 1 CNN rule.

Require: Let E be the whole considered set E_{c1} be the under-represented class and E_{c2} be the over-represented class .
$\hat{E} \leftarrow E_{c1}$
Randomly select $x \in E$
$\hat{E} \leftarrow \hat{E} \cup \{x\}$
Classify each remaining examples in E using 1-k-NN. All misclassified elements are noted X_f.

for all $x_f \in X_f$ **do**
$\hat{E} \leftarrow \hat{E} \cup \{x_f\}$
end for

As it was proven to be more efficient than only one of these heuristics,[34] we will consider this approach.

6.1 Proposed methods

The described above methods allow to under-sample the data under consistency consideration. However, two major pitfalls remain. First, they are likely to delete ambiguous examples, that might be noisy samples or outliers, but can also be interesting samples in crucial feature space zones. Second, there is still a significant amount of randomness in these algorithms. We propose a method to select the best samples that tackle those issues. The idea is to clusterize the potential support samples in the feature space provided by the prototypical neural network, then to keep only the closest elements to the prototype of each cluster. Using the prototype could be an option, but as it does not correspond to a real sample, this will increase the dependency on the encoding function (which is learnt), and the operation will not be under-sampling any more.

6.1.1 Meta-Data Filters on support dataset (MDF)

Since Deep Learning methods require thousands of data to be accurate, it is necessary to gather learning examples from a variety of sources (different airline companies, different aircraft models,

indoor and outdoor acquisitions, etc.). On the contrary, FSL is based on a handful of examples and it is therefore possible to avoid these mixtures to automatically consider problem-specific databases. We introduce meta data filters to create those adapted subsets for all categories. To improve the synergy between the deep model and the FSL method, we can also select examples to use as supports between two classes among those that could not be separated by the deep model during the learning phase. This is done using the learning confusion matrix. This makes it possible to obtain a matrix of the support examples to be selected as a priority for the separation of two given classes.

6.1.2 Cluster-based Medoid Prototypes (CMP)

To prepare *N*-shot learning task, we need to create prototypes from support samples. As we observed that classification accuracy growth with N, we set N to be equal to N_{c1} the cardinal of the under-represented class. We want to select the N best samples among N_{c2}, with $N_{c2} >> N_{c1}$. Prototypical Neural Network provides a feature space where euclidean distance can be used. We propose to create clusters in this space with Density-Based Spatial Clustering of Applications with Noise (DBSCAN) that is a very popular method,[11].[38] Once the clusters are created, we simply take the acquired sample that is the closest to the each cluster centroid (the medoid) as support sample for the class. The Figure 10 illustrates the difference between 9 randomly selected supports for the 'screw' class (left) where some screw models (numbered 1 and 2) can be represented more than once in the support set and medoids selection after clustering (right) where each support is a different type of screw.

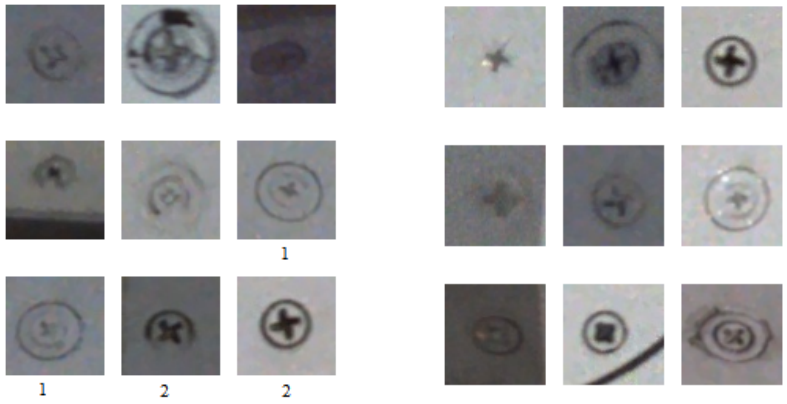

Fig 10 9 random screw supports (left) and 9 screw medoids selected after clustering (right).

The previously described heuristic can be used in many cases of extreme imbalance dataset. Another aspect, taking advantage of the specificity of our acquisitions, which consists of multiple shots of each fuselage area, is the linking of images corresponding to the same part of the aircraft. The following section discusses this approach.

6.1.3 Sequence-Wise Prototypical Network (SWPN)

In this section we discuss the possibility to use several views of the same object to create a prototype in the prototypical feature space for further classification. It should be remembered that the images acquired and all defects detected are geolocated relative to a 3D model of the aircraft: an example of such a model is given in Figure 2. In addition, a controlled overlay allows each of them to be visible on several of those. We propose to match these elements based on the position in space and the type of object predicted by the Deep Neural Network (referred to is previous sections). We can thus form sequence of images of the same object. From this point we embed all the elements of the sequence into the prototypical feature space, then create a query prototype to be classified by dedicated FSL classifier. Two views of the same objects are shown in Figure 11: here the objects are lightning strikes: in itself measures only a few square millimetres, but as it

was previously detected during a manual inspection, a black circular pad (much more visible in the image) with an identification number was added to the fuselage. The sequence embedding process is illustrated in Figure 12 with two extracted images of the same detected lighting strike.

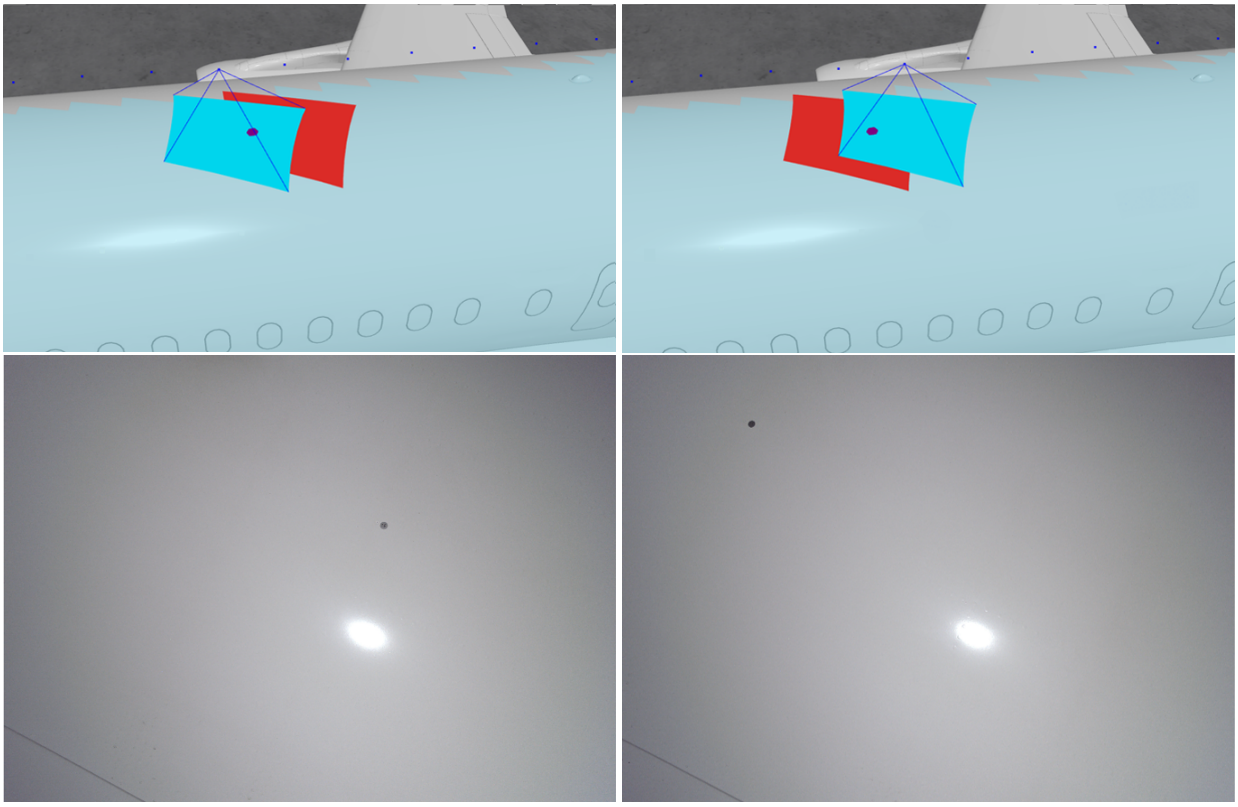

Fig 11 Lightning strike from two points of view.

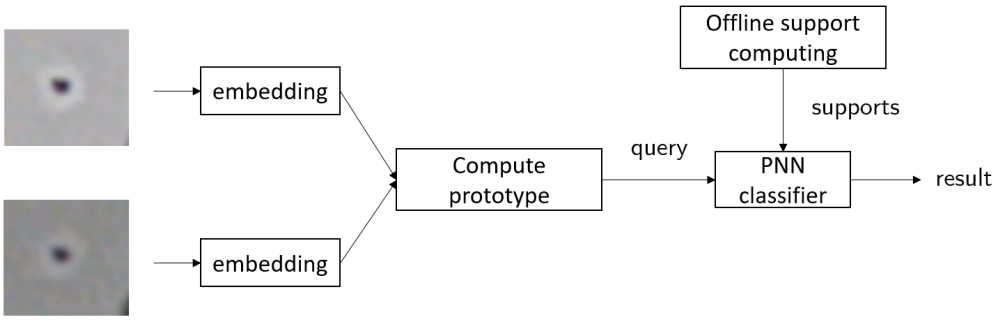

Fig 12 Sequential embedding for classification using prototypical neural network.

7 Experiment and analysis

In this section, we compare the results obtained by new materials combinations under one consistent dataset: Hybridization with Random Support Sampling (HRSS), Cluster-based Medoid Prototypes (CMP), One-Sided pre-processing for Support dataset cleaning (OSS), Meta-Data Filters (MDF) on support dataset and Sequence-Wise Prototypical Network (SWPN).

We sequentially apply the described algorithms to combine them, taking care of the order of operations. For example, the combination MDF + SWPN + OSS + CMP is obtained by the following steps:

1. Filter the elements using the available meta data: here we use the airline and aircraft model that correspond to the acquisitions after noticing that many elements are specific to them.

2. Associate views of the same objects using the SWPN.

3. Use OSS to reduce the size of over-represented class sets.

4. Select support prototypes using CMP.

This allows to evaluate performance gain of each of the proposed hybridization scheme. We compare the Deep Learning and FSL baselines with several hybrid models on both CIFAR-10 truncated dataset (we truncated each class then consider the mean average precision over those classes) and industrial dataset.

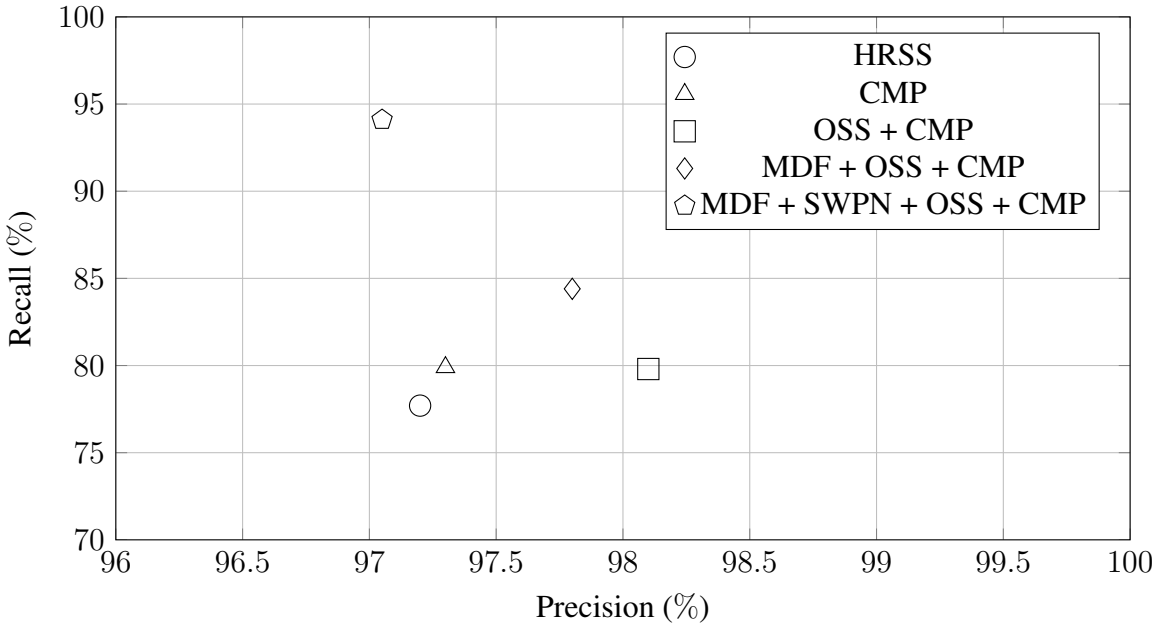

Fig 13 Precision and Recall for different methods.

In this figure we can see the significant improvements made by the different additive modules. In particular, the recall score is increased. It should be noted that the methods most specific to our application, namely the use of meta-data or sequencing of requests for the prototypical network, provide the most significant benefits. Nevertheless heuristics methods of sub-sampling which are model-agnostic are also effective. We have therefore presented general and effective methods as well as methods that are more specific to the application in question and even more effective.

8 Conclusion and prospect

In this paper, we analysed different machine learning approaches that give incomplete answer to a practical use-case of FSL due to extreme data imbalanced. As in many applications,[39,40] the rarest data are the most critical, so there is a need to be specifically accurate on those classes. We compared those methods with different training dataset sizes with balanced classes, then with a

training dataset that contains one few-sampled defect class. We proposed an hybrid method that performs better on our dataset for the under represented defect classification.

Some drawbacks of our approach might lead to future works on the following aspects:

Required data for training hybrid method. By now we needed to train multiple models with sub-sampled training sets to define boundaries (in term of training samples) that delimits which model is appropriate for which class. Using the observed logarithmic relationship between training set size and accuracy an analytical estimation of those boundaries might help reducing the number of needed training phases.

Few shot dataset representativeness. For FSL methods, the representativeness of the support set is not granted. Performances expectations are conditioned by the likelihood of not having an unrepresentative support set and the quality of these tiny data sets can be easily altered by the simplest noisy or outlier example.

Hybrid learning by confusion transfer. We introduced some methods to improve classification performance of an hybrid algorithm, however those methods are focusing on the choice of the support vector for prototypical network regardless of the existing confusion of tit could be possible to use samples not correctly learnt by the DL model as support vector for the FSL method, using the learning confusion matrix.

Overall, we discussed several machine approaches and proposed a data-aware hybridization method that applies to rare defects detection on aircraft fuselage and could be extended to many other fields with extreme imbalanced classes as it relies on the well-known effectiveness of DL approaches using transfer learning and simplicity of Prototypical Networks.The chosen method not only allows a significant improvement in the measured empirical risk, but also provides a real

mean of adapting the classification model through comprehensible actions in a semantic space. Opens the way to new prospects.

Online-adaptive classification model In the same vein, the presented work paves the way for online learning to complement the offline learning of the core model. Indeed, corrections made by a human operator can be considered as excellent candidates to support a prototype class with regard to the knowledge of effective confusions.

Using zero-shot learning for non-exclusive classification

Up to now we have discussed the case of classification of mutually exclusive classes. However, this assumption is not generally valid. Indeed, two distinct classes can be present in the same area of interest, in particular defects that tend to appear on protruding objects such as lightning strikes on screws. Various methods to tackle this issue can be found in the Machine Learning literature: it is possible to learn exclusive classes with a classic neural network classifier and modify the output interpretation function taking not only the class with the maximum probability but applying a threshold on the probabilities of all classes instead. We propose to use zero-shot learning to avoid those modifications: the idea is to create artificial prototypes that are the combination of N-shot embedding of real classes. By this we can produce all the class combinations automatically and classify them. This technique can be applied for refining classification results. In particular, defects that may occur preferentially on structural elements (such as lightning strikes on screws or rivets) can be categorized in this way. Figure 14 illustrates the creation of such a prototype: black filled circle and diamond are the medoids for respectively screws and lightning strikes and the black filled pentagon is the mean of those medoid and represent the theoretical prototype for screws struck by lightning.

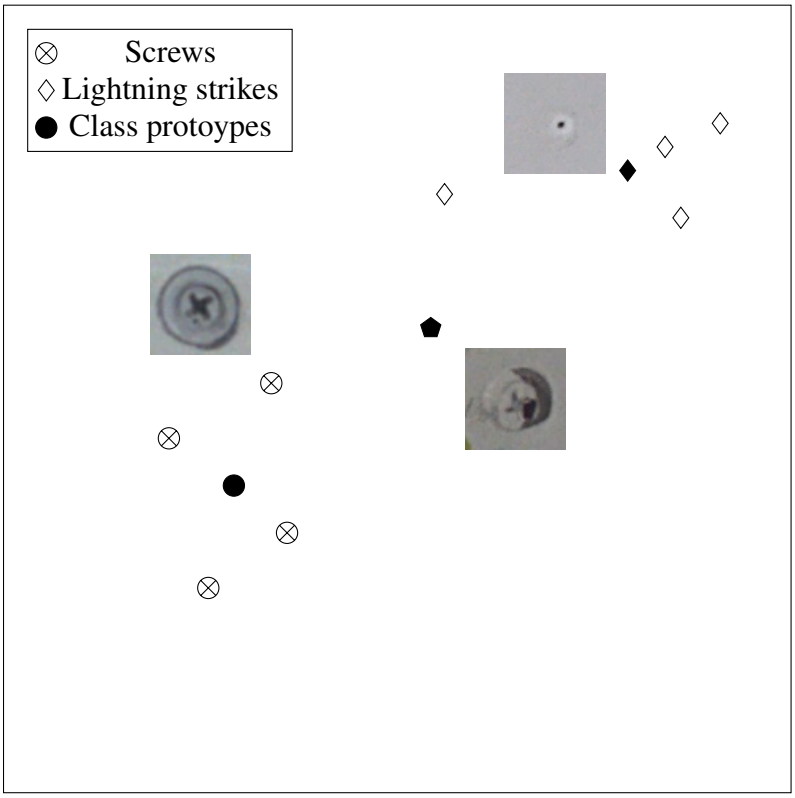

Fig 14 Zero-shot learning for class combination.

References

1 J. Miranda, S. Larnier, and M. Claybrough, "Caractérisation d'objets sur des images acquises par drone," in *Conférence Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP)*, 2018.

2 I. Jovančević, S. Larnier, J.-J. Orteu, and T. Sentenac, "Automated exterior inspection of an aircraft with a pan-tilt-zoom camera mounted on a mobile robot," *Journal of Electronic Imaging* **24**(6), p. 061110, 2015.

3 J. Miranda, S. Larnier, A. Herbulot, and M. Devy, "UAV-based inspection of airplane exterior screws with computer vision," in *14h International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, 2019.

4 J. Redmon and A. Farhadi, "Yolo v3: An incremental improvement," 2018. arXiv:1804.02767.

5 J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F. F. Li, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009.

6 A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," tech. rep., Citeseer, 2009.

7 J. Miranda, J. Veith, S. Larnier, A. Herbulot, and M. Devy, "Machine learning approaches for defect classification on aircraft fuselage images aquired by an UAV," in *Fourteenth International Conference on Quality Control by Artificial Vision*, **11172**, pp. 49–56, SPIE, 2019.

8 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision (ECCV)*, pp. 21–37, Springer, 2016.

9 A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Learnability and the Vapnik-Chervonenkis dimension," *Journal of the Association for Computing Machinery (JACM)* **36**(4), pp. 929–965, 1989.

10 G. K. Dziugaite and D. M. Roy, "Computing nonvacuous generalization bounds for deep (stochastic) neural networks with many more parameters than training data," 2017. arXiv:1703.11008.

11 C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 843–852, 2017.

12 J. Krause, B. Sapp, A. Howard, H. Zhou, A. Toshev, T. Duerig, J. Philbin, and L. Fei-Fei, "The unreasonable effectiveness of noisy data for fine-grained recognition," in *European Conference on Computer Vision*, pp. 301–320, Springer, 2016.

13 H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging* **35**(5), pp. 1285–1298, 2016.

14 B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence* **5**(4), pp. 221–232, 2016.

15 S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016. arXiv:1605.07146.

16 B. Graham, "Fractional max-pooling," 2014. arXiv:1412.6071.

17 Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," in *Advances in Neural Information Processing Systems*, pp. 4467–4475, 2017.

18 D. Mishkin and J. Matas, "All you need is a good init," 2015. arXiv:1511.06422.

19 C.-Y. Lee, P. W. Gallagher, and Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," in *Artificial intelligence and statistics*, pp. 464–472, 2016.

20 J. Snoek, O. Rippel, K. Swersky, R. Kiros, N. Satish, N. Sundaram, M. Patwary, M. Prabhat, and R. Adams, "Scalable bayesian optimization using deep neural networks," in *International Conference on Machine Learning (ICML)*, pp. 2171–2180, 2015.

21 C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and

the impact of residual connections on learning," in *31st AAAI Conference on Artificial Intelligence*, 2017.

22 C. Huang, Y. Li, C. Change Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

23 F. Konidaris, T. Tagaris, M. Sdraka, and A. Stafylopatis, "Generative adversarial networks as an advanced data augmentation technique for mri data," in *14h International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.*, 2019.

24 Y. Tang, "Deep learning using linear support vector machines," 2013. arXiv:1306.0239.

25 B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science* **350**(6266), pp. 1332–1338, 2015.

26 S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *5th International Conference on Learning Representations (ICLR)*, 2017.

27 N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, "A simple neural attentive meta-learner," 2017. arXiv:1707.03141.

28 C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *34th International Conference on Machine Learning-Volume 70*, pp. 1126–1135, 2017.

29 J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems*, pp. 4077–4087, 2017.

30 J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a"

siamese" time delay neural network," in *Advances in neural information processing systems*, pp. 737–744, 1994.

31 F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 815–823, 2015.

32 O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, *et al.*, "Matching networks for one shot learning," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 3630–3638, 2016.

33 X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 249–256, 2010.

34 G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter* **6**(1), pp. 20–29, 2004.

35 I. Tomek, "Two modifications of cnn," *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-6**, pp. 769–772, Nov 1976.

36 P. Hart, "The condensed nearest neighbor rule," *IEEE transactions on information theory* **14**(3), pp. 515–516, 1968.

37 M. Kubat, S. Matwin, *et al.*, "Addressing the curse of imbalanced training sets: one-sided selection," in *International Conference on Machine Learning*, **97**, pp. 179–186, Nashville, USA, 1997.

38 S. Chakraborty, N. Nagwani, and L. Dey, "Performance comparison of incremental k-means and incremental dbscan algorithms," 2014. arXiv:1406.4751.

39 B. Krawczyk, M. Galar, Ł. Jeleń, and F. Herrera, "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy," *Applied Soft Computing* **38**, pp. 714–726, 2016.

40 M. J. Siers and M. Z. Islam, "Software defect prediction using a cost sensitive decision forest and voting, and a potential solution to the class imbalance problem," *Information Systems* **51**, pp. 62–71, 2015.

Julien Miranda is a Ph.D student at LAAS -CNRS laboratory. He is working in partnership with the start-up Donecle on the subject of defect recognition on aircraft fuselages based on images acquired by a drone. He is interested in machine learning theory and in particular the expressiveness of models, decisional robustness and generalization theory.

Stanislas Larnier obtained a PhD degree in Applied Mathematics from Université Paul Sabatier, Toulouse, France, in 2011. Then he worked on Image and Video Processing for INRIA, LAAS-CNRS then AKKA Research. Now, he is the Computer Vision R&D Manager in Donecle. His field of expertise is scene analysis. He worked on collaborative projects with applications in various fields: coastal engineering, microbiology, visual grading, and robotics.

Ariane Herbulot is an Associate Professor from University Paul Sabatier (Toulouse, France). She obtained her PhD in 2007 at University of Nice-Sophia Antipolis (France) on image segmentation by active contours. She is interested on detection and object tracking for video-surveillance and robotics applications. Her research concerns segmentation, detection, object recognition and

motion estimation and tracking.

Michel Devy Michel DEVY is CNRS Research Director at /Laboratory of Analysis and Architecture of Systems/ at Toulouse, France. For more than 35 years, he has participated to the Robotics department, now member (as team leader until 2013) of the RAP team (/Robotics, Action and Perception)/. His research has been devoted to the application of computer vision in Automation and Robotics. He was PhD advisor or co-advisor for about 60 PhD students and co author of about 200 scientific communications.

List of Figures

List of Tables