AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

ENDIVE – ENcouraging DIVErsity in few-shot learning

Summary table of persons involved in the project

Partner	Name	First name	Current position	Role & responsibilities in the project	Involvement (person.month) throughout the project's total duration
IMT Atlantique	PASDELOUP	Bastien	Associate professor	Scientific coordinator WP lead	38
	GRIPON	Vincent	Professor	Expertise in FSL PhD direction	8
	FARRUGIA	Nicolas	Associate professor	Expertise in neuroscience	4
	REIFFERS-MASSON	Alexandre	Associate professor	Expertise in optimization	4
	To be hired		PhD student	Participant WP1, WP2, WP4	36
	To be hired		Intern	Participant WP3	6
	To be hired		Post-doc	Participant WP3, WP4	18
Gipsa-lab	TREMBLAY	Nicolas	Researcher (CRCN CNRS)	Expertise in DPP	6
CRIStAL	BARDENET	Rémi	Researcher (CPJ CNRS)	Expertise in DPP	6

Changes with respect to the pre-proposal

The requested total in the pre-proposal was estimated to 278k€. After carefully re-assessing the financial needs of the proposal, we have raised this amount to 296,555€. The reasons are a recent increase in PhD salary (from 132k€ to 138k€), adjusted functioning fees (from 34.7k€ to 37,555€) and addition of two GPUs and containing machines, that will be required to load large deep learning models (∼5k€ each). The total is below the allowed increase of 7% defined in the AAPG guide.

Contents

AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

1 Proposal's context, positioning and objective(s)

1.1 Objectives and research hypothesis

The objective of this project is to evaluate the benefits of diversity-aware sampling procedures – notably Determinantal Point Processes (DPPs) – within a Few-Shot Learning (FSL) context. We introduce in Section 1.1.1 why this is an interesting question, and then derive research questions in Section 1.1.2.

1.1.1 General context

FSL [1,2] arose as an interesting branch of machine learning, with the goal to study and reproduce a human's ability to learn from few examples, even in previously unseen problems, leveraging past experience [3]. More practically, FSL focuses on application cases where only small annotated datasets are available, due to reasons as diverse as ethics, legislation, high costs, etc. This encompasses many practical situations, such as classification of motor imagery data [4], action recognition [5], and more generally many applications in computer vision, robotics, audio processing, natural language processing and domain-specific cases. FSL also provides a great framework for pseudo-labeling data, thus reducing human annotation effort [6].

Typically, FSL classification is performed by leveraging pre-trained deep learning models through transfer. A chosen model is used to transform data into a feature representation, thus arranging data in a space learned by the model, for easier classification. Models can be trained on external datasets with the FSL task in mind, or taken off-the shelf. The former allows for more controlled settings, but generally leads to lower performance than the latter. Indeed, *foundation models* – that are behind most latest artificial intelligence breakthroughs such as large language models [7], vision transformers [8] or more recently video generation [9] – require very high resources (computing and datasets) to be trained. While this is a very active direction of research, being competitive on training such large models is realistically out of reach for us. However, using these models at inference is pretty straightforward, as most are calibrated to be usable on mainstream GPU devices.

Beyond the training aspect, FSL has intrinsic limitations. Indeed, the very low quantity of available labeled samples makes a FSL classification problem ambiguous by nature [10]. As an example, consider a FSL image classification task as described in Figure 1. In fact, even with supposedly canonical reference images, a FSL problem is still ill-posed, as the number of variables highly exceeds the amount of data available, and numerous spurious correlations can be made.

(a) Labeled image (class A)

(b) Labeled image (class B)

(c) Unlabeled image to classify

Figure 1: Example FSL image classification task from the miniImageNet dataset [11]. For each possible class, only one labeled example is provided. Here we can observe the ambiguity of the task, as expected class would be A, although the human, the room, or even records in the background can act as confusing elements.

Contrary to standard classification, a FSL problem is therefore very sensible to outliers, and an accurate estimation of which data to label, or which models to use for transfer, can have a huge impact on classification accuracy. In addition, a very low number of labeled data makes it hard to predict the performance one can achieve for a given FSL task, as reserving a part of data for validation – as is typically done in standard machine learning pipelines – is out of question [12].

While these concerns are not new, it is interesting to address them from the angle of random sampling. Indeed, sampling occurs frequently in FSL, whether for creating a FSL classification task, to choose which data to label from a larger unlabeled dataset, or to select a model from a collection of available pre-trained models. Since FSL is sensible to outliers, random sampling with diversity is a very promising approach. Indeed, if we can sample points such that we have a good estimate of the mean of a distribution – *i.e.*, be more representative of data as a whole – and a high variance among sampled points – *i.e.*, capture data aspects that may vary from a sample to another –, then we could probably obtain FSL classifiers that are robust and with high accuracy.

Among sampling approaches that encourage such properties, DPPs have known a recent rise in popularity since their integration in classical machine learning pipelines [13]. A DPP is a fully probabilistic approach, where subsets of points have a chance to be selected that is proportional to the diversity of points they contain. In short, a similarity is computed for all pairs of points, and used as a kernel by the DPP to find such a subset.

While DPPs have started to find a place in classical machine learning pipelines – for instance for constituting diverse batches [14], active learning [15], or hyperparameter tuning [16] – most approaches have focused on what DPPs can bring to training models. However, as mentioned earlier, training foundation models with state-of-the art performance is out of reach for a standard size research team. Therefore, in this project, we are interested in exploring what DPPs can bring to FSL, when such large and powerful models are used off-the-shelf and are only used for transfer. We therefore hope to develop FSL approach that leverage the best models possible, with a low carbon impact, thanks to diversity properties among data, features, or models.

1.1.2 Research hypothesis and objectives

Research hypothesis

The main research hypothesis in this project is that diversity can still help improving performance of FSL, while leveraging powerful state-of-the-art models, even when it is impossible to train them due to the requirements in computation power or data availability.

We therefore choose to focus on approaches that do not require training models, or for very small adaptations only. This is also in line with the philosophy behind the BRAIn (Better Representations for Artificial Intelligence) team(*), to which the project coordinator belongs, which aims at developing smaller and more efficient models for them to be exploitable to the general public, while reducing the carbon footprint.

In details, the points below identify some of the research questions that this project wants to address. We choose to group them according to where they occur in a standard transfer-based FSL pipeline. Details on this method and on other FSL approaches are provided in Section 1.2.

Data axis

- D1 — *Can we predict the difficulty of a FSL task?*

  In classical machine learning applications, it is common to reserve a part of the training set that is used as a validation set. The goal is to train on the remaining training data and to maximize accuracy on the validation set, with the purpose to learn how to generalize well to new data that are representative of the test set. Accuracy on the validation set is thus a good estimator for the model's performance in later applications.

  In FSL, reserving a part of labeled data for validation is either impossible, or severely harming the model's performance, due to the very low quantity of available data. Estimating FSL performance without a validation set is therefore a challenge, for which little has been done up to now [12, 17].

  Exploring diversity between the available labeled samples, unlabeled samples, and datasets used to train the model used for embedding data thus seems to be a promising direction to answer that question. Indeed, it seems reasonable to assume that high similarity between labeled samples and some training classes of the model, that are pairwise diverse to maximize separability, should indicate a good achievable performance for that particular FSL task. On the contrary, a FSL task with low diversity between labeled points should probably be an indication of low achievable performance.

(*) https://www.imt-atlantique.fr/en/research-innovation/teams/brain.

- D2 — *Given an unlabeled dataset, which few points should be labeled to maximize few-shot classification?*

  As mentioned earlier, active learning has shown gains in accuracy when enforcing constraints on uncertainty or diversity among sampled data when training models [17, 18]. In regard to these findings, we believe that these criteria may help selecting data to annotate in a pseudo-labeling process, without training.

  Additionally, some diversity-aware random sampling approaches have good properties as mean estimators, as illustrated in Section 1.2.2. Many FSL algorithms rely on precise estimation of prototypes of classes in a feature space. Therefore, diversity in sampled data may help on this aspect too.

- D3 — *Are all parts of data relevant?*

  When considering a FSL image classification problem, it is easy to understand that there may be some ambiguity. For instance, consider a labeled image of a dog playing with a person, and a second image of a person playing the guitar. If classes of interest are "dog" and "guitar" (*e.g.*, Figure 1), then the human acts as a confusing agent, and performance would probably benefit from cropping operations.

  Therefore, it may be beneficial to first identify parts of interest in images [10] – possibly using segmentation approaches [19](**) – and to select patches based on their diversity, to better characterize the feature space.

Feature axis

- F1 — *Are some features more relevant to FSL than others?*

  In transfer-based FSL, models are used to transport data into a feature space. Data are thus represented by a high-dimensional vector, that can be viewed as their decomposition in abstract concepts previously learned by the model on its training set [20]. As models are generally trained on datasets with high diversity of data, some of these concepts may not be useful for a specific task, and may as well bring confusion [21].

  It is thus interesting to evaluate if some features in the embedding space provide a more concise, diverse representation of data, and consequently improve classification when restricting the representation to them.

- F2 — *Can we exploit positional embedding for better FSL classification?*

  Contrary to more traditional embedding approaches based on multilayer perceptrons or convolutional models, transformers have a positional embedding [22]. The embedding of data depends on the order in which tokens are arranged in the sentence (for large language models) or patches are arranged in the image (for vision transformers). Thus, tokens are put into context for better representation of their associated concepts.

  This property may be interesting to explore in the context of FSL. Indeed, if data to classify are contextualized with labeled data (for instance by concatenating them, or using retrieval-augmented generation [23] to contextualize with labeled data), we may thus obtain various representations of a same piece of data to classify. Then, diversity of these representations in the feature space may indicate the target class, especially when multiple labeled data are available for each class.

- F3 — *Can we select good models for a FSL task?*

  Diverse models provide diverse representations of data. It is a known fact that the quality of transfer is dependent on the proximity between the model's training dataset and the application case [24]. This raises the following question: given a set of pre-trained models, can diversity between training sets and a target FSL task help selecting good models for providing embeddings that improve classification accuracy? Note that this is compatible with the previous research question, as some models could provide interesting features, complementary to features provided by other models. The same question can also be raised using diversity between embeddings instead of between datasets.

  Ensembling has already shown performance gains in FSL learning [25] when training multiple models. It may be interesting to see if some of these jointly trained models should be selected, rather than taking them all. Also – if we allow ourself to train them –, should we enforce some diversity criteria in their training procedure (for instance when defining batches [14]) to force them to learn various aspects of data, for a better selection of relevant one to provide a combined embedding for the FSL task?

(**) Two articles along these lines by the project coordinator and co-authors are under review at ECCV 2024 and EUSIPCO 2024.

1.2 Position of the project as it relates to the state of the art

The two main elements of this project are Few-Shot Learning (FSL) and diversity-aware sampling methods, notably Determinantal Point Processes (DPPs). In Section 1.2.1, we review standard approaches to FSL classification, highlight the domain's challenges, and identify to which of them this project contributes. Then, in Section 1.2.2, we review diversity-aware sampling approaches, and DPPs for machine learning.

In broader machine learning, the question of diversity has received some attention already. In that setting, diversity of training data has been shown as an important factor of performance when training models. In addition, diversity in trained model representations and inference results has also received some attention. Gong *et al.* [26] provide a nice review of diversity-aware approaches in traditional machine learning settings, and interestingly conclude that *"The training of machine learning models requires large amounts of labelled samples. However, the limited training samples constrain the performance of machine learning models. Therefore, effective diversity technology, which can encourage the model to be diversified and improve the representational ability of the model, is expected to be an active area of research in machine learning tasks"*.

1.2.1 Few-shot learning

Definition

A FSL classification task is generally defined as follows: given a very low number of labeled data per class (typically 1-5), maximize classification accuracy of new, unseen, samples.

More formally, we define a dataset $\mathcal{S} = \{(\mathbf{s}_i, y_i)\}_i$ of labeled samples (called *shots*), where $\mathbf{s}_i$ are the data, and $y_i \in \mathcal{Y}$ the associated labels, with values in a given set of candidate classes $\mathcal{Y}$. Additionally, we have a second dataset $\mathcal{Q} = \{\mathbf{q}_j\}_j$ of unlabeled samples (called *queries*), where $\mathbf{q}_j$ are the data for which we want to give a label in $\mathcal{Y}$. These datasets define a N*-way* K*-shot* FSL problem, *i.e.*, a classification problem of data in $\mathcal{Q}$, belonging to $|\mathcal{Y}| = N$ classes, for which we have K shots per class in $\mathcal{S}$. The most commonly encountered FSL classification problems in the literature are 5-way 1-shot, and 5-way 5-shot problems.

In FSL, we distinguish two settings:

- *Inductive setting* — Data from $\mathcal{S}$ are available right from the start, but elements in $\mathcal{Q}$ arrive sequentially, and need to be processed independently. This setting arises when data are rare or needs to be classified on the fly.
- *Transductive setting* — Data from $\mathcal{S}$ and $\mathcal{Q}$ are both available right from the start. This setting arises when data is abundant, but costly to label.

Depending on the setting, the quantity of information available is not the same, and allows one to choose adapted classifiers accordingly. Indeed, in the transductive setting, it is possible to exploit information such as the distribution of data in $\mathcal{Q}$ in a particular space [25].

Approaches to FSL

There are several approaches to FSL, some of which require training models from scratch, while some do not, or only for fine-tuning to adapt an existing model to the domain of the considered FSL task [27]. Among most common approaches, we distinguish *meta-learning*, *metric learning*, and *transfer-based* FSL [1].

Meta-learning commonly refers to approaches that train models, with the aim of "learning to learn" [28, 29]. The idea is to rely on episodic training, to learn from situations that resemble the target FSL task, hoping that the model will learn to solve similar problems, and will therefore adapt easily to new FSL problems. The most common technique in meta-learning branch is MAML [30] (and its numerous extensions [28]). MAML proposes to train easily adaptable models, so that their parameters end up in a state where only a few gradient steps are needed with few data to adapt to the task.

Metric learning approaches [31] aim at training models to provide a large distance between dissimilar data, and a small distances between similar ones [32]. Then, classification of unlabeled data in a FSL problem is made by evaluating the distance to the labeled data using the learned metric. In deep metric learning, a model is trained to learn the metric, sometimes using episodic training, which is why deep metric learning is sometimes considered a subset of meta-learning [1, 28].

Transfer-based learning aims at taking off-the shelf pre-trained (or sometimes training them) models to use them to provide embeddings to data. Due to the recent rise of very large and performing pre-trained models such

as foundation models, this approach has become more prevalent in the FSL literature [33]. Indeed, such models offer very high performance and are nearly impossible to train for the average research team due to the needed resource (in particular, the needed amount of data). However, using them (and fine-tuning them if needed) is quite easy, which allows one to profit from their high quality learned embeddings without any training effort.

To compare approaches, the standard method in FSL is to evaluate them on benchmarks. Benchmarking is performed by randomly generating independent FSL problems, *i.e.*, datasets $S = \{(\mathbf{s}_i, y_i)\}_i$ and $\mathcal{Q} = \{(\mathbf{q}_j, y_j)\}_j$ – where $y_j \in \mathcal{Y}$ are the ground truth target labels, given here for accuracy evaluation only – from a larger standard dataset (*e.g.*, Meta-Dataset [34], miniImageNet [11], and datasets from CoOp [35]). Accuracy is given as the ratio of correctly predicted labels, and is generally averaged over $\sim 1,000$ independent simulations or more.

Recent results on standard benchmarks show that transfer-based approaches have become more prevalent than meta-learning and metric learning [36]. For this reason, and because we want to limit the models training time in this project, transfer-based learning will be our direction of choice.

Transfer-based FSL

Transfer-based FSL consists in the following steps:

1. First, let us consider a deep learning model $\mathcal{M}$, trained on a dataset $\mathcal{D}$ distinct from $\mathcal{S}$ and $\mathcal{Q}$. Such a model can be seen as a composition of two elements: an embedding part, that turns input data into a feature vector; and a classifier, that associates a decision to the feature vector. Classically, the final layer of the model is interpreted as the classifier, but other choices can be made. The model is either trained with a FSL problem in mind, or taken off-the-shelf. In the latter case, similarity between $\mathcal{D}$, $\mathcal{S}$ and $\mathcal{Q}$ is preferable to make transferability efficient.

2. Let $\mathcal{M}(\mathbf{s}_i)$ and $\mathcal{M}(\mathbf{q}_j)$ be the feature representations of our shots and queries, where $\mathcal{M} : \mathbb{R}^D \to \mathbb{R}^E$ corresponds to the embedding part of the model, that transforms data of dimension D in an embedding of dimension E. Working with feature representations of data allows to easily inject a model's knowledge and provide a convenient space for their processing.

3. Finally, we use a semi-supervised classification algorithm to provide labels to queries, given the shots, in the feature space. Candidate algorithms depend on the FSL setting and can be diverse. Common choices are linear probing [37], Nearest Class Mean (NCM) algorithm (inductive setting), or a soft K-means (transductive setting). Prototype-based classifiers first define class anchors in the feature space, *i.e.*,

$$\forall y \in \mathcal{Y} : \overline{\mathbf{z}_y} \overset{\text{def}}{=} \frac{1}{K} \sum_{\substack{(\mathbf{s}_i, y_i) \in \mathcal{S} \\ \text{s.t. } y_i = y}} \mathcal{M}(\mathbf{s}_i) \ . \tag{1}$$

   Then, labels are attributed to the queries $\mathbf{q}_j \in \mathcal{Q}$ according to a measure of distance between their feature representations $\mathcal{M}(\mathbf{q}_j)$ to the class prototypes $\overline{\mathbf{z}_y}$ ($\forall y \in \mathcal{Y}$). Classical choices are Euclidean distance, of cosine similarity to match the loss function used when training the model.

This standard pipeline can be improved at numerous stages [25]. During training of the model, regularizations can be added to favor particular representations of data. Also, some preprocessing on data, or data augmentation techniques can be used to virtually increase the number of shots. In the embedding space, various normalizations can be made before classification, etc. Another interesting adaptation of that pipeline is ensembling. Indeed, multiple models can provide multiple points of view on data, and ensembling has been a standard technique in machine learning for a long time. Previous work has shown that multiple models may consistently achieve better performance than a single one, for the same total number of model parameters [25].

Challenges in FSL

There are many scientific challenges in FSL, such as integrating knowledge from models pre-trained on various datasets, adapting training procedures for models with the goal to use them for FSL tasks, developing data augmentation techniques to better estimate the distribution of data in a particular feature space, combining multiple modalities in data, etc. Some require training models, while some do not, with the aim to reduce the necessary training effort to reach good performance.

In addition, the apparition of multimodal foundation models – *e.g.*, CLIP [37], CLAP [38], Imagebind [39], Textbind [40] – has led to new perspectives in the FSL community [41]. In particular, models that incorporate a text dimension allow the use of the few available annotations associated with the labeled data to better characterize the feature space and improve FSL tasks.

Figure 2 shows an overview of current challenges and future directions (as of 2023) of FSL. In this figure, we added stars to indicate the directions to which this project contributes, as detailed in Sections 1.1.2 and 1.3.

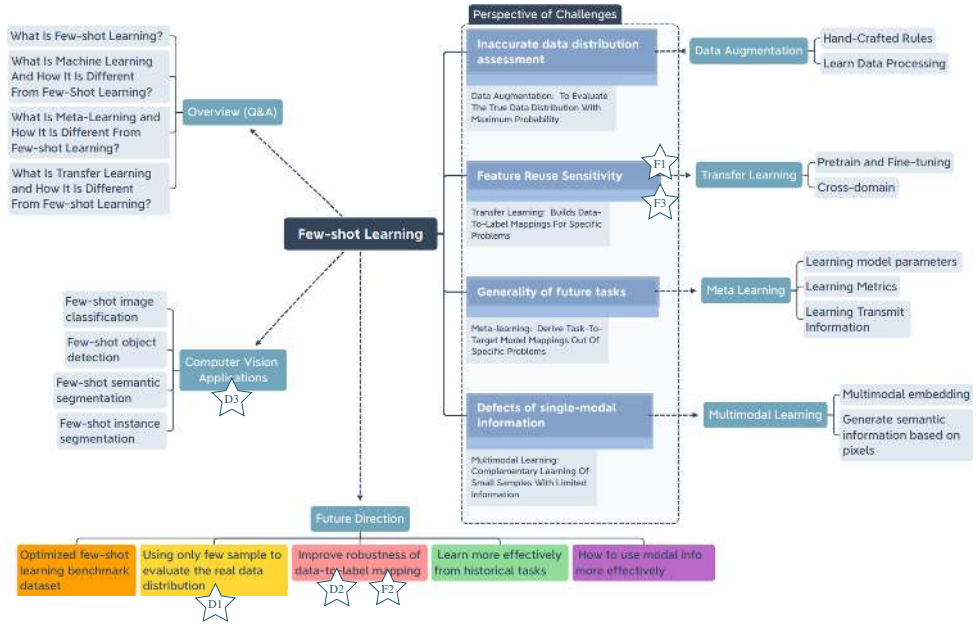


Figure 2: Illustration of the challenges in FSL. Stars indicate elements to which this project will contribute, with associated text corresponding to research questions in Section 1.1.2. Figure adapted from Song *et al.* [2]

1.2.2 Random sampling

Random sampling in deep learning models

Random sampling is a key element of machine learning. In fact, in most real-life datasets, data are not balanced, which could easily lead to training models strongly biased toward the majoritary class. A common technique to balance datasets relies on *oversampling* or *subsampling*. The idea is to give a probability to training data to be sampled that depends on their representativity in the dataset. One of the most common technique to achieve balancing is SMOTE [42] and its evolutions, *e.g.*, ADASYN [43].

Data augmentations are also frequently used to improve the balance in classes when datasets are imbalanced. Beyond that corrective usage, it has also become a standard when training deep learning models, as it allows to diversify the batches and enforce some desired invariance properties, *e.g.*, cropping, rotation, luminosity (for images [44]) or clipping, noise addition (for audio [45]). Integrating good data augmentations into training procedures has become a cornerstone to achieve state-of-the-art performance.

Sampling can also be found during training when batches are constituted. Indeed, training models on full datasets at once requires too much memory, and is generally done on smaller portions of the dataset. In most cases, these batches are created as a uniform random partition of training data, but some approaches have explored constitution of batches with desired properties such as diversity [46]. In addition to improving training efficiency, batches also allow one to perform batch normalization [47], in order to normalize activations and speed up the training procedure. Here again, sampling may be used to accelerate this operation [48].

More specifically, in FSL, random sampling has been used at many places. Episodic training [1, 28] in meta-learning and metric learning approaches require constitution of small FSL problems, which are drawn randomly from larger datasets. Also, estimating the distribution of data has received a lot of importance in FSL, as it allows better separability for classification, and sampling representative points [49] from that distribution.

Sampling is also at the core of active learning [17, 18]. Indeed, careful data selection for training models is key to select data that have a high impact when used to train the model. In addition, some works in the BRAIn team have started to explore active learning for good estimation of class prototypes in an embedding space [50].

Diversity-aware sampling

In many application cases where random sampling is needed, a first natural approach consists in using *i.i.d.* sampling, assuming that all points/data have an independent chance of being selected. In many situations however – *e.g.*, recommendation systems [51] or automatic summarization [13] – one may want to sample points that are diverse, to provide a larger panel of answers to a request or to improve user experience.

There have been a few approaches to encourage diversity in a subset of points. As reviewed in [13], most approaches are not fully probabilistic, and therefore suffer from limitations in their applications. From an optimization viewpoint, Batra *et al.* [52] introduce a methodology to obtain a diverse set of best solutions to a problem. More recently, diversity sampling and uncertainty sampling have been introduced as ways to select unlabeled data currently unknown to a machine learning model. This has received attention, notably in the context of active learning [17, 18, 53]. These approaches aim at estimating the decision boundary of a model, and to sample points that will improve its definition for better classification accuracy after a training phase.

It is also worth mentioning that some approaches have focused on sampling representative points (*i.e.*, from a typical set). Among these methods, stratified sampling [54] and cluster sampling [55] aim at better representing the population by sampling from subpopulations. While this can be seen as the opposite of this project, it may be interesting to couple some of these approaches with DPPs to sample both representative and diverse points.

Determinantal point processes

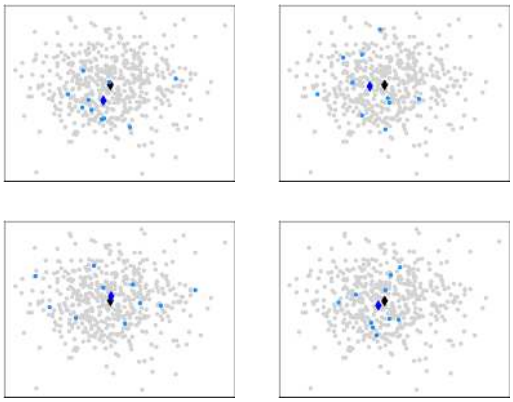
Determinantal Point Processes (DPPs) [56] are probabilistic models that represent the distribution of points across a given space, through a kernel matrix that captures the similarities between the points. Using a DPP, one is then able to sample random points with a guarantee of diversity among the samples. Figure 3 illustrates the difference between uniform and DPP samplings in a 2D space.

As illustrated, DPPs allow one to sample points with more diversity (given a particular metric) compared to *i.i.d.* sampling. This generalizes beyond the 2D case, therefore providing a simple tool to analyze the impact of diversity in many situations, notably in few-shot learning, as proposed in this project. Another observation is that DPPs have the ability to estimate the mean of a distribution more accurately than *i.i.d.* samplers. This is especially of interest in the context of this project, as accurate estimation of class prototypes is key for numerous classifiers in few-shot learning, *e.g.*, NCM or soft K-means.

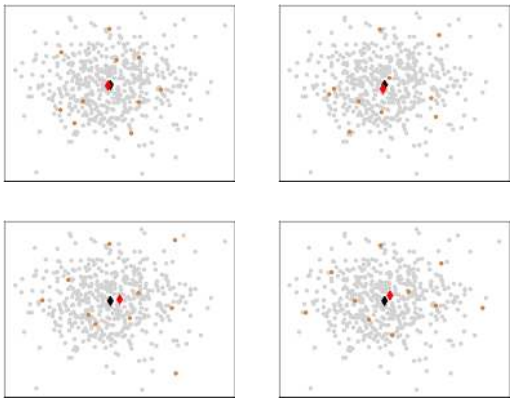
In more details, DPPs rely on computing a similarity kernel $\mathbf{K}$ that captures relations between data points. Provided that this kernel is positive semi-definite, one can show that the submatrix $\mathbf{K}_X$ of rows and columns of $\mathbf{K}$ of indices in a set X has a determinant that is proportional to the diversity of points associated with indices in X. Exploiting this, DPP sampling gives a probability to each subset of points to be sampled that is proportional to the determinant of the associated submatrix, hence its diversity [57]. An interesting extension to DPPs are *fixed-size DPPs*, that allow the sampled subsets of points to have a fixed size [56], contrary to standard DPPs that sample the desired number of points on average. This is especially useful in machine learning applications, for instance for creating fixed-size batches.

There are numerous algorithms to sample DPPs. Classically, exact approaches draw from a set of points using the Chain Rule algorithm [58] or its derivates. More recently, authors of [59] proposed a faster algorithm to sample from DPPs or fixed-size DPPs. It was later improved by authors of [60], who propose an algorithm of lower complexity to sample DPPs, depending of the kernel rank. In addition to exact approaches, approximate methods relying on Monte-Carlo Markov Chains (MCMC) have been developed [61, 62], notably for usage when one wants to sample continuous spaces. This variety of methods offer different strategies, depending on the rank of $\mathbf{K}$, the necessity for fixed-size sample, and the fact that we consider discrete or continuous samples.

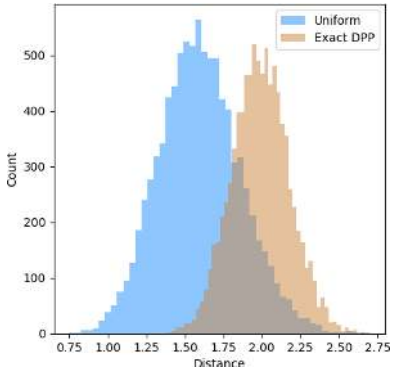
DPP sampling has applications to many aspects of machine learning [13, 63], *e.g.*, for batch diversity [14], active learning [15], hyperparameter tuning [16], neural architecture search [64], etc. However, while DPPs are being investigated at various stages of deep learning, little has been done for FSL. Although a few works can be found for meta-learning [65, 66], and a few others use DPPs for transfer learning in a broader sense [67, 68] or for a specific application [69], no work has studied the benefits of diversity induced by DPPs for FSL.



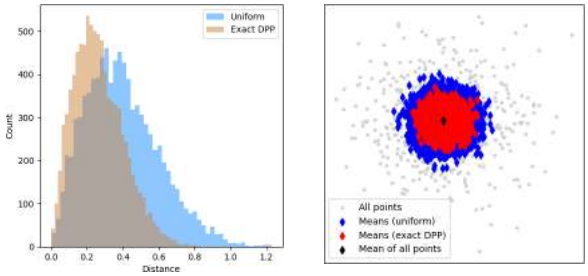
(a) Examples of a uniform sampling realization of 10 points (in blue) among 500 (in gray). The mean of the sampled points (resp. of the entire point cloud) is depicted with a blue (resp. black) diamond.



(b) Examples of a DPP sampling realization of 10 points (in orange) among 500 (in gray). The mean of the sampled points (resp. of the entire point cloud) is depicted with a red (resp. black) diamond.



(c) Distribution of the average inter-point distance, for $9,000$ realizations of each sampling procedure.



(d) Distribution of the distances between the mean of sampled points – *i.e.*, diamonds in (a) and (b) – and the mean of the point cloud, for $9,000$ realizations of each sampling procedure (left); Visual representation of the repartition of these means (right).

Figure 3: Illustration of the differences between uniform *i.i.d.* sampling (a) and DPP sampling (b). DPP-sampled points exhibit interesting properties in terms of diversity (c) and estimation of the mean of the distribution (d). Here, we use a Gaussian kernel as in Tremblay *et al.* [57] to model similarities between points as a decreasing function of the Euclidean distance. In practical cases, this should depend on the desired diversity criteria.

1.3 Methodology and risk management

We first introduce general elements in Section 1.3.1. Then, Sections 1.3.2 to 1.3.5 present the project work packages (WP), derived from objectives introduced in Section 1.1.2 as shown in the table below. Finally, Sections 1.3.6 to 1.3.8 present the deliverables, risk management, and temporal organization of the project.

	D1	D2	D3	F1	F2	F3
WP1	×					×
WP2			×		×	
WP3		×	×	×		
WP4	*Dissemination*					

1.3.1 Methodology

Metrics

- *Accuracy* — In standard FSL classification, the main goal is to develop approaches that maximize the ability to classify new, unseen, data. It is commonly defined as the percentage of data from a test set that are given the expected label. In this project, we will evaluate our approaches on standard benchmarks, associated with the modalities we will consider. Accuracy is important, as it provides an easy way to compare our results

with state-of-the-art methods in the literature. Additionally, it allows to evaluate whether proposed solutions can be put into practice, and to which extent one may expect them to work well.

- *Size* — As models and embeddings tend to grow in size, it is interesting to estimate the reduction allowed by our approaches. We will therefore quantify the embedding dimension reductions, as well as the number of total parameters among selected models. Size of the embeddings is more a concern that we believe is important to keep in mind, as it may open new directions on the dimensions of models needed to solve a particular FSL task. In addition, smaller models/embeddings allow results to be accessible to the general public, notably due to the costs of devices required to use models.

Datasets considered

All experiments will first be conducted on standard vision benchmarks such as Meta-Dataset [34], miniImageNet [11], and datasets from CoOp [35] – for which it is easy to find state-of-the-art models – to evaluate whether obtained results are consistent across datasets.

In a second step, we will evaluate how findings generalize to non-vision datasets. In particular, we want to focus on EEG and fMRI signals, for which we have a strong experience in the team [70–74]. In addition to providing a different modality, such signals often appear in few-shot applications – especially for brain-computer interface (BCI) motor imagery (MI) tasks – which makes it relevant for the project.

Ensuring that findings of the project extend to other types of data will allow our results to reach many possible application domains. Due to the interests of the principal investigator (PI) in medical applications [70,71,73–78], EEG and fMRI will be investigated first, but other types of signals should be considered as well in the follow-up of this project, and should lead to establishment of new partnerships. Going from vision datasets to more diverse medical data will not only cause the project to have an impact in the machine learning community, but also to lead to collaborations with practical outcomes in the medical field.

Concerns on open science

The BRAIn team has been engaged in promotion of open publication since its creation. Codes to reproduce experiments are systematically published in the team's GitHub account(***), articles are uploaded to HAL, and we target conferences or journals with open access and open review (*e.g.*, CVPR, ICML, NeurIPS, JMLR).

Regarding datasets, our approach is to work on open datasets whenever possible. This encourages comparison with state-of-the-art methods, and allows reproduction of our research by other teams. When datasets are constituted, our policy is to put them in open access under Creative Common CC-BY licence (Silent Cities [79]).

We believe that our approach is in line with ANR's concerns on open science.

Ecological considerations

As mentioned earlier in this document, we prefer to leverage pre-trained models than training new ones. Reasons are both practical – impossibility to train transformer models with the same performance level as those that achieve state-of-the art performance in natural language processing or vision – but also ecological. Indeed, training models from scratch receives a lot of attention already, and has a non-negligible carbon impact.

This is in line with the research objectives of the BRAIn team, and with many concerns in the machine learning community. With this in mind, most research questions we want to address do not require full training.

1.3.2 WP1 — Evaluating a FSL problem

Objectives

This WP aims at determining, for a given FSL task, the performance that one may achieve. This is a direct transposition of research objective D1 in Section 1.1.2. However, it also has strong links with research objective F3, as a careful selection of pre-trained models is an important aspect of performance prediction due to similarity between training datasets and FSL problems. This WP should thus help one answering the question: *Given a FSL task and a catalogue of models, which performance can we hope to achieve?*

(***) https://github.com/brain-bzh/.

AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

Preliminary work

In previous work, we have started to explore aspects of this question, although without the focus on diverse sampling. In [21], we showed that some classes in the training set of the models used to provide an embedding for FSL tasks can act as confusing factors. On the contrary, careful fine-tuning on selected training classes can lead to consistent improvement of performance. As a consequence, there is a strong link between the classes in the FSL problem and the datasets used for training off-the-shelf models.

In another work [12], we attacked that question from the feature angle, and studied the impact of data repartition in the feature space on predictability of performance. We demonstrated that, assuming a Gaussian distribution of features in that space, we can better predict generalization error compared to leave-one-out validation. In a later work [80], we showed that text could be used as a complementary modality to improve mean and covariance estimation of classes in the feature space, hence improving FSL classification accuracy.

Tasks

- WP1.1 — Identify DPP-based metrics that correlate with difficulty *(9m, medium risk, medium gain)*

  In a first task, we want to explore metrics that correlate with a problem's difficulty, exploiting DPPs. Promising directions are measures of diversity among 1) shots; 2) shots and queries; and 3) among shots, queries and datasets used to train models used for transfer. Among these directions, we expect that a higher diversity among shots corresponds to easier problems. We also hypothesize that similar diversity between data in the FSL problem, and between classes in the dataset used for embedding, should characterize a simpler problem.

- WP1.2 — Predict performance of a FSL pipeline and model on a new dataset *(9m, medium risk, high gain)*

  Another orthogonal task is to predict, for a FSL pipeline and a given model, the performance one may achieve when presented a new dataset. This can be achieved using DPPs to generate diverse tasks to solve, and therefore obtain an extensive profiling of the pipeline/model pair. Solving this task should lead to great outcomes, as it would provide guidelines to easily select a setting to attack a new FSL problem with easily achievable good performance.

- WP1.3 — Use previously identified metrics to select the best model for a FSL task *(10m, high risk, high gain)*

  This third task aims at integrating metrics found in previous tasks to predict, for a FSL task, which models should be used to obtain the best performance. This task has a higher associated risk, as it depends on the previous ones. The objective here is to answer the question that defines WP1, *i.e.*, to provide one with answers on the best achievable performance for a given FSL problem.

1.3.3 WP2 — Transformers & diversity

Objectives

This WP aims at exploring specificities of modern deep learning models based on transformers, and what diversity can bring to them. There are multiple directions to achieve that, corresponding to research questions D3 and F2: using transformers to improve class representativity (in particular, instance segmentation in images), or exploiting positional embedding for contextualization.

Preliminary work

In a first work, currently under review at EUSIPCO 2024 conference, we proposed a methodology (FICUS – Few-shot Image Classification with Unsupervised Segmentation) to disambiguate images in a FSL classification problem. To reach this goal, we leverage pre-trained transformer models for segmentation. The main idea is to rely on the Segment Anything Model (SAM) [81] to propose candidate masks, and to filter some out based on Deep Spectral Method (DSM) [82]. We observe that instance segmentation brings gains to FSL classification with this methodology. Details of the method are provided in Figure 4.

In addition, we pushed the analysis further to study the impact of instance segmentation in a FSL task. In particular, we study the gains obtained as a function of segmentation quality, either in the queries or in the shots. This second work has led to a submission at ECCV 2024, also currently under review.

AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

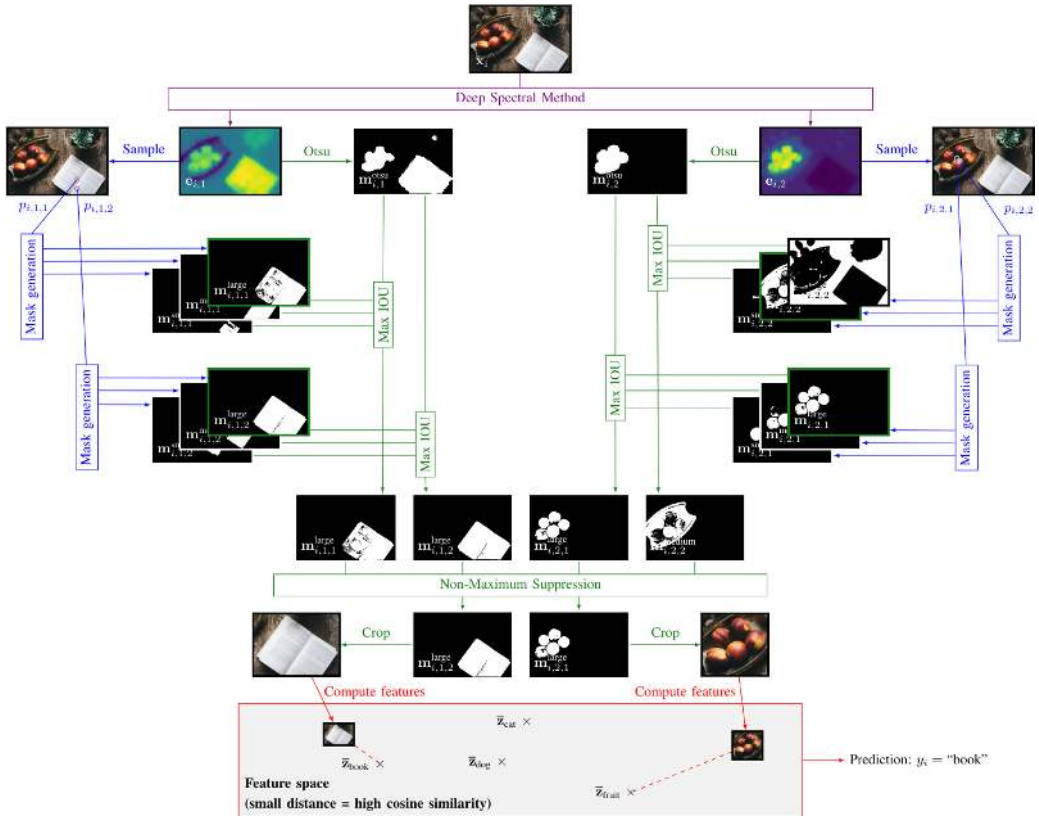


Figure 4: (violet) First, eigenmaps are produced using DSM. Each map is treated separately. (blue) Using the maps, random points are sampled and used to prompt SAM. For each point, we obtain 3 candidate masks. (green) Out of each group of 3 candidate masks, we keep the one that maximizes IOU with an Otsu thresholding of the map. Redundant masks are then filtered out using non-maximum suppression. (red) Finally, kept masks are used to compute feature representations of associated crops. A NCM is then used to return a label.

Tasks

- WP2.1 — Exploiting the diversity of instances *(6m, low risk, low gain)*

  In a direct continuation of the works submitted to EUSIPCO 2024 and ECCV 2024, we want to first address whether output of segmentation foundation models can be coupled with diversity for a better estimation of the prototypes of classes. Indeed, up to now, we have developed an ad-hoc methodology to determine and filter out segmentation masks, and have informally introduced some sort of diversity using non-maximum suppression to remove redundant masks. Here, DPPs may provide a more formal framework to provide a diverse set of elements of interest in an image, and identify more relevant points in the feature space.

  While incremental, this task gives an opportunity for the PhD student to get started efficiently, with easy potential for research publication. Therefore, this task will start right from the beginning of the project.

- WP2.2 — Exploration of positional embedding to contextualize FSL *(11m, high risk, high gain)*

  This task is much more exploratory than the previous one. The idea is to exploit positional embedding mechanisms of transformers to provide better embeddings of FSL data. In classical FSL pipelines, one generally embeds shots and queries independently, and then performs classification depending on proximity to class prototypes. Here, a first idea is to contextualize queries with the shots. In more details, for a given query $\mathbf{q}_j \in \mathcal{Q}$, a possible approach would be to create multiple sequences of tokens corresponding to pairs of tokens $\{(\mathbf{s}_i, \mathbf{q}_j)\}_i$, and to compute query embeddings created by these contextualizations. Here, we believe that deviation from a non-contextualized embedding may help define a DPP kernel that models similarity with labeled shots. Based on findings, this approach could be complemented with outputs of task WP2.1 to contextualize parts of data, and therefore propose multiple contextualized embeddings for a single query.

This task has the potential to open new directions for FSL, as it describes an approach that has not been explored in the literature before. Therefore, it has a higher risk, but can lead to very promising contributions.

1.3.4 WP3 — Selection of relevant parts in data and features

Objectives

The third WP aims at exploring which parts of data or features are useful to FSL classification. At a coarse scale, this is very related to the instance segmentation part of WP2, related to research question D3. In addition, at a finer grain, this WP tackles the question of selecting diverse data to label – related to active learning – but also of selecting specific features, with the expectation that specific features convey more useful information than other for a particular FSL task. Reformulating, this WP should help answer the question: *How can diversity help selecting parts of data and features in order to improve FSL?*

Preliminary work

Active few-shot classification [50] was introduced by collaborators of this project. In this paradigm, the goal is to select few data to label from a larger unlabeled dataset, in order to maximize performance. Authors proposed an approach that estimates the distribution of features, and then samples the most probable ones in addition to the least probable ones. Interestingly, this can be seen as a particular form of diversity. A natural question that arises is whether more diversity-oriented approaches can lead to more significant gains.

Tasks

- WP3.1 — Identify data to label from a large unlabeled dataset *(6m, low risk, low gain)*

  Similar to WP2.1, this task is a direct continuation of preliminary work in active few-shot classification [50]. Here, we want to explore the direct benefits one can gain using DPPs to select pertinent data to label from a large unlabeled dataset, with a fixed labeling budget. We naturally expect that DPPs select pertinent data for labeling, and easily lead to accuracy improvement on standard FSL tasks.

  As this task is very easy to address, this is a great opportunity to start a master student's internship. Indeed, existing codes and easily achievable experiments provide a facilitating context to familiarize with DPPs and FSL. Synchronizing the internship with the post-doc position, we believe this task will create positive interactions for efficient start of the latter too.

- WP3.2 — Determining features that are important for FSL *(12m, medium risk, medium gain)*

  Another direction that has the potential to reduce embedding size and to improve accuracy is to select particular features of data in the feature space. A model's feature representation can be seen as a decomposition of data along axes that represent contexts learned by the model. As foundation models are trained on gigantic quantities of data, they have necessary learned concepts that are not relevant to the FSL task. Therefore, we believe that correlations between feature values and shots may help defining a DPP kernel. Using this kernel, we could sample subsets of feature representations, and proceed to classification with some form of ensembling. Indeed, providing different viewpoints on data could potentially lead to better prediction.

  Additionally, this approach can be used with multiple models. Combining subsets of features from models trained on different datasets could lead to more diverse viewpoints, and allow for better ensembling classifiers.

1.3.5 WP4 — Dissemination

Objectives

Open research has become the norm in the machine learning community, and we want to contribute to this aspect. Therefore, in addition to research publications that will result from previous work packages, we want to contribute to existing libraries for DPPs.

Preliminary work

Actors of the project are involved in development of the Python library `DPPy` [83] and of the Julia library `Determinantal.jl` [84]. One objective of the project will be to extend one or both of these libraries with the findings from this work. We therefore hope that DPPs will become more prevalent in FSL literature.

AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

Tasks

- WP4.1 — Continuous integration *(continual)*

  With the help of the project partners, we will contribute to one or more of the existing DPP libraries. Choice of library will mostly depend on the language used for experiments. `DPPy` is a stronger candidate as `Python` and notably its library `pytorch` are prevalent in the deep learning community.

  As mentioned in Section 1.3.1, codes developed for experiments in research articles issued from the project will be systematically put on GitHub, and we will monitor issues to encourage the community to reuse, reproduce and compare to our work.

- WP4.2 — Transition toward medical applications *(6m, low risk, high gain)*

  While most experiments will be first made on controlled vision datasets to ease comparison with other works, we want to transition toward practical applications, notably using brain signals. Whenever pertinent, other tasks will be evaluated on such datasets. In addition, a final study of methods developed in the project on BCI MI datasets will be made to integrate findings of this project to practical applications.

1.3.6 Deliverables

- $T_0 + 6m$ — Website for the project.
- $T_0 + 18m$ — Intermediate report on findings with relation to WP2.
- $T_0 + 39m$ — Intermediate report on findings with relation to WP1.
- $T_0 + 48m$ — Final report, summarizing impact of previous work, plus new findings with relation to WP3.

1.3.7 Risks & Fallback solutions

- Late recruitment of the PhD student — In case of a late recruitment of the PhD student, permanent researchers will address WP2.1 so that the newly recruited PhD student can start on WP2.2 directly. To mitigate this risk, we have started to prospect identified cursus, notably the Computer Science department of ENS Rennes.
- Late recruitment of the post-doc researcher — In that situation, the master student will have to conduct experiments on WP3.1 alone, and will be seconded by the permanent members of the project. In case of a very late recruitment, the PhD student may also be interested to pursue the work after their defense. To avoid this situation, we will communicate on the position through our research networks.
- Impossibility to address WP1.3 — Most of the WPs in this project are independent, with the exception of WP1.3 that depends on outcomes of WP1.1 and 1.2. If findings of these tasks are not sufficient to address WP1.3, the PhD student will be re-affected to WP3.

1.3.8 Temporal organization of the project

2 Organisation and implementation of the project

2.1 Scientific coordinator and its team

2.1.1 Scientific coordinator

Bastien PASDELOUP

After defending my PhD in 2017, I joined the LTS4 laboratory at EPFL for a post-doc. I am now Associate Professor at IMT Atlantique and Lab-STICC since 2019, and joined the BRAIn team at its inception in 2022.

My research focuses on multiple aspects of deep learning, both fundamental – development of efficient models for FSL [25], understanding of how data impact FSL [10, 12, 21, 80] – and applied, generally to medical domains [70, 71, 73–78]. My interests also include graph signal processing [71, 85, 86], and I recently initiated preliminary research on DPPs with application to genetic algorithms [87]. Since my recruitment, I supervised 6 PhD theses – including 1 Cifre – (2 completed [88, 89]), 3 postdoctoral researchers – 2 funded through Carnot Foundation fundings, 1 through ANR (ASTRID) for which I am the main academic collaborator – (2 completed), 1 research engineer, and numerous master students, in addition to 2 medical theses [75, 77].

My role as a PI will be to define the important research questions for the project and determine the experimental protocols to answer them. Thanks to my expertise in FSL, I will provide evidence of the benefits of DPP approaches at different steps of FSL, and will supervise a research team built around the project.

With this project, I plan to develop research around sampling in the BRAIn team, leveraging on my expertise on FSL, and more broadly at the scale of the Lab-STICC laboratory. This project will allow me to increase my skills on DPPs, and to build a team around these aspects, to eventually be recognized as the expert on diverse sampling. Ensuring that findings generalize beyond classical vision datasets will first allow me to integrate the results in my collaborations with medical doctors, and then to develop new collaborations.

Implication of the scientific coordinator in on-going project(s)

Name of the researcher	Person.month	Call, funding agency, grant allocated	Project's title	Name of the scientific coordinator	Start – End
Bastien PASDELOUP	6.5	ANR-21-ASIA-0004-02 (142,184.00€)	Fusion de composantes hétérogènes par IA	Pierre ROMENTEAU (INPIXAL)	01/01/2022 – 31/03/2024
Bastien PASDELOUP	6	ANR-22-CE22-0016-04 (145,318.00€)	ML and metaheuristics algorithms for urban transportation	Romain BILLOT (IMT Atlantique)	01/02/2023 – 31/01/2027

2.1.2 Coordinator's team

In this project, I will be seconded by a carefully designed team, with consultative experts on all elements of the project, with their own individual interests, that will enrich discussions around the project. Additionally, expertise of the team members cover both application cases I want to address, *i.e.*, vision and brain signals.

Thanks to their strong expertise on DPPs, distant collaborators will provide important feedback on the research directions I propose. Besides, research stays will be organized in their teams for me and the PhD student recruited for the project, at key points of the projects.

Vincent GRIPON

Vincent is Professor at IMT Atlantique, Lab-STICC, and the head of the BRAIn team. He conducts research in many aspects of deep learning, including FSL [12, 21, 25, 80, 90, 91]. His expertise in training efficient models makes him a very valuable source of knowledge for all aspects of the project related to models.

Nicolas FARRUGIA

Nicolas is Associate Professor at IMT Atlantique, Lab-STICC, in the BRAIn team. He is a neuroscientist with a strong deep learning background. He will bring important insight to the project when transitioning from vision datasets to EEG and fMRI signals [70–74, 92–94].

Alexandre REIFFERS-MASSON

Alexandre is Associate Professor at IMT Atlantique, Lab-STICC in the MATHS&NET team. He has a strong interest in stochastic approximation for optimisation [95, 96] and machine learning [97, 98]. He an I had numerous discussions on DPPs for machine learning. Therefore, I know he will bring a solid mathematical viewpoint in the project and will help me start the project efficiently.

AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

Nicolas TREMBLAY

Nicolas is Researcher at CNRS, Gipsa-lab, in the GAIA team. He is a renowned expert on DPPs [57, 60, 63, 99], with interest in numerous aspects of machine learning [100–102], notably in the context of graphs. His project *GRANOLA – Determinantal Point Processes on Graphs for Randomized Numerical Linear Algebra* was supported by ANR JCJC in 2021;

Rémi BARDENET

Rémi is Researcher at CNRS, CRIStAL, in the SigMA team. He is also a renowned expert on DPPs [103, 104], with a strong interest in Monte-Carlo methods [105, 106] and in machine learning [107]. He is the PI of ERC grant *Blackjack – Fast Monte Carlo integration with repulsive point processes*, and holds a national chair in artificial intelligence: *Baccarat*. He received the CNRS bronze medal in 2021 for his work on DPPs.

2.1.3 Hired personnel

PhD student

The PhD student hired for the project will mainly work on WP1 and WP2, and will actively contribute to continuous implementation in WP4. To start efficiently, they will first address task WP2.1, for which there are already preliminary results. This should lead to quick results, as the PhD student will be provided material to apprehend deep learning models efficiently. The main effort will therefore be WP2.2 and WP1.

A good profile for a candidate would be a master student or engineer with a strong mathematical background and an interest in artificial intelligence. The typical PhD student would be from the Computer Science department of ENS, or from a machine learning-oriented cursus.

Post-doc

The post-doc researcher hired for the project will principally work on WP3, and will study parts of data and features that help FSL classification. As for the PhD student, task WP3.1 is already advanced, which should help them start efficiently and quickly upskill. In addition, they will also work on WP4.

The post-doc researcher will also second the PhD student on task WP1.3 at the end of the PhD, to increase the work force on that task. In addition, they will be seconded by the master student at the start.

The typical profile for this position is a researcher with preliminary experience on transformer models, ideally with experience in brain signals to help transitioning to such domains. Starting the work at the second quarter of 2027 has been chosen to correspond to standard PhD defense dates at the end of 2026, with a little margin.

Intern

Finally, the master student will address WP3.1 at the same time as the post-doc. The reasoning is to reinforce the team at the end of the PhD student's thesis, and to help the post-doc get started efficiently, while offering a full task to the student with potential for publication.

2.2 Implemented and requested resources to reach the objectives

Staff expenses

As detailed in Section 2.1.3, I plan to hire a PhD student (36 months, 138k€) and a post-doc researcher (18 months, 84.6k€). In addition, a master student (6 month, 3.9k€) will reinforce the team, by exploring questions related to WP3 and facilitating the work of the researchers invested on the question.

Instruments and material costs

The post-doc and PhD student will both receive a laptop and a screen for daily work (3k€ each). In addition, loading large models for experiments will require a medium-size GPU with enough VRAM, *e.g.*, NVIDIA RTX 4090(****). We plan to purchase two of these devices, to share among the PhD student, the post-doc researcher and the PI. With the rest of the computer to integrate it, this amounts to ∼5k€ each.

After the project, the GPU machines will enrich the local computing resources of the BRAIn team, and will serve subsequent projects. Laptops will be reused for new PhD students or post-doc researchers.

(****) https://www.nvidia.com/en-gb/geforce/graphics-cards/40-series/rtx-4090/.

AAPG2024	ENDIVE		JCJC
Coordinated by:	Bastien PASDELOUP	48 months	296,555€
Theme E.2 – Artificial intelligence and data science			

Overhead costs

We require 5k€ publication fees for open access publication, and 7.5k€ to cover attendance to conferences at which articles will be published. In addition, we require 4k€ to organise research stays with the research partners at CRIStAL, Lille an Gipsa-lab, Grenoble, for both the PI and the hired PhD student, at key points of the project, *i.e.* at the beginning to start efficiently, and close to publication deadlines to major machine learning conference. Finally, we add the administrative management fees and structure costs, which amount to 37,555€.

Requested means by item of expenditure and by partner

		Partner 1 – IMT Atlantique
Staff expenses, including costs of a partial release from teaching obligations	Permanent (66 p.m.)	278,917.97€
	Non-permanent (60 p.m.)	226,500€
Instruments and material costs		16,000€
Building and ground costs		0€
Outsourcing / subcontracting		5,000€
Overhead costs	Travel costs	11,500€
	Administrative management & & structure costs	37,555€
Sub-total		**575,472.97€**
Requested funding		**296,555€**

3 Impact and benefits of the project

Scientific dissemination

This project tackles a frequently disregarded aspect of deep learning, *i.e.*, the impact of sampling. It should therefore highlight the importance of carefully selected approaches, and open new research perspectives in the community. As FSL has received a lot of attention lately, contributing on fundamental aspects of the domain has the potential for publication in major conferences and journals in the deep learning community. Combined with the local expertise in FSL and with the expertise of collaborators in DPPs, plus their past experience in publishing at these venues [101, 107–109], we believe this project has the full potential to open the way to major conferences in deep learning, that have up to now been evading the BRAIn team.

Practically, contribution to existing libraries, as described in WP4 in Section 1.3.5, should allow researchers to quickly adopt solutions developed in the project, and to increase the concerns toward the use of diversity in deep learning pipelines. In addition, all codes to reproduce experiments in the articles issued from the project will be put on GitHub to encourage reproducible research, as done for years in the team.

Societal impact

Opening directions of this work toward medical domains should lead to societal improvements, as findings of this project could directly have implications to the practitioners of the domains considered. In addition, we believe that transitioning from classical vision datasets toward practical applications should lead to establishment of new partnerships with medical doctors, and have a positive impact on society.

Concretely, first openings toward brain data should lead to improvements on tasks such as motor imagery classification, that have practical applications to patients' everyday life. In addition to neuroscience, my collaborations with doctors from other domains (oncology, cardiology, ophtalmology) could possibly benefit from findings of this project, with direct implications on medical practices.

Impact on the PI position in the laboratory

Finally, this project will have a very positive impact on my positioning in the laboratory, as there is currently no expert at Lab-STICC on DPPs. This unicity in the laboratory will therefore encourage new collaborations in the future, as I would be identified as the local expert on these aspects. At the scale of the BRAIn team, I would also address complementary directions from the rest of the team, that would allow a broader coverage of efficient and low-regime deep learning. With more diversity of research directions, although with a common interest in FSL and learned representations, I believe this project has the potential to increase the team's visibility, and to contribute to its establishment as a reference in FSL worldwide.

It is also an opportunity for me to develop collaborations with Rémi BARDENET and Nicolas TREMBLAY, who are worldwide renown experts in DPPs. I hope that this project is only the first of a series.

To conclude, I believe that his project is just a first step of important developments in the FSL community. Indeed, FSL has become a very important domain, both for industrials and for academic researchers, and there are real direct consequences when one can use their data more efficiently. In addition, the rise of huge foundation models is very probably going to increase the interest in the domain. My position on sampling for FSL in such context will therefore provide me a very distinguishable profile in the worldwide research landscape.

4 References related to the project

[1] Archit Parnami and Minwoo Lee. Learning from few examples: A summary of approaches to few-shot learning. *arXiv:2203.04291*, 2022.
[2] Y. Song, T. Wang, P. Cai, S.K. Mondal, and J.P. Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*, 2023.
[3] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 2011.
[4] Sion An, Soopil Kim, Philip Chikontwe, and Sang Hyun Park. Few-shot relation learning with attention for eeg-based motor imagery classification. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.
[5] Xiang Wang, Shiwei Zhang, Hangjie Yuan, Yingya Zhang, Changxin Gao, Deli Zhao, and Nong Sang. Few-shot action recognition with captioning foundation models. *arXiv:2310.10125*, 2023.
[6] Rafael EP Ferreira, Yong Jae Lee, and João RR Dórea. Using pseudo-labeling to improve performance of deep neural networks for animal identification. *Scientific Reports*, 2023.
[7] HuggingFace. Lmsys chatbot arena leaderboard, 2024. https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard.
[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929*, 2020.
[9] OpenAI. Video generation models as world simulators, 2024. https://openai.com/research/video-generation-models-as-world-simulators.
[10] Y. Bendou, L. Drumetz, V. Gripon, G. Lioi, and B. Pasdeloup. Disambiguation of one-shot visual classification tasks: A simplex-based approach. *arXiv:2301.06372*, 2023.
[11] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 2016.
[12] Yassir Bendou, Vincent Gripon, Bastien Pasdeloup, Giulia Lioi, Lukas Mauch, Stefan Uhlich, Fabien Cardinaux, Ghouthi Boukli Hacene, and Javier Alonso Garcia. A statistical model for predicting generalization in few-shot classification. In *31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023.
[13] A. Kulesza and B. Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 2012.
[14] C. Zhang, H. Kjellstrom, and S. Mandt. Determinantal point processes for mini-batch diversification. *arXiv:1705.00607*, 2017.
[15] X. Zhan, Q. Li, and A.B. Chan. Multiple-criteria based active learning with fixed-size determinantal point processes. *arXiv:2107.01622*, 2021.
[16] J. Dodge, K. Jamieson, and N.A. Smith. Open loop hyperparameter optimization and determinantal point processes. *arXiv:1706.01566*, 2017.
[17] Thomas Müller, Guillermo Pérez-Torró, Angelo Basile, and Marc Franco-Salvador. Active few-shot learning with fasl. In *International Conference on Applications of Natural Language to Information Systems*. Springer, 2022.
[18] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM computing surveys (CSUR)*, 2021.
[19] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *International Conference on Computer Vision*, 2023.
[20] Yossi Gandelsman, Alexei A Efros, and Jacob Steinhardt. Interpreting clip's image representation via text-based decomposition. *arXiv:2310.05916*, 2023.
[21] Raphael Lafargue, Yassir Bendou, Bastien Pasdeloup, Jean-Philippe Diguet, Ian Reid, Vincent Gripon, and Jack Valmadre. Few and fewer: Learning better from few examples using fewer base classes. *arXiv:2401.15834*, 2024.
[22] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 2017.
[23] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 2020.
[24] Jesper Bäck. Domain similarity metrics for predicting transfer learning performance, 2019.
[25] Y. Bendou, Y. Hu, R. Lafargue, G. Lioi, B. Pasdeloup, S. Pateux, and V. Gripon. Easy—ensemble augmented-shot-y-shaped learning: State-of-the-art few-shot classification with simple components. *Journal of Imaging*, 2022.
[26] Zhiqiang Gong, Ping Zhong, and Weidong Hu. Diversity in machine learning. *Ieee Access*, 2019.
[27] Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim, and Timothy M Hospedales. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In *Conference on Computer Vision and Pattern Recognition*, 2022.
[28] Hassan Gharoun, Fereshteh Momenifar, Fang Chen, and Amir H Gandomi. Meta-learning approaches for few-shot learning: A survey of recent advances. *arXiv:2303.07502*, 2023.
[29] Jinfang Jia, Xiang Feng, and Huiqun Yu. Few-shot classification via efficient meta-learning with hybrid optimization. *Engineering Applications of Artificial Intelligence*, 2024.
[30] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*. PMLR, 2017.
[31] Xiaoxu Li, Xiaochen Yang, Zhanyu Ma, and Jing-Hao Xue. Deep metric learning for few-shot image classification: A review of recent developments. *Pattern Recognition*, 2023.

[32] Eric Xing, Michael Jordan, Stuart J Russell, and Andrew Ng. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*, 15, 2002.
[33] Chengming Xu, Siqian Yang, Yabiao Wang, Zhanxiong Wang, Yanwei Fu, and Xiangyang Xue. Exploring efficient few-shot adaptation for vision transformers. *arXiv:2301.02419*, 2023.
[34] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv:1903.03096*, 2019.
[35] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision (IJCV)*, 2022.
[36] X. Luo, H. Wu, J. Zhang, L. Gao, J. Xu, and J. Song. A closer look at few-shot classification again. *arXiv:2301.12246*, 2023.
[37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 2021.
[38] Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, and Shlomo Dubnov. Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.
[39] Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. Imagebind: One embedding space to bind them all. In *Conference on Computer Vision and Pattern Recognition*, 2023.
[40] Huayang Li, Siheng Li, Deng Cai, Longyue Wang, Lemao Liu, Taro Watanabe, Yujiu Yang, and Shuming Shi. Textbind: Multi-turn interleaved multimodal instruction-following in the wild. 2023.
[41] Fan Liu, Tianshu Zhang, Wenwen Dai, Wenwen Cai, Xiaocong Zhou, and Delong Chen. Few-shot adaptation of multi-modal foundation models: A survey. *arXiv:2401.01736*, 2024.
[42] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 2002.
[43] Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 2008.
[44] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 2019.
[45] Shengyun Wei, Shun Zou, Feifan Liao, et al. A comparison on data augmentation methods based on deep learning for audio classification. In *Journal of physics: Conference series*. IOP Publishing, 2020.
[46] Fedor Zhdanov. Diverse mini-batch active learning. *arXiv:1901.05954*, 2019.
[47] Nils Bjorck, Carla P Gomes, Bart Selman, and Kilian Q Weinberger. Understanding batch normalization. *Advances in neural information processing systems*, 2018.
[48] Zhaodong Chen, Lei Deng, Guoqi Li, Jiawei Sun, Xing Hu, Xin Ma, and Yuan Xie. Batch normalization sampling. *arXiv:1810.10962*, 2018.
[49] Jingyi Xu and Hieu Le. Generating representative samples for few-shot classification. In *Conference on Computer Vision and Pattern Recognition*, 2022.
[50] A. Abdali, V. Gripon, L. Drumetz, and B. Boguslawski. Active few-shot classification: a new paradigm for data-scarce learning settings. *arXiv:2209.11481*, 2022.
[51] Yuli Liu, Christian Walder, and Lexing Xie. Determinantal point process likelihoods for sequential recommendation. In *45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022.
[52] Dhruv Batra, Payman Yadollahpour, Abner Guzman-Rivera, and Gregory Shakhnarovich. Diverse m-best solutions in markov random fields. In *12th European Conference on Computer Vision*. Springer, 2012.
[53] Yingjie Gu, Zhong Jin, and Steve C Chiu. Active learning combining uncertainty and diversity for multi-class image classification. *IET Computer Vision*, 2015.
[54] Zdravko Botev and Ad Ridder. Variance reduction. *Wiley statsRef: Statistics reference online*, 2017.
[55] Raghunath Arnab. *Survey sampling theory and applications*. Academic Press, 2017.
[56] A. Kulesza and B. Taskar. k-dpps: Fixed-size determinantal point processes. In *International Conference on Machine Learning (ICML)*, 2011.
[57] Nicolas Tremblay, Simon Barthelmé, Konstantin Usevich, and Pierre-Olivier Amblard. Extended l-ensembles: a new representation for determinantal point processes. *The Annals of Applied Probability*, 2023.
[58] J Ben Hough, Manjunath Krishnapur, Yuval Peres, and Bálint Virág. Determinantal processes and independence. *math/0503110*, 2005.
[59] Michal Derezinski, Daniele Calandriello, and Michal Valko. Exact sampling of determinantal point processes with sublinear time preprocessing. *Advances in neural information processing systems*, 2019.
[60] S. Barthelme, N. Tremblay, and P.O. Amblard. A faster sampler for discrete determinantal point processes. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023.
[61] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. Monte carlo markov chain algorithms for sampling strongly rayleigh distributions and determinantal point processes. In *Conference on Learning Theory*. PMLR, 2016.
[62] Chengtao Li, Suvrit Sra, and Stefanie Jegelka. Fast mixing markov chains for strongly rayleigh measures, dpps, and constrained sampling. *Advances in Neural Information Processing Systems*, 2016.
[63] N. Tremblay, S. Barthelmé, and P.O. Amblard. Determinantal point processes for coresets. *J. Mach. Learn. Res.*, 2019.
[64] V. Nguyen, T. Le, M. Yamada, and M.A. Osborne. Optimal transport kernels for sequential and parallel neural architecture search. In *International Conference on Machine Learning (ICML)*. PMLR, 2021.
[65] S. Ravi and H. Larochelle. Meta-learning for batch mode active learning. 2018.
[66] R. Kumar, T. Deleu, and Y. Bengio. The effect of diversity in meta-learning. In *AAAI Conference on Artificial Intelligence*, 2023.
[67] Y. Lv, B. Zhang, X. Yue, Z. Xu, and W. Liu. Ensemble of adapters for transfer learning based on evidence theory. In *International Conference on Belief Functions*. Springer, 2021.
[68] K. Yang, J. Lu, W. Wan, G. Zhang, and L. Hou. Transfer learning based on sparse gaussian process for regression. *Information Sciences*, 2022.
[69] J. Ye, Z. Wu, J. Feng, T. Yu, and L. Kong. Compositional exemplars for in-context learning. *arXiv:2302.05698*, 2023.
[70] M. Ménoret, N. Farrugia, B. Pasdeloup, and V. Gripon. Evaluating graph signal processing for neuroimaging through classification and dimensionality reduction. In *Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017.
[71] Y. El Ouahidi, H. Tessier, G. Lioi, N. Farrugia, B. Pasdeloup, and V. Gripon. Pruning graph convolutional networks to select meaningful graph frequencies for fmri decoding. In *European Signal Processing Conference (EUSIPCO)*. IEEE, 2022.
[72] M. Bontonou, G. Lioi, N. Farrugia, and V. Gripon. Few-shot decoding of brain activation maps. In *European Signal Processing Conference (EUSIPCO)*. IEEE, 2021.

[73] Y. El Ouahidi, L. Drumetz, G. Lioi, N. Farrugia, B. Pasdeloup, and V. Gripon. Spatial graph signal interpolation with an application for merging bci datasets with various dimensionalities. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.
[74] Y. El El Ouahidi, V. Gripon, B. Pasdeloup, G. Bouallegue, N. Farrugia, and G. Lioi. A strong and simple deep learning baseline for bci mi decoding. *arXiv:2309.07159*, 2023.
[75] Victor Quéré. *Comparison of logistic regression analysis with random forest for prediciton of ICH early mortality in the population-based Brest Stroke Registry*. PhD thesis, Centre Hospitalier Universitaire de Brest, 2021.
[76] Y. El Ouahidi, M. Feller, M. Talagas, and B. Pasdeloup. Une approche pour regrouper des sujets atteints de cancers, sur la base des similarités des répartitions cellulaires au sein de leurs biopsies. *Extraction et Gestion des Connaissances (EGC)*, 2021.
[77] Amine El Ouahidi. *Utilisation de méthodes d'apprentissage automatique pour prédire l'implantation d'un pacemaker dans les suites d'une procédure de TAVI*. PhD thesis, Centre Hospitalier Universitaire de Brest, 2023.
[78] O Weizman, T Pezel, K Hamzi, G Schurtz, M Hauguel-Moreau, A Trimaille, S Toupin, T Bochaton, S Attou, C Meune, et al. Machine-learning score to predict in-hospital outcomes in patients hospitalized in intensive cardiac care unit. *European Heart Journal*, 2023.
[79] Samuel Challéat, Nicolas Farrugia, Amandine Gasc, Jeremy Froidevaux, Nicolas Pajusco, Jennifer Hatlauf, Frank Dziock, Adrien CHARBONNEAU, Juliette Linossier, Christopher Watson, and et al. Silent·cities, 2024.
[80] Yassir Bendou, Bastien Pasdeloup, Giulia Lioi, Vincent Gripon, Fabien Cardinaux, Ghouthi Boukli Hacene, and Lukas Mauch. Inferring latent class statistics from text for robust visual few-shot learning. In *R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Large Foundation Models*, 2023.
[81] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *International Conference on Computer Vision*, 2023.
[82] Luke Melas-Kyriazi, Christian Rupprecht, Iro Laina, and Andrea Vedaldi. Deep spectral methods: A surprisingly strong baseline for unsupervised semantic segmentation and localization. In *Conference on Computer Vision and Pattern Recognition*, 2022.
[83] Guillaume Gautier, Guillermo Polito, Rémi Bardenet, and Michal Valko. DPPy: DPP Sampling with Python. *Journal of Machine Learning Research - Machine Learning Open Source Software (JMLR-MLOSS)*, 2019.
[84] Simon Barthelmé, Nicolas Tremblay, and Guillaume Gautier. Determinantal.jl: Determinantal point processes in julia, 2023. https://github.com/dahtah/Determinantal.jl/tree/main.
[85] Bastien Pasdeloup. *Extending convolutional neural networks to irregular domains through graph inference*. PhD thesis, Ecole nationale supérieure Mines-Télécom Atlantique, 2017.
[86] C. Lassance, M. Bontonou, M. Hamidouche, B. Pasdeloup, L. Drumetz, and V. Gripon. Graphs as tools to improve deep learning methods. *Advances in Artificial Intelligence: Reviews*, 2022.
[87] M. Guzmán, K.N. Bakirci, M. Rouesne, H. Savatier-Dupré, B. Pasdeloup, and P. Meyer. Utilisation de processus ponctuels déterminantaux pour la sélection de parents dans un algorithme génétique. In *GRETSI*, 2023.
[88] Maryam Karimi Mamaghan. *Hybridizing metaheuristics with machine learning for combinatorial optimization: a taxonomy and learning to select operators*. PhD thesis, Ecole nationale supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire, 2022.
[89] Fred Michael Gonsalves. *Leveraging machine learning and operations research to enhance cruise ship energy efficiency*. PhD thesis, Ecole nationale supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire, 2023.
[90] Y. Hu, V. Gripon, and S. Pateux. Leveraging the feature distribution in transfer-based few-shot learning. In *International Conference on Artificial Neural Networks*. Springer, 2021.
[91] P.E. Novac, G. Boukli Hacene, A. Pegatoquet, B. Miramond, and V. Gripon. Quantization and deployment of deep neural networks on microcontrollers. *Sensors*, 2021.
[92] N. Mendes, S. Oligschläger, M.E. Lauckner, J. Golchert, J.M. Huntenburg, M. Falkiewicz, M. Ellamil, S. Krause, et al. A functional connectome phenotyping dataset including cognitive state and personality measures. *Scientific data*, 2019.
[93] A. Brahim and N. Farrugia. Graph fourier transform of fmri temporal signals based on an averaged structural connectome for the classification of neuroimaging. *Artificial Intelligence in Medicine*, 2020.
[94] Y. Zhang, N. Farrugia, and P. Bellec. Deep learning models of cognitive processes constrained by human brain connectomes. *Medical Image Analysis*, 2022.
[95] A. Reiffers-Masson, T. Chonavel, and Y. Hayel. Estimating fiedler value on large networks based on random walk observations. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.
[96] V.S. Borkar and A. Reiffers-Masson. Opinion shaping in social networks using reinforcement learning. *IEEE Transactions on Control of Network Systems*, 2021.
[97] S. Ghandi, A. Reiffers-Masson, S. Vaton, and T. Chonavel. Neural collaborative filtering for network delay matrix completion. In *International Conference on Network and Service Management (CNSM)*. IEEE, 2022.
[98] C. Troisemaine, V. Lemaire, S. Gosselin, A. Reiffers-Masson, J. Flocon-Cholet, and S. Vaton. Novel class discovery: an introduction and key concepts. *arXiv:2302.12028*, 2023.
[99] N. Tremblay, P.O. Amblard, and S. Barthelmé. Graph sampling with determinantal processes. In *European Signal Processing Conference (EUSIPCO)*. IEEE, 2017.
[100] N. Tremblay and A. Loukas. Approximating spectral clustering via sampling: a review. *Sampling Techniques for Supervised or Unsupervised Tasks*, 2020.
[101] N. Tremblay, G. Puy, R. Gribonval, and P. Vandergheynst. Compressive spectral clustering. In *International Conference on Machine Learning (ICML)*. PMLR, 2016.
[102] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst. Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis*, 2018.
[103] R. Bardenet and A. Hardy. Monte carlo with determinantal point processes. 2020.
[104] A. Belhadji, R. Bardenet, and P. Chainais. A determinantal point process for column subset selection. *Journal of Machine Learning Research*, 2020.
[105] R. Bardenet, A. Doucet, and C. Holmes. Towards scaling up markov chain monte carlo: an adaptive subsampling approach. In *International Conference on Machine Learning (ICML)*. PMLR, 2014.
[106] R. Bardenet, A. Doucet, and C. Holmes. On markov chain monte carlo methods for tall data. *Journal of Machine Learning Research*, 2017.
[107] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 2011.
[108] Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay. Revisiting the bethe-hessian: improved community detection in sparse heterogeneous graphs. *Advances in neural information processing systems*, 2019.
[109] Rémi Bardenet, Mátyás Brendel, Balázs Kégl, and Michele Sebag. Collaborative hyperparameter tuning. In *International conference on machine learning*. PMLR, 2013.