SORBONNE UNIVERSITÉ

# PROGRAMME INTITUTS ET INITIATIVES

## Appel à projet – campagne 2021

## Proposition de projet de recherche doctoral (PRD)

**Intitulé du projet de recherche doctoral (PRD):**

Memorization in Deep Learning

**Directeur.rice de thèse porteur.euse du projet (titulaire d'une HDR) :**


NOM : Lamprier Prénom : Sylvain

Titre : Assistant Professor

e-mail : sylvain.lamprier@lip6.fr

Adresse professionnelle : LIP6, Jussieu, Bat 26, 5eme étage


**Unité de Recherche : LIP6**

Intitulé : Laboratoire d'Informatique Parisien

Code : UMR7606

**École Doctorale de rattachement de l'équipe (future école doctorale du.de la doctorant.e) : EDITE**

**Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) :**

- Vincent Grari (début janvier 2019) en Cifre AXA - 50%
- Agnès Mustar (début septembre 2020) contrat ANR - 50%
- Jean-Yves Francheschi (début septembre 2018) bourse ministère - 50%
- Manon Cesaire (début Novembre 2020) IRT system X - 50%

SORBONNE UNIVERSITÉ

---

**Co-encadrant.e :**

NOM : Oyallon Prénom : Edouard

Titre : Junior Research Scientist HDR

e-mail : edouard.oyallon@lip6.fr

**Unité de Recherche : LIP6**

Intitulé : Laboratoire d'Informatique Parisien

Code *: UMR7606*

**École Doctorale de rattachement : EDITE**

Ou si ED non Alliance SU :

**Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) : Néant.**

**Selon vous, ce projet est-il susceptible d'intéresser une autre Initiative ou un autre Institut ?**
Oui, c'est une thématique nationale prioritaire de recherche

## Description du projet de recherche doctoral *(en français ou en anglais) :*

*Ce texte sera diffusé en ligne : il ne doit pas excéder 3 pages et est écrit en interligne simple.*

*Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.*

*Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le projet.*
*Préciser le profil d'étudiant(e) recherché.*

**Cotutelle internationale :** Non

**Merci d'enregistrer votre fichier au format PDF et de le nommer :**
**«ACRONYME de l'initiative/institut – AAP 2021 – NOM Porteur.euse Projet »**

***Fichier envoyer simultanément par e-mail à l'ED de rattachement et au programme : cd_instituts_et_initiatives@listes.upmc.fr avant le 20 février.***

# Memorization in Deep Learning

**Advisors** 1: Sylvain Lamprier (LIP6), 2: Edouard Oyallon (CNRS, LIP6),

**Key words** Statistical signal processing, Deep learning, Generalization

**To apply, send a CV + grade transcripts if relevant + recommendation letters to edouard.oyallon@lip6.fr and sylvain.lamprier@lip6.fr**

**Introduction** Deep Neural Networks obtain outstanding performances on many benchmarks, yet the key ingredient of their success remains unknown. This is mainly due to the high dimensional nature of those objects: they have a lot of parameters $D$ and use very large inputs $d$. By now, without loss in generality, we will focus on Neural Networks $\Phi$ learned for a classification task and which have been fed with $N$ samples. The weights of a Neural Network are specified via supervision and those networks tend to generalize well on a new test set: it implies those architectures have memorized important attributes from a dataset. During this PhD, we propose to study those attributes both from a theoretical and numerical point of view: what is their nature, how are they learned, how are they stored? We aims at studying two types of mechanisms which can be addressed independently while being neatly connected: the memorization through the symmetries of a supervised or unsupervised task, and the memorization through the data. Interestingly, any improvement concerning one aspect will benefit on the other aspect.

**Memorization?** It is an ill-defined concept that we would like to refine in a rigorous manner during this phd. First, one should point out that we do not refer to the memorization which can occur as a form of overfitting during a learning procedure. Instead, our objective is to derive a low-complexity class (in the sens of generalization) of models able to reach state-of-the-art performance on ImageNet: how do deep neural networks memorize the good attributes of the data? We propose to address it either via the notion of symmetry of the level sets of a supervised objective, either via a simplified model of the data.

**Memorizing symetries** A first type of mechanism that we propose to study is the idea of memorization through symmetries $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}^d$, which are operators that preserve the level sets learned by $\Phi$ [4], e.g.:

$$\forall x \in \mathbb{R}^d, \Phi \mathcal{L} x = \Phi x \, .$$

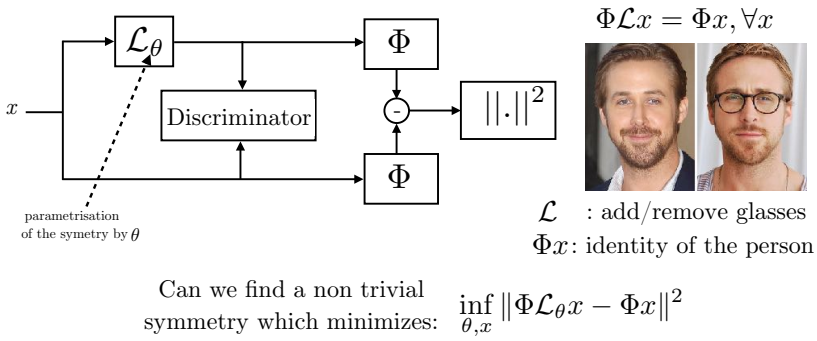


Figure 1: A *discriminator* tries to discriminate if the images $\mathcal{L}_\theta x$ or $x$ are real, while minimizing the distance between the representations $\Phi\mathcal{L}_\theta x$ and $\Phi x$ [9].

As a first step, we propose to exhibit such $\mathcal{L}$. Classical examples of symmetries are typically linear (on signals such as images): translation, rotation, scaling and any compositions of those; on the other hand, exhibiting non-linear symmetries is more complex because, in the current state of science for images, there is no explicit formulations of those and this should require learning. We propose to build, exhibit and study such symmetries. An initial line of work, through the Interferometric Graph Transform [7] seems to be a good framework to understand how those symmetries build invariance.

For instance, one could try to parametrize a symetry as a Convolutional Neural Networks using simplified data. Let us give an example of such dataset: for instance, consider faces images $x_1, ..., x_n$ related to individuals $f(x_1), ..., f(x_n)$ for which faces have or do not have glasses. Let us name $\mathcal{L}$ the involution that consists in adding or removing a glass, in this case, as it does preserve the identity:

$$f(\mathcal{L}x_i) = f(x_i), \forall i\,.$$

Assume we learn a model $\Phi$ to estimate $f$. The questions we would like to answer are as follow: can we estimate $\mathcal{L}$ from $f$? Can we estimate $\mathcal{L}$ from $\Phi$? An illustration is given Figure 1. Answering those questions would help to design a rigorous model of the data.

**Data memorization** Another approach consists in delimiting the complexity class spanned by a set of Neural Networks architectures, which leads to generalization bounds. The main principle is to reduce the complexity of such classes, without introducing a significant bias during the classification process. One can distinguish 3 types of methods: data-dependent, algorithm-agnostic (e.g., via Rademacher complexity), algorithm and data-agnostic (e.g., VC dimension, PAC-bayesian) and algorithm independent (e.g., Langevin). The community has focused a lot on those approaches to obtain some reasonable bounds [6, 1**?** ]. However, many recents work point out the fact that those bounds remain ac-

tually vacuous [12, 5, 3], and they are not numerically significant. Indeed, such approachs lack of specifity to the data structures. We propose to refine them by proposing some (simple) data models to explain memorization phenomenons which occurs in Deep Learning. An interesting initial model could be obtained via patch-based methods: those methods strive for simplicity, while allowing high-accuracy models [10].

This work could help us to address the following points: is the underlying data distribution low-dimensional? Is it possible to recover a low-dimensional embedding from a supervisedly learned Neural Network? We would like to propose a low-complexity data-dependent set of models $\mathcal{H}_N$, such that one can find a $\Phi \in \mathcal{H}_N$ leading to good performances. Note that this approach relies on a good modelization of the data.

For example, some models that we already know and which are appealing for addressing this task are the models learned via sparsity. The model of [11, 8] consists in a Scattering Transform applied on patches, followed by a local encoder of those patches. Another line of work [2] seems to suggest that patches, which are small projections of a signal, contain all the necessary information for state-of-the-art classification accuracies. However, the underlying data model remains unclear as well as the underlying mathematical principles which lead to such good performances. In a first step, one proposes to answer the following questions: do the patches of an image lie on some mathematical structure such as a manifold? Are they too high-dimensional and require some subsequent encoding? We would like to use classical tools from the signal processing and statistical literature to answer those questions.

**Outline** We will address those two lines of research, by first sudying the connexion between generalization bounds and the notion of invariance to symmetries (1 year). Then, we hope to obtain non-vacuous generalization bounds for images thanks to a simplified class of model $\mathcal{H}_N$ (1 year). At least 1 year will be as well dedicated to numerically verify our claims.

**Mini-bio** Edouard Oyallon is a junior researcher at CNRS and a specialist of the topic of this phd, who got recruited in fall 2019. Sylvain Lamprier is an assistant professor (HDR), specialist in bayesian inference for structured data, which will be particularly helpful for the first axis.

**Outcome** Addressing such issues can be useful for at least 2 major applications of machine learning: small data settings and interpretability, because they rely significantly on the complexity of the models used. A specific attention will be shade for obtaining a theory which reduces the gap between the numerical experiments and theoretical results, in order to avoid vacuous bounds.

**Desired profile** We are looking for a student with a strong record and who graduated in the field of applied mathematics, computer science or machine learning.

# References


[1] Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems*, pages 6240–6249, 2017.

[2] Wieland Brendel and Matthias Bethge. Approximating CNNs with bag-of-local-features models works surprisingly well on imagenet. In *International Conference on Learning Representations*, 2019.

[3] David Krueger, Nicolas Ballas, Stanislaw Jastrzebski, Devansh Arpit, Maxinder S Kanwal, Tegan Maharaj, Emmanuel Bengio, Asja Fischer, and Aaron Courville. Deep nets don't learn via memorization. 2017.

[4] Stéphane Mallat. Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150203, 2016.

[5] Vaishnavh Nagarajan and J Zico Kolter. Uniform convergence may be unable to explain generalization in deep learning. In *Advances in Neural Information Processing Systems*, pages 11611–11622, 2019.

[6] Behnam Neyshabur, Srinadh Bhojanapalli, David McAllester, and Nati Srebro. Exploring generalization in deep learning. In *Advances in Neural Information Processing Systems*, pages 5947–5956, 2017.

[7] Edouard Oyallon. Interferometric graph transform: a deep unsupervised graph representation. In *International Conference on Machine Learning*, pages 7434–7444. PMLR, 2020.

[8] Edouard Oyallon, Eugene Belilovsky, and Sergey Zagoruyko. Scaling the scattering transform: Deep hybrid networks. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[9] Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. Discriminative adversarial search for abstractive summarization. In *International Conference on Machine Learning*, pages 8555–8564. PMLR, 2020.

[10] Louis Thiry, Michael Arbel, Eugene Belilovsky, and Edouard Oyallon. The unreasonable effectiveness of patches in deep convolutional kernels methods. In *International Conference on Learning Representation (ICLR 2021)*, 2021.

[11] John Zarka, Louis Thiry, Tomas Angles, and Stephane Mallat. Deep network classification by scattering and homotopy dictionary learning. In *International Conference on Learning Representations*, 2020.

[12] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.