

# Perceptually Based Frugal Models for Low-Carbon AI

3-year PhD position

## General Context and Challenges

Artificial Intelligence (AI) tools have become ubiquitous in today society. AI allows computers to perform “intelligent” tasks such as decision making, problem solving, perception/identification and even understanding human communication and behavior. At the same time, the impact of AI on the environment has become non-negligible because of its carbon footprint. Indeed, the carbon footprint of computer systems is that of the part of fossil sources in the production and consumption of electricity. Moreover, fossil sources currently represent a significant part of the energy mix. For example, the rate of production of renewable electricity close to 100% is only reached in Iceland or Norway within Europe. In other countries, the decarbonization of electricity is only expected several decades into the future. When it comes to AI, its carbon footprint and role in climate change has recently come under intense scrutiny [1]. Awareness was triggered by the observation that training a single large language model (NLP) could approach 300 tons of emissions of carbon dioxide [2] – that is five times the lifetime emissions of an average car. Compared to other industrial sectors the AI impact is also non-negligible: for example, the AI environmental impact already globally exceeds that of air transport. Indeed, AI, being computationally intensive, is a major carbon emitter.

Rather than being data hungry, a frugal AI improves energy efficiency, thus addressing a significant challenge given the widespread use of machine learning. Neuromorphic computing is one of the many avenues explored by contemporary research. This approach, inspired by the efficient structure and functioning of the human brain, completely rethinks the physical architecture supporting deep learning. In this field, we focus on computer vision to draw inspiration from mammalian perception. A widely used comparison is that a child does not need to see thousands of cats to learn to recognize a cat. De facto, a less data-intensive algorithm would consume less energy, but the search for frugality goes even further. More concretely, energy consumption will be reduced on several levels: smaller model size, far shorter training time, and reduced inference time. As a side product, we also expect enhanced robustness to adversarial attacks.

## Project description :

The contemporary explosion of AI technology was unleashed by the availability of powerful computational facilities that allowed increasing the size of datasets to unprecedented sizes. AI models are tuned via gradient descent optimization techniques. Because these methods operate as searchlight devices, optimization is purely mathematical. Consequently, apparently highly scoring AI models may “pay attention” to surprisingly irrelevant parts of images. This flaw undermines the liability which is typically compensated by using ever larger datasets. This, in return, directly impacts the energetic consumption of such models.

We propose to address the above flaw more directly and effectively, via the development of perceptually based models. Like the human visual system, these models are specifically sensitive to perceptually pertinent features, such as textures, object contours, and their spatial arrangements. Research in perception has a long history. The links between perception and image processing started with detecting perceptually meaningful events. According to a long-standing principle in sensory processing, every large image deviation from “uniform noise” should be perceptible, provided this large deviation corresponds to an a priori fixed list of geometric structures (lines, curves, closed curves, convex sets, spots, local groups). Desolneux, et al. [3] explored the connection between this principle and image processing in a probabilistic setting for the detection of perceptual contours in natural images. Instead of using a prior to model the observations, they proceeded by modeling the noise. Structures that deviate largely from the noise model are deemed perceptually meaningful. This approach

supports detection of object boundaries from natural images. A link between this probabilistic approach and Mathematical Morphology has been proposed by Dokladal [4] to detect cracks in materials.

AI models too can be constrained to be sensitive to perceptually significant primitives, like lines or edges. This notion, however, runs counter to mainstream views that constrained models cannot match the performance of unconstrained models. However, at the negligible cost of small score reductions, one can obtain interesting properties when a model is tuned/constrained to target some desirable function. For example, the incorporation of modules inspired by biology has been shown to confer robustness to deep networks [5]. Further attempts at constraining AI models to perceptually significant features, developed in an effort to obtain invariance to rotation [6][7], produced interesting results in terms of 1) size of the model and 2) computational requirements. The latter factor is particularly interesting in relation to lowering the carbon footprint of AI models.

Visual processing of simple image elements (such as lines and edges) does not happen inside a cognitive vacuum: it may differ when those simple elements are embedded within natural scenes that look more like what we see every day, as opposed to the featureless backgrounds that are normally used in the laboratory. We know a good amount about the mechanisms that support vision in a simple setup (i.e. involving a simple stimulus with no natural meaningful content). We know virtually nothing about how those mechanisms may change and/or be augmented/replaced by new mechanisms under conditions that are closer to natural vision (i.e. when the image starts making sense and contains recognizable objects). In this thesis we will study how visual primitives (lines, edges, junctions) interact and how their spatial relations could be used by an AI model to efficiently use the semantic information [8] in the image to recognize objects and scenes.

Going towards these objectives we will constraint a model to operate on perceptual primitives such as lines. The features sensitive to these primitives will be fitted to data via learning. A promising tool for efficient encoding of spatial arrangements and relations are graph convolutional networks (GCN), introduced by Bruna et al. [9] and developed later by Kipf and Welling [10] to architecture that later became known as GCN. Since [10], the graph topology understanding remained on the level of immediate neighbors until Zhu et al. [11] proposed the H2GCN to encode a high-order network information from middle layers, and Qian et al. [12] explored that the performance of GCNs is related to the alignment among features, graph, and ground truth. Recently Wang et al. [13] proposed to integrate the graph motif-structure information into the convolution operation of each layer.

We will start on toy datasets and go progressively towards more complex datasets containing natural scenes. Various GCN architectures will be tested, and data ablation and model ablation will show how spatial relations can be learnt by the model to recognize objects. Further research on complex natural images will clarify how visual perception in mammals can help the design of frugal models requiring less data and time to learn.

The potential benefits of encoding the geometry and topology of perceptually significant primitives from image into a graph are significant. Indeed, a perceptually based model will not only be more efficient in terms of computational requirements, but it will also become data frugal, faster to train, and more robust to adversarial attacks. Such models will pave the way for a sustainable future with energy efficient, environment friendly AI.

This PhD is a new, emerging collaboration between two PSL institutions: the MINES Paris PSL, and the ENS PSL. These laboratories involved in this PhD support complementary expertise in artificial intelligence and visual perception.

**Acknowledgment:** The funding is provided by the TTI.5 Transition Institute

**PhD Advisors :**

**DOKLADAL Petr**, Domain : Intelligence Artificielle / Informatique

Affiliation : CMM / Mathématique et Systèmes / MINES Paris PSL

**NERI Peter**, Domain : Perception / Sciences cognitives

Affiliation : Laboratoire des systèmes perceptifs / Département d'études cognitives / ENS PSL

**Doctoral School :**

ISMEE Mathématiques et Systèmes, École doctorale 621, Ingénierie des Systèmes, Matériaux, Mécanique, Énergétique

**Application condition:** finished M2 programme, AI coding skills (tensorflow, pytorch), with excellent academic record

**Application procedure:** a prospective candidate should send his curriculum, academic track, motivation letter, two reference letters to : petr.dokladal@minesparis.psl.eu

#### References :

- [1] P. Dhar, The carbon impact of artificial intelligence, *Nature Machine Intelligence* 2, 423–425 (2020). <https://doi.org/10.1038/s42256-020-0219-9>
- [2] E. Strubell, A. Ganesh, and A. McCallum, Energy and policy considerations for deep learning in NLP, (2019) <https://doi.org/10.48550/arXiv.1906.02243>
- [3] A. Desolneux, L. Moisan and J.-M. Morel, Edge Detection by Helmholtz Principle, *Journal of Mathematical Imaging and Vision* 14: 271–284, 2001
- [4] P. Dokladal, <https://hal-mines-paristech.archives-ouvertes.fr/hal-01478089/document>
- [5] J. Dapello, T. Marques, M. Schrimpf, F. Geiger, D. Cox and J. Dicarolo, Simulating a Primary Visual Cortex at the Front of CNNs Improves Robustness to Image Perturbations, 2020, DOI: 10.1101/2020.06.16.154542.
- [6] R. Rodriguez Salas, P. Dokládál and E. Dokladalova, Rotation Invariant Networks for Image Classification for HPC and Embedded Systems, *Electronics*, 2021, DOI: <https://doi.org/10.3390/electronics10020139>
- [7] R. Rodriguez Salas, P. Dokládál and E. Dokladalova, A minimal model for classification of rotated objects with prediction of the angle of rotation, *J. of Visual Communication and Image Representation*, 2021, <https://doi.org/10.1016/j.jvcir.2021.103054>
- [8] P. Neri, Semantic control of feature extraction from natural scenes. *Journal of Neuroscience*, 34, 2374-2388, 2014
- [9] Bruna, J., Zaremba, W., Szlam, A. & LeCun, Y. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv: 1312.6203 (2013).
- [10] Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)* (2017).
- [11] Zhu, J. et al. Beyond homophily in graph neural networks: Current limitations and effective designs. *Adv. Neural Inf. Process. Syst.* 33, 7793–7804 (2020).
- [12] Qian, Y., Expert, P., Rieu, T., Panzarasa, P. & Barahona, M. Quantifying the alignment of graph and features in deep learning. *IEEE Transactions on Neural Networks and Learning Systems* (2021).
- [13] B. Wang, L. Cheng, J. Sheng, Z. Hou and Y. Chang, Graph convolutional networks fusing motif-structure information. *Scientific Reports*, 2022, vol. 12, no 1, p. 1-12. <https://doi.org/10.1038/s41598-022-13277-z>