Exploring the latent geometry for representation learning

Abstract

Deep Neural Networks (DNNs) are the state of the art in different tasks of computer vision. Although in continue development, neither their hidden structure is not yet fully understood, even for the first and simplest architectures.

This thesis aims to provide some instruments to understand the representation of some architectures and provide new techniques to improve them in different tasks, from classification to inverse problems.

These instruments come from neuroscience and geometry. Indeed neuroscience inspired artificial intelligence since its infancy, adapting the knowledge and the modelling of biological networks to build artificial networks. In particular a very popular field of study today is the so called Explainable Artificial Intelligence (AI), aiming to give an interpretation of artificial networks mechanisms. However sometimes these methods lead to contradictory results.

In this thesis, we propose a new explainability pipeline that resumes the inspiring principles of AI, i.e. neuroscience methods that served to understand neurons in the brain. With the same spirit, we are going to consider a DNN as an artificial brain and analyze single units to determine their role and give it a label to identify it. The project is composed of various sections, each offering a unique perspective from neuroscience that ultimately converges towards a shared interpretation.

The whole pipeline aims also to provide a benchmark that uses such networks to get predictions on biological networks. Indeed in the last part of the project we show some preliminary results from biological neurons of the visual cortex of a macaque.

Beyond understanding of the hidden structure of DNNs, this thesis shows how to explore and improve the representation in some vision models by studying the hidden geometrical structures.
This is the case of e-GLASS, that stands for "exploring the Gan LAtent Space Solutions", and is a framework that exploits the image prior learnt in the latent space of Generative Adversarial Networks (GANs) to provide sets of possible solution to linear inverse problems, such as super-resolution and inpainting. The method is entirely built upon the geometry of the latent space, providing useful directions to solutions perceptually different from each other more quickly than existing approaches.

While this method and in general most of the DNNs exploit the geometry induced by learning features in Euclidean space, in this thesis we study and propose new regularizations to learn features in a non Euclidean geometry, i.e. the hyperbolic space.
Even if most of the networks extract features and build representations in Euclidean space, spaces with more representative geometries may exist, especially when data have particular structures, e.g. images, graphs or molecules. It is the case of the hyperbolic space, a space that was already used to study the physics of the space-time in special relativity.
It turned out that the hyperbolic space is particularly relevant to embed data with hierarchical structures. Indeed it was demonstrated that tree graphs can be embedded with arbitrary low distortion in the hyperbolic space, a property that does not hold for flat spaces who distort the embeddings, losing the true distances in the graph.

In this thesis we propose a new methodology to represent the hierarchical compositionality of 3D objects, based on a regularization in the hyperbolic space. In fact 3D point clouds exhibit a part-whole hierarchy made by the parts composing the object, and capturing this property could reveal a better representation, leading to improvements in classification and segmentation.

These new methods revealed high adaptability to different architectures, tasks and datasets.
In the future, we'd like to generalize some of the techniques presented in this thesis to other problems and adapt to new state of the art models, e.g. vision transformers and diffusion models.