Towards latent space explorer

Scientific research is a human activity that often implies the production of new or improved measurement tools, leading to the generation of new data that needs to be analyzed, either to corroborate existing theories, or to support the generation of new ones. Next generation astronomical facilities, for instance, such as the Square Kilometre Array will generate an overwhelming volume of data at a rate that simply cannot be matched by our ability to make sense out of them.

Sometimes, moreover, it is impossible to use supervised techniques to support these researches: the studied phenomena are often object of intense study and classification schemes are still not agreed upon, they might be uncertain, or new data was acquired exactly with the goal of defining a classification scheme that was impossible to come up with using previously available data.

An array of techniques has been devised to support human researchers in the analysis of even large pools of data, even in absence of predefined knowledge structures or interpretation rules supporting this form of analysis. In particular, ***data visualization*** and ***clustering*** techniques are often employed when trying to make sense of data. But what if we are not talking about large tables but whole datasets of astronomical images? How to arrange a potentially large set of photos in an effective visualization? How to cluster images?

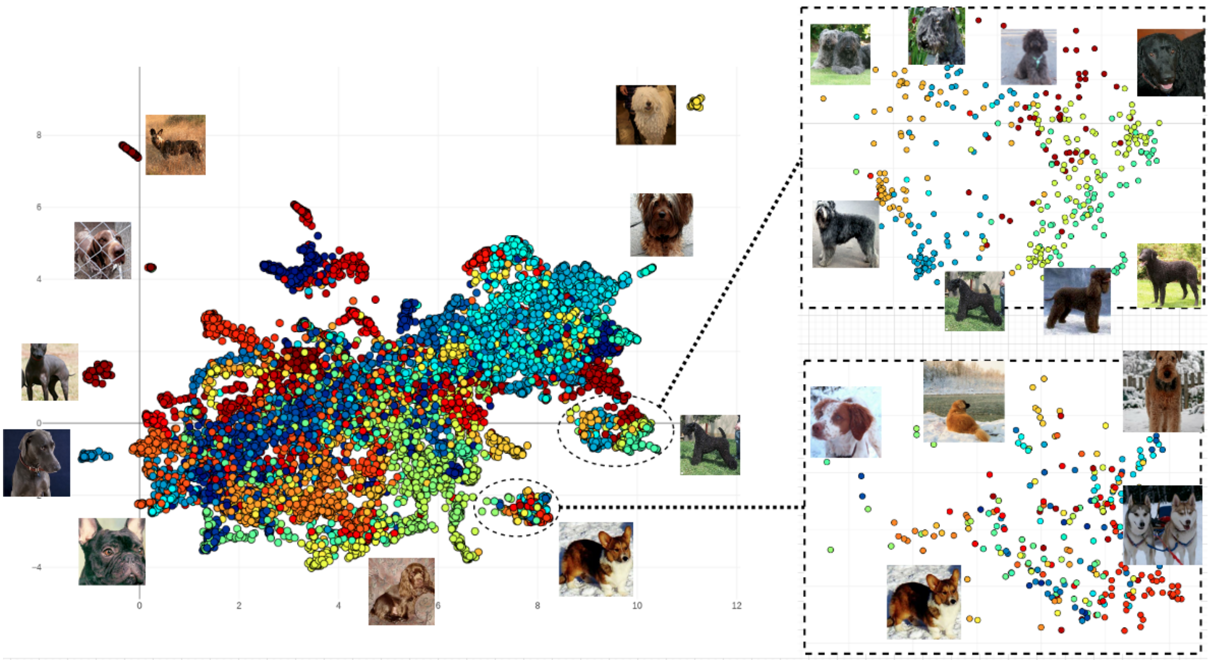

Fig.1: Stanford dogs[1] organized in a latent space

Let us get back to square one, and let us stress that automated/semi-automated tools are needed to support this kind of research. The second consideration is that machine learning (ML) can be an instrument in an overall workflow including domain experts and computer scientists: ML algorithms are sometimes (and often correctly) seen as potentially problematic, due to difficulties in understanding the reasons for some achieved results, something that calls for explainable forms of Artificial Intelligence.

[1] *Khosla, A., Jayadevaprakash, N., Yao, B., Li, F.F.: Novel dataset for fine-grained image cate-gorization: Stanford dogs. In: Proc. CVPR Workshop on Fine-Grained Visual Categorization(FGVC). vol. 2. Citeseer (2011)*
http://vision.stanford.edu/aditya86/ImageNetDogs/

The third one is that, although we want to keep the human domain expert in the loop, adopted ML approaches need to be essentially unsupervised, employing just the input data as a teacher.

Our proposal for discovering patterns in image datasets comprises two steps: (i) achieving a compact representation of the elements of the dataset by means of **representation learning techniques**, shifting the following analysis from cumbersome representations (i.e. bitmaps from FITS files) to compact vectors in a latent space, and (ii) visualizing the achieved space and support clustering of enclosed points, associated to instances of the starting dataset, to suggest patterns to the domain experts that will evaluate their potential meaning in the studied domain.

A representation of an image could be seen as a vector of features that we'll call latent vector. Our goal is to have a latent space where we could explore learned dataset features (without supervision), an example could be seen in the following figure. Fig.1 shows a latent space with different dog breeds organized in an unsupervised way. Exploring that latent space we could see that the neural network learned interesting visual features like context (snow, beach, cage), dog pose (frontal, side), fur length and more.

In this brief post we focus on describing different kinds of unsupervised representation learning techniques: generative approach and contrastive learning

Generative approach

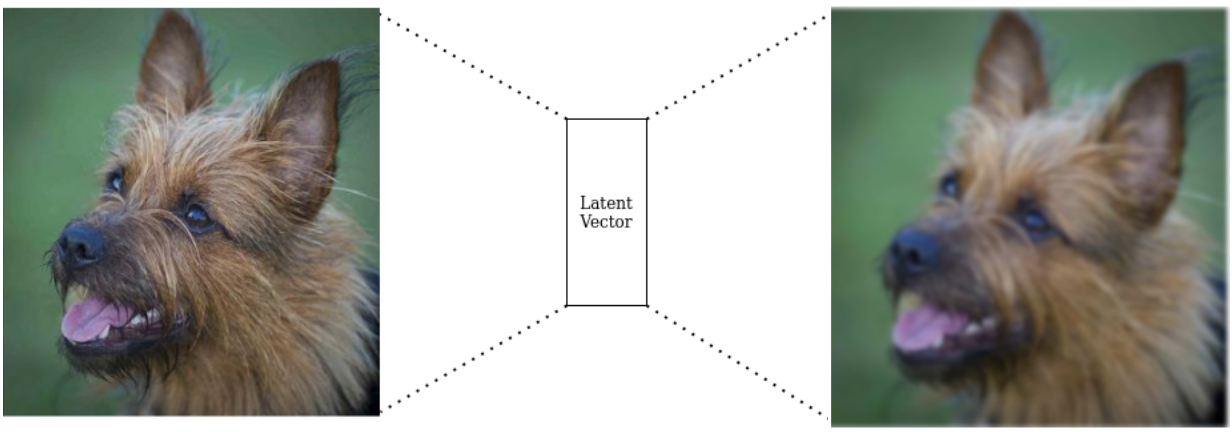

Fig.2: Autoencoder architecture

Neural networks are fascinating because with simple ideas or tricks (generally having serious but overlooked mathematical motivations) we can achieve different goals. For instance, we could build a clepsydra architecture (fig.2) where the input is an image, in the middle we have a custom sized vector of numbers, and the output an image again. If we set the goal of the training process to output an image as similar as possible to the one presented as input, then the network is forced to learn how to reproduce the image but using a much more compact representation, the above-mentioned vector that we will call *latent vector*. The name of this kind of architecture is *autoencoder*, and it was studied to be used for different tasks like image compression and noise removal. **Autoencoders** achieve tasks of reconstructing an image, but they don't generate totally new ones as instead done by variational autoencoder (VAE) and generative adversarial networks (GANs). In particular for GANs the task is harder: to generate never seen images that are plausible for the context.

To achieve that, an opponent (a classification network), is trained to classify real/fake images. Training two different models is a heavy task and convergence is not guaranteed. Using reconstruction or generative oriented architectures for representation learning is actually going a bit beyond the original goals, so it is not guaranteed that good results are at hand.

Contrastive learning

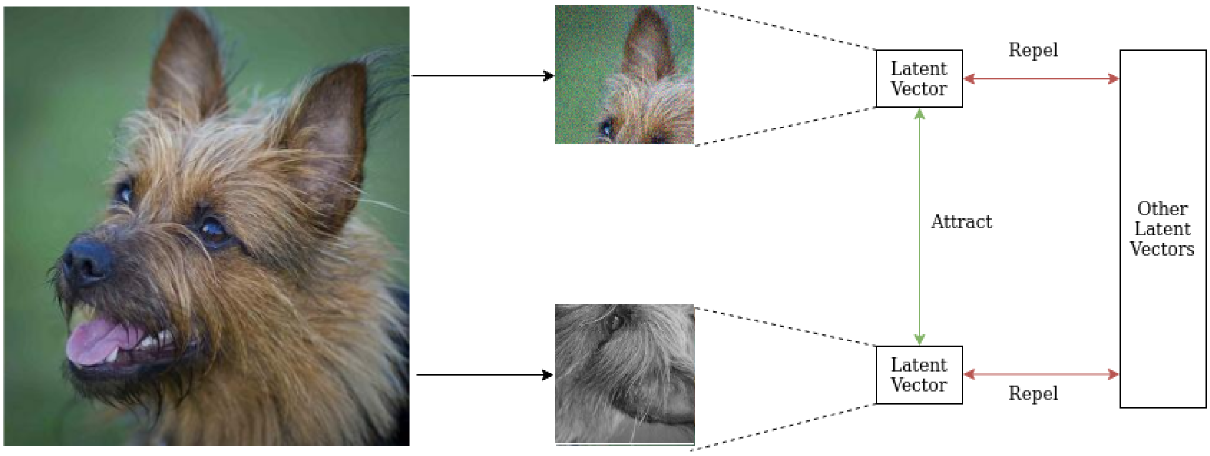

Fig.3: Simclr contrastive learning techniques

Self supervision is the keyword that describes the methods that utilize contrastive learning. Those methods have two steps, training on a so-called pretext task and, in a second step, training on a downstream task. The goal of a pretext task is to learn directly from the dataset, without supervision, in order to generalize better the concept adopting a bottom-up feature learning approach. In the later downstream task supervision is performed for a small subset of the data to fit the specific task. Those methods are competitive with supervised techniques (and sometimes better than them) on some benchmark datasets in different tasks like classification, segmentation for images and videos.

The first part of the contrastive learning process, that is basically unsupervised, could be used as a representation learning tool. A particular technique called **simclr**[2] is described in Fig. 3; the idea behind it is very simple: different augmented crops of the same image should produce similar representations, in other words similar latent vectors; on the contrary, representations of different images should disagree, be distant in the latent space.

The Fig.1 is a latent space, a spatial structure in which the set of latent vectors associated to images of the dataset can be arranged: in particular, it shows the latent space learned with contrastive learning on Imagenet, focusing only on the dogs dataset. We can see how the space is not finely divided by classes (breeds), but rather organized by features. In self supervised learning the second step is to fine-tune the model in order to divide (classify) by certain features, in that case in features that correspond to breeds.

[2] *Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.*
https://amitness.com/2020/03/illustrated-simclr/ a good visual explaination

Latent space explorer

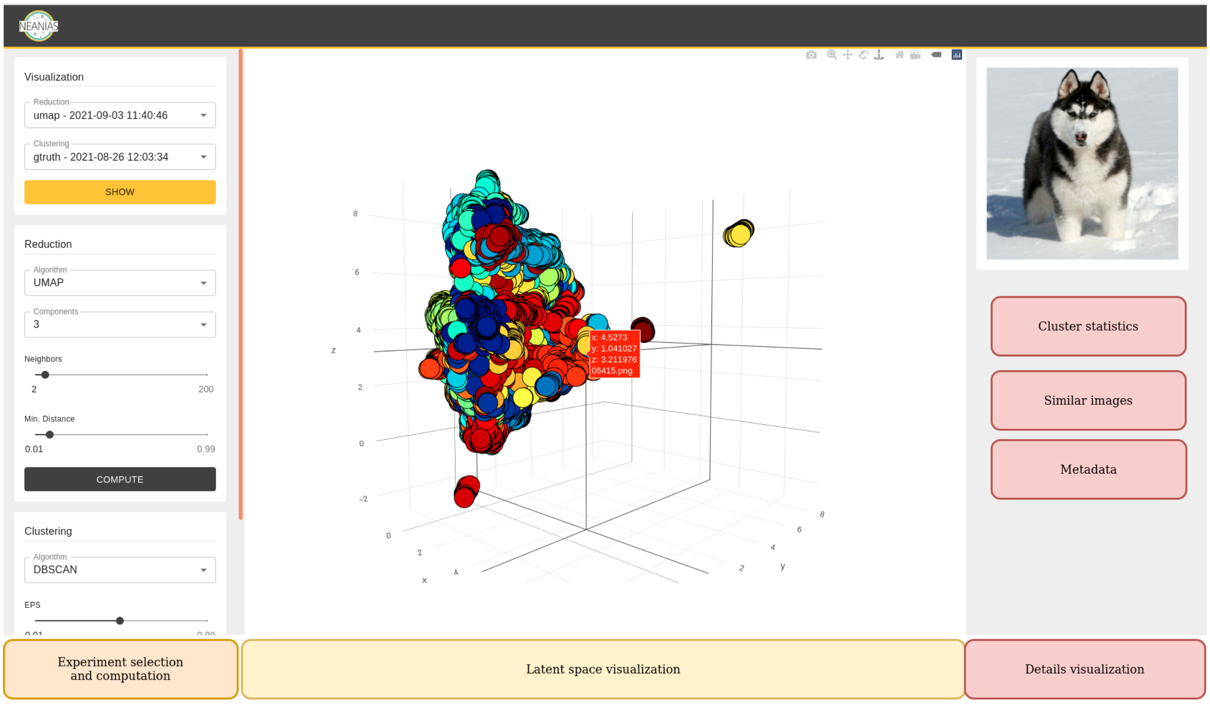

Fig.4: Snapshot of the latent space explorer

In neanias we are developing a service that could help the analysis of a latent space generated by a specific dataset, for instance of astronomical images. As it could be seen in Fig.4, the service is divided in three columns: the leftmost one is dedicated to experiment selection and computation. Since the latent vectors are multidimensional, a dimensionality reduction algorithm needs to be computed by selecting from the most famous ones (PCA, TSNE, UMAP). In addition to pure visualization, unsupervised analysis like clustering can be employed to suggest groupings, to support the analysis of the latent space when the domain expert is trying to define novel classification schemes based on available data.

In the center column it's possible to interact with the latent space. It's possible to navigate even in 3 dimensional space, and to perform basic operations like isolate classes, select a subset of points, take a screenshot; more functions will be developed as a consequence of the experimentation and feedbacks from users (by the way, if you're interested in trying this kind of tool, stay tuned on NEANIAS channels and maybe keep in touch).

The rightmost column shows information about an instance. In particular we want to develop functions that show information about the instance's cluster, show similar images (belonging to the cluster, but also the most similar ones from other clusters) and show additional metadata.

Soon the service will be available in the neanias catalogue, so… stay tuned!

Lucas Puerari, Thomas Cecconello, Giuseppe Vizzari
Università degli studi di Milano-Bicocca (UNIMIB)