# RecViz - Project Proposal - Topic D

Théo DI PIAZZA - theo.dipiazza@gmail.com

December 6, 2022

## 1 Introduction & Motivation

In computer vision, it is necessary to have a large amount of labelled data in order to develop supervised learning algorithms. However, the process of labelling this data is costly and time consuming, and can be biased towards the user responsible for the labels. To address this issue, contrast methods are developed using self-supervised learning to generate embeddings for a given image. For this project, we will focus on VICReg, a self-supervised learning method and also on the pre-training of a visual blackbone.

## 2 Plan of works

The first step of the project is to get familiar with self-supervised learning methods with paper [1], before studying VICReg in more detail with paper [2]. The main focus will be on the input/output data, the training process and the proposed model architectures.

In a second step, we will be able to take RecViz in hand thanks to the github code [3]. We will test the model on different open-source images. Then, we can proceed to the pre-training of the backbones on different datasets: CIFAR-10, CIFAR-100 and/or ImageNet-100 depending on the progress of the project. For blackbone, one idea is to use a ResNet as encoder (ResNet34 for example, to be inspired by VICReg). In order to evaluate the different performances of the models, a linear evaluation of the models will be carried out. This will consist of training a linear classifier on top of frozen representations learned by self-supervised methods. This step will be done on a different dataset from the pre-training one, probably by splitting CIFAR into different datasets.

Finally, the objective will be to compare the performance of these blackbone pretrain with the ResNet-50 associated with VICReg, once again via a linear evaluation protocol. Regarding the expected results, one can imagine that the larger the size of the training dataset, the better the model will perform. In addition, it can be expected that the complexity of the architecture of the models used for the backbone will have an influence on the performance, but also on the computation time. If time permits, it would be interesting to test the model representations on more mainstream tasks, such as classification.

The expected difficulties could be: the appropriation of the code, the training/testing time and the choice of architectures for the blackbone.

## 3 References

[1] *A Simple Framework for Contrastive Learning of Visual Representations, Chen et al, ICML 2020*
[2] VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning, Bardes et al, ICLR 2022
[3] VICReg Code : https://github.com/facebookresearch/vicreg