Project proposal

Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles

Team composition

Giovanni Cavallin, 1206693, giovanni.cavallin.1@studenti.unipd.it

Elena Mattiazzo, 1206695, elena.mattiazzo.1@studenti.unipd.it

Investigating problem

We are interested in the problem implementation of unsupervised learning of visual representation. This problem refers to a paper[1] that studies the problem of image representation learning without human annotation. They use the solution of a jigsaw puzzle as a pretext task, that requires no manual labeling, and then repurpose the neural network to solve object classification and detection. Our first aim is to setup the network for the classification problem and to extend it to the detection one later, if feasible.

Dataset

We would like to use three categories from ImageNet[2] since we plan to train the network on a local machine. Since those three categories are not well distributed (dogs' images are much more than fishes) we have decided to organize them in this way:

1. **dogs**: 9165 images (46%)
2. **cats**: 7882 images (39%)
3. **fishes**: 2953 images (15%)

We think that this repartition will give us the possibility to face some sub-optimal real condition in which some categories are less represented than others. In addition, three categories with a total of 20000 images are expected to give good inference results.

Algorithms and implementation

As regards the algorithm we are going to implement, we will completely refer to the original paper. Some changes are to be expected, mainly on the neural network we are going to fine-tune from. We are going to refer also to two existing implementations of this work:

- from the author's implementation in PyTorch[3], which is in PyTorch;
- another implementation of a Github user in Tensorflow and Keras[4].

The implementation will follow those steps:

1. understanding of the dataset and labels. Preparation and data augmentation;
2. set up of the net infrastructure in Tensorflow. We are still thinking about implementing the network with the Tensorflow 2.0 version, even if in an alpha stage;
3. tuning of the parameters.

---

[1] Unsupervised Learning of Visual Representation by Solving Jigsaw Puzzles (available at: https://arxiv.org/abs/1603.09246)

[2] Image Net, image database organized according to the WordNet hierarchy (available at: http://www.image-net.org)

[3] MehdiNoroozi's Github (available at: https://github.com/MehdiNoroozi/JigsawPuzzleSolver)

[4] Amobiny's Github (available at: https://github.com/amobiny/Self-Supervised_Jigsaw_Puzzle)