Paper Review: VoCo: A Simple-yet-Effective Volume Contrastive Learning Framework for 3D Medical Image Analysis

By: Lam Nguyen

1. SUMMARY

Deep learning for 3D medical imaging analysis is hampered by the expensive cost of requiring experts' annotations for 3D Data. There aren't that many experts and you can't just get any person off the street in a low wage area to do something as critical as medical image annotation where the cost of incorrect data could be a human life and the loss of reputation for a potential multi-million dollar contract. This is why SSL has received significant attention as it doesn't require annotations.

Self-Supervised-Learning (SSL) has shown good results in analyzing 3D medical images. There are still some problems with self-supervised learning. The problem that will be addressed in this paper is that there is a lack of high-level rules of thumb for pretraining a model. The human body is created a certain way and the parts of the body are generally arranged in the same pattern in space. This spatial information can be used to pretrain a consistent semantic representation.

Volume Contrast (VoCo) is a framework for leveraging these contextual priors for pretraining. Adding these contextual priors showed improvements on six downstream tasks.

This is how VoCo works:
1. Groups of non-overlapping volumes were cropped.
2. These cropped volumes are then represented as bases in high dimensional space where they are used as different class assignments.
3. Next randomly crop sub-volumes and predict which class they belong to by contrasting their similarity to different bases, which can be seen as predicting contextual positions of different subvolumes.
4. In this way, we formulate a contextual position data for pretraining medical image SSL.

2. STRENGTHS

- VoCo does not require expensive hand annotation of medical image data by experts. Eliminates the cost and time of having to pay for data annotation by hand.
- VoCo pretraining improves performance even on datasets that the model is not trained on.
- Pretraining is relatively inexpensive in terms of computing performance.

3. WEAKNESSES

- Could possibly still contain inaccuracies depending on how the code is implemented meaning that for medicine, human lives will be on the line for such mistakes. More testing has to be done or at least verification by a human should occur.
- Still requires extra time and computational power for dealing with 3D datasets.
- Paper could cover the minimum amount of pretraining required to achieve a sufficient improvement.

4. TECHNICAL EXTENSIONS

- Scale up the pretraining dataset to evaluate the maximum performance of VoCo
- Scale down the dataset to find the minimum amount of data to achieve a good enough improvement.
- Evaluate the performance on different datasets.
- The VoCo method relies on the fact that the human body and its organs generally have a very similar configuration. Organs are in the same general area, there are the same amount of limbs, etc… Could this method be used to pretrain other non-related data

besides the human body such as say a car, different types of animals or plants, or any object or being that generally has a similar structure?

5. OVERALL REVIEW

The VoCo method is a simple and effective method that pretrains a model to analyze 3D medical images. It pretrains a model by relying on the fact that the human body and its organs are generally in the same place across different people.

The strength of this method is that it reduces and/or eliminates the need for human annotation of the dataset, reducing the cost of training by a large margin.

However the weakness is that this method still requires extra training on a dataset and still will probably need human intervention or checking to ensure that the results of the SSL are fine.