Presented By : Kaivalya Anand Pandey

3D Self-Supervised Methods for Medical Imaging

Authors: Aiham Taleb, Winfried Loetzsch, Noel Danz

Novosibirsk State University

October 20, 2020

Introduction

- 3D imaging has numerous applications, such as in Robotic navigation, in CAD imaging, in Geology, and in Medical Imaging.
- Medical imaging plays a vital role in patient healthcare, as it aids in disease prevention, early detection, diagnosis, and treatment.
- Some difficulties in Medical 3D Imaging : 1. Generating Expert Annotation
  2. Annotations are expensive and time consuming
  3. Small Sample Sizes

- A widely used technique is transfer learning, which aims to reuse the features of already trained neural networks on different, but related, target tasks.
- A common example is adapting the features from networks trained on ImageNet, which can be reused for other visual tasks, e.g. semantic segmentation.
- To some extent, transfer learning has made it easier to solve tasks with limited number of samples.
- Despite attempts to leverage ImageNet features in the medical context, the difference in the distributions of natural and medical images is significant.
- Consequently, it is necessary to find better solutions for the mentioned challenges.

Self Supervised Method

- Viable Alternative - Self Supervised Methods.
- In these approaches, the supervisory signals are derived from the data
- Self Supervised Methods can offer cheaper solutions for the challenges faced by conventional supervised methods.
- Self-Supervision enables the models to derive notions with no additional annotation cost.

Proposed Self Supervised Task

- Five Self Supervised Methods : 1. 3D Contrastive Predictive Coding.
  2. 3D Rotation Prediction.
  3. 3D Jigsaw puzzles
  4. Relative 3D Patch Location.
  5. 3D Exemplar Networks
- Algorithms are inspired by their successful counterparts in 2D and except for 3D Jigsaw Puzzle, no method is extended to 3D.
- Several computational and methodological challenges arise when designing self-supervised tasks in 3D, due to the increased data dimensionality, which is addressed in this paper to ensure their efficiency.

3D Contrastive Predictive Coding (3D-CPC)

- Contrastive learning idea, first proposed by Michael Gutmann and others in A new estimation principle for unnormalized statistical models, this universal unsupervised technique predicts the latent space for future (next or adjacent) samples.
- CPC found success in multiple application fields, e.g. its 1D version in audio signals, and its 2D versions in images, and was able to bridge the gap between unsupervised and fully-supervised methods.
- Our proposed CPC version generalizes this technique to 3D inputs, and defines a proxy task by cropping equally- sized and overlapping 3D patches from each input scan. Then, the encoder model maps each input patch to its latent representation.

3D Contrastive Predictive Coding (3D-CPC)

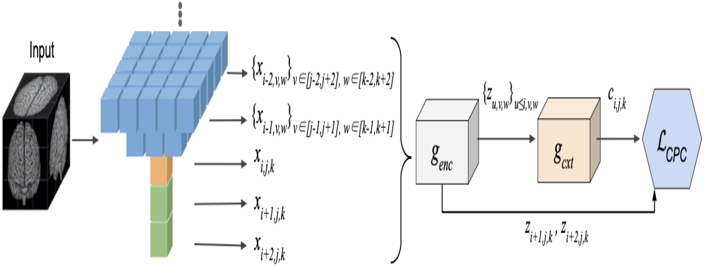


Figure 1: 3D-CPC

- Context vectors captures the high level content of the context that corresponds to input patch, it allows for predicting the latent representations of next (adjacent) patches.
- This prediction task is cast as an N-way classification problem by utilizing the InfoNCE loss, which takes its name from its ability to maximize the mutual information between context vectors and latent representation.
- Loss Function :

$$L_{CPC} = -\sum_{i,j,k,l} \log p(z_{i+l,_{j,k}} | \widetilde{z}_{i+l,_{j,k}} z_n) \tag{1}$$

Relative 3D patch location (3D-RPL)

- In this task, the spatial context in images is leveraged as a rich source of supervision, in order to learn semantic representations of the data. First proposed by Doersch for 2D images in his paper Unsupervised visual representation learning by context prediction in 2015.
- In our 3D version of this task, we leverage the full 3D spatial context in the design of our task. From each input 3D image, a 3D grid of N non-overlapping patches is sampled at random locations.
- Then, the patch in the center of the grid is used as a reference, and a query patch is selected from the surrounding $N - 1$ patches.

Relative 3D patch location (3D-RPL

- Formally, the cross-entropy loss in this task is written as:

$$L_{RPL} = -\sum_{k=1}^{K} \log P(y_q | \tilde{y_q}, y_n) \quad (2)$$

  where K is number of queries extracted from all samples
- In order to prevent the model from solving this task quickly by finding shortcut solutions, e.g. edge continuity, we follow in leaving random gaps (jitter) between neighbor 3D patches.

Relative 3D patch location (3D-RPL)

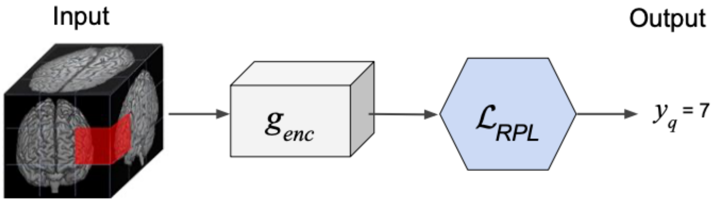


Figure 2: 3D-RPL

3D Jigsaw Puzzle Solutions (3D-Jig)

- We minimize the cross-entropy loss

$$L_{Jig}(y_p^k, \tilde{y}_p^k) \tag{3}$$

where k $\in$ 1, .., K is an arbitrary 3D puzzle from the list of extracted K puzzle.

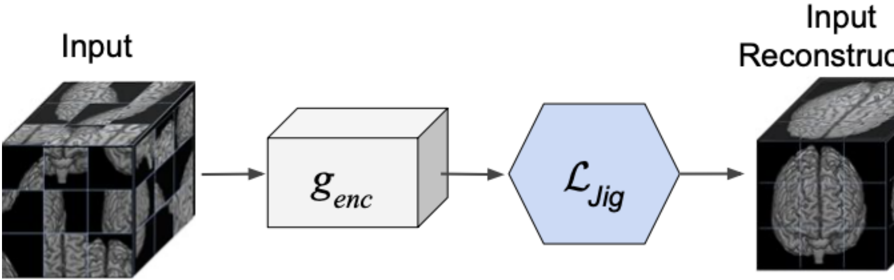


Figure 3: 3D Jig

3D Rotation Prediction(3D-Rot)

- Originally proposed by Gidaris in his paper Unsupervised Reperesenting Learning by predicting image rotation, the rotation prediction task encourages the model to learn visual representations by simply predicting the angle by which the input image is rotated.
- The intuition behind this task is to predict angle of rotation.
- In our 3D Rotation Prediction task, 3D input images are rotated randomly by a random degree r $\in$ 1, .., R out of R considered degree.
- Formally we minimize the cross entropy loss

$$L_{Rot}(r_k, \tilde{r_k}) \tag{4}$$

3D Rotation Prediction(3D-Rot)

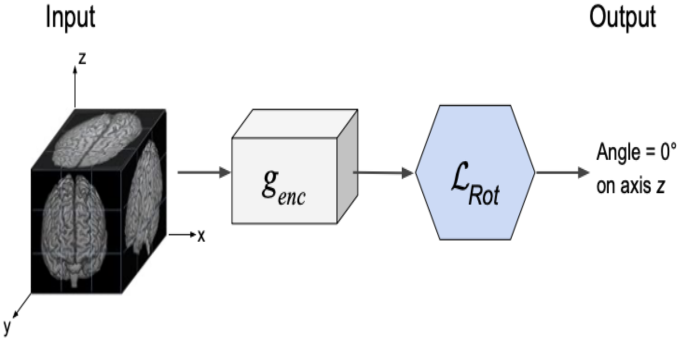


Figure 4: 3D Rot

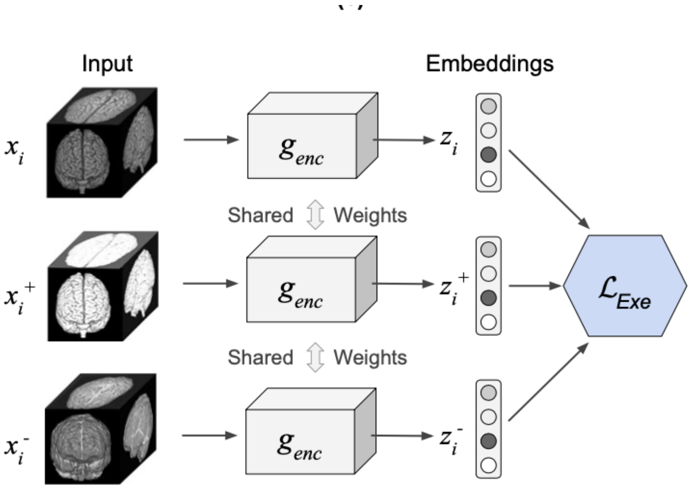
3D Exemplar Networks

- The task of Exemplar networks, proposed by Dosovitskiy in his paper Discriminative Unsupervised Feature Learning with CNN, is one of the earliest methods in the self-supervised family. To derive supervision labels, it relies on image augmentation techniques, i.e. transformations.
- In place of cross entropy loss we will use Triplet Loss here as CEL becomes expensive if dataset size grows larger as number of classes grows accordingly. The triplet loss is defined as

$$L_{Exe} = 1/N_T \sum i = 1 N_T max 0, D(z_i, z_i^+) - D(z_i, z_i^-) + \alpha \tag{5}$$

where D is pairwise distance function.

3D Exemplar Networks



(e)

Figure 5: 3D-Exemplar

Results

Table 1: BraTS segmentation results

Model	ET	WT	TC
3D-From scratch	76.38	87.82	83.11
2D-CPC	76.60	86.27	82.41
2D-RPL	77.53	87.91	82.56
2D-Jigsaw	76.12	86.28	83.26
2D-Rotation	76.60	88.78	82.41
2D-Exemplar	75.22	84.82	81.87
Popli *et al.* [58]	74.39	89.41	82.48
Baid *et al.* [59]	74.80	87.80	82.66
Chandra *et al.* [60]	74.06	87.19	79.89
Isensee *et al.* [57]	80.36	**90.80**	84.32
3D-CPC	80.83	89.88	85.11
3D-RPL	**81.28**	90.71	**86.12**
3D-Jigsaw	79.66	89.20	82.52
3D-Rotation	80.21	89.63	84.75
3D-Exemplar	79.46	**90.80**	83.87

Figure 6: BraTS Segmentation Results

Results

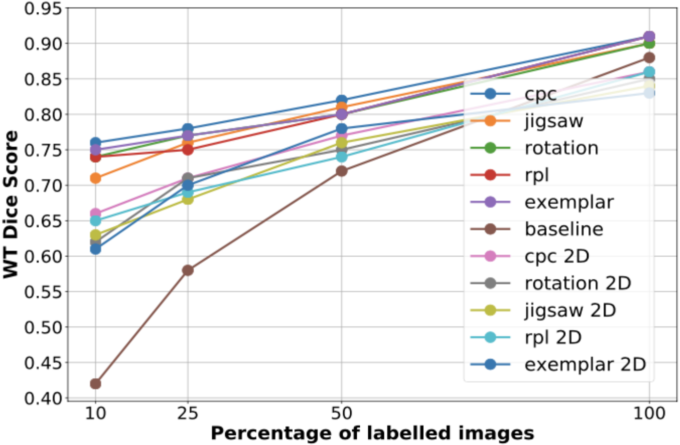


Figure 7: Data Efficient Segmentation Results in BraTS

Results

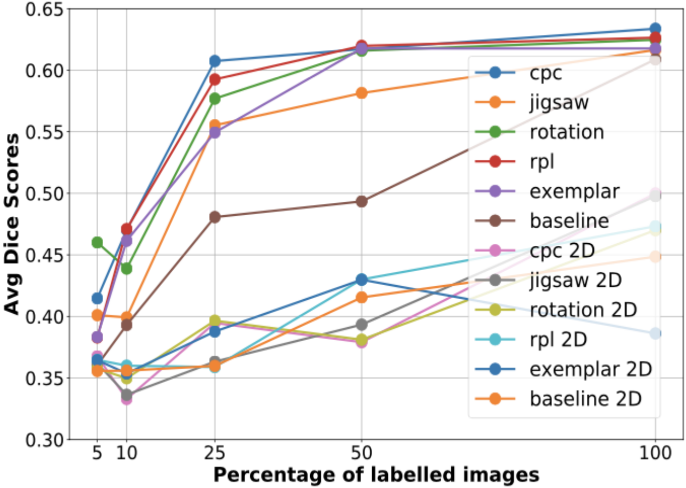


Figure 8: Pancreas Tumour Segmentation from 3D CT

Results

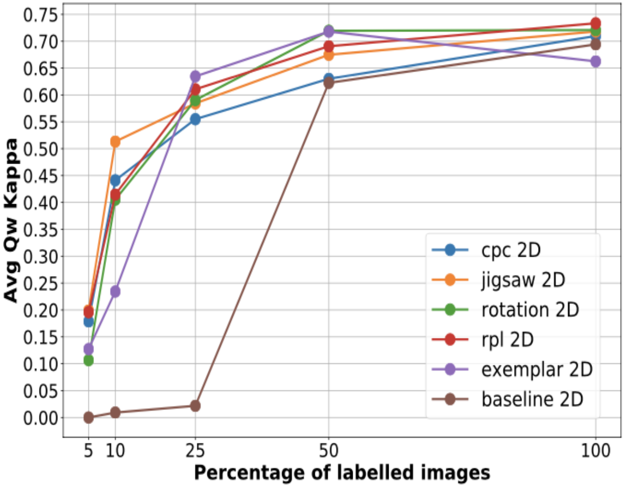


Figure 9: Diabetic Retinopath Detection

THANK YOU

Questions?