Supplementary Material for Context Matters: Graph-based Self-supervised Representation Learning for Medical Images

Li Sun[*], Ke Yu[*], and Kayhan Batmanghelich
University of Pittsburgh, USA
{lis118, key44, kayhan}@pitt.edu

Network Archtecture

In the tables below, we show the detailed architectures of conditional encoder $E(\cdot, \cdot)$, including $C(\cdot)$ and $f_l(\cdot)$, and graph convolutional network $G(\cdot, \cdot)$.

Table 1: Architecture of the C Network

Layer	Filter size, stride	Output size(C, D, H, W)
Input	-	1×32×32×32
Conv3D	3×3×3, 1	8×32×32×32
BatchNorm+ELU	-	8×32×32×32
Conv3D	3×3×3, 2	8×16×16×16
BatchNorm+ELU	-	8×16×16×16
Conv3D	3×3×3, 1	16×16×16×16
BatchNorm+ELU	-	16×16×16×16
Conv3D	3×3×3, 1	16×16×16×16
BatchNorm+ELU	-	16×16×16×16
Conv3D	3×3×3, 2	16×8×8×8
BatchNorm+ELU	-	16×8×8×8
Conv3D	3×3×3, 1	32×8×8×8
BatchNorm+ELU	-	32×8×8×8
Conv3D	3×3×3, 1	32×8×8×8
BatchNorm+ELU	-	32×8×8×8
Conv3D	3×3×3, 2	32×4×4×4
BatchNorm+ELU	-	32×4×4×4
Conv3D	3×3×3, 1	64×4×4×4
BatchNorm+ELU	-	64×4×4×4
Conv3D	3×3×3, 1	64×4×4×4
BatchNorm+ELU	-	64×4×4×4
Conv3D	3×3×3, 2	64×2×2×2
BatchNorm+ELU	-	64×2×2×2
Conv3D	3×3×3, 1	128×2×2×2
BatchNorm+ELU	-	128×2×2×2
Conv3D	3×3×3, 2	128×1×1×1
BatchNorm+ELU	-	128×1×1×1
Reshape	-	1×128

Table 2: Architecture of the f_l Network

Layer	Filter size, stride	Output size(C, F)
Input	-	1×128,1×3
Concatenation	-	1×131
Dense	-	1×131
ReLU	-	1×131
Dense	-	1×131
ReLU	-	1×131
Dense	-	1×128

Table 3: Architecture of the G Network

Layer	Filter size, stride	Output size(C, N, F)
Input	-	1×581×128
GCNLayer	-	1×581×128
BatchNorm+ELU	-	1×581×128
AveragePooling	-	1×1×128
Dense	-	1×1×128
ReLU	-	1×1×128
Dense	-	1×1×128
ReLU	-	1×1×128
Dense	-	1×1×128
Reshape	-	1×128

Implementation Details (cont)

The patch size is set as $32 \times 32 \times 32$. Cosine schedule (**?**) is used to update the learning rate. For MoCo (**?**), we implement a 3D encoder to handle the 3D data and train the model on COPDGene and MosMed dataset. For ModelsGenesis (**?**), we train the model on COPDGene and MosMed dataset with the original setting. For MedicalNet (**?**), since it's training requires segmentation mask, we use pretrained weights provided by the authors.

[*] Equal contribution

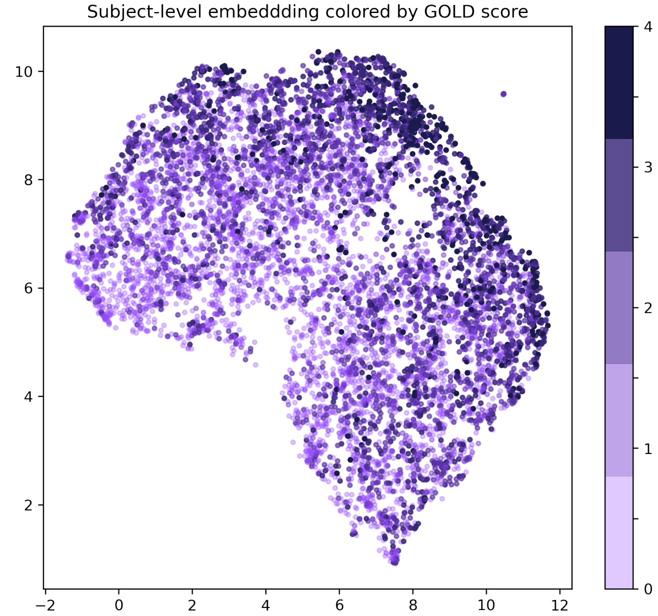

Figure 1: Embedding of subjects in 2D using UMAP. Each dot represents one subject colored by the GOLD score. We can find a trend, from lower-left to upper-right, along which we can see increasing GOLD score.

Model Visualization

To visualize the learned embedding and understand the model's behavior, we use two methods to visualize the model. The first one is embedding visualization, we use UMAP (**?**) to visualize the patient-level features extracted on the COPDGene dataset in two dimension. In Fig 1, we found that subjects with GOLD score of (0,1) and (3,4) are separable under two dimension. But subjects with GOLD score 2 are scattered. It requires further investigation to understanding of embedding pattern of subjects subjects with GOLD score 2. In Fig 1, we can find a trend, from lower-left to upper-right, along which we can see increasing GOLD score.

We use the model explanation method described before to visualize discriminative image regions used by our model for prediction in downstream task. In Fig. 2, we apply the explanation method using the target logit of GOLD score = 4 on a GOLD 4 subject in COPDGene dataset. The dark area on the right lung, where lung tissue is severely damaged, received highest activation value. Figure 3 (left) shows the axial view of the CT image of a COVID-19 positive patient, and Figure 3 (right) shows the corresponding activation map. The anatomical regions received high activation scores overlap with the peripheral ground glass opacities on the CT image, which is a known indicator of COVID-19. This result suggests that our model can highlight the regions that are clinically relevant to the prediction.

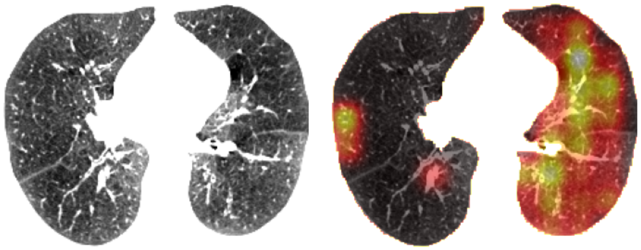

Figure 2: An axial view of the activation map on a $GOLD$ 4 subject in COPDGene dataset. Brighter color indicates higher relevance to the disease severity. The figure illustrates that high activation region overlaps with dark area on the right lung, where lung tissue is damaged.

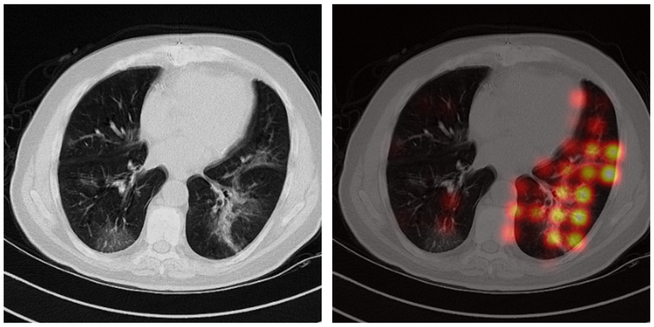

Figure 3: An axial view of the activation heatmap on a COVID-19 positive subject in the COVID-19 CT dataset. Brighter color indicates higher relevance to the disease severity. The figure illustrates that high activation region overlaps with the ground glass opacities.