LSOR: Longitudinally-Consistent Self-Organized Representation Learning

Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Wei Peng, Greg Zaharchuk, and Kilian M. Pohl(✉)

Stanford University, Stanford, CA 94305, USA
kilian.pohl@stanford.edu

Abstract. Interpretability is a key issue when applying deep learning models to longitudinal brain MRIs. One way to address this issue is by visualizing the high-dimensional latent spaces generated by deep learning via self-organizing maps (SOM). SOM separates the latent space into clusters and then maps the cluster centers to a discrete (typically 2D) grid preserving the high-dimensional relationship between clusters. However, learning SOM in a high-dimensional latent space tends to be unstable, especially in a self-supervision setting. Furthermore, the learned SOM grid does not necessarily capture clinically interesting information, such as brain age. To resolve these issues, we propose the first self-supervised SOM approach that derives a high-dimensional, interpretable representation stratified by brain age solely based on longitudinal brain MRIs (i.e., without demographic or cognitive information). Called **L**ongitudinally-consistent **S**elf-**O**rganized **R**epresentation learning (LSOR), the method is stable during training as it relies on soft clustering (vs. the hard cluster assignments used by existing SOM). Furthermore, our approach generates a latent space stratified according to brain age by aligning trajectories inferred from longitudinal MRIs to the reference vector associated with the corresponding SOM cluster. When applied to longitudinal MRIs of the Alzheimer's Disease Neuroimaging Initiative (ADNI, $N = 632$), LSOR generates an interpretable latent space and achieves comparable or higher accuracy than the state-of-the-art representations with respect to the downstream tasks of classification (static vs. progressive mild cognitive impairment) and regression (determining ADAS-Cog score of all subjects). The code is available at https://github.com/ouyangjiahong/longitudinal-som-single-modality.

1 Introduction

The interpretability of deep learning models is especially a concern for applications related to human health, such as analyzing longitudinal brain MRIs. To avoid interpretation during post-hoc analysis [6,14], some methods strive for

G. Zaharchuk—Co-founder, equity Subtle Medical.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_27.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 279–289, 2023.
https://doi.org/10.1007/978-3-031-43907-0_27

an interpretable latent representation [9]. One example is self-organizing maps (SOM) [5], which cluster the latent space so that the SOM representations (i.e., the 'representatives of the clusters) can be arranged in a discrete (typically 2D) grid while preserving high-dimensional relationships between clusters. Embedded in unsupervised deep learning models, SOMs have been used to generate interpretable representations of low-resolution natural images [3,8].

Intriguing as it sounds, we found their application to (longitudinal) 3D brain MRIs unstable during training and resulted in uninformative SOMs. These models get stuck in local minima so that only a few SOM representations are updated during backpropagation. The issue has been less severe in prior applications [3,8] as their corresponding latent space is of much lower dimension than the task at hand, which requires a high dimension latent space so that it can accurately encode the fine-grained anatomical details in brain MRIs [12,17]. To ensure all SOM representations can be updated during backpropagation, we propose a soft weighing scheme that not only updates the closest SOM representation for a given MRI but also updates all other SOM representations based on their distance to the closest SOM representation [3,8]. Moreover, our model relies on a stop-gradient operator [16], which sets the gradient of the latent representation to zero so that it only focuses on updating the SOM representations. It is especially crucial at the beginning of the training when the (randomly initialized) SOM representations are not good representatives of their clusters. Finally, the latent representations of the MRIs are updated via a commitment loss, which encourages the latent representation of an MRI sample to be close to its nearest SOM representation. In practice, these three components ensure stability during the self-supervised training of the SOM on high-dimensional latent spaces.

To generate SOMs informative to neuroscientists, we extend SOMs to the longitudinal setting such that the latent space and corresponding SOM grid encode brain aging. Inspired by [12], we encode pairs of MRIs from the same longitudinal sequence (i.e., same subject) as a trajectory and encourage the latent space to be a smooth trajectory (vector) field. We enforce smoothness by computing for each SOM cluster a reference trajectory, which represents the average aging of that cluster with respect to the training set. The reference trajectories are updated by the exponential moving average (EMA) such that, in each iteration, it aggregates the average trajectory of a cluster with respect to the corresponding training batch (i.e., batch-wise average trajectory). In doing so, the model ensures longitudinal consistency as the (subject-specific) trajectories of a cluster are maximally aligned with the reference trajectory of that cluster.

Named **L**ongitudinally-consistent **S**elf-**O**rganized **R**epresentation learning (LSOR), we evaluate our method on a longitudinal T1-weighted MRI dataset of 632 subjects from ADNI to encode the brain aging of Normal Controls (NC) and patients diagnosed with static Mild Cognitive Impairment (sMCI), progressive Mild Cognitive Impairment (pMCI), and Alzheimer's Disease (AD). LSOR clusters the latent representations of all MRIs into 32 SOM representations. The resulting 4-by-8 SOM grid is organized by both chronological age and cognitive measures that are indicators of brain age. Note, such an organization solely relies

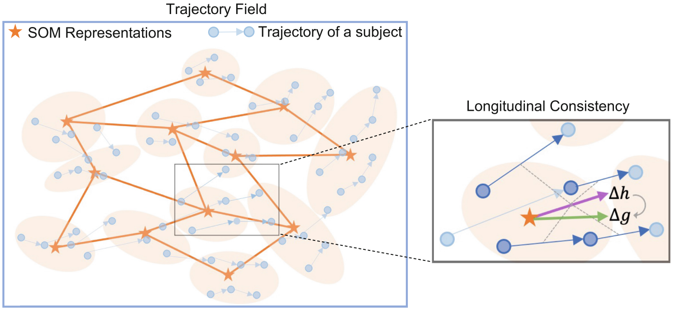

Fig. 1. Overview of the latent space derived from LSOR. All trajectories (Δz) form a trajectory field (blue box) modeling brain aging. SOM representations in $\mathcal{G}$ (orange star) are organized as a 2D grid (orange grid). As shown in the black box, reference trajectories $\Delta\mathcal{G}$ (collection of all Δg, green arrow) are iteratively updated by EMA using the aggregated trajectory Δh (purple arrow) across all trajectories of the corresponding SOM cluster within a training batch. (Color figure online)

on longitudinal MRIs, i.e., without using any tabular data such as age, cognitive measure, or diagnosis. To visualize aging effects on the grid, we compute (post-hoc) a 2D similarity grid for each MRI that stores the similarity scores between the latent representation of that MRI and all SOM representations. As the SOM grid is an encoding of brain aging, the similarity grid indicates the likelihood of placing the MRI within the “spectrum” of aging. Given all MRIs of a longitudinal scan, the change across the corresponding similarity grids over time represents the brain aging process of that individual. Furthermore, we infer brain aging on a group-level by first computing the average similarity grid for an age group and then visualizing the difference of those average similarity grids across age groups. With respect to the downstream tasks of classification (sMCI vs. pMCI) and regression (i.e., estimating the Alzheimer’s Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) on all subjects), our latent representations of the MRIs is associated with comparable or higher accuracy scores than representations learned by other state-of-the-art self-supervised methods.

2 Method

As shown in Fig. 1, the longitudinal 3D MRIs of a subject are encoded as a series of trajectories (blue vectors) in the latent space. Following [12,17], we consider a pair of longitudinal MRIs (that corresponds to a blue vector) as a training sample. Specifically, let $\mathcal{S}$ denote the set of image pairs of the training cohort, where the MRIs x^u and x^v of a longitudinal pair (x^u, x^v) are from the same subject and x^v was acquired Δt years after x^u. For simplicity, $\times$ refers to u or v when a function is separately applied to both time points. The MRIs are then mapped to the latent space by an encoder F, i.e., $z^\times := F(x^\times)$. On the latent space, the trajectory of the pair is denoted as $\Delta z := (z^v - z^u)/\Delta t$, which represents morphological changes. Finally, decoder H reconstructs the input MRI $x^{\times}$ from the latent representation $z^{\times}$, i.e., $\tilde{x}^{\times} := H(z^{\times})$. Next, we describe LSOR, which generates interpretable SOM representations, and the post-hoc analysis for deriving similarity grids.

2.1 LSOR

Following [3,8], SOM representations are organized in a N_r by N_c grid (denoted as SOM grid) $\mathcal{G} = \{g_{i,j}\}_{i=1,j=1}^{N_r,N_c}$, where $g_{i,j}$ denotes the SOM representation on the i-th row and j-th column. This easy-to-visualize grid preserves the high-dimensional relationships between the clusters as shown in by the orange lines in Fig. 1. Given the latent representation $z^{\times}$, its closest SOM representation is denoted as $g_{\epsilon^{\times}}$, where $\epsilon^{\times} := argmin_{(i,j)} \parallel z^{\times} - g_{i,j} \parallel_2$ is its 2D grid index in $\mathcal{G}$ and $\parallel \cdot \parallel_2$ is the Euclidean norm. This SOM representation is also used to reconstruct the input MRI by the decoder, i.e., $\tilde{x}_g^{\times} = H(g_{\epsilon^{\times}})$. To do so, the reconstruction loss encourages both the latent representation $z^{\times}$ and its closet SOM representation $g_{\epsilon^{\times}}$ to be descriptive of the input MRI $x^{\times}$, i.e.,

$$L_{recon} := \mathbb{E}_{(x^u,x^v)\sim\mathcal{S}} \left(\sum_{\times\in\{x,v\}} \parallel x^{\times} - \tilde{x}^{\times} \parallel_2^2 + \parallel x^{\times} - \tilde{x}_g^{\times} \parallel_2^2 \right), \qquad (1)$$

where $\mathbb{E}$ defines the expected value. The remainder describes the three novel components of our SOM representation.

Explicitly Regularizing Closeness. Though L_{recon} implicitly encourages close proximity between $z^{\times}$ and $g_{\epsilon^{\times}}$, it does not inherently optimize $g_{\epsilon^{\times}}$ as $z^{\times}$ is not differentiable with respect to $g_{\epsilon^{\times}}$. Therefore, we introduce an additional 'commitment' loss explicitly promoting closeness between them:

$$L_{commit} := \mathbb{E}_{(x^u,x^v)\sim\mathcal{S}} \left(\parallel z^u - g_{\epsilon^u} \parallel_2^2 + \parallel z^v - g_{\epsilon^v} \parallel_2^2 \right).$$

Soft Weighting Scheme. In addition to update $z^{\times}$'s closest SOM representation $g_{\epsilon^{\times}}$, we also update all SOM representations $g_{i,j}$ by introducing a soft weighting scheme as proposed in [10]. Specifically, we design a weight $w_{i,j}^{\times}$ to regularize how much $g_{i,j}$ should be updated with respect to $z^{\times}$ based on its proximity to the grid location $\epsilon^{\times}$ of $g_{\epsilon^{\times}}$, i.e.,

$$w_{i,j}^{\times} := \delta\left(e^{-\frac{\parallel \epsilon^{\times} - (i,j) \parallel_1^2}{2\tau}} \right), \qquad (2)$$

where $\delta(w) := \frac{w}{\sum_{i,j} w_{i,j}}$ ensures that the scale of weights is constant during training and $\tau > 0$ is a scaling hyperparameter. Now, we design the following loss L_{som} so that SOM representations close to $\epsilon^{\times}$ on the grid are also close to $z^{\times}$ in the latent space (measured by the Euclidean distance $\parallel z^{\times} - g_{i,j} \parallel_2$):

$$L_{som} := \mathbb{E}_{(x^u,x^v)\sim\mathcal{S}} \left(\sum_{g_{i,j}\sim\mathcal{G}} \left(w_{i,j}^u \cdot \parallel z^u - g_{i,j} \parallel_2^2 + w_{i,j}^v \cdot \parallel z^v - g_{i,j} \parallel_2^2 \right) \right). \qquad (3)$$

To improve robustness, we make two more changes to Eq. 3. First, we account for SOM representations transitioning from random initialization to becoming meaningful cluster centers that preserve the high-dimensional relationships within the 2D SOM grid. We do so by decreasing τ in Eq. 2 with each iteration so that the weights gradually concentrate on SOM representations closer to $g_{\epsilon^\times}$ as training proceeds: $\tau(t) := N_r \cdot N_c \cdot \tau_{max} \left(\frac{\tau_{min}}{\tau_{max}}\right)^{t/T}$ with τ_{min} being the minimum and τ_{max} the maximum standard deviation in the Gaussian kernel, and t represents the current and T the maximum iteration.

The second change to Eq. 3 is to apply the stop-gradient operator $sg[\cdot]$ [16] to $z^\times$, which sets the gradients of $z^\times$ to 0 during the backward pass. The stop-gradient operator prevents the undesirable scenario where $z^\times$ is pulled towards a naive solution, i.e., different MRI samples are mapped to the same weighted average of all SOM representations. This risk of deriving the naive solution is especially high in the early stages of the training when the SOM representations are randomly initialized and may not accurately represent the clusters.

Longitudinal Consistency Regularization. We derive a SOM grid related to brain aging by generating an age-stratified latent space. Specifically, the latent space is defined by a smooth trajectory field (Fig. 1, blue box) characterizing the morphological changes associated with brain aging. The smoothness is based on the assumption that MRIs with similar appearances (close latent representations on the latent space) should have similar trajectories. It is enforced by modeling the similarity between each subject-specific trajectory Δz with a reference trajectory that represents the average trajectory of the cluster. Specifically, $\Delta g_{i,j}$ is the reference trajectory (Fig. 1, green arrow) associated with $g_{i,j}$ then the reference trajectories of all clusters $\mathcal{G}_\Delta = \{\Delta g_{i,j}\}_{i=1,j=1}^{N_r,N_c}$ represent the average aging of SOM clusters with respect to the training set. As all subject-specific trajectories are iteratively updated during the training, it is computationally infeasible to keep track of $\mathcal{G}_\Delta$ on the whole training set. We instead propose to compute the exponential moving average (EMA) (Fig. 1, black box), which iteratively aggregates the average trajectory with respect to a training batch to $\mathcal{G}_\Delta$:

$$\Delta g_{i,j} \leftarrow \begin{cases} \Delta h_{i,j} & t = 0 \\ \Delta g_{i,j} & t > 0 \text{ and } |\Omega_{i,j}| = 0 \\ \alpha \cdot \Delta g_{i,j} + (1-\alpha) \cdot \Delta h_{i,j} & t > 0 \text{ and } |\Omega_{i,j}| > 0 \end{cases}$$

$$\text{with } \Delta h_{i,j} := \frac{1}{|\Omega_{i,j}|} \sum_{k=1}^{N_{bs}} \mathbb{1}[\epsilon_k^u = (i,j)] \cdot \Delta z_k \text{ and } |\Omega_{i,j}| := \sum_{k=1}^{N_{bs}} \mathbb{1}[\epsilon_k^u = (i,j)].$$

α is the EMA keep rate, k denotes the index of the sample pair, N_{bs} symbolizes the batch size, $\mathbb{1}[\cdot]$ is the indicator function, and $|\Omega_{i,j}|$ denotes the number of sample pairs with $\epsilon^u = (i,j)$ within a batch. Then in each iteration, $\Delta h_{i,j}$ (Fig. 1, purple arrow) represents the batch-wise average of subject-specific trajectories for sample pairs with $\epsilon^u = (i,j)$. By iteratively updating $\mathcal{G}_\Delta$, $\mathcal{G}_\Delta$ then approximate the average trajectories derived from the entire training set. Lastly, inspired by [11,12], the longitudinal consistency regularization is formulated as

$$L_{dir} := \mathbb{E}_{(x^u,x^v)\sim\mathcal{S}}\left(1 - cos(\theta[\Delta z, sg[\Delta g_{\epsilon^u}]])\right),$$

where $\theta[\cdot,\cdot]$ denotes the angle between two vectors. Since Δg is optimized by EMA, the stop-gradient operator is again incorporated to only compute the gradient with respect to Δz in L_{dir}.

Objective Function. The complete objective function is the weighted combination of the prior losses with weighing parameters λ_{commit}, λ_{som}, and λ_{dir}:

$$L := L_{recon} + \lambda_{commit} \cdot L_{commit} + \lambda_{som} \cdot L_{som} + \lambda_{dir} \cdot L_{dir}$$

The objective function encourages a smooth trajectory field of aging on the latent space while maintaining interpretable SOM representations for analyzing brain age in a pure self-supervised fashion.

2.2 SOM Similarity Grid

During inference, a (2D) similarity grid ρ is computed by the closeness between the latent representation z of an MRI sample and the SOM representations:

$$\rho := softmax(- \parallel z - \mathcal{G} \parallel_2^2 /\gamma) \text{ with } \gamma := std(\parallel z - \mathcal{G} \parallel_2^2)$$

std denotes the standard deviation of the distance between z to all SOM representations. As the SOM grid is learned to be associated with brain age (e.g., represents aging from left to right), the similarity grid essentially encodes a "likelihood function" of the brain age in z. Given all MRIs of a longitudinal scan, the change across the corresponding similarity grids over time represents the brain aging process of that individual. Furthermore, brain aging on the group-level is captured by first computing the average similarity grid for an age group and then visualizing the difference of those average similarity grids across age groups.

3 Experiments

3.1 Experimental Setting

Dataset. We evaluated the proposed method on all 632 longitudinal T1-weighted MRIs (at least two visits per subject, 2389 MRIs in total) from ADNI-1 [13]. The data set consists of 185 NC (age: 75.57 ± 5.06 years), 193 subjects diagnosed with sMCI (age: 75.63 ± 6.62 years), 135 subjects diagnosed with pMCI (age: 75.91 ± 5.35 years), and 119 subjects with AD (age: 75.17 ± 7.57 years). There was no significant age difference between the NC and AD cohorts (p = 0.55, two-sample *t*-test) as well as the sMCI and pMCI cohorts (p = 0.75). All MRI images were preprocessed by a pipeline including denoising, bias field correction, skull stripping, affine registration to a template, re-scaling to 64 × 64 × 64 volume, and transforming image intensities to z-scores.

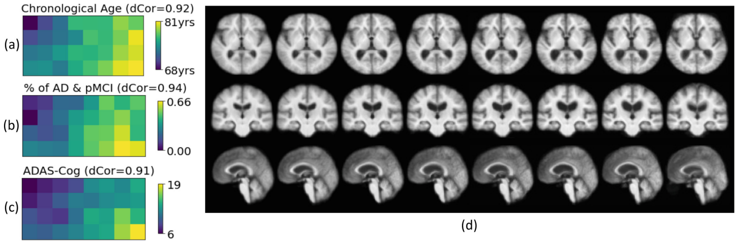

Fig. 2. The color at each SOM representation encodes the average value of (a) chronological age, (b) % of AD and pMCI, and (c) ADAS-Cog score across the training samples of that cluster; (d) Confined to the last row of the grid, the average MRI of 20 latent representations closest to the corresponding SOM representation. (Color figure online)

Implementation Details. Let C_k denote a Convolution(kernel size of $3 \times 3 \times 3$, Conv_k)-BatchNorm-LeakyReLU(slope of 0.2)-MaxPool(kernel size of 2) block with k filters, and CD_k an Convolution-BatchNorm-LeakyReLU-Upsample block. The architecture was designed as C_{16}-C_{32}-C_{64}-C_{16}-Conv_{16}-CD_{64}-CD_{32}-CD_{16}-CD_{16}-Conv_1, which results in a latent space of 1024 dimensions. The training of SOM is difficult in this high-dimensional space with random initialization in practice, thus we first pre-trained the model with only L_{recon} for 10 epochs and initialized the SOM representations by doing k-means of all training samples using this pre-trained model. Then, the network was further trained for 40 epochs with regularization weights set to $\lambda_{recon} = 1.0$, $\lambda_{commit} = 0.5$, $\lambda_{som} = 1.0$, $\lambda_{dir} = 0.2$. Adam optimizer with learning rate of 5×10^{-4} and weight decay of 10^{-5} were used. τ_{min} and τ_{max} in L_{som} were set as 0.1 and 1.0 respectively. An EMA keep rate of $\alpha = 0.99$ was used to update reference trajectories. A batch size $N_{bs} = 64$ and the SOM grid size $N_r = 4, N_c = 8$ were applied.

Evaluation. We performed five-fold cross-validation (folds split based on subjects) using 10% of the training subjects for validation. The training data was augmented by flipping brain hemispheres and random rotation and translation. To quantify the interpretability of the SOM grid, we correlated the coordinates of the SOM grid with quantitative measures related to brain age, e.g., chronological age, the percentage of subjects with severe cognitive decline, and Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog). We illustrated the interpretability with respect to brain aging by visualizing the changes in the SOM similarity maps over time. We further visualized the trajectory vector field along with SOM representations by projecting the 1024-dimensional representations to the first two principal components of SOM representations. Lastly, we quantitatively evaluated the quality of the representations by applying them to the downstream tasks of classifying sMCI vs. pMCI and ADAS-Cog prediction. We measured the classification accuracy via Balanced accuracy (BACC) and Area Under Curve (AUC) and the prediction accuracy via R2 and root-mean-square error (RMSE). The classifier and predictor were multi-layer per-

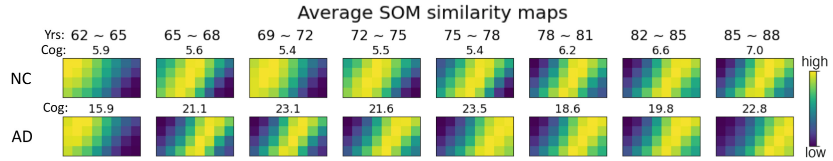

Fig. 3. The average similarity grid ρ over subjects of a specific age and diagnosis (NC vs AD). Each grid encodes the likelihood of the average brain age of the corresponding sub-cohort. Cog denotes the average ADAS-Cog score.

ceptrons containing two fully connected layers of dimensions 1024 and 64 with a LeakyReLU activation. We compared the accuracy metrics to models using the same architecture with encoders pre-trained by other representation learning methods, including unsupervised methods (AE, VAE [4]), self-supervised method (SimCLR [1]), longitudinal self-supervised method (LSSL [17]), and longitudinal neighborhood embedding (LNE [12]). All comparing methods used the same experimental setup (e.g., encoder-decoder, learning rate, batch size, epochs, etc.), and the method-specific hyperparameters followed [12].

3.2 Results

Interpretability of SOM Embeddings. Fig. 2 shows the stratification of brain age over the SOM grid $\mathcal{G}$. For each grid entry, we show the average value of chronological age (Fig. 2(a)), % of AD & pMCI (Fig. 2(b)), and ADAS-Cog score (Fig. 2(c)) over samples of that cluster. We observed a trend of older brain age (yellow) from the upper left towards the lower right, corresponding to older chronological age and worse cognitive status. The SOM grid index strongly correlated with these three factors (distance correlation of 0.92, 0.94, and 0.91 respectively). Figure 2(d) shows the average brain over 20 input images with representations that are closest to each SOM representation of the last row of the grid (see Supplement Fig. S1 for all rows). From left to right the ventricles are enlarging and the brain is atrophying, which is a hallmark for brain aging effects.

Interpretability of Similarity Grid. Visualizing the average similarity grid ρ of the NC and AD at each age range in Fig. 3, we observed that higher similarity (yellow) gradually shifts towards the right with age in both NC and AD (see Supplemental Fig. S2 for sMCI and pMCI cohorts). However, the shift is faster for AD, which aligns with AD literature reporting that AD is linked to accelerated brain aging [15]. Furthermore, the subject-level aging effects shown in Supplemental Fig. S3 reveal that the proposed visualization could capture subtle morphological changes caused by brain aging.

Interpretability of Trajectory Vector Field. Fig. 4 plots the PCA projections of the latent space in 2D, which shows a smooth trajectory field (gray arrows) and reference trajectories $\mathcal{G}_\Delta$ (blue arrows) representing brain aging.

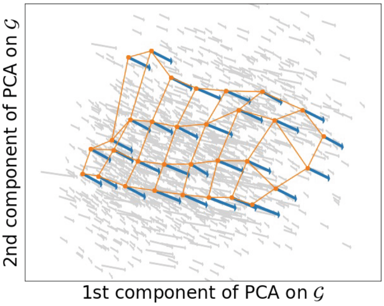

Fig. 4. 2D PCA of the LSOR's latent space. Light gray arrows represent Δz. The orange grid represents the relationships between SOM representations and associated reference trajectory $\Delta\mathcal{G}$ (blue arrow). (Color figure online)

Table 1. Supervised downstream tasks using the learned representations z (without fine-tuning the encoder). LSOR achieved comparable or higher accuracy scores than other state-of-the-art self- and un-supervised methods.

Methods	sMCI/pMCI		ADAS-Cog	
	BACC	AUC	R2	RMSE
AE	62.6	65.4	0.26	6.98
VAE [4]	61.3	64.8	0.23	7.17
SimCLR [1]	63.3	66.3	0.26	6.79
LSSL [17]	69.4	71.8	0.29	6.49
LNE [12]	**70.6**	72.1	0.30	6.46
LSOR	69.8	**72.4**	**0.32**	**6.31**

This projection also preserved the 2D grid structure (orange) of the SOM representations suggesting that aging was the most important variation in the latent space.

Downstream Tasks. To evaluate the quality of the learned representations, we froze encoders trained by each method without fine-tuning and utilized their representations for the downstream tasks (Table 1). On the task of sMCI vs. pMCI classification (Table 1 (left)), the proposed method achieved a BACC of 69.8 and an AUC of 72.4, a comparable accuracy ($p > 0.05$, DeLong's test) with LSSL [17] and LNE [12], two state-of-the-art self-supervised methods on this task. On the ADAS-Cog score regression task, the proposed method obtained the best accuracy with an R2 of 0.32 and an RMSE of 6.31. It is worth mentioning that an accurate prediction of the ADAS-Cog score is very challenging due to its large range (between 0 and 70) and its subjectiveness resulting in large variability across exams [2] so that even larger RMSEs have been reported for this task [7]. Furthermore, our representations were learned in an unsupervised manner so that further fine-tuning of the encoder would improve the prediction accuracy.

4 Conclusion

In this work, we proposed LSOR, the first SOM-based learning framework for longitudinal MRIs that is self-supervised and interpretable. By incorporating a soft SOM regularization, the training of the SOM was stable in the high-dimensional latent space of MRIs. By regularizing the latent space based on longitudinal consistency as defined by longitudinal MRIs, the latent space formed a smooth trajectory field capturing brain aging as shown by the resulting SOM grid. The interpretability of the representations was confirmed by the correlation between the SOM grid and cognitive measures, and the SOM similarity map.

When evaluated on downstream tasks sMCI vs. pMCI classification and ADAS-Cog prediction, LSOR was comparable to or better than representations learned from other state-of-the-art self- and un-supervised methods. In conclusion, LSOR is able to generate a latent space with high interpretability regarding brain age purely based on MRIs, and valuable representations for downstream tasks.

Acknowledgement. This work was partly supported by funding from the National Institute of Health (MH113406, DA057567, AA017347, AA010723, AA005965, and AA028840), the DGIST R&D program of the Ministry of Science and ICT of KOREA (22-KUJoint-02), Stanford's Department of Psychiatry & Behavioral Sciences Faculty Development & Leadership Award, and by Stanford HAI Google Cloud Credit.

References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
2. Connor, D.J., Sabbagh, M.N.: Administration and scoring variance on the ADAS-Cog. J. Alzheimers Dis. **15**(3), 461–464 (2008)
3. Fortuin, V., Hüser, M., Locatello, F., Strathmann, H., Rätsch, G.: SOM-VAE: interpretable discrete representation learning on time series. In: International Conference on Learning Representations (2019)
4. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
5. Kohonen, T.: The self-organizing map. Proc. IEEE **78**(9), 1464–1480 (1990)
6. Li, O., Liu, H., Chen, C., Rudin, C.: Deep learning for case-based reasoning through prototypes: a neural network that explains its predictions. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
7. Ma, D., Pabalan, C., Interian, Y., Raj, A.: Multi-task learning and ensemble approach to predict cognitive scores for patients with Alzheimer's disease. bioRxiv, pp. 2021–12 (2021)
8. Manduchi, L., Hüser, M., Vogt, J., Rätsch, G., Fortuin, V.: DPSOM: deep probabilistic clustering with self-organizing maps. In: Conference on Neural Information Processing Systems Workshop on Machine Learning for Health (2019)
9. Molnar, C.: Interpretable machine learning (2020)
10. Mulyadi, A.W., Jung, W., Oh, K., Yoon, J.S., Lee, K.H., Suk, H.I.: Estimating explainable Alzheimer's disease likelihood map via clinically-guided prototype learning. Neuroimage **273**, 120073 (2023)
11. Ouyang, J., Zhao, Q., Adeli, E., Zaharchuk, G., Pohl, K.M.: Self-supervised learning of neighborhood embedding for longitudinal MRI. Med. Image Anal. **82**, 102571 (2022)
12. Ouyang, J., et al.: Self-supervised longitudinal neighbourhood embedding. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 80–89. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_8
13. Petersen, R.C., et al.: Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology **74**(3), 201–209 (2010)
14. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. **1**(5), 206–215 (2019)

15. Toepper, M.: Dissociating normal aging from Alzheimer's disease: a view from cognitive neuroscience. J. Alzheimers Dis. **57**(2), 331–352 (2017)
16. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. Adv. Neural Inf. Process. Syst. **30** (2017)
17. Zhao, Q., Liu, Z., Adeli, E., Pohl, K.M.: Longitudinal self-supervised learning. Med. Image Anal. **71**, 102051 (2021)