(MICCAI 2021)
Self-Supervised Longitudinal Neighbourhood Embedding

Jiahong Ouyang[1], Qingyu Zhao[1], Ehsan Adeli[1], Edith V Sullivan[1], Adolf Pfefferbaum[1,2], Greg Zaharchuk[1], and Kilian M Pohl[1,2]

[1] Stanford University, Stanford CA 94305, USA
[2] SRI International, Menlo Park CA, 94025, USA

Motivation

Longitudinal MRIs are often used to capture the gradual deterioration of brain structure and function caused by aging or neurological diseases. Analyzing this data via machine learning generally requires a large number of ground-truth labels, which are often missing or expensive to obtain. **Reducing the need for labels**, they propose a self-supervised strategy for representation learning named Longitudinal Neighborhood Embedding (LNE).

Method

LNE explicitly models the similarity between trajectory vectors across different subjects. They do so by building a graph in each training iteration defining neighborhoods in the latent space **so that the progression direction of a subject follows the direction of its neighbors.**

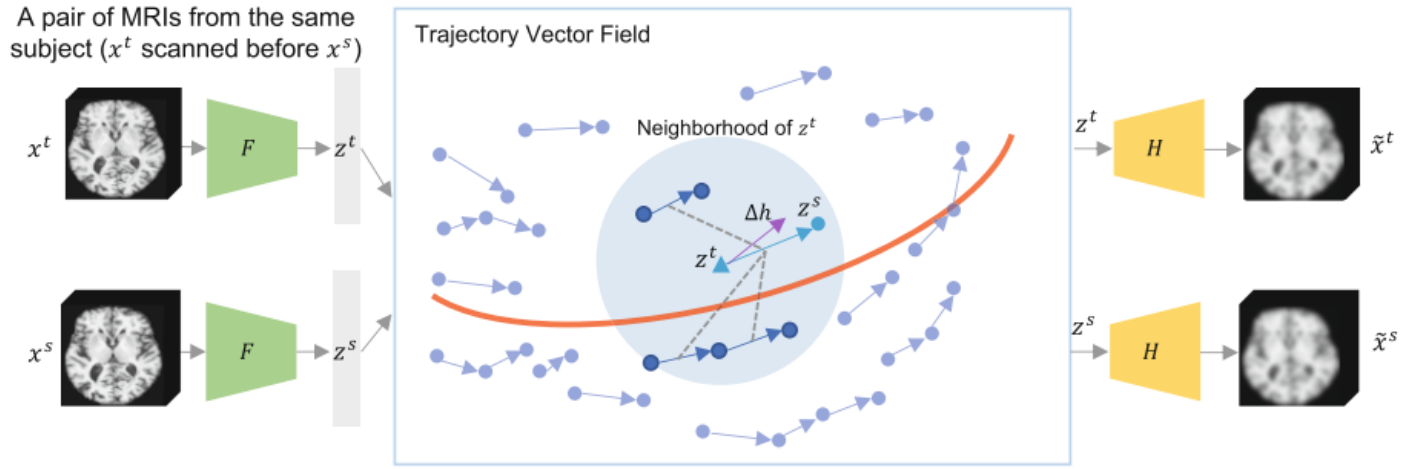
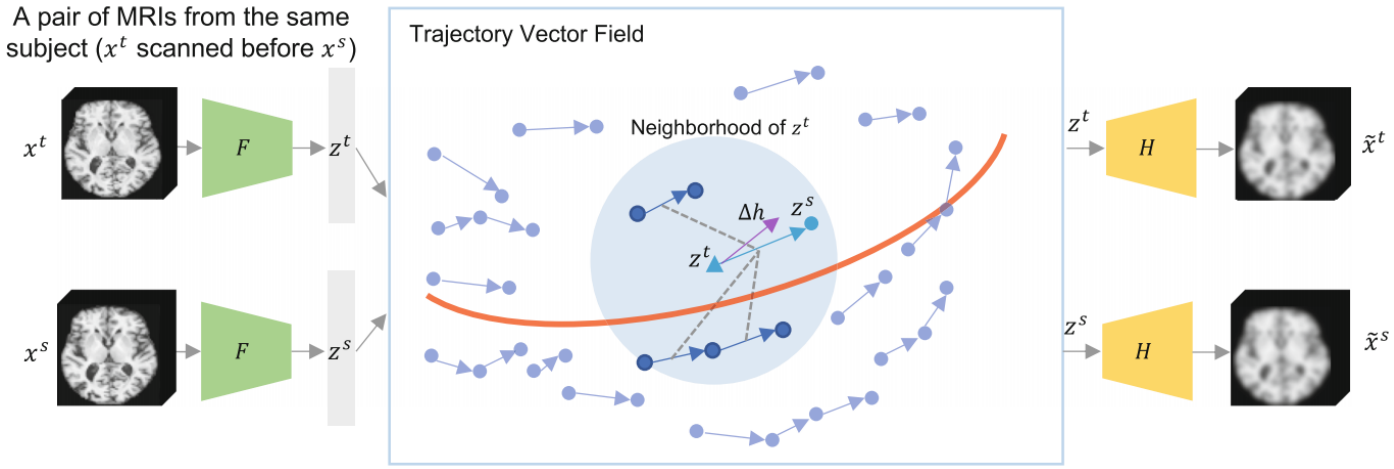


Fig. 1. Overview of the proposed method: an encoder projects a subject-specific image pair (x^t, x^s) into the latent space resulting in a trajectory vector (cyan). We encourage the direction of this vector to be consistent with Δh (purple), a vector pooled from the neighborhood of z^t (blue circle). As a result, the latent space encodes the global morphological change linked to aging (red curve). (Color figure online)

Method



Objective Function.

$$L := \mathbf{E}_{(x^t,x^s)\sim\mathcal{S}} \left(\| x^t - \tilde{x}^t \|_2^2 + \| x^s - \tilde{x}^s \|_2^2 - \lambda \cdot \cos(\theta_{\langle \Delta z, \Delta h \rangle}) \right),$$

1. Pairwise Training Strategy.

i. subject-specific image pairs

$$(x^t, x^s)$$

ii. latent representations

$$z^t = F(x^t),\ z^s = F(x^s)$$

iii. normalized trajectory vector

$$\Delta z^{(t,s)} = (z^s - z^t)/\Delta t^{(t,s)}$$

2. Longitudinal Neighbourhood Embedding.

i. adjacency matrix

$$P_{i,j} = \| z_i^t - z_j^t \|_2$$

$$A_{i,j} := \begin{cases} exp(-\frac{P_{i,j}^2}{2\sigma_i^2}) & j \in \mathcal{N}_i \\ 0, & j \notin \mathcal{N}_i \end{cases}.$$

with $\sigma_i := max(P_{i,j\in\mathcal{N}_i}) - min(P_{i,j\in\mathcal{N}_i})$

ii. longitudinal neighbourhood embedding

$$\Delta h_i := \sum_{j\in\mathcal{N}_i} A_{i,j} D_{i,j}^{-1} \Delta z_j,$$

Experiments

1. Predicting age

Lab data set: 582 MRIs of 274 healthy individuals with the age ranging from 20 to 90. Each subject had 1 to 13 scans with an average of 2.3 scans spanning an average time interval of 3.8 years.

2. Classification

The second data set comprised 2389 longitudinal T1-weighted MRIs (at least two visits per subject) from ADNI, which consisted of 185 NC (age: 75.57 ± 5.06 years), 119 subjects with AD (age: 75.17 ± 7.57 years), 193 subjects diagnosed with sMCI (age: 75.63 ± 6.62 years), and 135 subjects diagnosed with pMCI (age: 75.91 ± 5.35 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 ($p = 0.75$).

Result

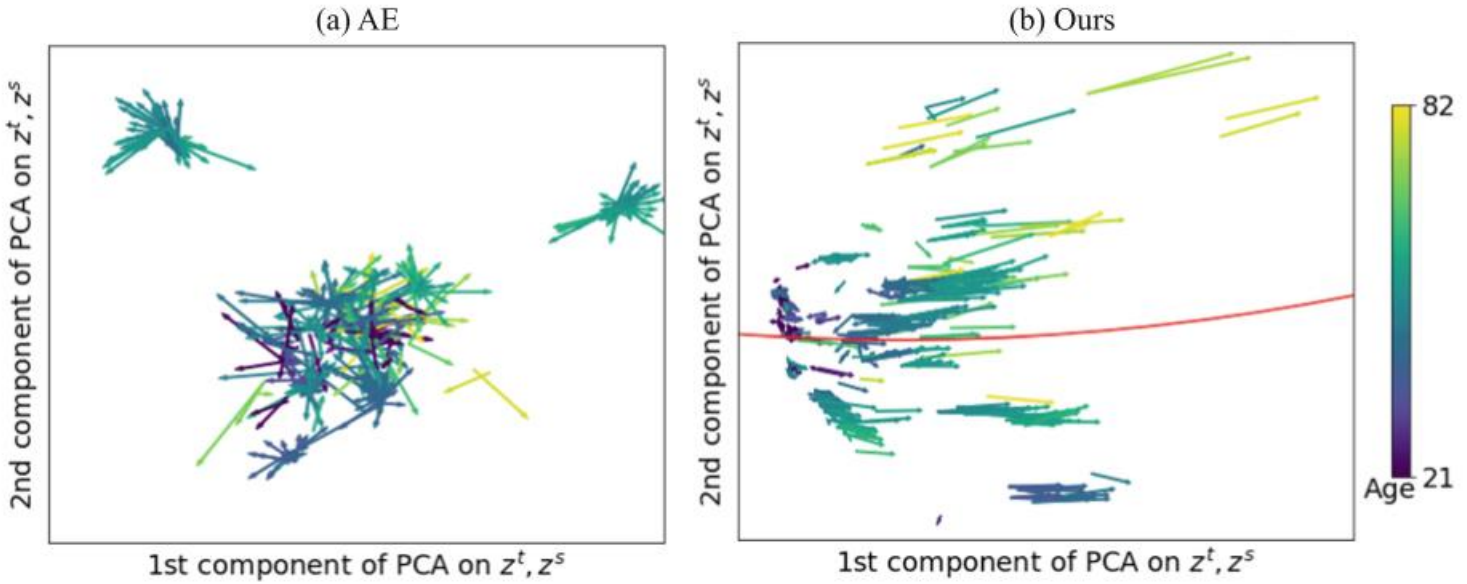


Fig. 2. Experiments on healthy aging: Latent space of AutoEncoder (AE) (a) and the proposed LNE (b) projected into 2D PCA space of z^t and z^s. Arrows represent Δz and are color-coded by the age of z^t. The global trajectory in (b) is fitted by robust linear mixed effect model (red curve). (Color figure online)

Result

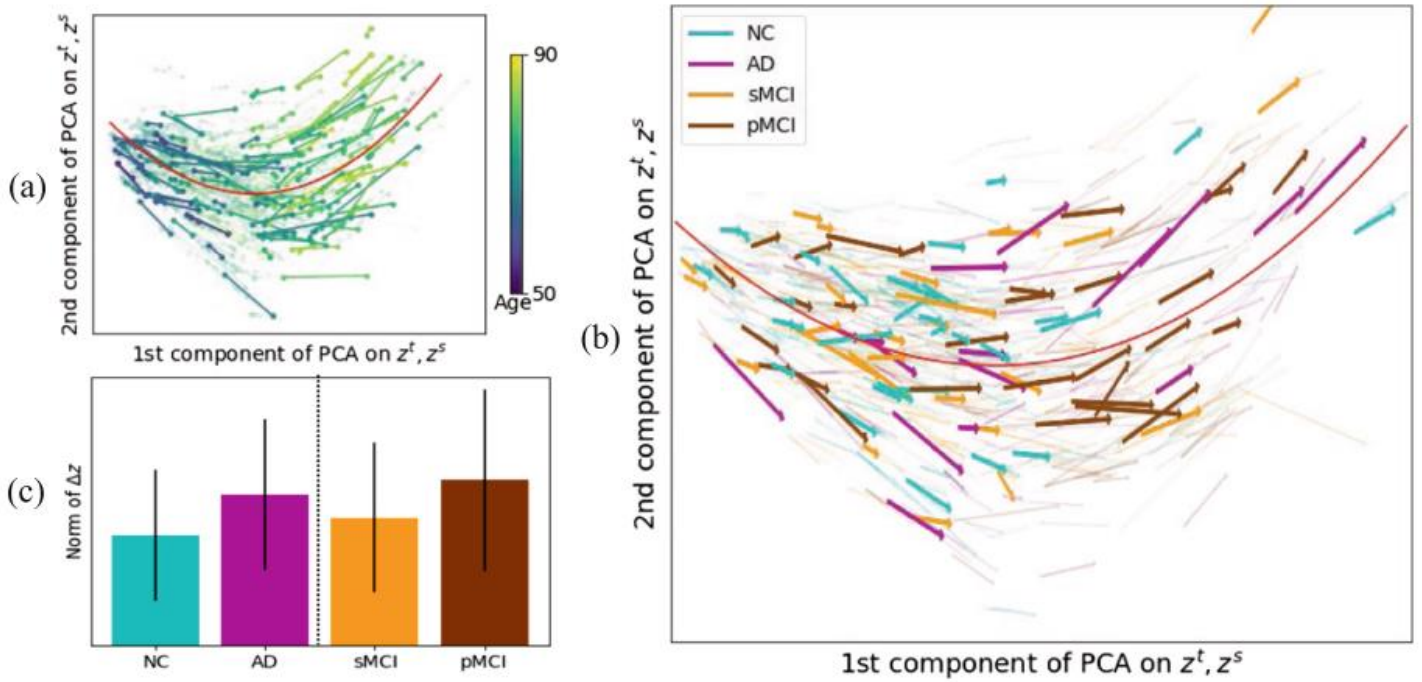


Fig. 3. Experiments on ADNI: (a) The age distribution of the latent space. Lines connecting z^t and z^s are color-coded by the age of z^t; Red curve is the global trajectory fitted by a robust linear mixed effect model. (b) Trajectory vector field color-coded by diagnosis groups; (c) The norm of Δz encoding the speed of aging for 4 different diagnosis groups. (Color figure online)

Result

Table 1. Supervised downstream tasks in frozen or fine-tune scenarios. Left: Age regression on healthy subjects with R2 as an evaluation metric. Right: classification on ADNI dataset with BACC as the metric.

Methods	Health aging (**R2**)		ADNI (**BACC**)			
	Age		NC vs AD		sMCI vs pMCI	
	Frozen	Fine-tune	Frozen	Fine-tune	Frozen	Fine-tune
No pretrain	–	0.72	–	79.4	–	69.3
AE	0.53	0.69	72.2	80.7	62.6	69.5
VAE [12]	0.51	0.69	66.7	77.0	61.3	63.8
SimCLR [6]	0.56	0.73	72.9	82.4	63.3	69.5
LSSL [24]	0.59	**0.74**	74.2	82.1	69.4	71.2
Ours (LNE)	**0.62**	**0.74**	**81.9**	**83.6**	**70.6**	**73.4**

(MICCAI 2021)
Contrastive Learning with Continuous Proxy Meta-Data for 3D MRI Classification

Benoit Dufumier[1,2], Pietro Gori[2], Julie Victor[1], Antoine Grigis[1], Michel Wessa[3], Paolo Brambilla[4], Pauline Favre[1], Mircea Polosan[5], Colm McDonald[6], Camille Marie Piguet, and Edouard Duchesnay[1] for the Alzheimer's Disease Neuroimaging Initiative

[1] NeuroSpin, CEA Saclay, Université Paris-Saclay, France
benoit.dufumier@cea.fr
[2] LTCI, Télécom Paris, IPParis, France
[3] Dep. of Neuropsychology, Johannes-Gutenberg Univ. of Mainz, Germany
[4] Dep. of Neurosciences, Fondazione IRCCS, University of Milan, Italy
[5] Université Grenoble Alpes, Inserm U1216, CHU Grenoble Alpe, France
[6] Centre for Neuroimaging & Cognitive Genomics (NICOG), Galway, Ireland

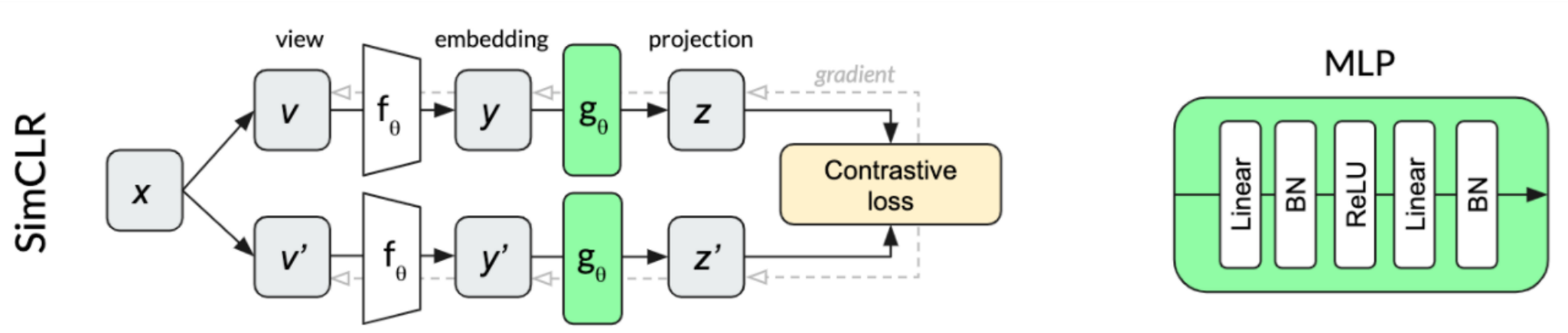
Motivation

Most of recent works do not take advantage of available meta-data, such as participant’s age.

Goal

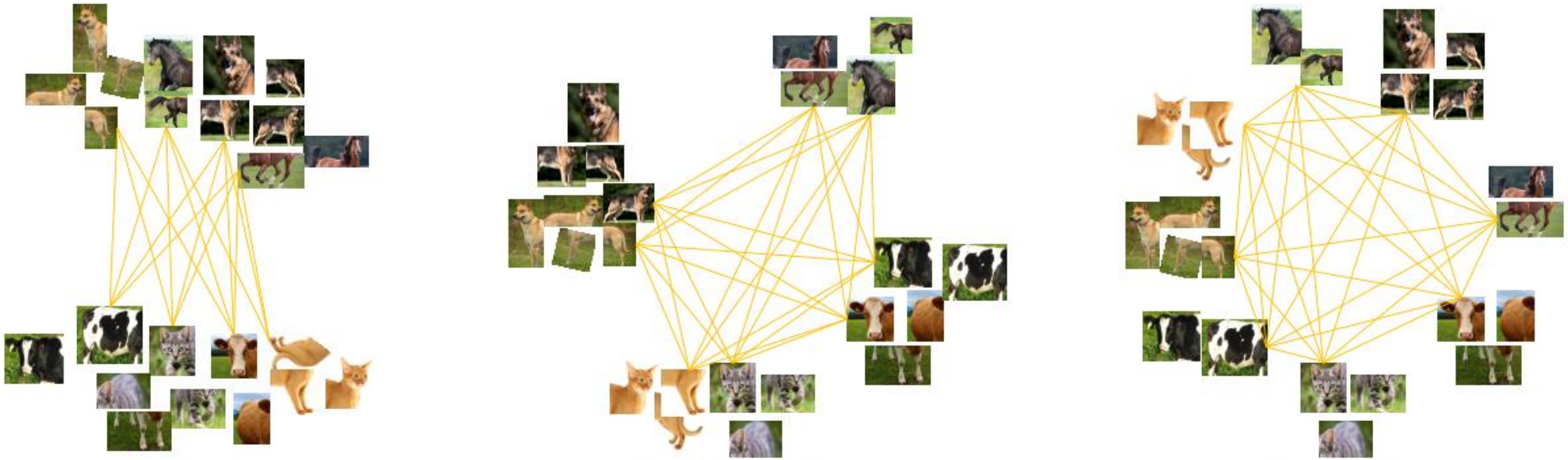
Propose a new **y-Aware InfoNCE loss** inspired from the Noise Contrastive Estimation loss that aims at improving the positive sampling according to the similarity between two proxy meta-data

SimCLR



$$L_N = -\mathbb{E}_X \left[\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp(f(x)^T f(x^+)) + \sum_{j=1}^{N-1} \exp\left(f(x)^T f(x_j^-)\right)} \right]$$

Problems



(a) under-clustering (b) ideal clustering (c) over-clustering

Figure 2. Illustration of under-clustering and over-clustering. Each sample pair connected by a yellow line represents a negative pair.

[1] (arXiv 2104) Solving Inefficiency of Self-supervised Representation Learning

Method

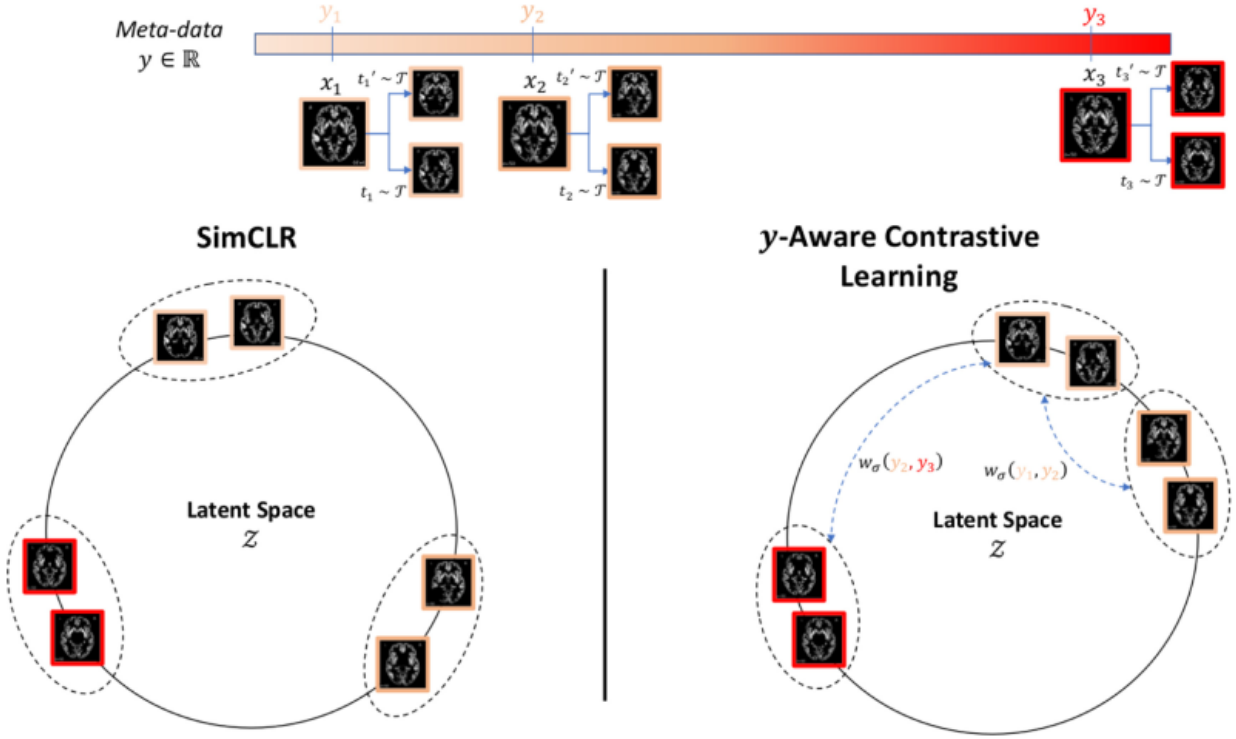


Fig. 1. Differently from SimCLR [5], our new loss can handle meta-data $y \in \mathbb{R}$ by redefining the notion of similarity between two images in the latent space $\mathcal{Z}$. For an image x_i, transformed twice through two augmentations $t_1, t_1' \sim \mathcal{T}$, the resulting views $(t_1(x_i), t_2(x_i))$ are expected to be close in the latent space through the learnt mapping f_θ, as in SimCLR. However, we also expect a different input $x_{k \neq i}$ to be close to x_i in $\mathcal{Z}$ if the two *proxy* meta-data y_i and y_k are similar. We define a similarity function $w_\sigma(y_i, y_k)$ that quantifies this notion of similarity.

InfoNCE loss:

$$\mathcal{L}_{NCE} = -\log \frac{e^{f_\theta(v_1^i, v_2^i)}}{\frac{1}{n}\sum_{j=1}^{n} e^{f_\theta(v_1^i, v_2^j)}}$$

y-Aware InfoNCE loss:

$$\mathcal{L}^y_{NCE} = -\sum_{k=1}^{n} \frac{w_\sigma(y_k, y_i)}{\sum_{j=1}^{n} w_\sigma(y_j, y_i)} \log \frac{e^{f_\theta(v_1^i, v_2^k)}}{\frac{1}{n}\sum_{j=1}^{n} e^{f_\theta(v_1^i, v_2^j)}}$$

Radius Basis Function (RBF) kernel

$$w_\sigma = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

Experiments

- Datasets

  - **Big Healthy Brains (BHB) dataset** We aggregated 13 publicly available datasets of 3D T1 MRI scans of healthy controls (HC) acquired on more than 70 different scanners and comprising *N* = 10 samples. We use this dataset only to pre-train our model with the **participant's age as the *proxy* meta-data**. The learnt representation is then tested on the following four data-sets using as final task a binary classification between HC and patients.
  - **SCHIZCONNECT-VIP** It comprises *N* = 605 multi-site MRI scans including 275 patients with strict schizophrenia (SCZ：精神分裂) and 330 HC.
  - **BIOBD** This dataset includes *N* = 662 MRI scans acquired on 8 different sites with 356 HC and 306 patients with bipolar disorder (BD：躁郁症).
  - **BSNIP** [25] It includes *N* = 511 MRI scans with *N* = 200 HC, *N* = 194 SCZ and *N* = 117 BD.
  - **Alzheimer's Disease Neuroimaging Initiative (ADNI-GO)** We use *N* = 387 co-registered T1-weighted MRI images divided in *N* = 199 healthy controls and *N* = 188 Alzheimer's patients (AD：阿兹海默). We only included one scan per patient at the first session (baseline).

Result

Evaluation of the representation

Classification is performed using a linear layer on top of the pre-trained frozen encoders.

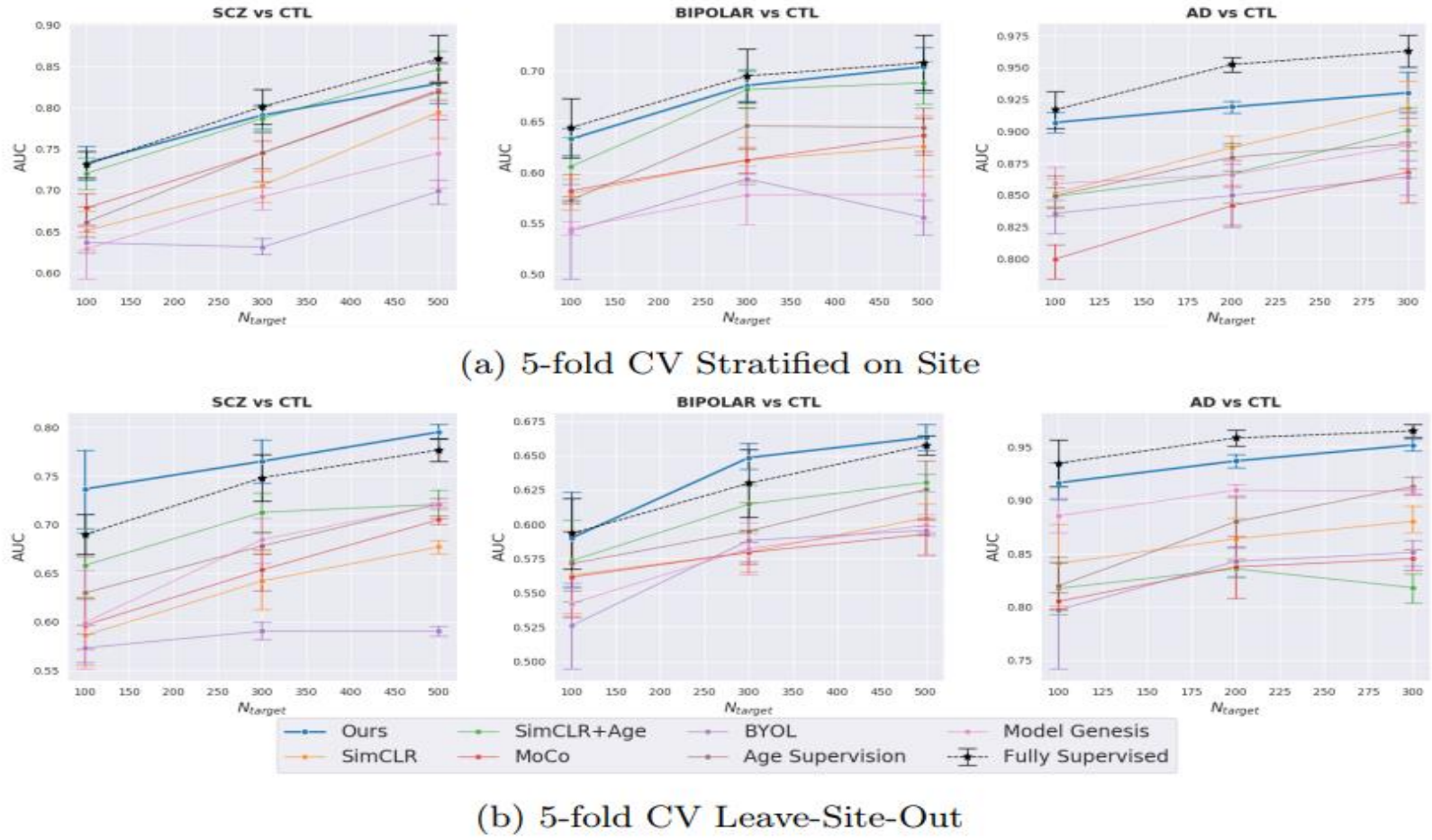


Fig. 2: Comparison of different representations in terms of classification accuracy (downstream task) on three different data-sets (one per column). Classification is performed using a linear layer on top of the pre-trained frozen encoders. (a) Data for training/validation and test come from the the same acquisition sites (b) Data for training/validation and test come from different sites.

Result

Fine-tuning Results

Backbone	Pre-training	SCZ vs HC		BD vs HC		AD vs HC	
		$N_{train} = 100$	$N_{train} = 500$	$N_{train} = 100$	$N_{train} = 500$	$N_{train} = 100$	$N_{train} = 300$
UNet	None	$72.62_{\pm 0.9}$	$76.45_{\pm 2.2}$	$63.03_{\pm 2.7}$	$69.20_{\pm 3.7}$	$88.12_{\pm 3.2}$	$94.16_{\pm 3.9}$
	Model Genesis [29]	$73.00_{\pm 3.4}$	$81.8_{\pm 4.7}$	$60.96_{\pm 1.8}$	$67.04_{\pm 4.4}$	$89.44_{\pm 2.6}$	$95.16_{\pm 3.3}$
	SimCLR [4]	$73.63_{\pm 2.4}$	$80.12_{\pm 4.9}$	$59.89_{\pm 2.6}$	$66.51_{\pm 4.3}$	$90.60_{\pm 2.5}$	$94.21_{\pm 2.7}$
	Age Prediction w/ D.A	$\underline{75.32}_{\pm 2.2}$	$\underline{85.27}_{\pm 2.3}$	$\mathbf{64.6}_{\pm \mathbf{1.6}}$	$\mathbf{70.78}_{\pm \mathbf{2.1}}$	$\underline{91.71}_{\pm 1.1}$	$\underline{95.26}_{\pm 1.5}$
	Age-Aware Contrastive Learning (ours)	$\mathbf{75.95}_{\pm 2.7}$	$\mathbf{85.73}_{\pm 4.7}$	$\underline{63.79}_{\pm 3.0}$	$\underline{70.35}_{\pm 2.7}$	$\mathbf{92.19}_{\pm \mathbf{1.8}}$	$\mathbf{96.58}_{\pm \mathbf{1.6}}$
DenseNet	None	$73.09_{\pm 1.6}$	$85.92_{\pm 2.8}$	$64.39_{\pm 2.9}$	$70.77_{\pm 2.7}$	$92.23_{\pm 1.6}$	$93.68_{\pm 1.7}$
	None w/ D.A	$\underline{74.71}_{\pm 1.3}$	$86.94_{\pm 2.8}$	$64.79_{\pm 1.3}$	$72.25_{\pm 1.5}$	$92.10_{\pm 1.8}$	$94.16_{\pm 2.5}$
	SimCLR [5]	$70.80_{\pm 1.9}$	$86.35_{\pm 2.2}$	$60.57_{\pm 1.9}$	$67.99_{\pm 3.3}$	$91.54_{\pm 1.9}$	$94.26_{\pm 2.9}$
	Age Prediction	$72.90_{\pm 4.6}$	$\underline{87.75}_{\pm 2.0}$	$64.60_{\pm 3.6}$	$72.07_{\pm 3.0}$	$92.07_{\pm 2.7}$	$\underline{96.37}_{\pm 0.9}$
	Age Prediction w/ D.A	$74.06_{\pm 3.4}$	$86.90_{\pm 1.6}$	$\mathbf{65.79}_{\pm 2.0}$	$\underline{73.02}_{\pm 4.3}$	$\mathbf{94.01}_{\pm \mathbf{1.4}}$	$96.10_{\pm 3.0}$
	Age-Aware Contrastive Learning (ours)	$\mathbf{76.33}_{\pm 2.3}$	$\mathbf{88.11}_{\pm 1.5}$	$\underline{65.36}_{\pm 3.7}$	$\mathbf{73.33}_{\pm \mathbf{4.3}}$	$\underline{93.87}_{\pm 1.3}$	$\mathbf{96.84}_{\pm \mathbf{2.3}}$

Table 1: Fine-tuning results using 100 or 500 (300 for AD vs HC) training subjects. For each task, we report the AUC (%) of the fine-tuned models initialized with different approaches with 5-fold cross-validation. For age prediction, we employ the same transformations as in contrastive learning for the Data Augmentation (D.A) strategy. Best results are in **bold** and second bests are underlined.

Thanks for listening