OCL: Ordinal Contrastive Learning for Imputing Features with Progressive Labels

Seunghun Baek [1*] Jaeyoon Sim[1*] Guorong Wu[2] Won Hwa Kim[1]

[*]Equal Contribution

[1]Pohang University of Science and Technology, South Korea

[2]University of North Carolina at Chapel Hill, USA

Main Idea

- Problem: Accurately discriminating progressive stages of Alzheimer's Disease (AD) is crucial, but missing data from multiple imaging modalities hinder robust analyses.
- Question: How can we impute missing neuroimaging features **while maintaining disease progression information for Alzheimer's diagnosis**?
- Solution: Holistic imaging feature imputation method leveraging Ordinal Contrastive Learning (OCL) to align modality-independent embeddings with disease progression.

Contribution 1. Our method accurately estimates unobserved imaging measures for individual subjects using their existing data to solidify downstream analyses.

Contribution 2. We introduce ordinal contrastive learning, which aligns samples in the embedding space based on their disease severity.

Contribution 3. The experiments on ADNI data show that our method accurately translates data, capturing realistic information for subsequent analyses.

Ordinal Contrastive Learning (OCL)

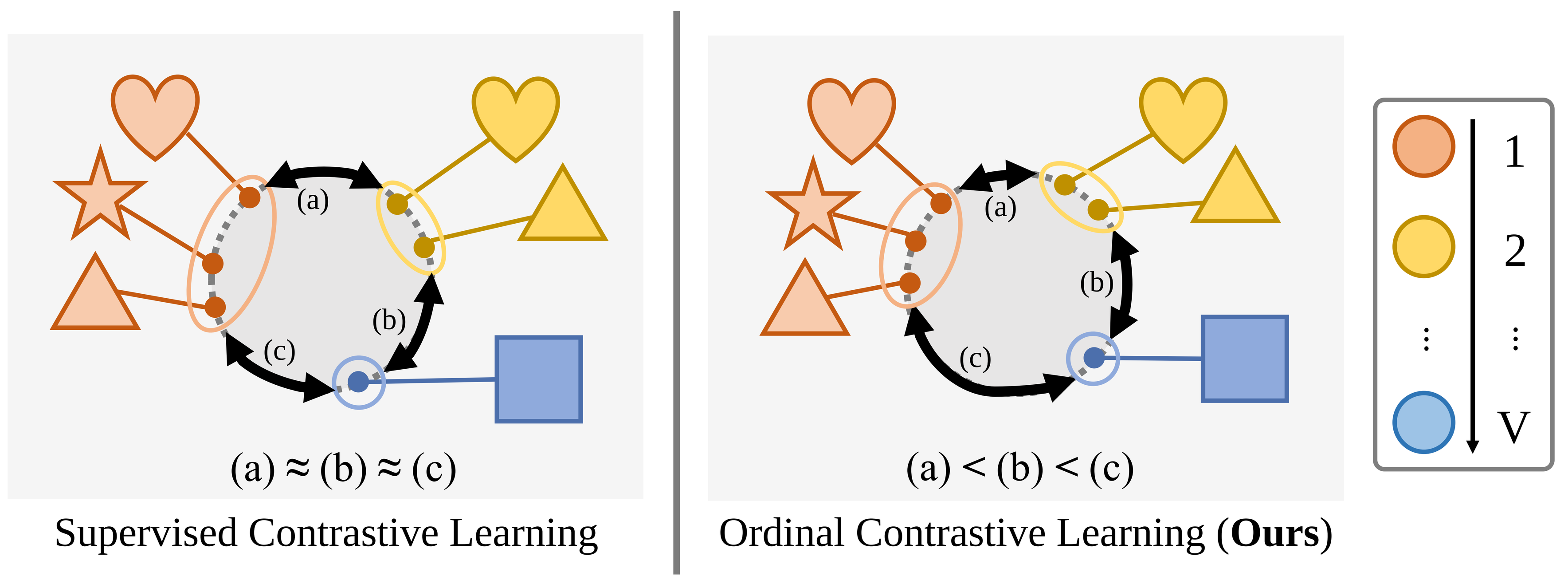


Figure: Comparison of supervised (left) and ordinal (right) contrastive learning: Both approaches contrast the set of all samples from the same class as positives against the negatives from the rest of the batch. While supervised contrastive learning repels each negative without differentiation on labels denoted as $(a) \approx (b) \approx (c)$, ordinal contrastive learning assigns the penalizing strength based on the label distance.

- In Supervised Contrastive Learning (SCL), single τ controls the strength of separation, ignoring the degree of differences between each label. The SCL loss is given by:

$$\mathcal{L}_{SC} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau)}{\sum\limits_{p \in P(i)} \exp(z_i \cdot z_p/\tau) + \sum\limits_{n \in N(i)} \exp(z_i \cdot z_n/\tau)} \tag{1}$$

- Considering that values of diagnostic label $y \in \{1, \cdots, V\}$ are aligned according to their severity (e.g., i-th subject is more severe than n-th subject if $y_i > y_n$), we define a function $d(y_i, y_n)$ measuring the distance between two labels as $|y_i - y_n|$.
- We make τ_i, n dependent on $y_{i,\cdot}$ and $y_{n,\cdot}$ as $\tau/d(i,n)$ to penalize greater label distance.
- To prevent the collapse or dispersion of the embedding space, the magnitude of gradient w.r.t positives and negatives should be the same. By the gradient analysis detailed in the supplementary material, $\tau_{i,P}$ between z_i and z_p is set as

$$\tau_{i,P} = \frac{\sum\limits_{n \in N(i)} \exp(z_{i,\cdot} \cdot z_{n,\cdot}/\tau_{i,n})}{\sum\limits_{n \in N(i)} \exp(z_{i,\cdot} \cdot z_{n,\cdot}/\tau_{i,n})/\tau_{i,n}}. \tag{2}$$

- By setting adaptive $\tau_{i,n}$ for each $z_{n,\cdot}$ and unique $\tau_{i,P}$ for every $z_{p,\cdot}$, we formulate our ordinal contrastive loss $\mathcal{L}_{OC}$ as

$$\mathcal{L}_{OC} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau_{i,P})}{\sum\limits_{q \in P(i)} \exp(z_i \cdot z_q/\tau_{i,P}) + \sum\limits_{n \in N(i)} \exp(z_i \cdot z_n/\tau_{i,n})}. \tag{3}$$

Embedding Visualization

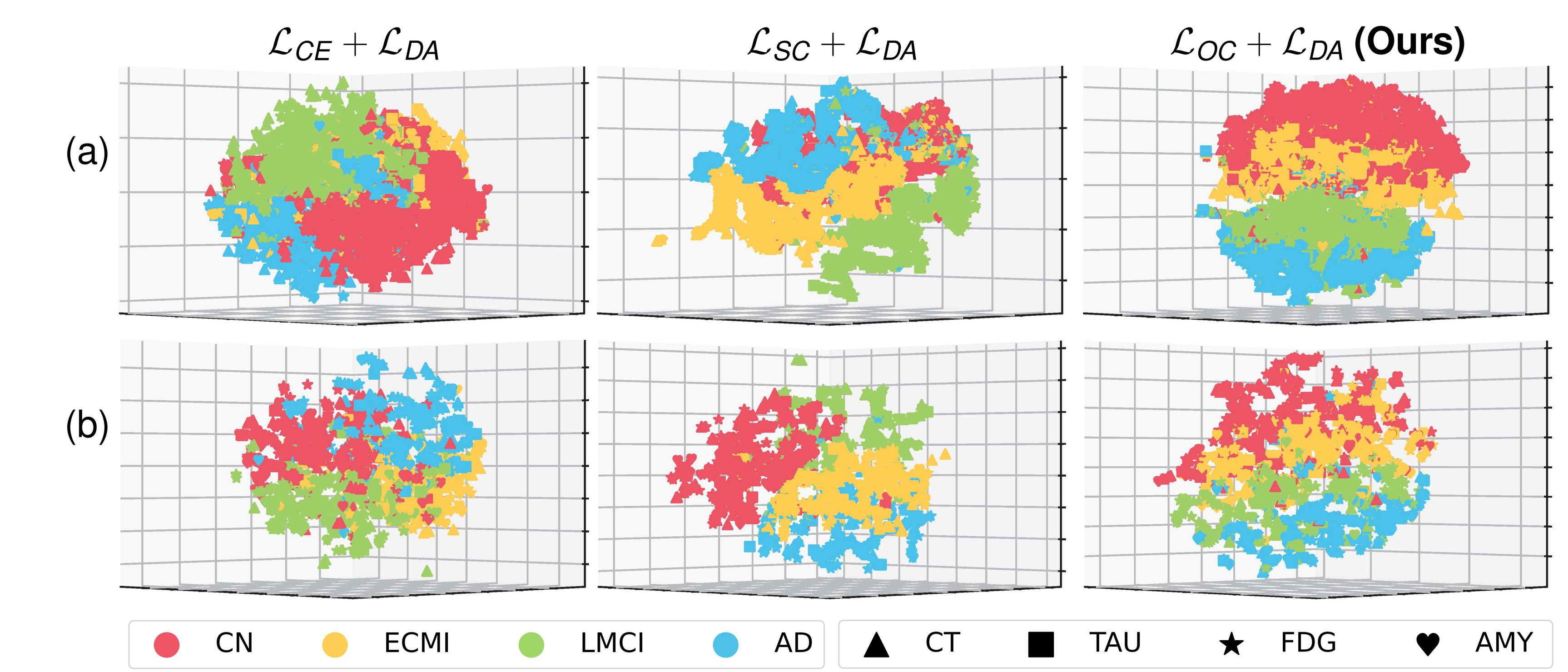


Figure: Visualizations of embeddings under each loss by t-SNE. Each individual encoder is trained with three distinct losses including Cross-Entropy $\mathcal{L}_{CE}$ (left), Supervised Contrastive Loss $\mathcal{L}_{SC}$ (center) and our Ordinal Contrastive Loss $\mathcal{L}_{OC}$ (right) along with domain adversarial loss $\mathcal{L}_{DA}$. (a) and (b) correspond to training and testing data respectively. (Color: AD-stage labels, Shape: imaging scan types.)

Acknowledgment

This research was supported by NRF-2022R1A2C2092336 (50%), RS-2022-II2202290 (20%), RS-2019-II191906 (AI Graduate Program at POSTECH, 10%) funded by MSIT, RS-2022-KH127855 (10%), RS-2022-KH128705 (10%) funded by MOHW from South Korea.

Imputation Framework

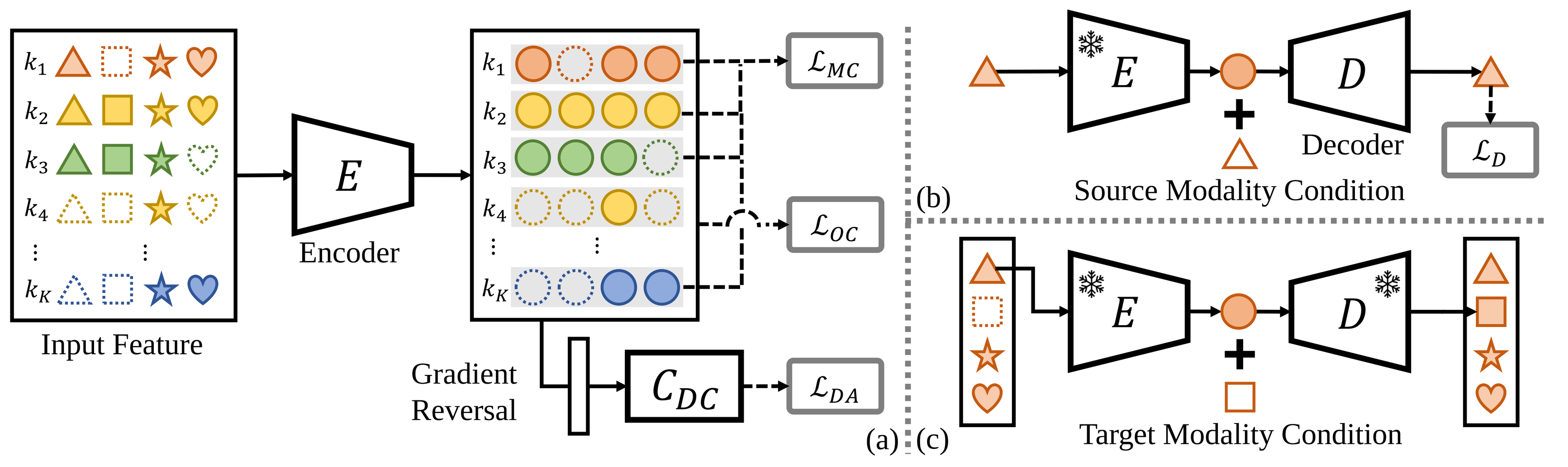


Figure: Overall scheme of our multi-scale learning network. Input X is transformed to a high-dimensional space with kernels $g(s)$ and Principal Components U (i.e., convolution) and fed to a downstream classifier (solid line). The S and classifier are trained to obtain the optimal task-specific multi-scale representation (dashed line).

- **Ordinal Contrastive Learning ($\mathcal{L}_{OC}$):** Train E to arrange each sample in the embedding space by the orders to accurately characterize disease progression.
- **Domain Adversarial Training ($\mathcal{L}_{DA}$):** Train E to eliminate modality-specific information associated with s from $z_{k,s} = E(x_{k,s})$. The modality adversarial loss $\mathcal{L}_{DA}$ is defined as

$$\mathcal{L}_{DA} = \mathcal{J}(s, C_{DC}(E(x_{k,s}))), \tag{4}$$

where $\mathcal{J}$ represents a suitable loss function (e.g., Cross-entropy).

- **Modality-wise coherence within a subject maximization ($\mathcal{L}_{MC}$):** Train E using a similarity function $sim(\cdot,\cdot)$ (e.g., cosine similarity) as

$$\mathcal{L}_{MC} = \frac{\sum_{k=1}^{K} \sum_{\substack{i,j \in \{1,\cdots,S\} \\ i \neq j}} -\delta_k(i,j) \cdot sim(x_{k,i}, x_{k,j})}{\sum_{k=1}^{K} \sum_{\substack{i,j \in \{1,\cdots,S\} \\ i \neq j}} \delta_k(i,j)} \tag{5}$$

where $\delta_k(i,j)$ is an indicator function defined as $\delta_k(i,j) = 1$ if both $x_{k,i}$ and $x_{k,j}$ exist for subject k, and $\delta_k(i,j) = 0$ otherwise.

$$\textbf{Encoder Loss: } \mathcal{L}_E = \mathcal{L}_{DA} + \mathcal{L}_{OC} + \mathcal{L}_{MC} \tag{6}$$

- Due to $\mathcal{L}_{DA}$ and $\mathcal{L}_{MC}$, the loss $\mathcal{L}_D$ for modality translation can be approximated as

$$\textbf{Decoder Loss: } \mathcal{L}_D(x_{k,t}) = ||x_{k,t} - D([E(x_{k,t}), c_t])||^2 \tag{7}$$

ADNI Dataset

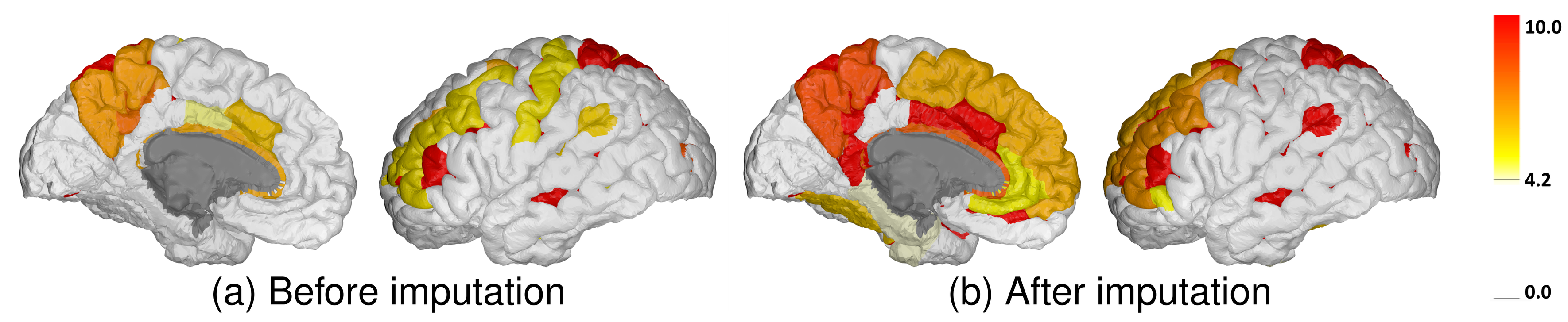
Table: Sample-size per modality of ADNI dataset.

Label	CT	TAU	FDG	AMY	Common
CN	844	237	861	735	123
EMCI	490	186	597	833	102
LMCI	250	105	1138	447	40
AD	240	85	755	422	10
Total	1824	613	3351	2437	275

- The Alzheimer's Disease Neuroimaging Initiative (ADNI) study provides magnetic resonance image (MRI) and positron emission tomography (PET).
- Images were partitioned into 148 cortical and 12 sub-cortical regions using Destrieux atlas.
- 4 AD-specific progressive groups: cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI) and Alzheimer's Disease (AD).

Experimental Results

Experiment 1: Group Comparisons



Modality	CN vs EMCI		EMCI vs LMCI		LMCI vs AD	
	(a)	(b)	(a)	(b)	(a)	(b)
Cortical Thickness	59	88 (57)	24	64 (20)	55	131 (55)
Tau	0	84 (0)	1	22 (1)	9	99 (9)
FDG	48	83 (44)	77	94 (75)	139	119 (119)
β-Amyloid	32	70 (27)	6	78 (6)	144	152 (144)

Figure: p-values from group comparisons with Bonferroni correction at $\alpha = 0.01$: (a) before imputation, (b) after imputation from our model. Top: Resutant p-value maps on a brain surface (left hemisphere) in a $-log_{10}$ from CN and EMCI comparison with cortical thickness, and (b) shows higher sensitivity. Bottom: Number of significant ROIs. Number of common ROIs before-and-after imputation are in ().

Experiment 2: Classifcation Performance

Table: Classification performance on ADNI data with all imaging features.

Classifier	MLP (2 layers)			MLP (4 layers)		
Method	Accuracy	Precision	Recall	Accuracy	Precision	Recall
No Imputation	0.673±0.030	0.659±0.025	0.673±0.030	0.698±0.048	0.707±0.047	0.698±0.048
Class-wise Mean	0.753±0.050	0.778±0.041	0.753±0.050	0.775±0.036	0.771±0.032	0.775±0.036
MICE	0.739±0.043	0.761±0.046	0.739±0.043	0.814±0.043	0.761±0.046	0.814±0.043
MissForest	0.721±0.061	0.753±0.080	0.721±0.061	0.832±0.025	0.844±0.024	0.832±0.025
Sinkhorn	0.776±0.044	0.799±0.041	0.776±0.044	0.829±0.033	0.847±0.041	0.829±0.033
GAIN	0.752±0.029	0.766±0.022	0.752±0.029	0.795±0.054	0.805±0.050	0.795±0.054
Pair-wise MLPs	0.756±0.030	0.782±0.036	0.756±0.030	0.782±0.062	0.799±0.060	0.782±0.062
SCL	0.813±0.042	0.812±0.051	0.813±0.042	0.845±0.020	0.851±0.038	0.845±0.020
Ours ($\mathcal{L}_{OC}$)	0.826±0.029	0.829±0.021	0.826±0.029	0.851±0.046	0.862±0.051	0.851±0.046
Ours ($\mathcal{L}_{OC} + \mathcal{L}_{MC}$)	**0.829±0.042**	**0.839±0.041**	**0.829±0.042**	**0.854±0.025**	**0.862±0.024**	**0.854±0.025**