Supplementary Material

Proof. The loss of ordinal contrasitve learning $\mathcal{L}_{OC}$ can be conceptualized as:

$$
\begin{aligned}
\mathcal{L}_{OC} &= \sum_{i\in I} \frac{-1}{|P(i)|} \sum_{p\in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau_{i,P})}{\sum\limits_{q\in P(i)} \exp(z_i \cdot z_q/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(z_i \cdot z_n/\tau_{i,n})} \\
&= \sum_{i\in I} \frac{-1}{|P(i)|} \sum_{p\in P(i)} \left\{ s_{i,p}/\tau_{i,P} - \log\left\{ \sum_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n}) \right\} \right\} \\
&= \sum_{i\in I} \frac{-1}{|P(i)|} \sum_{p\in P(i)} \mathcal{L}_{OC}^{i,p}
\end{aligned}
\tag{1}
$$

where $s_{i,j}$ denotes the inner product of two embeddings z_i and z_j, and $\mathcal{L}_{OC}^{i,p}$ represents the contrastive loss when the i-th sample serves as an anchor and the p-th sample is considered as a positive. Then, gradient of loss $\mathcal{L}_{OC}^{i,p}$ toward a positive sample z_p can be derived as:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_{OC}^{i,p}}{\partial z_p} &= z_i/\tau_{i,P} - \frac{\exp(s_{i,p}/\tau_{i,P}) \cdot z_i/\tau_{i,P}}{\sum\limits_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})} \\
&= \frac{z_i}{\tau_{i,P}} \left\{ \frac{\sum\limits_{q'\in P(i)\setminus\{p\}} \exp(s_{i,q'}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})}{\sum\limits_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})} \right\}
\end{aligned}
\tag{2}
$$

For the other positives, indexed as $q' \in P(i) \setminus \{p\}$, each of their gradients is:

$$
\frac{\partial \mathcal{L}_{OC}^{i,q'}}{\partial z_{q'}} = -\frac{\exp(s_{i,q'}/\tau_{i,P}) \cdot z_i/\tau_{i,P}}{\sum\limits_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})}
\tag{3}
$$

Likewise, gradient toward a negative sample $z_{n'}$ for $n' \in N(i)$ is denoted as:

$$
\frac{\partial \mathcal{L}_{OC}^{i,n'}}{\partial z_{n'}} = -\frac{\exp(s_{i,n'}/\tau_{i,n'}) \cdot z_i/\tau_{i,n'}}{\sum\limits_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})}
\tag{4}
$$

Then, the magnitude of gradient w.r.t positives and negatives are calculated as:

$$
\begin{aligned}
Grad_{pos} &= \frac{\partial \mathcal{L}_{OC}^{i,p}}{\partial z_p} + \sum_{q'\in P(i)\setminus\{p\}} \frac{\partial \mathcal{L}_{OC}^{i,q'}}{\partial z_{q'}} \\
&= \frac{z_i}{\tau_{i,P}} \left\{ \frac{\sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})}{\sum\limits_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})} \right\} \\
Grad_{neg} &= \sum_{n'\in N(i)} -\frac{\exp(s_{i,n'}/\tau_{i,n'}) \cdot z_i/\tau_{i,n'}}{\sum\limits_{q\in P(i)} \exp(s_{i,q}/\tau_{i,P}) + \sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n})}
\end{aligned}
\tag{5}
$$

To make the magnitude of $Grad_{pos}$ and $Grad_{neg}$ same, $\tau_{i,P}$ is determined as:

$$
\tau_{i,P} = \frac{\sum\limits_{n'\in N(i)} \exp(s_{i,n'}/\tau_{i,n'})}{\sum\limits_{n\in N(i)} \exp(s_{i,n}/\tau_{i,n}) \cdot 1/\tau_{i,n}}
\tag{6}
$$

Table 1: Detailed sample-size per each modality pair. Limited pair-wise sample-size demonstrates the need for our holistic translation model over one-to-one translations.

CT	✓	✗	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓	✗	✓
TAU	✗	✓	✗	✗	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓
FDG	✗	✗	✓	✗	✗	✓	✗	✓	✗	✓	✓	✗	✓	✓	✓
AMY	✗	✗	✗	✓	✗	✗	✓	✗	✓	✓	✗	✓	✓	✓	✓
# Sample	957	49	1626	478	31	3	177	25	36	954	4	53	324	140	275

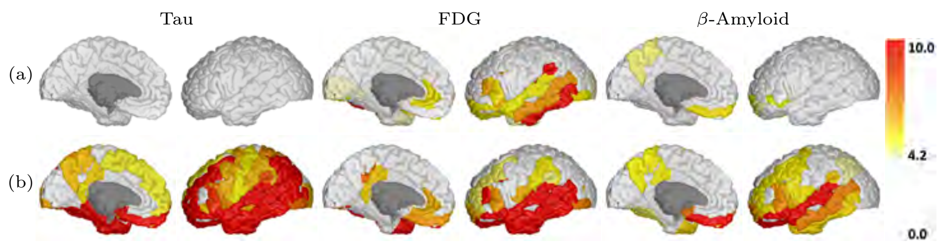

Fig. 1: p-values from group comparisons with Bonferroni correction at $\alpha = 0.01$: (a) before imputation, (b) after imputation from our model. Resutant p-value maps on a brain surface (left hemisphere) in a $-log_{10}$ from CN and EMCI comparison with Tau, FDG, and β-Amyloid. (b) shows higher sensitivity compared to (a).

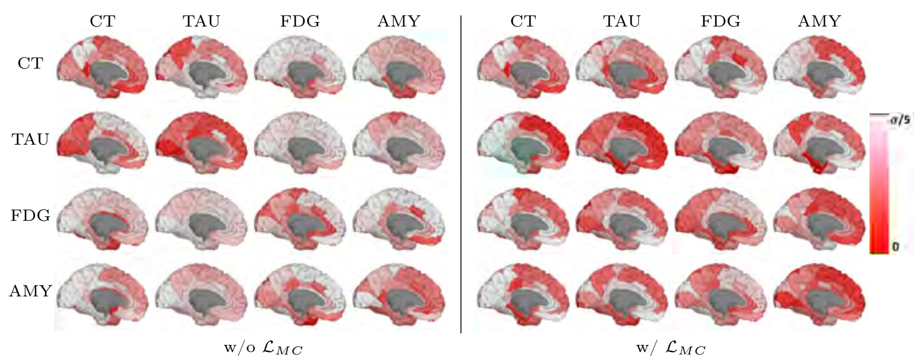

Fig. 2: Visualization of ROI-wise disparities between the real (target: Column) measure and the generated measure from each modality (source: Row) for the subject '009_S_1030', illustrating the impact of $\mathcal{L}_{MC}$. Each disparity is normalized with the ROI-wise mean and variance of the entire dataset. While self-reconstructions (diagonal entries) are consistently achieved regardless of the adoption of $\mathcal{L}_{MC}$, yielding more regions with small disparities (below $\sigma/5$) when adopting $\mathcal{L}_{MC}$ in translations (non-diagonal entries) suggests the effectiveness of maximizing the modality-wise coherence.