



T : distribution of transformations

t, t' : random transformations

f_{θ}, f'_{θ} : encoders

h_{ϕ}, h'_{ϕ} : intra-modality projector

g_{γ}, g'_{γ} : cross-modality projector

I : batch data

X, X' : batches augmented data

Y, Y' : batches of representations

Z, Z' : batches of embeddings intra-mod

W, W' : batches of embeddings cross-mod

$L_{i,i}$: intra-modality loss for modality (i)

$L_{i,j}$: cross-modality loss for modalities (i, j)

$L_{i,k}$: cross-modality loss for modalities (i, k)