CMIM: Cross-modal mutual information

Tristan Sylvain[1,2], Francis Dutil[3], Tess Berthier[3], Lisa Di Jorio[3], Margaux Luck[1], R Devon Hjelm[4], Yoshua Bengio[1,2]
[1]MILA [2]University of Montreal, [3]Imagia Cybernetics, [4]Microsoft Research

Motivations

Improving multi-modal learning.

- We showcase the efficiency of cross-modal representation learning.
- Different types of cross-modal and local/global losses are considered.

Improving a common clinical setup.

- Often, in medical datasets, different modalities are present at test and train times, and across patients.
- Our approach alleviates this issue by making the best of what modalities are present for training.

Other Methods

Open-I

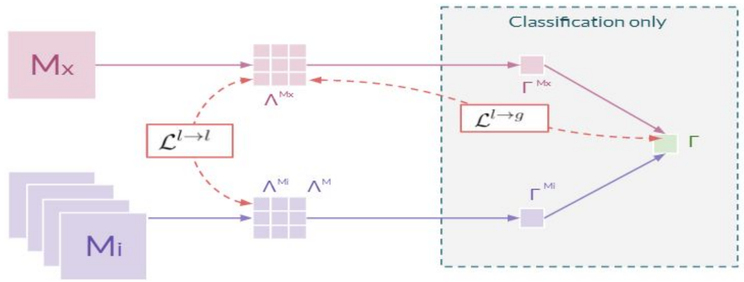
- Classical methods: Fully supervised ResNet, TieNet.
- We also experiment with adding synthesized text modalities at test-time.

BRATS

- HeMIS (previous SOTA from cross-modal training on BRATS), simpler mean and MLP baselines.
- At test time, a single modality is present.

Full Architecture

Model Diagram



- Optimization of different local/local, local/global, and cross-modal losses.
- Minimal architecture changes required for classification.
- We experimented with adding synthetic modalities at test-time for the classification task.

Mutual information estimation

- We rely on recent advances in neural mutual information estimation.
- We consider cross-modal losses: losses between representations of different modalities, and mutual information losses between combinations of different local and global representations of our inputs.

Results

Datasets Used

- Open-I (2D classification)
- BRATS 2015 (3D semantic segmentation)

Quantitative Results on Open-I

Method	Training I	Training T	Testing I	Testing T*	AUC
ResNet	•	○	•	○	0.785
TieNet	•	•	•	•	0.741
CMIM	•	•	•	○	**0.793**

- Our model outperforms the baselines.
- Contrary to the baselines, we actually benefit from the present at train-time of a dropped modality.

Quantitative Results on BRATS

Test-time modalities F	T_1	T_1c	T_2	CMIM	*HeMIS	*Mean (baseline)	*MLP (baseline)
•	○	○	○	**23.37**	5.57	6.25	15.90
○	•	○	○	**14.15**	4.67	6.25	10.78
○	○	•	○	49.00	**49.93**	30.02	32.92
○	○	○	•	**29.56**	20.31	6.25	18.62

- Our model outperforms the baselines.
- The approach is strongest for modalities with low predictive performance, which motivates the interest of the approach.