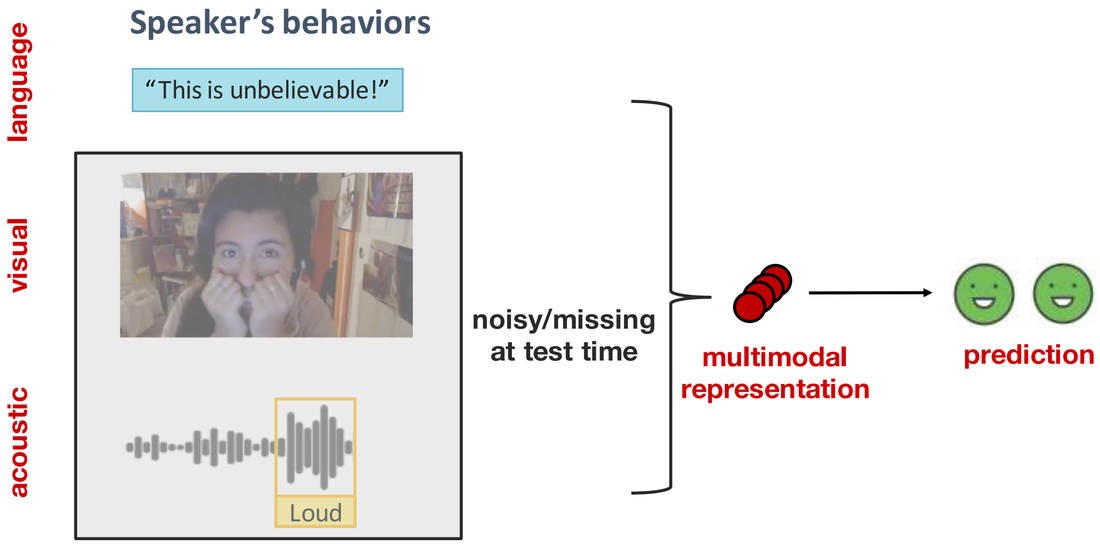
# Found in Translation:

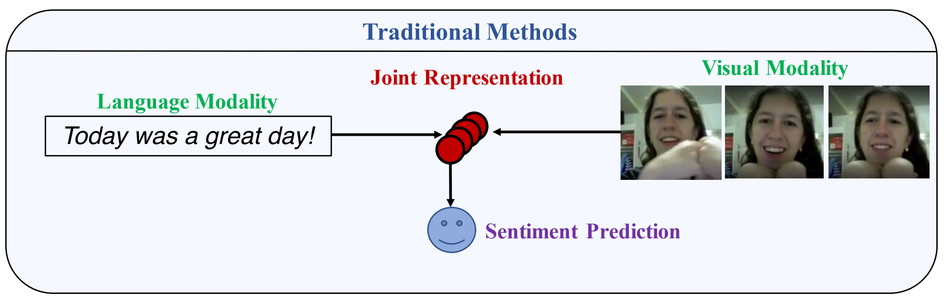
## Learning Robust Joint Representations by Cyclic Translations Between Modalities

Hai Pham*, Paul Pu Liang*, Thomas Manzini, Louis-Philippe Morency, Barnabás Póczos
(*equal contributions)
School of Computer Science, Carnegie Mellon University

## OVERVIEW

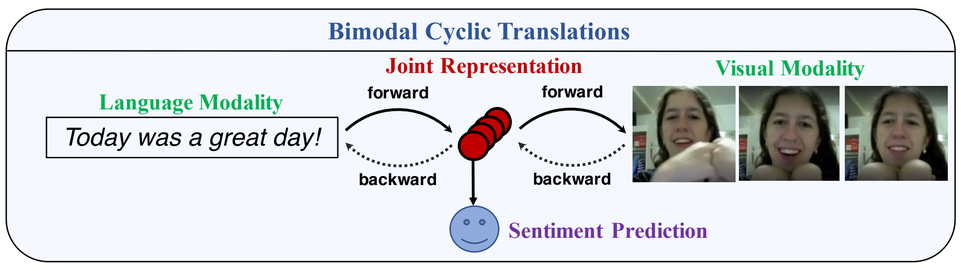



- Traditional approaches



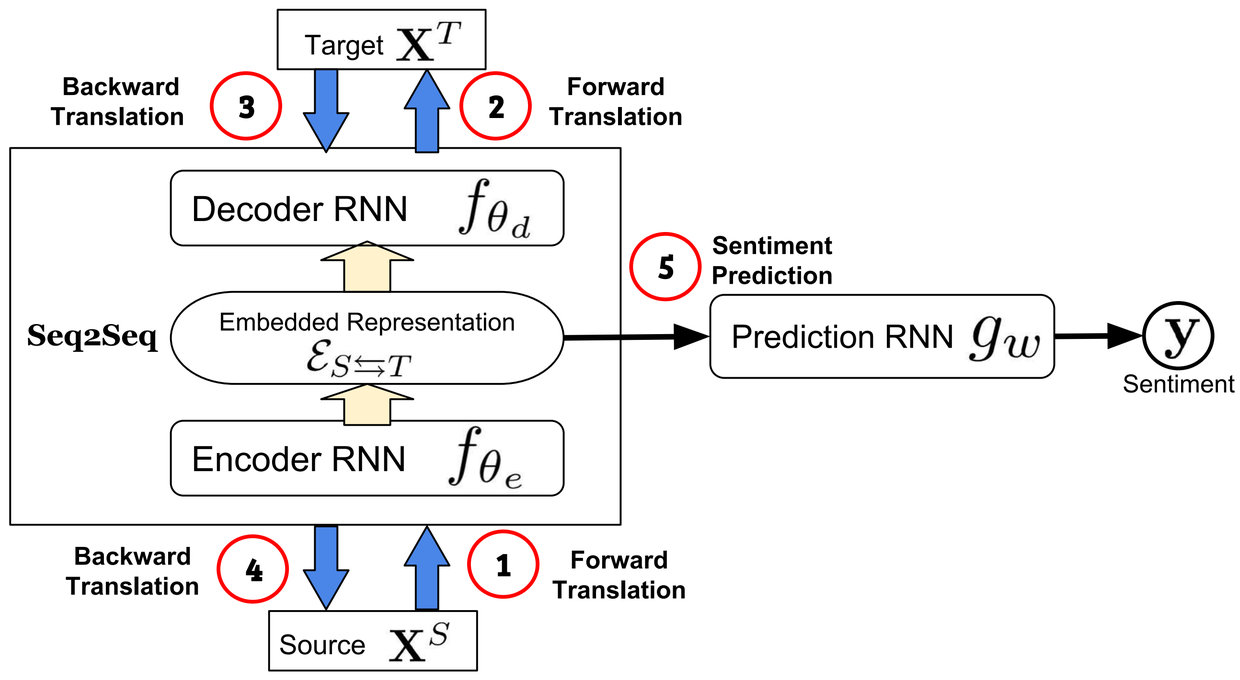

Both modalities required at test time!
Sensitive to missing/noisy visual modality.

- Our approach: Found in Translation




Only language modality required at test time!

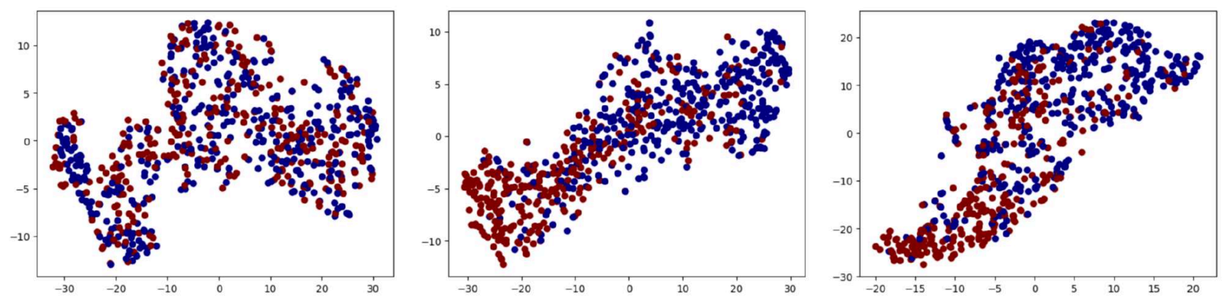
## MULTIMODAL CYCLIC TRANSLATION NETWORK (MCTN)




- Forward translation loss $\mathcal{L}_t = \mathbb{E}[\ell_{\mathbf{X}^T}(\hat{\mathbf{X}}^T, \mathbf{X}^T)]$
- Cycle consistent loss $\mathcal{L}_c = \mathbb{E}[\ell_{\mathbf{X}^S}(\hat{\mathbf{X}}^S, \mathbf{X}^S)]$
- Prediction loss $\mathcal{L}_p = \mathbb{E}[\ell_{\mathbf{y}}(\hat{\mathbf{y}}, \mathbf{y})]$

$$\mathcal{L} = \lambda_t \mathcal{L}_t + \lambda_c \mathcal{L}_c + \mathcal{L}_p$$

## EMBEDDED REPRESENTATION WITH t-SNE



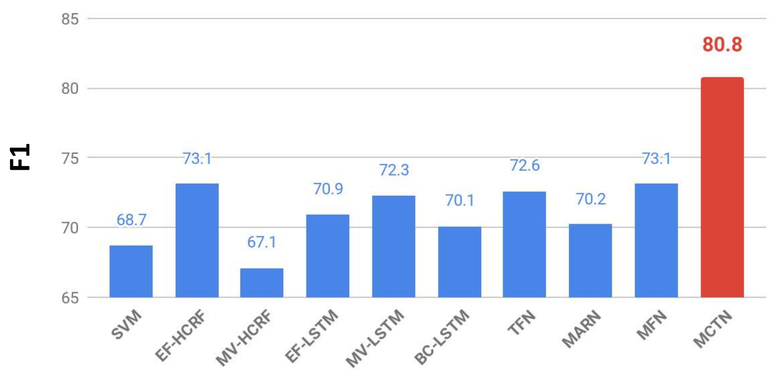
**MCTN:**
No cycle consistent loss
2 modalities

**MCTN:**
With cycle consistent loss
2 modalities

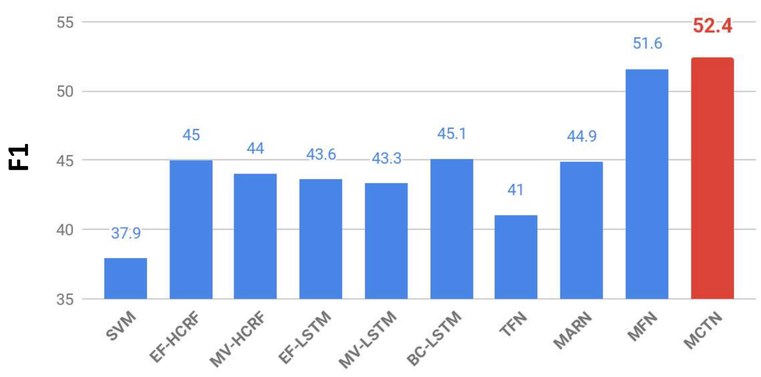
**MCTN:**
With cycle consistent loss
3 modalities

## STATE-OF-THE-ART PREDICTION RESULTS

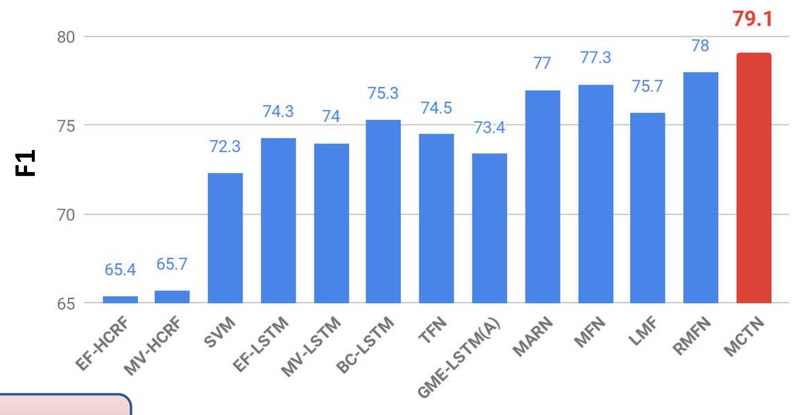
**ICT-MMMO dataset**




**YouTube dataset**

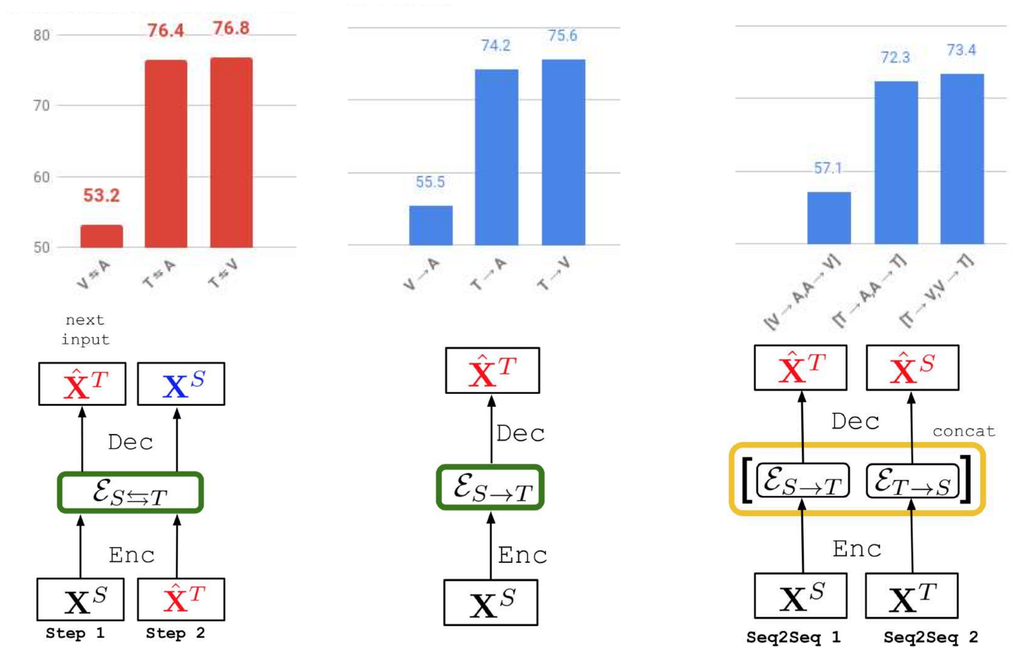



**CMU-MOSI dataset**




MCTN uses only language modality at test time!

## ABLATION STUDY




**MCTN:**
With cycle consistent loss
With parameter sharing
2 modalities

**MCTN:**
No cycle consistent loss
With parameter sharing
2 modalities

**MCTN:**
With cycle consistent loss
No parameter sharing
2 modalities

1. Use language as source modality
2. Use cyclic translations
3. Share parameters in seq2seq models

Code and Models:
http://github.com/hainow/MCTN