MACHINE LEARNING DEPARTMENT

Language Technologies Institute

Carnegie Mellon University

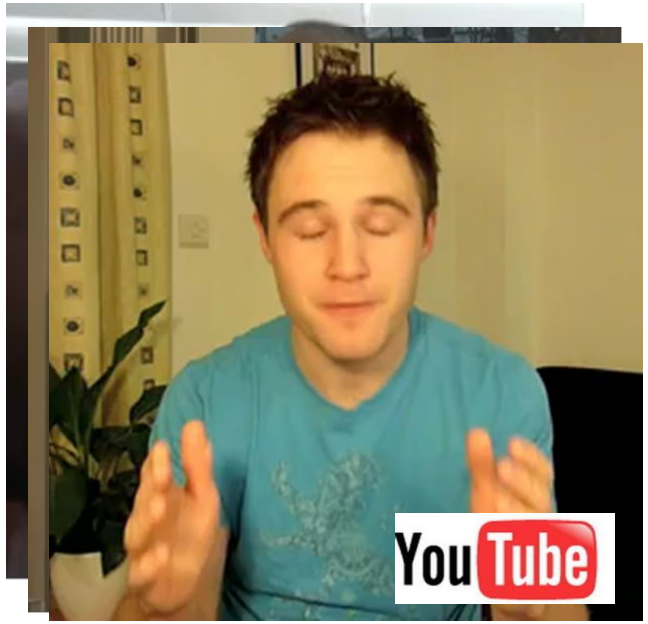
Found in Translation: Learning Robust Joint Representations by Cyclic Translations Between Modalities

Presenter: Hai Pham

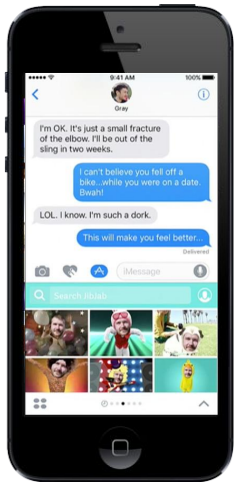
Hai Pham*, Paul Pu Liang*, Thomas Manzini, Louis-Philippe Morency, Barnabás Póczos

Progress of Artificial Intelligence

Multimedia Content



Intelligent
Personal Assistants



Robots and
Virtual Agents

Multimodal Language Modalities

Language

- Lexicon
- Syntax
- Pragmatics

Visual

- Gestures
- Body language
- Eye contact
- Facial expressions

Acoustic

- Prosody
- Vocal expressions

Multimodal Language Modalities

Language

- Lexicon
- Syntax
- Pragmatics

Visual

- Gestures
- Body language
- Eye contact
- Facial expressions

Acoustic

- Prosody
- Vocal expressions

Sentiment

- Positive
- Negative

Emotion

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise

Personality

- Confidence
- Persuasion
- Passion

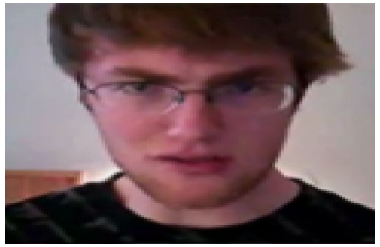
Challenge 1: Intra-modal Interactions

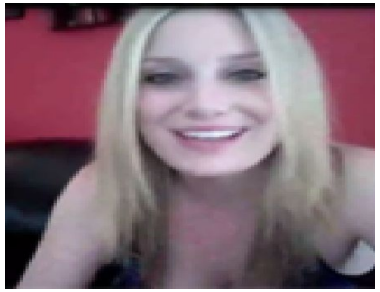
a) Temporal sequences

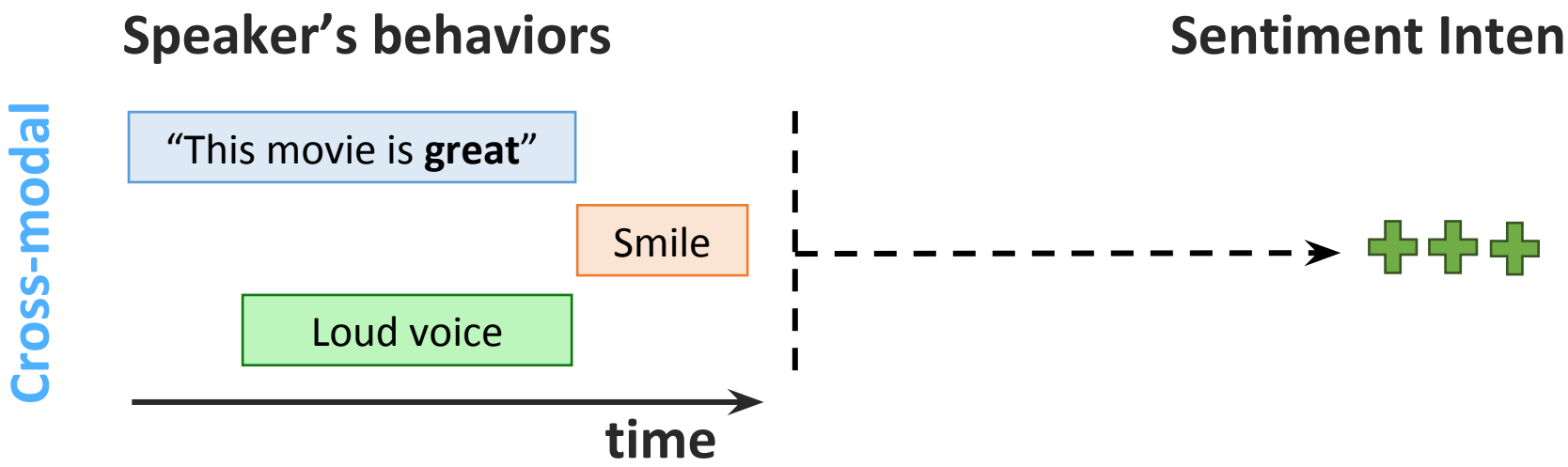
Challenge 2: Cross-modal Interactions

a) Multiple co-occurring interactions

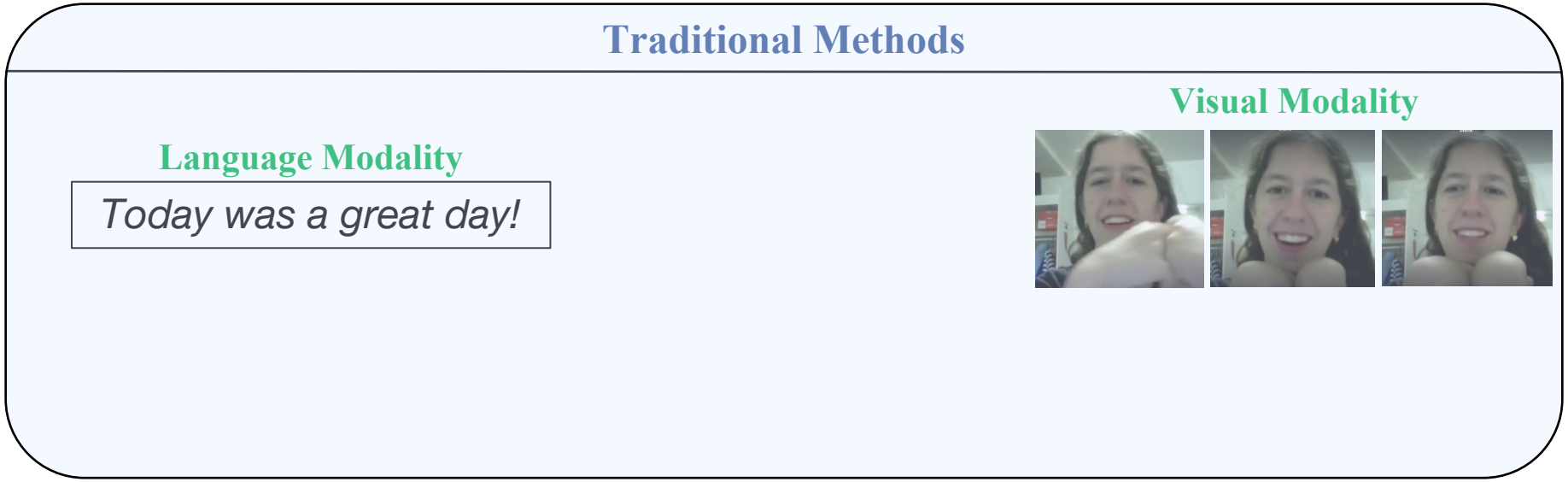
b) Different weighted combinations



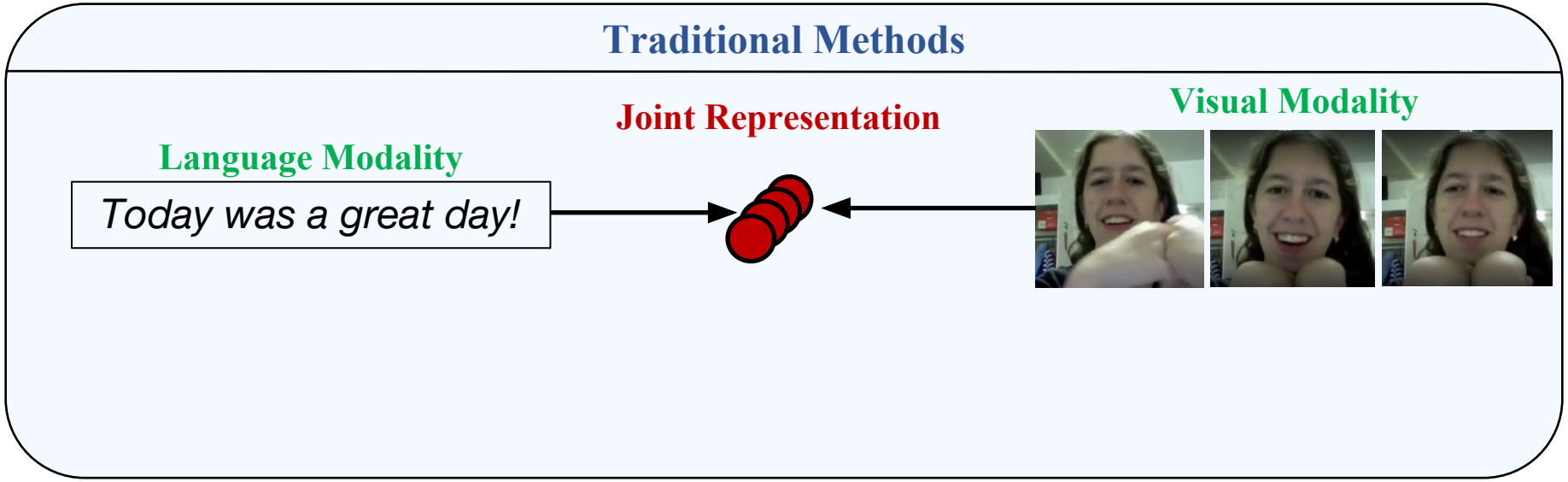




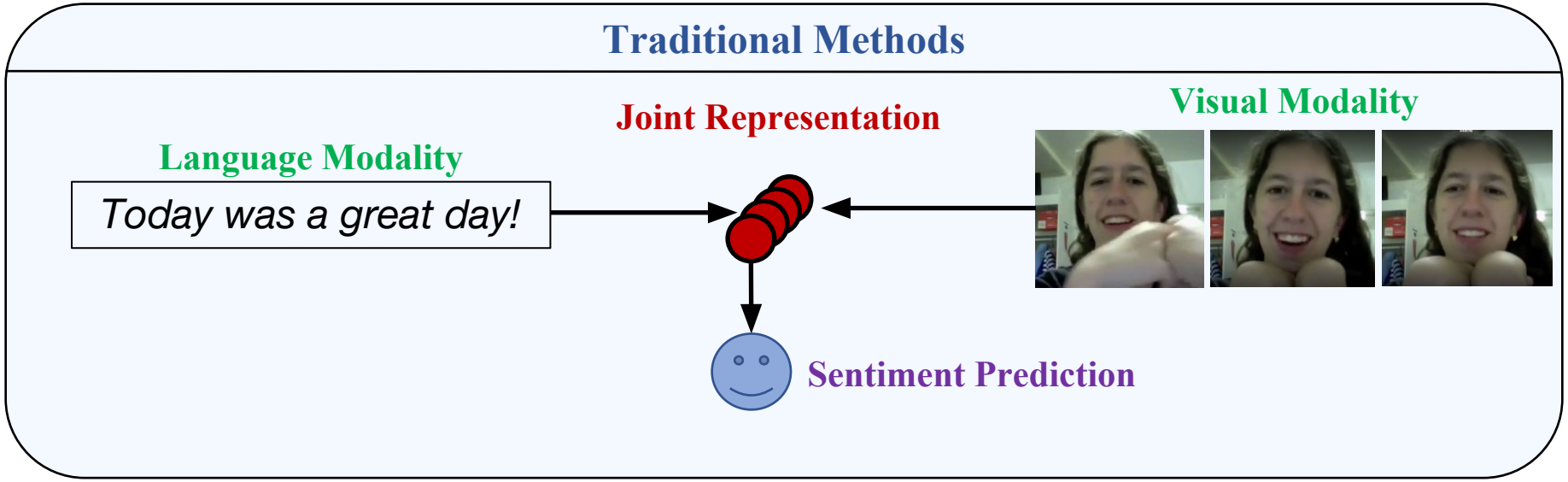
Learning Joint Representations: 2 modalities



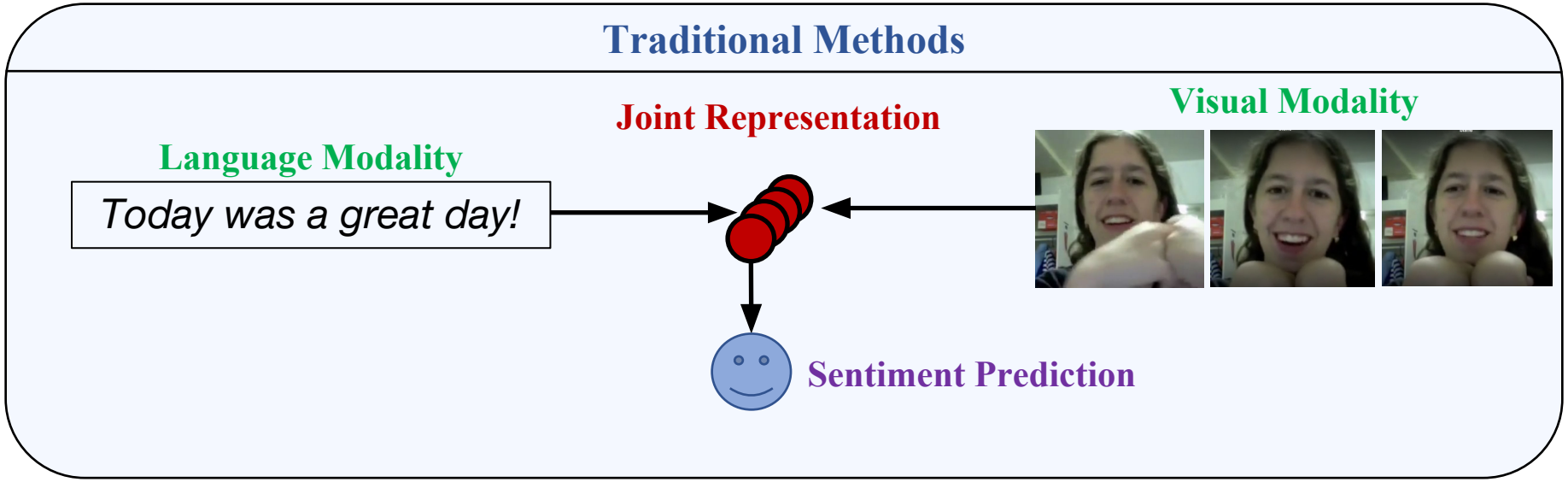
Learning Joint Representations: 2 modalities



Learning Joint Representations: 2 modalities

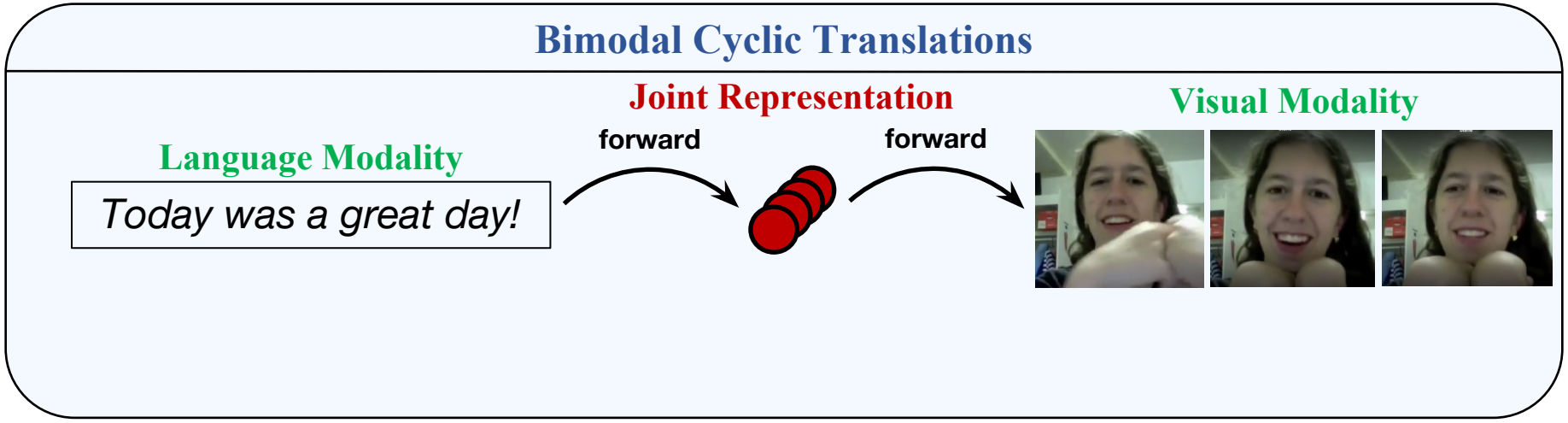


Learning Joint Representations: 2 modalities

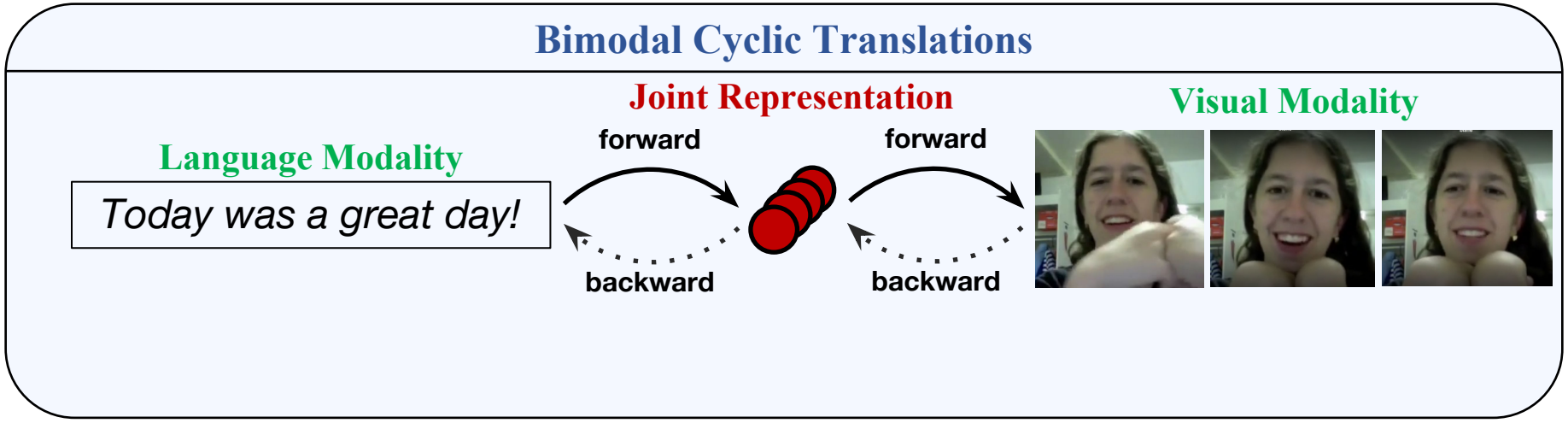


Both modalities required at test time!
Sensitive to missing/noisy visual modality.

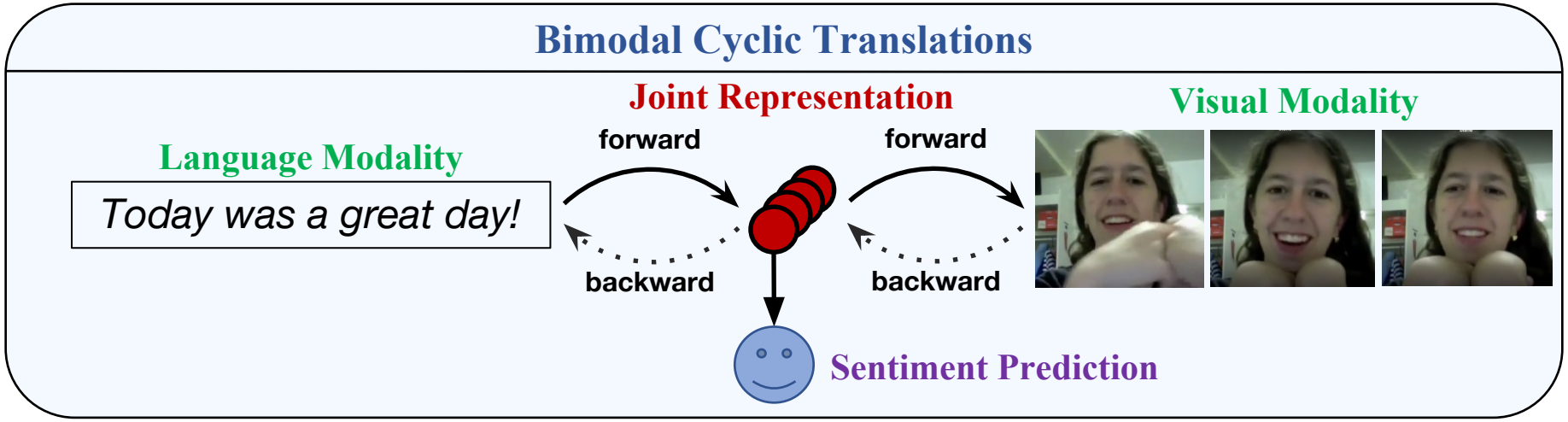
Learning Robust Joint Representations: 2 modalities



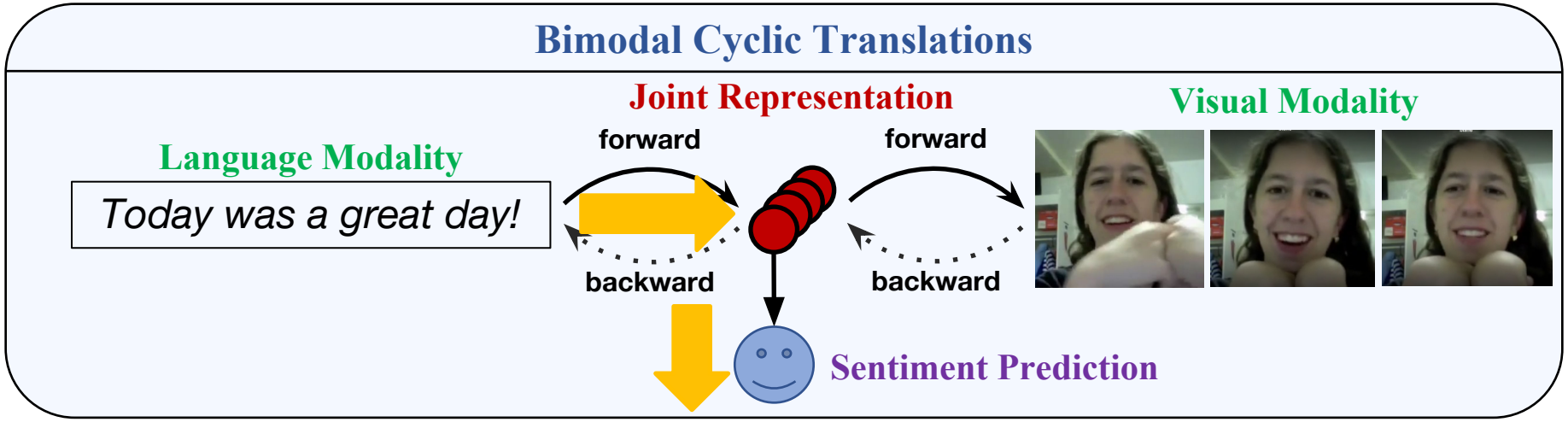
Learning Robust Joint Representations: 2 modalities



Learning Robust Joint Representations: 2 modalities

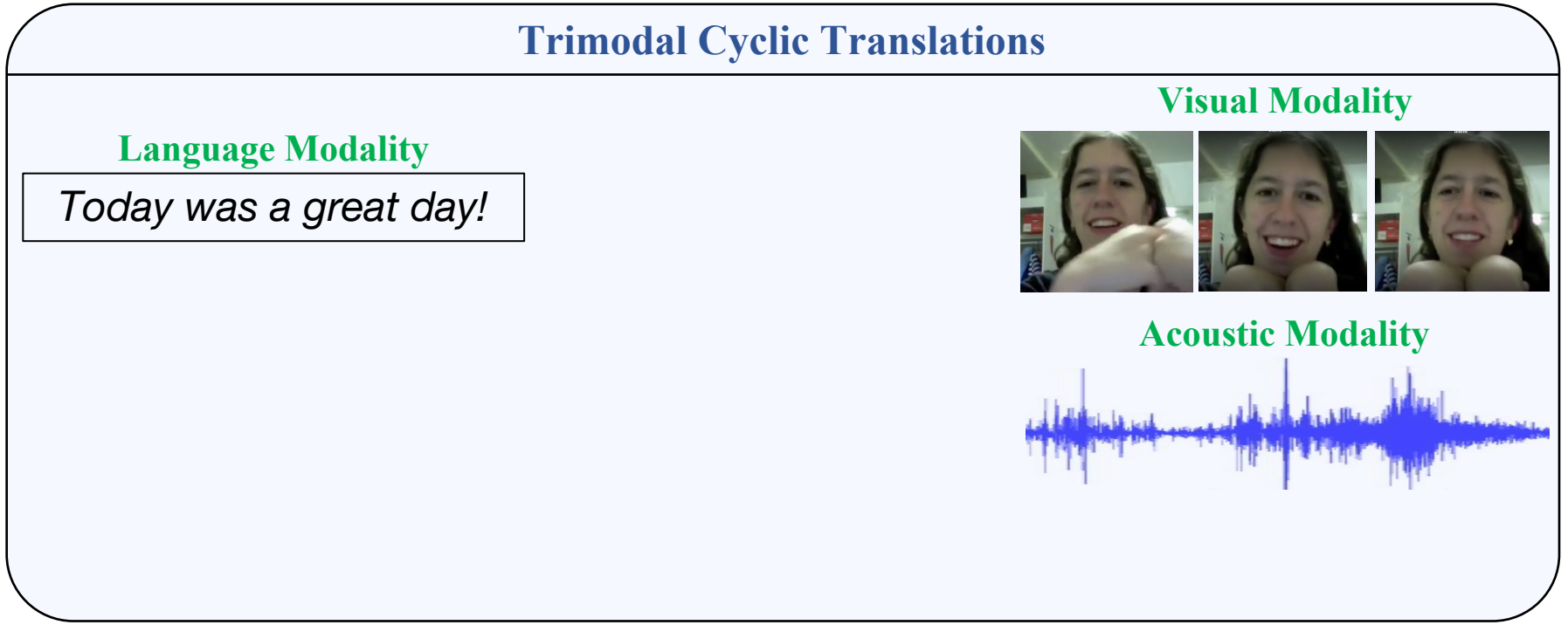


Learning Robust Joint Representations: 2 modalities

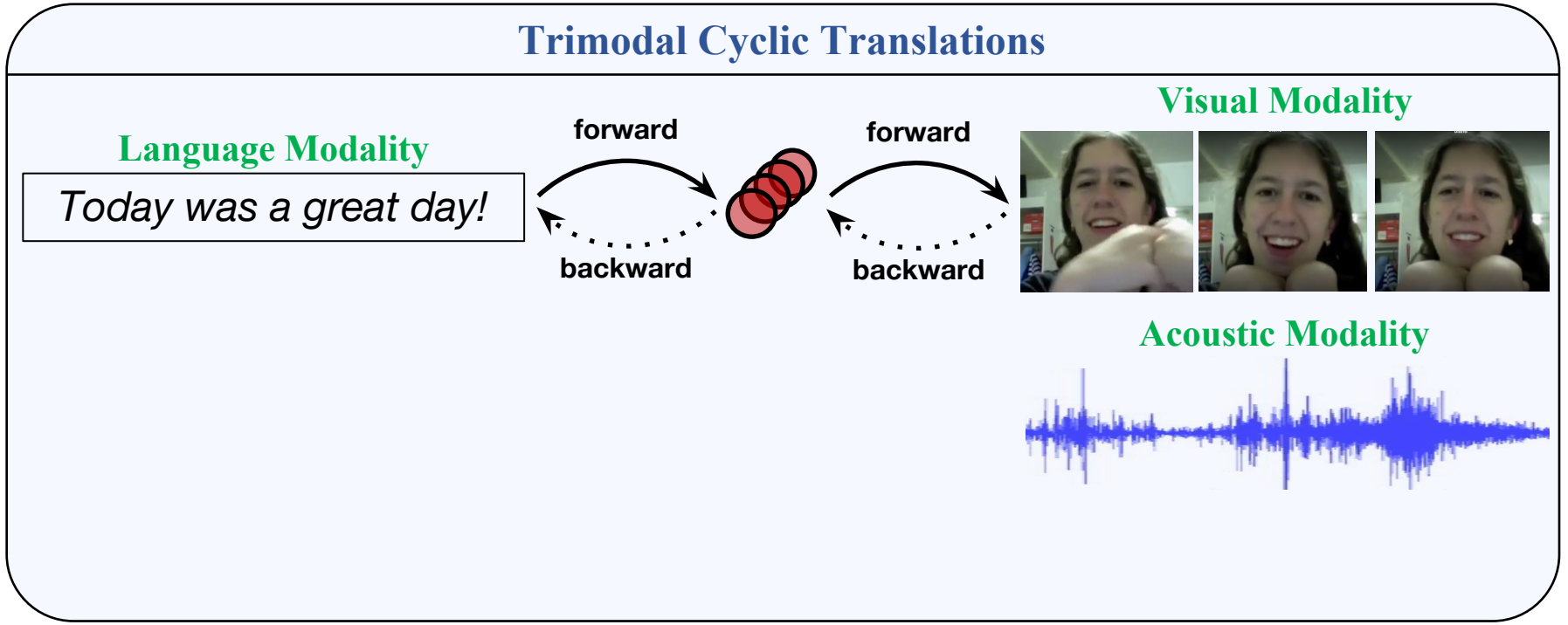


Only language modality required at test time!

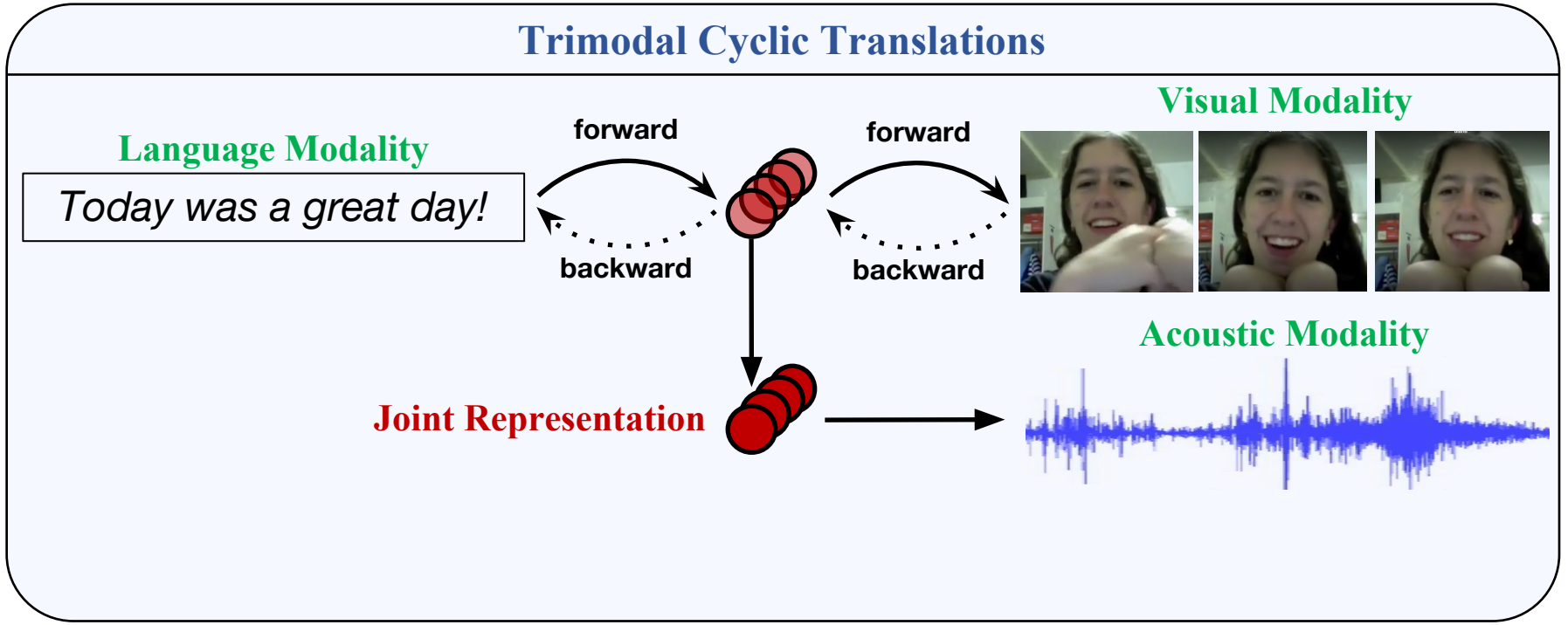
Learning Robust Joint Representations: 3 modalities



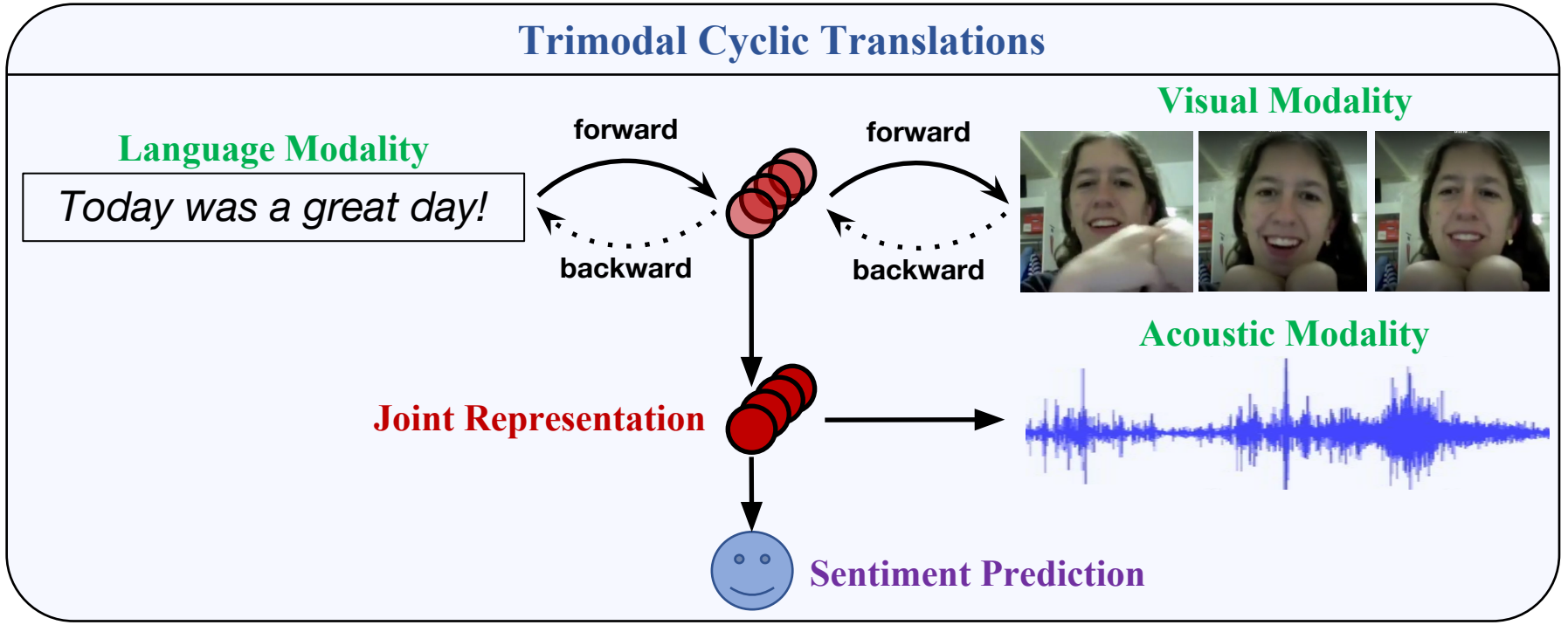
Learning Robust Joint Representations: 3 modalities



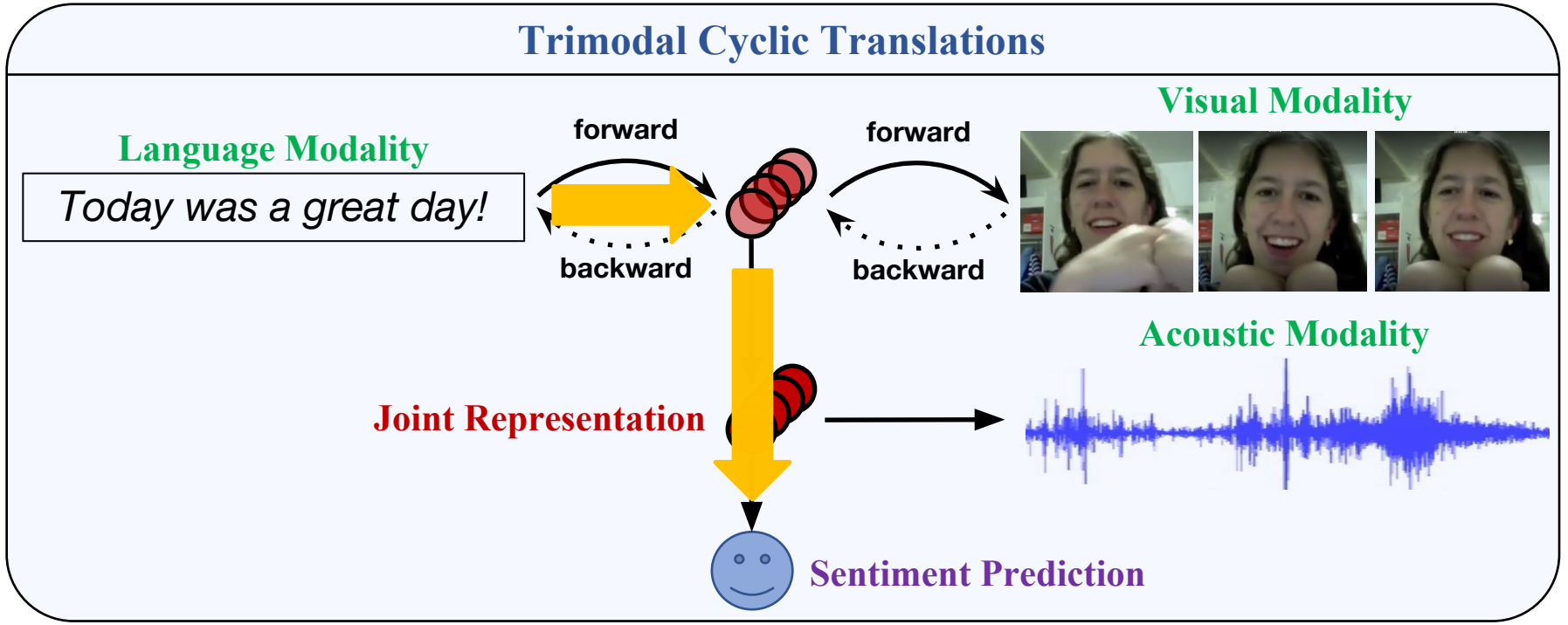
Learning Robust Joint Representations: 3 modalities



Learning Robust Joint Representations: 3 modalities

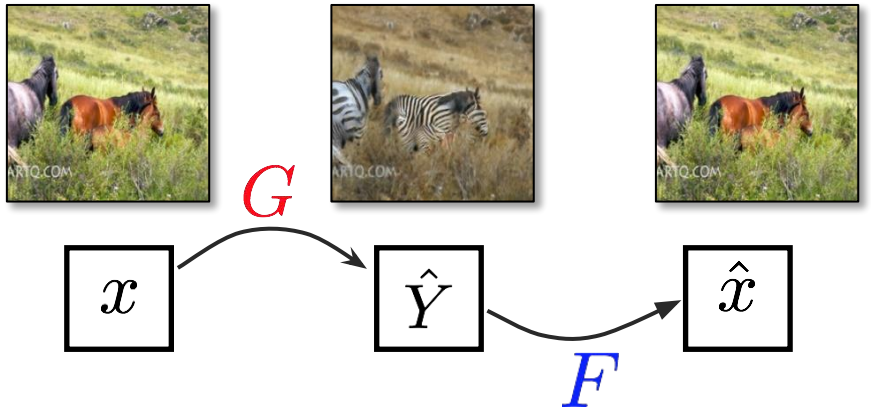


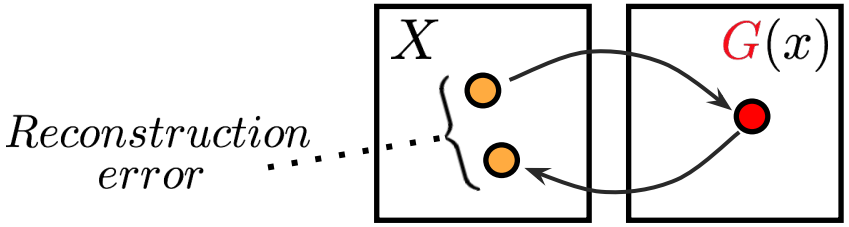
Learning Robust Joint Representations: 3 modalities



Only language modality required at test time!

Cyclic Translations





$$\|F(G(x)) - x\|_1$$

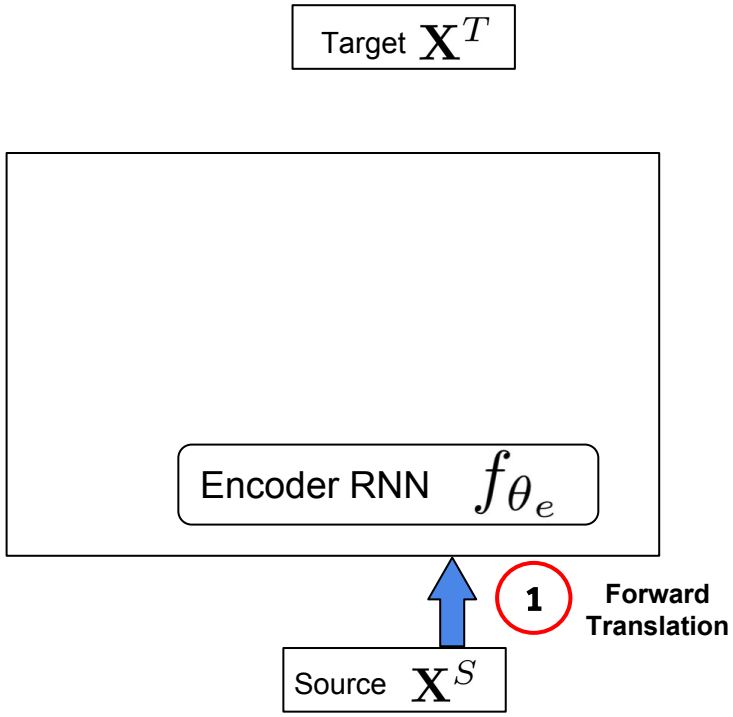
[Zhu*, Park*, Isola, and Efros, ICCV 2017]

Multimodal Cyclic Translation Network

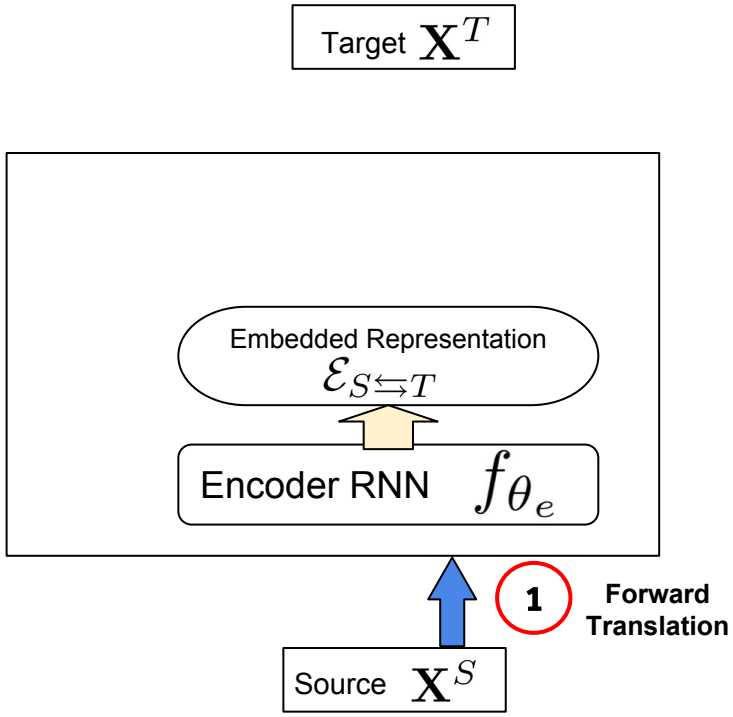
Target $\mathbf{X}^T$

Source $\mathbf{X}^S$

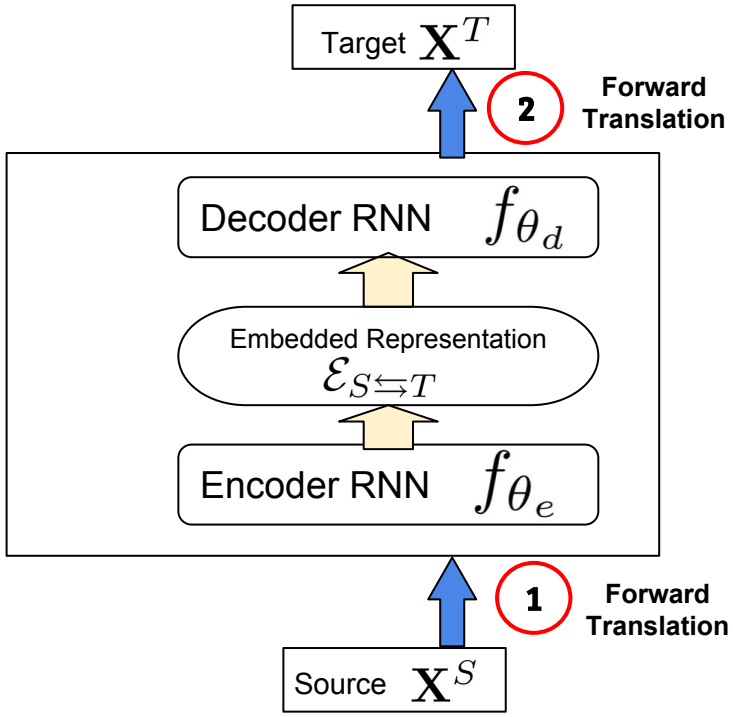
Multimodal Cyclic Translation Network



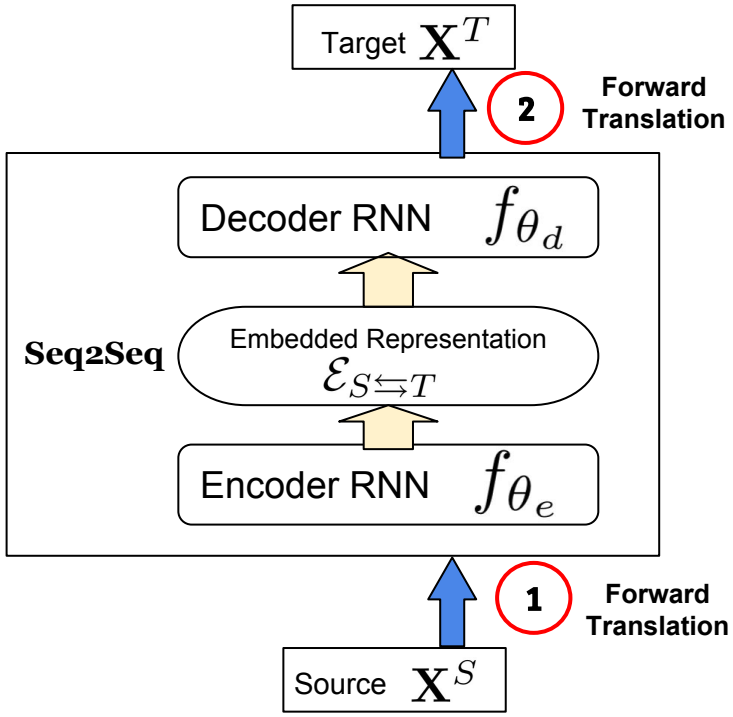
Multimodal Cyclic Translation Network



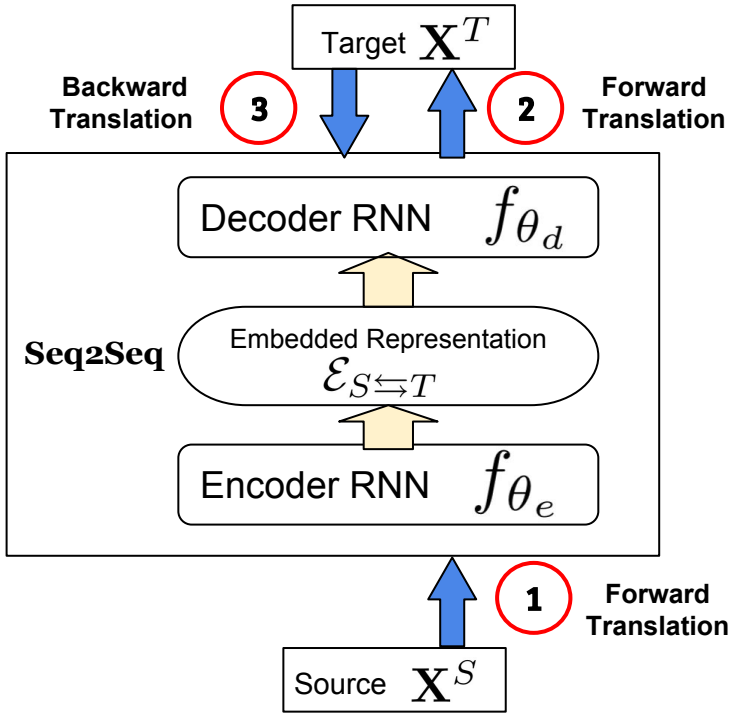
Multimodal Cyclic Translation Network



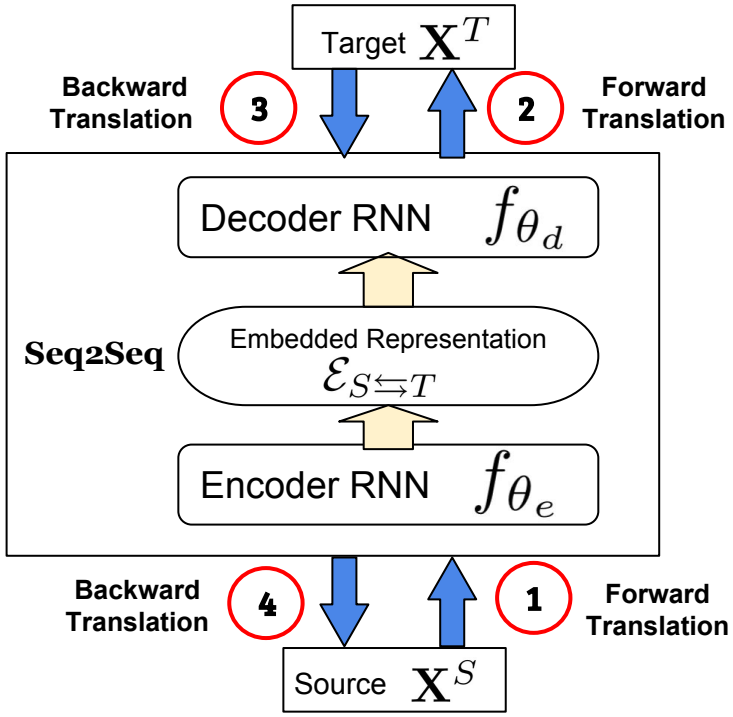
Multimodal Cyclic Translation Network



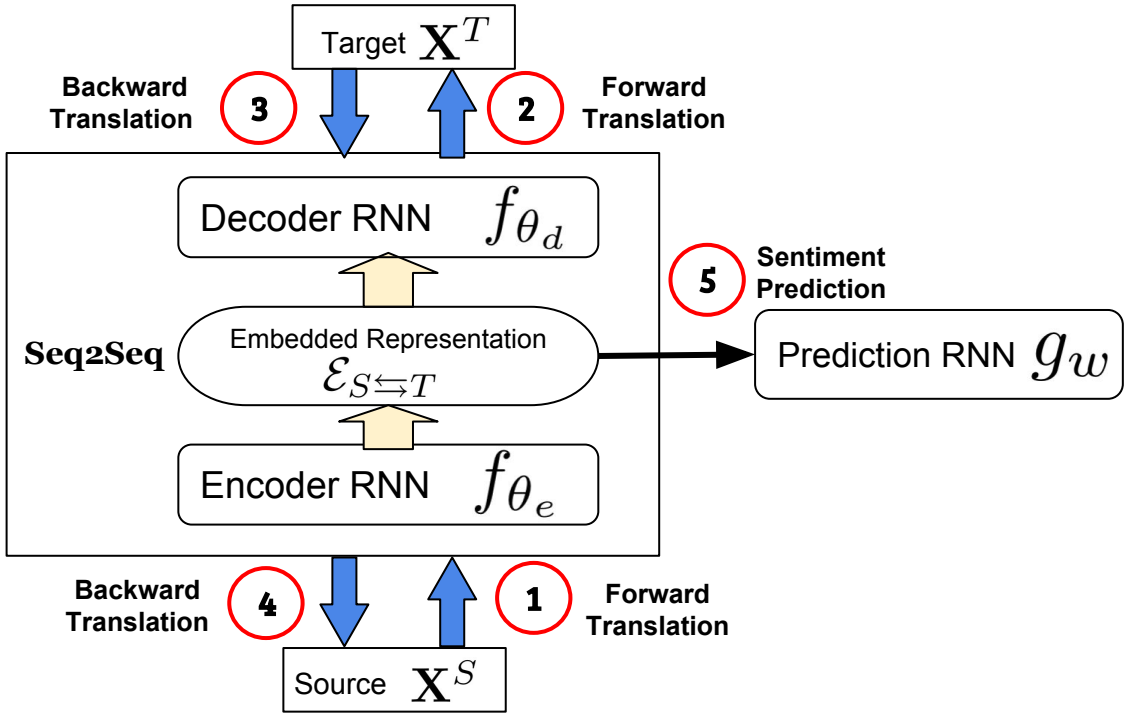
Multimodal Cyclic Translation Network



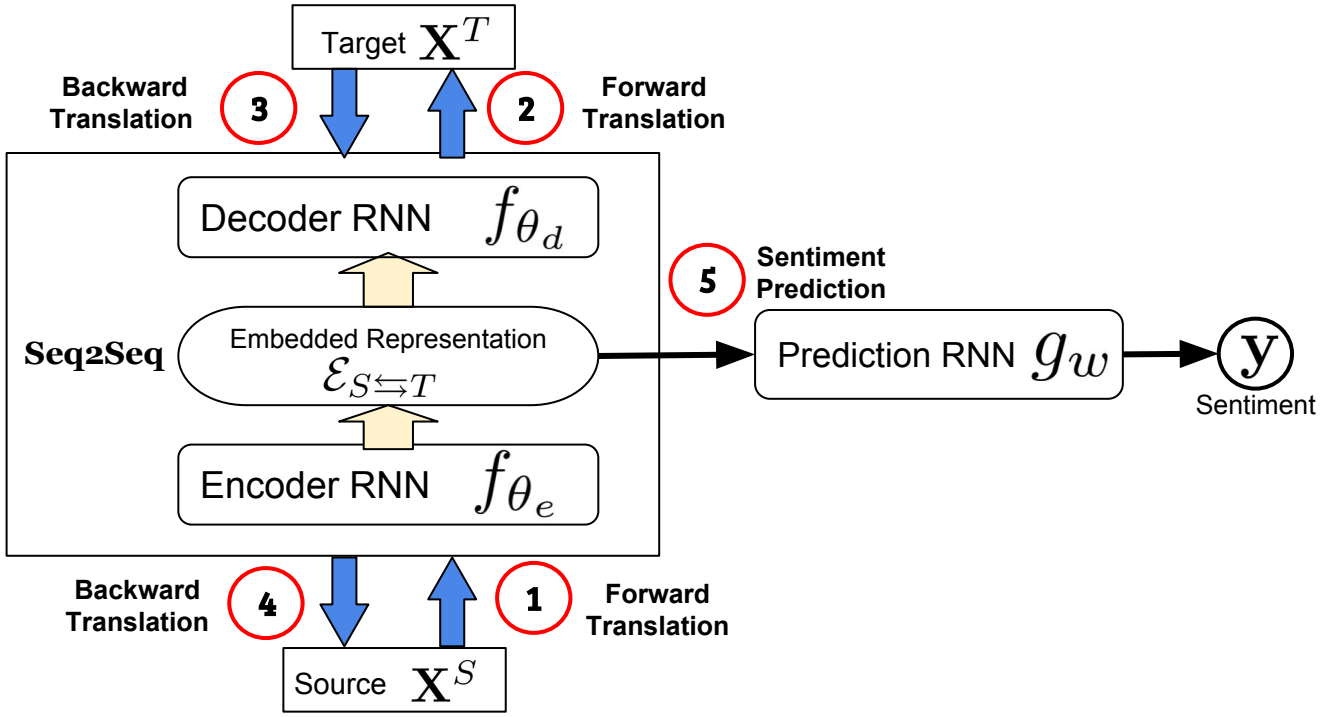
Multimodal Cyclic Translation Network



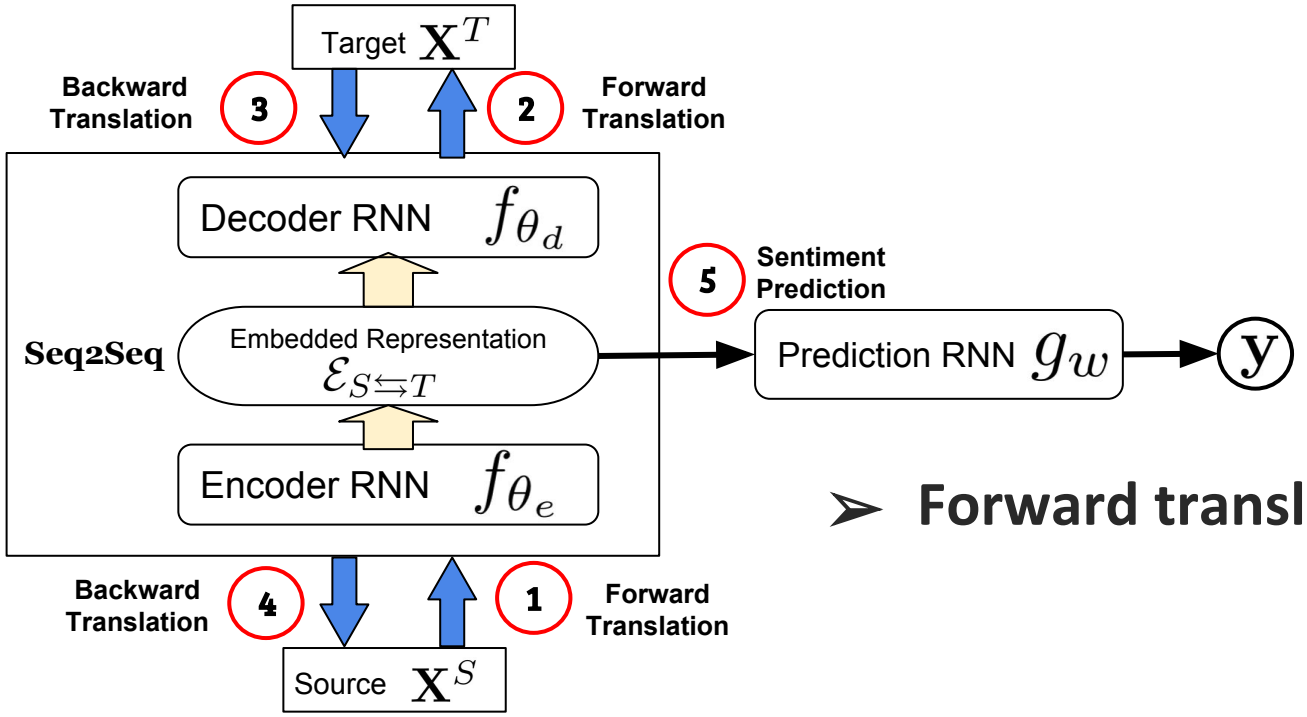
Multimodal Cyclic Translation Network



Multimodal Cyclic Translation Network

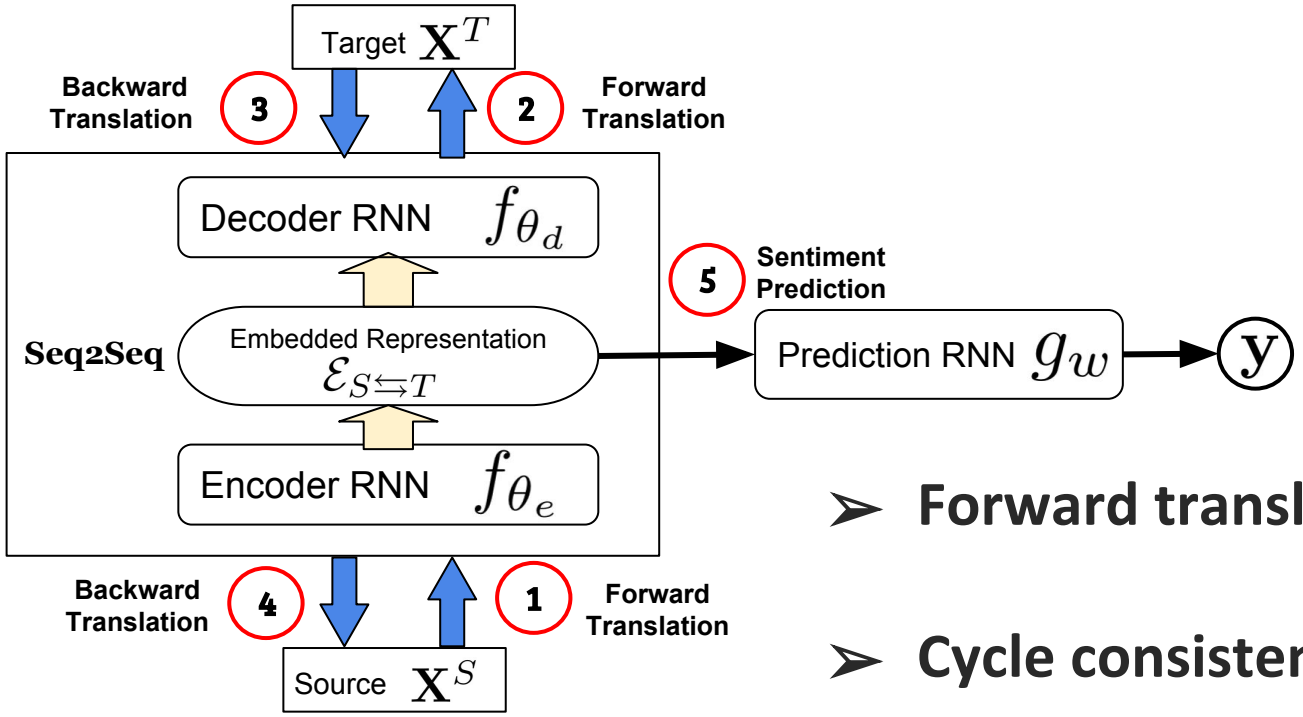


Coupled Translation-Prediction Objective



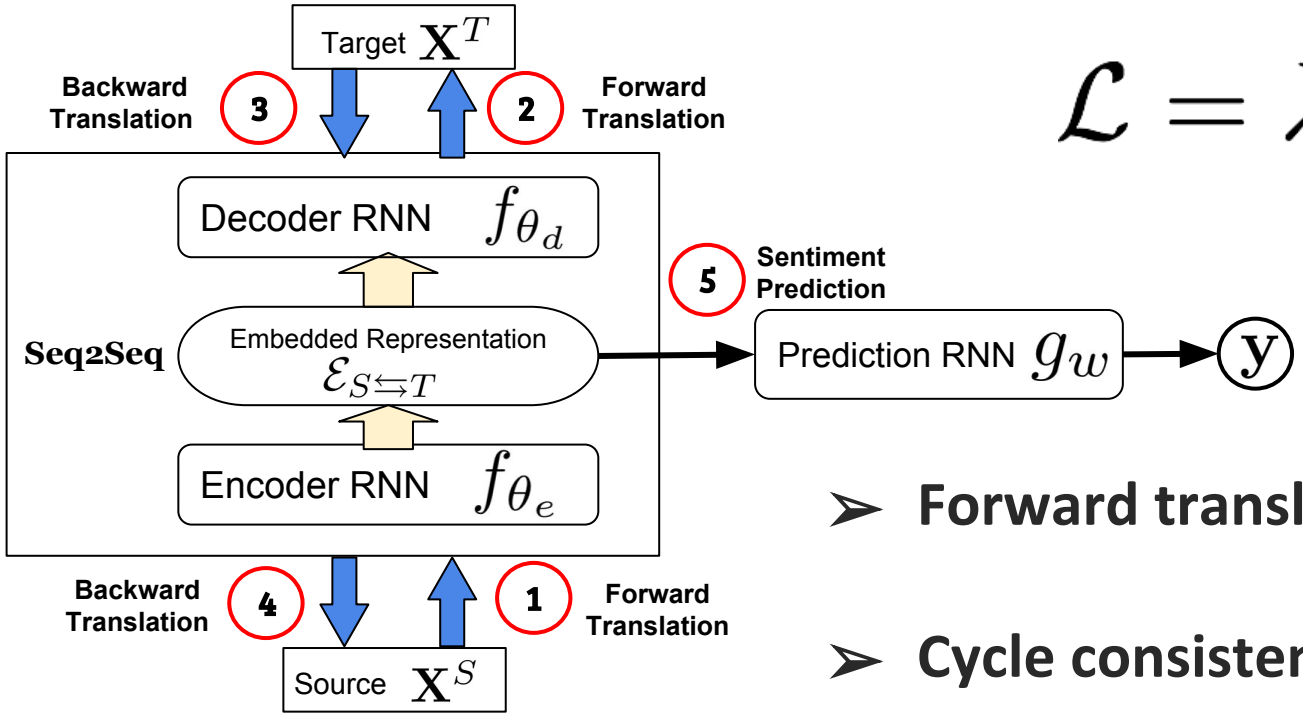
- **Forward translation loss** $\mathcal{L}_t = \mathbb{E}[\ell_{\mathbf{X}^T}(\hat{\mathbf{X}}^T, \mathbf{X}^T)]$

Coupled Translation-Prediction Objective



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

Coupled Translation-Prediction Objective

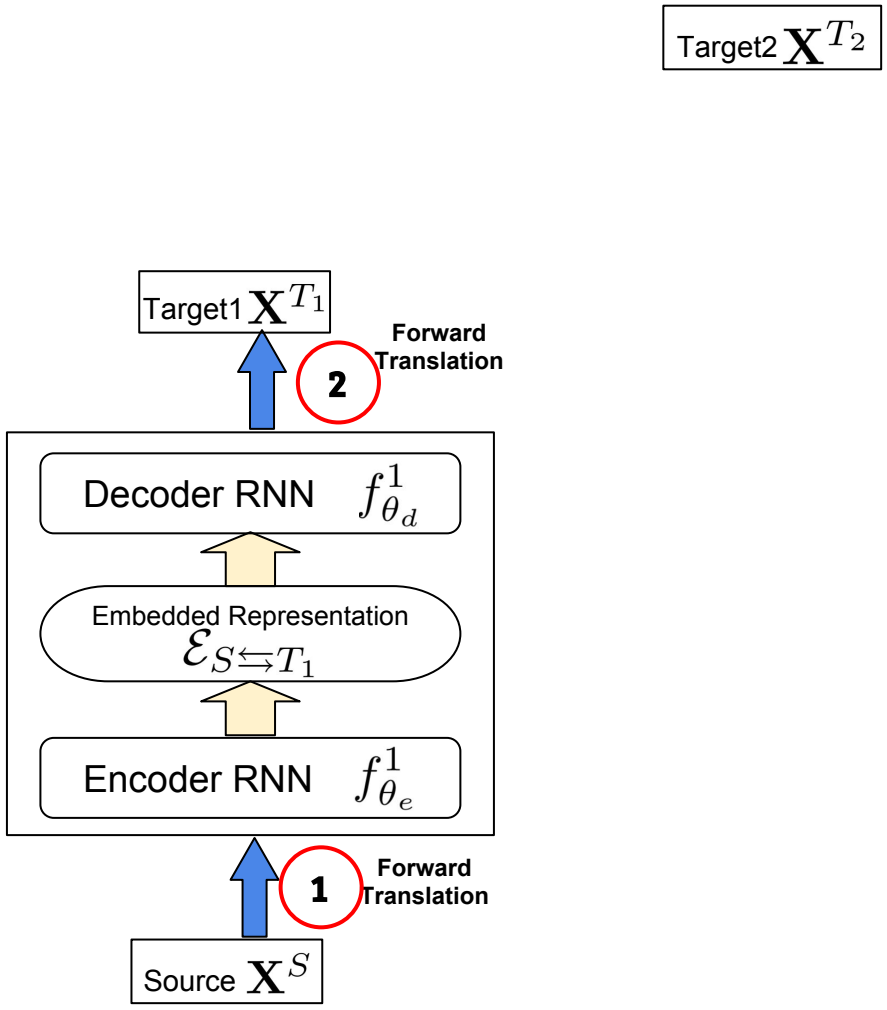


$$\mathcal{L} = \lambda_t \mathcal{L}_t + \lambda_c \mathcal{L}_c + \mathcal{L}_p$$

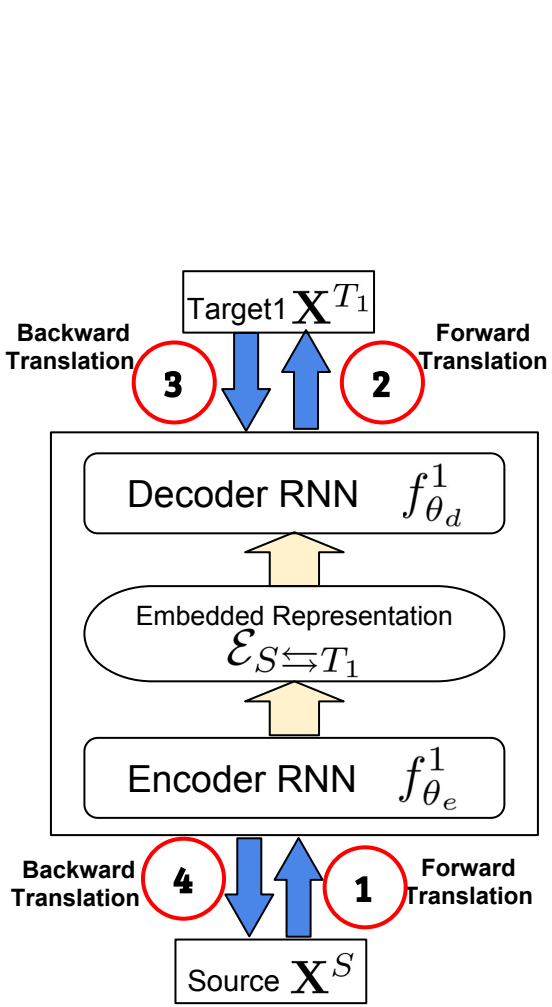
- **Forward translation loss** $\mathcal{L}_t = \mathbb{E}[\ell_{\mathbf{X}^T}(\hat{\mathbf{X}}^T, \mathbf{X}^T)]$
- **Cycle consistent loss** $\mathcal{L}_c = \mathbb{E}[\ell_{\mathbf{X}^S}(\hat{\mathbf{X}}^S, \mathbf{X}^S)]$
- **Prediction loss** $\mathcal{L}_p = \mathbb{E}[\ell_{\mathbf{y}}(\hat{\mathbf{y}},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

Hierarchical Multimodal Cyclic Translation Network

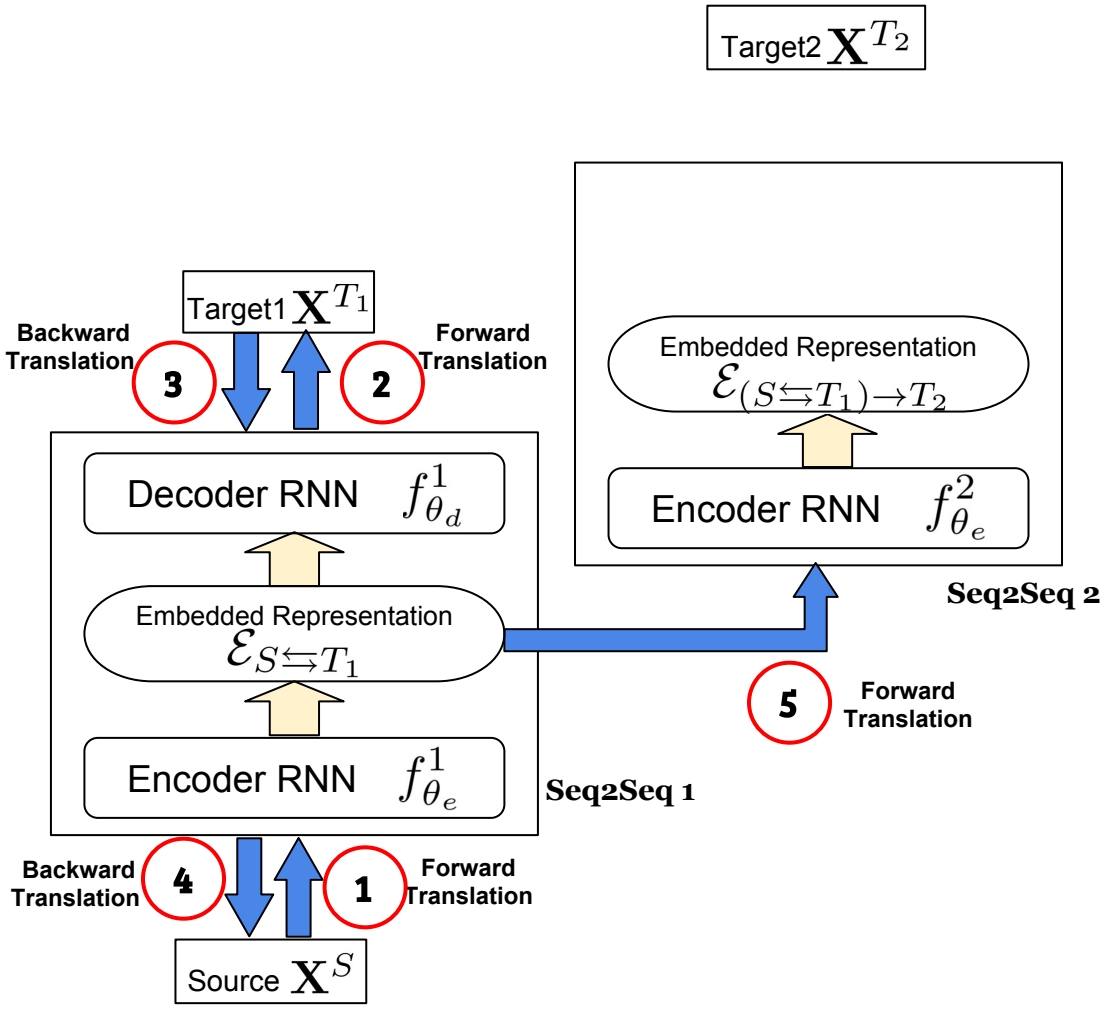
Hierarchical Multimodal Cyclic Translation Network



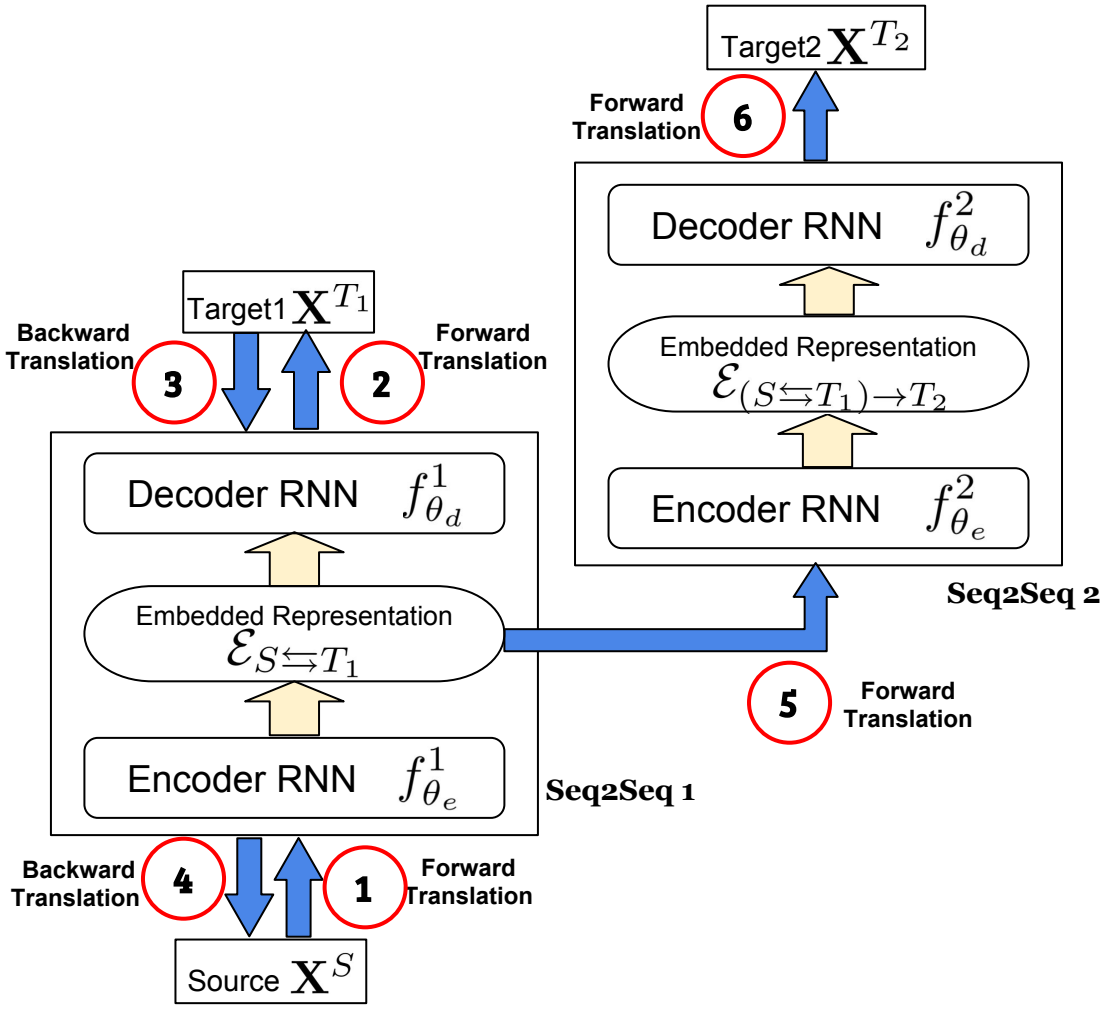
Hierarchical Multimodal Cyclic Translation Network



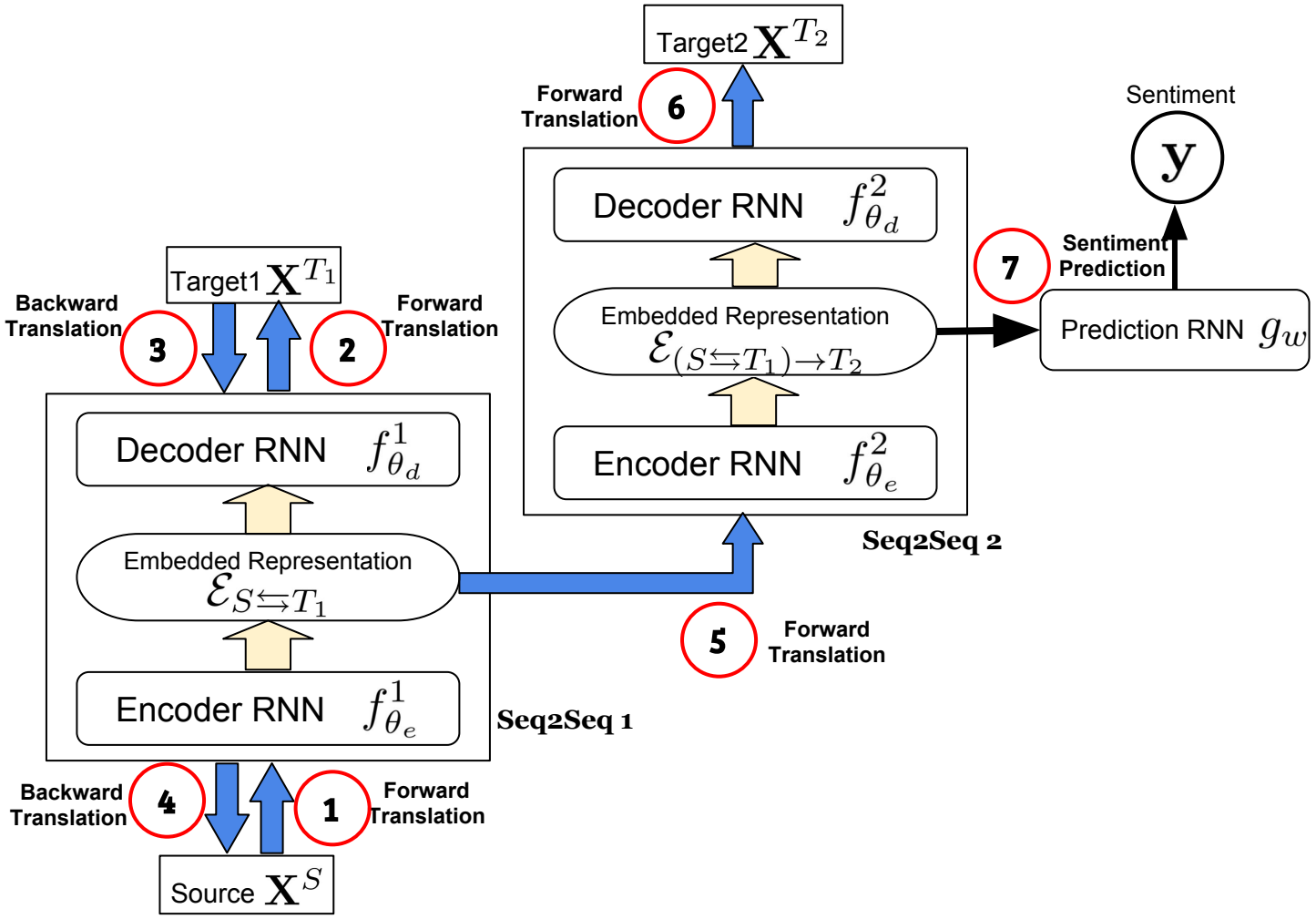
Hierarchical Multimodal Cyclic Translation Network



Hierarchical Multimodal Cyclic Translation Network



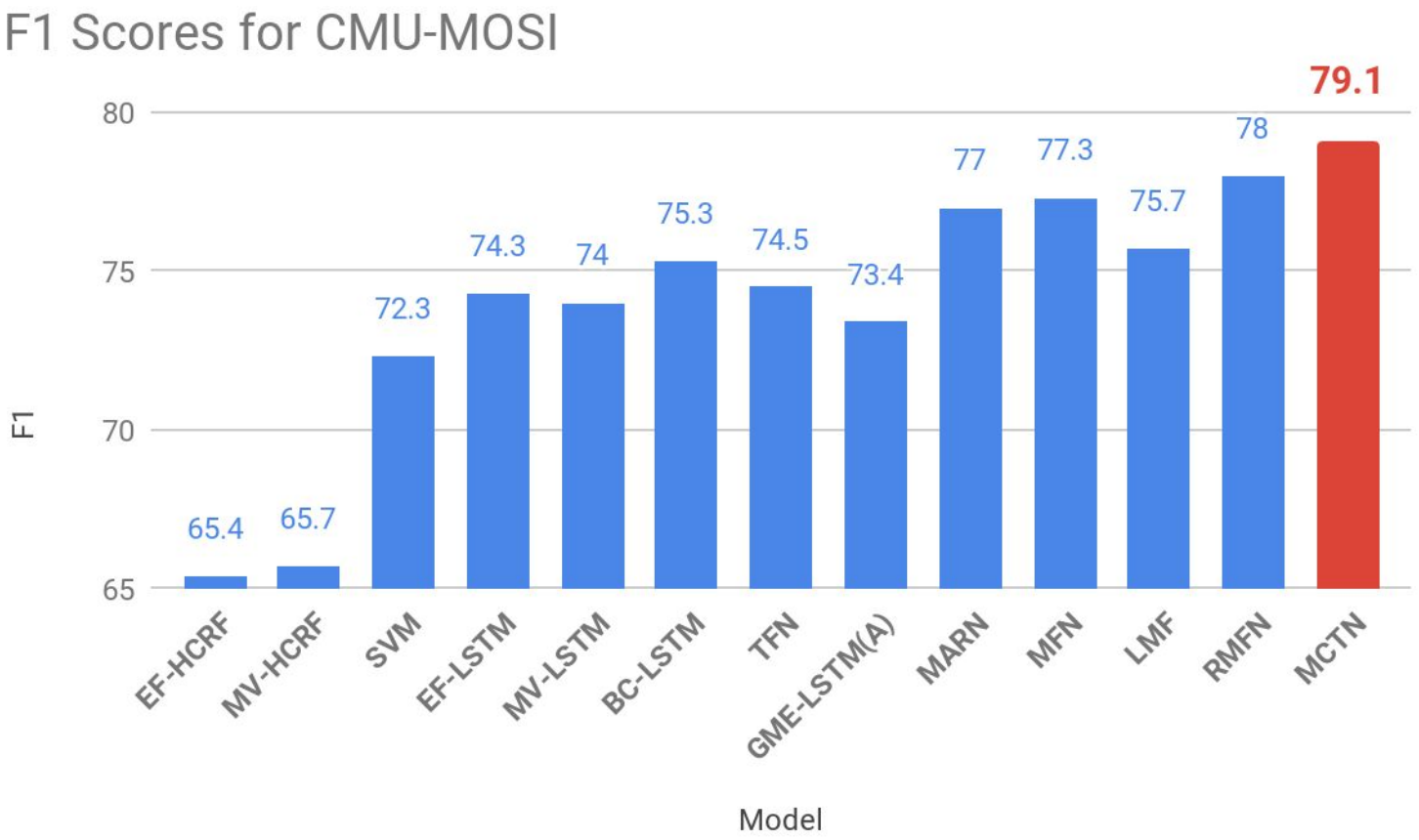
Hierarchical Multimodal Cyclic Translation Network



Baseline Models

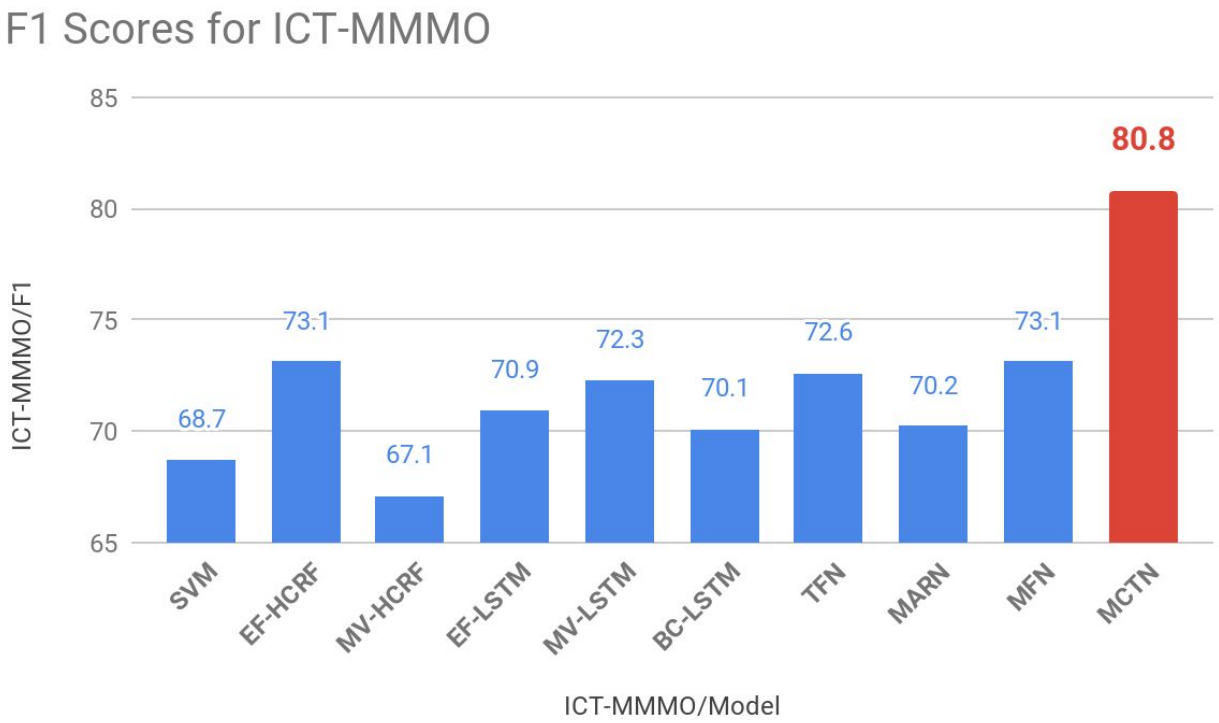
1. Non-temporal models: SVM (Cortes and Vapnik, 1995), DF (Nojavanasghari et al., 2016)
2. Early fusion: EF-LSTM (Hochreiter and Schmidhuber, 1997), EF-RHN (Zilly et al., 2016)
3. Late fusion: LMF (Liu et al., 2018), TFN (Zadeh et al., 2017), BC-LSTM (Poria et al., 2017)
4. Multi-view learning: MV-LSTM (Rajagopalan et al., 2016)
5. Memory-based models: MARN, MFN (Zadeh et al., 2018)
6. Multi-stage model: RMFN (Liang et al., 2018)

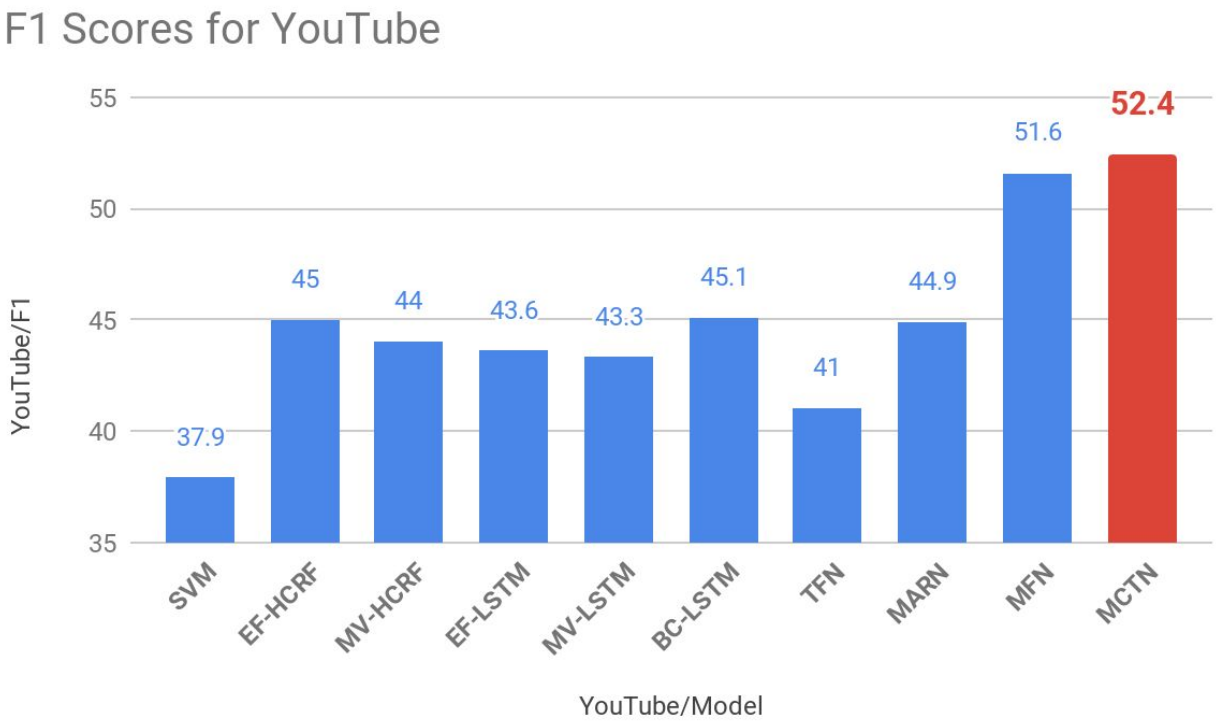
State-of-the-art Results: CMU-MOSI



MCTN: Only language modality required at test time!

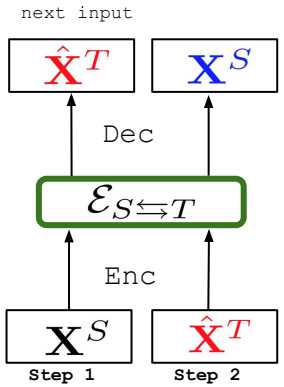
State-of-the-art Results: ICT-MMMO and YouTube



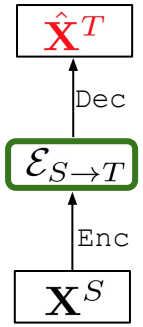


MCTN: Only language modality required at test time!

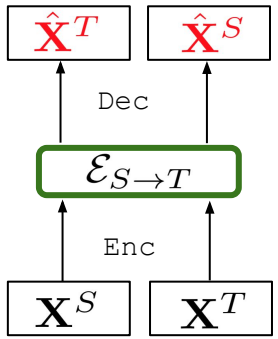
Bimodal Variations



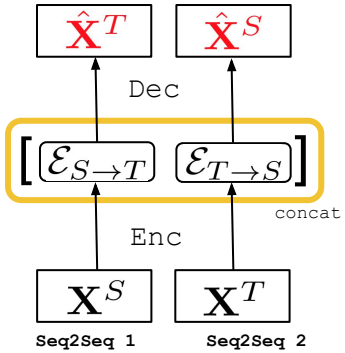
MCTN Bimodal



Simple Bimodal



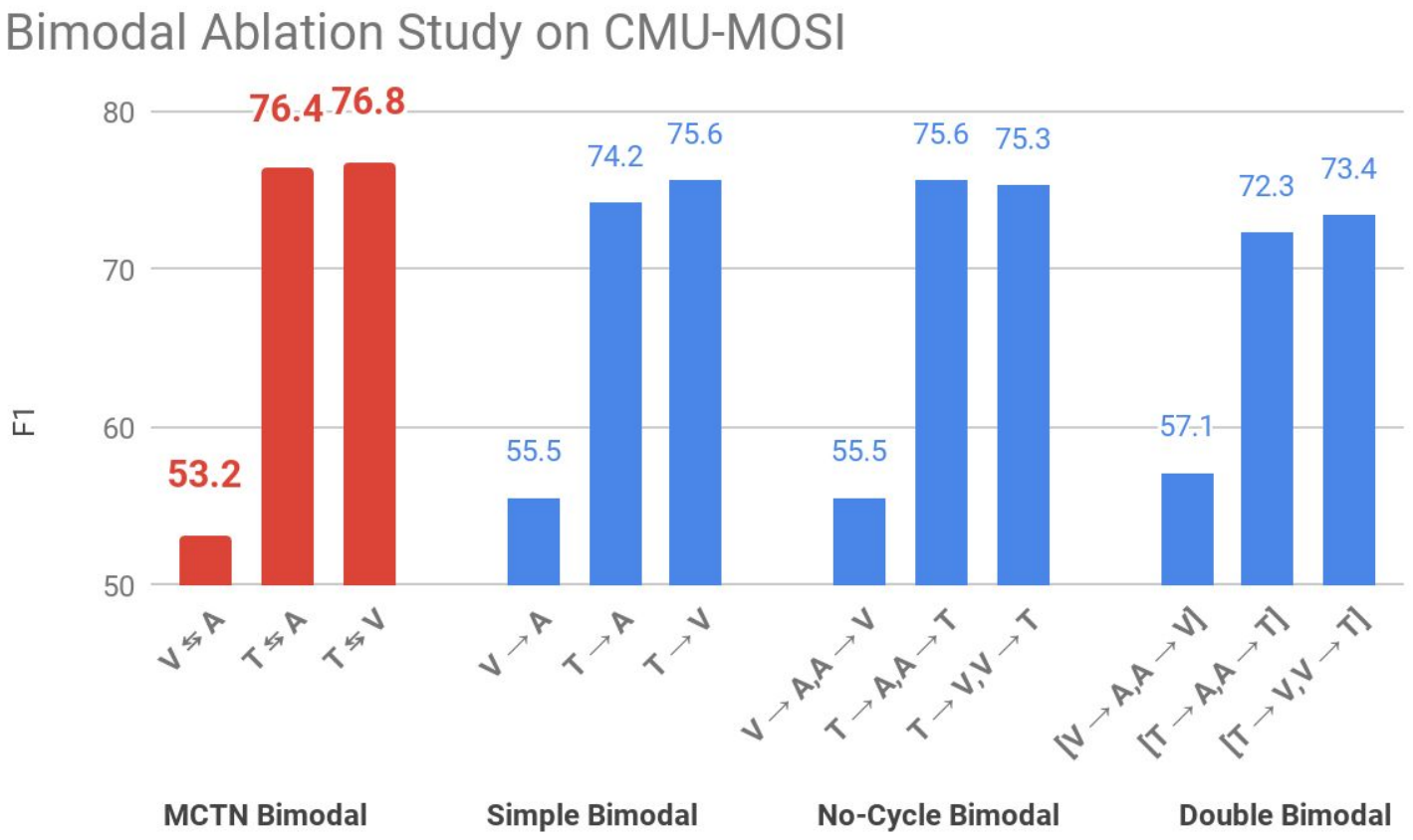
No-Cycle Bimodal



Double Bimodal

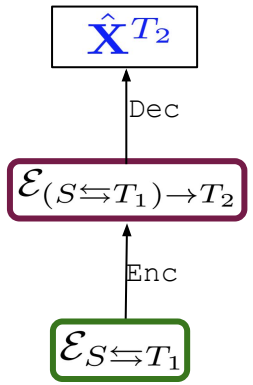
Test: use of cyclic translations, modality ordering, and hierarchical structure

Bimodal Variations Results

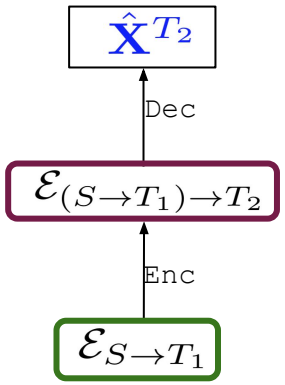


1. Use cyclic translations
2. Use language as source modality
3. Share parameters in seq2seq models

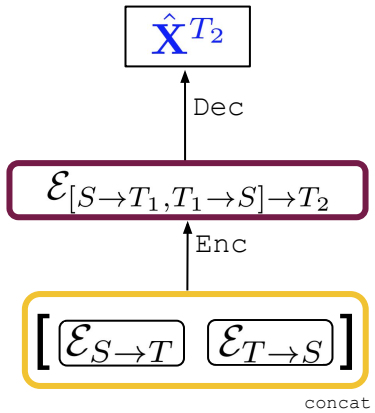
Trimodal Variations



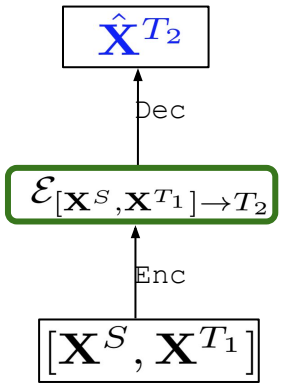
MCTN Trimodal



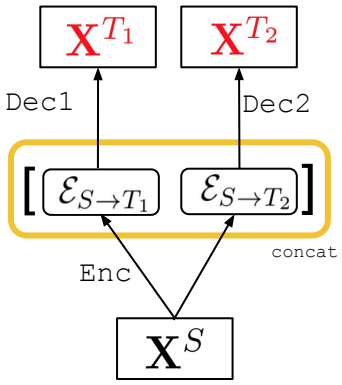
Simple Trimodal



Double Trimodal



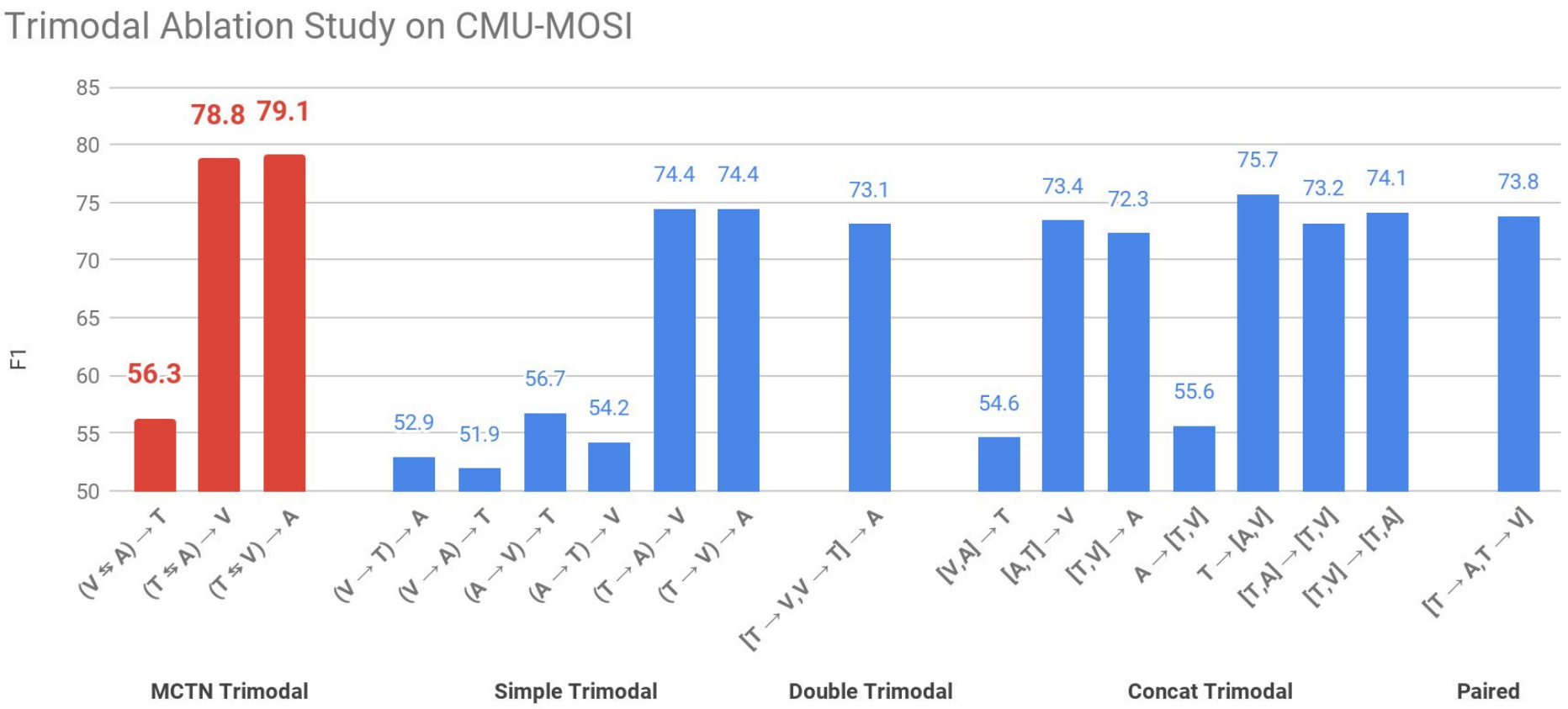
Concat Trimodal



Paired Trimodal

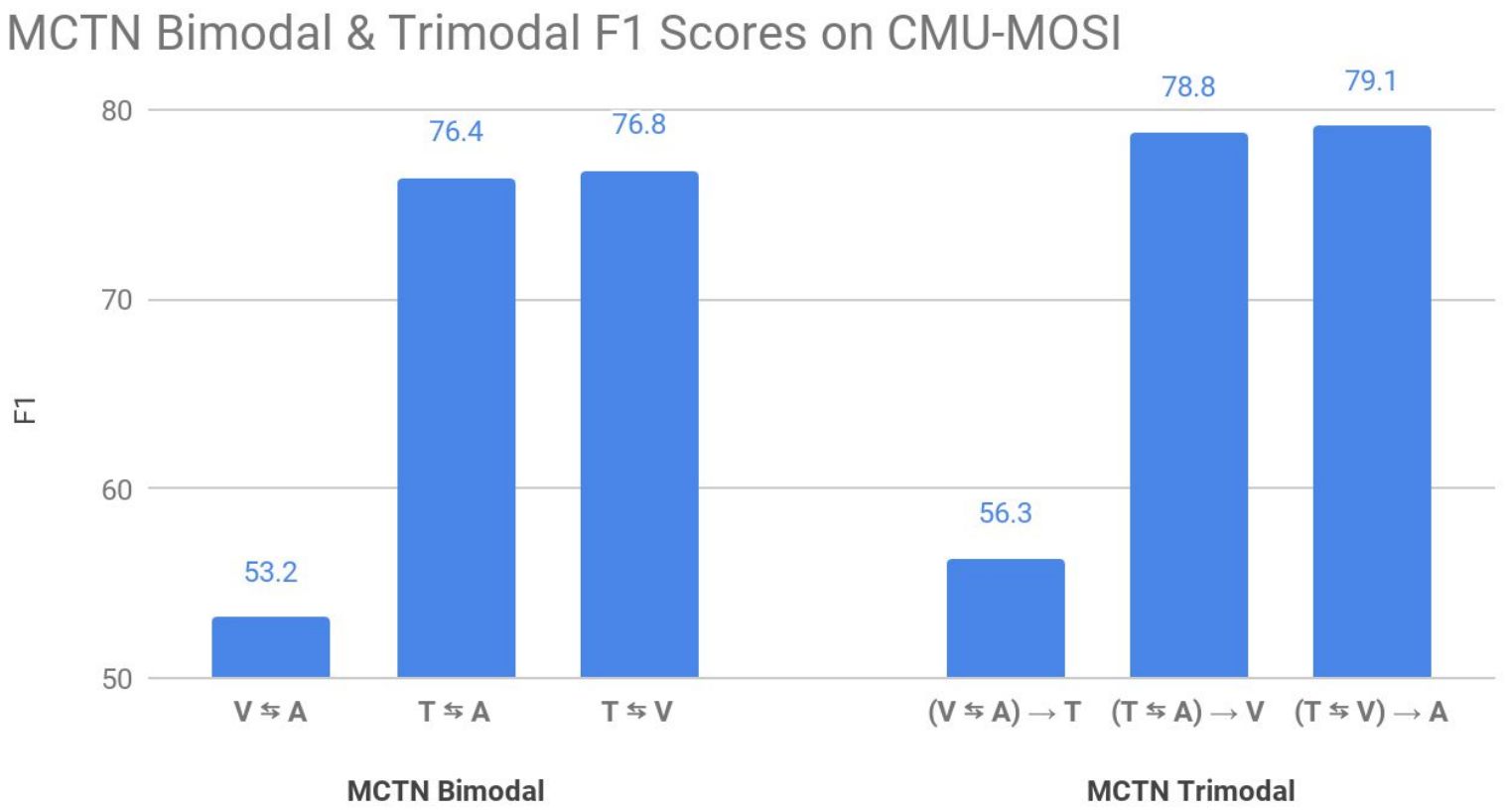
Test: use of cyclic translations, modality ordering, and hierarchical structure

Trimodal Variations Results

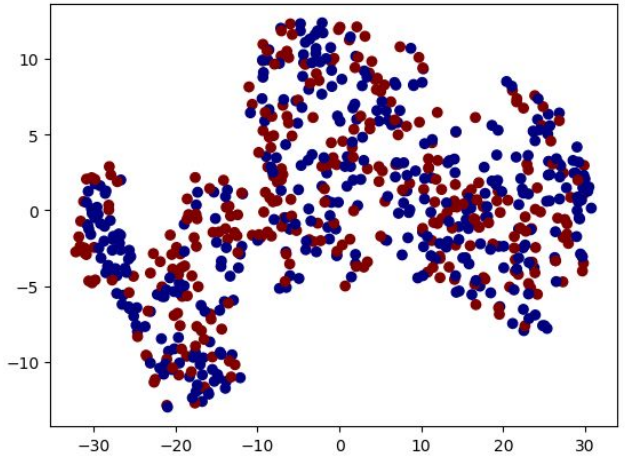


1. Use hierarchical translations
2. Use cyclic translations
3. Use language as source modality
4. Share parameters in seq2seq models

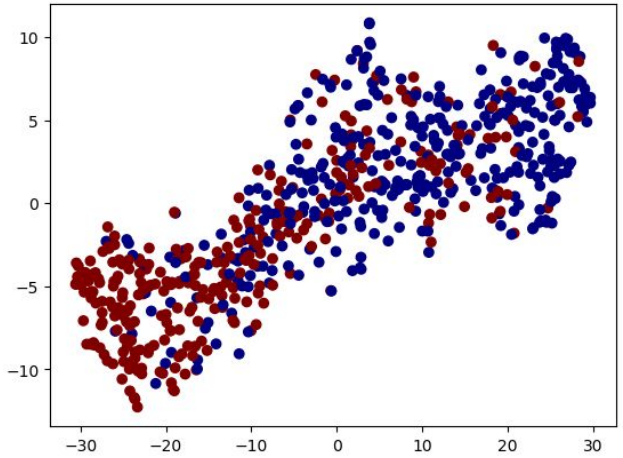
Adding More Modalities



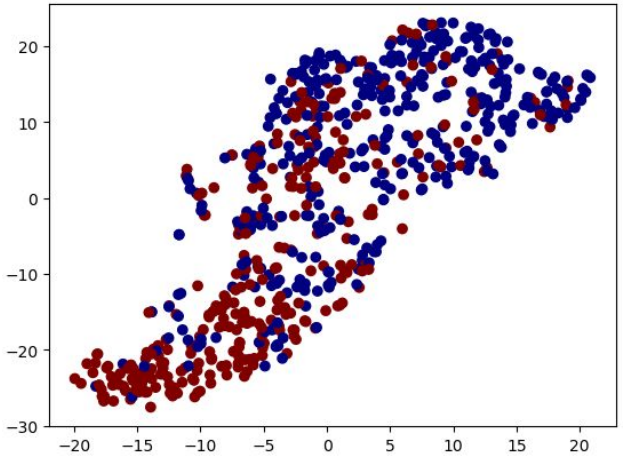
Adding More Modalities



Bimodal MCTN
without
cyclic translation



Bimodal MCTN
with
cyclic translation



Trimodal MCTN
with
cyclic translation

Thank you for your attention!

Code: https://github.com/hainow/MCTN/

Email: htpham@cs.cmu.edu
Twitter: @hai_t_pham

Email: plinag@cs.cmu.edu
Twitter: @pliang279

Email: tmanzini@cs.cmu.edu
Twitter: @Tom_Manzini