Language Technologies Institute

MACHINE LEARNING DEPARTMENT

Carnegie Mellon University

NANYANG TECHNOLOGICAL UNIVERSITY

Multi-attention Recurrent Network for Human Communication Comprehension

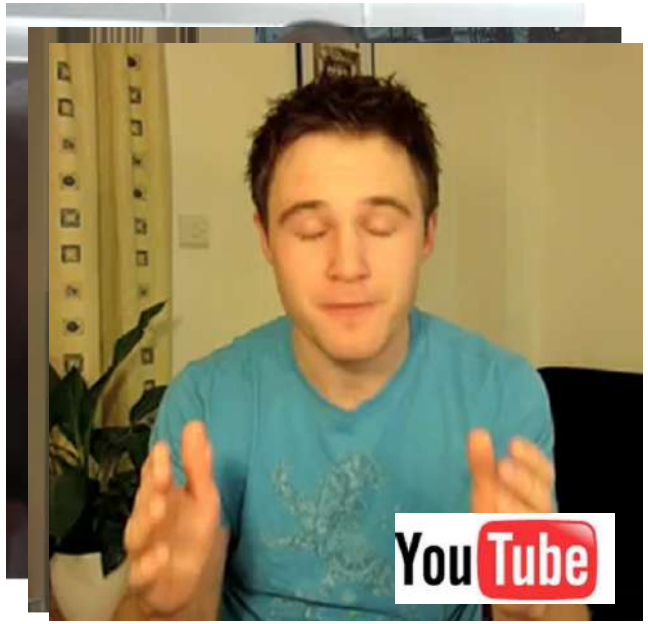
Amir Zadeh, Paul Pu Liang, Soujanya Poria,

Prateek Vij, Erik Cambria, Louis-Philippe Morency

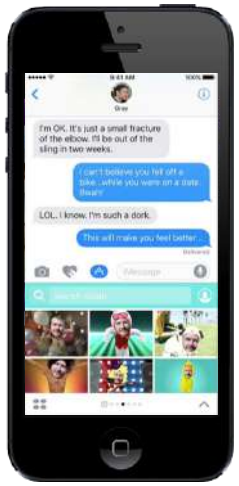
Presenter: Paul Pu Liang

Progress of Artificial Intelligence

Multimedia Content



Intelligent
Personal Assistants



Robots and
Virtual Agents

Multimodal Communicative Behaviors

Language

- **Lexicon**
  - Words
- **Syntax**
  - Part-of-speech
  - Dependencies
- **Pragmatics**
  - Discourse acts

Acoustic

- **Prosody**
  - Intonation
  - Voice quality
- **Vocal expressions**
  - Laughter, moans

Visual

- **Gestures**
  - Head gestures
  - Eye gestures
  - Arm gestures
- **Body language**
  - Body posture
  - Proxemics
- **Eye contact**
  - Head gaze
  - Eye gaze
- **Facial expressions**
  - FACS action units
  - Smile, frowning

Sentiment

- Positive
- Negative

Emotion

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

Personality

- Confidence
- Persuasion
- Passion

Challenge 1: Intra-modal dynamics

Intra-modal

Speaker's behaviors | **Sentiment Intensity**

"This movie is great" → ++

time

Smile | Head nod → +

time

Challenge 1: Intra-modal dynamics

Intra-modal

Speaker's behaviors **Sentiment Intensity**

"This movie is great"

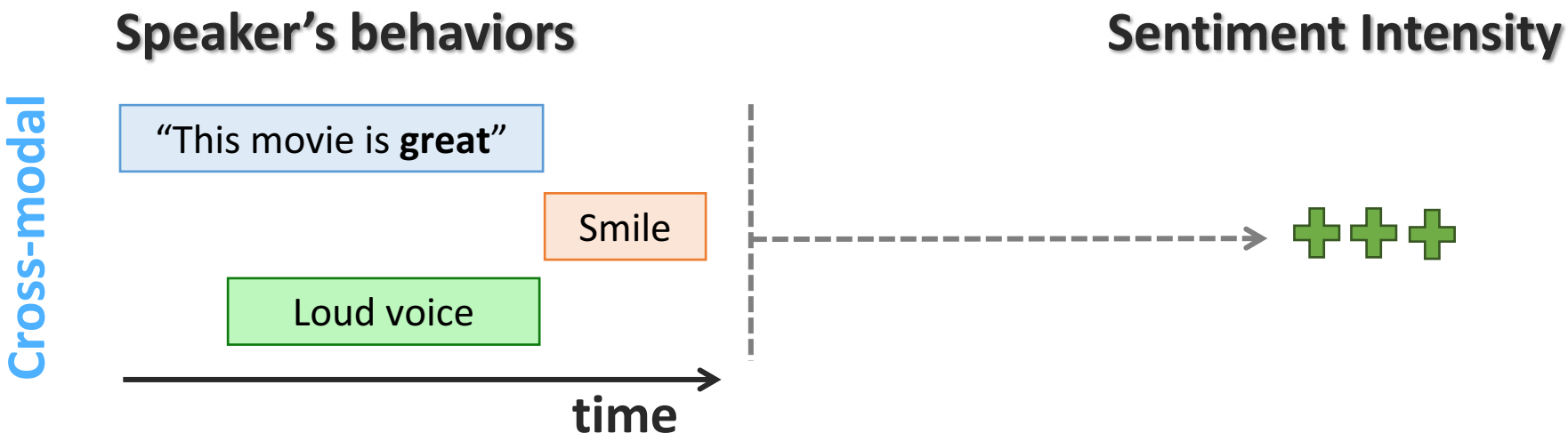
time

Smile Head nod

time

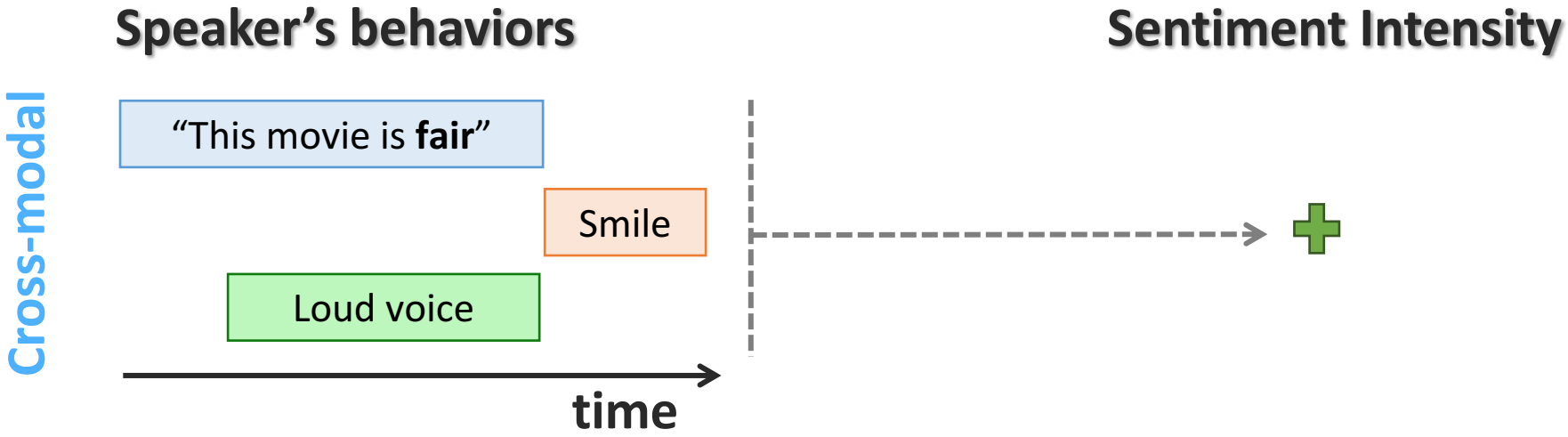
Challenge 2: Cross-modal Dynamics

a) Multiple co-occurring interactions



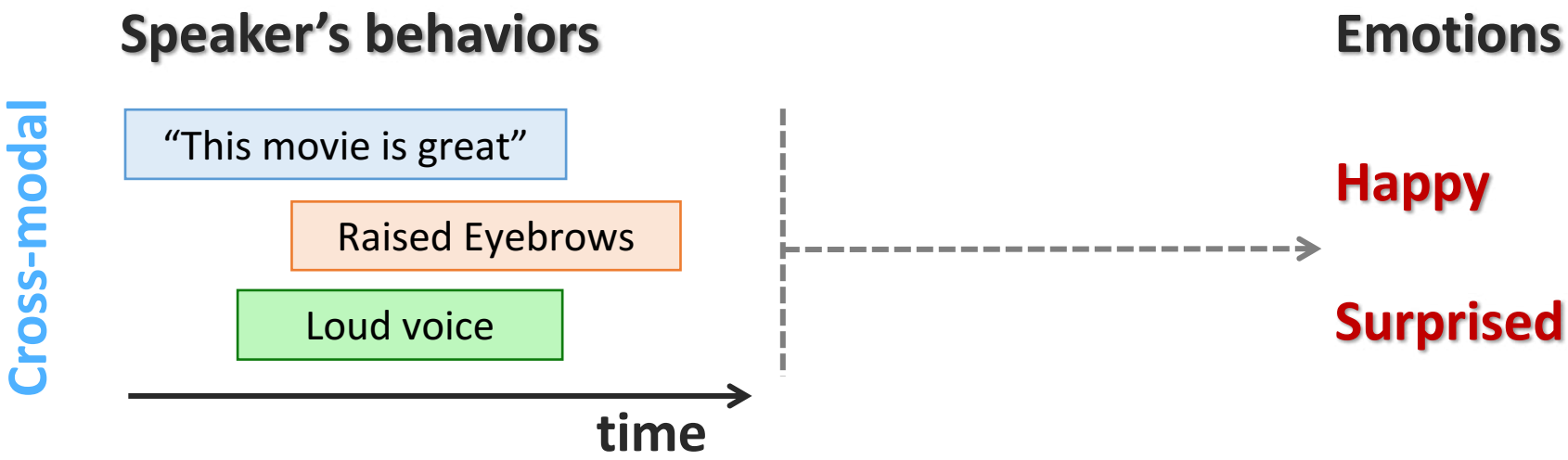
Challenge 2: Cross-modal Dynamics

a) Multiple co-occurring interactions
b) Different weighted combinations



Challenge 2: Cross-modal Dynamics

a) Multiple co-occurring interactions
b) Different weighted combinations
c) Multiple prediction targets



Multi-attention Recurrent Network (MARN)

1 **Modeling intra-modal dynamics**

→ Set of Long-short Term Memories

Multi-attention Recurrent Network (MARN)

1 **Modeling intra-modal dynamics**

→ Set of Long-short Term Memories

2 **Modeling cross-modal dynamics**

→ Set of Long-short Term **Hybrid** Memories + Single-attention Block

Multi-attention Recurrent Network (MARN)

1 **Modeling intra-modal dynamics**

→ Set of Long-short Term Memories

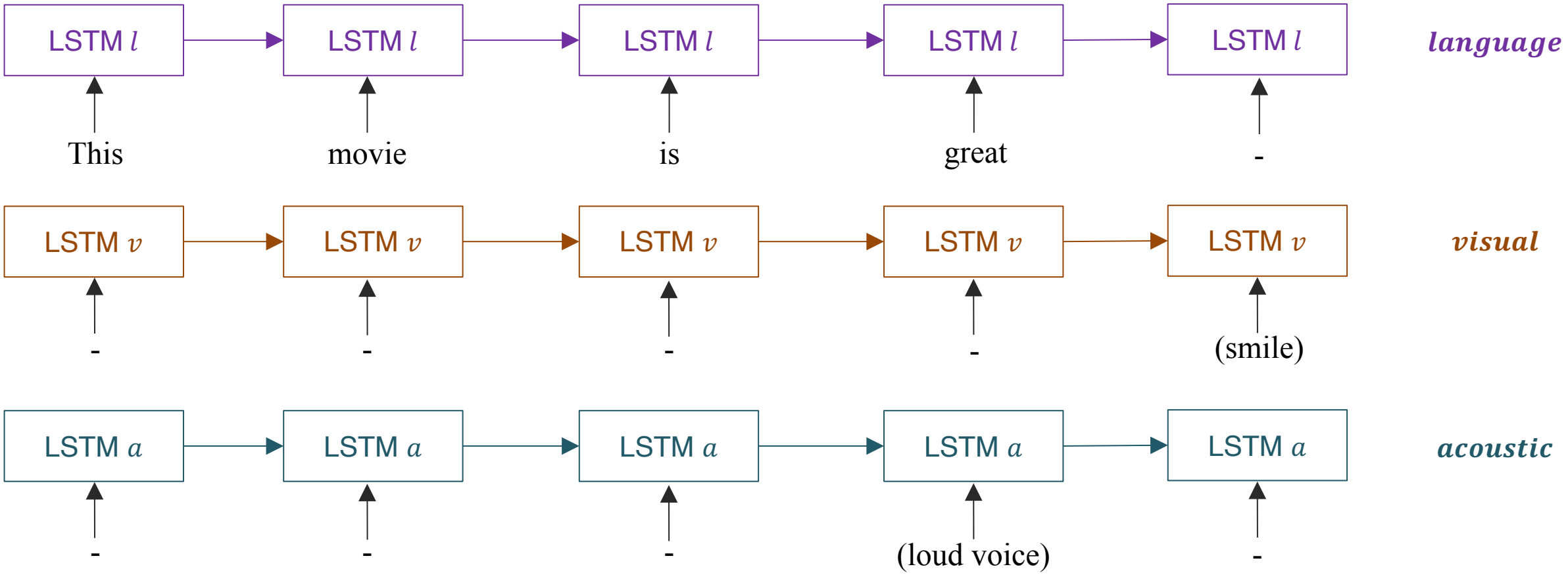
2 **Modeling cross-modal dynamics**

→ Set of Long-short Term **Hybrid** Memories + Single-attention Block

Modeling multiple cross-modal dynamics

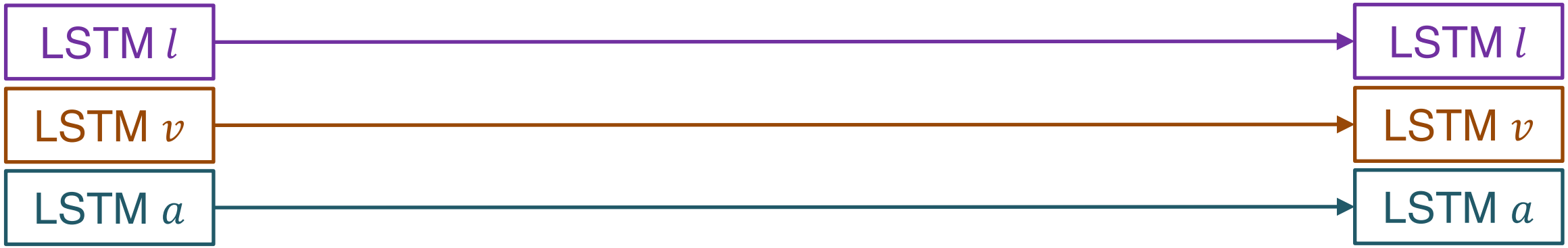
→ Set of Long-short Term **Hybrid** Memories + **Multi-attention** Block

Challenge 1: Intra-modal Dynamics

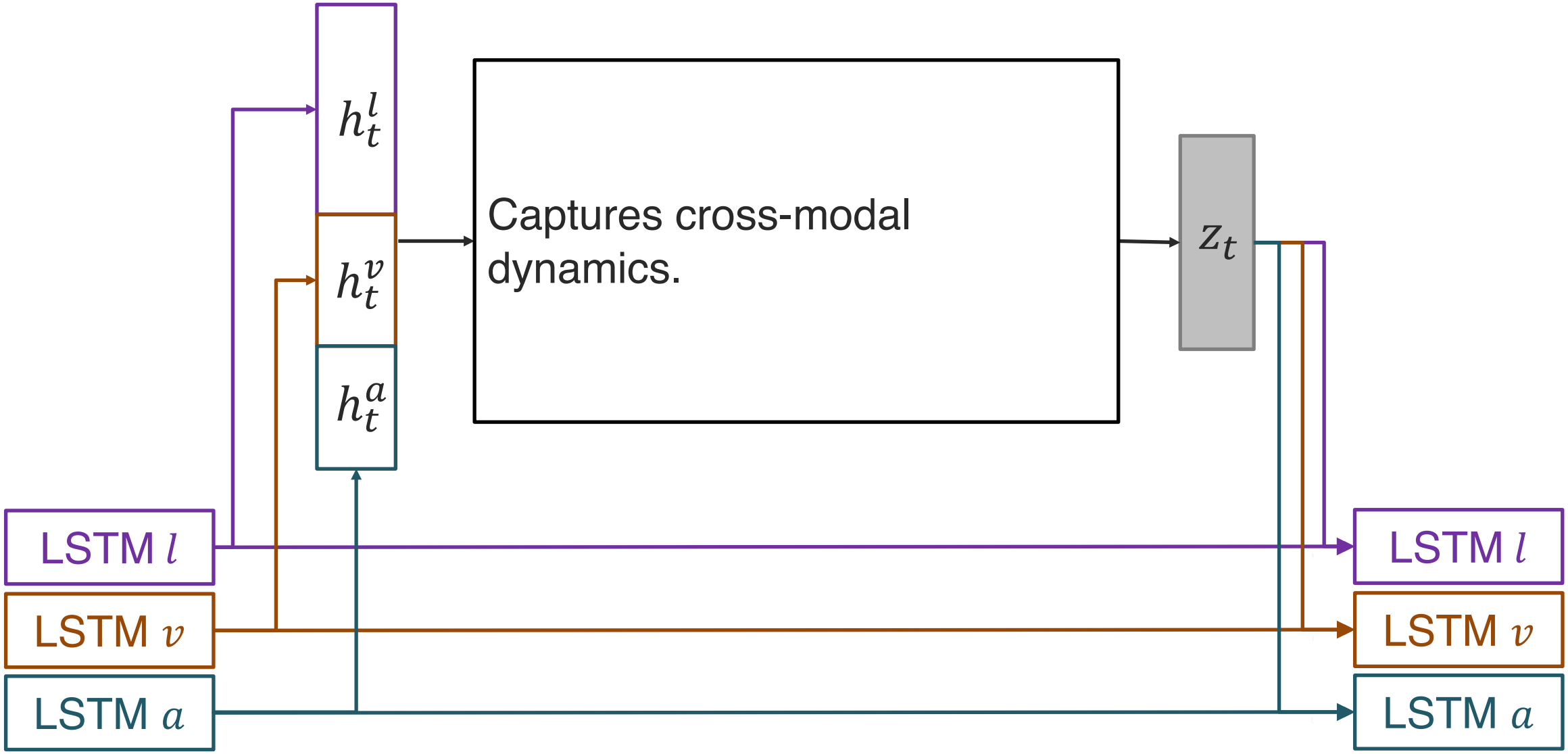


Challenge 2: Cross-modal Dynamics

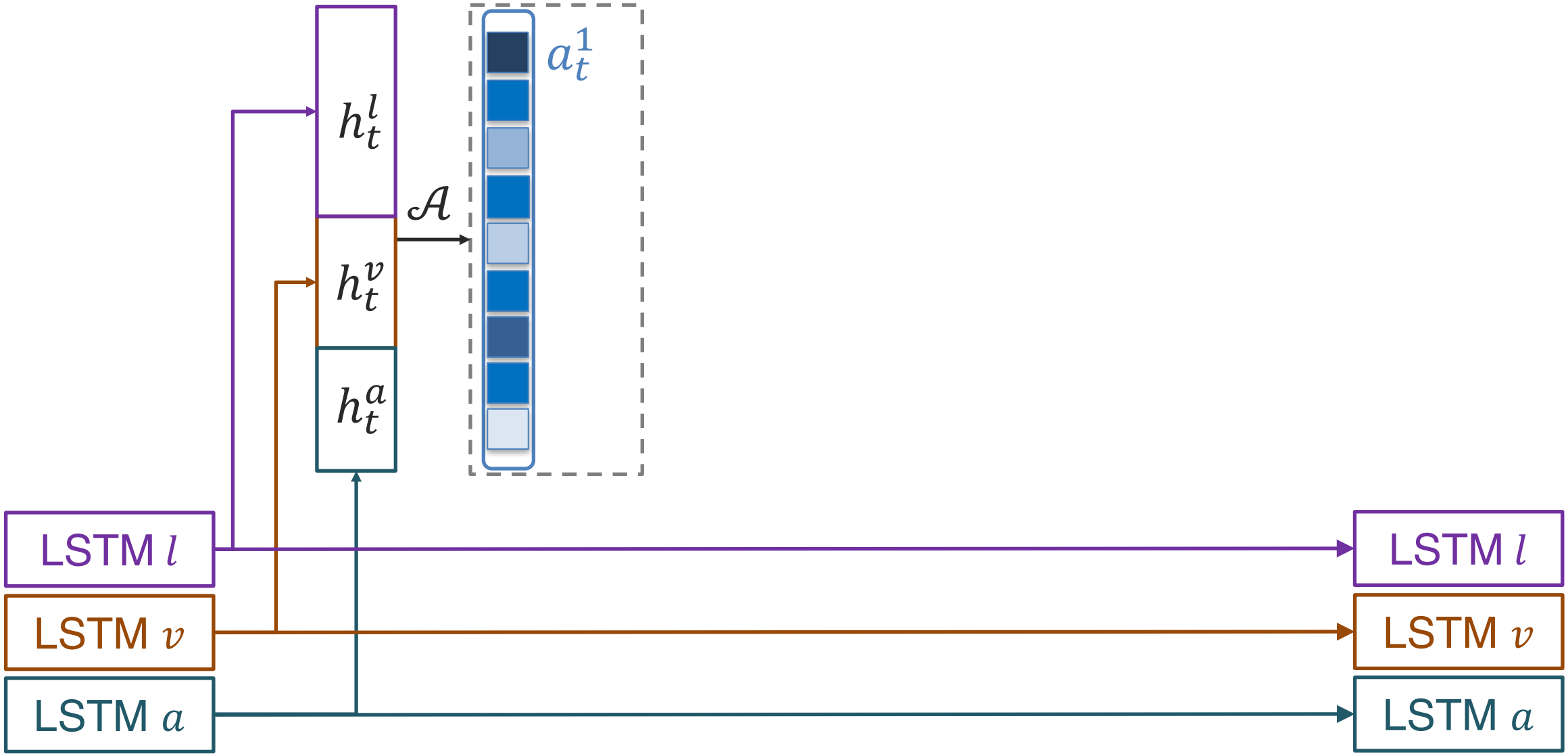
- How do we capture cross-modal dynamics continuously across time?



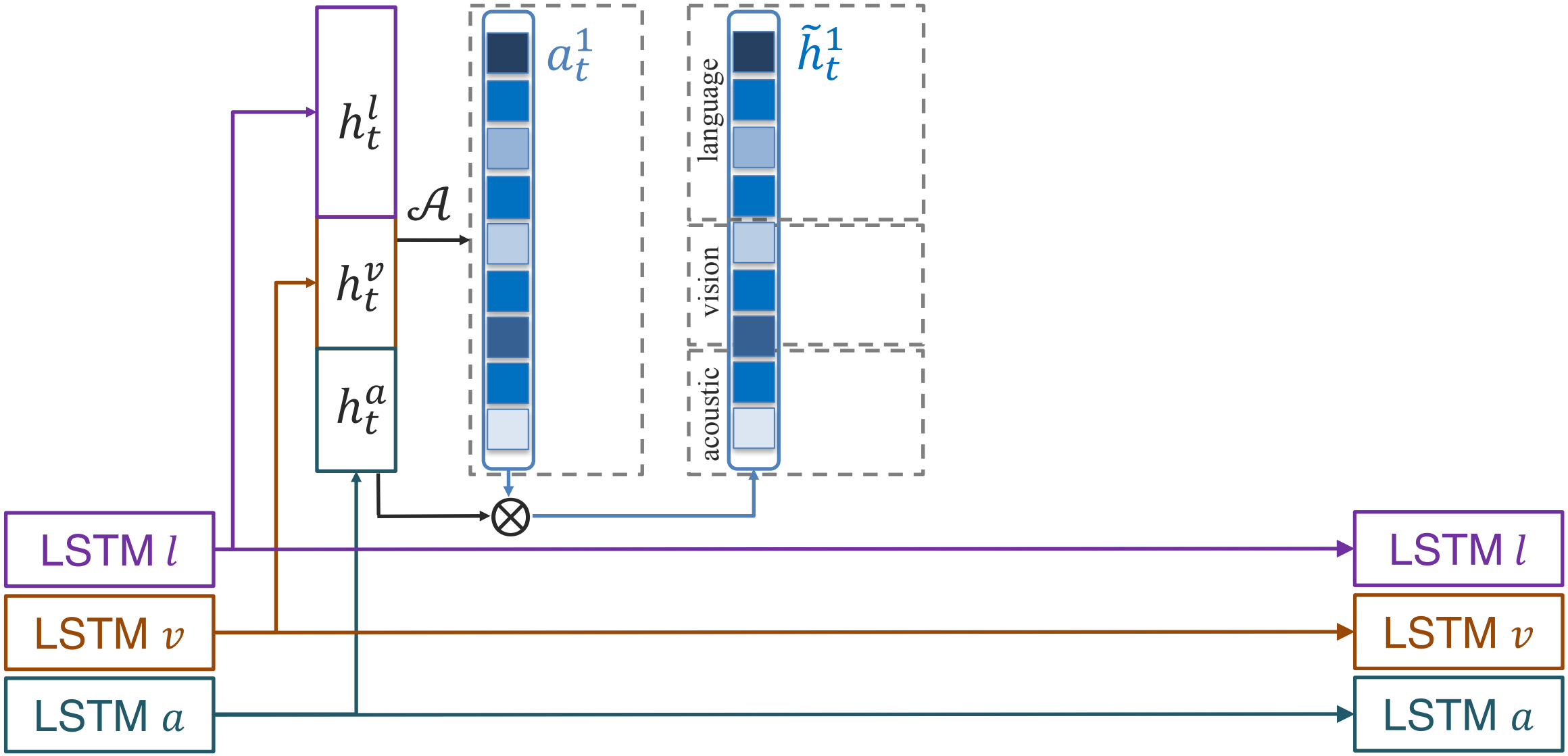
Challenge 2: Single-attention Block



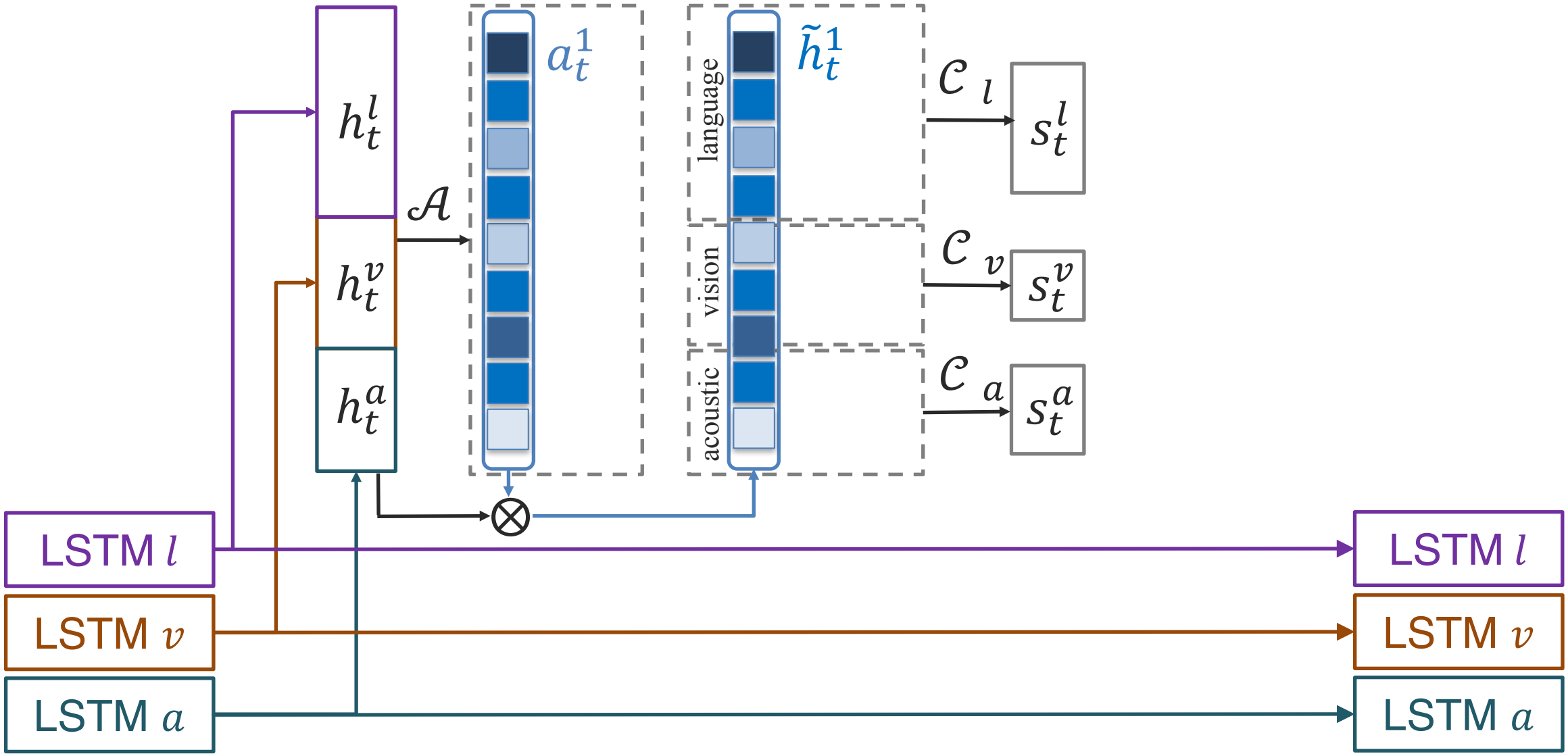
Challenge 2: Single-attention Block



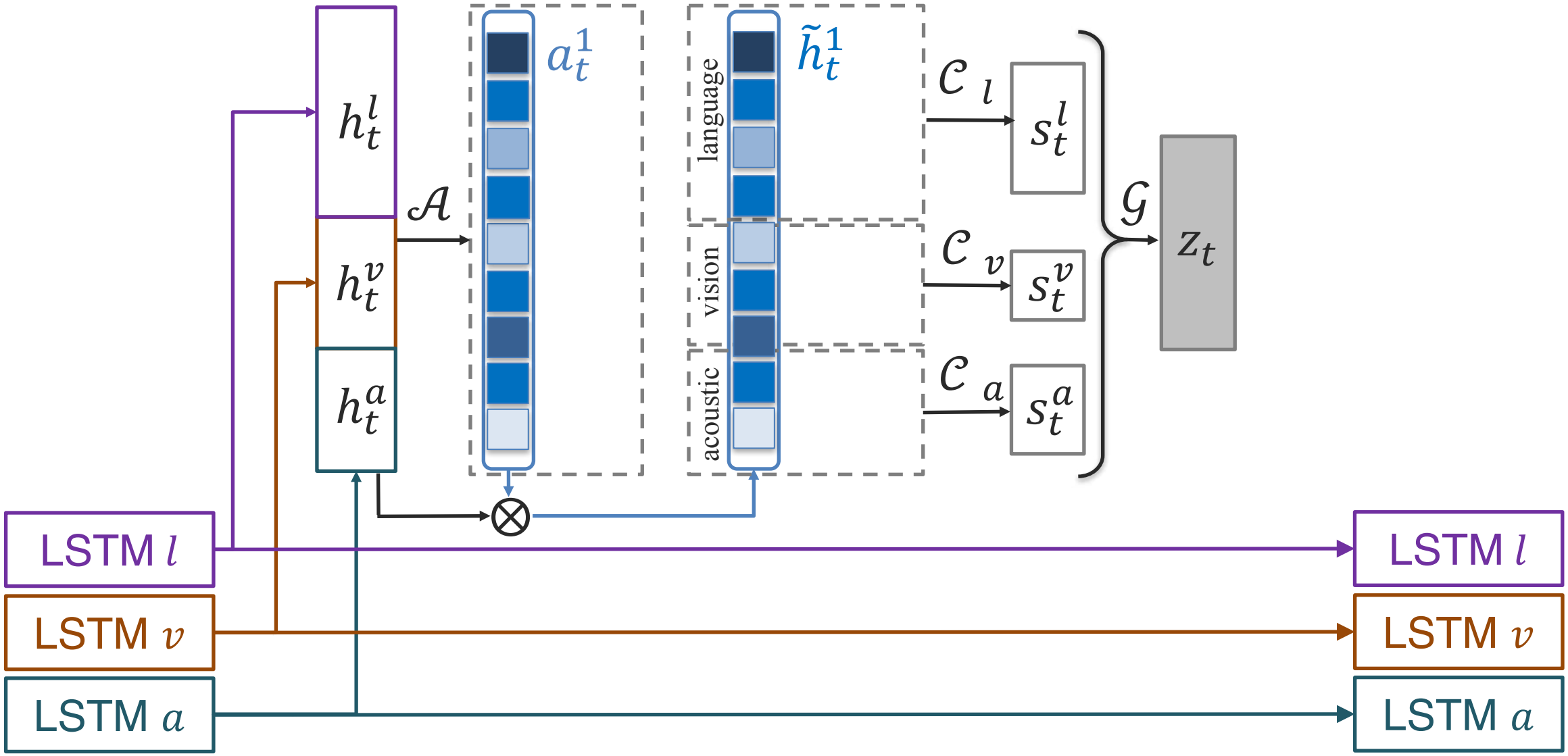
Challenge 2: Single-attention Block



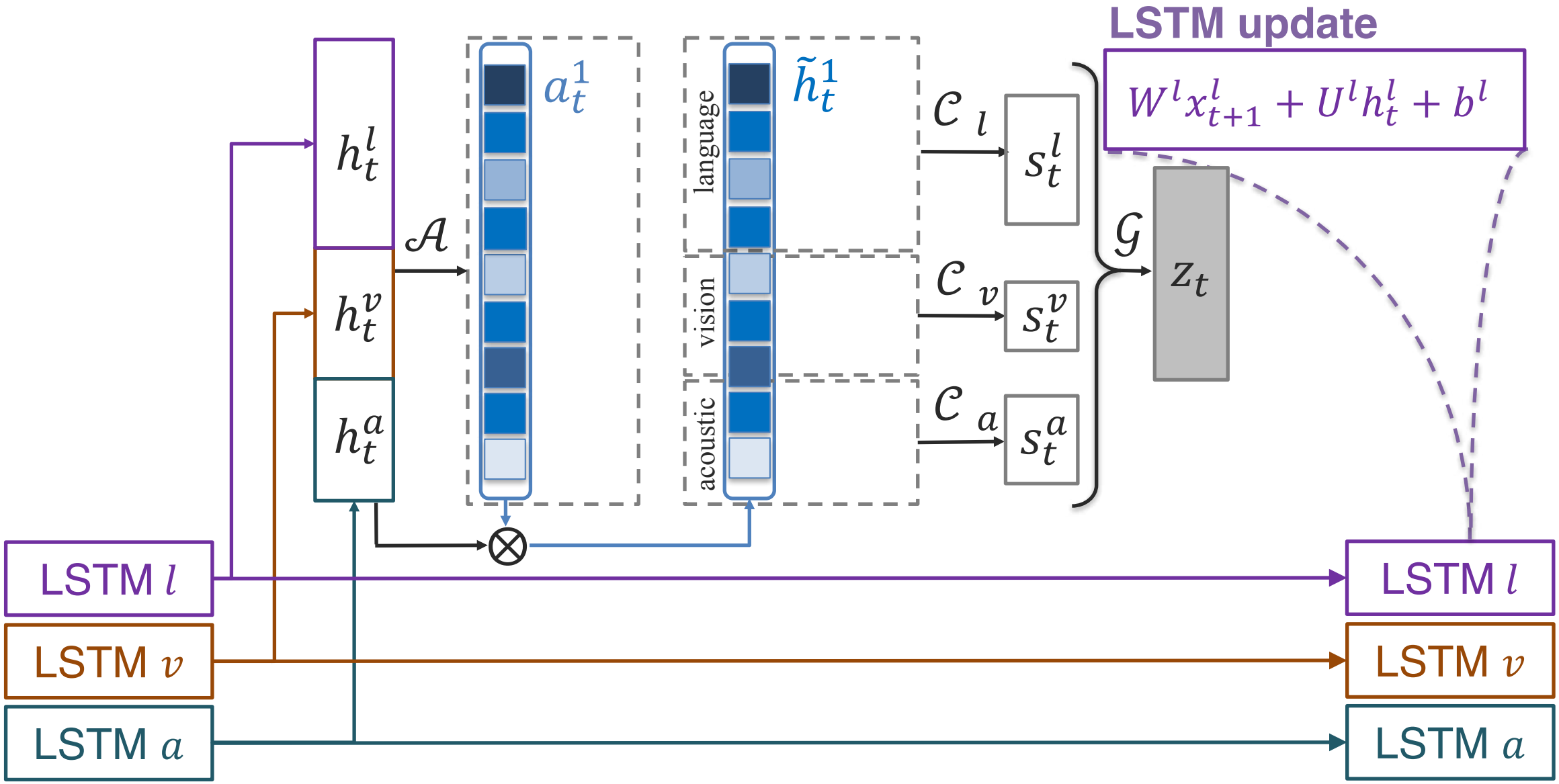
Challenge 2: Single-attention Block



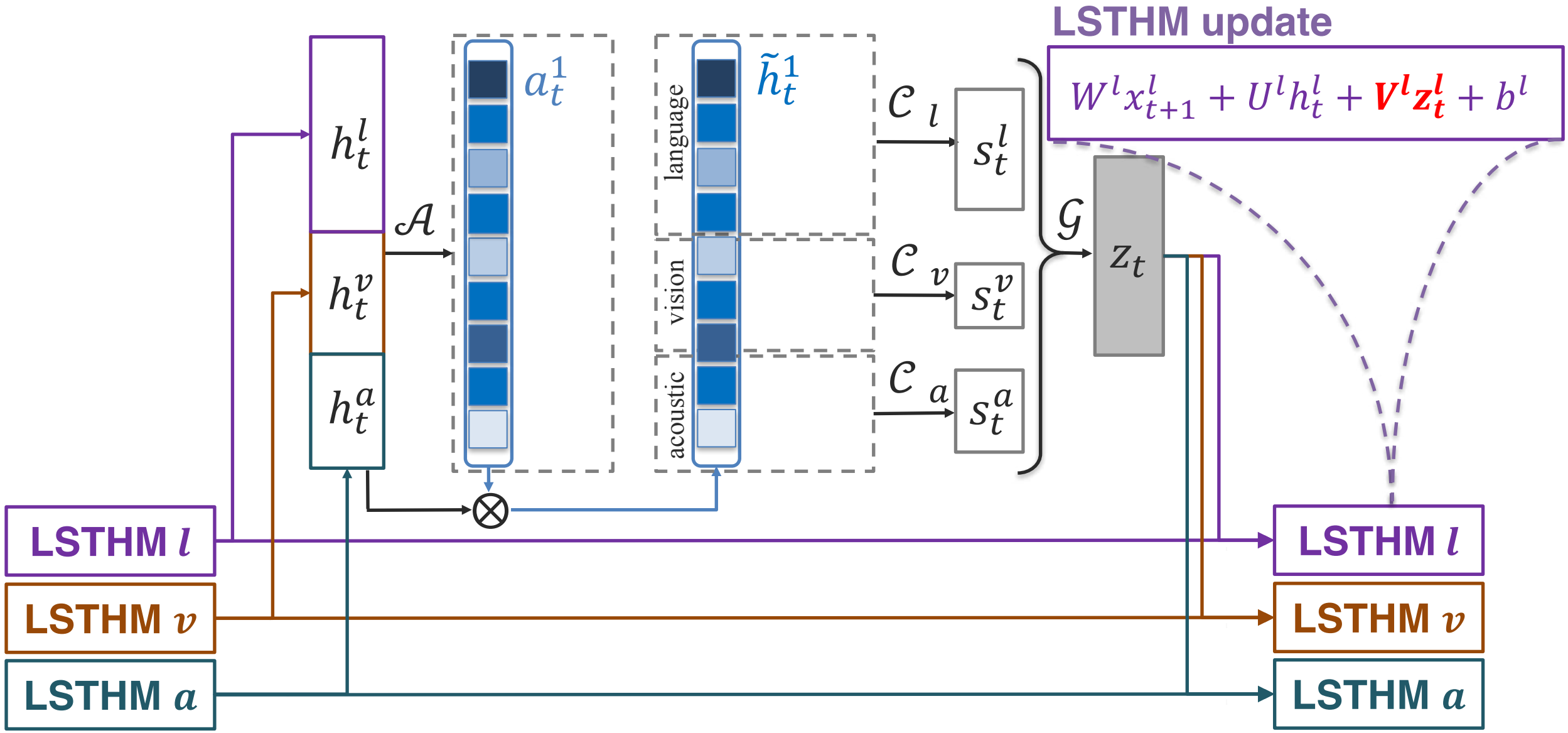
Challenge 2: Single-attention Block



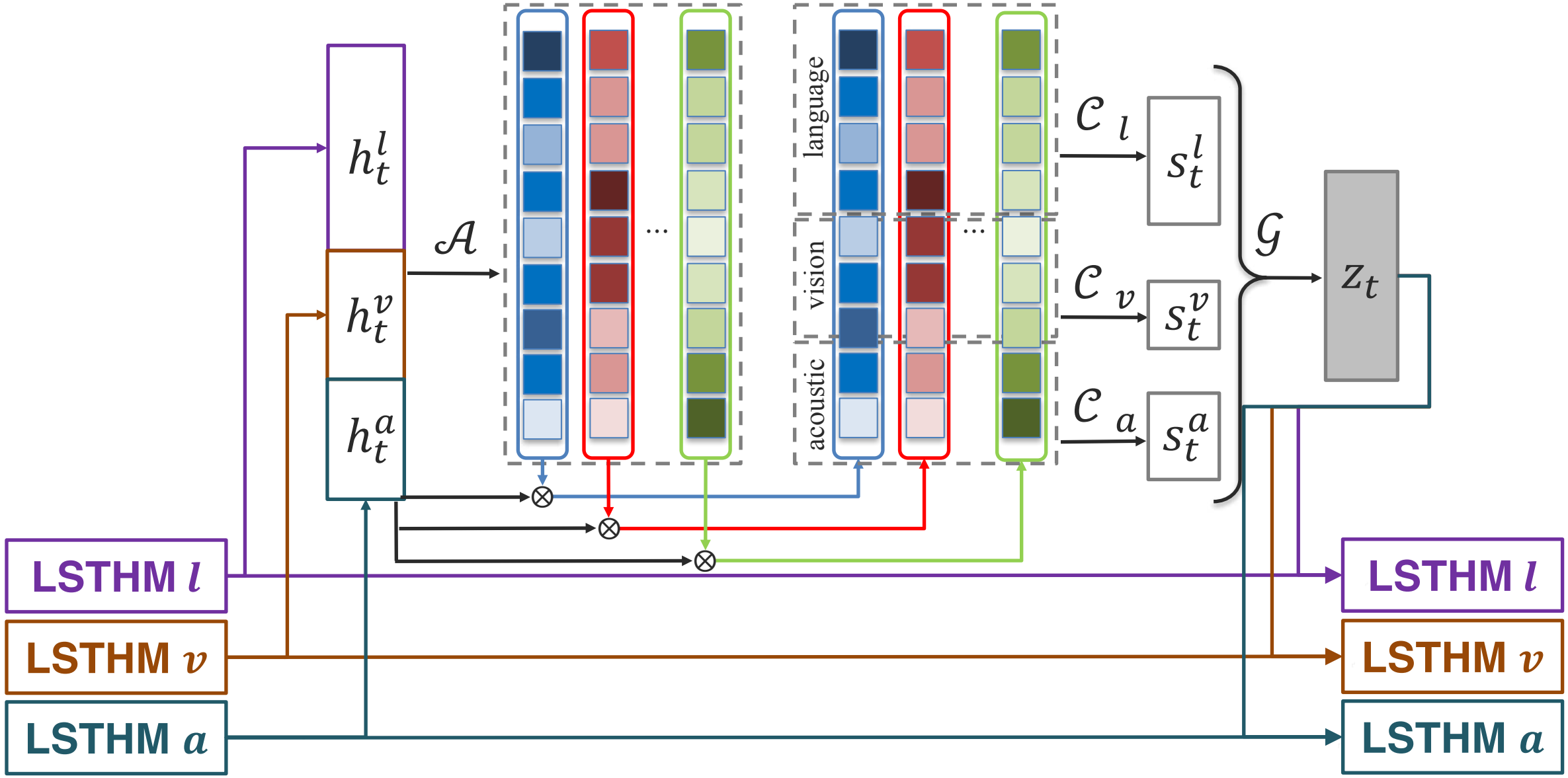
Challenge 2: Single-attention Block



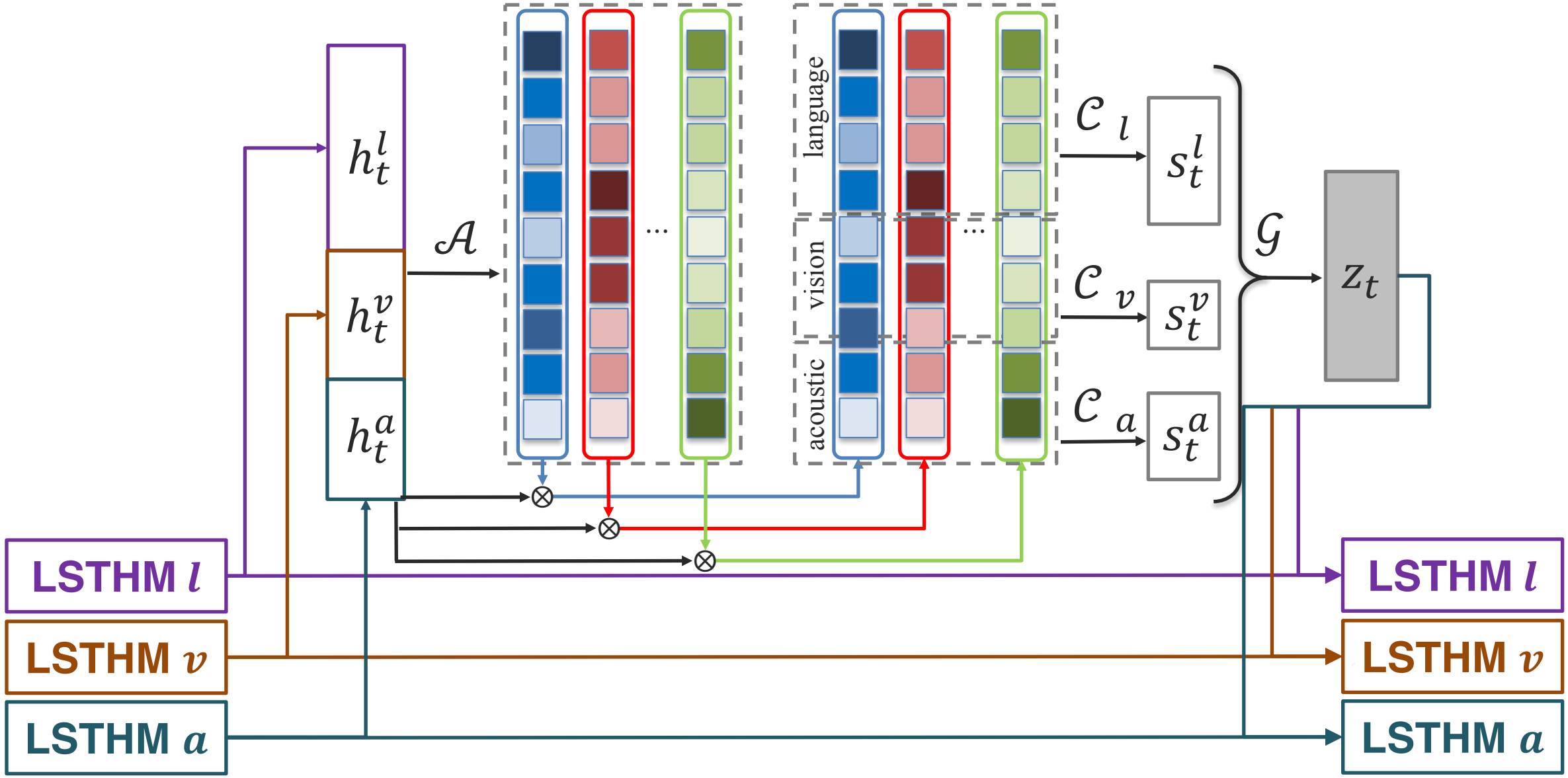
Challenge 2: Long-short Term Hybrid Memory



Challenge 2: Multi-attention Block



Multi-attention Recurrent Network (MARN)



Experiments

Language

- **Glove word embeddings**

Visual

- **Facet features**
  - FACS action units
  - Emotions

Acoustic

- **COVAREP features**
  - MFCCs
  - Pitch tracking

Alignment

- **Word level**
- **P2FA**

Sentiment

- Positive
- Negative

Emotion

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

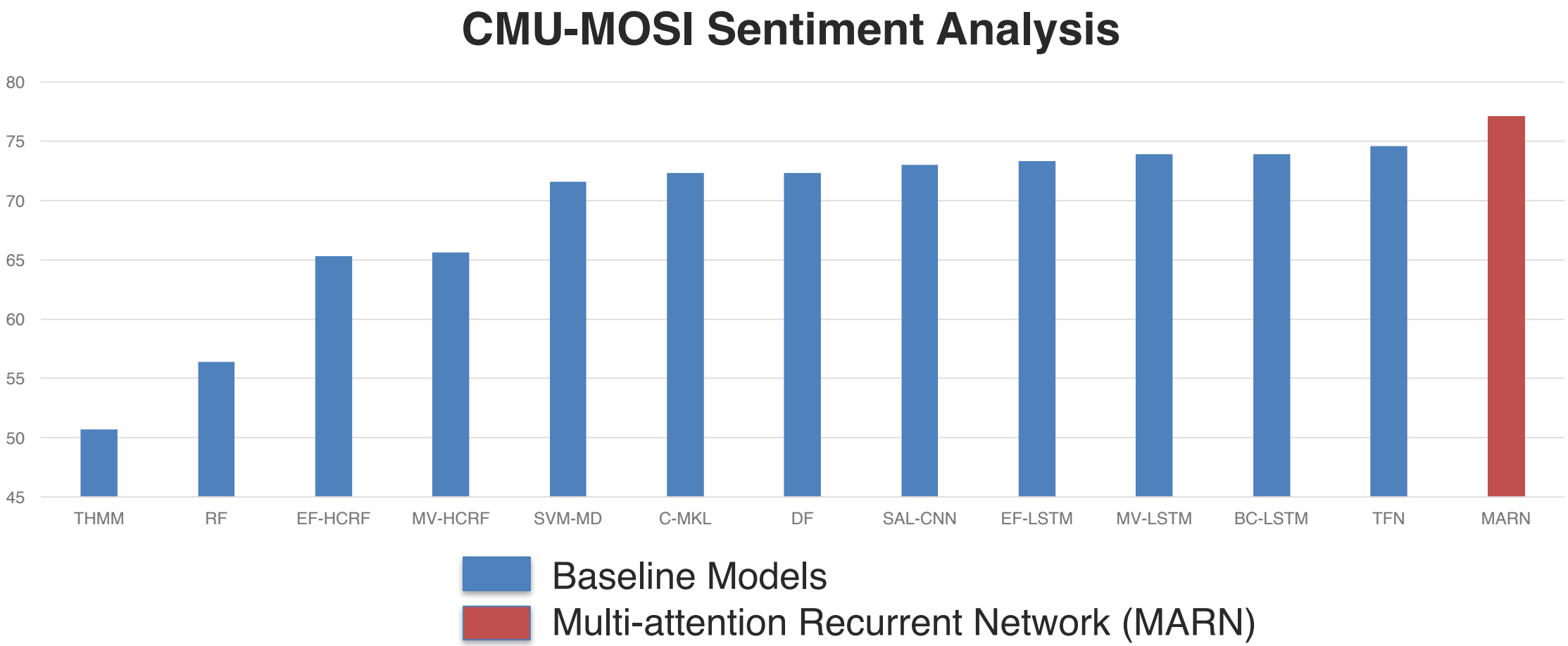
Personality

- Confidence
- Persuasion
- Passion

Baseline Models

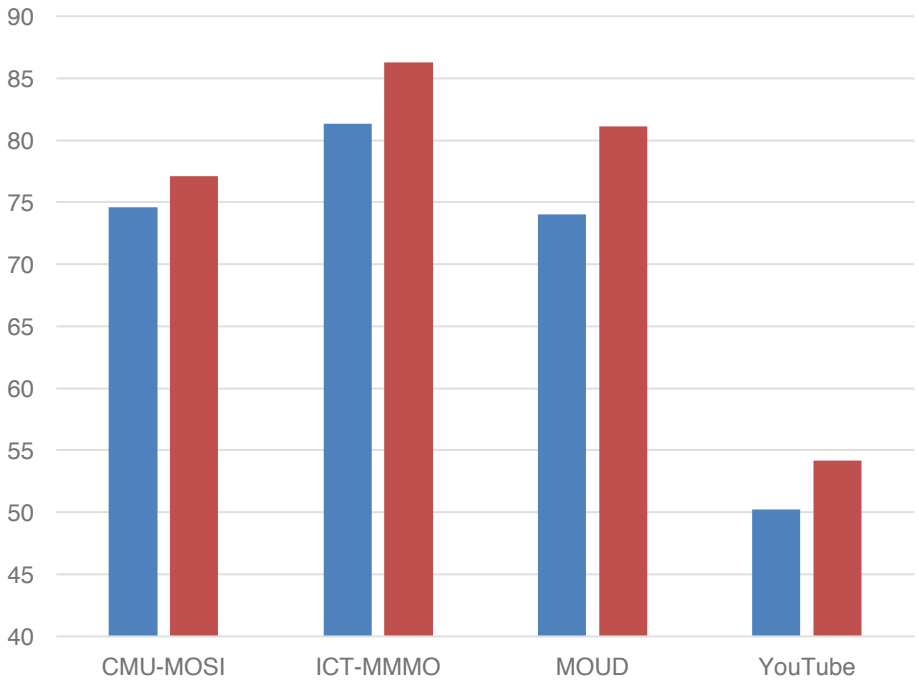
1. Non-temporal Models
   - SVM-MD, RF
2. Early Fusion
   - HMM, EF-LSTM, EF-HCRF, C-MKL, SAL-CNN
3. Late Fusion
   - DF, TFN, BC-LSTM
4. Multi-view Learning
   - MV-HMMs, MV-HCRFs, MV-LSTM

State-of-the-art Results

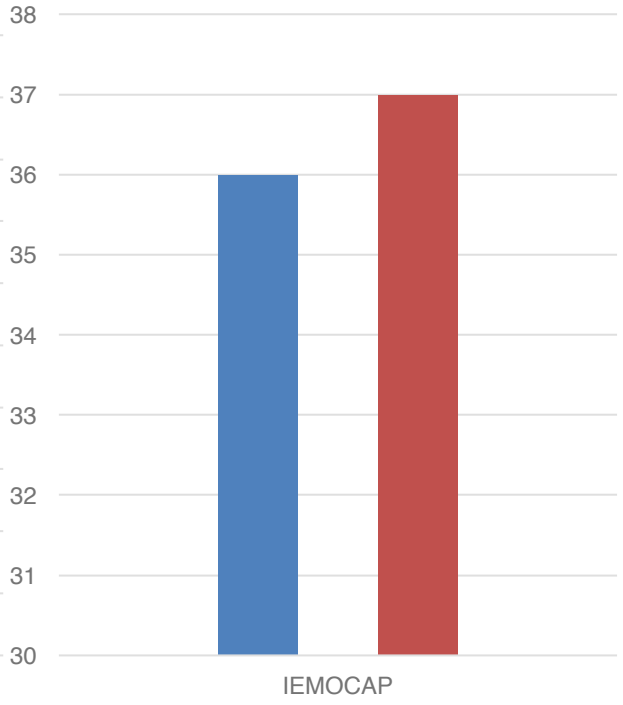


State-of-the-art Results

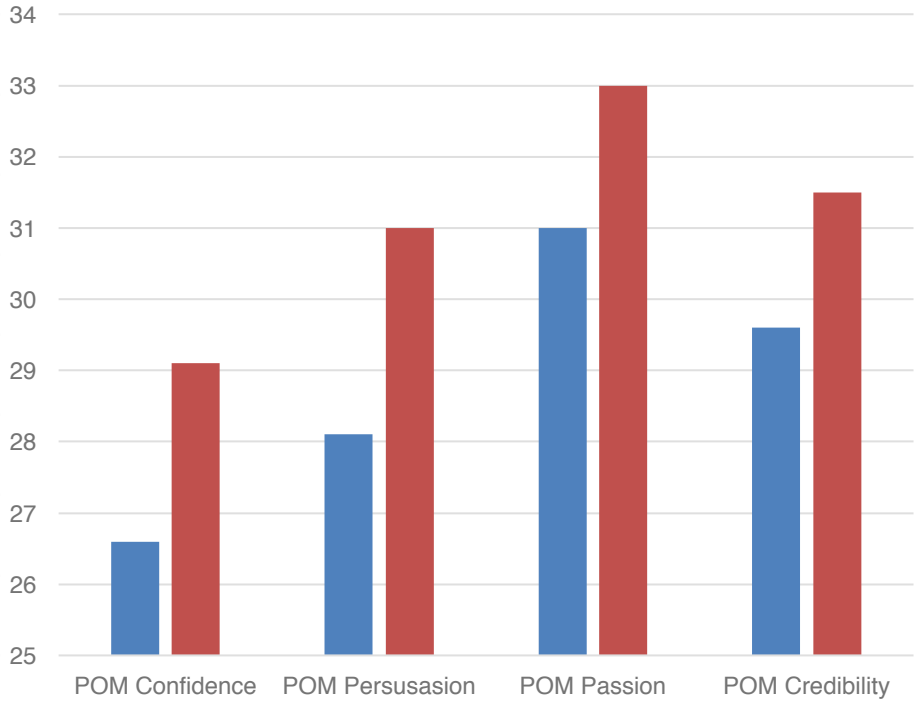
Sentiment Analysis



Emotion Recognition



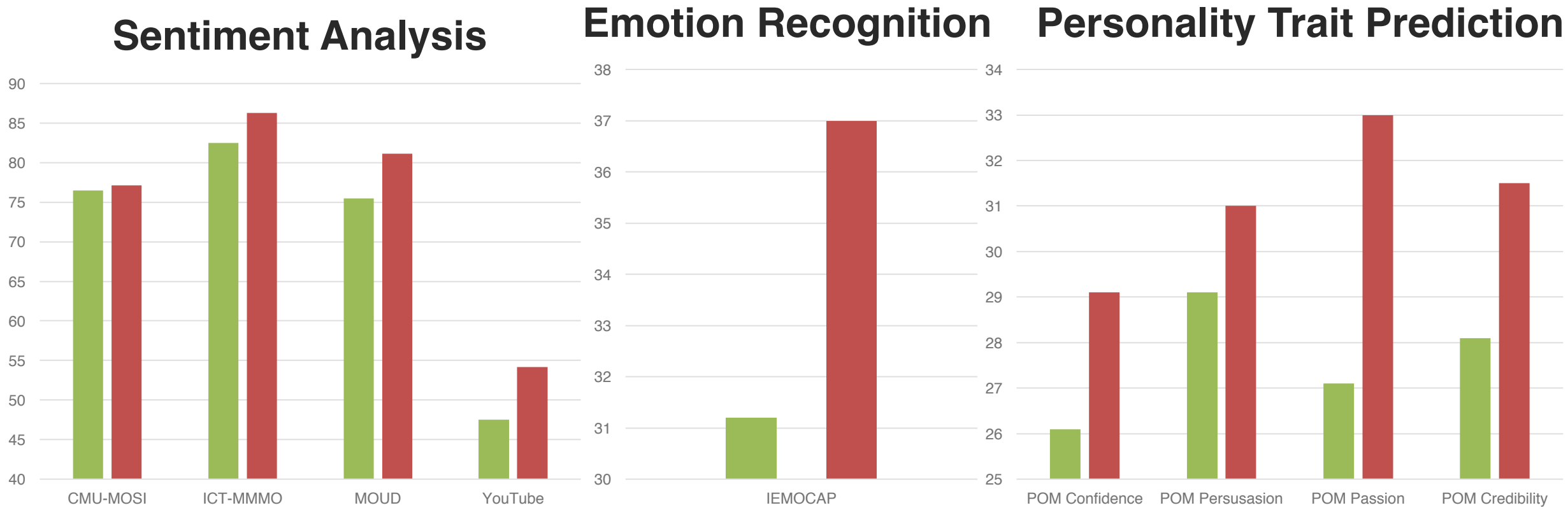
Personality Trait Prediction



State-of-the-art Baseline

Multi-attention Recurrent Network (MARN)

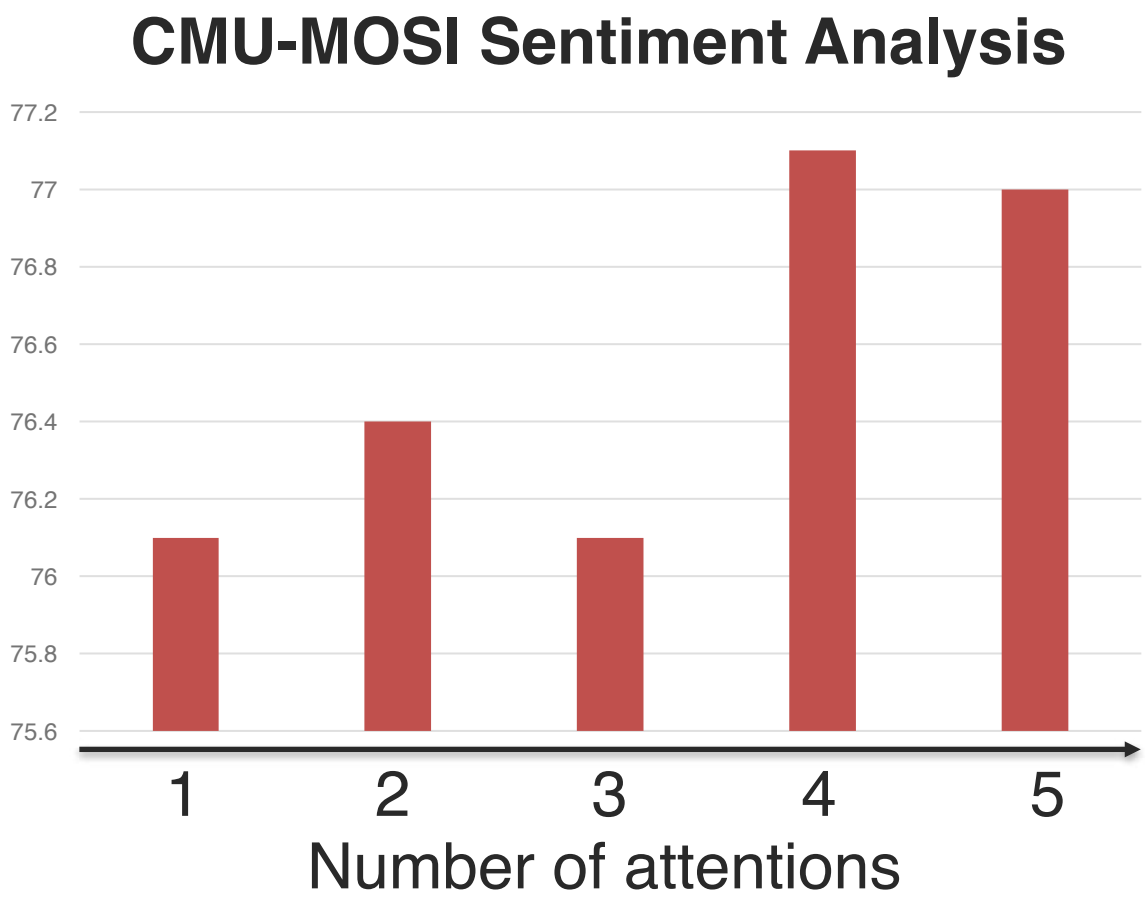
Multi-attention Block is Important

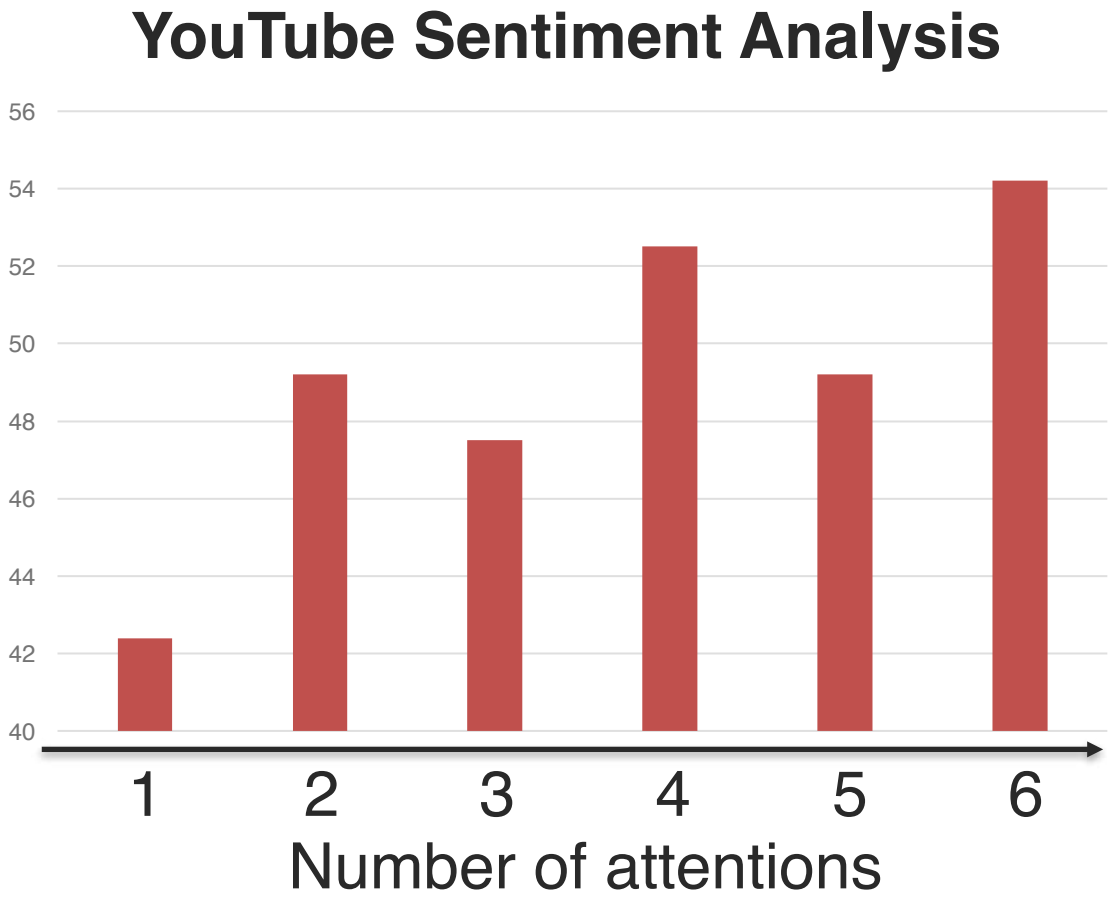


No Multi-attention Block

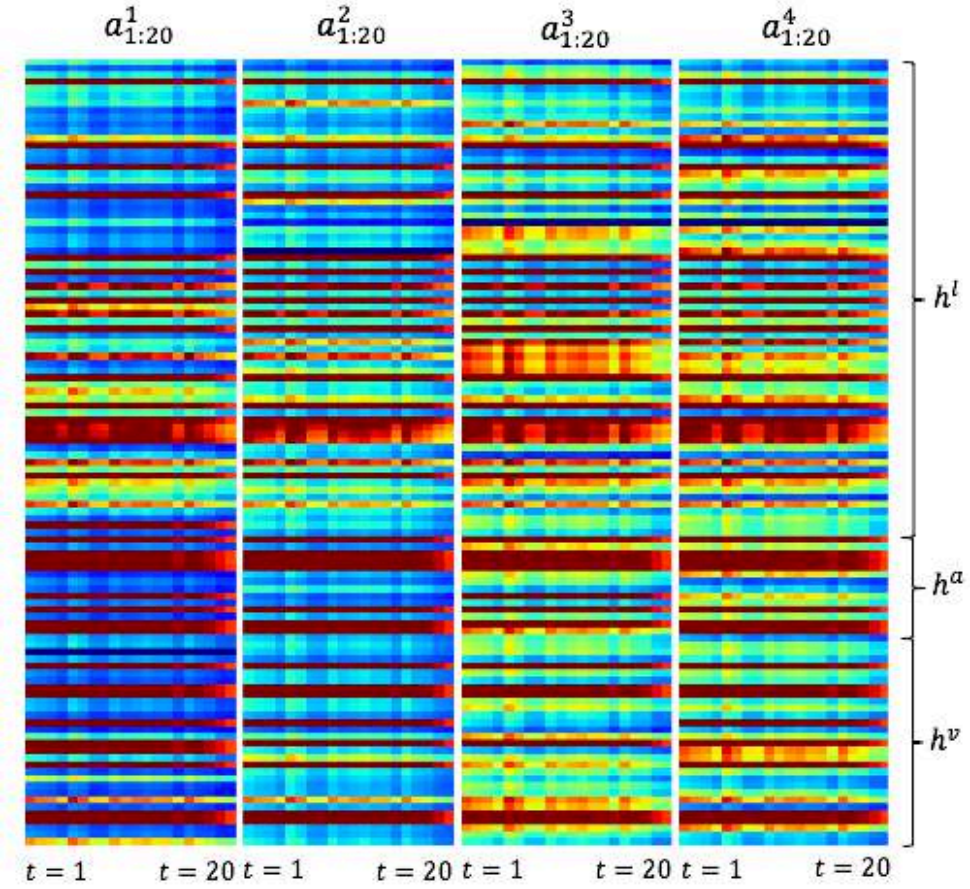
Multi-attention Recurrent Network (MARN)

Multiple Attentions are Important



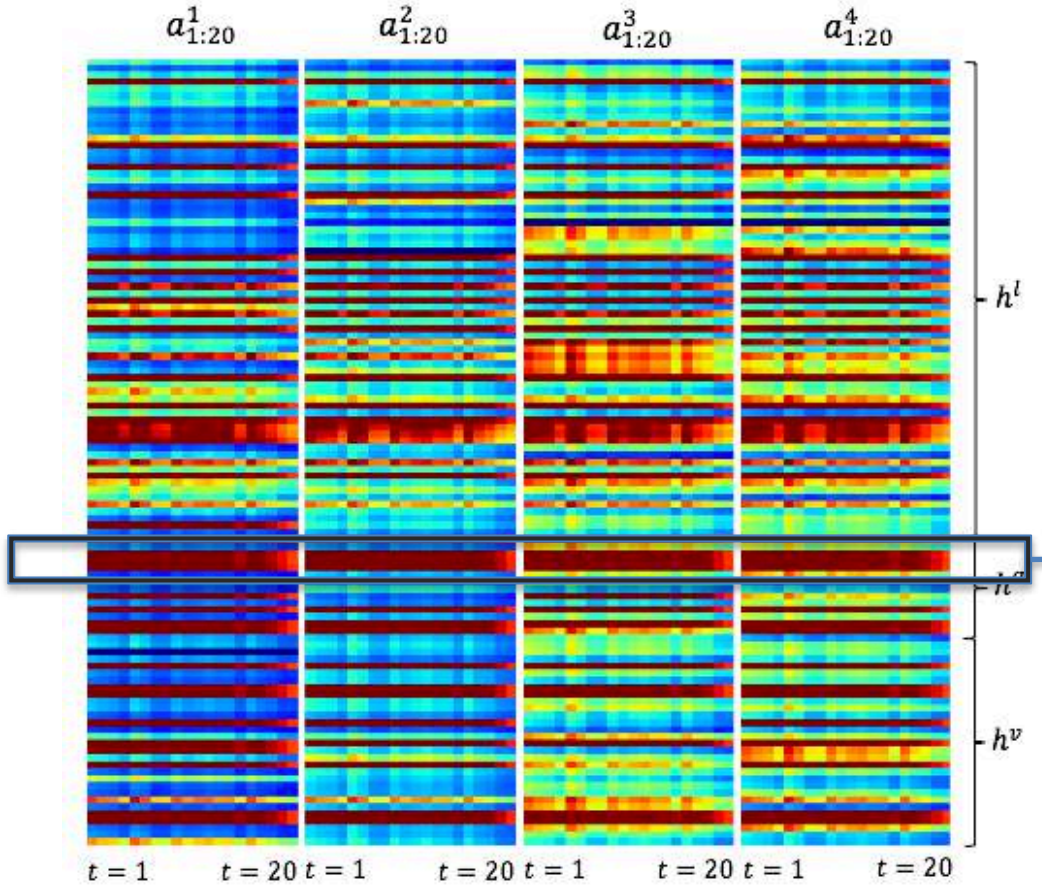


Visualization



Attentions *show diversity* and are sensitive to different cross-modal dynamics

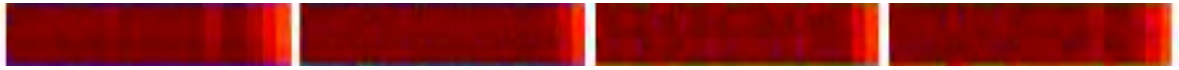
Visualization



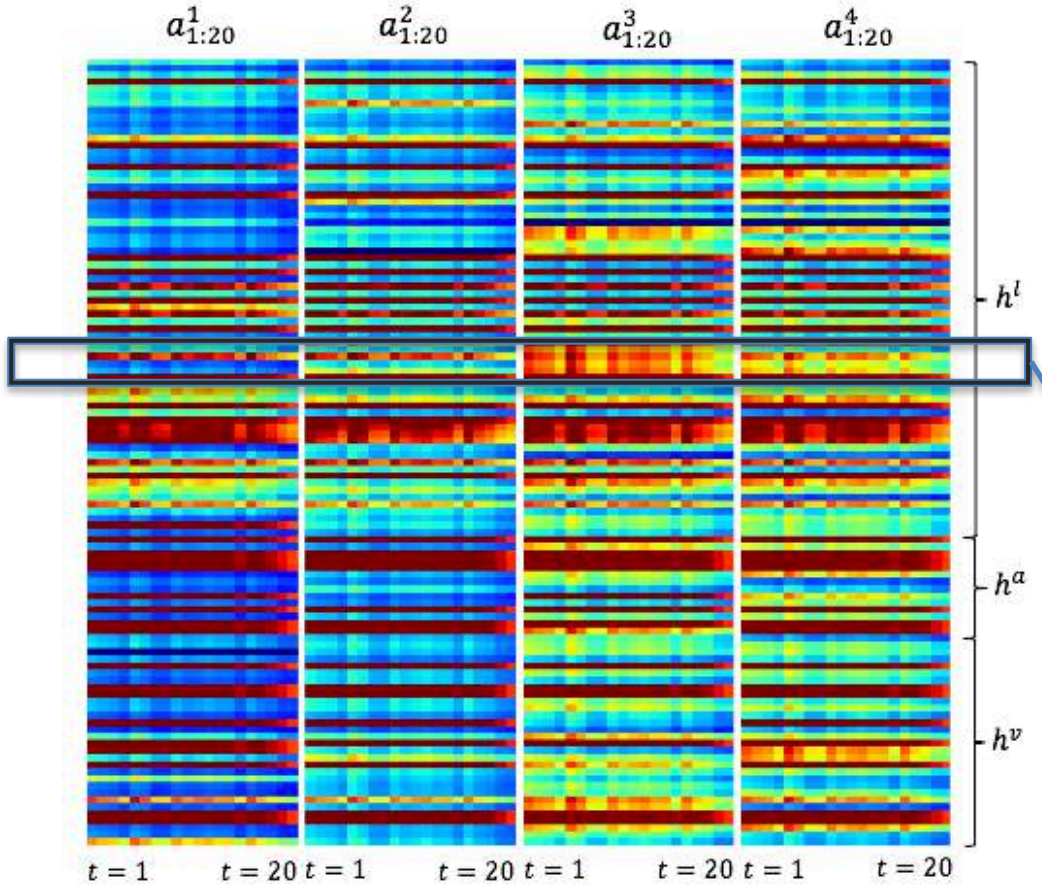
active **inactive**

Some attentions *always inactive*

- Carry only intra-modal dynamics
- No cross-modal dynamics

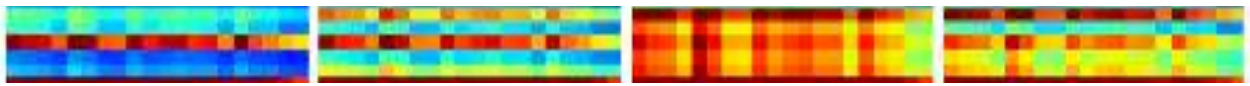


Visualization



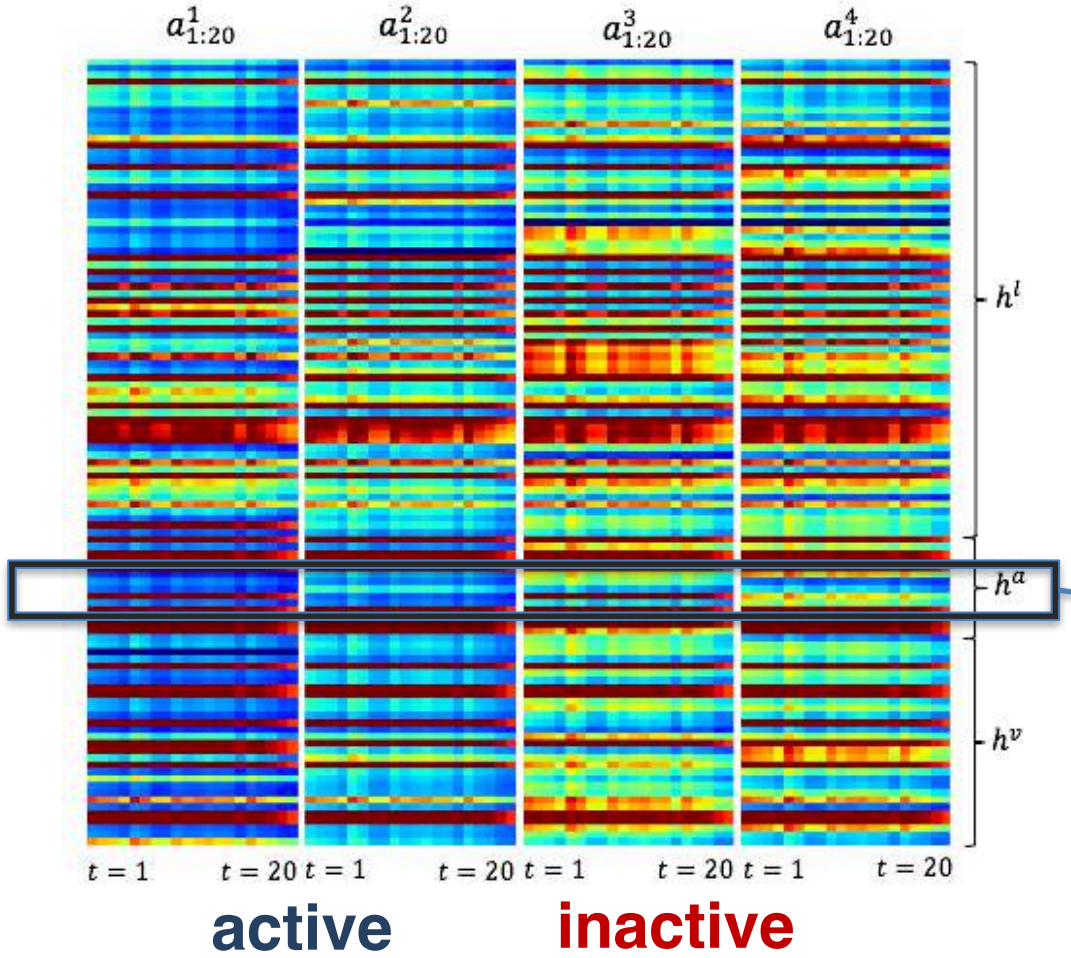
active **inactive**

Attentions *change behaviors across time*, some changes are more drastic than others.

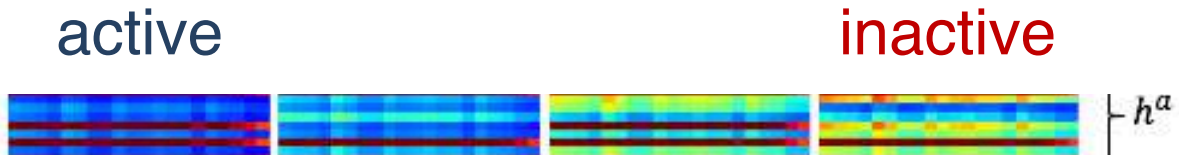


time ⟶

Visualization



Different attentions *focus on different modalities.*



Multi-attention Recurrent Network (MARN)

1 **Modeling intra-modal dynamics**

→ Set of Long-short Term Memories

2 **Modeling cross-modal dynamics**

→ Set of Long-short Term **Hybrid** Memories + Single-attention Block

Modeling multiple cross-modal dynamics

→ Set of Long-short Term **Hybrid** Memories + **Multi-attention** Block

The End!

Code: https://github.com/A2Zadeh/MARN

Email: pliang@cs.cmu.edu

The End!

Code: https://github.com/A2Zadeh/MARN

Email: pliang@cs.cmu.edu

Workshop @ ACL 2018

First Workshop on Computational Modeling of Human Multimodal Language

multicomp.cs.cmu.edu/acl2018multimodalchallenge/