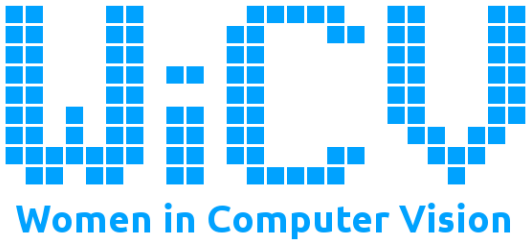


Nonverbal communication cue recognition

A pathway to more accessible communication

Zoya Shafique, Haiyan Wang, Yingli Tian

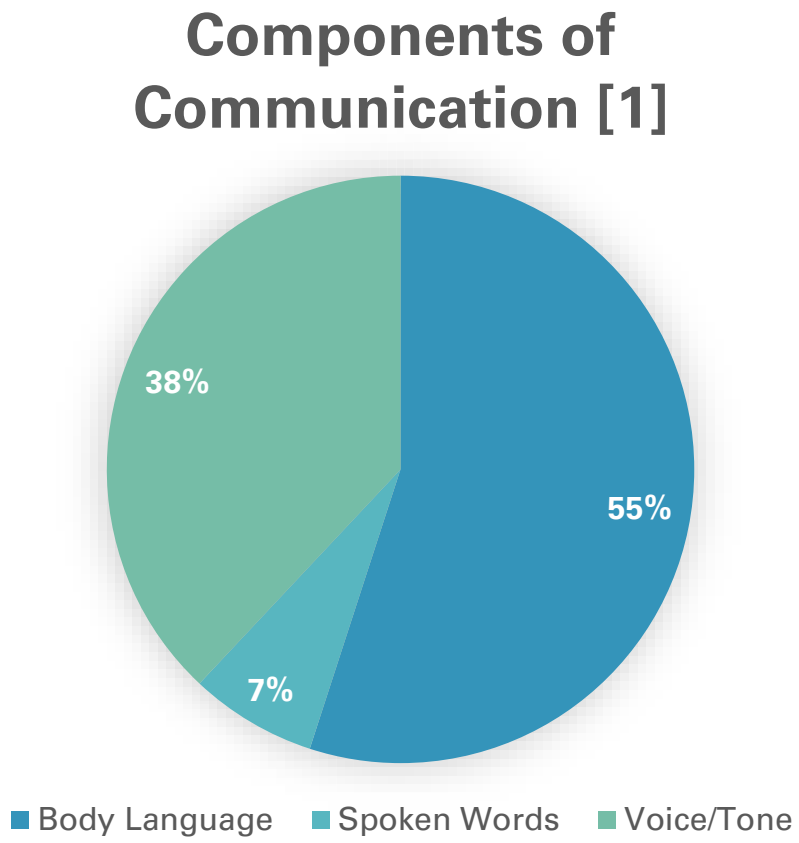
zshafiq001@citymail.cuny.edu

The City College of New York, New York

Introduction

Motivation

- **2.2 billion** people worldwide have some form of vision impairment [2]
- The blind and low vision (BLV) community may understand other people's feelings differently than sighted people as they cannot perceive nonverbal communication cues (NVCs) [3]



[1] Mark L Knapp et al. Nonverbal communication in human interaction. 2013
[2] Blindness and Vision Impairment. WHO.
[3] Jolanta Sak-Wernicka. Exploring theory of mind use in blind adults during natural communication. 2016

Introduction

Related Works

Existing NVC Recognition Methods [4-7]

- Mainly focus on facial expression recognition (FER)
- Not multimodal
- Limited scale
- Distracting feedback mechanisms

Existing NVC Datasets [8, 9]

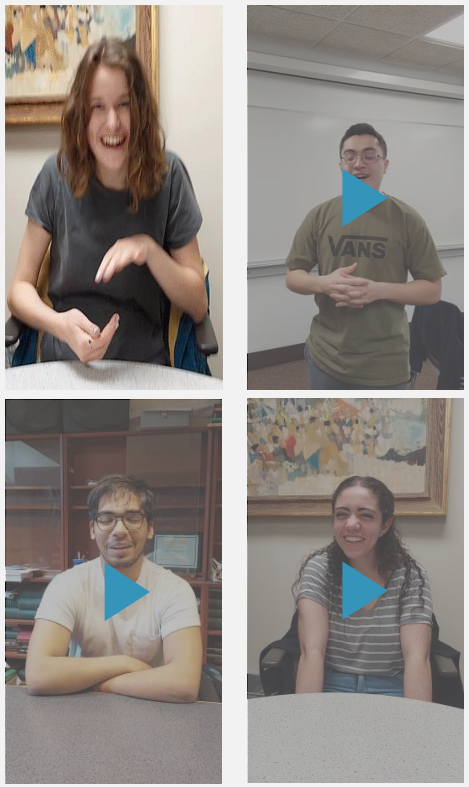
- Focus on seven basic emotions and FER
- Very few are multimodal
- Do not annotate common NVCs
- Many datasets are posed

[4] Anam et al. ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication, 2014.
[5] Astler et al. ACM SIGACCESS Conf. on Computers and Accessibility, 2011.
[6] Krishna et al. CHI'10 Extended Abstracts on Hum. Factors in Computing Systems, 2010.
[7] Shi et al. Proc. of the ACM on Hum.-Comput. Interact., 3(CSCW), 2019.
[8] Liu et al. imigue: An identity-free video dataset for micro-gesture understanding and emotion analysis. CVPR 2021.
[9] Kollias et al. Analysing affective behavior in the second abaw2 competition. ICCV 2021.

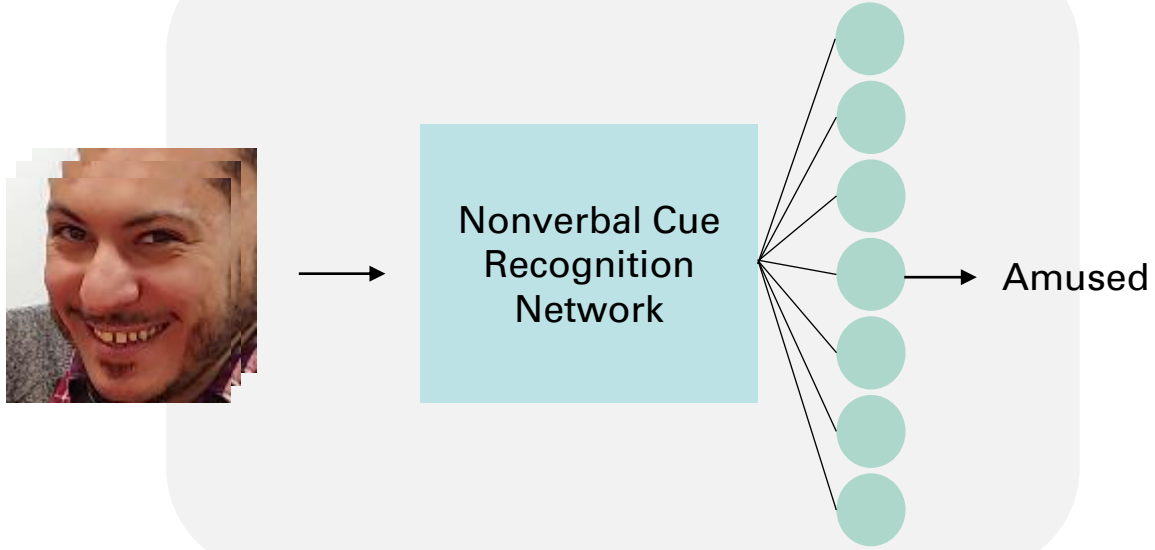
Introduction

Contributions

CCNY NVC Dataset



Baseline Model for NVC Recognition



CCNY NVC Dataset

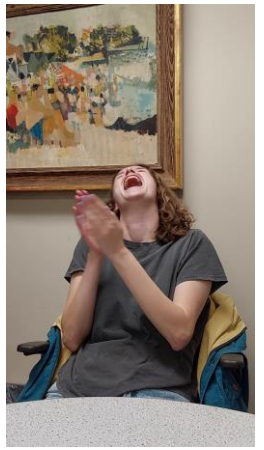
Developing a dataset for nonverbal cue recognition and localization

128 videos

- 12 Subjects
- 10 classes:
  - Thinking, Agreement/Understanding, Surprised, Happy, Upset, Exasperated, Dislike, Disagreement, Amused, Confused

Multimodal annotations

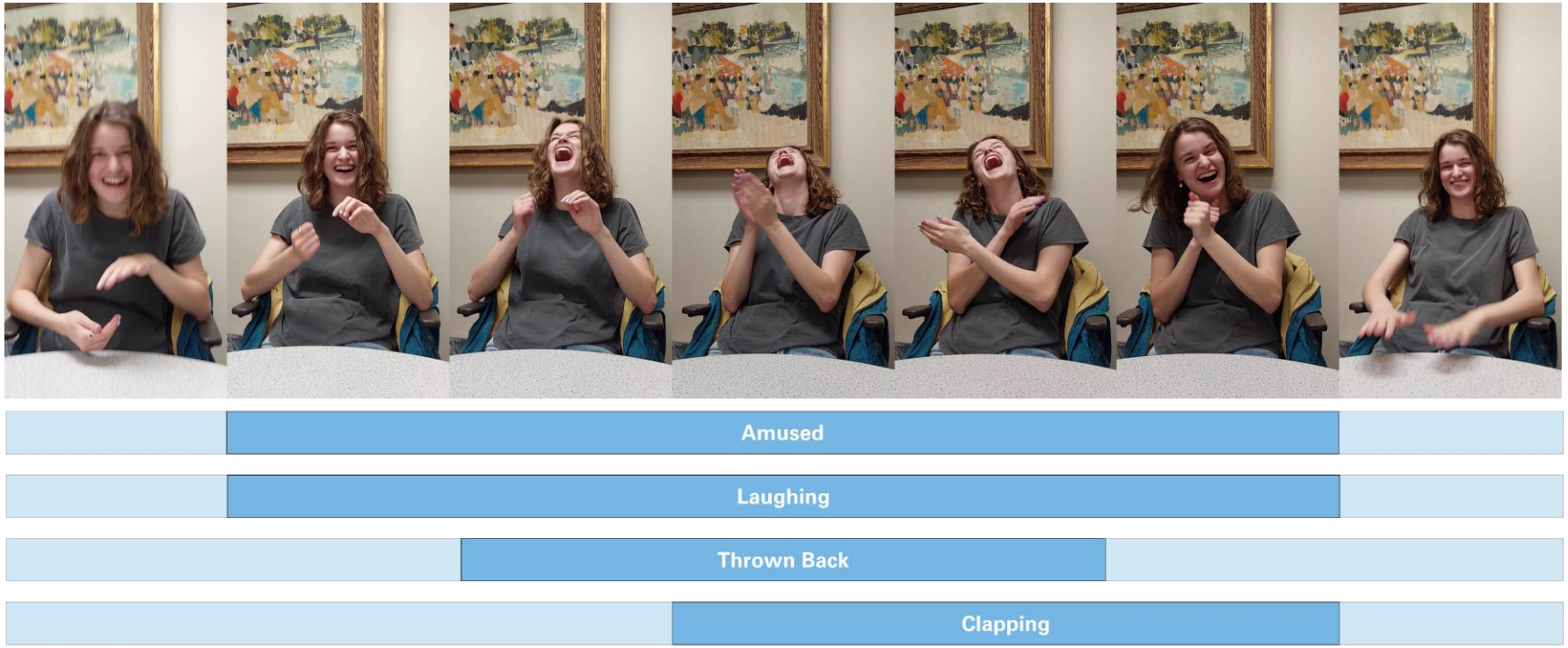
- High level Emotion
- Fine grained actions
  - Facial expressions, Head movements, Hand gestures



CCNY NVC Dataset

Developing a Dataset for Nonverbal Cue Recognition

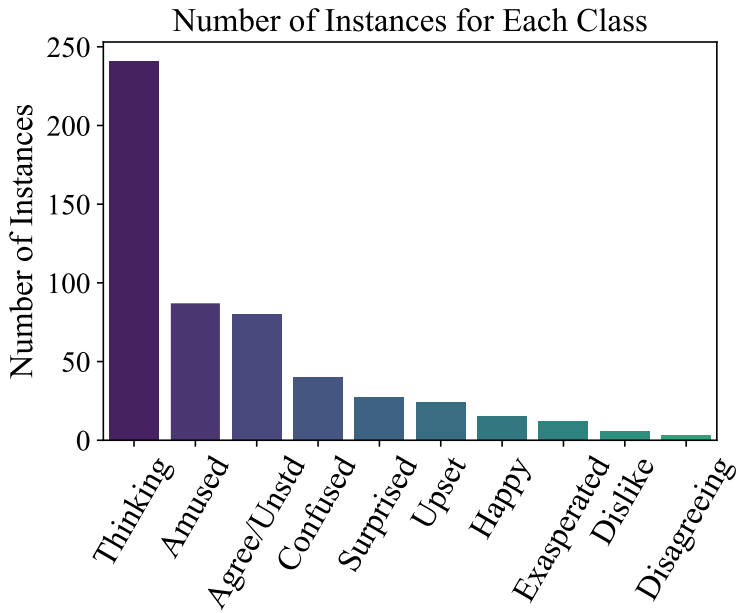
As nonverbal cues are expressed as a mixture of facial expressions, hand movements, and head/body movements, we provide both fine grained action labels in addition to the high-level emotion annotation. Actions are annotated from a start to an end time.

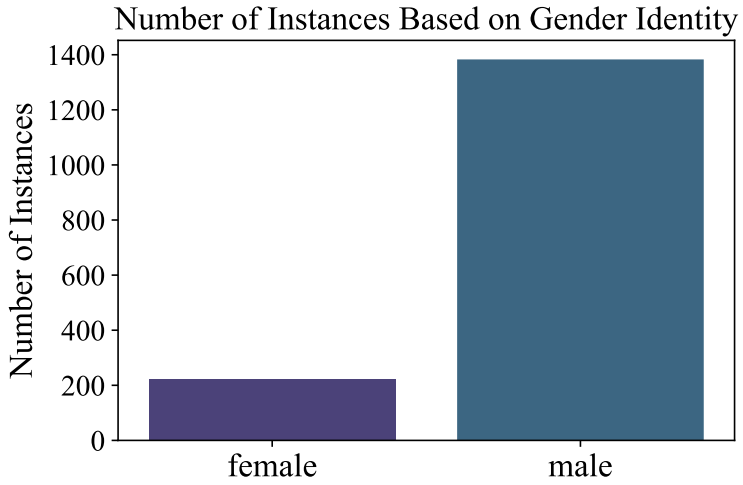


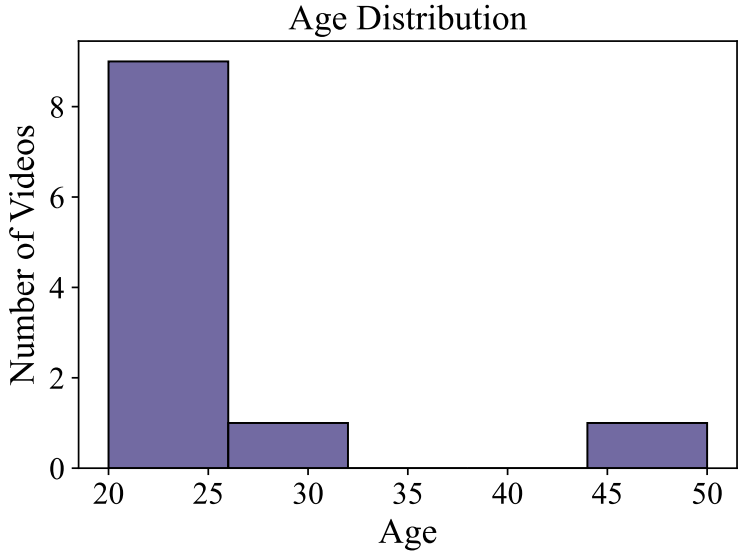
CCNY NVC Dataset

Challenges in Data Curation

It is important to balance not only the classes, but also the gender identities, age, and racial/ethnic backgrounds of the study participants





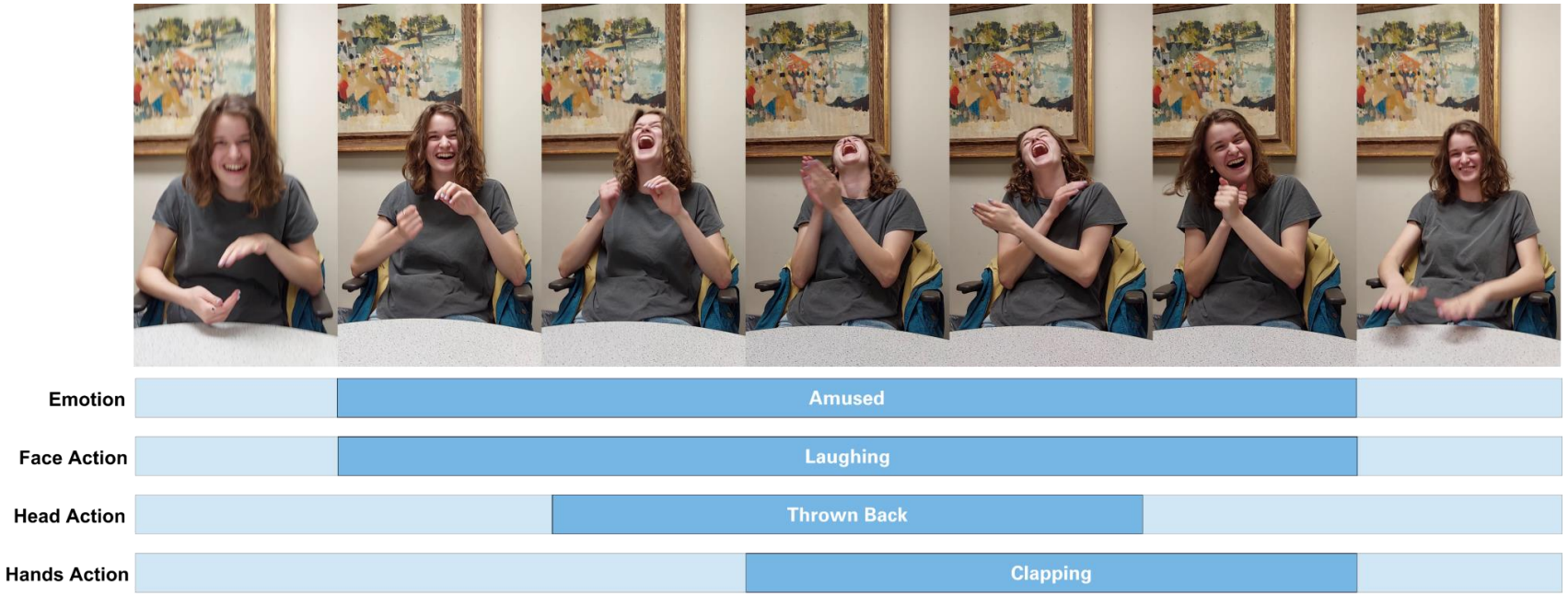


NVC Recognition

Task Definition

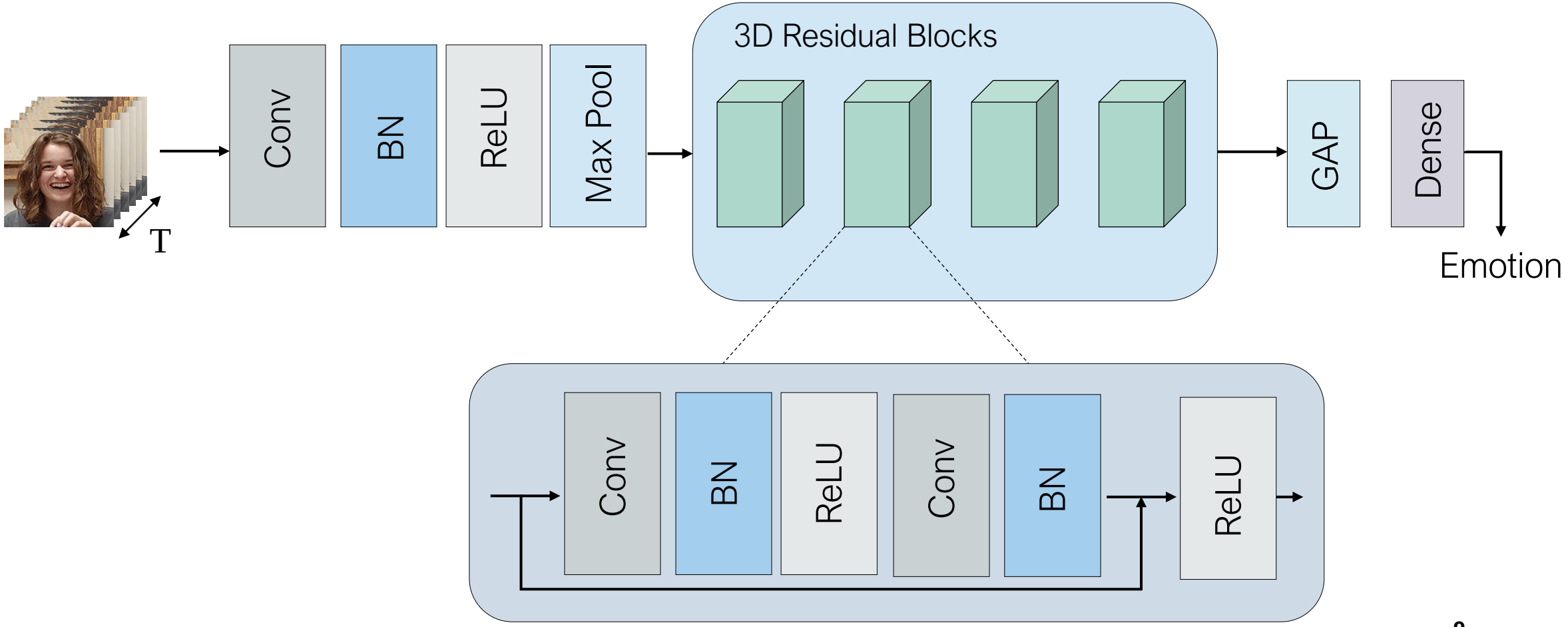
Given a video, can we classify

- the nonverbal gestures (facial expressions, hand gestures, body posture, etc.) and
- the emotions expressed by the nonverbal gestures?



NVC Recognition

Proposed Method: Baseline



NVC Recognition

Proposed Method: Results on Aff-Wild2 Dataset [9]

Our model achieves comparable results with previous SOTA methods on the validation set of the **Aff-Wild2 dataset** [9]

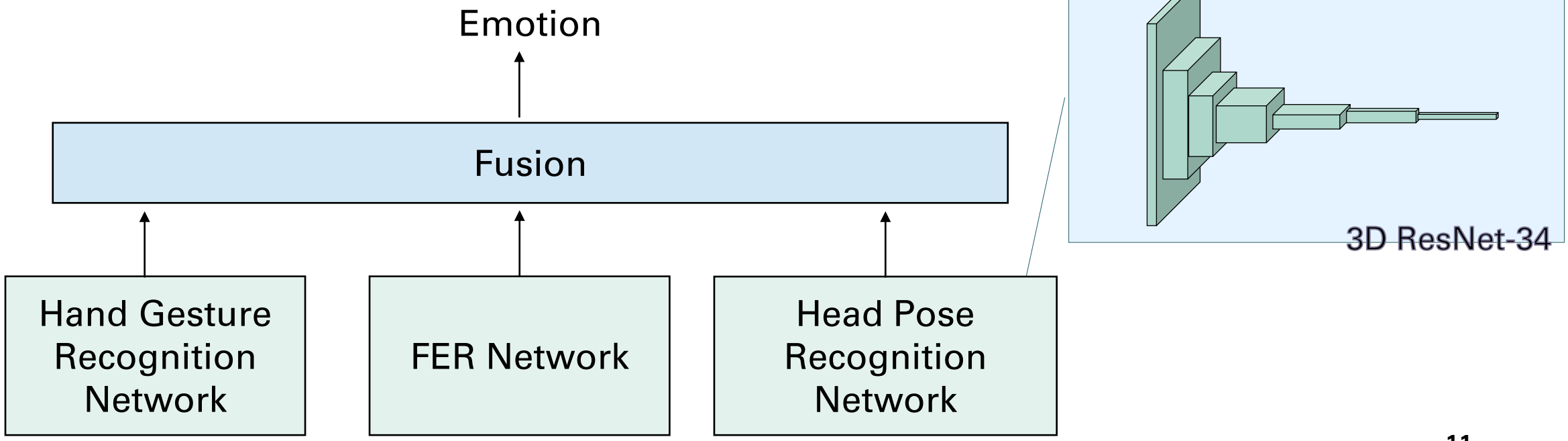
$$\epsilon_{total} = 0.67 \times F_1 + 0.33 \times TAcc$$

Method	F1 Score	Accuracy	ABAW2 Metric
Baseline [9]	30	50	36.6
CPIC-FIR2021 [10]	40.2	63	47.7
Netease Fuxi Virtual Human [11]	**75.7**	**85.6**	**79**
Ours	<u>64.3</u>	<u>68.2</u>	<u>65.6</u>

Conclusion and Future Work

Proposed Method

We aim to extend the FER pipeline to a multimodal framework for NVC cue recognition



Conclusion and Future Work

Closing Remarks

Achieving comparable results on the Aff-Wild2 dataset showcases the validity of our model

We aim to extend the FER network into a multimodal framework for NVC recognition

Our end goal is to create a real time NVC recognition aid for the BLV community

Thank you!

Visit our poster for more details!

Paper ID: 62

We would like to thank NSF for funding this research.