

# MMChat: Multi-Modal Chat Dataset on Social Media

Yinhe Zheng, Guanyi Chen, Xin Liu and Jian Sun

## Multi-modal Dialogues

### Image/Dialogue-grounded

### Dialogue-grounded



Person 1: What is this?

Person 2: I'm cleaning the paintbox that has been put for many years and the dried paint

Person 1: So you are a beauty painter?

Person 1: I learned it when I was young ~ Haven't painted for a long time.

Person 2: Looking forward to your masterpiece.

## MMChat: Basic Information

MMChat is a Multi-Modal Chat Dataset

- Aiming at *sparse image-grounded dialogues*;
- that is in Chinese;
- that consists of dialogues from social media;
- each dialogue in which is associated with multiple images;
- that was manually filtered.

Not every utterance is image-grounded!

## Rule-filtered MMChat

Sessions	4.26M
Sessions (> 4 utterances)	2.30M
Utterances	18.59M
Images	4.87M
Utterance/Session	4.37
Image/Session	1.67
Character/Utterance	14.10

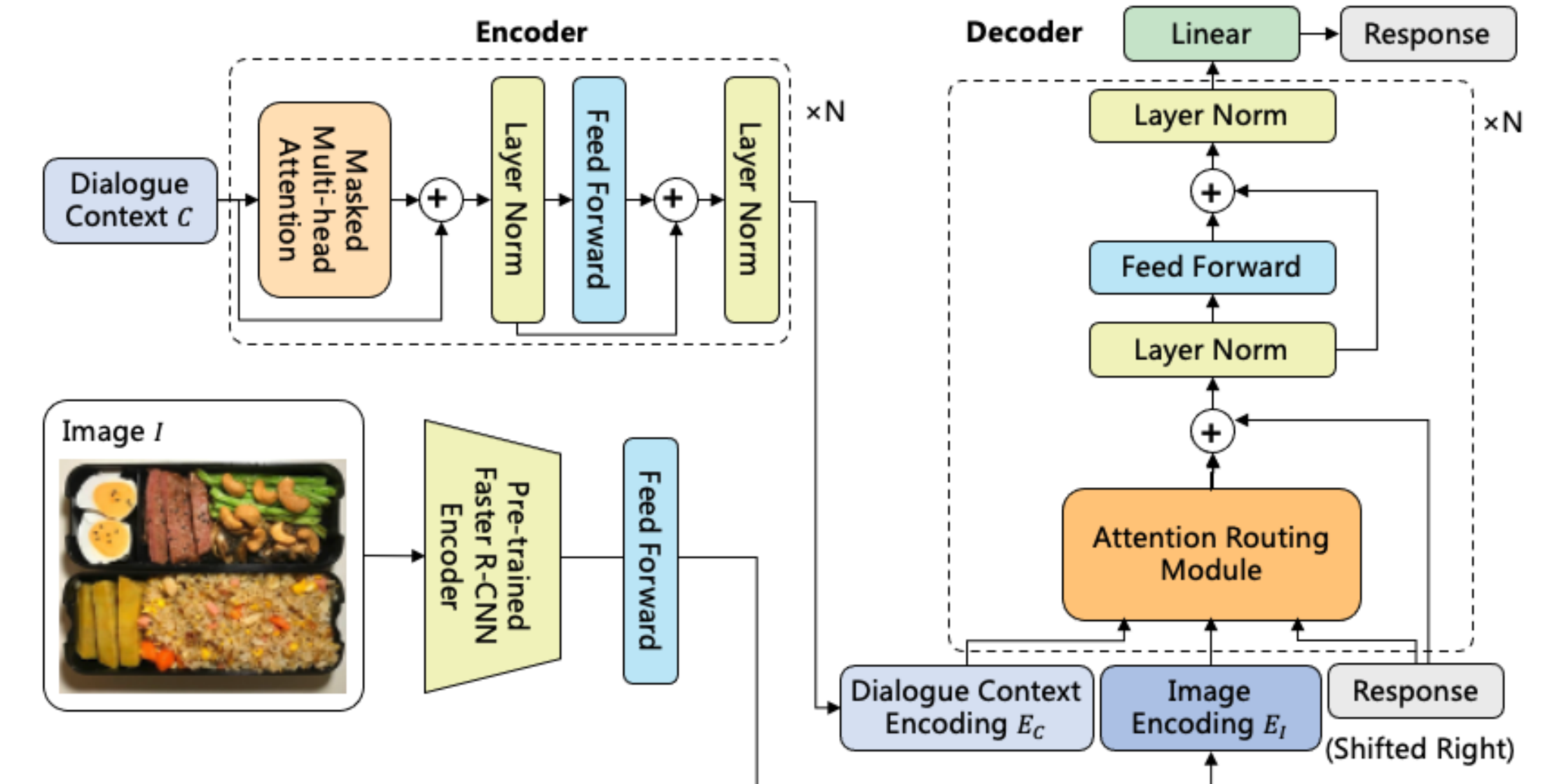
## Rule-filtered MMChat

We sampled 100k dialogues and asked:

- Whether the associated images are qualified?
- Whether the dialogue contents are non-offensive?
- Whether the dialogue contents are strongly correlated with the associated images?

Sessions	19.90K
Sessions (> 4 utterances)	8.91K
Utterances	81.06K
Images	52.66K
Utterance/Session	4.07
Image/Session	2.70
Character/Utterance	11.93

## Modeling MMChat



Please check our dataset at:

<https://github.com/silverriver/MMChat>