Dublin Royal Convention Centre

#ACMMM25

ACM multimedia

Dublin, Ireland **27-31.10.2025**

IntentVC:

Intention-Oriented Controllable Video Captioning

Takahiro Komamizu
Nagoya U.
Japan

Marc A. Kastner
Hiroshima City U.
Japan

Yasutomo Kawanishi
RIKEN
Japan

Trung Thanh Nguyen
Nagoya U.
Japan

Junan Chen
Nagoya U.
Japan

Video is a major data in the world.

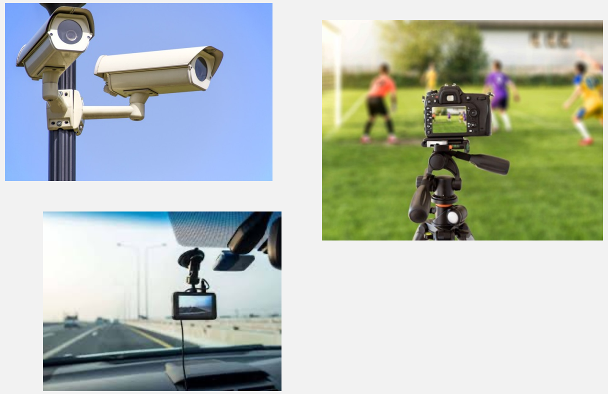
- IDC in 2020* forecasted 80% of data will be video in 2025.

User-generated Contents

YouTube

bilibili

Monitoring & Surveillance



AI-generated Contents

Fi

HeyGen

synthesia

invideo

*Potnis, A., 2024. Managing unstructured data growth requires a fresh approach.
URL: https://www. quantum. com/globalassets/documents/idc-vendor-spotlight. pdf

With such massive video data, what we should do...

Real-Time Performance

Reducing latency in distributed systems.

Video Enhancement & Restoration

Developing adaptive enhancement algorithms.

Big Data Scalability

Managing massive video volumes efficiently for storage and retrieval.

Semantic Understanding & Retrieval

Overcoming unstructured, high-dimensional video data.

Specialized Domains (e.g., Sports)

Addressing domain-specific challenges and datasets.

High Computational Cost

Deep learning models (CNNs, RNNs, Transformers) require heavy resources.

Privacy & Security

Developing secure yet efficient processing techniques.

Human-Centered Optimization

Aligning machine video analysis with human visual perception.

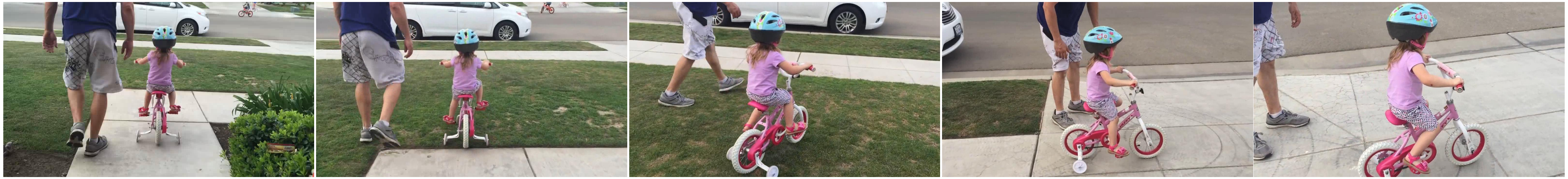
Intention-oriented Video Captioning

General-purpose VC

Describe all the contents
as much as a model can.

Intention-oriented VC

Describe the content specified
by a user in a video.



(a) **Non-Intention-Oriented**
Caption: A child rides a bicycle with an adult walking alongside on a sunny day in a neighborhood.

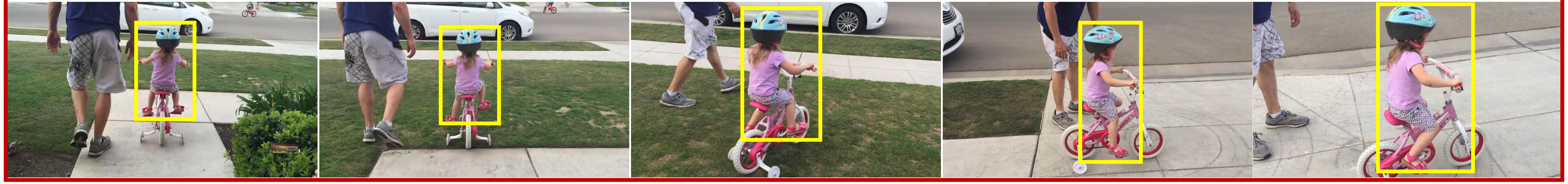
(b) **Intention-Oriented Object:** Person
Caption: *A young child wearing a helmet* learns to ride a bicycle, guided by an adult for support.

(c) **Intention-Oriented Object:** Bicycle
Caption: *A small bicycle with training wheels* is ridden by a child, carefully supported by an adult along a sidewalk.

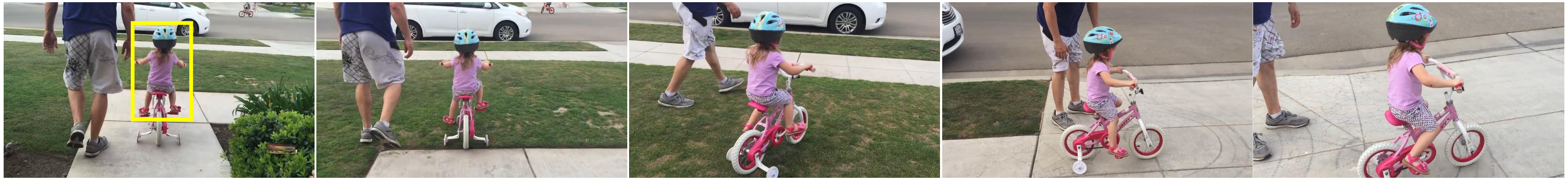
IntentVC – Our Challenge

2025 Edition

Level 1: Given a video and an intention tracklet (a series of bounding boxes for the same object)



Level 2: Given a video and an intention bounding box in the first frame



Level 3: Given a video and an intention text specification (e.g., **'A child'**)

Dataset and Leaders Board

Dataset

- Based on LaSOT dataset*.
  - http://vision.cs.stonybrook.edu/~lasot/

- 70 classes of objects
- Semi-manual Annotation
- Test sets: public and private

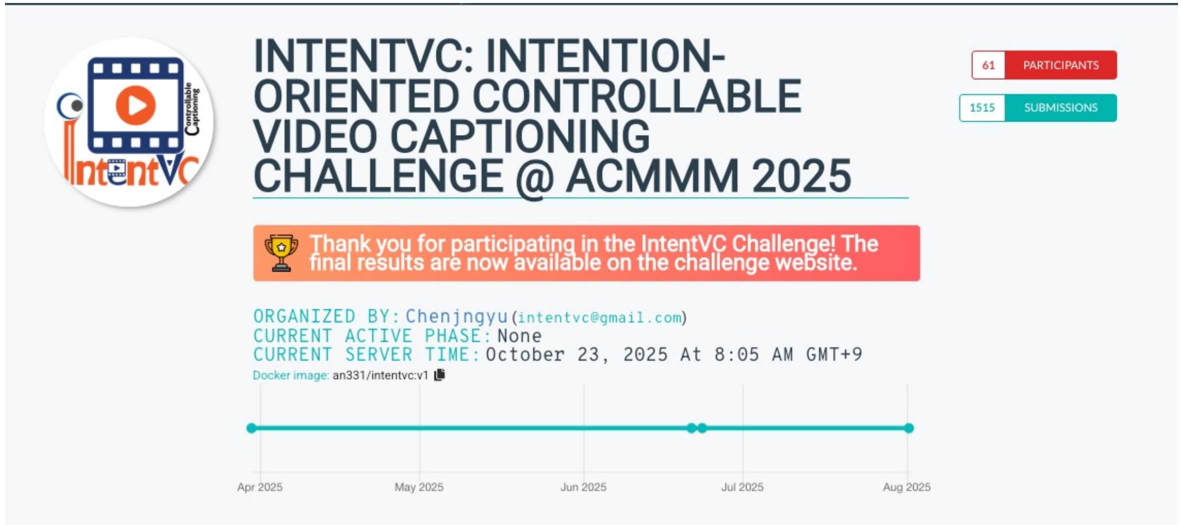
Metrics

- CIDEr, METEOR, BLEU@4

* H. Fan, H. Bai, L. Lin, F. Yang, P. Chu, G. Deng, S. Yu, Harshit, M. Huang, J Liu, Y. Xu, C. Liao, L Yuan, and H. Ling, LaSOT: A High-quality Large-scale Single Object Tracking Benchmark, International Journal of Computer Vision (IJCV), 129: 439–461, 2021.

Leaders Board: Codebench

https://www.codabench.org/competitions/6923/



INTENTVC: INTENTION-ORIENTED CONTROLLABLE VIDEO CAPTIONING CHALLENGE @ ACMMM 2025

61 PARTICIPANTS

1515 SUBMISSIONS

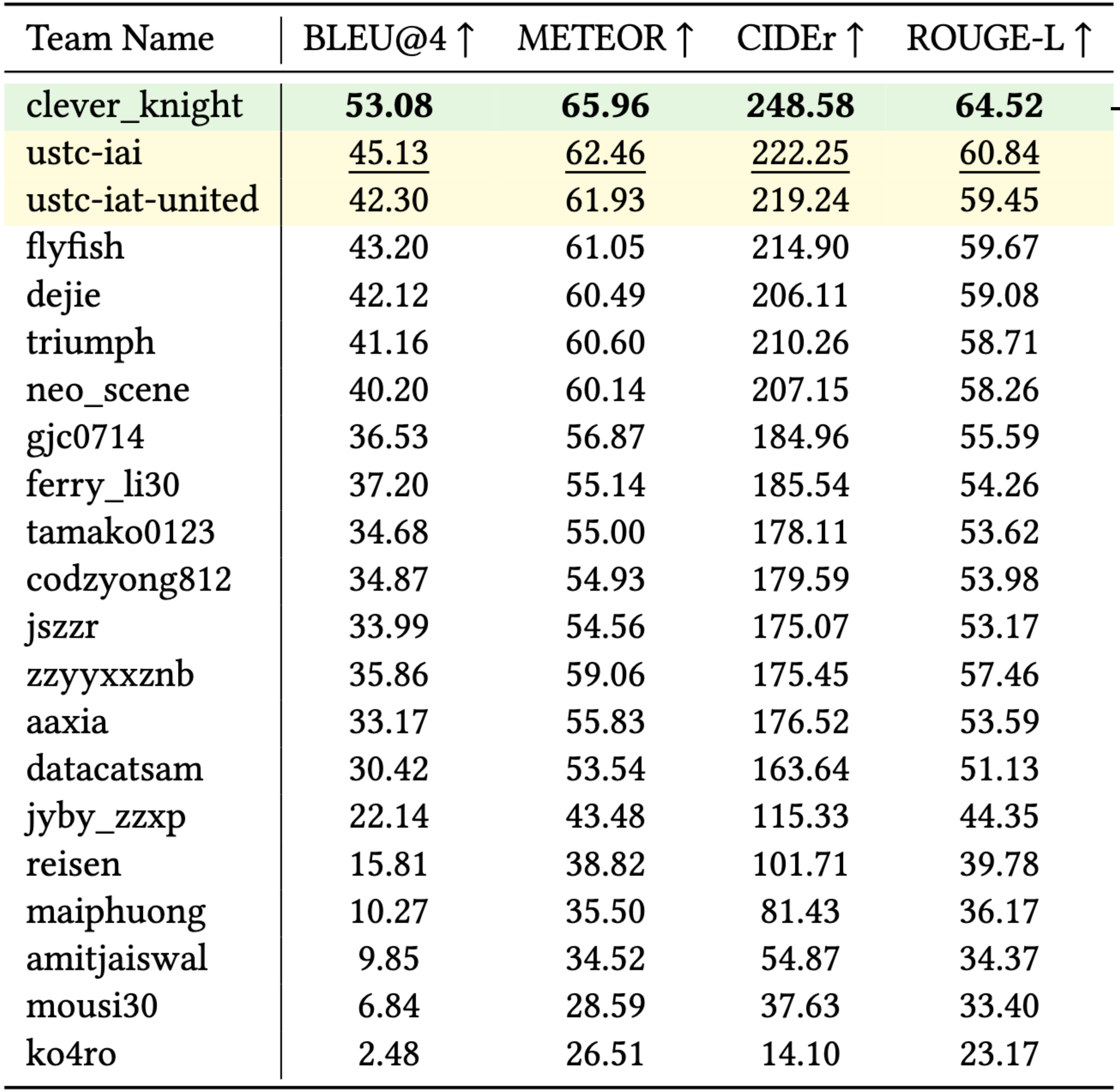
Thank you for participating in the IntentVC Challenge! The final results are now available on the challenge website.

ORGANIZED BY: Chenjngyu (intentvc@gmail.com)
CURRENT ACTIVE PHASE: None
CURRENT SERVER TIME: October 23, 2025 At 8:05 AM GMT+9
Docker image: an331/intentvc:v1

Participants

- 23 Teams, 58 participants
- 1,443 entries

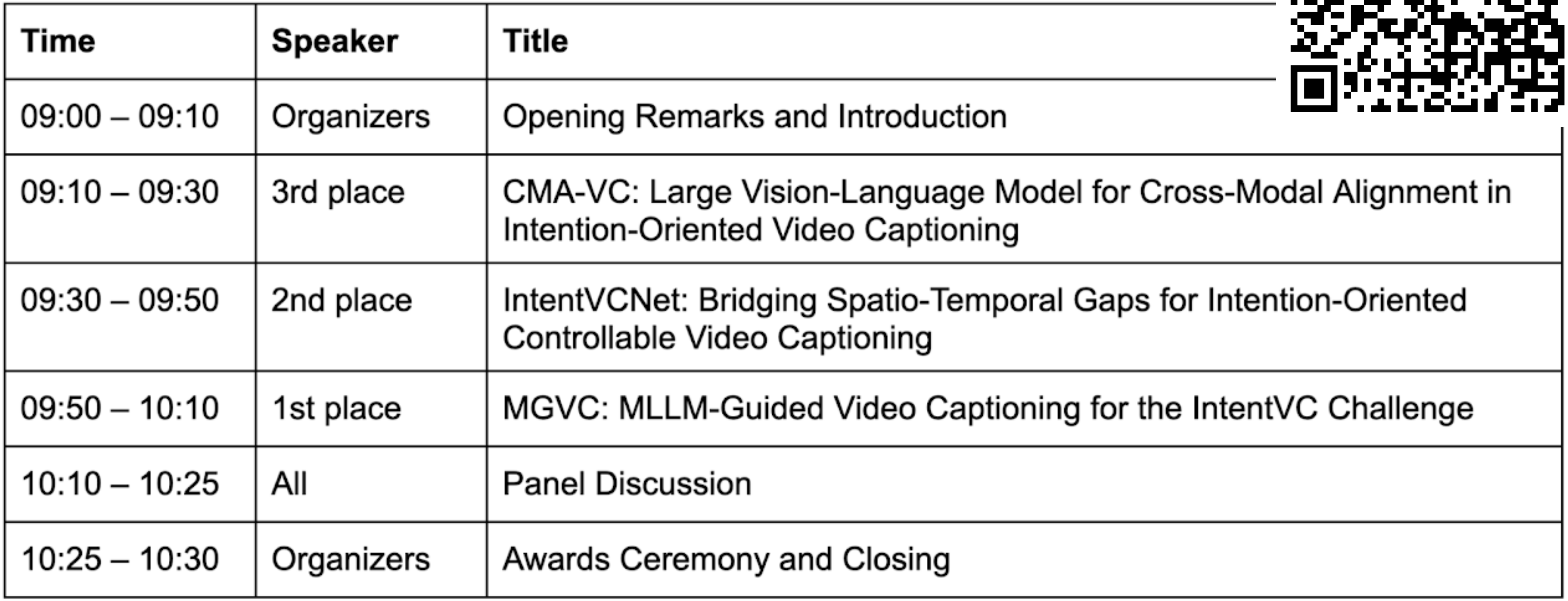
Results



Team Name	BLEU@4 ↑	METEOR ↑	CIDEr ↑	ROUGE-L ↑
clever_knight	**53.08**	**65.96**	**248.58**	**64.52**
ustc-iai	45.13	62.46	222.25	60.84
ustc-iat-united	42.30	61.93	219.24	59.45
flyfish	43.20	61.05	214.90	59.67
dejie	42.12	60.49	206.11	59.08
triumph	41.16	60.60	210.26	58.71
neo_scene	40.20	60.14	207.15	58.26
gjc0714	36.53	56.87	184.96	55.59
ferry_li30	37.20	55.14	185.54	54.26
tamako0123	34.68	55.00	178.11	53.62
codzyong812	34.87	54.93	179.59	53.98
jszzr	33.99	54.56	175.07	53.17
zzyyxxznb	35.86	59.06	175.45	57.46
aaxia	33.17	55.83	176.52	53.59
datacatsam	30.42	53.54	163.64	51.13
jyby_zzxp	22.14	43.48	115.33	44.35
reisen	15.81	38.82	101.71	39.78
maiphuong	10.27	35.50	81.43	36.17
amitjaiswal	9.85	34.52	54.87	34.37
mousi30	6.84	28.59	37.63	33.40
ko4ro	2.48	26.51	14.10	23.17

renamed → MR-CAS

Session Schedule



Time	Speaker	Title
09:00 – 09:10	Organizers	Opening Remarks and Introduction
09:10 – 09:30	3rd place	CMA-VC: Large Vision-Language Model for Cross-Modal Alignment in Intention-Oriented Video Captioning
09:30 – 09:50	2nd place	IntentVCNet: Bridging Spatio-Temporal Gaps for Intention-Oriented Controllable Video Captioning
09:50 – 10:10	1st place	MGVC: MLLM-Guided Video Captioning for the IntentVC Challenge
10:10 – 10:25	All	Panel Discussion
10:25 – 10:30	Organizers	Awards Ceremony and Closing

https://us02web.zoom.us/j/82619678583?pwd=LHPCP7avg0xVo0kFyJqD78pZgpdkKC.1

IntentVC
Controllable Captioning

Panel Discussion

Takahiro Komamizu
Nagoya U.
Japan

Marc A. Kastner
Hiroshima City U.
Japan

Yasutomo Kawanishi
RIKEN
Japan

Trung Thanh Nguyen
Nagoya U.
Japan

Junan Chen
Nagoya U.
Japan

Problem settings?

Ethical Concerns?

Dataset annotation?

What is the core difficulty in VC?

Foundation models?

Real-time processing?

Evaluation criteria?

IntentVC: Award Ceremony

Takahiro Komamizu
Nagoya U.
Japan

Marc A. Kastner
Hiroshima City U.
Japan

Yasutomo Kawanishi
RIKEN
Japan

Trung Thanh Nguyen
Nagoya U.
Japan

Junan Chen
Nagoya U.
Japan

Third Place Certificate of IntentVC 2025

In recognition of outstanding achievement, this certificate is presented to this team for securing Third Place in IntentVC 2025, one of the Grand Challenges at ACM Multimedia 2025.

Team ustc-iat-united

CMA-VC: Large Vision-Language Model for Cross-Modal Alignment in Intention-Oriented Video Captioning

Jun Yu
University of Science and Technology of China

Xilong Lu
University of Science and Technology of China

Yunxiang Zhang
University of Science and Technology of China

Qiang Ling
University of Science and Technology of China

Dr. Takahiro Komamizu
Lead Organizer, IntentVC 2025
Nagoya University

Dr. Marc A. Kastner
Co-Organizer, IntentVC 2025
Hiroshima City University

Dr. Yasutomo Kawanishi
Co-Organizer, IntentVC 2025
RIKEN

Mr. Trung Thanh Nguyen
Co-Organizer, IntentVC 2025
Nagoya University

Mr. Junan Chen
Co-Organizer, IntentVC 2025
Nagoya University

No.2
RANKING

Second Place Certificate of IntentVC 2025

Controllable Captioning
IntentVC

In recognition of outstanding achievement, this certificate is presented to this team for securing Second Place in IntentVC 2025, one of the Grand Challenges at ACM Multimedia 2025.

Team ustc-iai

IntentVCNet: Bridging Spatio-Temporal Gaps for Intention-Oriented Controllable Video Captioning

Tianheng Qiu
University of Science and Technology of China

Jingchun Gao
University of Science and Technology of China

Jingyu Li
Nanjing University

Huiyi Leong
University of Chicago

Xuan Huang
Chinese Academy of Sciences

Xi Wang
National University of Defense Technology

Xiaocheng Zhang
Harbin Institute of Technology

Kele Xu
National University of Defense Technology

Lan Zhang
University of Science and Technology of China

Dr. Takahiro Komamizu
Lead Organizer, IntentVC 2025
Nagoya University

Dr. Marc A. Kastner
Co-Organizer, IntentVC 2025
Hiroshima City University

Dr. Yasutomo Kawanishi
Co-Organizer, IntentVC 2025
RIKEN

Mr. Trung Thanh Nguyen
Co-Organizer, IntentVC 2025
Nagoya University

Mr. Junan Chen
Co-Organizer, IntentVC 2025
Nagoya University

No.1
RANKING

First Place Certificate of IntentVC 2025

Controllable Captioning
IntentVC

In recognition of outstanding achievement, this certificate is presented to this team for securing First Place in IntentVC 2025, one of the Grand Challenges at ACM Multimedia 2025.

Team MR-CAS

MGVC: MLLM-Guided Video Captioning for the IntentVC Challenge

Zhipeng Yu
SEECE, UCAS

Qianqian Xu
IIP, ICT, CAS

Yangbangyan Jiang
SCST, UCAS

Pinci Yang
SEECE, UCAS

Qingming Huang
SCST, UCAS; IIP, ICT, CAS

Dr. Takahiro Komamizu
Lead Organizer, IntentVC 2025
Nagoya University

Dr. Marc A. Kastner
Co-Organizer, IntentVC 2025
Hiroshima City University

Dr. Yasutomo Kawanishi
Co-Organizer, IntentVC 2025
RIKEN

Mr. Trung Thanh Nguyen
Co-Organizer, IntentVC 2025
Nagoya University

Mr. Junan Chen
Co-Organizer, IntentVC 2025
Nagoya University

IntentVC: Closing

Takahiro Komamizu
Nagoya U.
Japan

Marc A. Kastner
Hiroshima City U.
Japan

Yasutomo Kawanishi
RIKEN
Japan

Trung Thanh Nguyen
Nagoya U.
Japan

Junan Chen
Nagoya U.
Japan