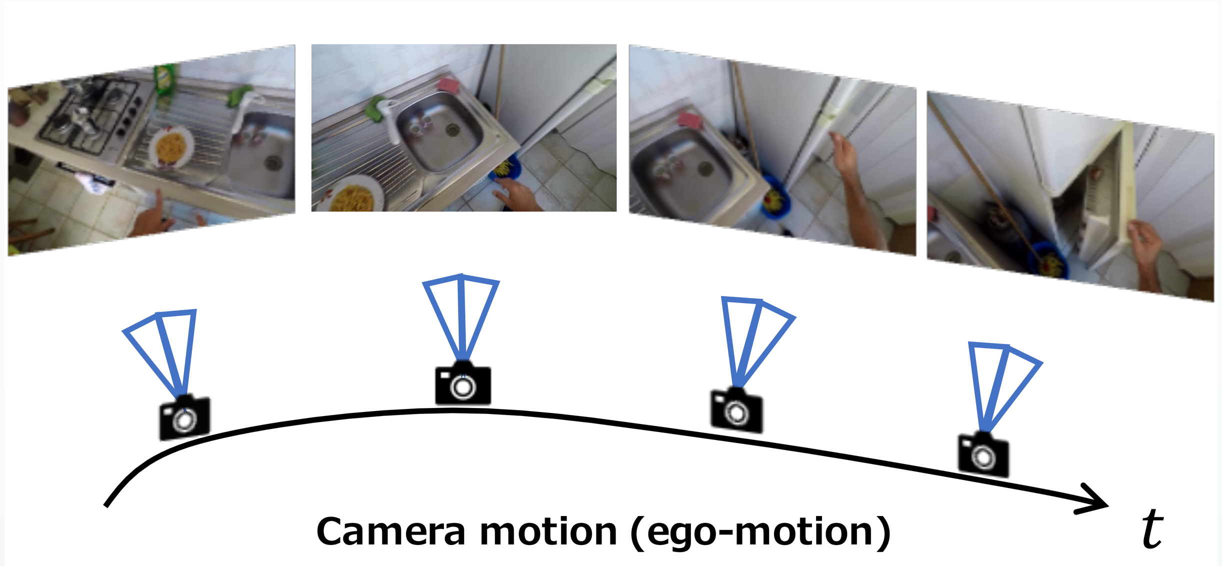
EMAG: Ego-motion Aware and Generalizable 2D Hand Forecasting from Egocentric Videos

Masashi Hatano[1] Ryo Hachiuma[2] Hideo Saito[1]
[1]Keio University [2]NVIDIA

Overview

- ➢ First work that investigates the potential benefits of incorporating ego-motion in egocentric 2D hand forecasting task
- ➢ Propose simple but effective approach, EMAG
- ➢ Validated on Ego4D and EPIC-Kitchens in intra and cross-dataset scenarios

Background



Issues

1. Ego-motion incorporation
- Take ego-motion as input
- Forecasting future ego-motion

2. Generalization ability
- Robustness to novel scene

Task Definition

Task: 2D Hand Forecasting
Observation: 2s, Forecsting: 1s

x_{p1}: frame 0.25s after the last observed frame
x_{p2}: frame of 0.5s after
x_{p3}: frame of 0.75s after
x_{p4}: frame of 1s after

Predict hand location $(\hat{h}_i^l, \hat{h}_i^r)$, where $i \in \{p1, p2, p3, p4\}$, on 2D image coordinate

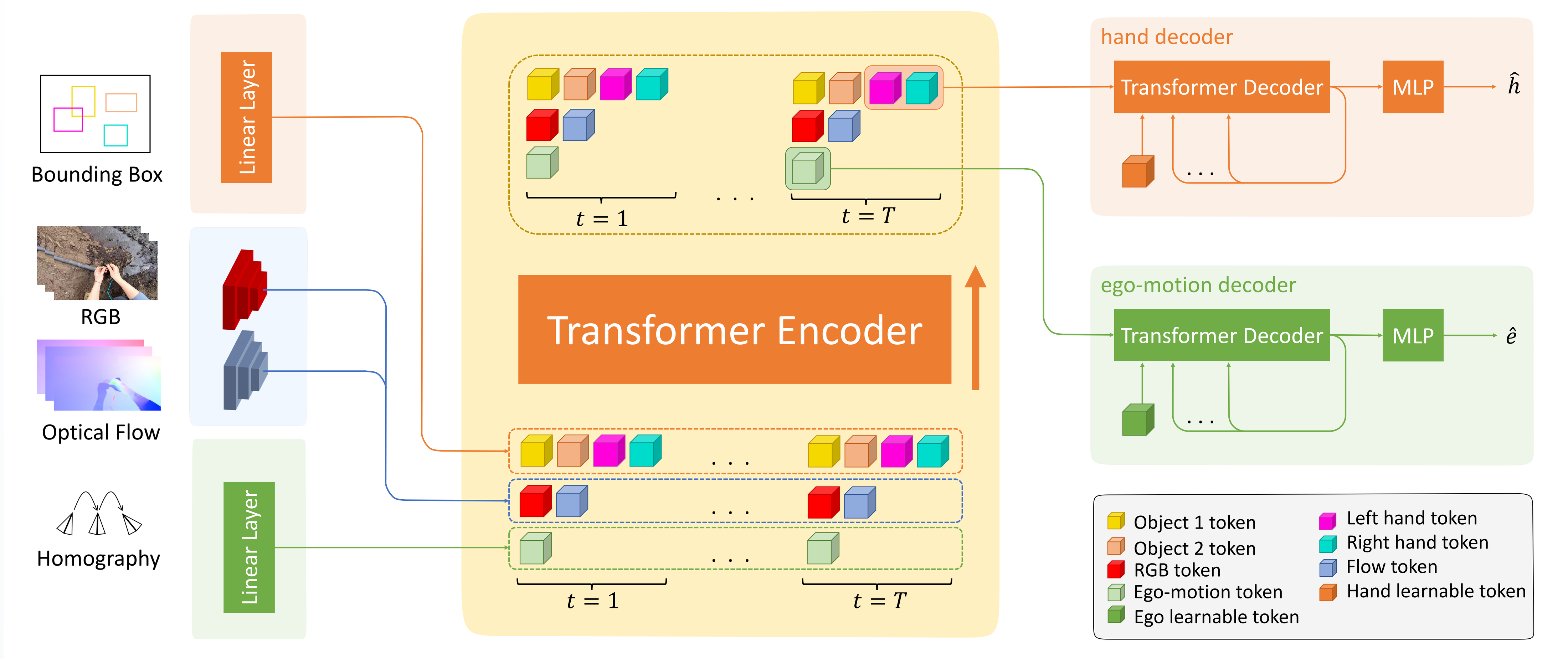
Limitations & Future Work

- ❖ Forecasting in 3D
- ❖ More interesting scenario where hands are out-of-view in several frames during observation

References

[1] Schöller et al, What the constant velocity model can teach us about pedestrian motion prediction, RA-L (2020)
[2] Kalman et al, A new approach to linear filtering and prediction problems, Journal of Basic Engineering (1960)
[3] Sutskever et al, Sequence to sequence learning with neural networks, NeurIPS (2014)
[4] Liu et al, Joint Hand Motion and Interaction Hotspots Prediction from Egocentric Videos, CVPR(2022)
[5] Grauman et al, Ego4D: Around the World in 3,000 Hours of Egocentric Video, CVPR(2022)

Proposed Method



Experimental Results

Intra & Cross-Dataset Evaluation

Method	Ego4D → Ego4D		EPIC → EPIC		EPIC → Ego4D		Ego4D → EPIC	
	ADE ↓	FDE ↓	ADE ↓	FDE ↓	ADE ↓	FDE ↓	ADE ↓	FDE ↓
CVM [1]	108.11	143.23	141.70	155.40	108.11	143.23	141.70	155.40
KF [2]	71.23	72.87	70.58	75.60	71.23	72.87	70.58	75.60
Seq2Seq[3]	55.91	60.72	62.24	67.85	62.43	67.85	67.97	72.26
OCT [4]	49.40	54.73	53.85	59.06	57.74	59.10	64.97	65.84
I3D + Regression [5]	49.27	53.04	49.64	54.83	59.72	61.72	51.70	58.37
Ours	**48.99**	**52.83**	**48.78**	**54.03**	**53.67**	**56.36**	**51.03**	**56.78**

Input Ablation

Object	RGB	Flow	Ego	Intra		Cross	
				ADE ↓	FDE ↓	ADE ↓	FDE ↓
	✓	✓	✓	48.76	53.79	52.78	57.02
✓		✓	✓	50.08	54.83	53.30	57.54
✓	✓		✓	51.00	54.78	54,74	57.93
✓	✓	✓		**48.35**	**53.24**	52.89	57.02
✓	✓	✓	✓	48.89	53.43	**52.35**	**56.57**

Loss Ablation

Method	Intra		Cross	
	ADE ↓	FDE ↓	ADE ↓	FDE ↓
w/o $\mathcal{L}_{ego}$	49.66	54.26	52.84	57.08
w/ $\mathcal{L}_{ego}$ (Ours)	**48.89**	**53.43**	**52.35**	**56.57**

Qualitative Results