# Leveraging Temporal Context in Low Representational Power Regimes WED-AM-235

#### Camilo Fosco, SouYoung Jin, Emilie Josephs, Aude Oliva **Olivalab**, MIT CSAIL





# The problem

Using temporal information is crucial to understand videos. Yet, current models don't explicitly attempt to leverage temporal regularities in datasets with long videos.

- Can we leverage the statistics in temporal sequences of video datasets to improve performance in downstream tasks?
- Can we build richer embeddings with this information?
- Where does this type of information help the most?



#### The idea

- We propose to build an **Event Transition Matrix**: a representation that captures typical transition probabilities between actions in long video sequences
- We use this matrix as supervision in a new training protocol to generate **strong embeddings for video snippets**
- We leverage these embeddings to **improve action recognition and action anticipation performance,** especially on low complexity models.



### Key results

Our model-agnostic framework helps **low complexity models improve performance** on action recognition and action anticipation across 3 datasets.





# The Event Transition Matrix

- Computed by looking at **all actions** happening after a given action, weighted by a decay function
- Square matrix, **not symmetric**
- Several postprocessing steps:
  - Dimensionality reduction
  - Decay definition
  - Normalization



#### How can we leverage this ETM?

- We propose to use the rows and columns of the matrix as targets in a regression problem
- An encoder is tasked to generate an embedding that can predict the action + regress the past and future



# Building the ETM

- Large matrix with sparse entries
- **Dimensionality reduction:** do we use the full matrix, or reducing the number of actions considered help?
- **Decay:** how to we weight the contributions of actions happening later in the video?
- **Distance metric:** how do we measure the distance between two actions in a video?

open door	-49.3	1.0 1	14.0 24	5 0.8	2.2	07.3	16.5 0	8 31.	0.4	14 0	3 1	0.2	5.6	5.0 19.5	30.5	0.0	0.1 16	8 0.0	1.4	5.2 (	0.0	14	0.0 0	0.0 0.0	0.0	0.0 7	2 0.0	0.0	1.6 0.0	0.0	0.0 0.0	0.0	0.0	0.0	0.0	0.9 0.	0 1.0	0.0
turn on light	- 0.0	0.0	1.0 5.	2 0.8	0.8	0.0	2.2 0	7 81	0.5	7.9 6	0.3 0.9	2.2	0.3	19 24	13.3	0.0	0.0 6	2 0.0	0.0	43 0	0.0	3.5	0.0 0	0.0 0.0	0.0	0.0 0	0.0	0.0	16 17	0.0	00 0	0.0	0.0	0.0	0.0	0.0 0.	0 0.4	0.3
close door	- 39.0	0.0 2	2.5 15	6 0.9	27	0.8	12.6 1	1 19.9	0.5	11.7 1	18 1	6.4	5.6	2.9 12.9	27.3	01	0.3 16	2 0.0	2.5	0.9 0	0.0	0.3	0.0 0	0.0 0.0	0.6	0.0 3	1 0.0	0.0	10 0.8	0.0	0.0 0.	0.0	0.9 0	10 0.0	0.0	1.2 0.	0 01	0.0
open hidge	10.0	48	6.6 78	2 1.0	5.3	0.9 4	68.1 6	3 139	1 2.6	115.6 3	15 63	61.7	19.0	19.9 88.9	9336.0	2.0	1.4 65	1 30	11	5.6 1	3 0.7	7.0	0.0 0	0.0 0.0	40	2.0 0	5 0.8	0.0	14 42	0.0	0.0 0.0	0.0	0.9 0	0.0	0.9	91 4	9 5.5	4.9
take cellery	0.0	0.0	0.0 1	8 0.1	1.0	1.0	2.5 0	9 12	0.6	0.5 0	15 0.3	0.4	10	0.6 0.3	1.2	0.3	0.0 0	7 0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0 0.0	1.1	0.0 0	0.0	0.0	20 0.6	19	07 1	1 0.2	0.2 1	.0 0.9	0.7	0.5 1	0 0.4	0.4
take container	0.0	0.0	0.0 E	1 0.0	7.2	1.0	10.6 2	4 9.7	0.7	11.8 1	17 1	10.4	2.1	41 57	2.5	0.4	0.0 2	8 21	0.2	01 0	10 0.0	0.0	0.0 0	0.0 0.0	5.3	0.0 0	0.0	0.0	10 32	0.0	00 0	0.0	0.0 0	10 0.0	0.0	9.3 2	6 0.0	55
take tota	0.0	0.0	0.0 1	0.0	0.0	0.0	18 0	9 14	0.7	0.6 0	0.6 0.9	0.5	1.8	0.7 0.4	0.3	05	0.0 0	0 0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0 0.0	0.0	0.0 0	0.0	0.0	20 0.3	0.0	0.0 0.0	0.0	0.0 0	10 0.0	0.0	0.0 0.	0 0.0	0.0
close fridge	-15.4	6.3	16 TI	4 0.0	0.3	0.0 1	5.2 2	6 225	3 2.7	111.7 5	5.4 6	69.1	22.5	5.0 63.0	5 11 2 .0	3.7	1.5 71	5 4.0	1.0	8.6 3	5 0.5	7.1	0.0 0	0.0 0.0	43	16 0	6 0.5	0.0	12 34	0.0	00 0	0.0	10 0	10 0.0	0.9	55 4	9 6.5	4.4
take carrot	0.0	0.0	0.0 2	3 0.0	0.0	0.0	40 1	5.0 2.9	0.8	1.9 0	0.7 0.	3.0	2.6	27 33	6.6	0.7	3.4 3	8 0.5	0.0	0.0 0	9 19	0.0	0.0 0	0.0 0.0	0.0	0.0 0	0.0	0.0	20 0.0	0.0	0.0 0.0	0.0	0.0	10 0.0	0.0	0.0 0.	0 02	0.0
open drawer	-14.5	12.6	8.3 09	7 0.5	6.3	0.0	13.2 4	2 291	7 1 9	154.9 1	х эл	5 135.9	30.8	72 2 703	7158.2	4.9	2.0 22	14 15	3.0	15.8 2	8 2.8	23.2	0.0 0	0.0 0.0	15.1	92 1	7 4.0	0.7	9.8 20.	9 0.7	01 0	0.0	0.2 0	10 0.0	0.0	1.7 1	9 5.9	11.4
put down vegetable	- 0.0	0.0	0.0 0	0 0.0	0.0	0.0	0.7 0	0 40	0.8	3.4 0	0.9 0.	2.6	1.4	2.0 3.4	0.6	11	0.6 0	2 0.9	0.8	11 0	07 11	0.4	0.0	0.0 0.0	0.0	0.0 0	0.0	0.0	2.0 0.0	0.0	0.0 0.	0.0	0.0	0.0	0.0	0.0 0.	0 0.0	0.0
open cupboard	- 4.0	8.7	3.9 140	e 0.0	13.7	02	6.7 5	1 275	9 0 2	961.1 S	5.2 21	4 805 2	28.2	0.2174	5296.1	2.6	0.8 12	7.0 4.0	12.8	9.0 0	9 4.2	24.2	1.0 0	0.0 0.0	7.7	8.0 2	6 0.5	0.0 2	1.3 19.	7 0.0	0.0 0.	0.0	0.0	0.0	0.0	16.9 5.	9 7.7	10.5
take board cutting	0.6	0.0	14 1	1 0.0	0.4	0.0	0.9 0	0 6.0	0.3	12.3 1	1.0 16.	9.5	9.2	7.8 3.0	5.9	1.4	0.9 1	6 2.9	1.3	0.5 0	0.0	0.0	0.0	0.0 0.0	3.2	0.0 0	6 0.0	0.0	20 24	0.0	0.0 0.0	0.0	0.9 0	10 0.0	0.0	0.0 0.	9 0.0	0.0
put-down bears cutting	3.2	31	21 5	7 0.0	0.4	0.0	42 1	3 27 5	5 0.3	20.5 0	4 10.	2 18 7	23	57.0 30.	40.0	1.9	19 21	2 43	0.3	5.4 0	10 0.9	0.0	0.0 0	0.0 0.0	0.5	10 2	5 10	0.0	14 61	0.0	00 0	0.0	0.0 0	10 0.0	0.0	16 0	0 52	0.2
close cupboard	19	8.9	0.9 103	8 0.0	7.3	0.0	56.2 4	0 201	5 0.3	258.5 1	10 18	6 217.3	22.0	33.6 147.	7238.4	2.1	10 13	6.3 1.5	5.7	26.0 2	.0 4.8	13.5	31 0	0.0 0.0	2.6	67 0	1 0.1	0.0	64 13.	5 0.0	0.0 0.0	0.0	0.0	10 0.0	0.0	9.7 2	8 5.9	6.9
take kinfe	6.4	0.0	4.7 14	2 0.0	3.8	0.0	11.0 2	5 11.4	0.9	14.1 12	2.0 B.	11.4	21.9	58 7 21 5	47.3	3.4	3.2 32	4 21	5.9	2.8 0	2 0.2	0.2	05 0	.0 0.0	2.0	0.0 2	6 0.6	0.0	20 65	0.0	00 0	0.0	0.9 0	10 0.0	0.0	3.1 12	5 1.0	1.3
pat-down in te	4.8	3.4	4.8 44	5 0.0	6.7	0.3	50.6 3	9 62 5	5 1.0	70.7 1	7.8 24	5 48.3	29.1 2	46.5 52.3	139.7	43	3.7 12	3.3 1.3	7.6	4.4 0	3 50	4.9	0.8 0	0.0 0.0	5.8	7.5 2	3 2.0	0.0	14 15	s 0.0	0.0 0.0	0.0	0.8	10 0.0	0.0	9.6 5.	1 20.4	7.0
close drawer	12.1	91	5.8 GT	6 0.5	41	0.0	19.9 Z	2 262	6 2.9	562.9 2	2.4 20	0 12 4.6	34.7	17 5 179	6342.9	3.7	17 90	9 1.0	3.7	26.6 2	1 0.9	11.5	0.0 0	0.0 0.0	5.8	95 0	8 42	0.8	5.0 21	3 0.8	01 0	0.0	0.2 0	10 0.0	0.0	83 1	3 42	7.2
turn on tap	-34.9	4.6	8.4 77	6 0.2	4.5	0.0 5	2.3 2	3 141	9 0.7	207.8 4	19 35	9134.0	48.9 1	98.8	795.9	49		0. 31	10.9	2.2 0	9 9.9	14.0	0.0	0.0 0.0	7.4	12.7 1	3 0.7	0.0	3.6 35.	1 1.0	07 1	2 0.3	0.6 0	1 0.0	0.0	7.1 6.	6 51	6.7
wash courgette	- 0.0	0.0	0.0 0.	0.0	0.0	0.0	0.0 0	0 0.6	1.4	2.0 0	0.0	2.8	0.8	0.7 0.0	1.5	1.5	35 6	6 1.2	1.7	1.5 1	3 2.9	0.1	01 0	0.0 0.0	0.0	0.0 0	0.0	0.0	20 0.0	0.0	0.0 0.0	0.0	0.0	0.0	0.0	0.0 0.	0 0.0	0.0
wash carrot	0.0	0.0	0.0 O	0.0	0.0	0.0	0.0 0	6 0.9	1.8	2.5 0	0.0 0.	3.1	1.8	0.0 0.9	0.0	0.0	10 3	9 2.3	1.5	13 2	1 2.2	0.5	0.0 0	0.0 0.0	0.0	0.0 0	0.0	0.0	20 0.0	0.0	0.0 0.0	0.0	0.0 0	10 0.0	0.0	0.0 0.	0 0.0	0.0
turn-off tap	40.6	3.3	7.3 88	3 0.4	8.5	0.0	3.0 3	4 359	7 1.0	233.1 2	2.6 24	8 156.5	50.5	57.0115	7904 5	1.5	0.8 65	14 28	11.8	4.3 1	9 12	23.0	0.2 0	0.0 0.0	5.5	13.9 4	8 19	0.0	7.0 31	2 0.0	10 1	0.5	0.9 0	13 0.2	01	7.0 3.	5 65	10.2
take grater	0.0	0.0	0.0 C	0.0	0.0	0.0	0.0 3	0 2.0	0.0	3.4 0	15 0.	5.9	3.2	15 18	2.5	0.0	0.0 1	9 0.1	1.0	0.9 0	18 1.5	0.5	0.0 0	0.0 0.0	0.8	0.0 0	0.0	0.0	20 0.0	0.0	0.0 0.0	0.0	0.0	10 0.0	0.0	0.0 0.	0.0	0.0
take pan	- 14	0.0	7.4 4	9 0.0	1.4	0.0	3.4 0	3 40	0.0	9.6 1	11 0	15.5	62	2.4 2.0	10.6	0.0	0.2 8	2 0.0	7.7 :	13.7 0	9 1.8	1.3	1.2 0	0.0 0.0	0.9	00 0	5 0.0	0.0		0.0	00 0	0.0	0.0	10 0.0	0.0	0.0 Q.	0 0.0	0.0
pat.down.pan	3.3	0.3	0.0 15	1 0.0	0.0	0.0	2.9 0	1 37)	5 00	30.2 2	2.1 10.	25.6	4.9	12.5 28.3	62.7	0.6	0.1 53	1 0.7	2.1	8.6 3	0 2.5	25.7	1.4 0	0.0 0.0	0.0	0.8 0	3 0.0	0.0	2.6 8.1	0.0	00 0	0.0	0.0	10 0.0	0.0	0.7 0.	7 0.6	2.5
take courgette	- 0.0	0.0	0.0 0	3 0.0	0.0	0.0	0.6 0	0 0.1	0.0	0.0 0	0.0 0.0	0.0	0.0	0.5 0.3	1.6	1.9	0.0 2	5 0.0	0.0	0.0 0	9 3.8	0.7	02 0	0.0 0.0	0.0	0.0 0	.0 0.0	0.0		0.0	00 0	0.0	0.0 0	0.0	0.0	0.0 Q	0 03	0.0
cut courgette	- 0.0	0.0	0.0 0.	0 0.0	0.0	0.0	0.0 0	0 6.8	0.0	0.8 0	0.0	0.0	0.0	12.1 3.8	7.0	0.0	1.9 7	1 0.0	0.0	2.7 3	0 56.	5 3.8	45 0	0.0 0.0	0.0	17 0	0 0.3	0.0	10 0.0	0.0	0.0 0.0	0.0	0.0	10 0.0	0.0	0.0 0.	0 0.0	0.0
turn-on hob	- 2.8	33	0.4 14	0.0	0.0	0.0	5.8 0	0 17.3	0.0	29.6 0	0.0	12.9	21	5.6 7.8	14.2	0.0	0.0 10	e 0.0	0.8	43.0	10 2.6	11.2	0.9 0	0.0 0.0	1.2	0.0 0	0 19	0.0	2.3 0.0	0.0	0.0 0.0	0.0	0.0	0.0	0.0	0.9 0.	2 1.5	0.0
dice courgette	0.0	0.6	0.0 O.	0 0.2	0.0	0.0	0.0 0	0 3.6	0.0	0.0 0	0.0	0.0	0.0	4.6 3.4	0.1	0.0	0.0 0	0 0.0	0.0	0.0 0	0.0	0.1	19.1 1	10 0.8	0.6	4.0 0	4 0.4	0.4	13 0.2	0.1	0.0 0.0	0.0	0.0	10 0.0	0.0	0.0 0.	0 0.0	0.0
pour-onto courgette par	- 0.0	0.0	0.0 0.0	0 0.3	0.0	0.0	0.0 0	0 0.9	0.0	0.0 0	0.0	0.0	0.0	0.0 0.8	0.2	0.0	0.0 0	0 0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0 0.9	0.8	12 0	6 0.6	0.6	15 0.3	0.4	0.0 0.0	0.0	0.0	0.0	0.0	0.0 0.	0 0.0	0.0
into courgette bin rubbish	0.0	0.0	0.0 C	0 0.4	0.0	0.0	0.0 0	0 10	0.0	0.0 0	0.0	0.0	0.0	0.0 0.9	0.3	0.0	0.0 0	1 0.0	0.0	0.0 0	10 0.0	0.0	0.0 0	0.0 0.0	0.9	14 0	7 0.7	0.7	16 05	0.6	0.0 0.0	0.0	0.0	10 0.0	0.0	0.0 0.	0 0.0	0.0
take spatula	0.0	0.0	0.0 0	7 0.6	2.7	0.6	05 1	1 59	0.0	51 0	15 1	40	7.4	7.3 2.5	7.0	0.0	0.0 4	3 2.4	1.7	0.7 0	.0 0.0	0.0	0.0 0	0.0 0.0	11.3	17 2	3 2 2	0.8	2.1 25	7 0.9	02 0	1 0.0	0.3 0	10 0.0	1.0	2.6 1	5 07	1.9
stir courgette	0.0	0.6	0.0 2	9 1.7	0.0	0.0	0.2 0	0 47	0.0	35 0	20 2	2.2	0.0	11 37	16.9	0.0	0.0 11	9 0.0	0.0	03 0	10 0.0	0.3	0.0 0	0.0 0.0	0.0	31.5 1	0 42	0.9	2.2 8.4	2.8	08 1	2 01	0.1 0	10 0.0	0.0	0.0 0.	0 32	0.0
take soft	- 16	0.0	0.0 1	4 0.7	0.9	0.0	06 0	3 40	0.0	0.4 0	.9 0.	16	2.1	0.0 1.8	3.8	0.0	0.0 1	4 0.0	1.2	0.8 0	0.0	0.0	0.0 0	0.0 0.0	1.3	0.9 6	6.9	1.0	11 0.0	12	03 0	4 0.0	0.0 0	10 0.0	0.0	0.0 0.	0 00	0.0
open sait	- 14	0.0	11 0	0.8	0.0	0.0	00 0	0 31	0.0	0.9 0	0.8 0.	0.5	0.9	0.0 0.8	3.6	0.0	0.0 1	3 0.0	0.0	0.0 0	.0 0.0	0.0	0.0 0	0.0 0.0	1.1	48 0	0 17	1.0 1	0.4 0.2	13	0.4 0	5 0.0	0.0 0	10 0.0	0.0	0.0 Q.	0 00	0.0
pour-onto salt pan	- 0.0	0.0	0.0 O	0.8	0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0	0.0	0.0	0.0 0.0	0.7	0.0	0.0 0	5 0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0 0.0	0.0	0.9 0	0.0	0.0	1.0 0.0	1.4	0.4 0.	5 0.0	0.0	10 0.0	0.0	0.0 0.	0.00	0.0
put-down sait	0.6	1.6	0.4 5	7 0.8	0.0	0.0	43 0	0 10.0	0.0	10.3 1	1 0.	16.2	1.5	13 7.4	12.4	0.0	0.0 7	5 0.0	0.0	2.0 0	0.0	2.1	0.0 0	0.0 0.0	2.4	3.6 0	0.9	0.0	43 14	1.4	0.4 0.	6 0.0	0.1 0	10 0.0	0.0	0.0 0.	0 0.0	0.8
put-down spatula	- 0.0	0.4	0.0 9	7 11	5.4	1.2	5.1 0	0 26.	0.0	25.6 0	12 63	14.0	63 :	8.8 15.	59.0	0.0	0.0 45	2 0.0	2.0	9.7 0	0.0	2.8	0.0	0.0 0.0	21.1	18 1	0 1.0	0.0	3.3 37	4 18	0.6 0.	9 0.2	21 0	10 0.0	0.0	7.6 1	9 2.4	5.8
wash celety	0.0	0.0	0.0 O.	4 0.9	0.0	0.0	0.3 0	0 0.0	0.0	0.0 0	0.0	0.0	0.0	0.0 0.0	0.0	0.0	0.0 2	0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0 0.0	0.1	0.0 0	0.0	0.0	20 0.0	0.0	19 3.	3 11	0.9 0	16 0.4	01	0.0 0.	0 0.0	0.0
pat-down celery	- 0.0	0.0	0.0 O	3 0.5	0.0	0.0	02 0	0.0	0.0	0.0 0		0.0	0.0	0.0 0.0	0.0	0.0	0.0 0	0 0.0	0.0	0.0 0	.0 0.0	0.0	0.0 0	.0 0.0	0.1	0.0 0	.0 0.0	0.0	10 0.0	0.0	00 1	0.0	0.5 0	14 0.3	01	0.0 Q	0 0.0	0.0
cut celery	0.0	0.0	0.0 1	0 15	0.0	0.0	0.9 0	0.0	0.0	0.0 0	0.0 0.0	0.0	0.0	0.0 0.0	0.0	0.0	0.0 0	0.0	0.0	0.0 0	.0 0.0	0.0	0.0 0	0.0 0.0	0.8	0.0 0	.0 0.0	0.0	10 0.4	0.0	00 1	0 17	15 1	2 1.0	0.7	0.3 0.	4 02	0.1
pour-onto celery pan	0.0	0.0	0.0 0	7 1.0	0.0	0.0	07 0	0 0.0	0.0	0.0 0	0.0 0.0	0.0	0.2	0.0 0.0	0.0	0.0	0.0 0	0.0	0.0	0.0 0	.0 0.0	0.0	0.0 0	0.0 0.0	0.8	0.0 0	.0 0.0	0.0	20 04	0.0	00 0	0.0	10 0	19 0.7	0.6	0.4 0.	2 0.3	0.3
r. down boerd outling knife	0.0	0.0	0.0 0	9 1.0	0.0	0.0	0.5 0	0 2.3	0.0	14 0		1.0	1.4	0.7 2.1	2.3	0.0	0.0 1	8 0.0	1.0	0.9 0	.0 0.0	0.6	0.0 0	0.0 0.0	2.7	0.0 0	0 0.9	0.0	15 42	0.0	00 0	0.0	0.0	19 0.8	0.6	0.7 0.	7 0.6	0.5
throw into bin garbage	- 0.0	0.0	0.0 1	0.0	0.0	0.0	0.9 0	0 0.0	0.0	0.0 0	0.0	0.0	0.4	0.0 0.0	0.0	0.0	0.0 0	0 0.0	0.0	0.0 0	0.0	0.0	0.0 0	0.0 0.0	1.1	0.0 0	0 0.0	0.0	10 01	0.0	0.0 0.0	0.0	0.0	10 1.0	0.8	0.6 1	1 05	0.5
put-into celery fridge	0.0	0.0	0.0 Q.	0.0	0.0	0.0	10 0	0 0.0	0.0	0.0 0	0.0	0.0	0.5	0.0 0.0	0.0	0.0	0.0 0	0 0.0	0.0	0.0 0	0.0	0.0	0.0	0.0 0.0	1.4	0.0 0	0.0	0.0	10 0.7	0.0	0.0 0.0	0.0	0.0	10 0.0	0.9	0.7 1	3 0.6	0.6
str spetule	- 0.0	0.0	0.0 0.0	0.0	0.0	0.0	0.0 0	0 0.0	0.0	0.0 0	0.0	0.0	0.8	0.2 0.0	0.1	0.0	0.0 0	0 0.0	0.0	0.0 0	0.0	0.0	0.0	0.0 0.0	0.8	0.0 0	0.0	0.0	10 10	0.0	0.0 0.0	0.0	0.0	0.0	0.0	10 1	8 0.9	0.9
open container	0.0	0.3	0.0 5	9 0.0	6.9	0.0	45 0	0 13.3	0.0	6.4 0	13 1	6.0	2.9	2.0 8.1	15.6	0.0	0.0 6	6 0.3	0.0	0.8 0	0.0	0.7	0.0 0	0.0 0.0	3.4	0.0 0	0.0	0.0	14 13	0.0	0.0 0.0	0.0	0.5 0	10 0.0	0.0	9.8 7.	5 4.8	12.8
take origin	- 3.2	0.0	1.5 5.	4 0.0	0.0	0.0	48 0	6 3.8	0.0	19 0	0.0	3.4	15.4	13 3.8	6.7	0.0	0.0 8	9 0.0	0.0	0.0	0.0	0.0	0.0	0.0 0.0	1.8	0.0 0	0 0.0	0.0	2.0 0.0	0.0	0.0 0.	0.0	0.0	0.0	0.0	0.0 30	6 6.4	5.7
put down onion	0.8	2.2	0.8 3	5 0.0	1.0	0.0	3.9 0	0 10.3	0.0	35 0	20 2	17	55	13.0 6.6	5.8	0.7	0.0 5	1 0.0	0.0	0.0 0	18 0.0	0.4	0.0 0	0.0 0.0	0.9	3.0 0	.0 0.0	0.0	16 13	0.0	00 0	0.0	0.5 0	10 0.0	0.0	11 4	3 21.3	4.6
close container	- 0.0	19	0.0 16	8 0.0	0.0	0.0	14.2 0	0 11.		18.2 0	20 3	34.4	3.4	25 63	7.7	0.0	0.0 G	2 0.0	0.0	16 0	.0 0.0	0.0	0.0 0	0.0 0.0	4.2	00 0	.0 0.0	0.0	1.7 2.1	0.0	00 Q	0.0	0.7 0	10 0.0	0.0	2.4 0.	0 2.1	8.4
		- tab	doct .	- 548	- 124	- 140	- after		able .	- part	1 20	- pas	- auto	nte.	- 190-	ette .	- tour	ter.	-		i i	in the second	÷	i s	oute -	i i	i i	- 10	tin.	- Line	Sery -	ind.	- inter	- ače -	- 810	1	100-	ine.
	openi	n on	ciater -	take ce	conta	Takes	hose fr.	en dia	vepets	cupp	and cut	r cuption	Taken k	Stern >	um en	duna i	mark Cl	a and	take	down	duno.	10.11	Court o	an rubt	to the	round raise	6	to set	moto :	est ce	Cutt Cel	celery.	1 Out	ing the	11. Spo	n conta Take of	town of	e contra
		2			toke		Ŭ .		-dzen	ado .	over 00	chile		5 G	-	ALC: N				2	¥ 8	8	do	gefte t	8	th.		ue-1100	put dia	,	5 ct -5	th cerbo	cerci ci	inte ce	ĵ.	ado .	p.u.d	chu
									μ															a contra								rod	queo	ind ind				
																																	-					

### Types of distance metrics

#### Distance in frames/seconds between actions



- Takes into account temporal difference
- Can differentiate between end/start (but which one to choose?)
- But dependent on length of actions

# Distance as index difference in ordinal sequence



- Only considers ordering
- Might be more adapted to causality concepts
- Doesn't depend on length of actions

#### Testing ETM design choices

We test several design choices, including different decay functions, ETM sizes and distance metrics.

Size	Decay	Temp Metric	Present (top-1 accuracy)						
SIZC	Decay	remp. wieute	Verb	Noun	Action				
13k	linear	time	0.551	0.462	0.288				
13k	exponential	time	0.556	0.477	0.291				
2.5k	exponential	time	0.586	0.488	0.313				
2.5k	no decay	-	0.581	0.480	0.305				
2.5k	linear	index	0.601	0.493	0.319				
2.5k	exponential	index	0.603	0.503	0.324				

#### Datasets used



700 videos depicting cooking actions, totalling 100 hours.



3670 hours of video from 71 different participants.

#### EGTEA Gaze+



10k segments annotated with 19 verbs, 51 nouns and 106 unique actions.

#### Tasks where we test our embeddings



#### Experimental results - Action recognition

#### Results with MoViNet A0

Dataset	Model	Present					
Dataset	WIGHEI	Verb	Noun	Action			
EK100 [14]	Baseline	64.8	47.4	36.8			
	ETM(Ours)	67.9	51.2	40.2			
EGO4D	Baseline	32.3	23.5	21.1			
LTA [16]	ETM(Ours)	32.9	24.2	22.0			
EGTEA	Baseline	81.2	71.7	60.4			
Gaze+ [28]	ETM(Ours)	83.4	72.9	62.5			

#### **Cross Model**

Results on EK100 across a wide variety of models

Model	w/o ETM	w/ ETM
MoviNet A0 [24]	36.8	40.2
MoviNet A2 [24]	41.2	43.4
X3D-XS [11]	35.5	38.1
X3D-S [11]	40.5	42.2
ConvNeXt-S 224 [31]	20.1	32.4
LambdaResNet-50 [4]	26.6	27.1
EfficientNet-B0 [57]	25.3	26.3
EfficientNet-B4 [11]	29.2	29.4
AVT-b [14]	30.4	30.7

#### **Experimental results - Action Anticipation**

		Cross	Model	S						
		Results on EK								
			Baseline			ETM (Ours	)	Encoder	w/o ETM	V
Dataset	Frozen Encoder?	Verb ↑	Noun ↑	Action ↑	Verb ↑	Noun ↑	Action ↑	MoViNet A0 [24]	8.0	
Inclusion of Concession	1	19.9	20.4	7.2	21.5	20.5	8.1	MoviNet A2 [24]	10.2	
EK100		20.8	21.3	8.0	22.4	22.7	9.1	X3D-X5 [11]	0.3 9.4	
ECO4D LTA	~	17.1	16.6	10.3	18.1	17.8	11.4	ConvNeXt-S 224 [31]	4.1	
EGO4D LIA		18.2	17.5	11.1	19.9	19.1	12.9	EfficientNet B0 [57]	7.2	
EGTEA Gaze+	$\checkmark$	42.1	37.6	28.9	43.4	38.9	31.3	EfficientNet B4 [57]	9.4	
		43.5	38.5	30.3	46.5	40.7	34.1	AVT-b [14]	13.4	

with ETM 9.1

10.8

7.4 9.9

5.0

8.0

10.1 13.5

#### Larger gains on smaller models!



#### Performance on architecture families



#### **Baseline tests and ablations**

**Shuffle Rows** 



Row shuffling in ETM matrix: Distribution of **future actions** doesn't match the action index at a given **row**.



Row shuffling in ETM matrix: Distribution of **past actions** doesn't match the action index at a given **column**.

#### Full Shuffle



Full shuffling in ETM matrix: no transition probability estimation matches its original action pair.

#### **Co-occurrence**

Co-occurrence frequency between a1 and a2



Using a co-occurrence matrix: cells correspond to **co-occurrence frequencies** instead of transition probabilities

#### Baseline tests and ablations

Model		Present		MAE on	MAE on
Widder	Verb ↑	Noun ↑	Action ↑	<b>Past</b> $\downarrow$	<b>Future</b> $\downarrow$
Baseline	64.8	47.4	36.8	-	-
Full shuffle	64.1	47.2	36.3	4.117	4.012
Columns/rows shuffle	64.7	47.6	36.7	3.254	3.101
Co-occurrence	65.3	49.0	37.9	1.211	1.115
Only past vector	65.7	49.3	38.2	0.901	-
Only future vector	65.5	49.8	38.3	-	0.898
ETM (Ours)	67.9	51.2	40.2	0.882	0.859

# Conclusions

- We introduce a new training regime that uses external temporal regularities to boost video understanding.
- Using our ETM as a training target enhances action recognition and anticipation, particularly on **low representational power models.**
- Our ETM protocol's key benefits: flexibility, simplicity, cost-effectiveness, and easy integration.





# Thank you!

# Leveraging Temporal Context in Low Representational Power Regimes

Camilo Fosco, SouYoung Jin, Emilie Josephs, Aude Oliva

Project page: <u>camilofosco.com/etm\_website</u>

Contact: camilolu@mit.edu



