

# Punching Bag vs. Punching Person: Motion Transferability in Videos

## Supplementary Material

The supplementary material provides additional information to complement the main paper.

- Sec. A includes a comprehensive description of the unimodal and multimodal models used in our experiments, along with implementation details.
- Sec. B presents detailed performance tables for models on known and unknown classes, along with more examples, failure cases and analysis figures.
- Sec. C shows the architecture of our proposed approach.
- Sec. D provides the complete lists of coarse and fine classes for the Syn-TA, SSv2-TA, and K400-TA datasets.
- Datasets and relevant code are available at: <https://github.com/raiyaan-abdullah/Motion-Transfer>.

### A. Models

We experimented with several unimodal and multimodal models. This includes traditional convolutional neural networks such as ResNet50 [24], I3D [23] and X3D [17]. We also experimented with SlowFast [18] which utilizes slow and fast pathways. Then we explored models based on the Vision Transformer [13] such as MViTv2 [32] and Rev-MViT [36] which combine multiscale features with the transformer architecture. UniFormerV2 [30] combines pre-trained ViTs with efficient Uniformer [29] designs. AIM [63] utilizes the frozen parameters of pre-trained image models and trains various adapters. Among multimodal models, we experimented with different variations of CLIP [46] designed for activity recognition. ActionCLIP [57] adapts a “pre-train, prompt, and fine-tune” approach. X-CLIP [41] proposes a cross-frame module and a video specific prompting scheme to adapt pre-trained language image models. ViFi-CLIP [47] shows that simple fine-tuning can achieve similar results to using specific temporal components. EZ-CLIP [1] uses temporal visual prompting and spatial adapters to efficiently prepare CLIP for downstream tasks while keeping original model weights frozen. FROSTER [25] utilizes the frozen CLIP model as a teacher for adapting to activity recognition using residual feature distillation. We also experimented with domain generalization methods such as VideoDG [64], STDN [34], and CIR [44] on coarse classes.

**Implementation details:** For training ResNet50, I3D, X3D, MViTv2, Rev-MViT, and SlowFast we utilized the [PySlowFast repository](#) [15] from Meta Research. For other models and domain generalization methods, we used the code from their respective GitHub repositories. We particularly used the model versions: I3D R50, X3D-M, SlowFast R50, MViTv2-S, Rev-MViT-B-16, AIM ViT-B/16,

UniFormerV2-B/16, ActionCLIP ViT-B/16, X-CLIP-B/16, EZ-CLIP ViT-B/16, and FROSTER-B/16. The model hyperparameters for training were kept similar to their configuration for Something-something-v2 and Kinetics400. For Syn-TA, we followed the respective hyperparameters of Kinetics400 for each model. The learning rate was slightly tuned in some cases. We also modified the configurations dependent on the compute machine like batch size, number of GPUs, number of workers, etc to adjust to our resources. The number of epochs was varied by model and dataset depending on how fast the model converges. The models were trained on 1-4 NVIDIA GPUs. The memory of GPUs varied from 11 GB to 80 GB. The configuration files for training the models are available in our [GitHub](#).

### B. Benchmark results

Along with  $D_{abs}$  and HM, we show an additional metric:

**Relative Drop:** The percentage decrease in performance when shifting to an unknown context:

$$D_{rel} = \left| \frac{Known - Unknown}{Known} \right| \times 100$$

**Performance on known vs. unknown classes:** The known and unknown accuracies for both Set 1 and Set 2, covering coarse and fine classes, along with other metrics, are detailed in Tab. 6, Tab. 7 (Syn-TA); Tab. 8, Tab. 9 (K400-TA), and Tab. 10, Tab. 11 (SSv2-TA). For metrics such as known accuracy, unknown accuracy, and harmonic mean (HM), higher values indicate better performance, whereas lower values are desirable for  $D_{abs}$  and  $D_{rel}$ . As discussed in the main paper, there is a noticeable drop in performance for both coarse and fine motions across all models, illustrated more clearly in Fig. 11.

**Performance on coarse vs fine classes:** Fig. 11 also shows that fine classes are generally more challenging than coarse classes. However, notable exceptions include the performance of EZ-CLIP and FROSTER in K400-TA unknown classes.

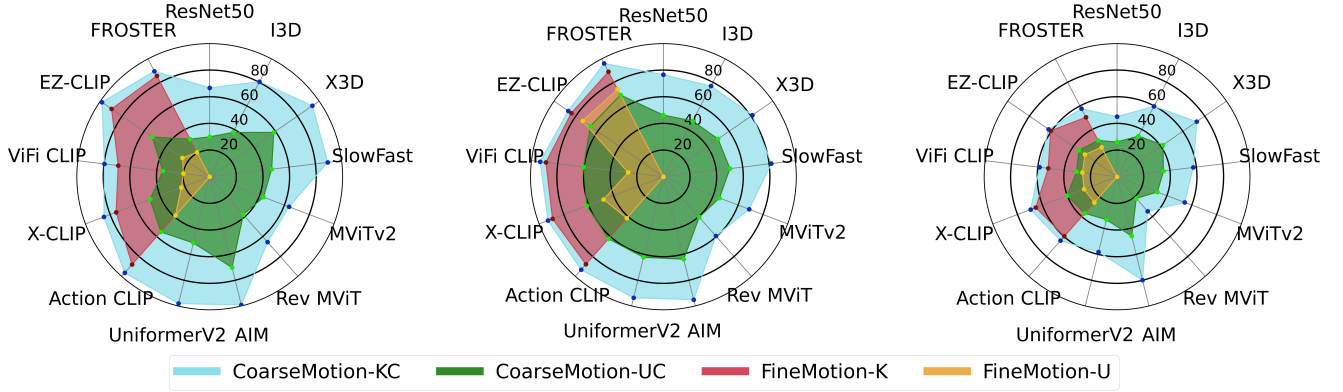


Figure 11. Left: Syn-TA, middle: K400-TA, and right: SSv2-TA. Average detection scores of both sets are given for the three datasets. Performance drop is observed for both coarse (blue to green for all models) and fine motions (red to yellow for multimodal models).

Model	Set 1					Set 2				
	Known (Set 1) ↑	Unknown (Set 2) ↑	D <sub>abs</sub> ↓	D <sub>rel</sub> ↓	HM ↑	Known (Set 2) ↑	Unknown (Set 1) ↑	D <sub>abs</sub> ↓	D <sub>rel</sub> ↓	HM ↑
<b>Unimodal models</b>										
ResNet50 [24]	67.59	29.26	38.33	56.70	40.84	65.74	30.61	35.13	53.43	41.77
I3D [23]	83.40	37.77	45.63	54.71	51.99	77.61	37.26	40.35	51.99	50.34
X3D [17]	94.76	52.71	42.05	44.37	67.74	92.66	64.20	28.46	30.71	75.84
SlowFast [18]	90.94	46.70	44.24	48.64	61.71	87.61	47.03	40.58	46.31	61.20
MViTv2 [32]	57.50	38.72	<b>18.78</b>	32.66	46.27	69.89	47.74	<b>22.15</b>	31.69	56.73
Rev-MViT [36]	62.45	40.48	21.97	35.18	49.12	68.62	35.57	33.05	48.16	46.85
AIM [63]	<b>99.81</b>	<b>70.85</b>	28.96	<b>29.01</b>	<b>82.87</b>	98.46	<b>69.48</b>	28.98	<b>29.43</b>	<b>81.47</b>
UniformerV2 [30]	96.93	50.32	46.61	48.08	66.24	<b>98.99</b>	52.08	46.91	47.38	68.25
<b>Multimodal models</b>										
ActionCLIP [57]	97.74	55.43	42.31	43.28	70.74	94.84	55.24	39.60	41.75	69.81
X-CLIP [41]	87.54	48.03	39.51	45.13	62.02	82.55	47.64	34.91	42.28	60.41
ViFi-CLIP [47]	81.32	39.04	42.28	51.99	52.75	78.03	31.88	46.15	59.14	45.26
EZ-CLIP [1]	98.38	51.92	46.46	47.22	67.96	98.23	52.94	45.29	46.10	68.80
FROSTER [25]	91.13	31.44	59.69	65.49	46.75	87.71	32.17	55.54	63.32	47.07
<b>Domain Generalization Methods</b>										
VideoDG [64]	98.34	45.69	52.65	53.53	62.39	97.81	41.17	56.64	57.90	57.94
STDN [34]	70.80	26.48	44.32	62.59	38.54	70.53	21.46	49.07	69.57	32.90
CIR [44]	62.01	7.42	54.59	88.03	13.25	58.25	11.76	46.49	79.81	19.56

Table 6. Known and unknown accuracy of coarse motions on Syn-TA

Model	Set 1					Set 2				
	Known (Set 1) ↑	Unknown (Set 2) ↑	D <sub>abs</sub> ↓	D <sub>rel</sub> ↓	HM ↑	Known (Set 2) ↑	Unknown (Set 1) ↑	D <sub>abs</sub> ↓	D <sub>rel</sub> ↓	HM ↑
ActionCLIP [57]	89.58	<b>40.74</b>	48.84	<b>54.52</b>	<b>56.00</b>	86.44	<b>36.89</b>	<b>49.55</b>	<b>57.32</b>	<b>51.71</b>
X-CLIP [41]	75.99	26.43	49.56	65.21	39.21	74.41	19.38	55.03	73.95	30.75
ViFi-CLIP [47]	72.21	23.88	<b>48.33</b>	66.92	35.89	66.33	15.94	50.39	75.96	25.70
EZ-CLIP [1]	<b>91.69</b>	30.23	61.46	67.03	45.46	<b>87.39</b>	19.56	67.83	77.61	31.96
FROSTER [25]	87.26	22.82	64.44	73.84	36.17	83.62	18.54	65.08	77.82	30.35

Table 7. Known and unknown accuracy of fine motions on Syn-TA

Model	Set 1					Set 2				
	Known (Set 1) $\uparrow$	Unknown (Set 2) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$	Known (Set 2) $\uparrow$	Unknown (Set 1) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$
<b>Unimodal models</b>										
ResNet50 [24]	79.27	49.74	29.53	37.25	61.12	73.71	42.68	31.03	42.09	54.05
I3D [23]	80.09	51.76	28.33	35.37	62.88	73.69	42.75	30.94	41.98	54.10
X3D [17]	80.66	51.42	29.24	36.25	62.80	81.80	48.34	33.46	40.90	60.76
SlowFast [18]	81.61	52.40	29.21	35.79	63.82	81.80	48.27	33.53	40.99	60.71
MViTv2 [32]	70.46	47.41	23.05	32.71	56.68	67.30	42.72	24.58	36.52	52.26
Rev-MViT [36]	58.62	41.80	16.82	28.69	48.80	60.18	39.28	<b>20.90</b>	34.72	47.53
AIM [63]	<b>95.00</b>	64.77	30.23	31.82	77.02	95.09	62.70	32.39	34.06	75.57
UniformerV2 [30]	92.81	63.81	29.00	31.24	75.62	94.32	60.77	33.55	35.57	73.91
<b>Multimodal models</b>										
ActionCLIP [57]	92.75	65.20	27.55	29.70	76.57	93.74	59.29	34.45	36.75	72.63
X-CLIP [41]	92.54	63.36	29.18	31.53	75.21	92.85	59.59	33.26	35.82	72.59
ViFi-CLIP [47]	92.66	62.30	30.36	32.76	74.50	93.83	58.58	35.25	37.56	72.12
EZ-CLIP [1]	85.70	68.39	<b>17.31</b>	<b>20.19</b>	76.07	88.07	65.01	23.06	<b>26.18</b>	74.80
FROSTER [25]	92.68	<b>69.38</b>	23.30	25.14	<b>79.35</b>	<b>99.31</b>	<b>69.09</b>	30.22	30.43	<b>81.48</b>
<b>Domain Generalization Methods</b>										
VideoDG [64]	84.94	56.63	28.31	33.32	67.95	87.28	51.28	36.00	41.24	64.60
STDN [34]	65.17	47.96	17.21	26.40	55.25	71.05	44.24	26.81	37.73	54.52
CIR [44]	68.53	13.82	54.71	79.83	23.00	47.22	11.50	57.03	83.21	19.69

Table 8. Known and unknown accuracy of coarse motions on K400-TA

Model	Set 1					Set 2				
	Known (Set 1) $\uparrow$	Unknown (Set 2) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$	Known (Set 2) $\uparrow$	Unknown (Set 1) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$
ActionCLIP [57]	86.84	45.86	40.98	47.19	60.02	88.66	37.18	51.48	58.06	52.39
X-CLIP [41]	<b>88.98</b>	51.95	37.03	41.61	65.60	89.14	44.27	44.87	50.33	59.15
ViFi-CLIP [47]	88.44	29.63	58.81	66.49	44.38	89.38	23.77	65.61	73.40	37.55
EZ-CLIP [1]	82.24	<b>77.20</b>	<b>5.04</b>	<b>6.12</b>	79.64	85.29	70.70	<b>14.59</b>	<b>17.10</b>	77.31
FROSTER [25]	88.35	76.97	11.38	12.88	<b>82.26</b>	<b>89.52</b>	<b>71.26</b>	18.26	20.39	<b>79.35</b>

Table 9. Known and unknown accuracy of fine motions on K400-TA

Model	Set 1					Set 2				
	Known (Set 1) $\uparrow$	Unknown (Set 2) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$	Known (Set 2) $\uparrow$	Unknown (Set 1) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$
<b>Unimodal models</b>										
ResNet50 [24]	47.02	25.67	21.35	45.40	33.21	43.13	26.50	16.63	38.55	32.82
I3D [23]	63.19	33.20	29.99	47.46	43.53	56.02	35.61	20.41	36.43	43.54
X3D [17]	74.43	40.11	34.32	46.11	52.12	71.04	43.52	27.52	38.73	53.97
SlowFast [18]	61.47	34.48	26.99	43.90	44.17	53.87	35.83	18.04	33.48	43.03
MViTv2 [32]	59.46	32.71	26.75	44.98	42.20	49.16	32.04	17.12	34.82	38.79
Rev-MViT [36]	38.32	22.71	<b>15.61</b>	<b>40.73</b>	28.51	30.97	20.74	<b>10.23</b>	<b>33.03</b>	24.84
AIM [63]	<b>81.62</b>	<b>43.52</b>	38.10	46.67	<b>56.77</b>	<b>78.27</b>	<b>48.13</b>	30.14	38.50	<b>59.60</b>
UniformerV2 [30]	59.06	32.47	26.59	45.02	41.90	57.26	33.93	23.33	40.74	42.61
<b>Multimodal models</b>										
ActionCLIP [57]	66.44	34.67	31.77	47.81	45.56	61.77	38.66	23.11	37.41	47.55
X-CLIP [41]	72.50	37.80	34.70	47.86	49.69	66.49	42.41	24.08	36.21	51.78
ViFi-CLIP [47]	60.17	28.49	31.68	52.65	38.67	57.21	32.90	24.31	42.49	41.77
EZ-CLIP [1]	64.86	34.26	30.60	47.17	44.83	60.24	35.43	24.81	41.18	44.61
FROSTER [25]	59.01	28.68	30.33	51.39	38.60	56.30	32.69	23.61	41.93	41.36
<b>Domain Generalization Methods</b>										
VideoDG [64]	59.21	30.69	28.52	48.16	40.42	55.30	32.39	22.91	41.42	40.85
STDN [34]	37.44	22.33	15.11	40.35	27.97	34.43	22.29	12.14	35.25	27.06
CIR [44]	48.80	31.84	16.96	34.75	38.53	47.22	32.10	15.12	32.02	38.21

Table 10. Known and unknown accuracy of coarse motions on SSv2-TA

Model	Set 1					Set 2				
	Known (Set 1) $\uparrow$	Unknown (Set 2) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$	Known (Set 2) $\uparrow$	Unknown (Set 1) $\uparrow$	$D_{abs} \downarrow$	$D_{rel} \downarrow$	HM $\uparrow$
ActionCLIP [57]	60.12	24.10	36.02	59.91	34.40	59.33	27.58	31.75	53.51	37.65
X-CLIP [41]	<b>66.01</b>	24.92	41.09	62.24	36.18	<b>64.61</b>	<b>28.14</b>	36.47	56.44	<b>39.20</b>
ViFi-CLIP [47]	54.02	26.59	<b>27.43</b>	50.77	35.63	50.24	25.97	24.27	48.30	34.24
EZ-CLIP [1]	61.08	<b>31.83</b>	29.25	<b>47.88</b>	<b>41.85</b>	58.58	27.64	30.94	52.81	37.55
FROSTER [25]	51.90	23.58	28.32	54.56	32.42	48.79	26.40	<b>22.39</b>	<b>45.89</b>	34.26

Table 11. Known and unknown accuracy of fine motions on SSv2-TA

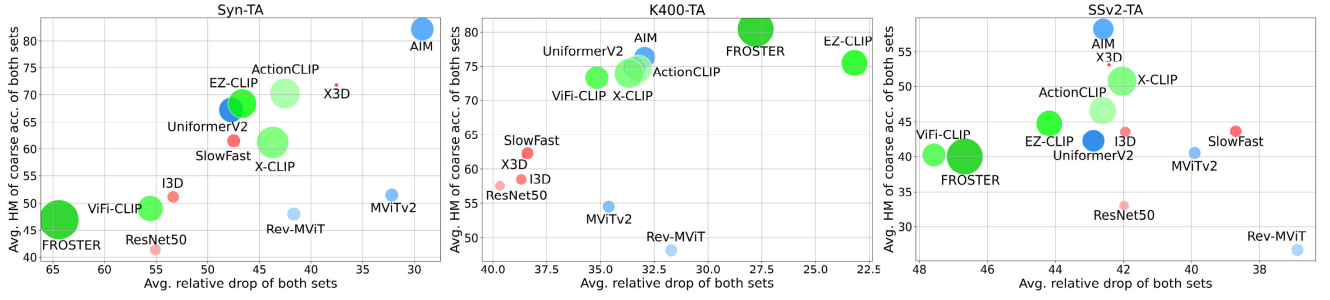


Figure 12. **Effect of model size for each dataset:** Average harmonic mean of coarse accuracy vs. relative drop  $D_{rel}$ . Bubble sizes correspond to the total number of model parameters, with colors indicating architecture types (red: CNN, blue: transformer - unimodal, green: transformer - multimodal). Models with larger parameter counts perform better on K400-TA, where videos contain rich spatial cues. This effect is less pronounced in Syn-TA, which requires some temporal understanding. In SSv2-TA, which is heavily reliant on temporal information, model size does not show a clear correlation with performance.

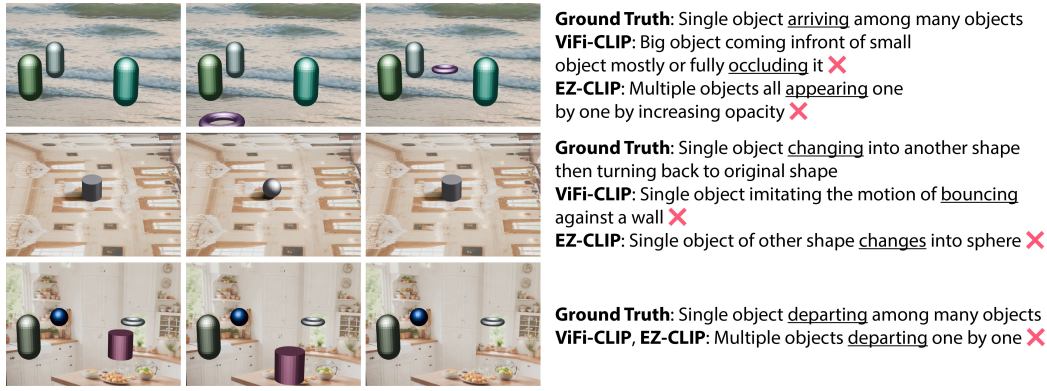


Figure 13. **Example of failure cases for FineMotion-U in Syn-TA with ViFi-CLIP, EZ-CLIP:** (Top) For the arrival of single objects among many objects, ViFi-CLIP is confusing the motion with occlusion as the arriving pink torus temporarily occludes the objects before completing its path. EZ-CLIP hallucinates the multiple objects are appearing in the scene. (Middle) The object in the scene changes its shape to a sphere and turns back. EZ-CLIP thinks the object did not transform back. (Bottom) Both models are mispredicting that the other objects are departing as well after the first object.

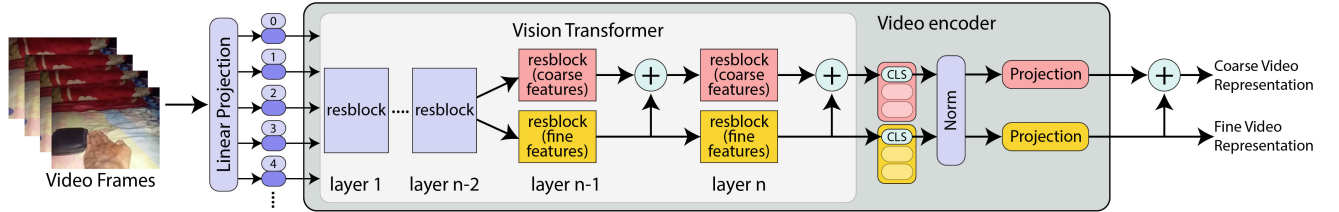


Figure 14. **Disentanglement of coarse and fine video features:** In the final two layers of the vision transformer, two branches extract coarse and fine motions simultaneously. The fine embeddings are added to the high-level embedding at each step via residual connections, combining detailed context with the broader motion features. The fine projection layer is trainable. Overall, this enables each branch to focus on disentangling the features most relevant to its specific role.

Set	Model	Syn-TA		K400-TA		SSv2-TA	
		Known/Unknown	HM	Known/Unknown	HM	Known/Unknown	HM
Coarse motions							
Set 1	Base-Coarse	98.38/51.92	67.96	85.70/68.39	76.07	64.86/34.26	44.83
	Ours	99.76/54.75	70.69	84.15/69.39	76.06	69.17/35.47	46.89
Set 2	Base-Coarse	98.23/52.94	68.80	88.07/65.01	74.80	60.24/35.43	44.61
	Ours	99.62/56.41	72.03	88.53/65.36	75.20	65.01/37.82	47.82

Table 12. **Performance comparison for disentanglement approach:** Comparison of baseline vs. our proposed method for coarse motions.

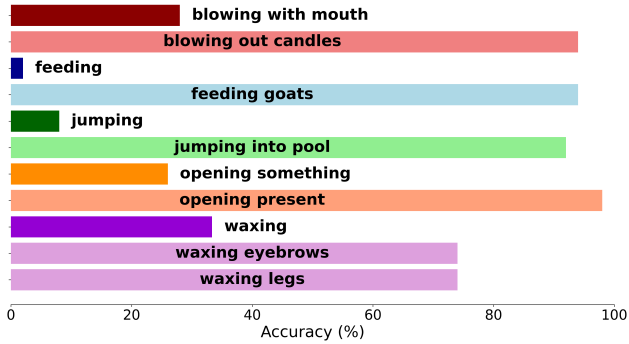


Figure 15. *Example cases of K400-TA where unknown fine performance is higher*: The bars with a darker color denote coarse classes while their corresponding fine class is shown in a lighter color. The accuracy of coarse classes is noticeably lower than fine counterparts. Results are shown for EZ-CLIP model.

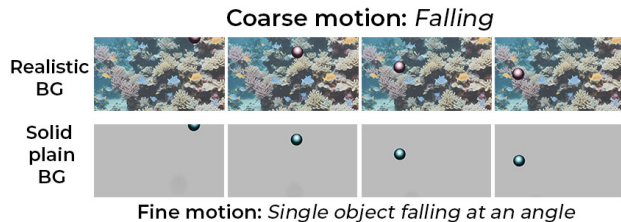


Figure 16. *Preview of Syn-TA realistic vs. solid plain background*: The absence of complex textures in the background enhances the model’s ability to comprehend object motion. This suggests that even for relatively simple motions, models are often challenged by textured backgrounds.

### C. Disentanglement architecture

We tested our approach on EZ-CLIP [1], chosen for its efficiency due to fewer learnable parameters, achieving competitive results with reduced time and computation. We modified EZ-CLIP’s video encoder by adding two branches in the final layers, duplicating the transformer blocks in the last two layers to separate feature learning. The earlier layers capture low and mid-level features, while one branch of the final layers specializes in high-level (coarse) concepts (e.g., “Pushing”) and the other focuses on fine-grained context (e.g., “Pushing something from left to right”). As shown in Fig. 14, the fine feature is combined with the coarse feature at every step. The class token (CLS) from each branch flows through a shared normalization layer and distinct projection layers (fine one is trainable), generating two embeddings. We also disabled passing EZ-CLIP’s temporal prompt in the final two layers. This setup allows the coarse branch to focus on high-level features while the fine branch selectively contributes necessary detail, improving overall coarse accuracy and maintaining focused feature disentanglement in each branch. As shown in Tab. 12, our approach improves coarse motion accuracy on temporal

datasets such as Syn-TA and SSv2-TA.

## D. Detailed list of classes

The full list of coarse classes and fine classes for each dataset is provided in this section. The split files for each of the dataset is also provided in our [GitHub](#).

### D.1. Syn-TA

In our newly proposed dataset ‘Syn-TA’, we generated videos of plain 3D objects, such as cubes, spheres, cylinders, etc, performing various motions using the 3D modeling software Blender [10]. Each video features a realistic background depicting either outdoor scenes (e.g., desert, forest, sunset) or indoor settings (e.g., coffee shop, kitchen, library), where 3D objects perform various motions. The videos are rendered at 24 frames per second (FPS) with a resolution of 1920x1080 pixels. The camera view is either from the front or the top.

The dataset includes 20 coarse motions, which are further subdivided into 100 fine motions. Each video contains, on average, 105 frames with a standard deviation of approximately 46. The fine motions are split into two subsets:  $S1$  (53 classes) and  $S2$  (47 classes). Each set has its own collection of object shapes and backgrounds, creating a more challenging setting for evaluating model performance on novel scenarios. Each fine motion is labeled with its corresponding coarse class and additional contextual details. Both subsets include at least two fine classes under each coarse motion. If a coarse class contains an even number of fine classes, they are evenly distributed between  $S1$  and  $S2$ . For coarse classes with an odd number of fine motions,  $S1$  is assigned one extra class. For example, for the coarse class “*departing*”, the fine class “*multiple objects departing simultaneously*” is in  $S1$  while “*multiple objects departing one by one*” is in  $S2$ .

This carefully constructed dataset is intended to serve as a diagnostic benchmark for evaluating how models adapt to detecting high-level actions under varying contexts. The Blender Python API code used to generate the videos is available in our [GitHub](#), and samples of each class are provided. The videos are stored in standard .mp4 format and follow the structure of Kinetics400 [26], ensuring compatibility with existing dataloader implementations.

Table 13. List of coarse classes for Syn-TA

Coarse ID	Coarse motion
0	appearing
1	arriving
2	bouncing
3	changing color
4	changing shape
5	cloning
6	departing
7	disappearing
8	expanding
9	falling
10	following
11	merging
12	moving in a path
13	not colliding
14	occlusion
15	orbiting
16	rising
17	shooting projectile
18	shrinking
19	teleporting



Table 14. List of the fine classes in Syn-TA - Set 1

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
0	appearing	0	Single object appearing by increasing opacity
		1	Multiple objects all appearing simultaneously by increasing opacity
1	arriving	2	Single object arriving
		3	Multiple objects arriving simultaneously
		4	Single object arriving beside another single object
2	bouncing	5	Single ball bouncing on a plain surface eventually stops
		6	Multiple balls bouncing on plain surface at very different heights eventually stop
		7	Single object bouncing and going forward
		8	Single object imitating the motion of bouncing and going down stairs
3	changing color	9	Single object keeps changing colors
		10	Two objects changing colors by switching with each other
		11	Single object changing from bright to dark color
4	changing shape	12	Single object of other shape changes into pyramid
		13	Single object of other shape changes into cube
		14	Two objects changing shapes by switching with each other
5	cloning	15	Single object cloning into two identical objects both going in same direction
		16	Single object cloning into multiple identical objects one by one
		17	Single object cloning into one identical object and another object with different colour
6	departing	18	Single object departing
		19	Multiple objects departing simultaneously
		20	Single object departing beside another single object
7	disappearing	21	Single object disappearing by decreasing opacity
		22	Multiple objects all disappearing simultaneously by decreasing opacity
8	expanding	23	Small single object expanding into medium object
		24	Two different small objects expanding into similar larger size
9	falling	25	Single object falling straight down without changing direction
		26	Single object falling down in a zigzag pattern
10	following	27	Single object following another object all the time while moving in a straight line
		28	Single object following another object for some time while moving in straight line then follows another path
		29	Multiple objects following one object all the time in a straight line
11	merging	30	Two objects merging into one similar sized object while moving towards each other

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
12	moving in a path	31	Multiple objects merging into one big object
		32	Single object moving slowly from one position to another and stopping
		33	Single object moving slowly from one position to another and coming back to original position
		34	Single object moving from left to right
13	not colliding	35	Two objects moving towards each other but not colliding
		36	Single object moving towards single stationary object but not colliding with it
		37	Multiple objects moving towards each other but not colliding among themselves
14	occlusion	38	Small object coming in front of big object barely occluding it
		39	Object which is occluding another object switches places and goes behind it
15	orbiting	40	Single object orbiting another object in the center in clockwise direction on horizontal plane
		41	Single object orbiting another object in the center in clockwise direction first and then counter clockwise direction on horizontal plane
		42	Single object orbiting another object in the center in clockwise direction on vertical plane
16	rising	43	Single object rising straight up without changing direction
		44	Single object rising in a zigzag pattern
17	shooting projectile	45	Shooting projectile from one corner to opposite corner
		46	Shooting projectile from right corner to left corner
		47	Shooting two projectiles from same position to opposite corner
18	shrinking	48	Large single object shrinking into medium sized object
		49	Two different large objects shrinking into same smaller size
19	teleporting	50	Single object teleporting from one corner to another
		51	Single object teleporting from one corner to another while another object is in the center
		52	Two objects in opposite corners switch places by teleporting

Table 15. List of the classes in Syn-TA - Set 2

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
0	appearing	0	Multiple objects all appearing one by one by increasing opacity
		1	Two objects appearing among two other objects
1	arriving	2	Multiple objects arriving one by one
		3	Single object arriving among many objects

Continued on next page



Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
2	bouncing	4	Single object arriving beside another single object, pausing then going away
		5	Multiple objects bouncing on plain surface at almost similar heights eventually stop
		6	Single object imitating the motion of bouncing against a wall
		7	Single object bouncing on a plain surface perpetually
		8	Single object imitating the motion of bouncing and going up stairs
3	changing color	9	Single object changing from dark to bright color
		10	Single object changing color then returning to previous one
4	changing shape	11	Single object of other shape changes into sphere
		12	Single object of other shape changes into cylinder
		13	Single object changing into another shape then turning back to original shape
5	cloning	14	Single object cloning into two identical objects both going in different directions
		15	Single object cloning into multiple identical objects at once
6	departing	16	Multiple objects departing one by one
		17	Single object departing among many objects
		18	Single object departing beside another single object, then coming back shortly
7	disappearing	19	Multiple objects all disappearing one by one by decreasing opacity
		20	Multiple objects some disappearing and some remaining
8	expanding	21	Small single object expanding into large object
		22	Two same small objects expanding into different bigger sizes
9	falling	23	Single object falling at an angle
		24	Single object falling at a curved path
10	following	25	Single object following another object all the time in a complex path
		26	Single object not following another object at first but later starts following
11	merging	27	Faster object merging with slower object into one similar sized object while moving in same direction
		28	Multiple objects merging into multiple big objects
12	moving in a path	29	Single object moving quickly from one position to another and stopping
		30	Two objects moving and switching their positions
		31	Single object moving from right to left
13	not colliding	32	Two objects moving in opposite directions in circular path but not colliding later
		33	One object moving towards multiple stationary objects but not colliding
Continued on next page			

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
14	occlusion	34	Big object coming in front of small object mostly or fully occluding it
		35	Two objects partially occluding two other objects switch positions with the latter two and go behind them
15	orbiting	36	Single object orbiting another object in the center in counter clockwise direction on horizontal plane
		37	Single object orbiting another object in clockwise direction then second object orbiting first object in clockwise direction on horizontal plane
		38	Single object orbiting another object in the center in counter clockwise direction on vertical plane
16	rising	39	Single object rising at an angle
		40	Single object rising at a curved path
17	shooting projectile	41	Shooting projectile from left corner to right corner
		42	Shooting two projectiles from opposite corners switching their positions
18	shrinking	43	Large single object shrinking into very small object
		44	Two different large objects shrinking into different smaller sizes
19	teleporting	45	Four objects in four corners switch places by teleporting
		46	Two objects together in one corner then one teleports to other corner

## D.2. Kinetics400 - Transferable Activity

Table 16. List of coarse classes for K400-TA

Coarse ID	Coarse motion
0	blowing with mouth
1	building a structure
2	catching or throwing other objects than ball
3	cleaning
4	climbing
5	cooking
6	dancing
7	drinking
8	eating
9	feeding
10	fishing
11	folding
12	gardening
13	grooming animal
14	hair care
15	juggling
16	jumping
17	kicking
18	maintaining vehicle
19	massaging

Coarse ID	Coarse motion
20	music without instrument
21	opening something
22	petting animal
23	playing by hitting ball with something
24	playing instrument
25	playing with ball in hand
26	preparing fruit or vegetable
27	punching
28	pushing
29	reading
30	riding animal
31	riding on something over water
32	riding vehicle
33	shaving
34	skiing
35	smoking
36	swimming
37	toddler interaction
38	tying
39	washing
40	waxing

Table 17. List of the fine classes in K400-TA Set 1

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
0	blowing with mouth	0	balloon blowing
1	building a structure	1	building cabinet
2	catching or throwing other objects than ball	2	catching or throwing frisbee
		3	throwing axe
		4	cleaning floor
3	cleaning	5	cleaning gutters
		6	cleaning pool
		7	climbing a rope
4	climbing	8	climbing ladder
		9	baking cookies
		10	barbequing
5	cooking	11	cooking chicken
		12	cooking egg
		13	belly dancing
6	dancing	14	breakdancing
		15	country line dancing
		16	dancing ballet
		17	dancing charleston
		18	dancing gangnam style
		19	dancing macarena
		20	jumpstyle dancing
		21	drinking beer
7	drinking	22	drinking shots
8	eating	23	dining

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
		24	eating burger
		25	eating cake
		26	eating carrots
		27	eating chips
		28	eating doughnuts
9	feeding	29	feeding birds
10	fishing	30	feeding fish
11	folding	31	catching fish
		32	folding clothes
12	gardening	33	folding napkins
		34	arranging flowers
		35	planting trees
13	grooming animal	36	stomping grapes
14	hair care	37	grooming dog
		38	braiding hair
		39	brushing hair
		40	curling hair
		41	dying hair
15	juggling	42	contact juggling
16	jumping	43	juggling balls
		44	high jump
17	kicking	45	drop kicking
		46	high kick
		47	kicking field goal
		48	kicking soccer ball
18	maintaining vehicle	49	changing oil
		50	changing wheel
19	massaging	51	massaging back
		52	massaging feet
20	music without instrument	53	air drumming
		54	beatboxing
21	opening something	55	opening bottle
22	petting animal	56	petting animal (not cat)
23	playing by hitting ball with something	57	golf chipping
		58	golf driving
		59	hitting baseball
24	playing instrument	60	busking
		61	drumming fingers
		62	playing accordion
		63	playing bagpipes
		64	playing bass guitar
		65	playing cello
		66	playing clarinet
		67	playing cymbals
		68	playing didgeridoo
		69	playing drums
		70	playing flute
		71	playing guitar
		72	playing harmonica
25	playing with ball in hand	73	catching or throwing baseball
		74	dodgeball

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
		75	dribbling basketball
		76	dunking basketball
		77	passing American football (in game)
		78	throwing ball
26	preparing fruit or vegetable	79	cutting pineapple
		80	peeling apples
27	punching	81	punching bag
		82	pushing car
28	pushing	83	pushing cart
29	reading	84	reading book
		85	riding camel
30	riding animal	86	riding elephant
		87	crossing river
31	riding on something over water	88	sailing
		89	surfing water
		90	biking through snow
		91	driving car
32	riding vehicle	92	motorcycling
		93	riding scooter
		94	snowmobiling
33	shaving	95	shaving head
		96	ski jumping
34	skiing	97	skiing (not slalom or crosscountry)
		98	water skiing
35	smoking	99	smoking cigarette or short object
		100	scuba diving
36	swimming	101	swimming backstroke
		102	swimming breast stroke
		103	baby waking up
37	toddler interaction	104	carrying baby
		105	tying bow tie
38	tying	106	tying knot (not on a tie)
		107	washing dishes
39	washing	108	washing feet
		109	waxing back
40	waxing	110	waxing chest

Table 18. List of the fine classes in K400-TA Set 2

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
0	blowing with mouth	0	blowing out candles
1	building a structure	1	building shed
2	catching or throwing other objects than ball	2	throwing discus
		3	cleaning shoes
3	cleaning	4	cleaning toilet
		5	cleaning windows
		6	climbing tree
4	climbing	7	ice climbing
5	cooking	8	cooking on campfire

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
6	dancing	9	cooking sausages
		10	flipping pancake
		11	frying vegetables
		12	krumping
		13	robot dancing
		14	salsa dancing
		15	swing dancing
		16	tango dancing
7	drinking	17	tap dancing
		18	zumba
8	eating	19	drinking water or soft drinks
		20	tasting beer
		21	eating hotdog
		22	eating ice cream
		23	eating spaghetti
9	feeding	24	eating watermelon
		25	tasting food
10	fishing	26	feeding goats
11	folding	27	ice fishing
12	gardenning	28	folding paper
		29	trimming trees
13	grooming animal	30	watering plants
		31	grooming horse
14	hair care	32	fixing hair
		33	getting a haircut
		34	trimming or shaving beard
15	juggling	35	juggling fire
		36	juggling soccer ball
16	jumping	37	jumping into pool
		38	playing kickball
17	kicking	39	shooting goal (soccer)
		40	side kick
18	maintaining vehicle	41	checking tires
		42	pumping gas
19	massaging	43	massaging legs
		44	massaging person's head
20	music without instrument	45	singing
21	opening something	46	opening present
22	petting animal	47	petting cat
23	playing by hitting ball with something	48	golf putting
		49	playing cricket
24	playing instrument	50	playing harp
		51	playing keyboard
		52	playing organ
		53	playing piano
		54	playing recorder
		55	playing saxophone
		56	playing trombone
		57	playing trumpet
		58	playing ukulele

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
		59	playing violin
		60	playing xylophone
		61	strumming guitar
		62	tapping guitar
25	playing with ball in hand	63	catching or throwing softball
		64	passing American football (not in game)
		65	playing basketball
		66	playing volleyball
		67	shooting basketball
26	preparing fruit or vegetable	68	cutting watermelon
		69	peeling potatoes
27	punching	70	punching person (boxing)
28	pushing	71	pushing wheelchair
29	reading	72	reading newspaper
30	riding animal	73	riding mule
		74	riding or walking with horse
31	riding on something over water	75	water sliding
		76	windsurfing
32	riding vehicle	77	driving tractor
		78	riding a bike
		79	riding mountain bike
		80	riding unicycle
		81	using segway
33	shaving	82	shaving legs
34	skiing	83	skiing crosscountry
		84	skiing slalom
35	smoking	85	smoking hookah
36	swimming	86	snorkeling
		87	swimming butterfly stroke
37	toddler interaction	88	crawling baby
38	tying	89	tying tie
39	washing	90	washing hair
		91	washing hands
40	waxing	92	waxing eyebrows
		93	waxing legs

### D.3. Something-something-v2 - Transferable Activity

Table 19. List of coarse classes for SSv2-TA

Coarse ID	Coarse motion
0	Bending
1	Dropping
2	Holding
3	Letting to roll
4	Lifting
5	Moving
6	Plugging
7	Poking
8	Pouring



Coarse ID	Coarse motion
9	Pretending
10	Pulling
11	Pushing
12	Putting
13	Showing
14	Colliding deflected
15	Falling
16	Spilling
17	Spinning
18	Taking
19	Tearing
20	Throwing
21	Tilting
22	Tipping
23	Trying but failing
24	Turning
25	Twisting

Table 20. List of the fine classes in SSv2-TA Set 1

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
0	Bending	0	Bending something so that it deforms
1	Dropping	1	Dropping something behind something
		2	Dropping something in front of something
		3	Dropping something into something
2	Holding	4	Holding something
		5	Holding something behind something
		6	Holding something in front of something
3	Letting to roll	7	Letting something roll along a flat surface
		8	Letting something roll down a slanted surface
4	Lifting	9	Lifting a surface with something on it but not enough for it to slide down
		10	Lifting a surface with something on it until it starts sliding down
		11	Lifting something up completely without letting it drop down
		12	Lifting something up completely, then letting it drop down
5	Moving	13	Moving away from something with your camera
		14	Moving part of something
		15	Moving something across a surface until it falls down
		16	Moving something across a surface without it falling down
		17	Moving something and something away from each other
		18	Moving something and something closer to each other

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
6	Plugging	19	Moving something and something so they collide with each other
		20	Plugging something into something
7	Poking	21	Poking a hole into some substance
		22	Poking a hole into something soft
		23	Poking a stack of something so the stack collapses
		24	Poking a stack of something without the stack collapsing
8	Pouring	25	Pouring something into something
		26	Pouring something into something until it overflows
9	Pretending	27	Pretending or failing to wipe something off of something
		28	Pretending or trying and failing to twist something
		29	Pretending to be tearing something that is not tearable
		30	Pretending to close something without actually closing it
		31	Pretending to open something without actually opening it
		32	Pretending to pick something up
		33	Pretending to poke something
		34	Pretending to pour something out of something, but something is empty
		35	Pretending to put something behind something
		36	Pretending to put something into something
10	Pulling	37	Pretending to put something next to something
		38	Pulling something from behind of something
		39	Pulling something from left to right
		40	Pulling something from right to left
		41	Pulling something onto something
11	Pushing	42	Pushing something from left to right
		43	Pushing something from right to left
		44	Pushing something off of something
		45	Pushing something onto something
		46	Pushing something so it spins
12	Putting	47	Putting number of something onto something
		48	Putting something and something on the table
		49	Putting something behind something
		50	Putting something in front of something
		51	Putting something into something
		52	Putting something next to something
		53	Putting something on a flat surface without letting it roll
		54	Putting something on a surface
		55	Putting something on the edge of something so it is not supported and falls down
13	Showing	56	Putting something onto a slanted surface but it doesn't glide down
		57	Showing a photo of something to the camera
		58	Showing something behind something

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
14	Colliding deflected	59	Showing something next to something
		60	Showing something on top of something
		61	Something being deflected from something
		62	Something colliding with something and both are being deflected
15	Falling	63	Something falling like a feather or paper
16	Spilling	64	Spilling something behind something
		65	Spilling something next to something
17	Spinning	66	Spinning something so it continues spinning
18	Taking	67	Taking one of many similar things on the table
		68	Taking something from somewhere
19	Tearing	69	Tearing something into two pieces
20	Throwing	70	Throwing something
		71	Throwing something against something
		72	Throwing something in the air and catching it
21	Tilting	73	Tilting something with something on it slightly so it doesn't fall down
22	Tipping	74	Tipping something over
23	Trying but failing	75	Trying but failing to attach something to something because it doesn't stick
		76	Trying to bend something unbendable so nothing happens
24	Turning	77	Turning something upside down
		78	Turning the camera downwards while filming something
		79	Turning the camera left while filming something
25	Twisting	80	Twisting (wringing) something wet until water comes out

Table 21. List of the fine classes in SSv2-TA Set 2

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
0	Bending	0	Bending something until it breaks
1	Dropping	1	Dropping something next to something
		2	Dropping something onto something
2	Holding	3	Holding something next to something
		4	Holding something over something
3	Letting to roll	5	Letting something roll up a slanted surface, so it rolls back down
4	Lifting	6	Lifting something with something on it
		7	Lifting up one end of something without letting it drop down
		8	Lifting up one end of something, then letting it drop down
5	Moving	9	Moving something and something so they pass each other
		10	Moving something away from something
		11	Moving something away from the camera
		12	Moving something closer to something
		13	Moving something down

Continued on next page

Coarse ID	Coarse motion	Fine ID	Coarse motion + fine motion
6	Plugging	14	Moving something towards the camera
		15	Moving something up
		16	Plugging something into something but pulling it right out as you remove your hand
7	Poking	17	Poking something so it slightly moves
		18	Poking something so lightly that it doesn't or almost doesn't move
		19	Poking something so that it falls over
		20	Poking something so that it spins around
8	Pouring	21	Pouring something onto something
		22	Pouring something out of something
9	Pretending	23	Pretending to put something on a surface
		24	Pretending to put something onto something
		25	Pretending to put something underneath something
		26	Pretending to scoop something up with something
		27	Pretending to spread air onto something
		28	Pretending to sprinkle air onto something
		29	Pretending to squeeze something
		30	Pretending to take something from somewhere
		31	Pretending to take something out of something
		32	Pretending to throw something
		33	Pretending to turn something upside down
10	Pulling	34	Pulling something out of something
		35	Pulling two ends of something but nothing happens
		36	Pulling two ends of something so that it gets stretched
		37	Pulling two ends of something so that it separates into two pieces
11	Pushing	38	Pushing something so that it almost falls off but doesn't
		39	Pushing something so that it falls off the table
		40	Pushing something so that it slightly moves
		41	Pushing something with something
12	Putting	42	Putting something onto something
		43	Putting something onto something else that cannot support it so it falls down
		44	Putting something similar to other things that are already on the table
		45	Putting something that can't roll onto a slanted surface, so it slides down
		46	Putting something that can't roll onto a slanted surface, so it stays where it is
		47	Putting something that cannot actually stand upright upright on the table, so it falls on its side
		48	Putting something underneath something
		49	Putting something upright on the table
		50	Putting something, something and something on the table
13	Showing	51	Showing something to the camera

Continued on next page

<b>Coarse ID</b>	<b>Coarse motion</b>	<b>Fine ID</b>	<b>Coarse motion + fine motion</b>
14	Colliding deflected	52	Showing that something is empty
		53	Showing that something is inside something
		54	Something colliding with something and both come to a halt
15	Falling	55	Something falling like a rock
16	Spilling	56	Spilling something onto something
17	Spinning	57	Spinning something that quickly stops spinning
18	Taking	58	Taking something out of something
19	Tearing	59	Tearing something just a little bit
20	Throwing	60	Throwing something in the air and letting it fall
		61	Throwing something onto a surface
21	Tilting	62	Tilting something with something on it until it falls off
22	Tipping	63	Tipping something with something in it over, so something in it falls out
23	Trying but failing	64	Trying to pour something into something, but missing so it spills next to it
24	Turning	65	Turning the camera right while filming something
		66	Turning the camera upwards while filming something
25	Twisting	67	Twisting something