

UniFormerV2: Unlocking the Potential of Image ViTs for Video Understanding

Supplementary Material

Dataset	Training #Samples	Validation #Samples	Average Length	#Actions
Kinetics-710 (ours)	658,340	66,803	10s	710
Kinetics-400 [36]	240,436 246,245*	19,787 20,000*	10s	400
Kinetics-600 [9]	366,006 392,622*	27,935 30,000*	10s	600
Kinetics-700 [10]	529,573 545,317*	33,861 35,000*	10s	700
MiT V1 [50]	802,244 802,264*	33,899 33,900*	3s	339
SthSth V1 [24]	86,017	11,522	4.0s	174
SthSth V2 [24]	168,913	24,777	4.0s	174
ActivityNet [27]	10,024	4,926	117s	200
HACS [89]	37,452	5,953	149s	200

Table 14: **Dataset descriptions.** * indicates the original video number.

A. Additional implement details

Datasets. In Table 14, we give more details of our datasets. *Kinetics* family [36] is the most widely-used benchmark, includes Kinetics-400 [36], Kinetics-600 [9] and Kinetics-700 [10]. Since some videos are unavailable on YouTube, the Kinetics datasets are gradually shrinking over time. We report the video number of our version for a more fair comparison. *Moments in Time V1* [50] contains 0.8M 3-second video clips annotated with 339 classes, which suggests capturing the gist of a dynamic scene. *Something-Something V1/V2* [24] consist of 174 actions interacted with everyday objects. They require strong temporal modeling to distinguish confusing actions such as opening/closing something. *ActivityNet* [27] and *HACS* [89] are two large-scale untrimmed video benchmark. They respectively contain about 20K and 50K videos in 200 human daily living actions. For these two datasets, we sample those video clips containing action for training, thus we do not add another background class. When testing, we sample the frames sparsely from the whole untrimmed videos.

Implementation Details. For the scene-related datasets, we only insert the global UniBlocks in the last 4 layers of ViT-B/L to perform multi-stage fusion, since the local UniBlocks and temporal downsampling do not further improve the results in Table 1. But for Something-Something V1/V2, we adopt all the designs and insert the global UniBlocks in the last 8/16 layers of ViT-B/L for better temporal modeling. Besides, when finetuning those models with large-scale dataset pretraining, it is necessary to initialize the new parameters properly. For stable training, we zero initialize some of the layers, including the

	K710	K	M	A&H	SS
<i>Optimization</i>					
Optimizer		AdamW [47]			
Momentum		$\beta_1, \beta_2 = 0.9, 0.999$			
Weight decay		0.05			
Learning rate schedule		cosine decay [48]			
Start learning rate		1e-6			
End learning rate		1e-6			
Batch size	512	256	512	64	128
Learning rate (Base)	2e-5	2e-6	2e-5	-	4e-5
Learning rate (Large)	1e-5	1.5e-6	1e-5	5e-6	2e-5
Warmup epochs [23]	5	1	5	5	5
Total epochs (Base)	55	5	24	-	22
Total epochs (Large)	40	5	18	20	15
<i>Data augmentation</i>					
Inception-style cropping [62]					
Scale		[0.08, 1.00]			
Jitter aspect ratio		[0.75, 1.33]			
Color jitter		0.4			
Rand augment [15]		rand-m7-n4-mstd0.5-inc1			
Repeated sampling [29]	1	1	1	2	2
<i>Regularisation</i>					
Dropout [61]					
Backbone			0.5		
Global branch			0.5		
Drop path [31]					
Backbone	-	-	-	0.2	0.2
Global branch	-	-	-	0.4	0.4

Table 15: **Training hyperparameters for our experiments.** “-” indicates that the related method is not used. Values constant in all the datasets are listed once. Datasets are denoted as follows: K (Kinetics), M (MiT V1), A&H (ActivityNet&HACS), SS (SthSth V1&V2).

last point-wise convolutions in the local temporal MHRA, the query tokens and output projection layers in the query-based cross MHRA, the last linear layers in the FFN of the global UniBlock, and the learnable fusion weights. What’s more, we provide the detailed hyperparameters in Table 15. Most of the training scripts follow UniFormer [39], but differently, we do not apply Mixup [87], CutMix [84], Label Smoothing [63] and Random Erasing [91]. When finetuning the full models on Kinetics directly from image pre-training, we adopt the same hyperparameters as in K710 pretraining. If the backbone is frozen, we use a larger learning rate (4e-4) without warmup.

Training Cost. In Table 6, we compare different training scripts. When finetuning Kinetics-400, 600 and 700 individually, we train the models for 55 epochs, and the total training data is about $0.24 + 0.366 + 0.529 \approx 1.14\text{M}$. When pretraining with Kinetics-710 (0.66M), we only finetune the models for 5 epochs. Thus the percentage of saving cost is

Method	Frame× Crop×Clip	Param (M)	FLOPs (T)	SSV1		SSV2	
				Top-1	Top-5	Top-1	Top-5
UniFormerV2-B/16	16×3×1	163	0.6	56.8	84.2	69.5	92.3
UniFormerV2-B/16	16×3×2	163	1.1	57.2	84.3	69.7	92.5
UniFormerV2-B/16	32×3×1	163	1.1	59.4	86.2	70.7	93.2
UniFormerV2-B/16	32×3×2	163	2.2	59.5	86.2	71.0	93.2
UniFormerV2-L/14	16×3×1	574	2.6	60.5	86.5	72.1	93.6
UniFormerV2-L/14	16×3×2	574	5.2	60.9	86.8	72.2	93.7
UniFormerV2-L/14	32×3×1	574	5.2	62.7	88.0	73.0	94.5
UniFormerV2-L/14	32×3×2	574	10.3	62.9	88.3	73.1	94.5

Table 16: **More results on Something-Something.** All models are directly finetuned from CLIP.

Method	Frame× Crop×Clip	K400
Only Global	8×3×4	81.8
Local+Global	8×3×4	84.4

Table 17: **Output token combination.**

Method	Frame× Crop×Clip	SSV1	SSV2
CLIP-400M	16×3×1	56.8	69.5
CLIP-400M+K400	16×3×1	55.8	68.4

Table 18: **K400 pretraining.**

Method	Param (M)	SSV2
Mean Pooling [72]	86	45.1
Temporal Shift [43]	86	65.7
Divided Space-Time MHSA [4]	114	63.4
Joint Space-Time MHSA [4]	86	65.8
Temporal Transformer [49]	128	61.5
Temporal Convolution [69]	86	65.6
Our Local MHRA	97	67.3

Table 19: **Different local modules.**

as follows,

$$1 - \frac{0.66 \times 55 + 1.14 \times 5}{1.14 \times 55} \approx 0.33 \quad (14)$$

Thus we save almost 33% of the training cost. More importantly, for the models with more frames (16, 32, or even 64), we only need to finetune the K710 pretrained models with 8 frames. Our training scripts are very efficient and effective for the Kinetics family.

B. More ablation studies

We conduct more ablation studies based on CLIP-ViT-B/16 [56].

Output token combination. When only using the global token for classification, the top-1 accuracy drops from 84.4% to 81.8% in Table 17. It shows that both local and global output tokens are essential for maintaining performance.

Kinetics pretraining for Something-Something. Different from the prior works [39, 20], in Table 18, we find that extra Kinetics pretraining harms the representation inherited from CLIP, leading to lower performance.

Different modules. In Table 19, we compare our local MHRA with popular temporal modules, including simple mean pooling [72], divided and joint space-time MHSA [4], temporal convolution [69], temporal shift [43] and temporal transformer [59]. All the modules are inserted before all the spatial MHSA, except that the 6-layer temporal transformer is added after the backbone. The results show that our local MHRA beats the previous methods, achieving 1.5% to

22.2% higher top-1 accuracy. It demonstrates the effectiveness of our local MHRA for temporal modeling.

C. Additional results

In Table 20, Table 21, Table 16 and Table 22, we give more results on the 8 video benchmarks, i.e., Kinetics-400/600/700, Moments in Time V1, Something-Something V1/V2, ActivityNet and HACs.

D. Label list of Kinetics-710

To generate our Kinetics-710, we align labels in different Kinetics datasets by filtering symbols and replacing synonyms. The final label list is shown in Table 23. Compared with Kinetics-700, there are 8 and 2 unique labels in Kinetics-400 and Kinetics-600 respectively. When finetuning the models pretrained on Kinetics-710, it is vital to load the pretrained weight of the classification layer, thus we map the weight according to the label list.

Method	Pretraining Data	Frame× Crop×Clip	Param (M)	FLOPs (T)	K400		K600		K700	
					Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
UniFormerV2-B/16	CLIP-400M	8×1×3	115	0.4	84.0	96.3	84.8	96.8	75.4	92.6
UniFormerV2-B/16		8×3×4	115	1.6	84.4	96.3	85.0	97.0	75.8	92.8
UniFormerV2-L/14		8×1×3	354	2.0	87.3	97.7	87.8	97.6	80.0	95.0
UniFormerV2-L/14		8×3×4	354	8.0	87.7	97.9	88.0	97.7	80.3	95.2
UniFormerV2-B/16	CLIP-400M+K710	8×1×3	115	0.4	85.2	96.7	85.6	97.0	75.8	92.4
UniFormerV2-B/16		8×3×4	115	1.8	85.6	97.0	86.1	97.2	76.3	92.7
UniFormerV2-L/14		8×1×3	354	2.0	88.4	97.9	88.6	98.1	80.4	95.2
UniFormerV2-L/14		8×3×4	354	8.0	88.8	98.1	89.0	98.2	80.8	95.4
UniFormerV2-L/14		16×3×1	354	4.0	88.9	98.0	89.2	98.2	80.9	95.4
UniFormerV2-L/14		16×3×4	354	16.0	89.1	98.2	89.4	98.3	81.2	95.6
UniFormerV2-L/14		32×3×1	354	16.0	89.2	98.2	89.3	98.2	81.3	95.6
UniFormerV2-L/14		32×3×2	354	16.0	89.3	98.2	89.5	98.3	81.5	95.7
UniFormerV2-L/14		32×3×4	354	32.0	89.5	98.2	89.5	98.3	81.4	95.8
UniFormerV2-L/14 336↑		32×3×2	354	37.6	89.7	98.3	89.9	98.5	82.1	96.1
UniFormerV2-L/14 336↑		32×3×4	354	75.3	89.7	98.3	89.9	98.5	82.2	96.1
UniFormerV2-L/14 336↑		64×3×2	354	75.3	90.0	98.4	90.1	98.5	82.7	96.2
UniFormerV2-L/14 336↑		64×3×4	354	150.6	90.0	98.4	90.1	98.5	82.7	96.3
UniFormerV2-L/14 (frozen) 336↑	CLIP-400M	8×1×3	51	4.7	86.7	93.4	87.4	97.7	79.6	94.6
UniFormerV2-L/14 (frozen) 336↑	CLIP-400M+K710	8×1×3	51	4.7	87.8	98.0	88.2	98.0	79.7	94.7
UniFormerV2-L/14 (frozen) 336↑		32×3×1	51	18.8	88.8	98.1	89.1	98.2	80.6	95.2
UniFormerV2-L/14 (frozen) 336↑		32×3×4	51	75.3	88.9	98.2	89.2	98.2	80.8	95.4

Table 20: More results on Kinetics-400, 600 and 700.

Method	Pretraining Data	Frame× Crop×Clip	Param (M)	FLOPs (T)	MiT	
					Top-1	Top-5
UniFormerV2-B/16	CLIP-400M	8×3×4	115	1.8	42.2	71.3
UniFormerV2-B/16	CLIP-400M	32×3×4	115	7.2	42.2	71.5
UniFormerV2-B/16	CLIP-400M+K710	8×3×4	115	1.8	42.6	71.6
UniFormerV2-B/16	CLIP-400M+K710+K400	8×3×4	115	1.8	42.6	71.7
UniFormerV2-B/16	CLIP-400M+K710+K700	8×3×4	115	1.8	42.4	71.2
UniFormerV2-L/14	CLIP-400M	8×3×4	354	8.0	46.2	76.0
UniFormerV2-L/14	CLIP-400M	16×3×4	354	16.0	46.2	76.2
UniFormerV2-L/14	CLIP-400M	32×3×4	354	32.0	46.4	76.2
UniFormerV2-L/14	CLIP-400M+K710	8×3×4	354	8.0	46.7	76.2
UniFormerV2-L/14	CLIP-400M+K710+K400	8×3×4	354	8.0	47.0	76.1
UniFormerV2-L/14 336↑	CLIP-400M	8×3×4	354	18.8	47.2	76.5
UniFormerV2-L/14 336↑	CLIP-400M+K710	8×3×4	354	18.8	47.6	76.7
UniFormerV2-L/14 336↑	CLIP-400M+K710+K400	8×3×4	354	18.8	47.8	76.9

Table 21: More results on Moments in Time V1.

Dataset	Pretraining Data	#Frame	3×2		3×4		3×10	
			Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ActivityNet	CLIP-400M+K400	8	92.8	99.0	92.8	99.1	93.0	99.1
	CLIP-400M+K400	16	93.5	99.4	93.5	99.5	93.6	99.5
	CLIP-400M+K710+K400	16	93.9	99.4	94.1	99.5	94.3	99.5
	CLIP-400M+K710+K700	16	94.0	99.4	94.2	99.5	94.3	99.6
	CLIP-400M+K710+K400	32	94.3	99.6	94.5	99.6	94.7	99.5
HACS	CLIP-400M+K400	16	94.7	99.8	94.7	99.8	94.9	99.9
	CLIP-400M+K710+K400	16	95.3	99.9	95.2	99.8	95.5	99.8
	CLIP-400M+K710+K700	16	94.7	99.7	94.7	99.8	94.9	99.8
	CLIP-400M+K710+K400	32	95.2	99.8	95.3	99.8	95.4	99.8

Table 22: More results on ActivityNet and HACS. All models are based on UniFormerV2-L/14.

Table 23: Labels of Kinetics-710.

Label	K4	K6	K7	Label	K4	K6	K7	Label	K4	K6	K7
luge	×	✓	✓	krumping	✓	✓	✓	skiing mono	✓	✓	✓
yoga	✓	✓	✓	slapping	✓	✓	✓	ski jumping	✓	✓	✓
vault	✓	×	×	decoupage	×	×	✓	driving car	✓	✓	✓
squat	✓	✓	✓	arresting	×	×	✓	tap dancing	✓	✓	✓
lunge	✓	✓	✓	surveying	×	×	✓	hockey stop	✓	✓	✓
zumba	✓	✓	✓	fly tying	×	✓	✓	tobogganing	✓	✓	✓
situp	✓	✓	✓	capsizing	×	✓	✓	cooking egg	✓	✓	✓
sewing	×	✓	✓	tiptoeing	×	✓	✓	slacklining	✓	✓	✓
cumbia	×	✓	✓	using atm	×	✓	✓	pushing car	✓	✓	✓
crying	✓	✓	✓	waking up	×	✓	✓	ice skating	✓	✓	✓
dining	✓	✓	✓	fidgeting	×	✓	✓	ice fishing	✓	✓	✓
digging	✓	×	✓	tie dying	×	✓	✓	celebrating	✓	✓	✓
chasing	×	×	✓	wrestling	✓	✓	✓	windsurfing	✓	✓	✓
sieving	×	×	✓	whistling	✓	✓	✓	riding mule	✓	✓	✓
staring	×	✓	✓	high kick	✓	✓	✓	waxing legs	✓	✓	✓
karaoke	×	✓	✓	abseiling	✓	✓	✓	deadlifting	✓	✓	✓
burping	×	✓	✓	high jump	✓	✓	✓	bee keeping	✓	✓	✓
packing	×	✓	✓	trapezing	✓	✓	✓	pumping gas	✓	✓	✓
licking	×	✓	✓	skydiving	✓	✓	✓	tapping pen	✓	✓	✓
winking	×	✓	✓	bandaging	✓	✓	✓	headbanging	✓	✓	✓
arguing	×	✓	✓	side kick	✓	✓	✓	bookbinding	✓	✓	✓
ironing	✓	✓	✓	jetskiing	✓	✓	✓	flying kite	✓	✓	✓
drawing	✓	✓	✓	long jump	✓	✓	✓	fixing hair	✓	✓	✓
archery	✓	✓	✓	hopscotch	✓	✓	✓	egg hunting	✓	✓	✓
jogging	✓	✓	✓	dodgeball	✓	✓	✓	mowing lawn	✓	✓	✓
singing	✓	✓	✓	crocheting	×	×	✓	triple jump	✓	✓	✓
yawning	✓	✓	✓	ski ballet	×	×	✓	milking cow	✓	✓	✓
writing	✓	✓	✓	geocaching	×	✓	✓	doing nails	✓	✓	✓
push up	✓	✓	✓	bulldozing	×	✓	✓	dyeing hair	✓	✓	✓
tai chi	✓	✓	✓	cosplaying	×	✓	✓	eating cake	✓	✓	✓
sailing	✓	✓	✓	spelunking	×	✓	✓	paragliding	✓	✓	✓
welding	✓	✓	✓	jaywalking	×	✓	✓	headbutting	✓	✓	✓
smoking	✓	✓	✓	head stand	×	✓	✓	bobsledding	✓	✓	✓
parkour	✓	✓	✓	contorting	×	✓	✓	kitesurfing	✓	✓	✓
texting	✓	✓	✓	plastering	✓	✓	✓	petting cat	✓	✓	✓
bowling	✓	✓	✓	bartending	✓	✓	✓	waxing back	✓	✓	✓
kissing	✓	✓	✓	beatboxing	✓	✓	✓	making slime	×	×	✓
busking	✓	✓	✓	applauding	✓	✓	✓	steering car	×	×	✓
gargling	✓	×	✓	pole vault	✓	✓	✓	rolling eyes	×	×	✓
spraying	✓	×	✓	barbequing	✓	✓	✓	moving child	×	×	✓
coughing	×	×	✓	snowkiting	✓	✓	✓	pouring milk	×	×	✓
saluting	×	×	✓	making tea	✓	✓	✓	grooming cat	×	×	✓
shouting	×	×	✓	auctioning	✓	✓	✓	doing sudoku	×	×	✓
sleeping	×	✓	✓	snorkeling	✓	✓	✓	closing door	×	×	✓
smashing	×	✓	✓	testifying	✓	✓	✓	pouring wine	×	×	✓
tackling	×	✓	✓	high fiving	×	×	✓	cutting cake	×	×	✓
shopping	×	✓	✓	moving baby	×	×	✓	milking goat	×	×	✓
pinching	×	✓	✓	shoot dance	×	×	✓	playing oboe	×	×	✓
huddling	×	✓	✓	pirouetting	×	✓	✓	filling cake	×	×	✓
bottling	×	✓	✓	coloring in	×	✓	✓	sanding wood	×	×	✓
drooling	×	✓	✓	sawing wood	×	✓	✓	jumping sofa	×	×	✓
tickling	✓	✓	✓	calculating	×	✓	✓	taking photo	×	×	✓
knitting	✓	✓	✓	waving hand	×	✓	✓	silent disco	×	×	✓
unboxing	✓	✓	✓	watching tv	×	✓	✓	ironing hair	×	✓	✓
shot put	✓	✓	✓	calligraphy	×	✓	✓	planing wood	×	✓	✓
marching	✓	✓	✓	carving ice	×	✓	✓	gold panning	×	✓	✓
capoeira	✓	✓	✓	bodysurfing	×	✓	✓	pillow fight	×	✓	✓
pull ups	✓	✓	✓	lifting hat	×	✓	✓	combing hair	×	✓	✓
laughing	✓	✓	✓	bathing dog	×	✓	✓	laying stone	×	✓	✓
hurdling	✓	✓	✓	chewing gum	×	✓	✓	photobombing	×	✓	✓
sneezing	✓	✓	✓	parasailing	✓	✓	✓	playing lute	×	✓	✓
clapping	✓	✓	✓	sipping cup	✓	✓	✓	land sailing	×	✓	✓

Label	K4	K6	K7	Label	K4	K6	K7	Label	K4	K6	K7
scrapbooking	×	✓	✓	washing feet	✓	✓	✓	ripping paper	✓	✓	✓
tasting wine	×	✓	✓	diving cliff	✓	✓	✓	crawling baby	✓	✓	✓
docking boat	×	✓	✓	golf putting	✓	✓	✓	cleaning pool	✓	✓	✓
photocopying	×	✓	✓	motorcycling	✓	✓	✓	brushing hair	✓	✓	✓
clam digging	×	✓	✓	breakdancing	✓	✓	✓	sanding floor	✓	✓	✓
ice swimming	×	✓	✓	drinking beer	✓	×	×	belly dancing	✓	✓	✓
roasting pig	×	✓	✓	swinging legs	✓	×	×	feeding goats	✓	✓	✓
pouring beer	×	✓	✓	bull fighting	×	✓	×	shaking hands	✓	✓	✓
smoking pipe	×	✓	✓	tossing salad	✓	×	✓	swing dancing	✓	✓	✓
lock picking	×	✓	✓	playing cards	✓	×	✓	carrying baby	✓	✓	✓
steer roping	×	✓	✓	slicing onion	×	×	✓	bending metal	✓	✓	✓
hugging baby	×	✓	✓	stacking dice	×	×	✓	playing poker	✓	✓	✓
embroidering	×	✓	✓	helmet diving	×	×	✓	grinding meat	✓	✓	✓
longboarding	×	✓	✓	dealing cards	×	×	✓	shining shoes	✓	✓	✓
laying tiles	×	✓	✓	treating wood	×	×	✓	folding paper	✓	✓	✓
playing gong	×	✓	✓	eating nachos	×	×	✓	blasting sand	✓	✓	✓
base jumping	×	✓	✓	being excited	×	×	✓	arm wrestling	✓	✓	✓
playing polo	×	✓	✓	vacuuming car	×	×	✓	rock climbing	✓	✓	✓
moon walking	×	✓	✓	petting horse	×	×	✓	catching fish	✓	✓	✓
opening door	×	✓	✓	stacking cups	×	×	✓	playing drums	✓	✓	✓
tasting food	✓	✓	✓	poaching eggs	×	×	✓	cracking neck	✓	✓	✓
shaving legs	✓	✓	✓	yarn spinning	×	✓	✓	tying necktie	✓	✓	✓
pumping fist	✓	✓	✓	card stacking	×	✓	✓	juggling fire	✓	✓	✓
making sushi	✓	✓	✓	rope pushdown	×	✓	✓	golf chipping	✓	✓	✓
snowmobiling	✓	✓	✓	smelling feet	×	✓	✓	javelin throw	✓	✓	✓
tasting beer	✓	✓	✓	card throwing	×	✓	✓	skateboarding	✓	✓	✓
golf driving	✓	✓	✓	playing darts	×	✓	✓	laying bricks	✓	✓	✓
waxing chest	✓	✓	✓	chopping meat	×	✓	✓	playing piano	✓	✓	✓
faceplanting	✓	✓	✓	making cheese	×	✓	✓	playing flute	✓	✓	✓
eating chips	✓	✓	✓	crossing eyes	×	✓	✓	salsa dancing	✓	✓	✓
playing harp	✓	✓	✓	cracking back	×	✓	✓	eating burger	✓	✓	✓
spinning poi	✓	✓	✓	building lego	×	✓	✓	skipping rope	✓	✓	✓
front raises	✓	✓	✓	using inhaler	×	✓	✓	climbing tree	✓	✓	✓
reading book	✓	✓	✓	jumping jacks	×	✓	✓	washing hands	✓	✓	✓
shaking head	✓	✓	✓	using puppets	×	✓	✓	playing chess	✓	✓	✓
snowboarding	✓	✓	✓	sucking lolly	×	✓	✓	tango dancing	✓	✓	✓
scuba diving	✓	✓	✓	cutting apple	×	✓	✓	using computer	✓	×	×
bending back	✓	✓	✓	lighting fire	×	✓	✓	cleaning floor	✓	×	×
drop kicking	✓	✓	✓	surfing water	✓	✓	✓	exercising arm	✓	×	✓
using segway	✓	✓	✓	playing organ	✓	✓	✓	baby waking up	✓	×	✓
ice climbing	✓	✓	✓	hoverboarding	✓	✓	✓	waxing armpits	×	×	✓
tossing coin	✓	✓	✓	feeding birds	✓	✓	✓	mixing colours	×	×	✓
cheerleading	✓	✓	✓	blowing glass	✓	✓	✓	carving marble	×	×	✓
blowing nose	✓	✓	✓	building shed	✓	✓	✓	peeling banana	×	×	✓
pushing cart	✓	✓	✓	setting table	✓	✓	✓	breaking glass	×	×	✓
water skiing	✓	✓	✓	doing laundry	✓	✓	✓	laying decking	×	×	✓
making pizza	✓	✓	✓	braiding hair	✓	✓	✓	brushing floor	×	×	✓
punching bag	✓	✓	✓	mopping floor	✓	✓	✓	herding cattle	×	×	✓
feeding fish	✓	✓	✓	tying bow tie	✓	✓	✓	blending fruit	×	×	✓
riding camel	✓	✓	✓	cutting nails	✓	✓	✓	seasoning food	×	×	✓
shaving head	✓	✓	✓	skiing slalom	✓	✓	✓	checking watch	×	×	✓
throwing axe	✓	✓	✓	making a cake	✓	✓	✓	massaging neck	×	✓	✓
grooming dog	✓	✓	✓	chopping wood	✓	✓	✓	leatherworking	×	✓	✓
curling hair	✓	✓	✓	somersaulting	✓	✓	✓	acting in play	×	✓	✓
air drumming	✓	✓	✓	riding a bike	✓	✓	✓	chiseling wood	×	✓	✓
training dog	✓	✓	✓	surfing crowd	✓	✓	✓	square dancing	×	✓	✓
disc golfing	✓	✓	✓	holding snake	✓	✓	✓	sausage making	×	✓	✓
hula hooping	✓	✓	✓	water sliding	✓	✓	✓	using a wrench	×	✓	✓
washing hair	✓	✓	✓	playing cello	✓	✓	✓	weaving fabric	×	✓	✓
cartwheeling	✓	✓	✓	throwing ball	✓	✓	✓	breathing fire	×	✓	✓
changing oil	✓	✓	✓	eating hotdog	✓	✓	✓	rolling pastry	×	✓	✓
hammer throw	✓	✓	✓	robot dancing	✓	✓	✓	cutting orange	×	✓	✓

Label	K4	K6	K7	Label	K4	K6	K7	Label	K4	K6	K7
needle felting	×	✓	✓	flipping bottle	×	×	✓	tagging graffiti	×	✓	✓
skipping stone	×	✓	✓	splashing water	×	×	✓	raising eyebrows	×	✓	✓
scrubbing face	×	✓	✓	carrying weight	×	×	✓	threading needle	×	✓	✓
flint knapping	×	✓	✓	spinning plates	×	×	✓	popping balloons	×	✓	✓
shuffling feet	×	✓	✓	fencing (sport)	×	✓	✓	cooking scallops	×	✓	✓
throwing knife	×	✓	✓	curling (sport)	×	✓	✓	backflip (human)	×	✓	✓
fixing bicycle	×	✓	✓	separating eggs	×	✓	✓	falling off bike	×	✓	✓
making bubbles	×	✓	✓	playing ocarina	×	✓	✓	playing scrabble	×	✓	✓
counting money	✓	✓	✓	playing netball	×	✓	✓	visiting the zoo	×	✓	✓
applying cream	✓	✓	✓	polishing metal	×	✓	✓	mosh pit dancing	×	✓	✓
blowing leaves	✓	✓	✓	jumping bicycle	×	✓	✓	shucking oysters	×	✓	✓
shoveling snow	✓	✓	✓	trimming shrubs	×	✓	✓	looking at phone	×	✓	✓
brush painting	✓	✓	✓	playing marbles	×	✓	✓	throwing tantrum	×	✓	✓
making the bed	✓	✓	✓	blowdrying hair	×	✓	✓	tying shoe laces	×	✓	✓
playing tennis	✓	✓	✓	dyeing eyebrows	×	✓	✓	dancing macarena	✓	✓	✓
playing violin	✓	✓	✓	laying concrete	×	✓	✓	playing bagpipes	✓	✓	✓
tapping guitar	✓	✓	✓	playing pinball	×	✓	✓	eating ice cream	✓	✓	✓
picking apples	✓	✓	✓	dumpster diving	×	✓	✓	playing monopoly	✓	✓	✓
doing aerobics	✓	✓	✓	putting on sari	×	✓	✓	flipping pancake	✓	✓	✓
drinking shots	✓	✓	✓	playing maracas	×	✓	✓	getting a tattoo	✓	✓	✓
bungee jumping	✓	✓	✓	delivering mail	×	✓	✓	building cabinet	✓	✓	✓
shearing sheep	✓	✓	✓	preparing salad	×	✓	✓	playing clarinet	✓	✓	✓
juggling balls	✓	✓	✓	vacuuming floor	×	✓	✓	eating spaghetti	✓	✓	✓
stretching arm	✓	✓	✓	chiseling stone	×	✓	✓	drumming fingers	✓	✓	✓
news anchoring	✓	✓	✓	breaking boards	×	✓	✓	eating doughnuts	✓	✓	✓
smoking hookah	✓	✓	✓	climbing ladder	✓	✓	✓	playing trombone	✓	✓	✓
massaging back	✓	✓	✓	hurling (sport)	✓	✓	✓	moving furniture	✓	✓	✓
weaving basket	✓	✓	✓	throwing discus	✓	✓	✓	contact juggling	✓	✓	✓
making snowman	✓	✓	✓	recording music	✓	✓	✓	playing recorder	✓	✓	✓
checking tires	✓	✓	✓	playing trumpet	✓	✓	✓	wrapping present	✓	✓	✓
planting trees	✓	✓	✓	sled dog racing	✓	✓	✓	hitting baseball	✓	✓	✓
spray painting	✓	✓	✓	stomping grapes	✓	✓	✓	playing kickball	✓	✓	✓
stretching leg	✓	✓	✓	carving pumpkin	✓	✓	✓	cleaning gutters	✓	✓	✓
clean and jerk	✓	✓	✓	unloading truck	✓	✓	✓	cleaning windows	✓	✓	✓
peeling apples	✓	✓	✓	watering plants	✓	✓	✓	peeling potatoes	✓	✓	✓
dancing ballet	✓	✓	✓	playing ukulele	✓	✓	✓	playing keyboard	✓	✓	✓
making jewelry	✓	✓	✓	cleaning toilet	✓	✓	✓	looking in mirror	×	×	✓
grooming horse	✓	✓	✓	folding napkins	✓	✓	✓	walking on stilts	×	×	✓
playing guitar	✓	✓	✓	playing cymbals	✓	✓	✓	playing billiards	×	×	✓
sword fighting	✓	✓	✓	riding unicycle	✓	✓	✓	curling eyelashes	×	×	✓
washing dishes	✓	✓	✓	playing cricket	✓	✓	✓	playing beer pong	×	✓	✓
roller skating	✓	✓	✓	climbing a rope	✓	✓	✓	directing traffic	×	✓	✓
massaging feet	✓	✓	✓	scrambling eggs	✓	✓	✓	twiddling fingers	×	✓	✓
cleaning shoes	✓	✓	✓	opening present	✓	✓	✓	marriage proposal	×	✓	✓
bench pressing	✓	✓	✓	folding clothes	✓	✓	✓	making horseshoes	×	✓	✓
riding scooter	✓	✓	✓	waiting in line	✓	✓	✓	cracking knuckles	×	✓	✓
sweeping floor	✓	✓	✓	finger snapping	✓	✓	✓	adjusting glasses	×	✓	✓
brushing teeth	✓	✓	✓	riding elephant	✓	✓	✓	tightrope walking	×	✓	✓
trimming trees	✓	✓	✓	waxing eyebrows	✓	✓	✓	playing laser tag	×	✓	✓
baking cookies	✓	✓	✓	shuffling cards	✓	✓	✓	installing carpet	×	✓	✓
massaging legs	✓	✓	✓	walking the dog	✓	✓	✓	lawn mower racing	×	✓	✓
crossing river	✓	✓	✓	driving tractor	✓	✓	✓	standing on hands	×	✓	✓
eating carrots	✓	✓	✓	strumming guitar	✓	×	×	playing pan pipes	×	✓	✓
taking a shower	✓	×	×	filling eyebrows	✓	×	✓	playing ping pong	×	✓	✓
cooking chicken	✓	×	✓	playing rounders	×	×	✓	falling off chair	×	✓	✓
shredding paper	✓	×	✓	squeezing orange	×	×	✓	playing blackjack	×	✓	✓
metal detecting	×	×	✓	making latte art	×	×	✓	mushroom foraging	×	✓	✓
lighting candle	×	×	✓	opening coconuts	×	×	✓	playing harmonica	✓	✓	✓
using megaphone	×	×	✓	playing checkers	×	×	✓	cutting pineapple	✓	✓	✓
playing piccolo	×	×	✓	sword swallowing	×	✓	✓	sharpening knives	✓	✓	✓
entering church	×	×	✓	playing dominoes	×	✓	✓	playing badminton	✓	✓	✓
playing mahjong	×	×	✓	putting on shoes	×	✓	✓	getting a haircut	✓	✓	✓

Label	K4	K6	K7	Label	K4	K6	K7	Label	K4	K6	K7
playing saxophone	✓	✓	✓	swimming backstroke	✓	✓	✓	putting wallpaper on wall	×	×	✓
making a sandwich	✓	✓	✓	skiing crosscountry	✓	✓	✓	playing american football	×	×	✓
playing xylophone	✓	✓	✓	answering questions	✓	✓	✓	carving wood with a knife	×	×	✓
reading newspaper	✓	✓	✓	assembling computer	✓	✓	✓	bouncing on bouncy castle	×	✓	✓
jumping into pool	✓	✓	✓	sticking tongue out	✓	✓	✓	putting in contact lenses	×	✓	✓
arranging flowers	✓	✓	✓	biking through snow	✓	✓	✓	archaeological excavation	×	✓	✓
frying vegetables	✓	✓	✓	playing bass guitar	✓	✓	✓	swimming butterfly stroke	✓	✓	✓
sharpening pencil	✓	✓	✓	shooting basketball	✓	✓	✓	tying knot (not on a tie)	✓	✓	✓
playing accordion	✓	✓	✓	blowing out candles	✓	✓	✓	person collecting garbage	✓	✓	✓
eating watermelon	✓	✓	✓	rock scissors paper	✓	✓	✓	trimming or shaving beard	✓	✓	✓
jumpstyle dancing	✓	✓	✓	riding mountain bike	✓	×	×	giving or receiving award	✓	✓	✓
playing paintball	✓	✓	✓	playing slot machine	×	×	✓	breeding or breadcrumbing	✓	✓	✓
playing nose flute	×	×	✓	swimming with sharks	×	×	✓	opening bottle (not wine)	✓	✓	✓
getting a piercing	×	✓	✓	playing shuffleboard	×	×	✓	sign language interpreting	✓	✓	✓
wading through mud	×	✓	✓	using a paint roller	×	✓	✓	mountain climber (exercise)	×	✓	✓
wood burning (art)	×	✓	✓	home roasting coffee	×	✓	✓	playing hand clapping games	×	✓	✓
using circular saw	×	✓	✓	battle rope training	×	✓	✓	presenting weather forecast	✓	✓	✓
assembling bicycle	×	✓	✓	changing gear in car	×	✓	✓	bouncing ball	×	×	✓
blowing bubble gum	×	✓	✓	swimming front crawl	×	✓	✓	(not juggling)	×	×	✓
repairing puncture	×	✓	✓	wading through water	×	✓	✓	changing wheel	✓	✓	✓
poking bellybutton	×	✓	✓	walking through snow	×	✓	✓	(not on bike)	✓	✓	✓
putting on mascara	×	✓	✓	attending conference	×	✓	✓	catching or throwing	✓	✓	✓
throwing snowballs	×	✓	✓	casting fishing line	×	✓	✓	frisbee	✓	✓	✓
riding snow blower	×	✓	✓	opening refrigerator	×	✓	✓	riding or walking	✓	✓	✓
shining flashlight	×	✓	✓	hand washing clothes	×	✓	✓	with horse	✓	✓	✓
using a microscope	×	✓	✓	playing field hockey	×	✓	✓	catching or	✓	✓	✓
kicking field goal	✓	✓	✓	juggling soccer ball	✓	✓	✓	throwing softball	✓	✓	✓
playing ice hockey	✓	✓	✓	dribbling basketball	✓	✓	✓	playing squash	✓	✓	✓
playing controller	✓	✓	✓	country line dancing	✓	✓	✓	or racquetball	✓	✓	✓
cutting watermelon	✓	✓	✓	canoeing or kayaking	✓	✓	✓	decorating the christmas	✓	✓	✓
dancing charleston	✓	✓	✓	running on treadmill	✓	✓	✓	tree	✓	✓	✓
hugging (not baby)	✓	✓	✓	walking with crutches	×	×	✓	catching or throwing	✓	✓	✓
springboard diving	✓	✓	✓	pulling espresso shot	×	×	✓	baseball	✓	✓	✓
playing basketball	✓	✓	✓	letting go of balloon	×	×	✓	exercising with	✓	✓	✓
dunking basketball	✓	✓	✓	being in zero gravity	×	×	✓	an exercise ball	✓	✓	✓
playing volleyball	✓	✓	✓	roasting marshmallows	×	✓	✓	passing American football	✓	✓	✓
playing didgeridoo	✓	✓	✓	using bagging machine	×	✓	✓	(in game)	✓	✓	✓
inflating balloons	✓	✓	✓	talking on cell phone	×	✓	✓	passing American football	✓	✓	✓
extinguishing fire	✓	✓	✓	putting on foundation	×	✓	✓	(not in game)	✓	✓	✓
pushing wheelchair	✓	✓	✓	using a sledge hammer	×	✓	✓				
chopping vegetables	×	✓	×	swinging baseball bat	×	✓	✓				
pulling rope (game)	×	×	✓	making balloon shapes	×	✓	✓				
picking blueberries	×	×	✓	dancing gangnam style	✓	✓	✓				
playing road hockey	×	×	✓	cooking sausages	✓	✓	✓				
uncorking champagne	×	×	✓	snatch weight lifting	✓	✓	✓				
polishing furniture	×	×	✓	swinging on something	✓	✓	✓				
playing with trains	×	✓	✓	swimming with dolphins	×	×	✓				
pushing wheelbarrow	×	✓	✓	shooting off fireworks	×	×	✓				
shaping bread dough	×	✓	✓	throwing water balloon	×	✓	✓				
alligator wrestling	×	✓	✓	historical reenactment	×	✓	✓				
building sandcastle	×	✓	✓	swimming breast stroke	✓	✓	✓				
doing jigsaw puzzle	×	✓	✓	bouncing on trampoline	✓	✓	✓				
opening wine bottle	×	✓	✓	shooting goal (soccer)	✓	✓	✓				
putting on eyeliner	×	✓	✓	riding mechanical bull	✓	✓	✓				
passing soccer ball	×	✓	✓	making paper aeroplanes	×	✓	✓				
playing rubiks cube	×	✓	✓	using remote controller	✓	✓	✓				
using a power drill	×	✓	✓	massaging person's head	✓	✓	✓				
putting on lipstick	×	✓	✓	gospel singing in church	×	✓	✓				
kicking soccer ball	✓	✓	✓	punching person (boxing)	✓	✓	✓				
cooking on campfire	✓	✓	✓	petting animal (not cat)	✓	✓	✓				
gymnastics tumbling	✓	✓	✓	pretending to be a statue	×	×	✓				
clay pottery making	✓	✓	✓	listening with headphones	×	×	✓				