1. Dataset

1.1. 10-selected classes

The mapping used for generating the classes used in the 10-class experiments is shown in Table 3. The 10 classes were defined by first selecting 41 classes by hand (based on class-definitions and on visual inspection of the videos) and subsequently remapping these into 10 groups.

Class names	Index
Dropping [something]	0
Holding [something]	1
Moving [something] from left to right	2
Moving [something] from right to left	3
Picking [something] up	4
Poking [something]	5
Pouring [something]	6
Putting [something]	7
Showing [something] (almost no hand):	8
Tearing [something]	9

Table 1: Subset of 10 selected classes used in some of the experiments.

Some example predictions from the 10-classes model are shown in the Figure 1

1.2. 40-selected classes

We took the above 10 selected classes and select 30 additional common classes to form this subset of data. The list of classes used is shown in the Table 2.

The confusion matrix for predictions on 40 selected classes using the best performing model is shown in Figure 2 and the corresponding dictionary to read classes off from the matrix is shown in the Table 2.

1.3. All data - 174 classes

The complete list of 174 classes and their corresponding action-groups is shown in Table 4.

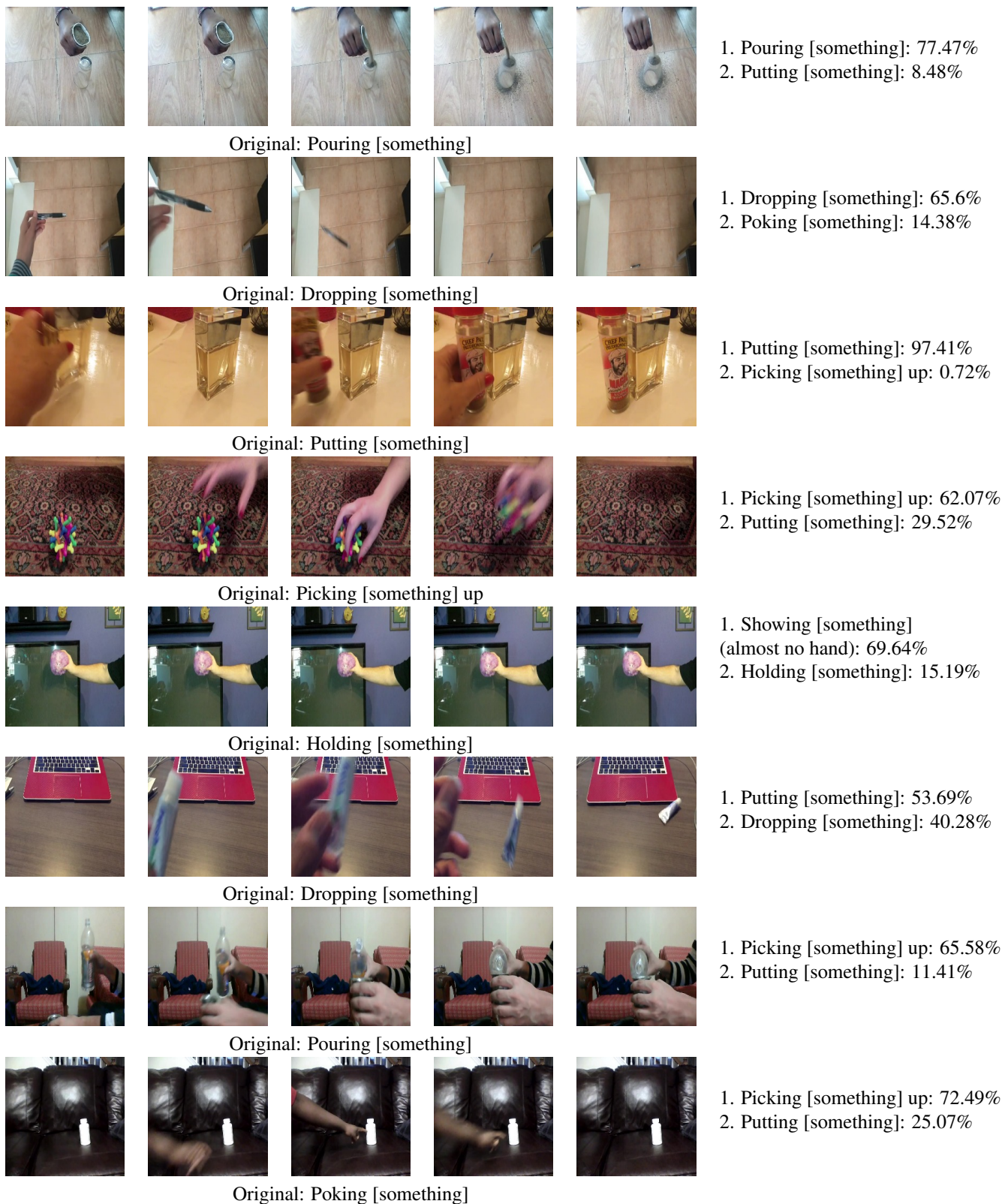| Class names	Index
Approaching [something] with your camera	0
Closing [something]	1
Dropping [something]	2
Folding [something]	3
Holding [something]	4
Holding [something] next to [something]	5
Moving [something] away from [something]	6
Moving [something] away from the camera	7
Moving [something] closer to [something]	8
Moving [something] down	9
Moving [something] from left to right	10
Moving [something] from right to left	11
Moving [something] towards the camera	12
Moving away from [something] with your camera	13
Opening [something]	14
Picking [something] up	15
Plugging [something] into [something]	16
Poking [something]	17
Pouring [something]	18
Pretending to pick [something] up	19
Pretending to put [something] next to [something]	20
Pretending to put [something] on a surface	21
Pretending to take [something] from [somewhere]	22
Pushing [something] so that it slightly moves	23
Pushing [something] with [something]	24
Putting [something]	25
Putting [something] into [something]	26
Showing [something] (almost no hand)	27
Showing a photo of [something] to the camera	28
Showing that [something] is empty	29
Stacking [number of] [something]	30
Tearing [something]	31
Throwing [something] against [something]	32
Turning [something] upside down	33
Turning the camera downwards while filming [something]	34
Turning the camera left while filming [something]	35
Turning the camera right while filming [something]	36
Turning the camera upwards while filming [something]	37
Uncovering [something]	38
Unfolding [something]	39

Table 2: Subset of 40 selected classes used in some of the experiments.

1. Pouring [something]: 77.47%
2. Putting [something]: 8.48%

Original: Pouring [something]

1. Dropping [something]: 65.6%
2. Poking [something]: 14.38%

Original: Dropping [something]

1. Putting [something]: 97.41%
2. Picking [something] up: 0.72%

Original: Putting [something]

1. Picking [something] up: 62.07%
2. Putting [something]: 29.52%

Original: Picking [something] up

1. Showing [something]
(almost no hand): 69.64%
2. Holding [something]: 15.19%

Original: Holding [something]

1. Putting [something]: 53.69%
2. Dropping [something]: 40.28%

Original: Dropping [something]

1. Picking [something] up: 65.58%
2. Putting [something]: 11.41%

Original: Pouring [something]

1. Picking [something] up: 72.49%
2. Putting [something]: 25.07%

Original: Poking [something]

Figure 1: Some example predictions from the best performing baseline model on 10 selected classes experiment
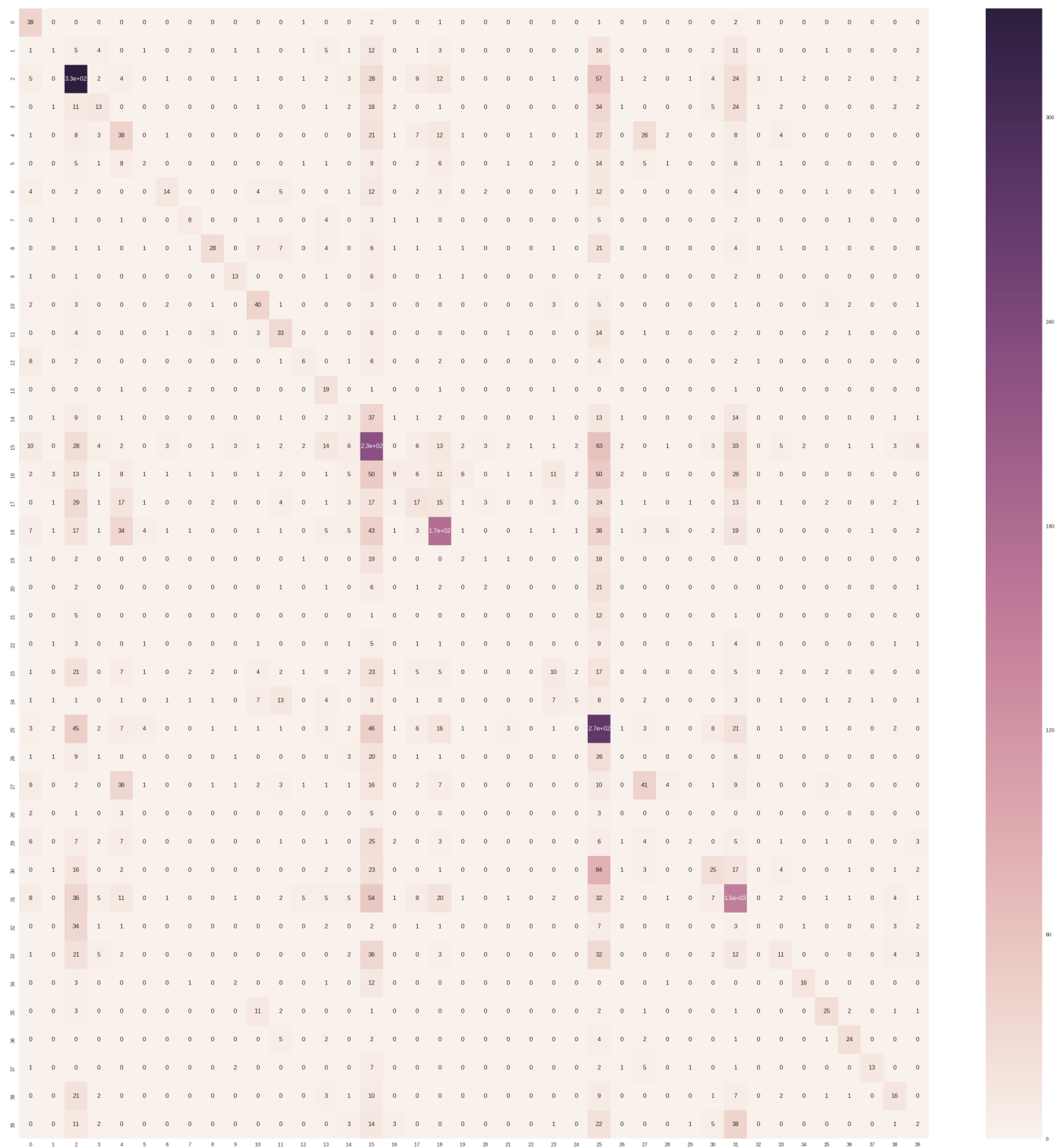
Figure 2: Confusion matrix for the best model trained on the 40 selected classes. Corresponding class-names are listed in Table 2.

Table 3: Mapping used for 10 selected classes

Actual class	Mapped class
[Something] falling like a rock [Something] falling like a feather or paper Throwing [something] Throwing [something] onto a surface Throwing [something] in the air and letting it fall	Dropping [something]
Pushing [something] from right to left Pulling [something] from right to left	Moving [something] from right to left
Pulling [something] from left to right Pushing [something] from left to right	Moving [something] from left to right
Picking [something] up Lifting [something] up completely without letting it drop down Moving [something] up Lifting [something] with [something] on it Taking [something] from [somewhere] Taking [one of many similar things on the table] Taking [something] out of [something]	Picking [something] up
Putting [something] next to [something] Putting [something] onto [something] Putting [something] on a surface Putting [something similar to other things that are already on the table] Putting [something] behind [something] Putting [something], [something] and [something] on the table Putting [something] and [something] on the table Putting [something] on a flat surface without letting it roll Putting [something] that can't roll onto a slanted surface, so it stays where it is	Putting [something]
Poking [something] so that it falls over Poking [something] so lightly that it doesn't or almost doesn't move Poking a stack of [something] so the stack collapses Poking a stack of [something] without the stack collapsing	Poking [something]
Tearing [something] into two pieces Tearing [something] just a little bit	Tearing [something]
Pouring [something] into [something] Pouring [something] onto [something] Pouring [something] out of [something] Pouring [something] into [something] until it overflows Trying to pour [something] into [something], but missing so it spills next to it	Pouring [something]
Holding [something] Holding [something] in front of [something]	Holding [something]
Showing [something] on top of [something] Showing [something] behind [something] Showing [something] next to [something]	Showing [something] (almost no hand)

Table 4: Class labels and their corresponding action-groups for all 174 classes

Class Labels	Action Groups
Trying but failing to attach [something] to [something] because it doesn't stick Attaching [something] to [something]	Attaching/Trying to attach
Bending [something] until it breaks Trying to bend [something unbendable] so nothing happens Bending [something] so that it deforms	Bending something
Digging [something] out of [something] Burying [something] in [something]	Burying or digging something
Moving away from [something] with your camera Turning the camera right while filming [something] Approaching [something] with your camera Turning the camera left while filming [something] Turning the camera upwards while filming [something] Moving [something] away from the camera Moving [something] towards the camera Turning the camera downwards while filming [something]	Camera motions
[Something] colliding with [something] and both are being deflected [Something] being deflected from [something] [Something] colliding with [something] and both come to a halt	Collisions of objects
Uncovering [something] Covering [something] with [something]	Covering
Putting [something similar to other things that are already on the table] Taking [one of many similar things on the table]	Crowd of things
Dropping [something] into [something] Dropping [something] onto [something] Dropping [something] next to [something] Dropping [something] in front of [something] Dropping [something] behind [something]	Dropping something
Showing [something] next to [something] Showing [something] on top of [something] Showing [something] behind [something]	Filming objects, without any actions
Folding [something] Unfolding [something]	Folding something
Hitting [something] with [something]	Hitting something with something
Holding [something] in front of [something] Holding [something] behind [something] Holding [something] next to [something] Holding [something] Holding [something] over [something]	Holding something
Lifting up one end of [something], then letting it drop down Lifting up one end of [something] without letting it drop down Lifting [something] up completely, then letting it drop down Lifting [something] up completely without letting it drop down	Lifting and (not) dropping something
Tilting [something] with [something] on it until it falls off Lifting [something] with [something] on it Tilting [something] with [something] on it slightly so it doesn't fall down	Lifting/Tilting objects with other objects on them
Moving [something] up Moving [something] down	Moving something

Moving [something] and [something] away from each other Moving [something] and [something] closer to each other Moving [something] closer to [something] Moving [something] away from [something]	Moving two objects relative to each other
Moving [part] of [something] Touching (without moving) [part] of [something]	Moving/Touching a part of something
Opening [something] Pretending to close [something] without actually closing it Pretending to open [something] without actually opening it Closing [something]	Opening or closing something
Picking [something] up Pretending to pick [something] up	Picking something up
Piling [something] up	Piles of stuff
Plugging [something] into [something] but pulling it right out as you remove your hand Plugging [something] into [something]	Plugging something into something
Poking [something] so it slightly moves Poking [something] so lightly that it doesn't or almost doesn't move Poking a stack of [something] without the stack collapsing Poking a hole into [something soft] Pretending to poke [something] Poking [something] so that it falls over Poking a stack of [something] so the stack collapses Poking a hole into [some substance] Poking [something] so that it spins around	Poking something
Trying to pour [something] into [something], but missing so it spills next to it Pretending to pour [something] out of [something], but [something] is empty Pouring [something] out of [something] Pouring [something] onto [something] Pouring [something] into [something] Pouring [something] into [something] until it overflows	Pouring something
Pulling [something] from behind of [something] Pulling [something] from right to left Pulling [something] out of [something] Pulling [something] onto [something] Pulling [something] from left to right	Pulling something
Pulling two ends of [something] but nothing happens Pulling two ends of [something] so that it separates into two pieces Pulling two ends of [something] so that it gets stretched	Pulling two ends of something
Pushing [something] onto [something] Pushing [something] from right to left Pushing [something] with [something] Pushing [something] so that it falls off the table Pushing [something] so that it almost falls off the table but doesn't Pushing [something] off of [something] Pushing [something] so that it slightly moves Pushing [something] from left to right	Pushing something
Pretending to put [something] on a surface Putting [something] on a surface	Putting something somewhere
Laying [something] on the table on its side, not upright Putting [something that cannot actually stand upright] upright on the table, so it falls on its side Putting [something] upright on the table	Putting something upright/on its side

Putting [something] underneath [something] Putting [something] onto [something else that cannot support it] so it falls down Failing to put [something] into [something] because [something] does not fit Putting [something], [something] and [something] on the table Pretending to put [something] behind [something] Putting [something] in front of [something] Taking [something] out of [something] Pretending to put [something] onto [something] Putting [something] and [something] on the table Pretending to take [something] out of [something] Putting [something] onto [something] Pretending to put [something] into [something] Pretending to put [something] underneath [something] Putting [something] next to [something] Putting [something] behind [something] Putting [something] on the edge of [something] so it is not supported and falls down Removing [something], revealing [something] behind Pretending to put [something] next to [something] Putting [something] into [something]	Putting/Taking objects into/out of/next to/ other objects
Letting [something] roll along a flat surface Rolling [something] on a flat surface Putting [something] that can't roll onto a slanted surface, so it slides down Lifting a surface with [something] on it until it starts sliding down Letting [something] roll down a slanted surface Lifting a surface with [something] on it but not enough for it to slide down Letting [something] roll up a slanted surface, so it rolls back down Putting [something] onto a slanted surface but it doesn't glide down Putting [something] that can't roll onto a slanted surface, so it stays where it is Putting [something] on a flat surface without letting it roll	Rolling and sliding something
Pretending to scoop [something] up with [something] Scooping [something] up with [something]	Scooping something up
Showing [something] to the camera Showing a photo of [something] to the camera	Showing objects and photos of objects
Showing that [something] is empty Showing that [something] is inside [something]	Showing that something is full/empty
Moving [something] across a surface without it falling down Moving [something] across a surface until it falls down	Something (not) falling over an edge
[Something] falling like a feather or paper [Something] falling like a rock	Something falling
Moving [something] and [something] so they collide with each other Moving [something] and [something] so they pass each other	Something passing/hitting another thing
Spilling [something] next to [something] Spilling [something] onto [something] Spilling [something] behind [something]	Spilling something
Spinning [something] so it continues spinning Spinning [something] that quickly stops spinning Pushing [something] so it spins	Spinning something
Spreading [something] onto [something] Pretending to spread 'air' onto [something]	Spreading something onto something
Pretending to sprinkle 'air' onto [something] Sprinkling [something] onto [something]	Sprinkling something onto something
Squeezing [something] Pretending to squeeze [something]	Squeezing something

Stacking [number of] [something] Putting [number of] [something] onto [something]	Stacking or placing N things
Stuffing [something] into [something]	Stuffing/Taking out
Taking [something] from [somewhere] Pretending to take [something] from [somewhere]	Taking something
Tearing [something] just a little bit Pretending to be tearing [something that is not tearable] Tearing [something] into two pieces	Tearing something
Throwing [something] against [something] Throwing [something] Throwing [something] in the air and catching it Pretending to throw [something] Throwing [something] in the air and letting it fall Throwing [something] onto a surface	Throwing something
Tipping [something] with [something in it] over, so [something in it] falls out Tipping [something] over	Tipping something over
Pretending to turn [something] upside down Turning [something] upside down	Turning something upside down
Twisting (wringing) [something] wet until water comes out Pretending or trying and failing to twist [something] Twisting [something]	Twisting something
Pretending or failing to wipe [something] off of [something] Wiping [something] off of [something]	Wiping something off of something

1.4. Data Collection Platform

We show some snapshots of our data collection platform in Figures 3 and 4. They demonstrate the platform used by the crowd-workers to select classes and to upload corresponding videos.

Submit Task

▸ Folding something (2)
▸ Stuffing/Taking out (1)
▾ Holding something (5)
   Holding [something]
   Holding [something] over [something]
   Holding [something] next to [something]
   Holding [something] in front of [something]
   Holding [something] behind [something]
▸ Crowd of things (2)
▸ Shadows (1)
▸ Collisions of objects (3)
▾ Tearing something (3)
   Tearing [something] into two pieces
   Pretending to be tearing [something that is not tearable]
   Tearing [something] just a little bit
▸ Lifting/Tilting objects with other objects on them (3)
▾ Moving two objects relative to each other (4)
   Moving [something] closer to [something]
   Moving [something] away from [something]
   Moving [something] and [something] closer to each other (fix the camera and use both hands to move both objects)
   Moving [something] and [something] away from each other (fix the camera and use both hands to move both objects)
▸ Attaching/Trying to attach (2)
▸ Spinning something (3)
▸ Something falling (2)
▸ Putting/Taking objects into/out of/next to/… other objects (19)
▾ Rolling and sliding something (10)
   Letting [something] roll down a slanted surface
   Letting [something] roll up a slanted surface, so it rolls back down
   Letting [something] roll along a flat surface
   Putting [something] on a flat surface without letting it roll
   Putting [something] that can't roll onto a slanted surface, so it stays where it is
   Putting [something] that can't roll onto a slanted surface, so it slides down
   Lifting a surface with [something] on it until it starts sliding down
   Lifting a surface with [something] on it but not enough for it to slide down
   Putting [something] onto a slanted surface but it doesn't glide down
   Rolling [something] on a flat surface
▸ Plugging something into something (2)

holding something over something

Upload Video

moving something away from something

Upload Video

lifting a surface with something on it but not enough for it to slide down

Upload Video

moving something away from something

Upload Video

tearing something into two pieces

Upload Video

Figure 3: Left Column: Crowd-workers can choose the classes they want to generate videos for (often the available classes are controlled so as to maintain class balance as much as possible). Right Column: An interface to upload videos and enter input-text.

Submit Task

▸ Folding something (2)
▸ Stuffing/Taking out (1)
▸ Holding something (5)
▸ Crowd of things (2)
▸ Shadows (1)
▸ Collisions of objects (3)
▸ Tearing something (3)
▸ Lifting/Tilting objects with other objects on them (3)
▸ Moving two objects relative to each other (4)
▸ Attaching/Trying to attach (2)
▸ Spinning something (3)
▸ Something falling (2)
▸ Putting/Taking objects into/out of/next to/... other objects (19)
▸ Rolling and sliding something (10)
▸ Plugging something into something (2)
▸ Twisting something (3)
▸ Opening or closing something (4)
▸ Pushing something (9)
▸ Tipping something over (2)
▸ Filming objects, without any actions (3)
▸ Spilling something (3)
▸ Turning something upside down (2)
▸ Putting something somewhere (2)
▸ Picking something up (2)
▸ Hitting something with something (1)
▸ Dropping something (5)
▸ Poking something (9)
▸ Throwing something (6)
▸ Wiping something off of something (2)
▸ Camera motions (8)
▸ Showing objects and photos of objects (2)
▸ Squeezing something (2)
▸ Bending something (3)

letting pen roll down a slanted surface

rolling pen on a flat surface

pushing post it so that it slightly moves

pushing pen onto folder
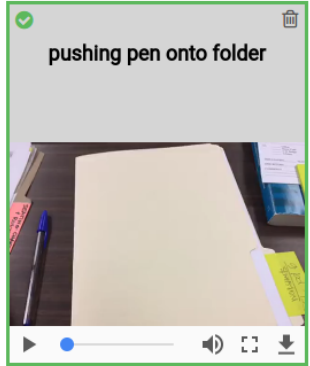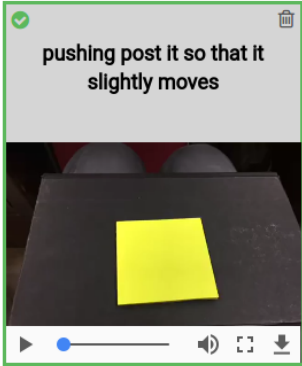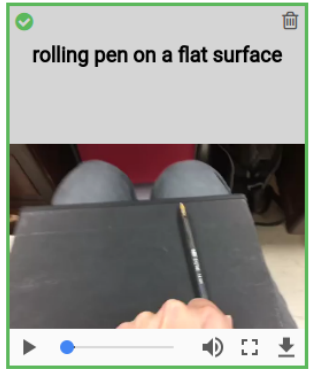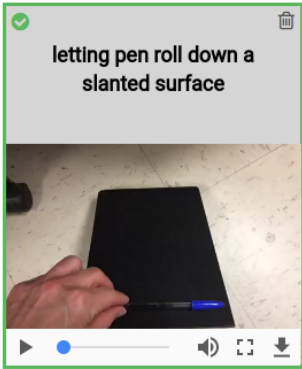
pushing pen off of folder

pushing post it with pen
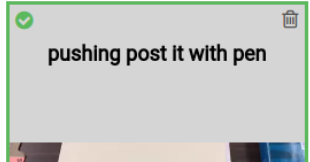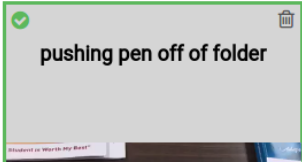
Figure 4: An example of the upload interface after uploading videos and entering input-text.