

# SUPPLEMENTARY MATERIAL - SWISS: FLEXIBLE, UNIFIED WEAKLY-SUPERVISED INSTANCE SEGMENTATION UTILIZING FOUNDATION MODELS

*Christian Wilms, Maris Hillemann, Ehsan Yaghoubi, and Simone Frinotrop*

Computer Vision Group, University of Hamburg

## 1. PER-CLASS RESULTS ON COCO

Table 1 shows the per-class results of  $\text{SWISS}_{\text{Mask R-CNN}}$  with box-, point-, and image-level supervision compared to the results of fully-supervised Mask R-CNN on the COCO *test-dev* split. All results are given in terms of AP.

## 2. DETAILS ON ANNOTATION SCHEMES

In the following, for each novel annotation scheme A, B, and C on the COCO dataset introduced in Sec. 4.5 of the main paper, the classes per supervision type (mask, box, or point) are given.

### Annotation Scheme A

**Mask supervision** Person, Bicycle, Train, Boat, Traffic light, Hydrant, Parking meter, Bench, Bird, Horse, Cow, Bear, Tie, Frisbee, Skis, Snowboard, Sports ball, Kite, Baseball bat, Baseball glove, Skateboard, Surfboard, Tennis racket, Bottle, Cup, Bowl, Banana, Sandwich, Orange, Broccoli, Carrot, Hot dog, Pizza, Donut, Cake, Chair, Couch, Potted plant, Bed, Dining table, Toilet, TV, Remote, Cell phone, Microwave, Oven, Toaster, Sink, Refrigerator, Clock, Vase, Scissors

**Box supervision** Car, Motorcycle, Airplane, Bus, Truck, Stop sign, Cat, Dog, Sheep, Elephant, Zebra, Giraffe, Backpack, Umbrella, Handbag, Suitcase, Wine glass, Fork, Knife, Spoon, Apple, Laptop, Mouse, Keyboard, Book, Teddy bear, Hair drier, Toothbrush

### Annotation Scheme B

**Mask supervision** Bicycle, Traffic light, Bench, Bird, Tie, Skis, Snowboard, Sports ball, Kite, Baseball bat, Bowl, Banana,

Sandwich, Broccoli, Hot dog, Cake, Couch, Potted plant, Bed, Dining table, Oven, Toaster, Scissors

**Box supervision** Person, Car, Motorcycle, Airplane, Bus, Train, Truck, Boat, Hydrant, Stop sign, Parking meter, Cat, Dog, Horse, Sheep, Cow, Elephant, Bear, Zebra, Giraffe, Backpack, Umbrella, Handbag, Suitcase, Frisbee, Baseball glove, Skateboard, Surfboard, Tennis racket, Bottle, Wine glass, Cup, Fork, Knife, Spoon, Apple, Orange, Carrot, Pizza, Donut, Chair, Toilet, TV, Laptop, Mouse, Remote, Keyboard, Cell phone, Microwave, Sink, Refrigerator, Book, Clock, Vase, Teddy bear, Hair drier, Toothbrush

### Annotation Scheme C

**Box supervision** Person, Bicycle, Bench, Bowl, Hot dog, Pizza, Chair, Couch, Potted plant, TV, Laptop, Bed, Dining table, Scissors

**Point supervision** Car, Motorcycle, Airplane, Bus, Train, Truck, Boat, Traffic light, Hydrant, Stop sign, Parking meter, Bird, Cat, Dog, Horse, Sheep, Cow, Elephant, Bear, Zebra, Giraffe, Backpack, Umbrella, Handbag, Tie, Suitcase, Frisbee, Skis, Snowboard, Sports ball, Kite, Baseball glove, Baseball bat, Skateboard, Surfboard, Tennis racket, Bottle, Wine glass, Cup, Fork, Knife, Spoon, Banana, Apple, Sandwich, Orange, Broccoli, Carrot, Donut, Cake, Toilet, Mouse, Remote, Keyboard, Cell phone, Microwave, Oven, Toaster, Sink, Refrigerator, Book, Clock, Vase, Teddy bear, Hair drier, Toothbrush

**Table 1.** Per-class AP results of Mask R-CNN and SWISS<sub>Mask R-CNN</sub> on the COCO *test-dev* split. The supervision types are full supervision with mask ( $\mathcal{M}$ ) and weak supervision with boxes ( $\mathcal{B}$ ), points ( $\mathcal{P}$ ), or image-level ( $\mathcal{I}$ ) annotations.

Class	Mask R-CNN	SWISS <sub>Mask R-CNN</sub>			Class	Mask R-CNN	SWISS <sub>Mask R-CNN</sub>		
	$\mathcal{M}$	$\mathcal{B}$	$\mathcal{P}$	$\mathcal{I}$		$\mathcal{M}$	$\mathcal{B}$	$\mathcal{P}$	$\mathcal{I}$
Person	45.3	43.8	14.3	12.7	Wine glass	30.2	30.8	25.3	17.7
Bicycle	17.0	13.7	8.2	4.6	Cup	41.2	40.1	35.1	29.0
Car	37.6	37.3	32.9	19.7	Fork	11.9	13.1	11.3	8.8
Motorcycle	30.4	30.6	22.6	15.4	Knife	11.4	11.9	8.3	5.6
Airplane	43.2	43.0	38.9	32.6	Spoon	10.1	10.0	8.6	5.6
Bus	65.4	65.7	49.3	47.8	Bowl	36.5	22.2	14.4	17.6
Train	61.7	60.0	52.2	50.4	Banana	20.8	19.1	10.8	7.8
Truck	30.6	31.7	24.8	19.7	Apple	24.7	24.6	23.2	19.1
Boat	23.3	22.8	19.8	13.3	Sandwich	32.2	29.7	9.0	7.5
Traffic light	30.4	28.8	21.9	16.2	Orange	28.9	28.4	25.4	19.5
Hydrant	59.0	57.1	48.5	46.6	Broccoli	25.5	24.0	18.8	14.6
Stop sign	68.6	68.0	60.4	46.9	Carrot	16.5	15.7	15.2	11.3
Parking meter	43.9	43.0	26.1	24.5	Hot dog	24.3	19.1	10.1	4.2
Bench	16.4	15.4	10.4	9.4	Pizza	55.9	54.8	44.7	49.8
Bird	30.1	28.5	27.0	19.5	Donut	50.0	48.3	33.8	19.6
Cat	60.0	60.2	58.0	55.6	Cake	30.4	28.5	18.9	16.6
Dog	51.1	51.6	47.4	45.7	Chair	17.9	17.6	13.9	7.8
Horse	40.8	38.8	38.3	30.6	Couch	30.3	24.1	4.6	3.0
Sheep	44.4	44.4	43.0	29.2	Potted plant	20.3	18.7	9.1	6.8
Cow	43.9	42.7	41.4	28.9	Bed	34.8	18.2	9.6	6.8
Elephant	59.7	58.7	56.9	43.8	Dining table	16.7	3.5	3.6	1.9
Bear	68.1	65.0	62.6	59.2	Toilet	58.0	56.0	21.2	21.2
Zebra	51.9	50.7	50.5	40.5	TV	55.3	54.3	28.4	40.8
Giraffe	49.8	49.8	48.4	44.3	Laptop	57.3	58.7	27.8	44.0
Backpack	18.9	19.0	16.7	12.4	Mouse	54.6	54.4	51.2	40.6
Umbrella	43.4	43.6	40.6	33.9	Remote	27.8	26.9	23.8	18.5
Handbag	13.4	14.5	12.3	7.3	Keyboard	49.1	50.5	36.7	30.1
Tie	32.6	30.5	26.3	17.4	Cell phone	30.1	29.3	20.0	19.7
Suitcase	34.3	34.5	31.7	22.1	Microwave	56.2	55.6	14.3	14.1
Frisbee	53.1	51.0	47.8	46.9	Oven	33.5	30.2	7.7	8.8
Skis	2.4	1.7	0.1	0.1	Toaster	21.2	15.7	16.0	7.0
Snowboard	21.7	19.9	15.6	7.8	Sink	34.2	33.6	26.3	21.7
Sports ball	49.7	45.7	42.9	35.7	Refrigerator	47.3	45.7	26.2	20.9
Kite	33.9	30.9	27.0	14.8	Book	8.9	9.4	6.9	4.0
Baseball bat	26.5	24.5	23.2	19.2	Clock	49.6	48.5	42.6	38.4
Baseball glove	38.3	36.8	35.4	24.1	Vase	39.6	38.7	34.6	15.5
Skateboard	25.8	25.5	18.7	9.6	Scissors	16.5	14.0	5.6	3.5
Surfboard	28.8	28.4	22.5	20.7	Teddy bear	37.2	37.7	23.9	19.2
Tennis racket	53.1	51.9	45.9	41.6	Hair drier	2.3	2.3	0.6	0.0
Bottle	36.1	35.2	25.5	21.2	Toothbrush	9.9	10.6	7.1	4.3