

Visual Sentences



Single images, e.g. LAION

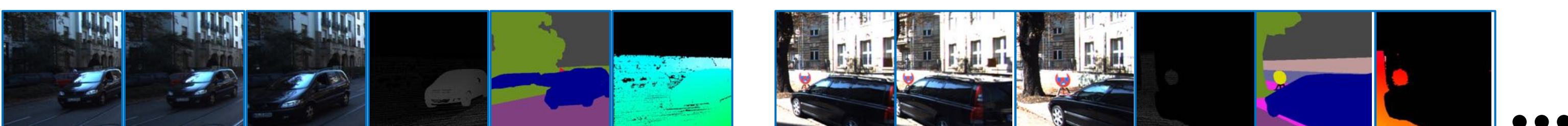
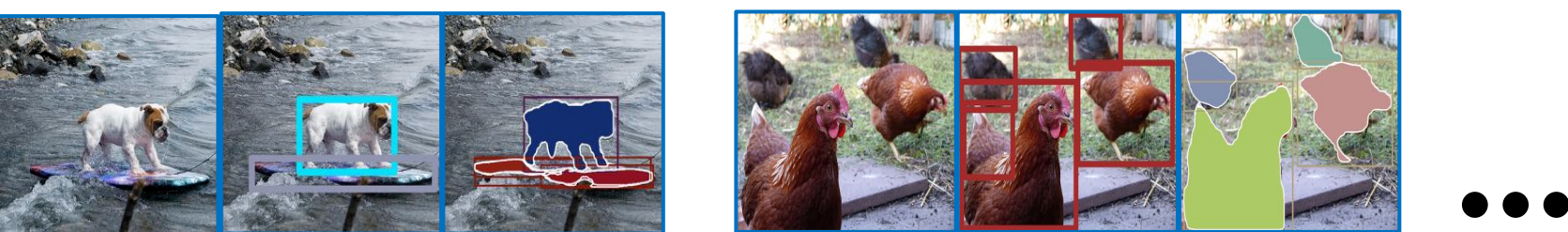
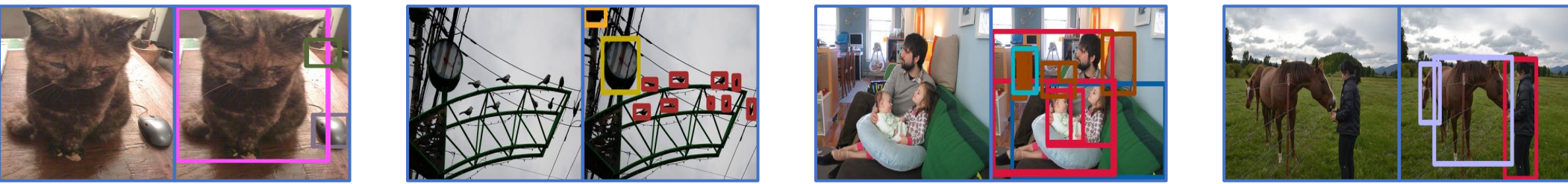
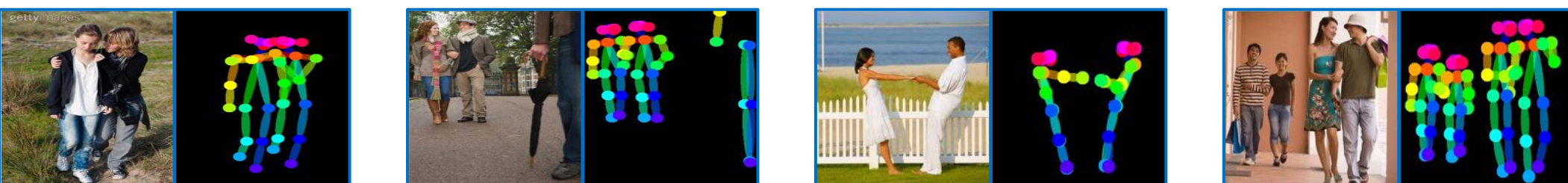


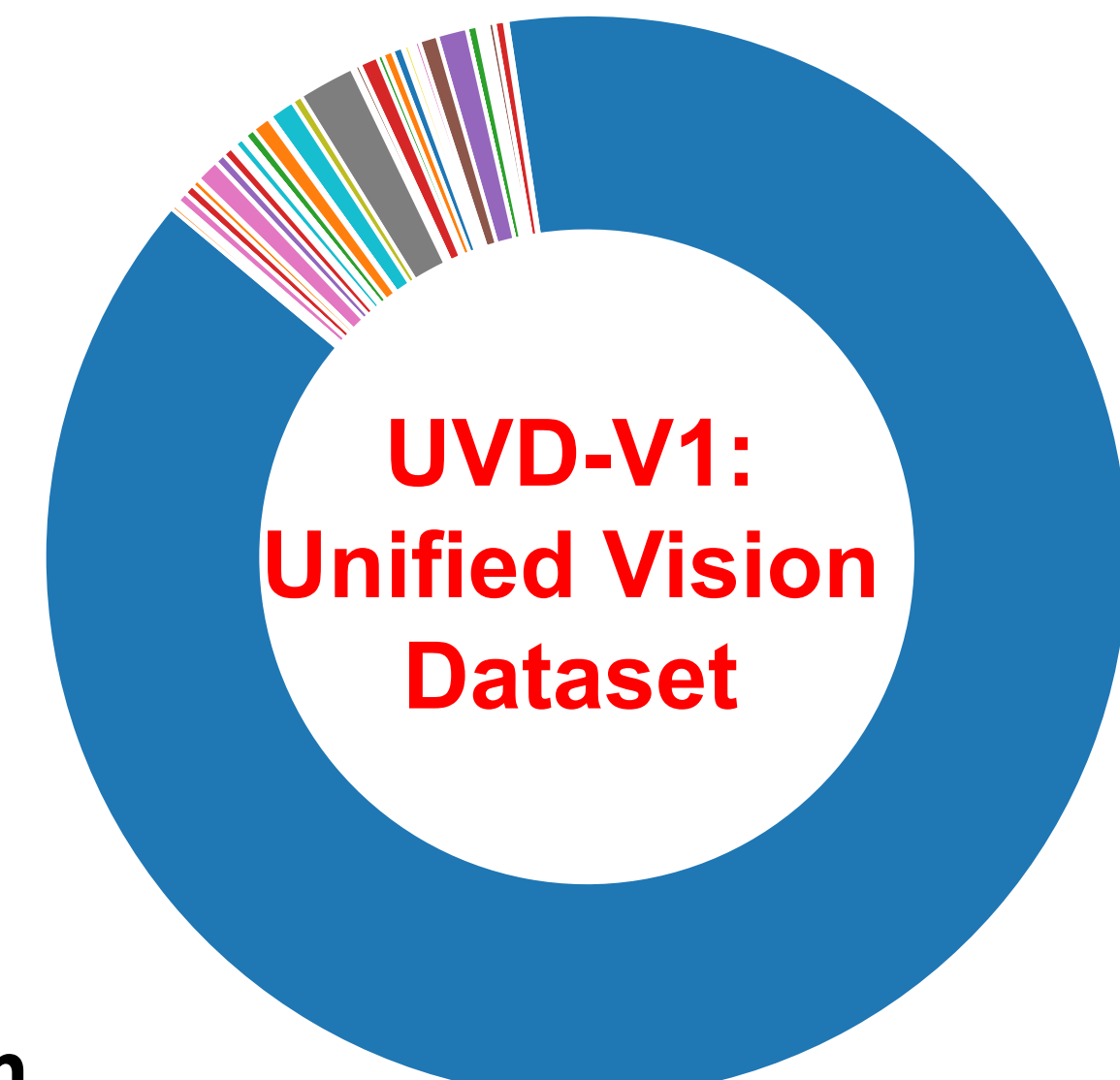
Image sequences,
e.g. videos, 3D
rotations, synthetic
viewpoints

Images with annotation,
e.g. style transfer, object
detection, low light
enhancement

Images with free form
annotation,
e.g. object detection +
instance segmentation etc

Videos with
annotation,
e.g. video
segmentation

Dataset Names	
laion (380.0000B tokens)	coco_generated_edge (0.3028B tokens)
multisports (0.0784B tokens)	kinetics700_s8 (2.3640B tokens)
denoise (0.2458B tokens)	i1k_edge (1.3211B tokens)
i1k_seg (1.3211B tokens)	kitti (0.0092B tokens)
mpii (0.0639B tokens)	ade20k (0.0517B tokens)
coco_generated_mixed (0.7570B tokens)	ava_detection (0.1229B tokens)
inpainting_coco (0.3028B tokens)	coco_generated_normal (0.3028B tokens)
cityscape (0.0152B tokens)	hand14k (0.0020B tokens)
coco_pose (0.3834B tokens)	DID_MDN_heavy (0.0081B tokens)
ucf101 (0.1091B tokens)	ego4d (1.1521B tokens)
charades_v1 (0.2415B tokens)	ava (0.1180B tokens)
DID_MDN_light (0.0081B tokens)	light_enhance (0.0005B tokens)
charades_ego (0.1931B tokens)	mpii_back_bg (0.0639B tokens)
diving48 (0.1507B tokens)	activity_net (0.3806B tokens)
i1k_category_edge (1.3195B tokens)	youtube_vos_annotation (0.0711B tokens)
jhmdb_optical_flow (0.0190B tokens)	i1k_depth (1.3211B tokens)
kinetics (3.8436B tokens)	i1k_colorization (1.3211B tokens)
kinetics700_s12 (2.3640B tokens)	msr_vtt (0.0573B tokens)
i1k_category_2s (0.6587B tokens)	moments_in_time (2.9790B tokens)
objaverse (0.1741B tokens)	hmd51 (0.0554B tokens)
coco_pose_generated (0.2123B tokens)	jhmdb_pose (0.0190B tokens)
coco_cot (0.3632B tokens)	you_cook (0.0031B tokens)
i1k_category_1s (0.6568B tokens)	3d_pose (0.0438B tokens)
coco_pose_generated_back_bg (0.2123B tokens)	lvmp (0.0394B tokens)
i1k_category_depth (1.3195B tokens)	derain (0.0351B tokens)
i1k_inpainting (1.3211B tokens)	sthv2 (0.9046B tokens)
coco_generated_seg_coco (0.7570B tokens)	lvmp_back_bg (0.0394B tokens)
kinetics700_s24 (2.3641B tokens)	i1k_category_normal (1.3195B tokens)
colorization (0.0768B tokens)	jhmdb_black_pose (0.0190B tokens)
i1k_category_4s (0.6598B tokens)	davis (0.0004B tokens)
instruct_pix2pix (0.4155B tokens)	i1k_normal (1.3211B tokens)
inpainting_ik (7.2689B tokens)	youtube_vos_clips (0.0637B tokens)
i1k_category_seg (1.3195B tokens)	coco_generated_depth (0.3028B tokens)
i1k_cot (3.3027B tokens)	vip_seg (0.0645B tokens)
mpii_cot (0.0500B tokens)	co3d_seq (0.2288B tokens)
DID_MDN_medium (0.0081B tokens)	jester (0.6065B tokens)



**UVD-V1:
Unified Vision
Dataset**

**420B tokens,
50 Datasets.**