Common Visual Data Foundation:

Enabling community-driven Research in Computer Vision

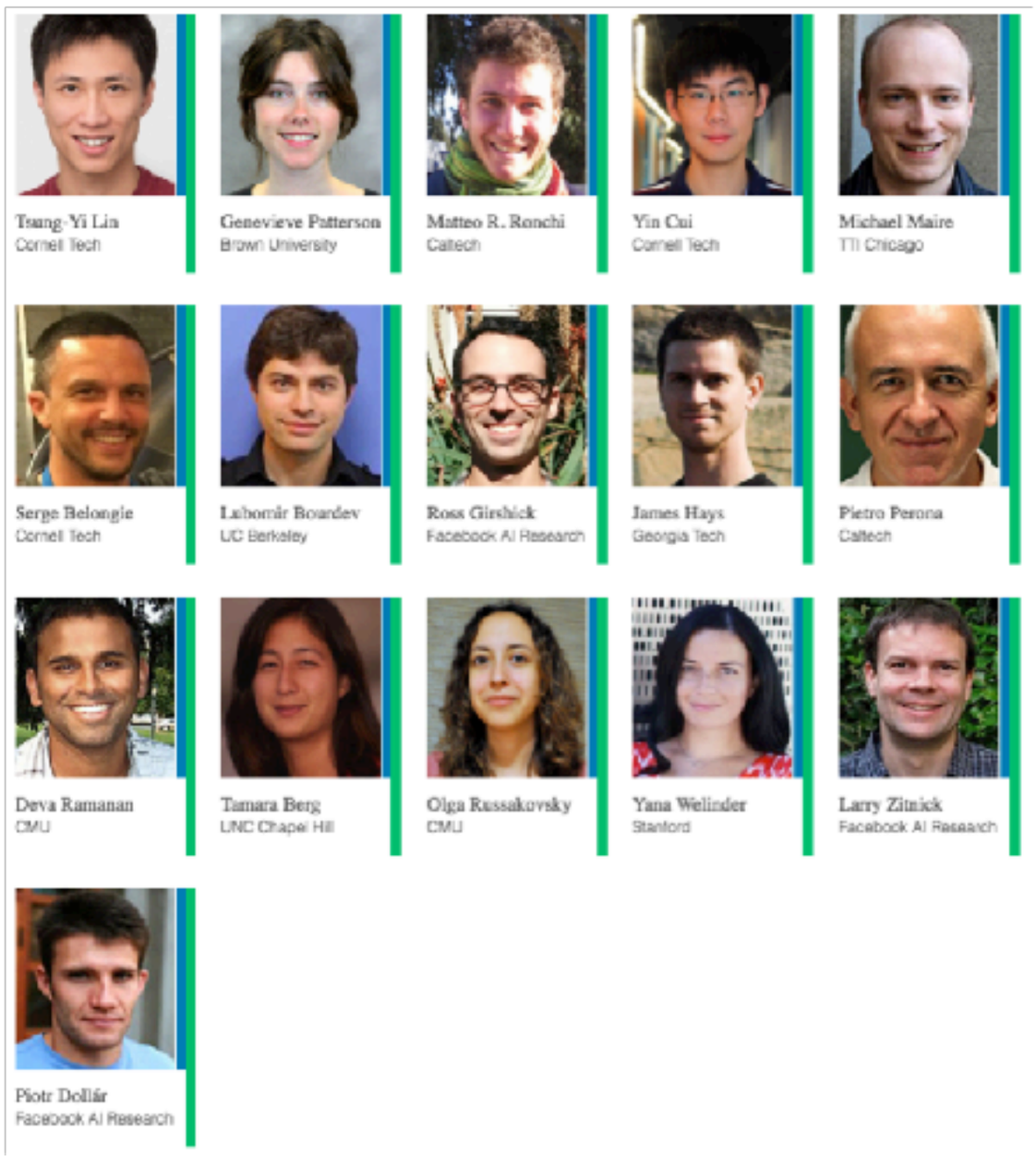
Tsung-Yi Lin
Cornell Tech, Cornell University





CORNELL
TECH

Common Visual Data Foundation (CVDF)



- President: Serge Belongie
- Vice President: Piotr Dollar
- Secretary: Tsung-Yi Lin
- Treasury: James Hays

Why CVDF?

- Enable community-driven research through **sharing data**
- **Why sharing data**?
  - new data enables new research direction
  - the progress of research community can be measured with the same metric







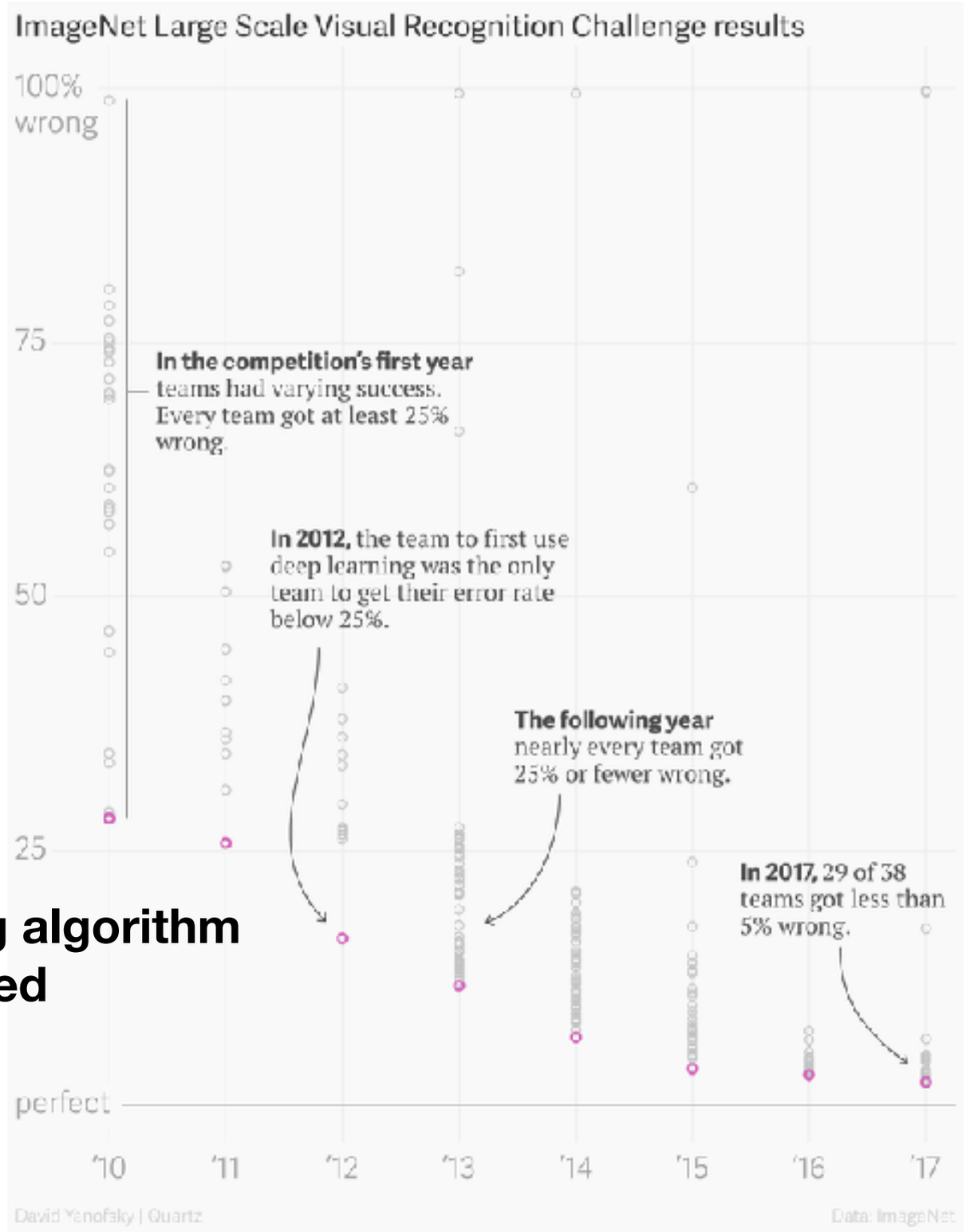






	AP	AP^{50}	AP^{75}	AP^{S}	AP^{M}	AP^{L}	AR^{1}	AR^{10}	AR^{100}	AR^{S}	AR^{M}	AR^{L}	date
G RMI[13]	0.413	0.62	0.45	0.231	0.436	0.547	0.34	0.548	0.604	0.424	0.641	0.748	2016-09-18
umd_det[25]	0.403	0.62	0.445	0.224	0.428	0.528	0.324	0.531	0.569	0.36	0.608	0.734	2017-06-10
MSRA[18]	0.394	0.613	0.424	0.216	0.431	0.521	0.322	0.483	0.502	0.314	0.565	0.65	2016-11-24
Deformable R-FCN[9]	0.373	0.58	0.407	0.185	0.402	0.518	0.317	0.473	0.483	0.277	0.541	0.673	2017-03-25
MSRA_2015[19]	0.371	0.588	0.393	0.173	0.415	0.525	0.319	0.475	0.489	0.267	0.552	0.679	2015-11-26
Trimps Soushen[22]	0.359	0.58	0.383	0.158	0.407	0.509	0.312	0.479	0.497	0.269	0.557	0.683	2016-09-16
FPN (single model)[11]	0.358	0.585	0.385	0.175	0.387	0.478	0.309	0.463	0.479	0.264	0.524	0.63	2016-11-04
Imagine Lab[15]	0.353	0.579	0.39	0.147	0.376	0.519	0.316	0.489	0.509	0.273	0.563	0.705	2016-12-24
Wal[27]	0.345	0.559	0.369	0.165	0.389	0.478	0.303	0.47	0.493	0.266	0.565	0.668	2016-09-19
FAIRCNN[16]	0.332	0.519	0.363	0.136	0.372	0.478	0.299	0.46	0.483	0.234	0.56	0.664	2015-11-26
CMU_A2_VGG16[4]	0.323	0.528	0.341	0.144	0.355	0.452	0.295	0.462	0.471	0.246	0.522	0.65	2016-09-19

IMAGENET



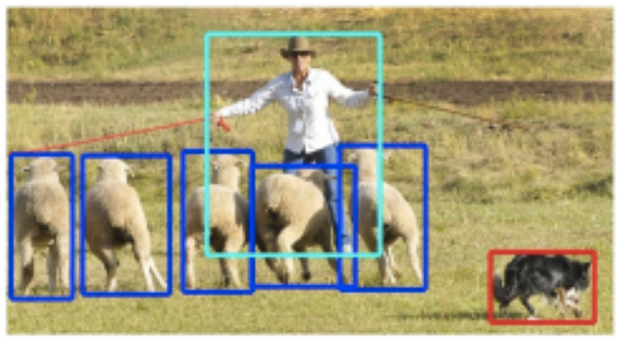
First deep learning algorithm was applied

COCO
Common Objects in Context

Image Classification



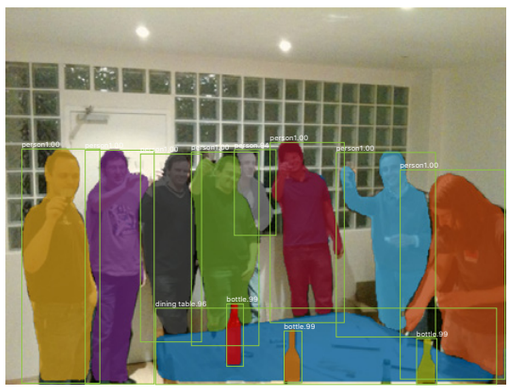
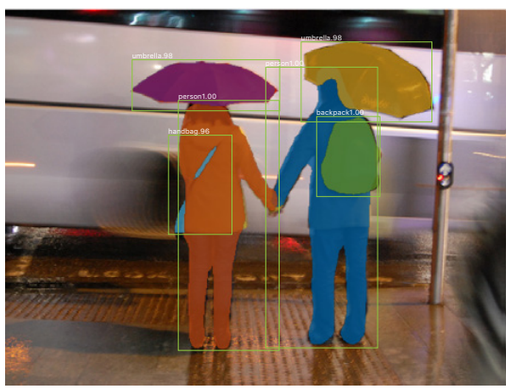
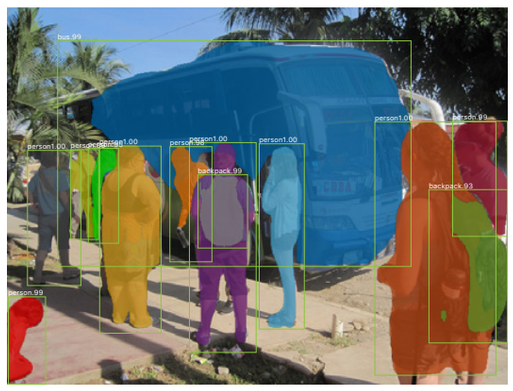
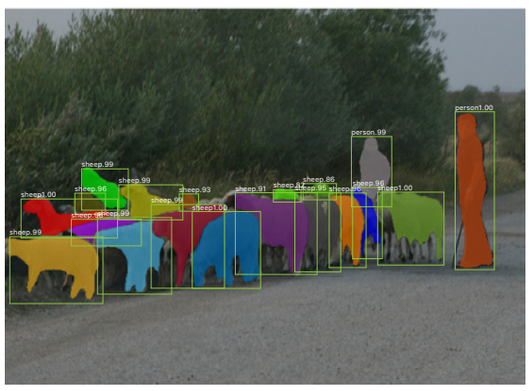
Box Detection

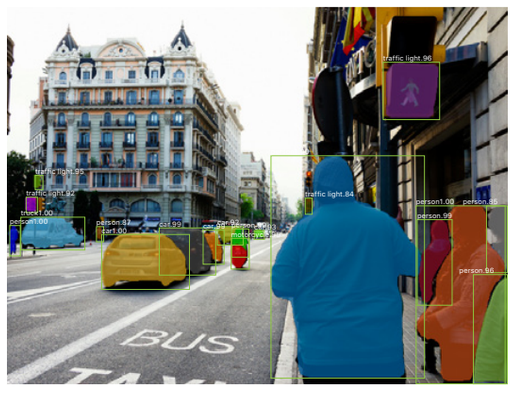


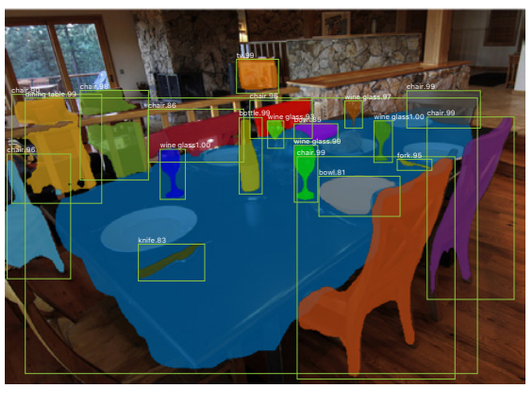
Instance Segmentation



Mask R-CNN (SOTA Algorithm)



Aren’t We Just Doing Fine?

- What we are doing well
  - the community is very open to share data when publishing!

- What are we **not** doing so well?

Academia

- Funding
- Long term viability

Industry

- Image copyright and license
- Take-down policy

What Are The Pain Points?

- **Where** to store pixels?
- What happens **after students graduate**?
- What if I need **funding** to expand my dataset?
- Could I get sued because of image **copyright** infringement?
- How should I design **take-down policy**?

Funding Model

- 501(c)(3) non-profit organization
- Funding from multiple sources

facebook Microsoft



Case Study

- Common Objects in Context (COCO) Dataset
- COCO-stuff (University of Edinburgh)
- Open Images Dataset (Google)



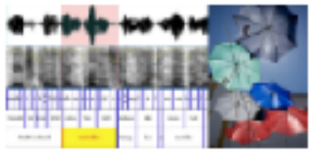
COCO



- Goal: build the next object detection dataset
- Results:
  - annual competition
  - Interdisciplinary dataset

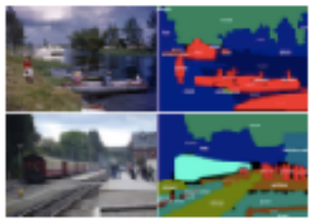
External Annotations on COCO and Related Datasets

Please contact us to add your dataset here! Do not release annotations on test set images under any circumstances to keep the integrity of the COCO challenges intact (contact us with any questions).



SPEECH-COCO

SPEECH-COCO augments the COCO dataset with speech captions generated using TTS synthesis. The corpus contains more than 600,000 spoken captions, allowing researches of language acquisition, unsupervised term discovery, keyword spotting, or semantic embedding using speech and vision.



COCO-Stuff

COCO-Stuff augments the COCO dataset with pixel-level stuff annotations for 10,000 images. The 91 stuff classes are carefully selected to have a similar level of granularity to the thing classes in COCO, allowing the study of stuff and things in context.



VISUAL GENOME

Visual Genome is a dataset, a knowledge base, an ongoing effort to connect structured image concepts to language.



RefCOCO

RefCOCO dataset was collected using Refer-it Game. Each expression aims to unambiguously indicate a particular person or object in an image.



COCO Attributes

COCO Attributes has over 3.5M attribute annotations for People, Animals, and Objects from the COCO training dataset.



G-Ref

Google referring expression dataset (G Ref) is a dataset focuses on unambiguous object text descriptions (i.e. referring expressions) that allow one to uniquely identify a single object or region within an image.



VerSe

VerSe annotates COCO images with OntoNotes senses for 90 verbs (actions) which have ambiguous visual cues. Along with the sense information we provide visualness labels for OntoNotes senses of 150 visual verbs.



COCO-Text

COCO Text is for both text detection and recognition. The dataset annotates scene text with transcriptions along with attributes such as legibility, printed or handwritten text.



FM-IQA

The Freestyle Multilingual Image Question Answering (FM-IQA) dataset contains over 120,000 images and 250,000 freestyle Chinese question-answer pairs and their English translations.



VQA

VQA is a new dataset containing open-ended questions about images. These questions require an understanding of vision, language and commonsense knowledge to answer.



VISUAL MADLIBS

Visual Madlibs is a new dataset consisting of focused natural language descriptions collected using automatically produced fill-in-the-blank templates. This dataset can be used for targeted generation or multiple-choice question-answering.



COCO-a

COCO-a annotates human actions and interactions with objects (or other people) with 140 visual actions (verbs with an unambiguous visual connotation), along with information such as emotional state and relative distance and position with the object.



SALICON

The SALICON dataset offers a large set of saliency annotations on the COCO dataset. This data complements the task-specific annotations to advance the ultimate goal of visual understanding.

COCO

- How do we achieve that?
  - **data sharing**
  - **challenging data** with real world applications
  - **generic annotation format** to be adopted by other research groups
  - committed group members






```
{
   "info"            : info,
   "images"          : [image],
   "annotations"     : [annotation],
   "licenses"        : [license],
}

info{
   "year"            : int,
   "version"         : str,
   "description"     : str,
   "contributor"     : str,
   "url"             : str,
   "date_created"    : datetime,
}

image{
   "id"              : int,
   "width"           : int,
   "height"          : int,
   "file_name"       : str,
   "license"         : int,
   "flickr_url"      : str,
   "coco_url"        : str,
   "date_captured"   : datetime,
}

license{
   "id"              : int,
   "name"            : str,
   "url"             : str,
}
```

```
{
  "info" : info,
  "images" : [image],
  "categories" : [category],
  "annotations" : [annotation],
  "licenses" : [license]
}

info{
  "year" : int,
  "version" : str,
  "description" : str,
  "contributor" : str,
  "url" : str,
  "date_created" : datetime,
}

image{
  "id" : int,
  "width" : int,
  "height" : int,
  "file_name" : str,
  "license" : int,
  "rights_holder" : str
}

category{
  "id" : int,
  "name" : str,
  "supercategory" : str,
}

annotation{
  "id" : int,
  "image_id" : int,
  "category_id" : int
}

license{
  "id" : int,
  "name" : str,
  "url" : str
}
```

COCO-stuff



- Initiated by the team at University of Edinburgh
- 10k images are annotated when first published

COCO-stuff

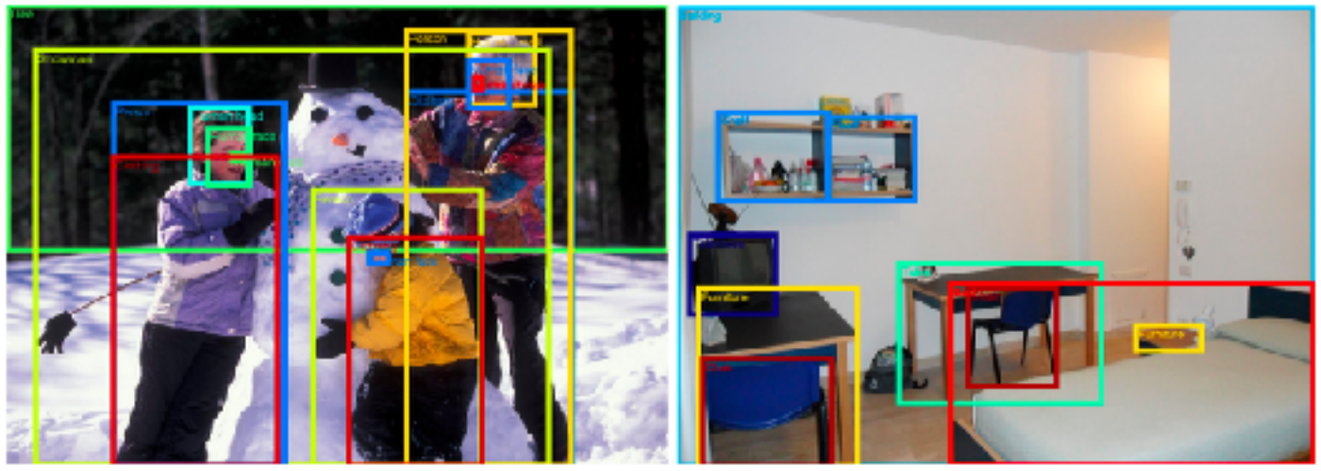
- + MightyAI





- COCO-stuff segmentation for COCO challenge this year
- $50k funding to collect annotations for 200k images

Open Images Dataset



- Large scale dataset: 8M images, only urls were released
- CVDF helps hosting the actual pixels
- CVDF is responsible for copyright issue of hosted images

How We Would Like to Work With You

- Hosting pixels/data
- Providing funding
- Advice on data collection and annotations
- Advice on organizing competitions

4D Light Field Dataset

Universität Konstanz

Heidelberg Collaboratory for Image Processing

Algorithm	Meta		Stratified				Test			
	MEDIAN	AVG	Backgammon	Dots	Pyramids	Stripes	Bedroom	Bicycle	Herbs	Origami
OBER-cross+ANP	3.077	3	6.548	1	3.403	2	0.574	1	0.364	8	1.055	2	3.025	2	6.601	3	2.121	1	8.565	1
EPO-MC	3.704	7	5.701	2	3.450	3	2.701	9	0.050	1	4.118	3	3.220	1	10.051	9	2.265	4	9.411	
OBER-cross	6.715	5	10.289	7	9.504	4	32.384	18	0.301	5	31.717	23	5.234	7	11.797	5	7.146	2	8.796	
PS_RF	7.539	4	8.578	4	7.142	15	7.975	5	0.107	2	2.964	1	8.013	9	17.179	13	10.480	10	13.573	1
RM3DE	7.592	5	10.216	6	7.510	12	31.735	16	0.682	16	4.132	4	8.099	6	16.089	9	10.144	9	13.559	
EPN+CS+GC	8.564	6	10.724	8	3.328	1	39.248	20	0.242	3	18.545	15	7.543	12	11.599	5	9.190	6	10.749	
SPO	8.779	7	8.461	3	8.781	7	16.274	12	0.861	19	14.987	10	4.864	4	10.907	4	8.260	3	11.695	
OBER	9.031	8	11.677	11	8.389	9	46.563	26	0.245	4	12.914	7	5.925	10	17.134	12	9.991	8	13.124	
RPR+	9.852	9	10.020	5	8.737	6	11.014	7	0.876	20	17.248	11	8.769	18	12.198	7	8.672	5	13.904	1
SIL_GC	10.705	10	14.259	18	4.791	13	40.513	22	1.203	22	25.207	25	5.889	5	12.207	8	12.610	12	10.672	
GC1_CV	10.750	11	12.156	13	7.897	16	12.671	14	0.505	14	26.659	22	9.163	19	18.158	16	13.048	13	10.014	
RM3_Swow	10.950	12	12.257	16	3.525	5	11.014	7	0.449	10	18.950	17	10.650	20	18.016	3	9.624	7	10.956	1
OFSY_330/DNR	11.329	13	12.066	12	6.928	12	37.670	19	0.336	7	18.640	15	4.749	3	16.184	10	12.304	11	15.685	1
EpiNet-stra+MF	12.940	14	10.029	9	12.603	20	12.799	10	0.435	11	9.465	5	7.403	11	17.312	14	13.511	18	14.985	1
EpiNet-stra	12.735	15	11.670	10	12.980	21	17.384	13	0.319	6	12.534	6	7.548	13	18.196	15	13.821	19	14.879	1
BSL_I	13.939	16	12.226	14	4.734	11	26.557	15	0.491	9	13.059	8	8.441	16	19.330	18	14.451	16	13.027	1
FBS*	13.980	17	13.291	17	10.307	19	12.833	11	0.649	13	19.117	18	5.773	8	19.650	21	14.764	17	22.864	3
BSL	13.409	18	12.748	16	6.560	10	32.407	17	0.480	12	13.094	9	8.496	17	19.410	19	14.237	14	13.724	1
OMG_bce	14.605	19	15.403	20	14.962	23	43.268	23	1.441	23	18.026	13	7.833	14	16.724	11	14.259	15	15.894	3
ZGTV	15.151	20	16.329	22	3.981	0	41.492	21	0.610	15	34.502	24	12.502	22	21.120	23	16.560	20	15.305	1
*FI	16.146	21	16.193	21	5.575	14	2.500	3	17.054	28	35.247	26	13.865	26	19.247	22	18.010	22	16.103	1
*MV	17.235	22	18.405	24	11.574	18	45.502	24	0.776	17	46.192	27	8.161	15	19.590	20	18.742	23	19.052	3
LF_OCC	17.016	23	15.074	19	13.502	22	9.595	6	1.450	24	18.734	16	10.344	20	18.965	17	17.729	21	15.729	1
*LF_OCC	18.451	24	17.579	23	19.005	25	5.822	4	3.172	26	18.408	14	17.503	27	21.502	24	20.850	26	22.431	2
MVCMO	21.952	25	21.050	25	15.190	24	12.047	9	8.515	27	26.316	21	12.920	23	22.821	25	21.157	28	25.680	2
*EPI1	22.891	26	24.324	27	21.331	26	62.001	28	0.838	18	25.809	20	13.588	25	25.213	26	47.062	27	28.901	2
*EPI2	22.942	27	22.651	26	22.075	27	46.534	25	1.083	21	23.508	19	13.557	24	25.585	27	30.944	24	27.121	2
EpiNet-cross-fast	31.136	28	26.511	28	31.960	28	42.905	22	1.941	26	45.076	27	13.472	21	30.287	28	28.076	25	33.161	3

