

Places2 Challenge

Held in conjunction with ILSVRC and COCO at ECCV 2016

Scene Parsing Challenge 2016

Challenges for Deep Scene Understanding

Bolei Zhou
MIT



Bolei
Zhou



Hang
Zhao



Xavier
Puig



Sanja
Fidler
(UToronto)



Adela
Barriuso



Aditya
Khosla

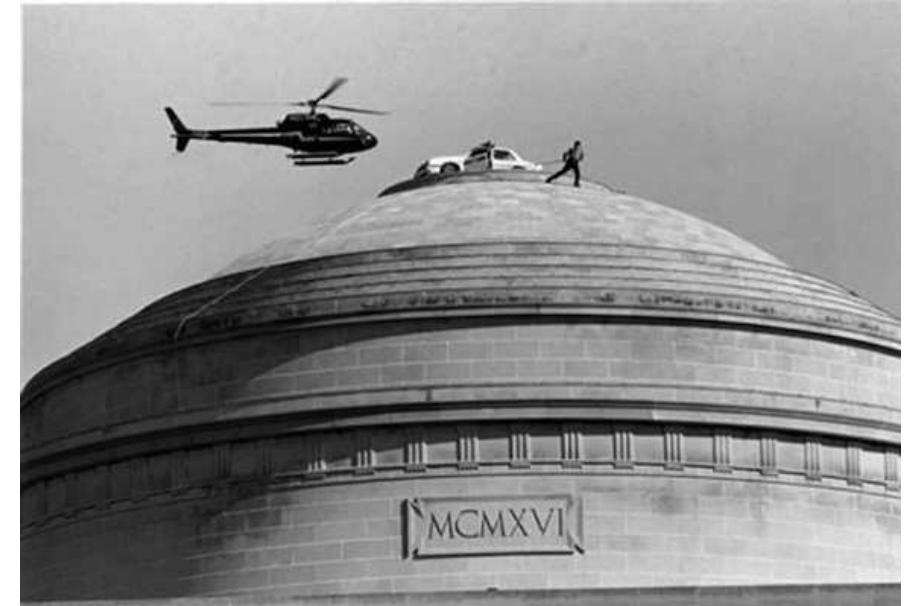


Antonio
Torralba



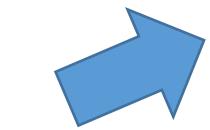
Aude
Oliva

Objects in the Scene Context

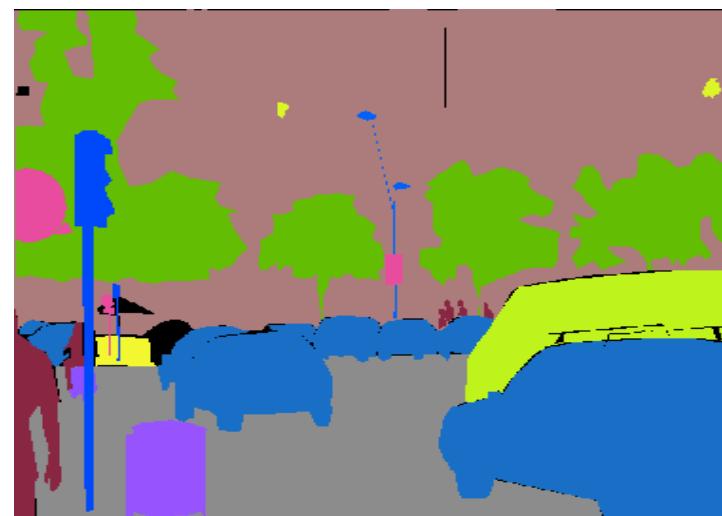


Challenge 1: Scene Classification

- Top1: street
- Top2: residential neighborhood
- Top3: crosswalk
- Top4: apartment building
- Top5: office building



Challenge 2: Scene Parsing



objects

tree
car
van
ashcan
person
streetlight
signboard
traffic light

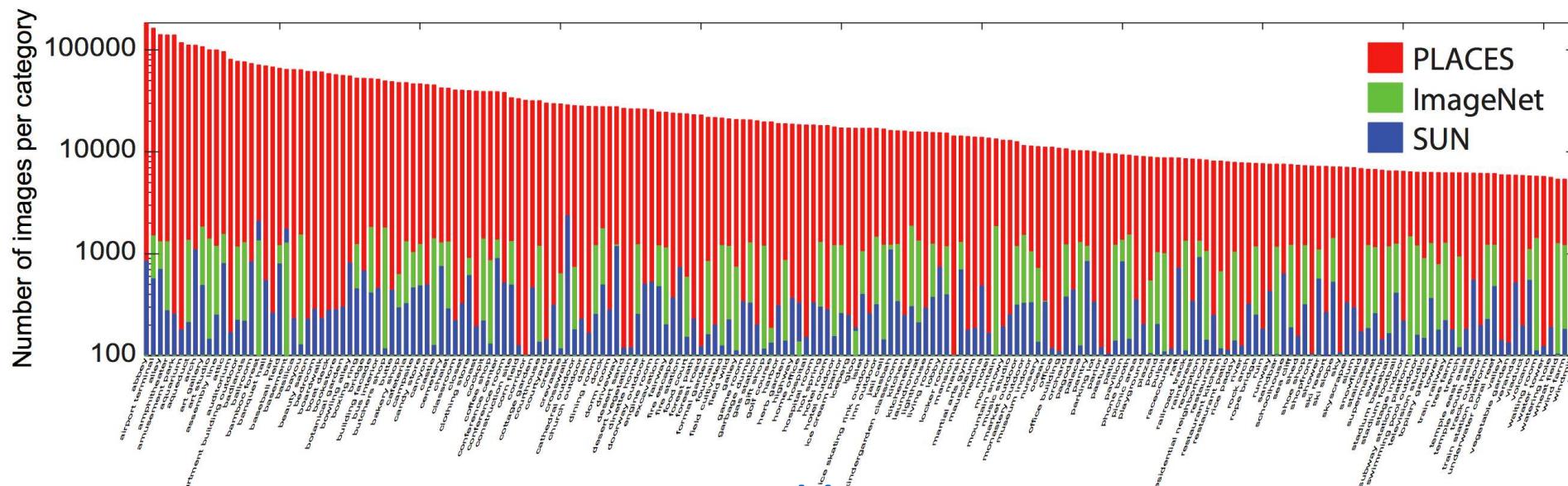
stuff

building
road
sidewalk

Places2 Challenge

Held in conjunction with ILSVRC and COCO at ECCV 2016

- 8 million training images from 365 categories of Places Database
- Test set: 900 images per category



Webpage: <http://places2.csail.mit.edu>

Constructing Places Database

1. Collect scene names from dictionary



~1000 scene names

2. Query and download images



696 adjectives + scene names
~ 90 million raw images downloaded

3. Annotate through Amazon Mechanical Turk

Instruction **Is this a cliff scene?** Submit (790 images left)
Definition: a high, steep or overhanging face of rock.



Three rounds of annotations

Indoor

bedroom



cafeteria



veterinarians office



elevator door

staircase



bar



conference center



shoe shop



Nature

fishpond



windmill



train station platform



watering hole



corral



amusement park



field road



arch



tower



soccer field

rainforest



swimming pool



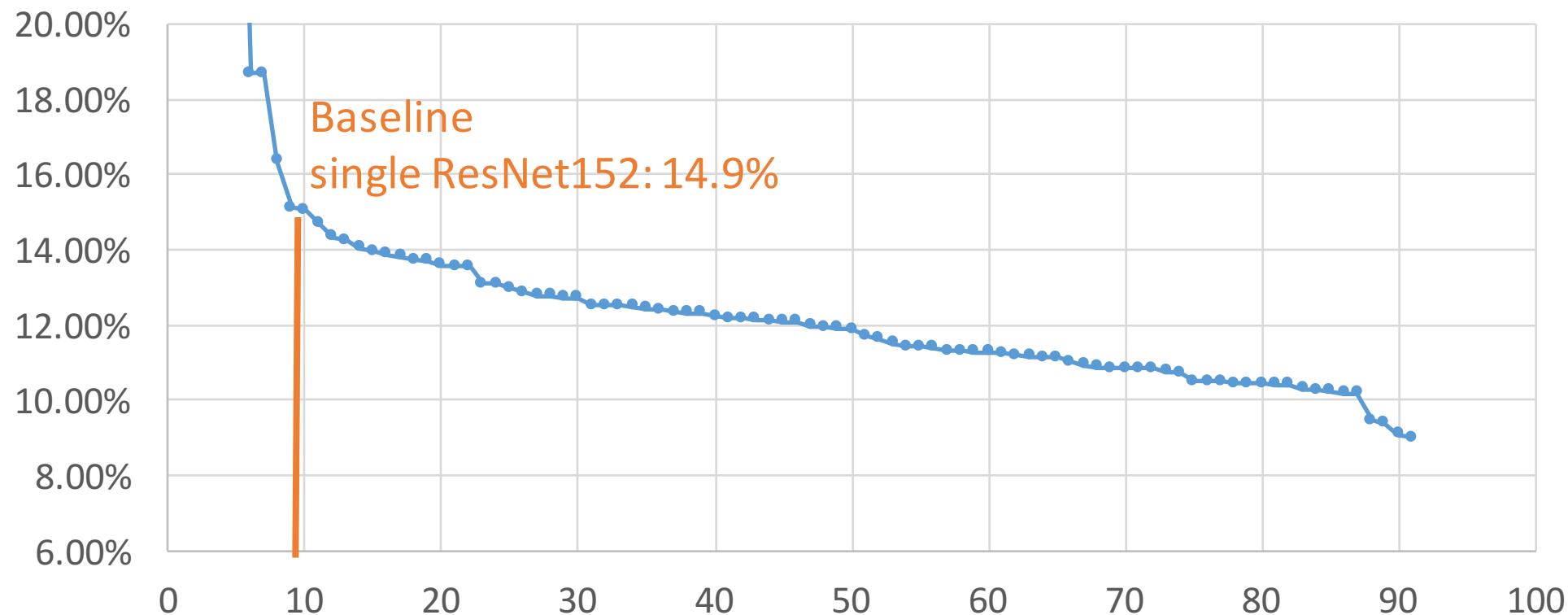
street



Results

92 valid submissions from **27** teams (each team allows to submit at most 5 submissions).

Top-5 errors of all the 92 submission (sorted)



Results

92 valid submissions from **27** teams.

Team Name	Top -5 Error
Hikvision	9.01%
MW	10.19%
Trimps-Soushen	10.30%
SIAT_MMLAB	10.43%
NTU-SC	10.85%
ResNet152	14.93%
VGG16	14.99%
AlexNet	17.25%

Single model
baselines

Hikvision
Qiaoyong Zhong, Chao Li, Yingying
Zhang, Haiming Sun, Shicai Yang, Di
Xie, Shiliang Pu.
Hikvision Research Institute

MW
Gang Sun and Jie Hu
**Chinese Academy of Sciences and
Peking University**

Trimps-Soushen
Jie Shao, Xiaoteng Zhang, Zhengyan
Ding, Yixin Zhao, Yanjun Chen, Jianying
Zhou, Wenfei Wang, Lin Mei,
Chuanping Hu
**The Third Research Institute of the
Ministry of Public Security, China**

Ambiguous predictions

1) Unusual activity in a scene

construction site



- top-1: martial arts gym
- top-2: stable
- top-3: boxing ring
- top-4: locker room
- top-5: basketball court

junkyard



- top-1: campsite
- top-2: sandbox
- top-3: beer garden
- top-4: market outdoor
- top-5: flea market indoor

2) Multiple scene parts

aquarium



- top-1: restaurant
- top-2: ice cream parlor
- top-3: coffee shop
- top-4: pizzeria
- top-5: cafeteria

lagoon



- top-1: balcony interior
- top-2: beach house
- top-3: boardwalk
- top-4: roof garden
- top-5: restaurant patio



Scene Parsing Challenge 2016

- New challenge this year
- Each pixel of the image is classified into some class



Scene
parsing



semantic mask

sky
road
sidewalk
escalator
building
car
bus
streetlight

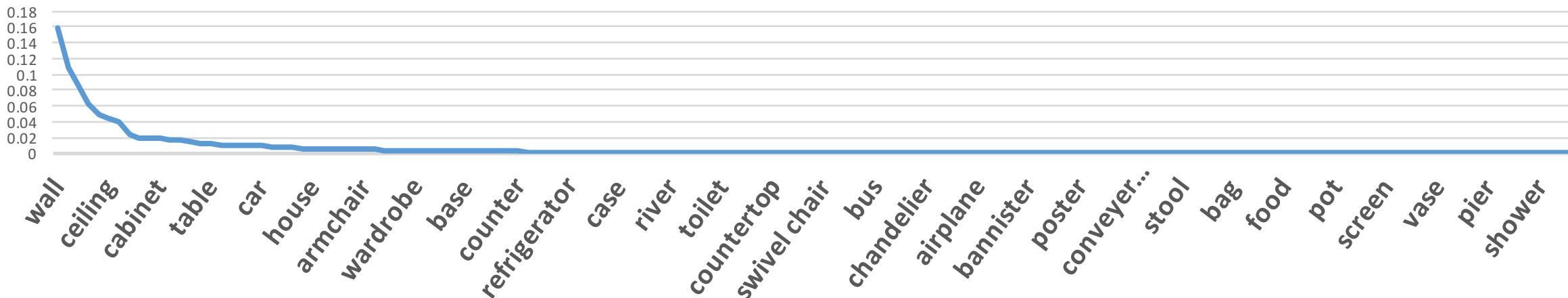
class label



Scene Parsing Challenge 2016

- 22,000 images for training and validation, 3,000 images for testing
- 150 classes of objects (car, person, table, etc) and stuff (sky, road, ceiling, etc)

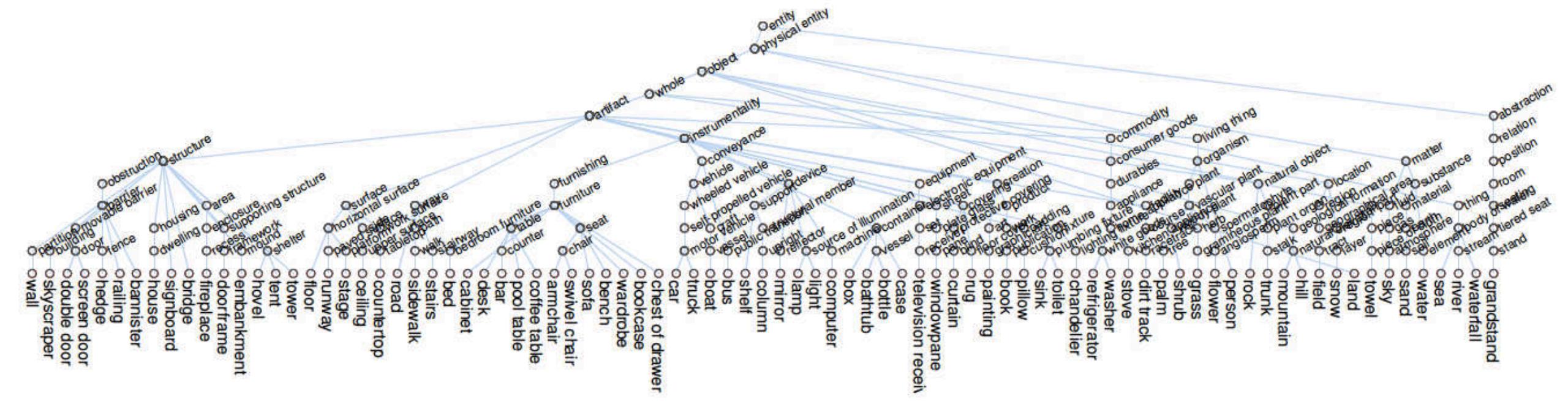
Pixel frequency in the training set





Scene Parsing Challenge 2016

- 22,000 images for training and validation, 3,000 images for testing
 - 150 classes of objects (car, person, table, etc) and stuff (sky, road, ceiling, etc)



Ms. Adela Barriuso

Constructing ADE Dataset

- Annotating each object instances in a scene
- Single expert annotator for a few years of work



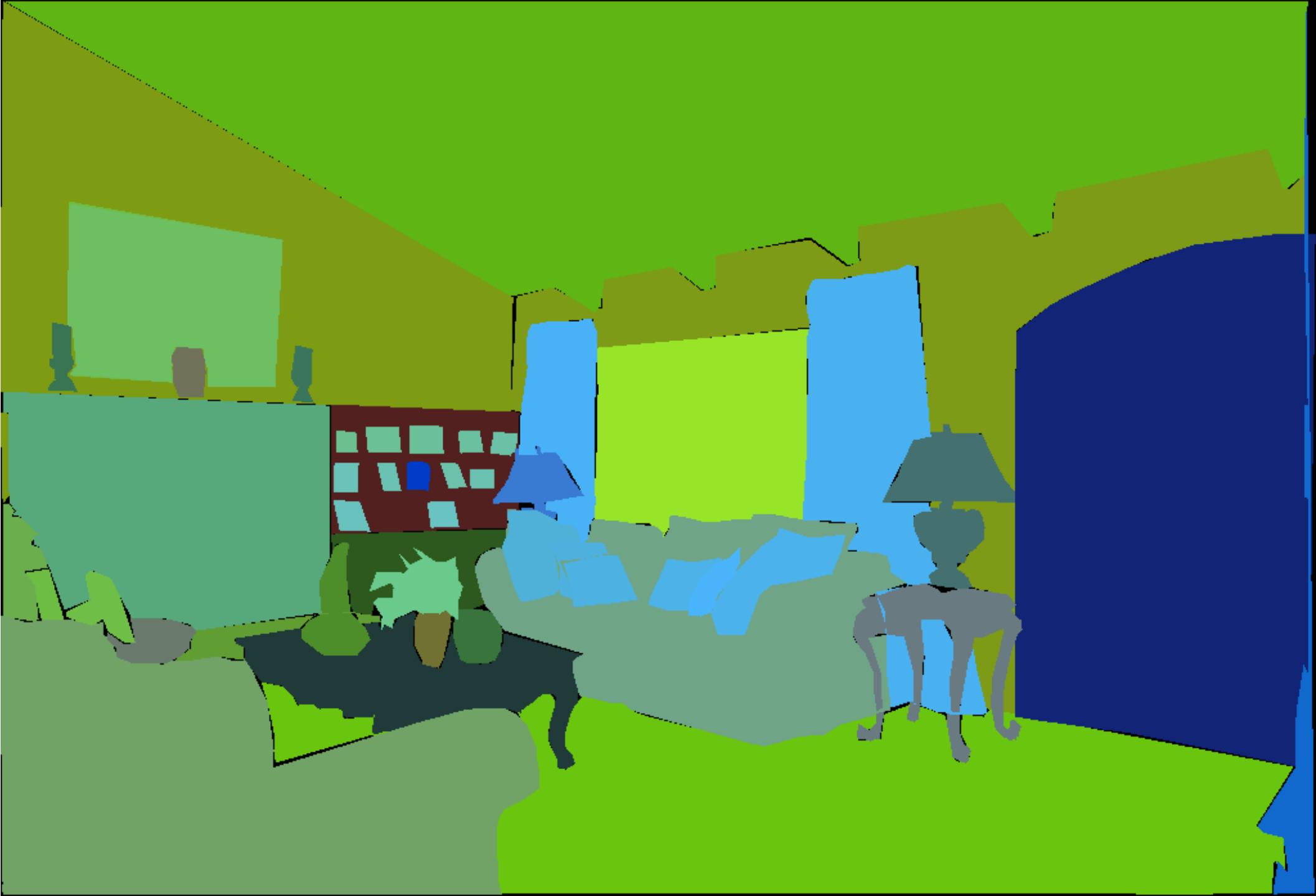
Labelme Annotation Tool

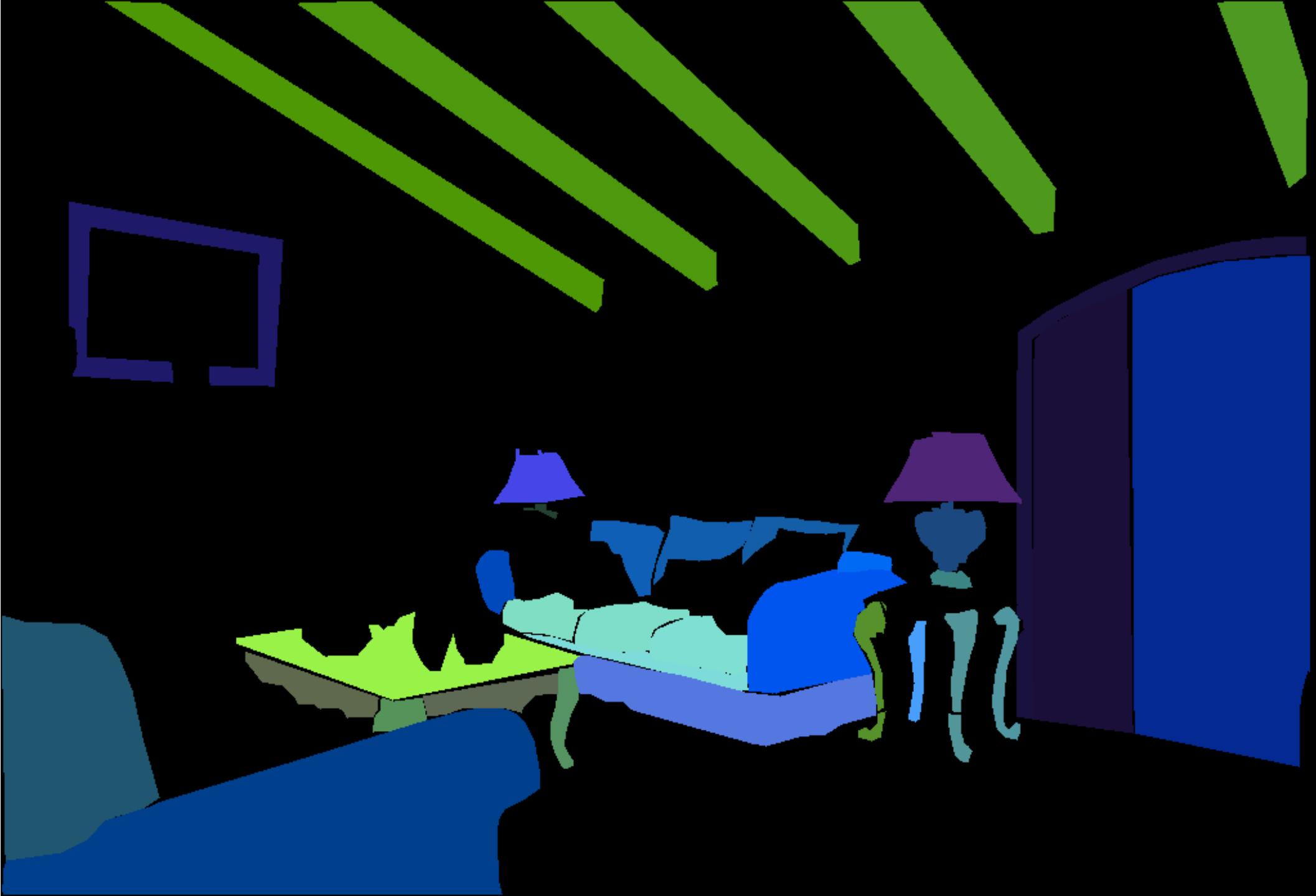
<http://groups.csail.mit.edu/vision/datasets/ADE20K/>



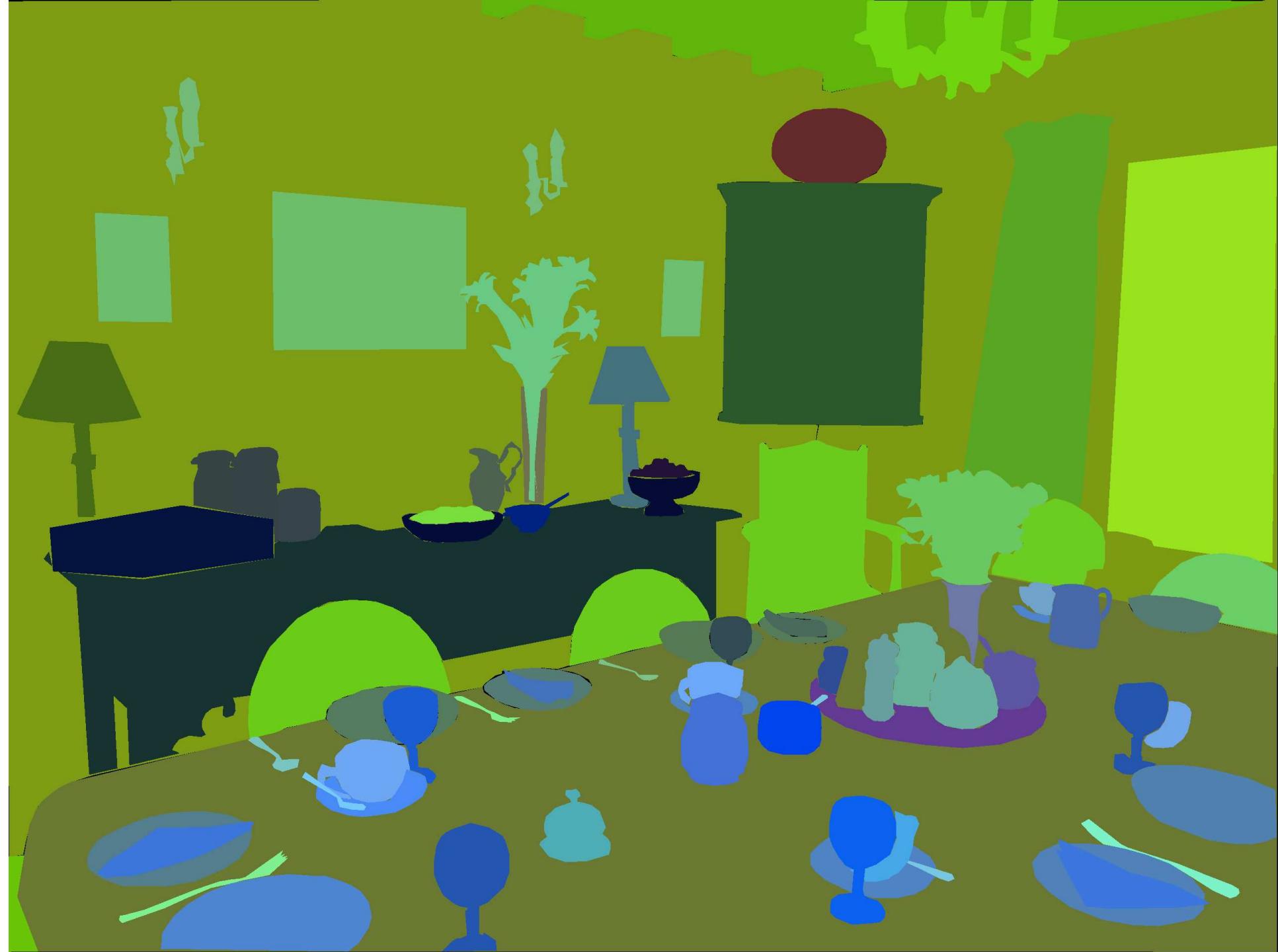


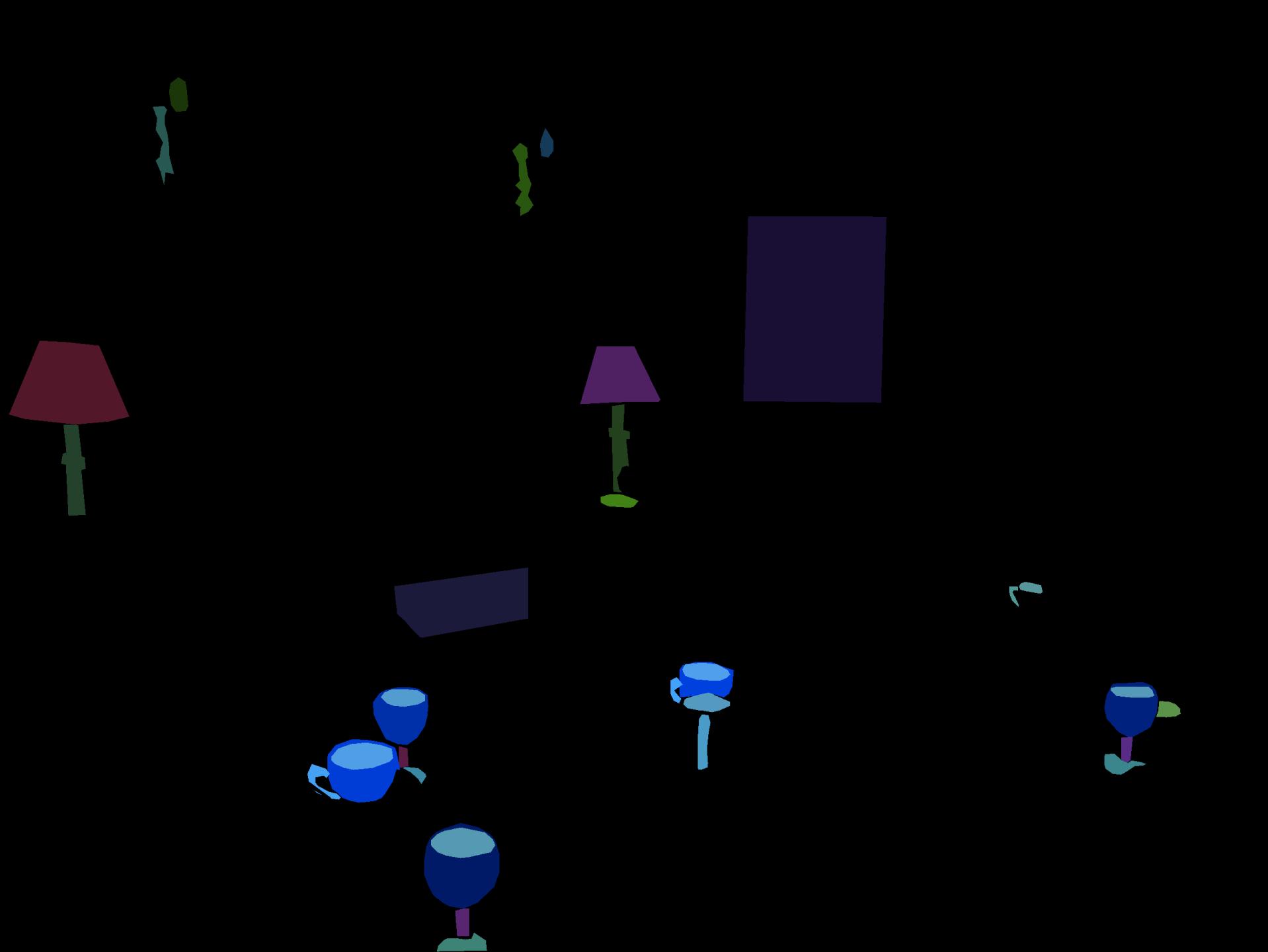
©davidpalermo.com



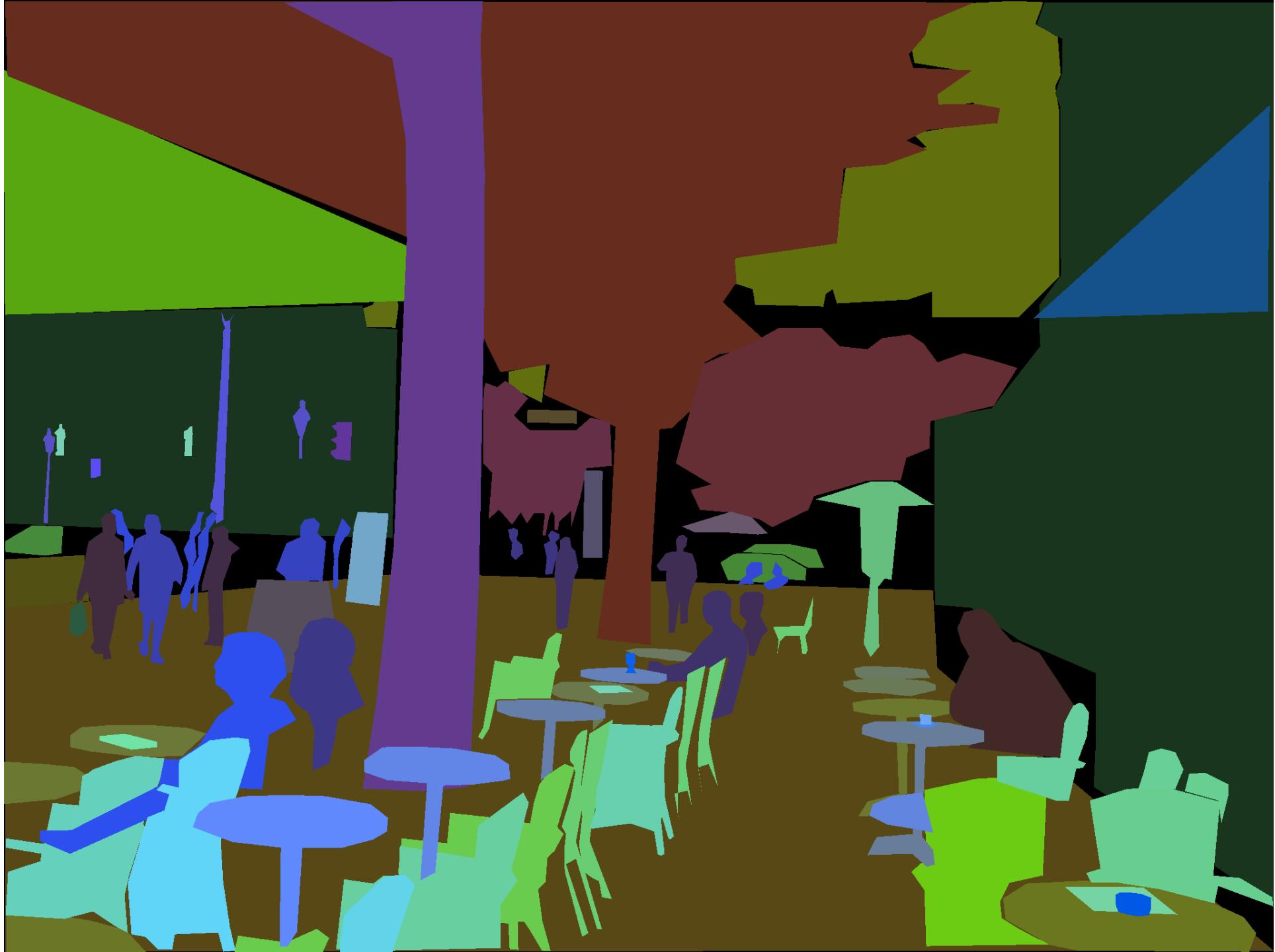








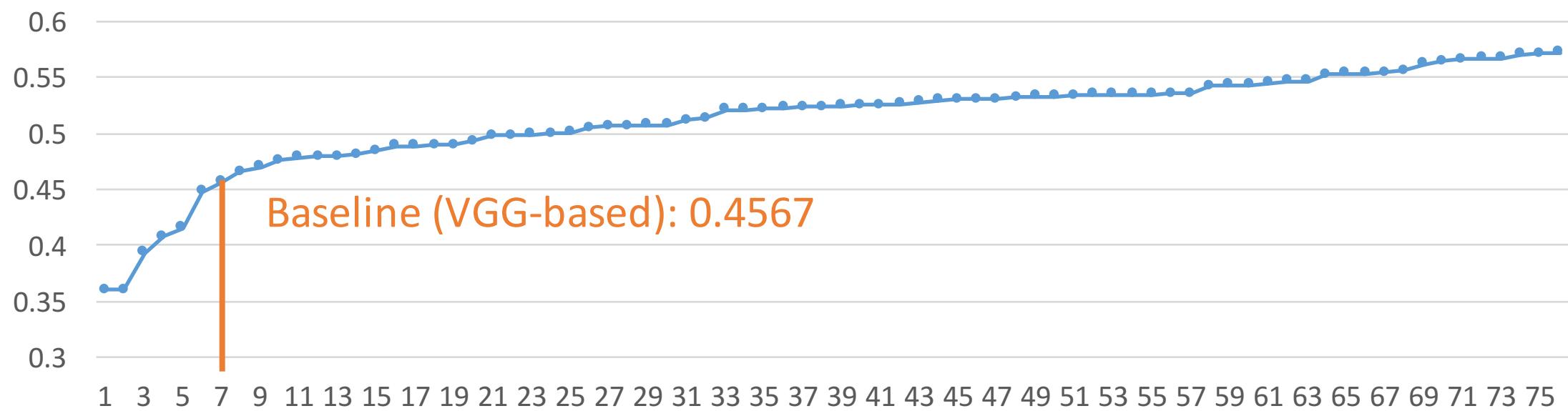




Results

75 valid submissions from 22 teams

Final score = (mean IoU + pixel accuracy)/2 for all the 75 submissions



Results

75 valid submissions from 22 teams

Final Score = (mean IoU + pixel accuracy) / 2

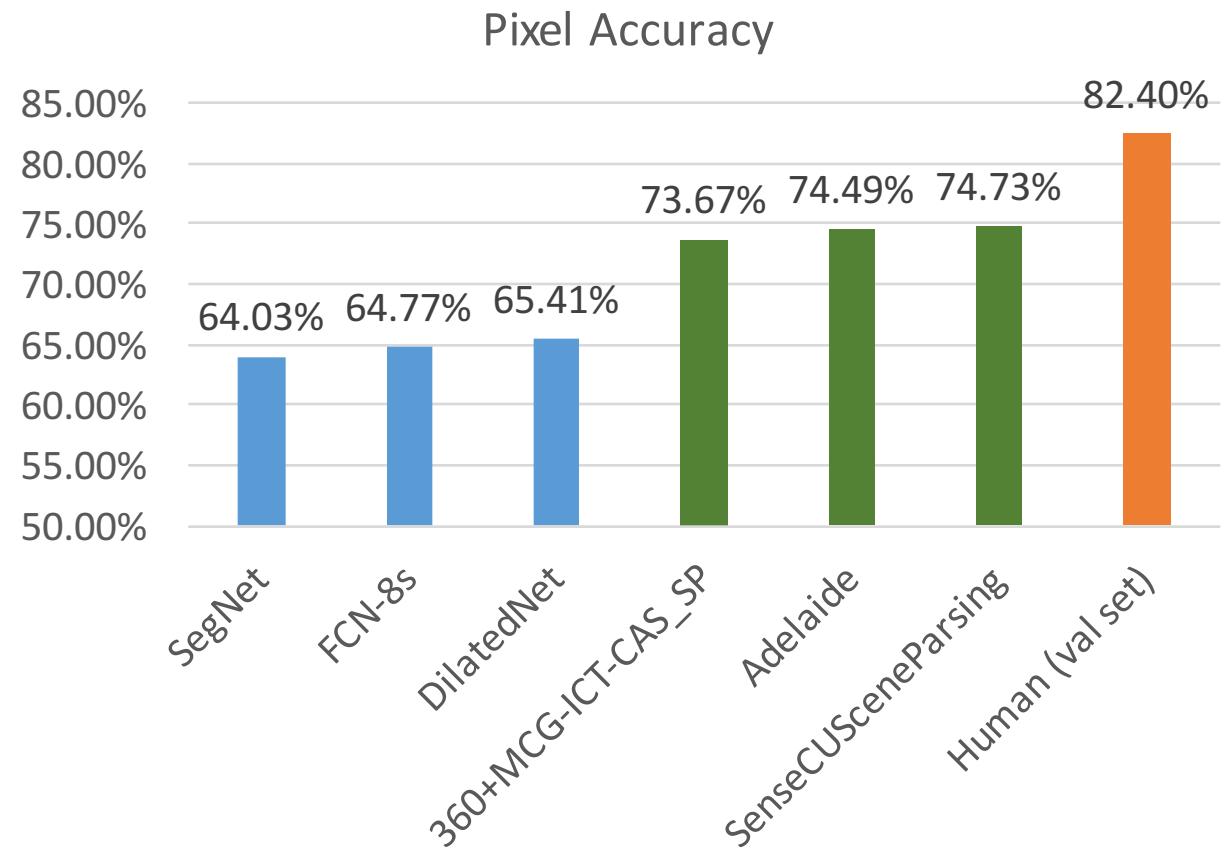
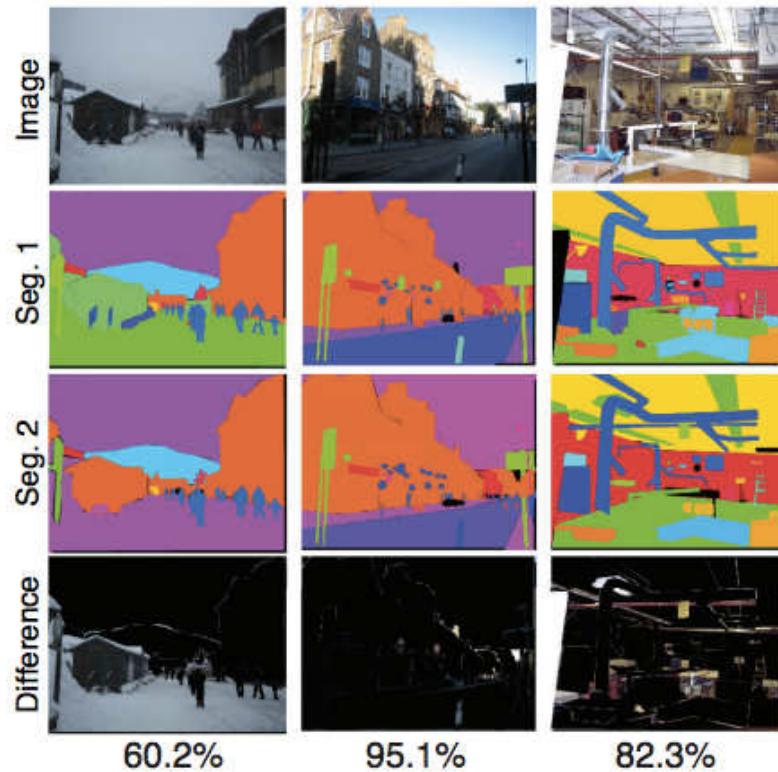
Team Name	Final Score
SenseCUSceneParsing	0.5721
Adelaide	0.5674
360+MCG-ICT-CAS_SP	0.5556
SegModel	0.5465
CASIA_IVA	0.5433
DilatedNet	0.4567
FCN-8s	0.4480
SegNet	0.4079

Single model
baselines

SenseCUSceneParsing Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Tong Xiao, Jiaya Jia Sensetime and CUHK, Hong Kong
Adelaide Zifeng Wu, Chunhua Shen, Anton van en Hengel University of Adelaide, Australia
360+MCG-ICT-CAS_SP Rui Zhang, Min Lin, Sheng Tang, Yu Li, YunPeng Chen, YongDong Zhang, JinTao Li, YuGang Han, ShuiCheng Yan Qihoo 360 ,Multimedia Computing Group,Institute of Computing Technology,Chinese Academy of Sciences (MCG- ICT-CAS), National University of Singapore (NUS)

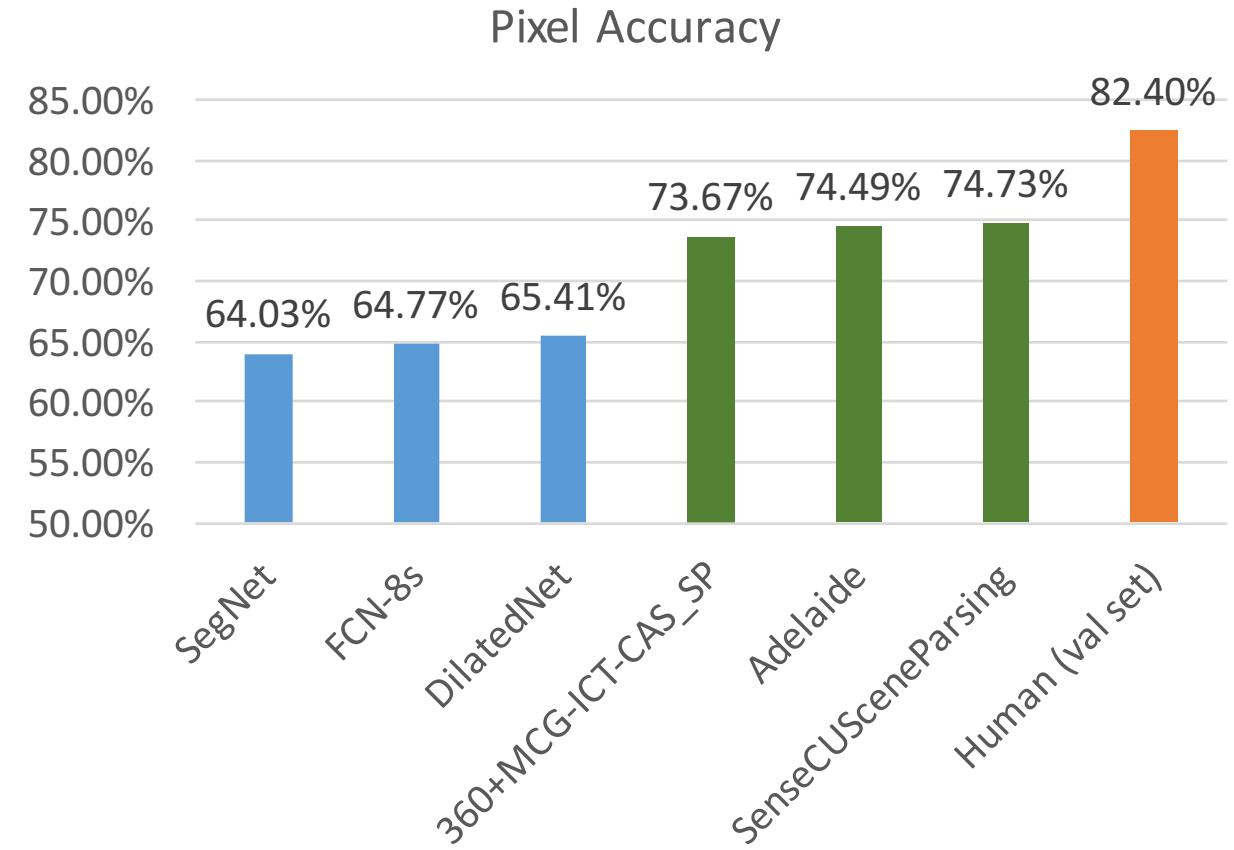
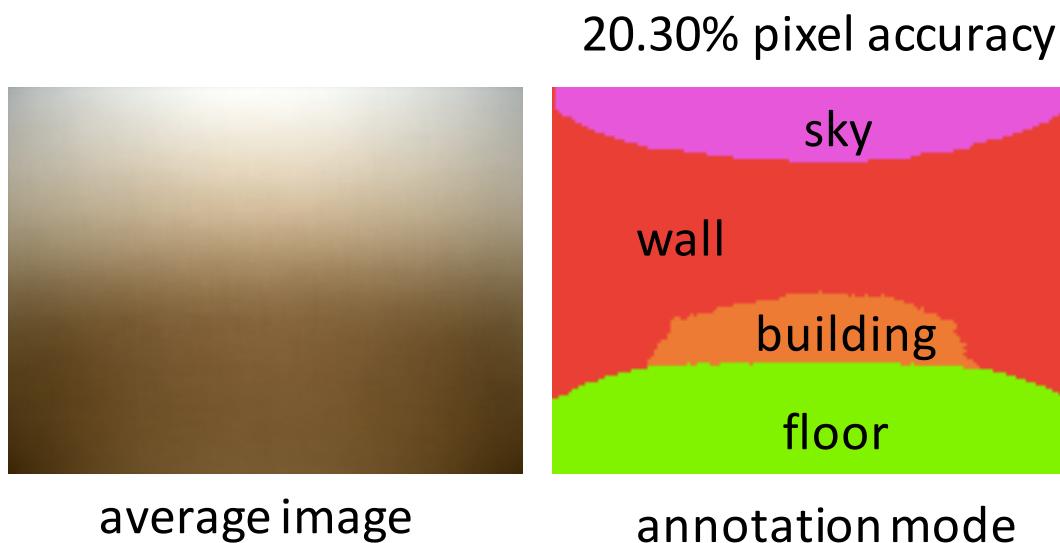
Data Consistency and Human Performance

- 61 images from val set are re-annotated after 6 months.
82.4% pixels got the same label.



Data Consistency and Human Performance

- 61 images from val set are re-annotated after 6 months.
82.4% pixels got the same label.



Image

Ground-truth

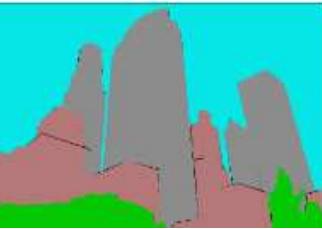
SenseCU...

Adelaide

360+MCG...

SegModel

CASIA_IVA



96.55%



96.36%



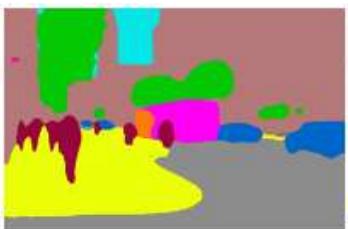
96.27%



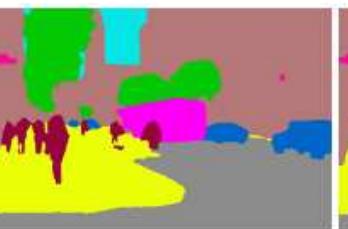
91.96%



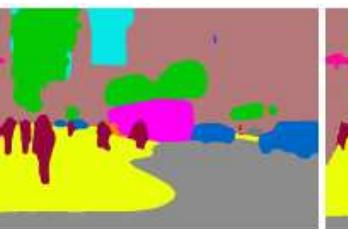
91.04%



92.73%



91.88%



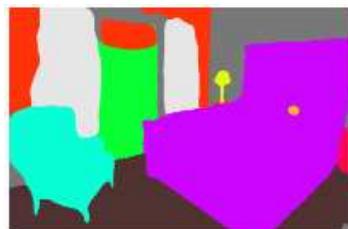
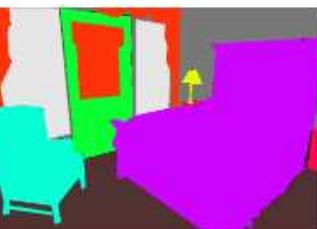
93.02%



92.62%



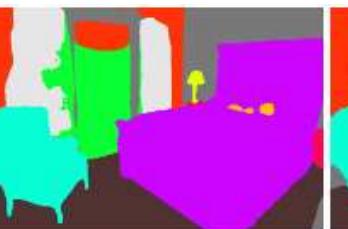
91.47%



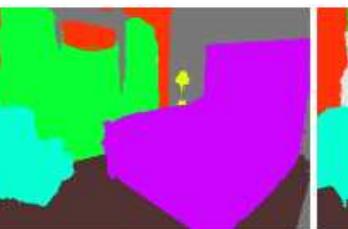
92.25%



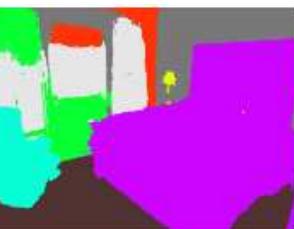
85.66%



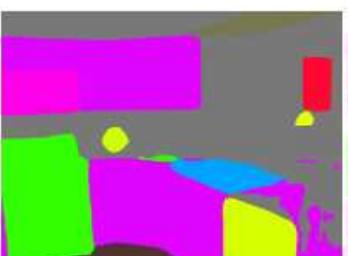
90.47%



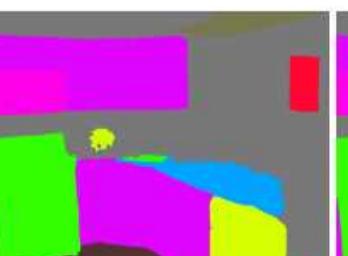
79.81%



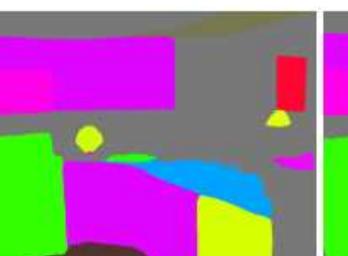
84.85%



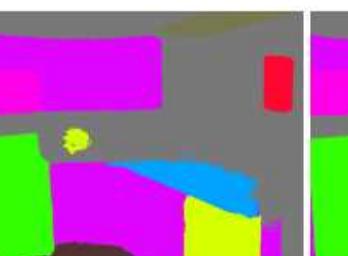
94.14%



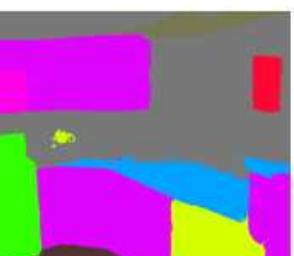
95.63%



96.55%



94.19%



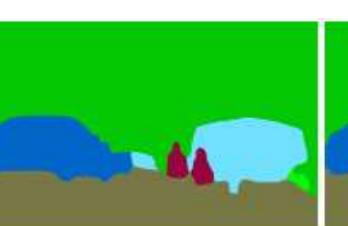
89.59%



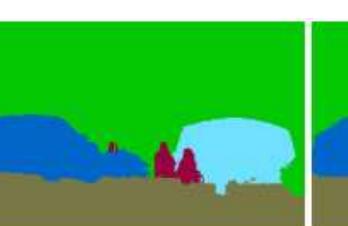
93.71%



93.51%



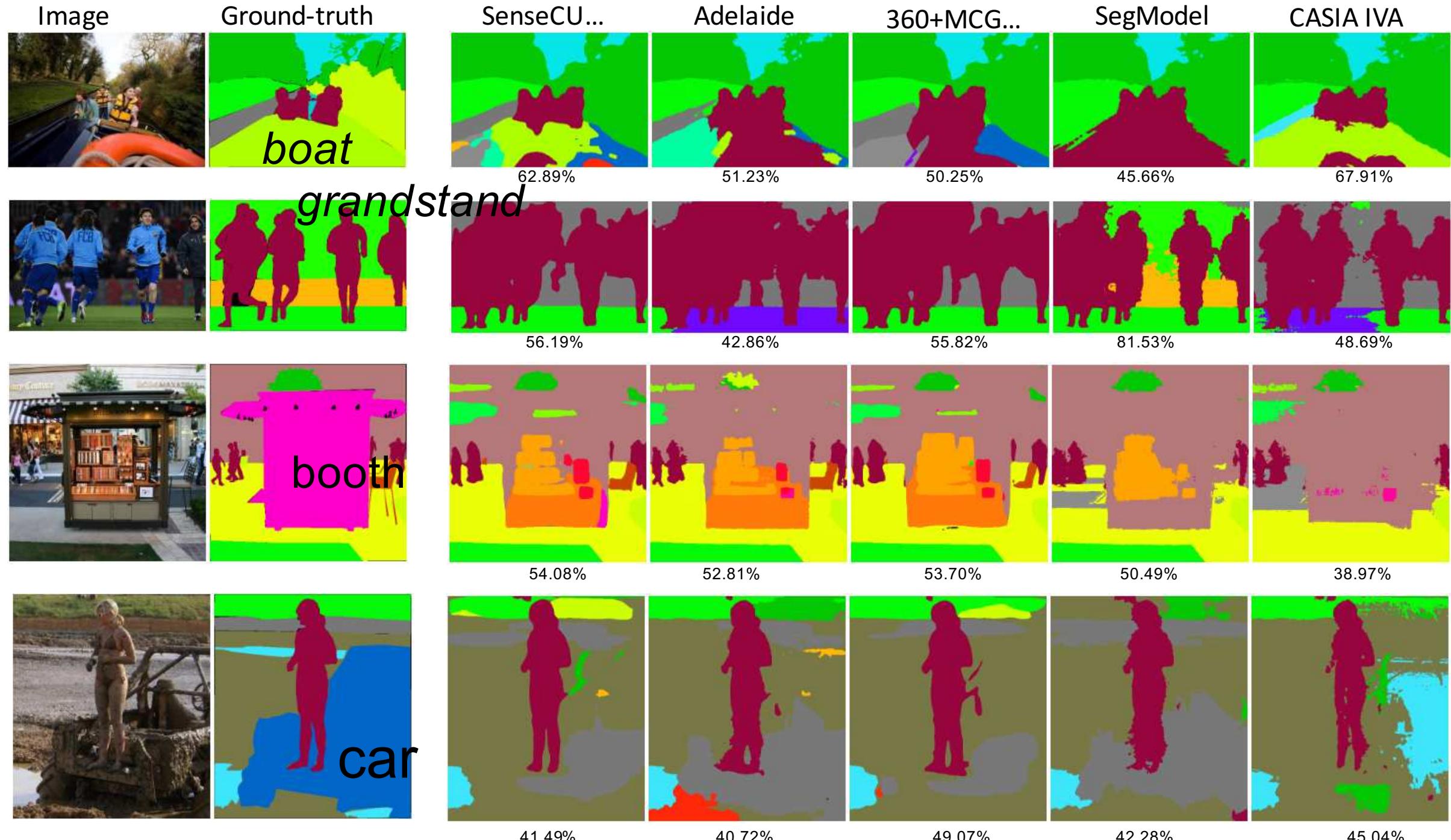
94.84%



94.89%



94.54%



Thanks all the Participants and Audiences!

<http://places2.csail.mit.edu>



<http://sceneparsing.csail.mit.edu>



Bolei
Zhou



Hang
Zhao



Xavier
Puig



Sanja
Fidler
(UToronto)



Adela
Barriuso



Aditya
Khosla



Antonio
Torralba



Aude
Oliva