



The COCO 2017 Stuff Challenge

H. Caesar, J. Uijlings, M. Maire, T.-Y. Lin, P. Dollar, V. Ferrari



Things vs. Stuff

Things	Stuff		
boat, person	sky, water		
size shape foreground composition of parts	no distinct spatial extent amorphous background materials, texture		



[Forsyth ORV 96, Heitz ECCV 08, Alexe CVPR 10, Tighe CVPR 13, Sun PAMI 13, Caesar arXiv 16, Shi ICCV 17]

Motivation



Motivation



Annotation protocol

- 1. Present predefined labels in a semantic hierarchy
 - \rightarrow Rapid access
 - → Avoid semantically overlapping free-form labels (*grass* vs. *field*)



Annotation protocol

- 2. Annotate superpixels with size-adjustable paintbrush tool
 - \rightarrow 2.2x speedup compared to polygon annotation
 - \rightarrow Similar quality since stuff boundaries are hard to delineate



Annotation protocol

3. Use existing COCO thing annotations \rightarrow Highly accurate stuff-thing transitions



Quality



Quality



COCO-Stuff 10K

- 10K images, 10 expert annotators
- For more information see

H. Caesar, J. Uijlings, V. Ferrari *COCO-Stuff: Thing and Stuff Classes in Context* arXiv preprint, March 2017.



Large-scale COCO-Stuff

- Join forces with
 - CVDF
 - Mighty Al
- Annotate entire COCO dataset (160k images)
- Generously sponsored by Mighty AI and CVDF



Dataset comparison

Dataset	Year	Images	Annotation type	Stuff Classes	Thing Classes	Classes / Usable
PASCAL Context	2014	10k	freeform, polygon	~ 152	~ 388	540 / 59
Cityscapes	2016	25k	predefined, polygon	~ 13	~ 14	30 / 19
ADE20K	2017	25k	freeform, polygon	~ 1242	~ 1451	2693 / 150
COCO-Stuff	ongoing	55k ⇒ 160k	predefined, superpixels	91	80	all

Challenge – Setup

• Semantic segmentation of 91 stuff and 1 'other' classes



COCO-Stuff Dataset

COCO 2017 Stuff Challenge



Challenge – Dataset splits



Challenge – No instances



Challenge – Metrics

Metrics on leaf categories and supercategories:

- Mean Intersection-over-Union (IOU) Primary challenge metric
- Frequency-weighted IOU
- Mean Accuracy
- Pixel Accuracy





Performance highly correlated across:

- Metrics (rel. to Mean IOU) (>= 95.4%)
- Leaf- and supercategories (>= 99.3%)
- Test-dev and test-challenge (>= 99.5%)

Supercategory results

IOU

^{1.00} ResNeXt-FPN G-RMI Active Vision Lab Baseline Deeplab



Category results

IOU

1.00

ResNeXt-FPN G-RMI Active Vision Lab Baseline Deeplab



Great results





Common mistakes – semantic ambiguity





Common mistakes – label inconsistency





Common mistakes – reflections / semi-transparency



door-stuff wall-conc wall-concre<mark>te</mark> othei wall-cone door-stuff wall mirror-stuff floor-marble wall-other loor-other rug furniture- floor-other floor-other

onc

paper

Challenge Ranking



Invited Speakers:

- ResNeXt-FPN / 1130 1145
- Active Vision Lab / 1145 1200
- G-RMI / 1445 1505 (Places)