Panoptic Segmentation

October 4th, 2021

Problem
Applications
Task Metric
Datasets
Human Consistency Study
Machine Performance

Semantic & Instance Segmentation

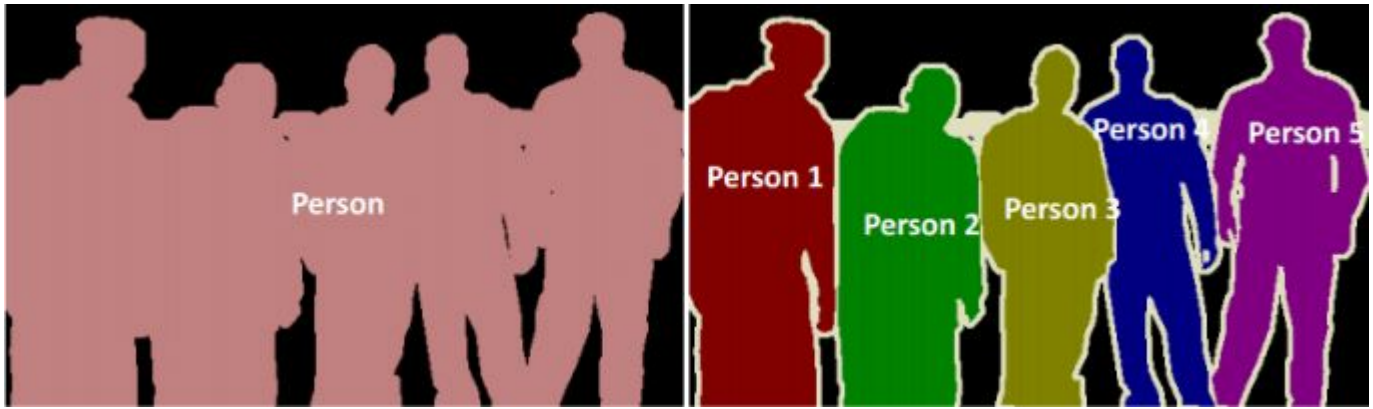


Figure from:
https://www.researchgate.net/figure/Semantic-segmentation-left-and-Instance-segmentation-right-8_fig1_339328277

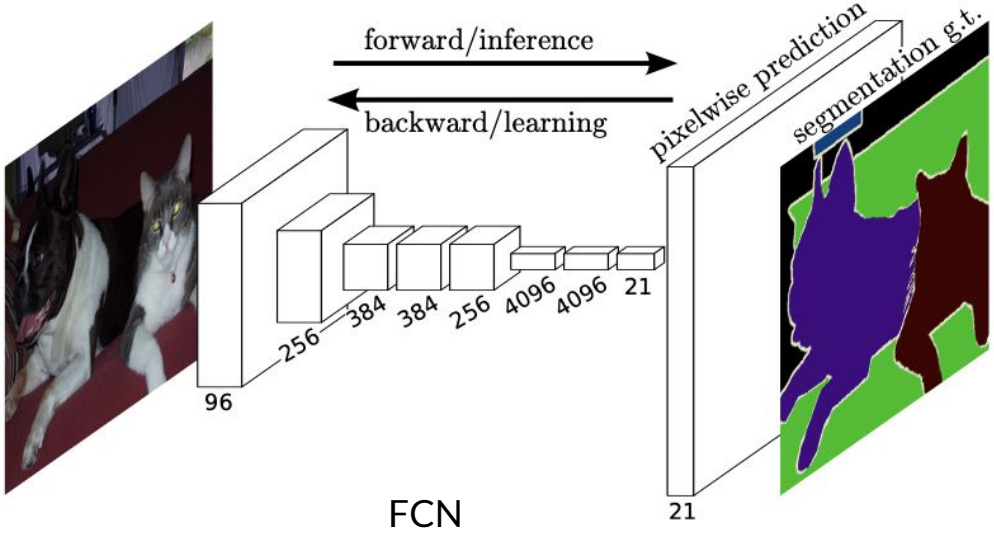
Semantic Segmentation

- Study of ***stuff***
- Assign one class label to each pixel in an image
- Treats *things* as stuff

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Semantic Segmentation

Typical model is fully convolutional



FCN

Figure from: Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

Semantic Segmentation

Evaluation Metrics

- Pixel accuracy
- Mean accuracy
- Mean IoU

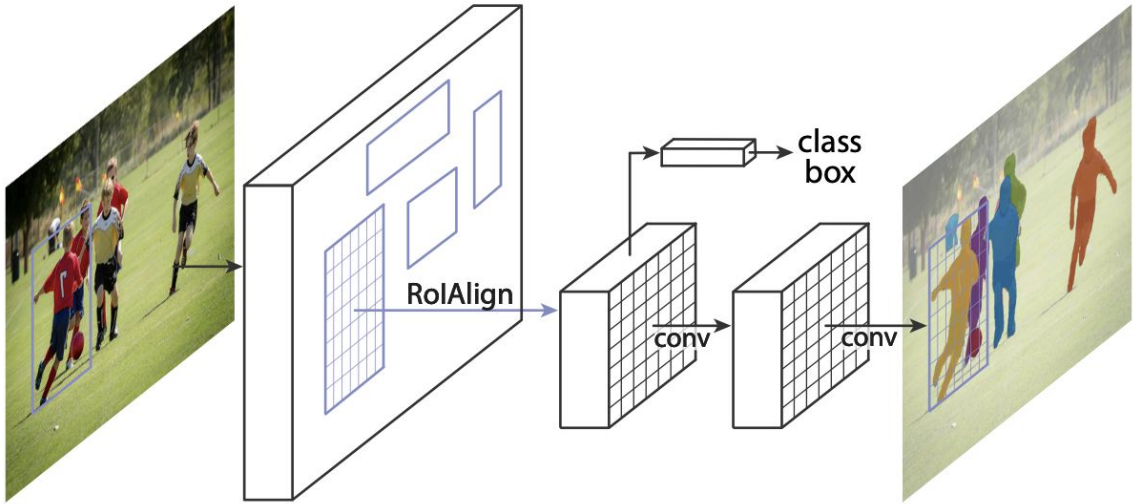
Instance Segmentation

- Study of ***things***
- Assign a class label and instance id to each pixel of an identified object
- Overlap allowed

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Instance Segmentation

Typical model includes object/region proposals



Mask R-CNN

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Instance Segmentation

Evaluation Metric

- Mean average precision

Schism of methods

	Semantic Segmentation	Instance Segmentation
Typically built on...	Fully convolutional networks	Object proposal and region-based methods
Evaluation metrics	• Pixel accuracy • Mean accuracy • Mean IoU	• Mean average precision

Can stuff and things be reconciled?

- Pre-deep learning researchers were interested in this problem
- Previously referred to by terms like *scene parsing* and *scene understanding*
- Direction is currently unpopular, and could be due to…
  - Lack of an appropriate metric
  - Recognition challenges

Revival of this direction

The authors propose a task that unifies segmentation by...

1. Encompassing both stuff and thing classes
2. Using a simple but general output format
3. Introducing a uniform evaluation metric

Panoptic Segmentation

- Study of ***stuff*** and ***things***
- Assign one class label and instance id to each pixel in an image

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Segmentation

Caveats

- No object overlap
- Not a multitask problem
- Confidence scores unpreferable

Figure from: https://ai-pool.com/d/why-do-the-masks-of-instances-overlap

Panoptic Segmentation

Panoptic Quality

- Metric that is simple, intuitive, and handles *things* and *stuff* uniformly
- Grounded via a human consistency and machine perf. study

Problem

Applications

Task Metric

Datasets

Human Consistency Study

Machine Performance

Assistive devices

Figure from: https://hscnews.usc.edu/augmented-reality-glasses-may-help-people-with-low-vision-better-navigate-their-environment

Lingual instructions for robots

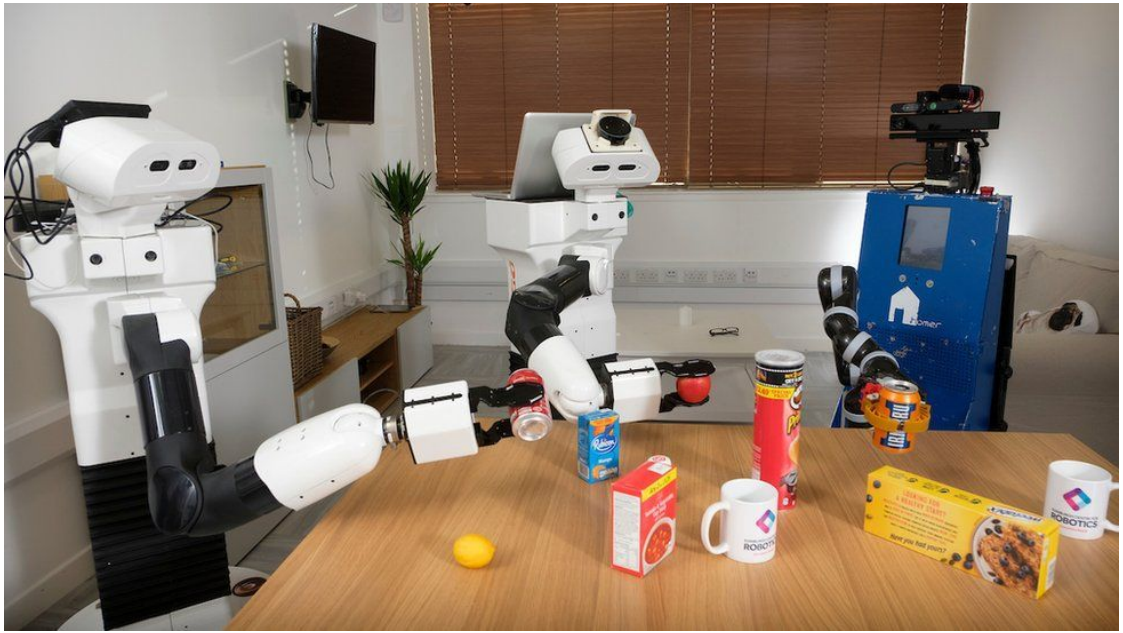


Figure from: https://www.bbc.com/news/uk-scotland-53241556

Map Building

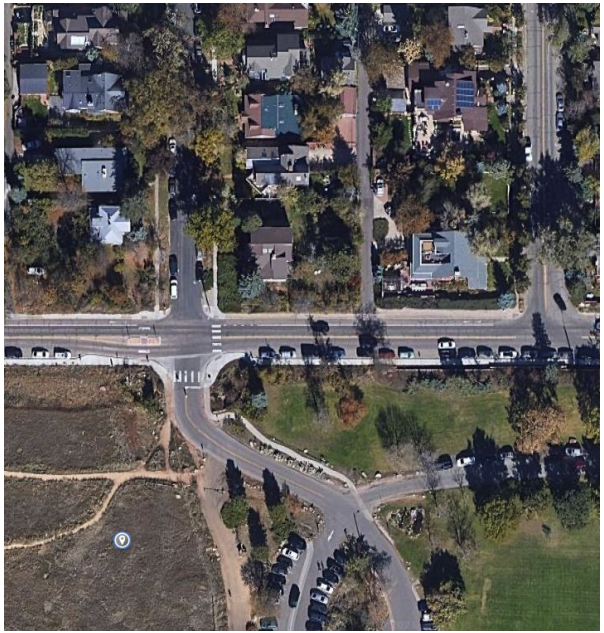


Figure from: Google Maps

Image Editing Software

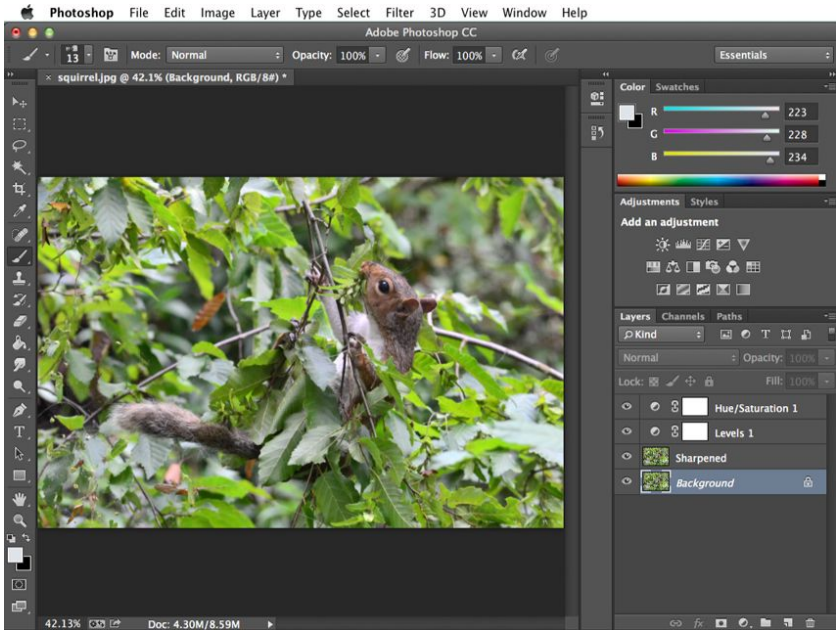


Figure from: https://edu.gcfglobal.org/en/photoshopbasics/getting-to-know-the-photoshop-interface/1/

Autonomous Vehicles

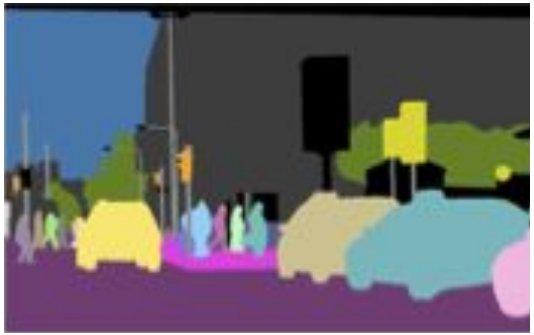


Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Problem
Applications
Task Metric
Datasets
Human Consistency Study
Machine Performance

Why a new metric?

- Recall:

	Semantic Segmentation	Instance Segmentation
Evaluation metrics	• Pixel accuracy • Mean accuracy • Mean IoU	• Mean average precision

Ignores instance metrics

Requires confidence scores

Why a new metric?

- No existing metric handles all classes (*things* and *stuff*) uniformly

Segment Matching

- Predicted segment and ground truth **match if their IoU > 0.5**
- Recall non-overlapping property: gives us a unique matching for each GT
- Splits segments into 3 sets: *TP, FP, and FN*

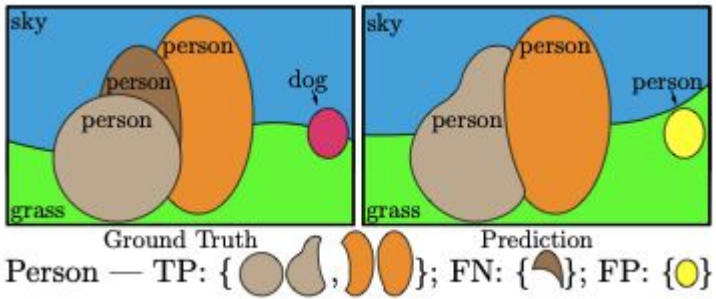


Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

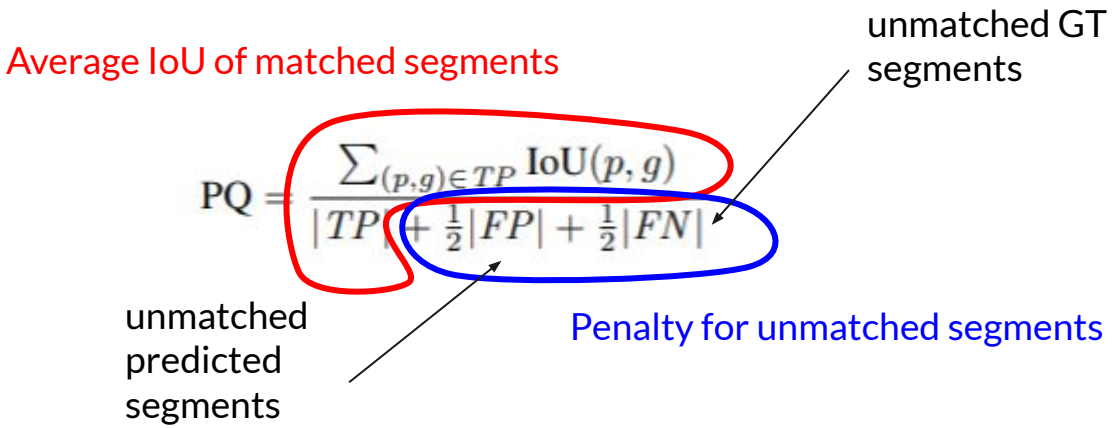


Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \times \frac{|TP|}{|TP|}$$

F1 score

$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}.$$

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

Ground truth

Prediction

$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

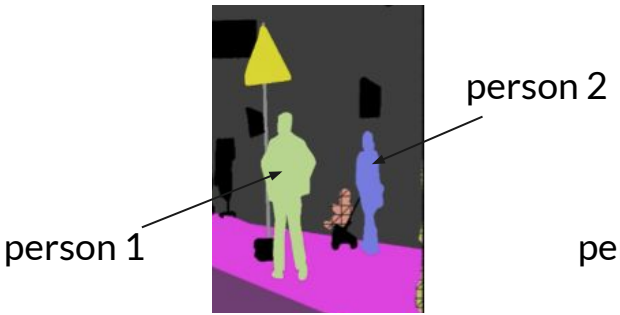
TP = ⌀

PQ = 0

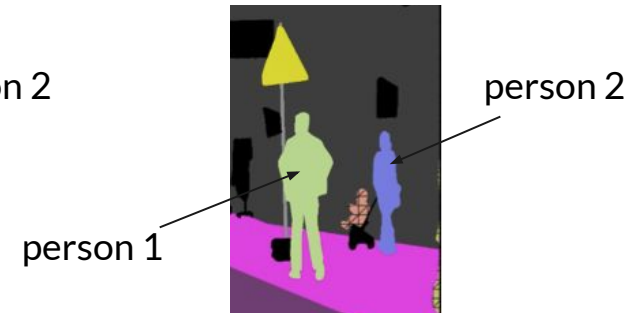
- What is the PQ for stuff class “road”?

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

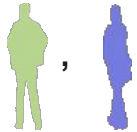


Ground truth



Prediction

$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

TP = {  }

FP = ⌀

FN = ⌀

PQ = 1

- What is the PQ for thing class “person”?

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

- Lower bound? 0
- Upper bound? 1

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

$$\mathrm{PQ} = \frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

- Computed independently for each class and then averaged

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Panoptic Quality

Final Comments

- Predictions are not evaluated for void labels:
  - out of class pixels
  - ambiguous/unknown pixels
- Group labels are not used during matching and do not result in FPs
  - Group labeling is a common annotation practice when delineation of instances is difficult

Problem
Applications
Task Metric
Datasets
Human Consistency Study
Machine Performance

Panoptic Segmentation Datasets

- Cityscapes
  - Egocentric driving scenarios
  - 5000 Images, 19 classes, 8 classes with instance level segmentation
- ADE20k
  - Over 25k Images. 100 thing and 50 stuff classes
- Mapillary Vistas
  - 25k Street view images. 28 stuff and 37 thing classes

These datasets contains all the information for a panoptic segmentation task.

COCO Dataset

- The COCO Dataset has 121,408 images.
- The COCO Dataset has 883,331 object annotations.
- The COCO Dataset has 80 classes.

Many of the Instance and Panoptic segmentation research at present relies on the COCO Dataset for generic objects training and validation

A peek into COCO Dataset structure


```
{
    "info": {
        "description": "COCO 2017 Dataset",
        "url": "http://cocodataset.org",
        "version": "1.0",
        "year": 2017,
        "contributor": "COCO Consortium",
        "date_created": "2017/09/01"
    },
    "licenses": [
        {
            "url": "http://creativecommons.org/licenses/by/2.0/",
            "id": 4,
            "name": "Attribution License"
        }
    ],
    "images": [
        {
            "id": 242287,
            "license": 4,
            "coco_url": "http://images.cocodataset.org/val2017/000000242287.jpg",
            "flickr_url": "http://farm3.staticflickr.com/2626/4072194513_edb6acfb2b_z.jpg",
            "width": 426,
            "height": 640,
            "file_name": "000000242287.jpg",
            "date_captured": "2013-11-15 02:41:42"
        },
        {
            "id": 245915,
            "license": 4,
            "coco_url": "http://images.cocodataset.org/val2017/000000245915.jpg",
            "flickr_url": "http://farm1.staticflickr.com/88/211747310_f58a16631e_z.jpg",
            "width": 500,
            "height": 333,
            "file_name": "000000245915.jpg",
            "date_captured": "2013-11-18 02:53:27"
        }
    ],
    "annotations": [
        {
            "id": 125686,
            "category_id": 2,
            "iscrowd": 0,
            "segmentation": [[164.81, 417.51, 164.81, 417.51, 164.81, 417.51, 159.31, 409.27, 159.31, 409.27, 155.19, 409.27, 155.19, 410.64, 152.45, 413.39, 144.21, 413.39, 140.09, 413.39, 137.34, 413.39, 134.59, 414
            "image_id": 242287,
            "area": 42061.80340000001,
            "bbox": [19.23, 383.18, 314.5, 244.46]
        },
        {
```


Figure from : https://www.youtube.com/watch?v=h6s61a_pqfM&t=109s

A peek into COCO Dataset structure


```
        "id": 125686,
        "category_id": 2,
        "iscrowd": 0,
        "segmentation": [[164.81, 417.51, 164.81, 417.51, 164.81, 417.51, 159.31, 409.27, 159.31, 409.27, 155.19, 409.27, 155.19,
        "image_id": 242287,
        "area": 42061.80340000001,
        "bbox": [19.23, 383.18, 314.5, 244.46]
    },
    {
        "id": 1409619,
        "category_id": 22,
        "iscrowd": 0,
        "segmentation": [[376.81, 238.8, 378.19, 228.91, 382.15, 216.06, 383.14, 210.72, 385.9, 207.56, 386.7, 207.16, 387.29, 2
        "image_id": 245915,
        "area": 3556.2197000000015,
        "bbox": [376.81, 189.76, 96.7, 56.95]
    },
    {
        "id": 1410165,
        "category_id": 22,
        "iscrowd": 0,
        "segmentation": [[486.34, 239.01, 477.88, 244.78, 468.26, 245.16, 464.41, 244.78, 458.64, 250.16, 451.72, 249.39, 445.56,
        "image_id": 245915,
        "area": 1775.8932499999994,
        "bbox": [445.56, 205.16, 54.44, 71.55]
    },
    {"id": 1410330, "category_id": 22, "iscrowd": 0, "segmentation": [[402.59, 196.81, 410.82, 177.35, 416.81, 171.36, 428.78, 171.36, 433.27, 166.87, 451.98, 164.63, 468.44, 165.38, 480.42, 172.11, 486.4, 181.84,
    {"id": 1410622, "category_id": 22, "iscrowd": 0, "segmentation": [[480.86, 166.6, 481.62, 161.78, 482.63, 158.99, 483.9, 156.2, 485.42, 151.63, 486.19, 147.32, 486.69, 143.0, 486.95, 139.7, 487.45, 135.13, 487
    {"id": 1410759, "category_id": 22, "iscrowd": 0, "segmentation": [[439.19, 122.67, 440.27, 115.81, 441.34, 106.38, 444.98, 100.16, 450.56, 100.38, 456.99, 101.23, 466.85, 103.59, 479.72, 108.09, 488.72, 114.53
    {"id": 1410834, "category_id": 22, "iscrowd": 0, "segmentation": [[331.42, 255.49, 322.64, 250.05, 318.04, 240.02, 318.46, 250.47, 308.01, 246.71, 306.33, 237.93, 303.82, 219.96, 294.63, 220.79, 284.18, 214.94
    {"id": 1410880, "category_id": 22, "iscrowd": 0, "segmentation": [[222.29, 191.74, 233.36, 181.64, 241.18, 167.63, 248.35, 154.93, 257.47, 153.95, 281.57, 149.39, 294.27, 149.39, 300.79, 152.98, 312.84, 160.14
    {"id": 1411090, "category_id": 22, "iscrowd": 0, "segmentation": [[245.09, 145.01, 245.09, 138.15, 247.12, 132.57, 250.93, 126.73, 258.29, 125.21, 267.68, 125.97, 281.64, 126.23, 290.78, 127.75, 299.91, 135.62
    {"id": 1411108, "category_id": 22, "iscrowd": 0, "segmentation": [[347.3, 157.93, 341.38, 156.45, 338.91, 159.9, 341.38, 167.31, 336.94, 167.8, 328.54, 155.95, 324.59, 151.02, 321.14, 159.9, 324.1, 162.86, 324
    {"id": 1411138, "category_id": 22, "iscrowd": 0, "segmentation": [[72.59, 177.35, 73.33, 169.12, 77.08, 160.89, 80.07, 156.4, 85.31, 149.66, 88.3, 145.17, 92.04, 139.19, 96.53, 133.95, 100.27, 127.96, 104.02,
    {"id": 1411160, "category_id": 22, "iscrowd": 0, "segmentation": [[389.47, 161.42, 391.37, 145.58, 400.25, 136.08, 416.09, 123.41, 438.26, 118.34, 464.88, 122.14, 472.48, 137.35, 480.08, 144.95, 486.42, 151.92
    {"id": 1411174, "category_id": 22, "iscrowd": 0, "segmentation": [[234.05, 180.69, 230.62, 180.38, 228.13, 183.18, 229.07, 185.37, 222.83, 190.04, 219.4, 183.18, 218.78, 186.61, 216.28, 185.37, 217.53, 179.13,
    {"id": 2210312, "category_id": 22, "iscrowd": 0, "segmentation": [[40.18, 136.53, 43.72, 132.05, 46.31, 112.74, 47.01, 107.08, 53.14, 101.43, 66.09, 101.67, 70.33, 103.32, 79.52, 110.85, 84.23, 116.51, 84.23,
    {
        "id": 902200245915,
        "category_id": 22,
        "iscrowd": 1,
        "segmentation": {
            "size": [333, 500],
            "counts": [26454, 2, 651, 3, 13, 1, 313, 12, 6, 3, 312, 21, 310, 23, 310, 22, 310, 21, 12, 4, 296, 20, 12, 6, 294, 19, 13, 8, 293, 18, 14, 9, 292, 16, 15, 10, 292, 14, 17, 10, 292, 13, 17, 11, 293, 11,
        },
        "image_id": 245915,
        "area": 3188,
        "bbox": [79, 127, 140, 57]
```


Figure from : https://www.youtube.com/watch?v=h6s61a_pqfM&t=109s

A peek into COCO Dataset structure

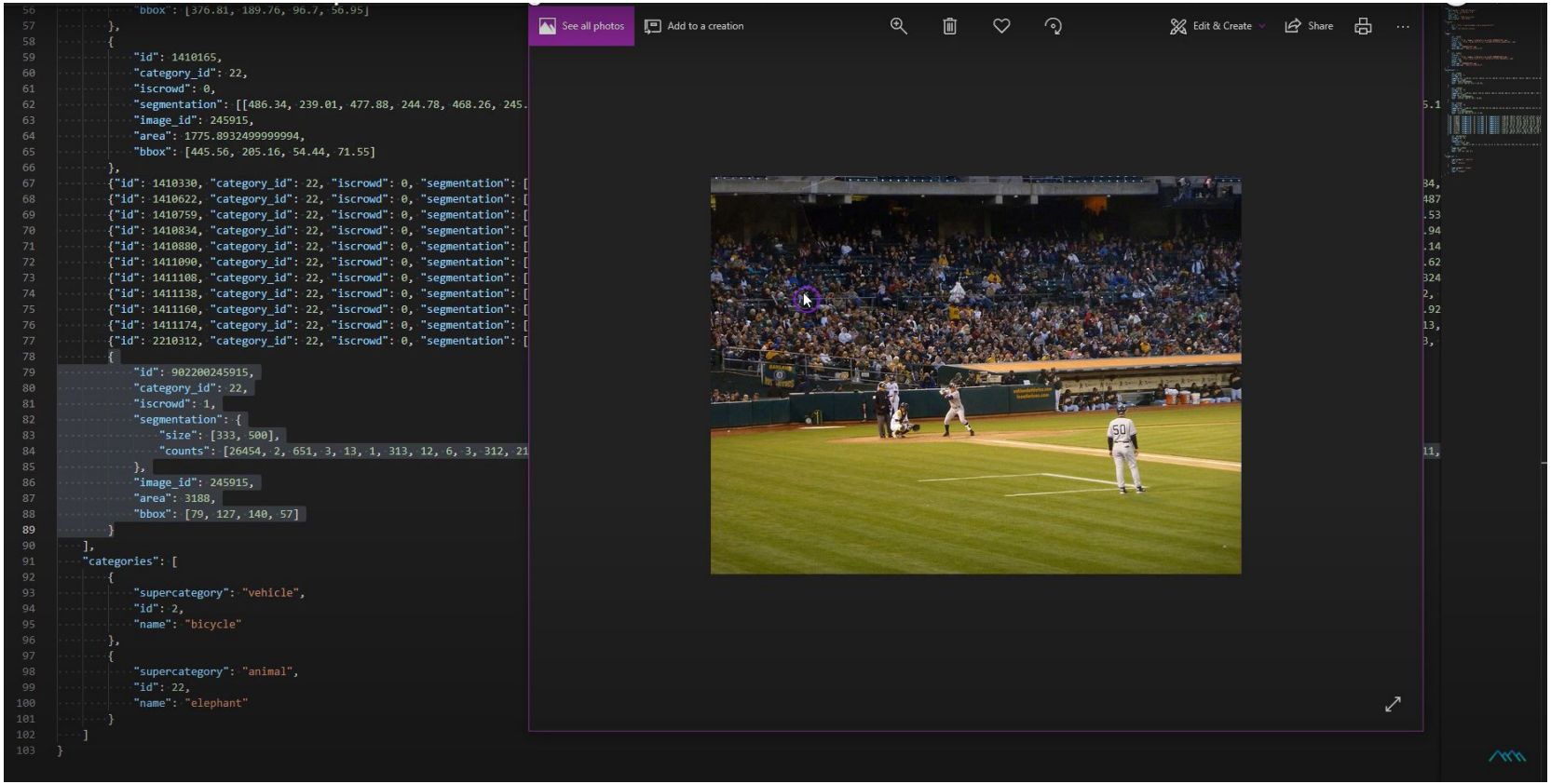


Figure from : https://www.youtube.com/watch?v=h6s61a_pqfM&t=109s

A peek into COCO Dataset structure

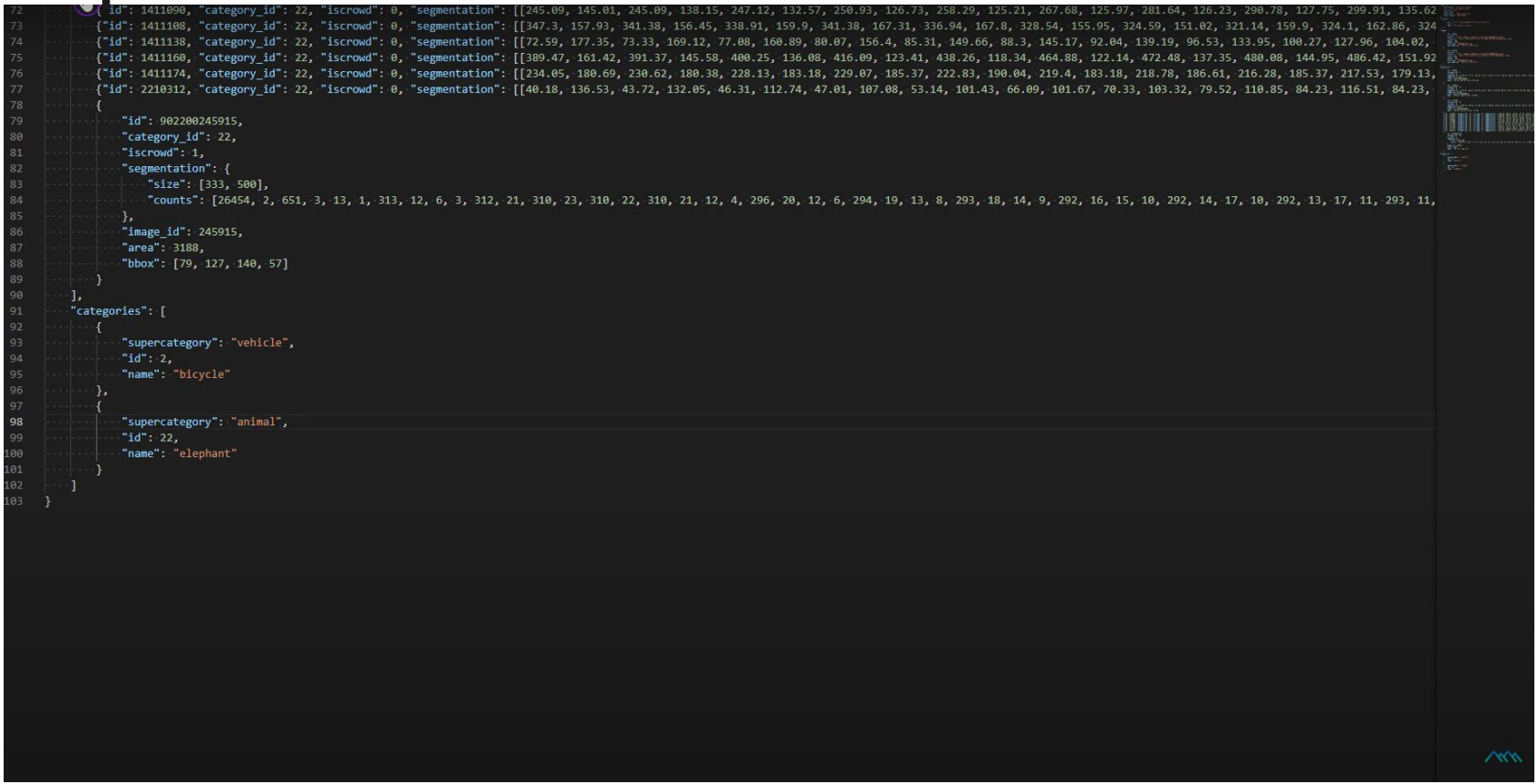


Figure from : https://www.youtube.com/watch?v=h6s61a_pqfM&t=109s

Problem
Applications
Task Metric
Datasets
Human Consistency Study
Machine Performance

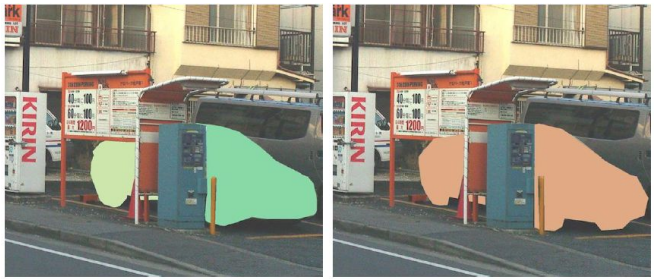
Human Consistency Study

Understanding the Panoptic Segmentation task with human annotations

Method:

- With doubly annotated images for Cityscapes, ADE20k and Vistas annotated independently by different annotators
- Considers one annotation for each image as ground truth and other as prediction

Original Image



Two annotated images of the same image

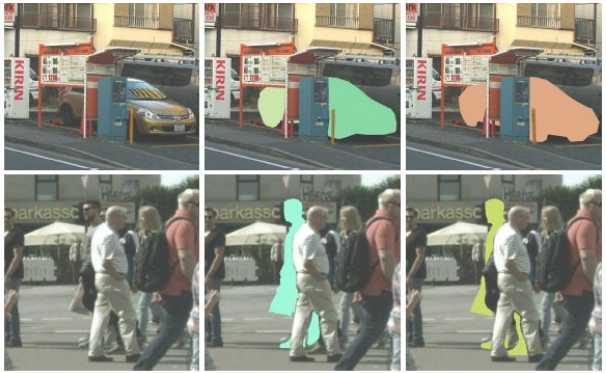
Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Human Consistency Study

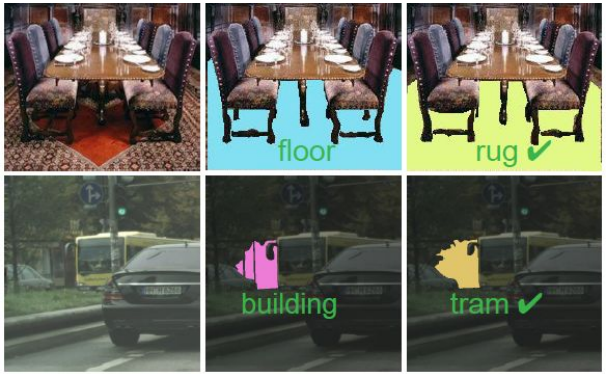
Helps understand

- The Panoptic Segmentation task in detail
- The details of PQ
- The breakdown of Human consistency along various axes(factors)

Errors visualization



Segmentation Error



Classification Error

How can we observe this in the PQ value?

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Stuff vs things

	PQ	PQ^{St}	PQ^{Th}	SQ	SQ^{St}	SQ^{Th}	RQ	RQ^{St}	RQ^{Th}
Cityscapes	69.7	71.3	67.4	84.2	84.4	83.9	82.1	83.4	80.2
ADE20k	67.1	70.3	65.9	85.8	85.5	85.9	78.0	82.4	76.4
Vistas	57.5	62.6	53.4	79.5	81.6	77.9	71.4	76.0	67.7

Human consistency for stuff vs things

Things can be difficult to annotate compared to stuff. But not by a big margin

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Human consistency vs scale

	PQ^S	PQ^M	PQ^L	SQ^S	SQ^M	SQ^L	RQ^S	RQ^M	RQ^L
Cityscapes	35.1	62.3	84.8	67.8	81.0	89.9	51.5	76.5	94.1
ADE20k	49.9	69.4	79.0	78.0	84.0	87.8	64.2	82.5	89.8
Vistas	35.6	47.7	69.4	70.1	76.6	83.1	51.5	62.3	82.6

Human consistency vs scale

Small size - > difficult to annotate

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Problem
Applications
Task Metric
Datasets
Human Consistency Study
Machine Performance

Machine Performance

There wasn’t an existing Algorithmic model to perform the Panoptic Segmentation task at the time of introduction of this idea

How to generate machine results?

Machine Performance

- By heuristic combinations of top-performing instance and semantic segmentations
  - How does this method perform?
  - How do the machine results compare to the human results that were presented before?

Datasets

Dataset	Instance and Semantic Segmentation outputs
Cityscapes	Generated from PSPNet and Mask R-CNN resp.
ADE20k	Output from the winners of 2017 places challenge
Mapillary Vistas	Output from the winners of LSUN'17 segmentation challenge

Results for Semantic and Instance segmentation are disjoint in these outputs.

Heuristic combination

How to combine?

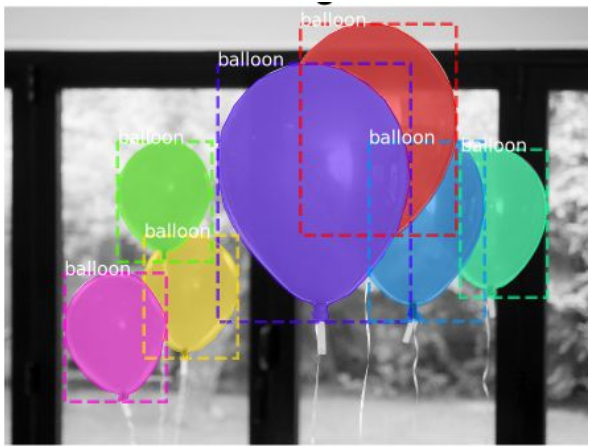
~~Panoptics Segments = Instance Segments + Semantic Segments of stuff~~

Why?

Heuristic combination

Instance segmentation allows overlapped segments.

But the proposed Panoptic segmentation idea doesn't allow this

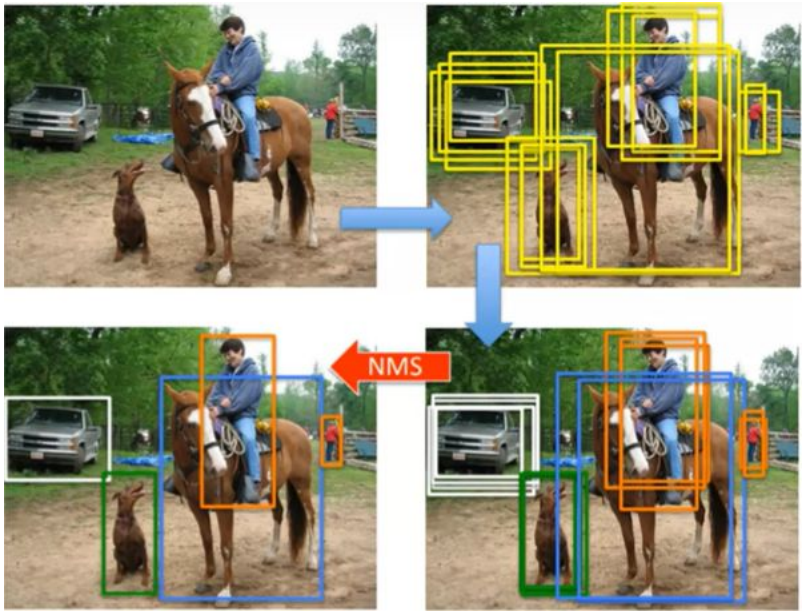


Panoptic segments = Non overlapping instance segments + Semantic Segments of stuff

How to create non overlapping instance segments? - NMS like procedure

Recap on NMS

- Sorts the bounding boxes based on confidence scores
- Eliminates bounding boxes with higher IoU than a threshold with the bounding box with highest confidence score



Figure_from:https://www.youtube.com/watch?v=L69fw3s63HU

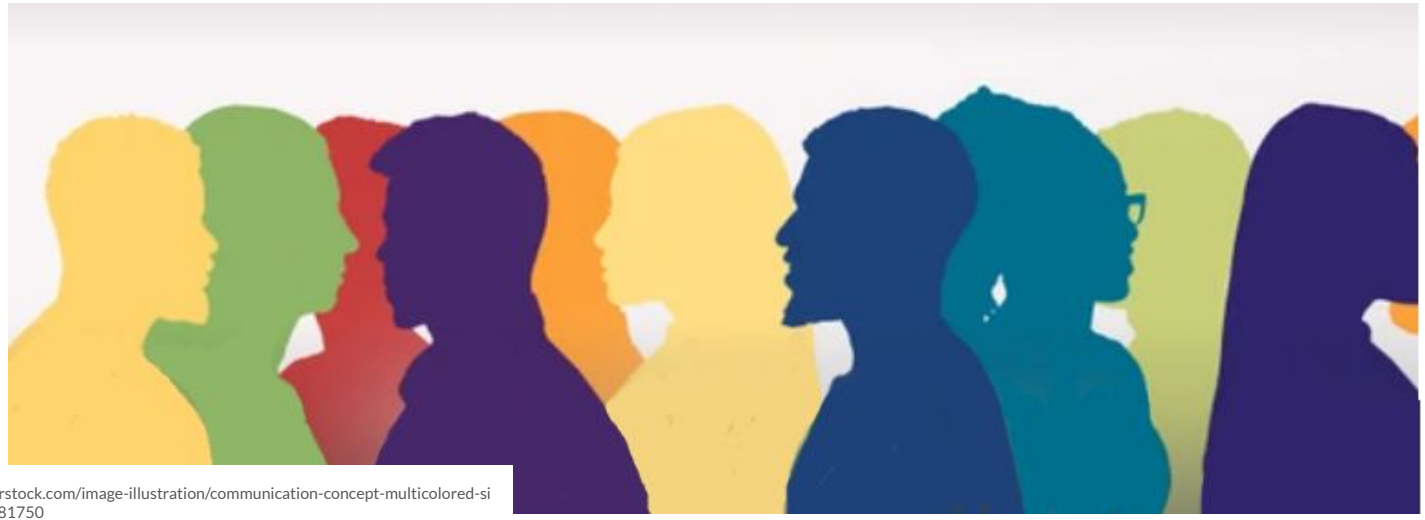
Heuristic combination

Step 1: Sort the predicted segments based on their confidence scores

Figure_from:https://www.shutterstock.com/image-illustration/communication-concept-multicolored-silhouettes-people-talking-1606081750

Heuristic combination

Step 2: For each instance, remove pixels which were assigned to a previous segment



Figure_from:https://www.shutterstock.com/image-illustration/communication-concept-multicolored-silhouettes-people-talking-1606081750

Heuristic combination

Step 3: If the area of an instance is less than a threshold, remove them

Figure_from:https://www.shutterstock.com/image-illustration/communication-concept-multicolored-silhouettes-people-talking-1606081750

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

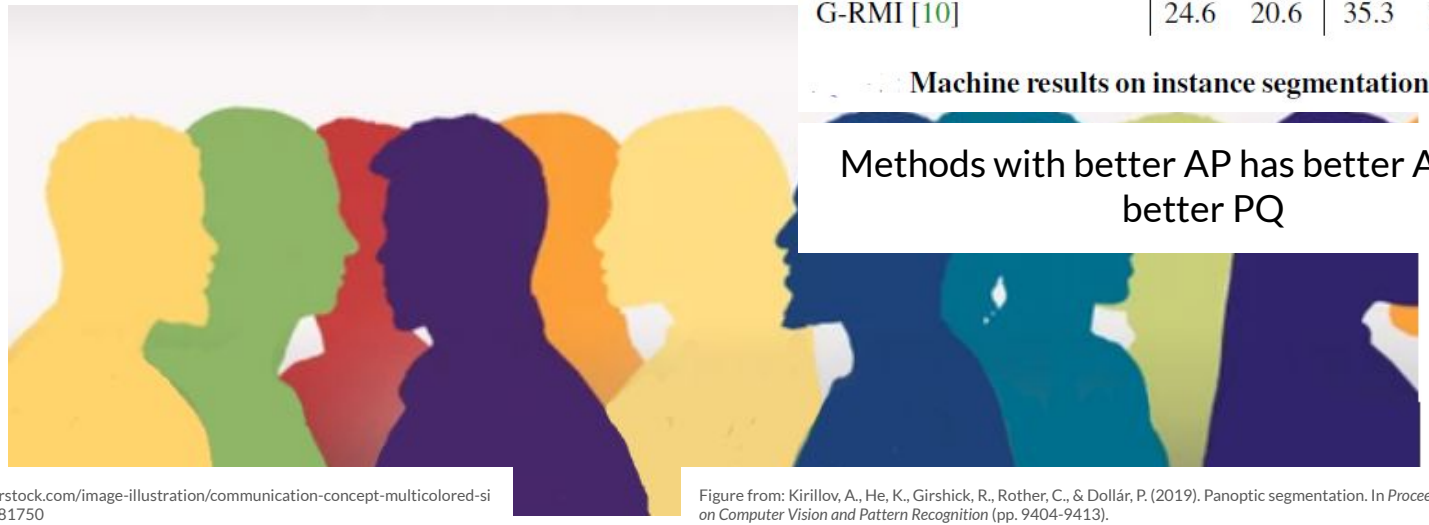
Heuristic combination

Step 3: If the area of an instance is less than a threshold, re

Cityscapes	AP	AP^{NO}	PQ^{Th}	SQ^{Th}	RQ^{Th}
Mask R-CNN+COCO [14]	**36.4**	**33.1**	**54.0**	**79.4**	**67.8**
Mask R-CNN [14]	31.5	28.0	49.6	78.7	63.0
ADE20k	AP	AP^{NO}	PQ^{Th}	SQ^{Th}	RQ^{Th}
Megvii [31]	**30.1**	**24.8**	**41.1**	**81.6**	**49.6**
G-RMI [10]	24.6	20.6	35.3	79.3	43.2

Machine results on instance segmentation

Methods with better AP has better AP^{NO} and better PQ



Figure_from:https://www.shutterstock.com/image-illustration/communication-concept-multicolored-silhouettes-people-talking-1606081750

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Heuristic combination

Step 4: Add the semantic classes. If stuff and thing masks coincide, preference is given to thing

Figure_from:https://www.shutterstock.com/image-illustration/communication-concept-multicolored-silhouettes-people-talking-1606081750

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Heuristic combination

Step 4: Add the semantic classes. If stuff and thing masks coincide, preference is given to thing

Cityscapes	IoU	PQ^{St}	SQ^{St}	RQ^{St}
PSPNet multi-scale [53]	**80.6**	**66.6**	**82.2**	**79.3**
PSPNet single-scale [53]	79.6	65.2	81.6	78.0
ADE20k	IoU	PQ^{St}	SQ^{St}	RQ^{St}
CASIA_IVA_JD [12]	**32.3**	**27.4**	**61.9**	**33.7**
G-RMI [11]	30.6	19.3	58.7	24.3

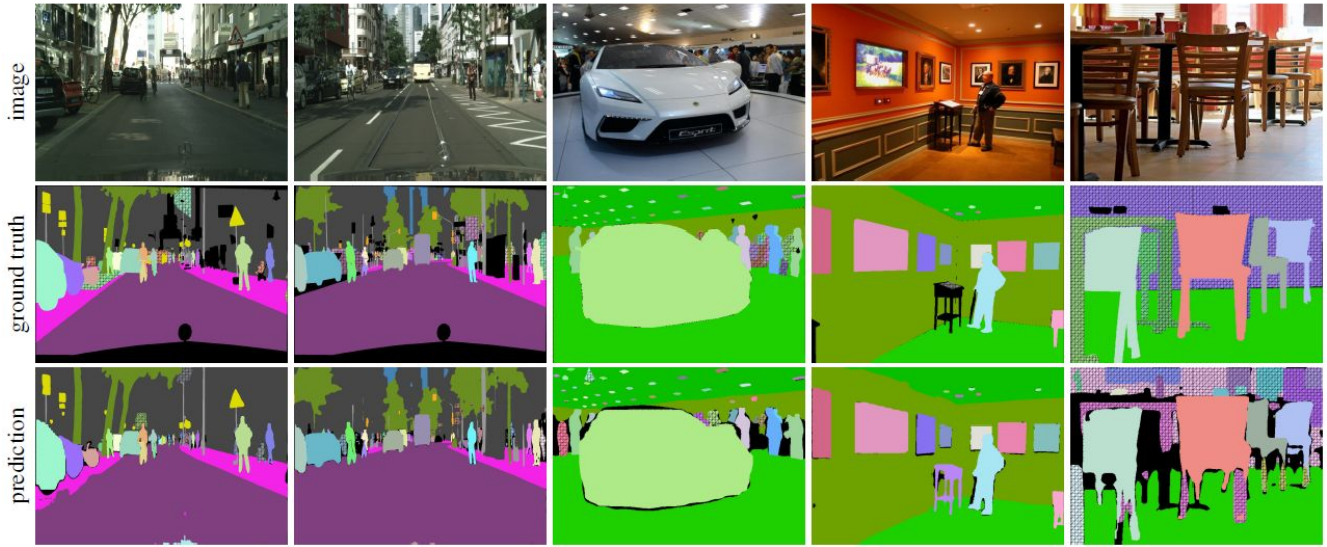
Machine results on semantic segmentation

Methods with better IoU has better PQ

Figure_from:https://www.shutterstock.com/image-illustration/communication-concept-multicolored-silhouettes-people-talking-1606081750

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

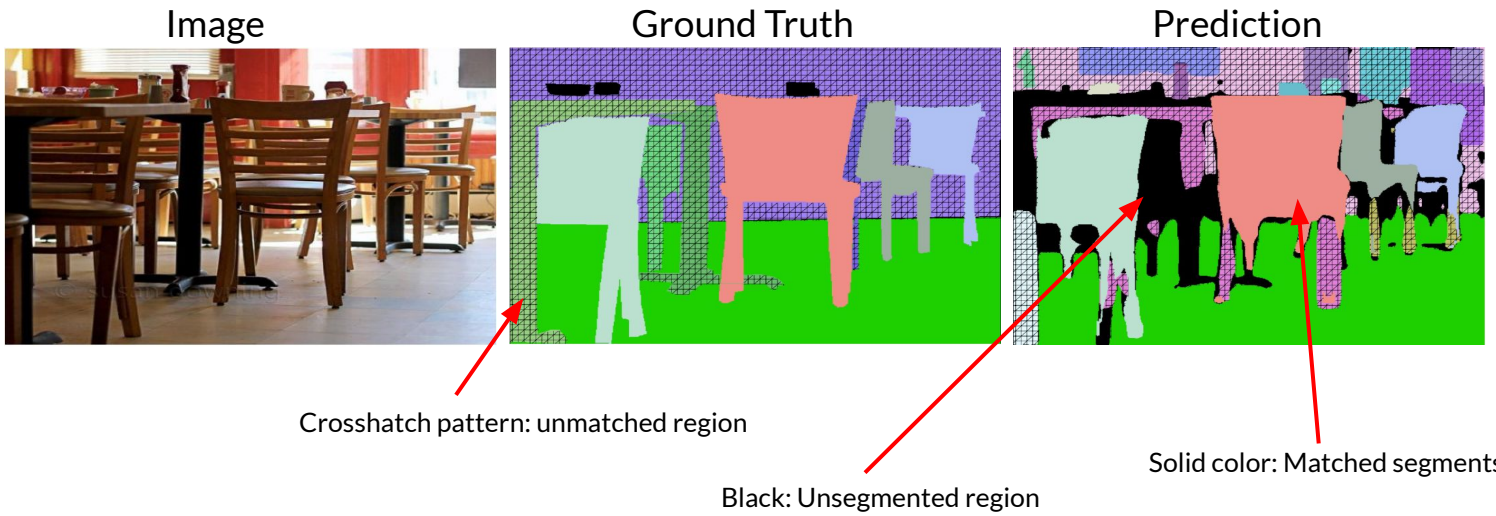
Segmentation Results



Predictions based on merged outputs of Instance and semantic segmentation tasks. Segments matched only if IoU > 0.5

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Segmentation Results



IoU > 0.5 makes sure that only one predicted segment matches with each ground truth segment

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Inferences

Cityscapes	PQ	PQ^{St}	PQ^{Th}
machine-separate	n/a	66.6	54.0
machine-panoptic	61.2	66.4	54.0
ADE20k	PQ	PQ^{St}	PQ^{Th}
machine-separate	n/a	27.4	41.1
machine-panoptic	35.6	24.5	41.1
Vistas	PQ	PQ^{St}	PQ^{Th}
machine-separate	n/a	43.7	35.7
machine-panoptic	38.3	41.8	35.7

Panoptic *vs.* independent predictions.

PQ of things are consistent but PQ for stuff is slightly low - Reason???

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Inferences

Cityscapes	PQ	SQ	RQ	PQ^{St}	PQ^{Th}
human	$69.6^{+2.5}_{-2.7}$	$84.1^{+0.8}_{-0.8}$	$82.0^{+2.7}_{-2.9}$	$71.2^{+2.3}_{-2.5}$	$67.4^{+4.6}_{-4.9}$
machine	61.2	80.9	74.4	66.4	54.0
ADE20k	PQ	SQ	RQ	PQ^{St}	PQ^{Th}
human	$67.6^{+2.0}_{-2.0}$	$85.7^{+0.6}_{-0.6}$	$78.6^{+2.1}_{-2.1}$	$71.0^{+3.7}_{-3.2}$	$66.4^{+2.3}_{-2.4}$
machine	35.6	74.4	43.2	24.5	41.1
Vistas	PQ	SQ	RQ	PQ^{St}	PQ^{Th}
human	$57.7^{+1.9}_{-2.0}$	$79.7^{+0.8}_{-0.7}$	$71.6^{+2.2}_{-2.3}$	$62.7^{+2.8}_{-2.8}$	$53.6^{+2.7}_{-2.8}$
machine	38.3	73.6	47.7	41.8	35.7

Human *vs.* machine performance

SQ is closer but human consistency is much higher in RQ

Figure from: Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9404-9413).

Future

Goals when the idea was introduced:

While the authors of the paper uses certain heuristics to produce PS outputs, in the future they are excited to see actual Panoptic Segmentation models

Thanks

Questions?