Panoptic Segmentation

Alexander Kirillov[1,2] Kaiming He[1] Ross Girshick[1] Carsten Rother[2] Piotr Dollár[1]

[1]Facebook AI Research (FAIR) [2]HCI/IWR, Heidelberg University, Germany

김 성 철

Contents

Introduction & Related Work

- **Introduction**
  - **Semantic Segmentation**
    - Stuff (such as grass, sky, road, ..)
    - Simply assign a class label to each pixel in an image (treat thing classes as stuff)
    - FCN으로 object instance를 나누기 힘듬
  - **Instance Segmentation**
    - Things (such as people, animals, tools, ..)
    - Detect each object and delineate it with a bbox or mask
    - Region-based method로 overlapping을 피하기 힘듬

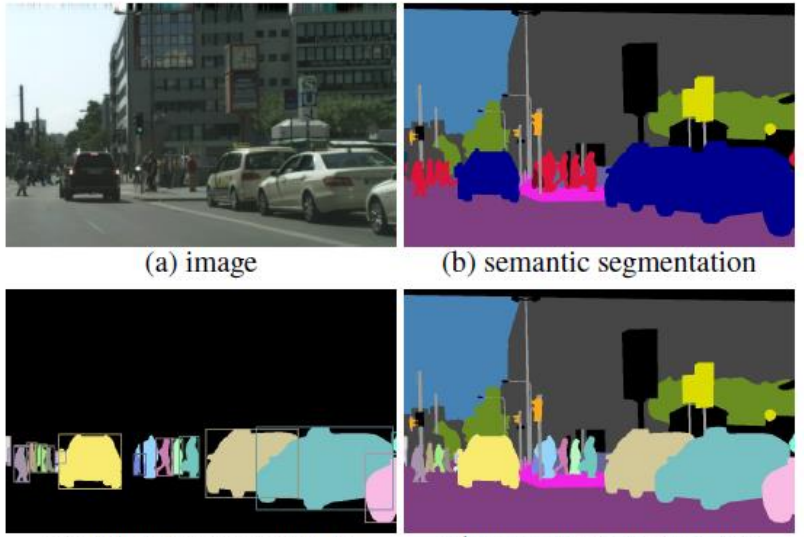

Figure 1: For a given (a) image, we show *ground truth* for: (b) semantic segmentation (per-pixel class labels), (c) instance segmentation (per-object mask and class label), and (d) the proposed *panoptic segmentation* task (per-pixel class+instance labels). The PS task: (1) encompasses both stuff and thing classes, (2) uses a simple but general format, and (3) introduces a uniform evaluation metric for all classes. Panoptic segmentation generalizes both semantic and instance segmentation and we expect the unified task will present novel challenges and enable innovative new methods.

Introduction & Related Work

- **Introduction**
  - **Panoptic Segmentation**
    - Encompass both stuff and thing classes

      → **Panoptic : including every thing visible in one view**
    - Use a simple but general output format

      → 각 pixel은 semantic label과 instance id를 할당받음
    - Introduce a uniform evaluation metric

      → Semantic, Instance 모두 적용 가능한 **Panoptic Quality**

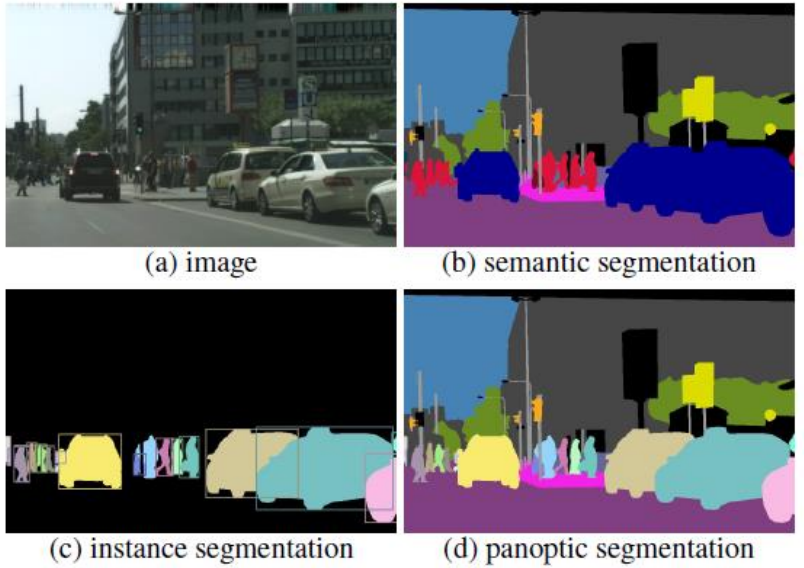

Figure 1: For a given (a) image, we show *ground truth* for: (b) semantic segmentation (per-pixel class labels), (c) instance segmentation (per-object mask and class label), and (d) the proposed *panoptic segmentation* task (per-pixel class+instance labels). The PS task: (1) encompasses both stuff and thing classes, (2) uses a simple but general format, and (3) introduces a uniform evaluation metric for all classes. Panoptic segmentation generalizes both semantic and instance segmentation and we expect the unified task will present novel challenges and enable innovative new methods.

Panoptic Segmentation

- **Format**
  - **Task format**
    - Semantic classes $\mathcal{L} \coloneqq \{0, \dots, \mathcal{L}-1\}$
    - Each pixel i is mapped to a pair $(\boldsymbol{l_i}, \boldsymbol{z_i}) \in \boldsymbol{\mathcal{L}} \times \mathbb{N}$
      - l_i : the semantic class of pixel i
      - z_i : its instance id. Group pixels of the same class into distinct segments
    - 애매하거나 out-of-class pixel들은 void label로 배정

      → 모든 pixel이 semantic label을 갖는 건 아님

Panoptic Segmentation

- **Format**
  - **Stuff and thing labels**
    - Semantic label set은 subset $\mathcal{L}^{St}$와 $\mathcal{L}^{Th}$로 구성
      - $\mathcal{L} = \mathcal{L}^{St} \cup \mathcal{L}^{Th}$, $\mathcal{L}^{St} \cap \mathcal{L}^{Th} = \emptyset$
    - Pixel이 $l_i \in \mathcal{L}^{St}$로 레이블 될 때, z_i는 관계 없음

      → 모든 stuff classes는 같은 instance를 갖기 때문 (*e.g.*, the same *sky*)
    - 모든 pixel이 같은 $(\boldsymbol{l_i}, \boldsymbol{z_i})$ 일 때 ($l_i \in \mathcal{L}^{Th}$), 같은 instance에 속함 (*e.g.*, the same *car*)
      - Single instance에 속하는 경우 같은 $(\boldsymbol{l_i}, \boldsymbol{z_i})$을 가짐

Panoptic Segmentation

- **Format**
  - **Relationship to semantic segmentation**
    - Panoptic segmentation은 semantic segmentation의 strict generalization
      - 각 pixel에 semantic label을 할당하는 것은 동일, thing class가 들어가는 순간 달라짐
  - **Relationship to instance segmentation**
    - Instance segmentation은 overlapping segment 발생

      → Panoptic segmentation은 각 pixel에 semantic label과 instance id 하나씩 할당

      → Overlapping 발생 X

Panoptic Segmentation

- **Metric**
  - 기존 방법들은 semantic or instance segmentation에 특화

    → stuff와 thing을 동시에 측정 불가
  - Stuff와 thing을 동시에 측정할 수 있는 metric 제시
    - **Completeness :** Treat stuff and thing classes in a uniform way
    - **Interpretability :** A metric with identifiable meaning
    - **Simplicity :** Simple to define and implement

→ Panoptic Quality (PQ)

Panoptic Segmentation

- **Metric**
  - **Segment Matching**
    - $IoU(p_i, g) = \frac{|p_i \cap g|}{|p_i \cup g|} \leq \frac{|p_i \cap g|}{|g|} \quad for\ i \in \{1, 2\}$
      - $p_1 \cap p_2 = \emptyset,\ |p_i \cup g| \geq |g|$
    - $IoU(p_1, g) + IoU(p_2, g) \leq \frac{|p_1 \cap g| + |p_2 \cap g|}{|g|} \leq 1$
      - $|p_1 \cap g| + |p_2 \cap g| \leq |\mathrm{g}|$
  - **Simple and Interpretable!**

Panoptic Segmentation

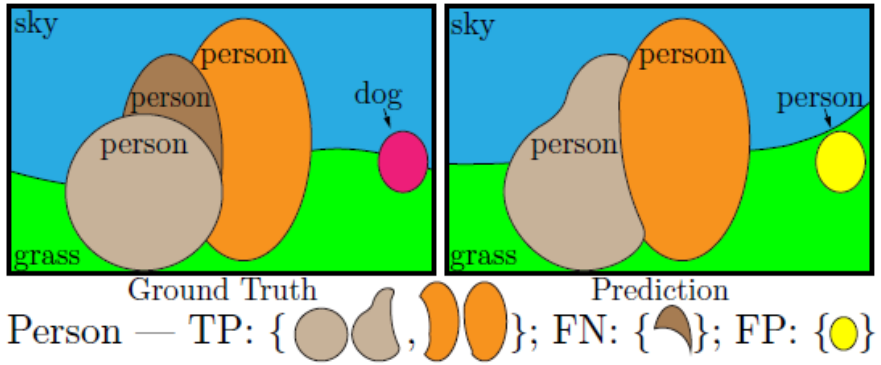

Figure 2: Toy illustration of ground truth and predicted panoptic segmentations of an image. Pairs of segments of the same color have IoU larger than 0.5 and are therefore matched. We show how the segments for the *person* class are partitioned into true positives TP, false negatives FN, and false positives FP.

- Metric
  - PQ Computation
    - Class별로 PQ 계산 후 평균 → **class imbalance에 insensitive하게 만듦**
    - $\boldsymbol{PQ} = \frac{\sum_{(p,g)\in TP} \boldsymbol{IoU}(\boldsymbol{p},\boldsymbol{g})}{|\boldsymbol{TP}| + \frac{1}{2}|\boldsymbol{FP}| + \frac{1}{2}|\boldsymbol{FN}|}$
      - $\frac{1}{|TP|}\sum_{(p,g)\in TP} IoU(p,g)$: the average IoU of matched segments
      - $\frac{1}{2}|FP| + \frac{1}{2}|FN|$: to penalize segments without matches
    - **All segments receive equal importance regardless of their area!**

Panoptic Segmentation

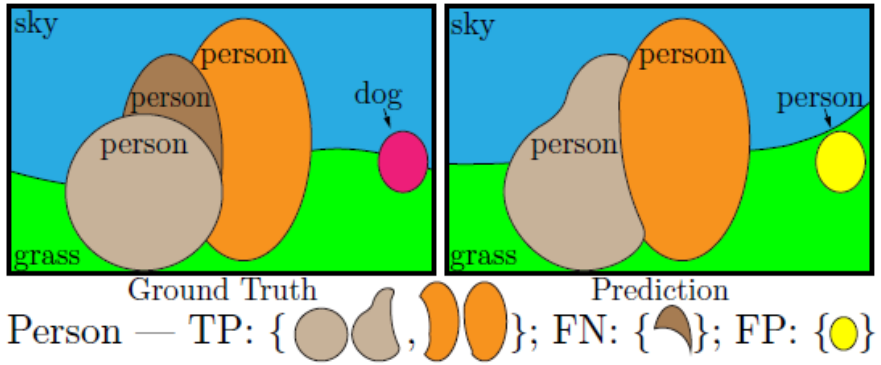

Figure 2: Toy illustration of ground truth and predicted panoptic segmentations of an image. Pairs of segments of the same color have IoU larger than 0.5 and are therefore matched. We show how the segments for the *person* class are partitioned into true positives TP, false negatives FN, and false positives FP.

- Metric
  - PQ Computation
    - $\boldsymbol{PQ} = \frac{\sum_{(p,g)\in TP} \boldsymbol{IoU(p,g)}}{|\boldsymbol{TP}| + \frac{1}{2}|\boldsymbol{FP}| + \frac{1}{2}|\boldsymbol{FN}|} = \frac{\sum_{(p,g)\in TP} \boldsymbol{IoU(p,g)}}{|\boldsymbol{TP}|} \times \frac{|\boldsymbol{TP}|}{|\boldsymbol{TP}| + \frac{1}{2}|\boldsymbol{FP}| + \frac{1}{2}|\boldsymbol{FN}|} (\boldsymbol{SQ} \times \boldsymbol{RQ})$
      - $\frac{\sum_{(p,g)\in TP} IoU(p,g)}{|TP|}$: segmentation quality (SQ) ← the average IoU of matched segments
      - $\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$: recognition quality (RQ) ← F_1 score과 흡사

Panoptic Segmentation

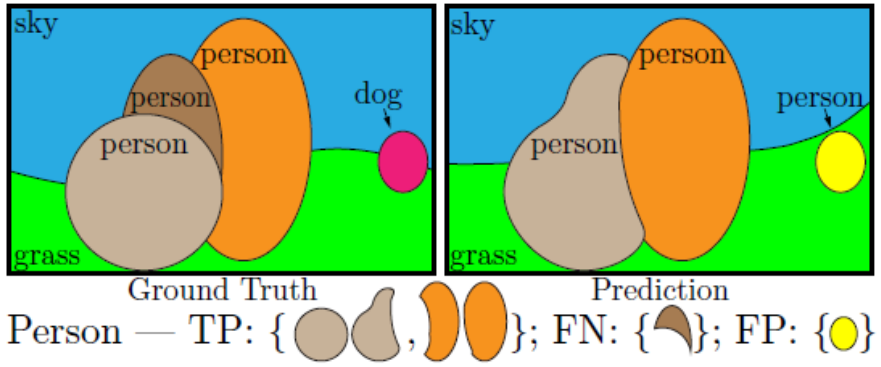

Figure 2: Toy illustration of ground truth and predicted panoptic segmentations of an image. Pairs of segments of the same color have IoU larger than 0.5 and are therefore matched. We show how the segments for the *person* class are partitioned into true positives TP, false negatives FN, and false positives FP.

- **Metric**
  - **PQ Computation**
    - Void labels in the ground truth
      - Out of class pixels / ambiguous or unknown pixels → 평가하지 않음
      - Ground truth에서 void labels인 pixel들은 제외하고 IoU 계산
      - Void pixel의 일부가 포함된 unmatched predicted segments는 제거 & FP로 처리X
    - Group labels
      - Cityscapes, COCO는 instance id 대신 group label을 사용
      - Matching 중에는 group region 사용 X
      - Matching 이후, matching threshold를 넘는 same class의 일부를 포함하는

Panoptic Segmentation

- **Metric**
  - **Comparison to Existing Metrics**
    - Semantic segmentation metrics
      - IoU → pixel output/labels 기반 계산 & object-level labels 무시
    - Instance segmentation metrics
      - Average Precision : Confidence score 필요 → semantic segmentation에 적합하지 않음
    - Panoptic quality
      - Treat all classes (stuff and things) in a uniform way
      - SQ, RQ로 decomposing 할 수 있지만, PQ가 semantic과 instance의 metric은 아님
        → 오히려 SQ와 RQ가 각각 semantic과 instance의 metric

Panoptic Segmentation

- **Datasets**
  - **Cityscapes**
    - 5000 images (2975 / 500 / 1525)
    - 19 classes among which 8 have instance-level
  - **ADE20k**
    - 25k images (20k / 2k / 3k)
    - 100 thing and 50 stuff classes in the 2017 Places Challenge
  - **Mapillary Vistas**
    - 25k street-view images (18k / 2k / 5k)
    - 28 stuff and 37 thing classes in ‘research edition’

Human Consistency Study

- **Human annotation**
  - 하나를 ground truth, 다른 하나를 prediction으로 설정 후 PQ 측정
- **Human consistency**

	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

Table 1: **Human consistency for stuff *vs.* things.** Panoptic, segmentation, and recognition quality (PQ, SQ, RQ) averaged over classes (PQ=SQ×RQ per class) are reported as percentages. Perhaps surprisingly, we find that human consistency on each dataset is relatively similar for both stuff and things.

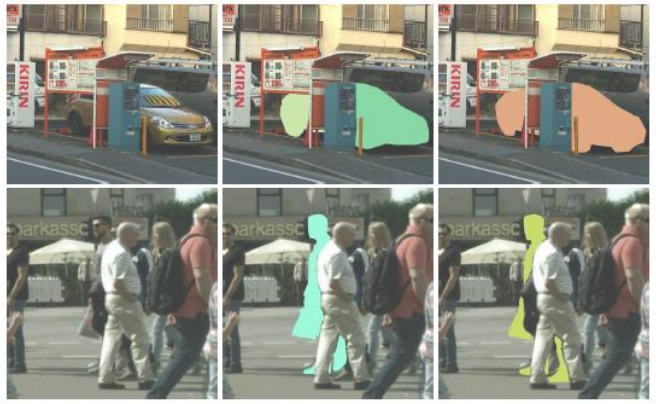

Figure 3: **Segmentation flaws.** Images are zoomed and cropped. Top row (Vistas image): both annotators identify the object as a car, however, one splits the car into two cars. Bottom row (Cityscapes image): the segmentation is genuinely ambiguous.

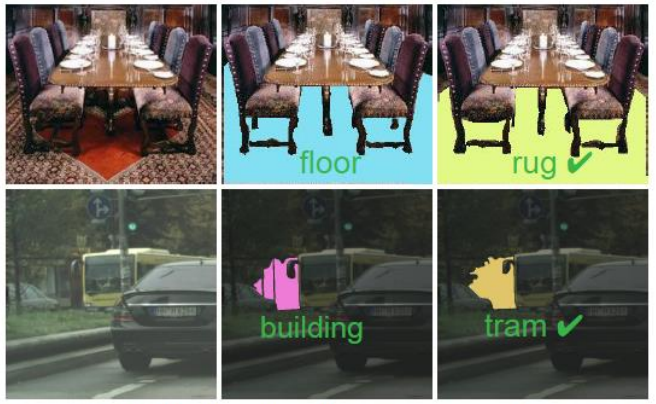

Figure 4: **Classification flaws.** Images are zoomed and cropped. Top row (ADE20k image): simple misclassification. Bottom row (Cityscapes image): the scene is extremely difficult, tram is the correct class for the segment. Many errors are difficult to resolve.

Human Consistency Study

- **Stuff *vs.* Things**
  - 모든 stuff and things classes가 고르게 분포
- **Small *vs.* large objects**
  - Small 25% | middle 50% | largest 25%

	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

Table 2: **Human consistency *vs.* scale**, for small (S), medium (M) and large (L) objects. Scale plays a large role in determining human consistency for panoptic segmentation. On large objects both SQ and RQ are above 80 on all datasets, while for small objects RQ drops precipitously. SQ for small objects is quite reasonable.

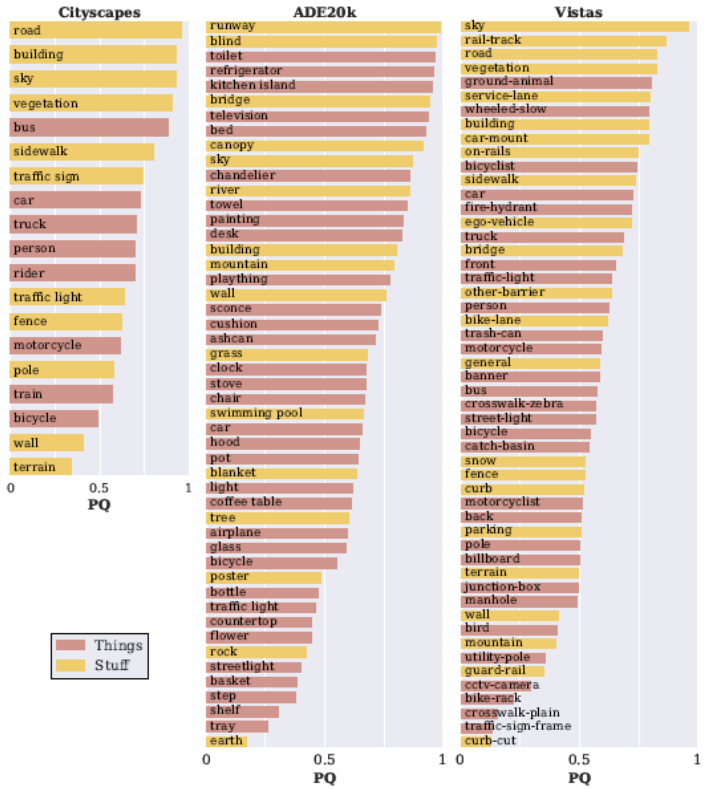

Figure 5: **Per-Class Human consistency, sorted by PQ**. Thing classes are shown in red, stuff classes in orange (for ADE20k every other class is shown, classes without matches in the dual-annotated tests sets are omitted). Things and stuff are distributed fairly evenly, implying PQ balances their performance.

Human Consistency Study

- IoU threshold

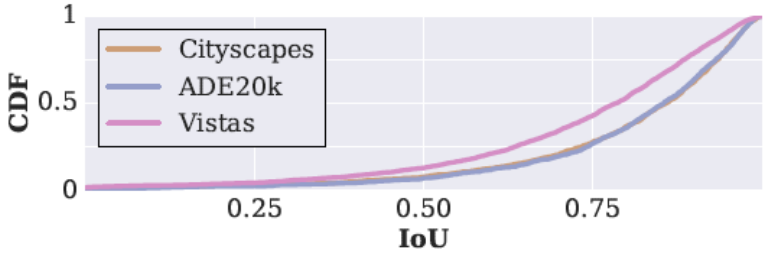

Figure 6: **Cumulative density functions of overlaps** for matched segments in three datasets when matches are computed by solving a maximum weighted bipartite matching problem [47]. After matching, less than 16% of matched objects have IoU below 0.5.

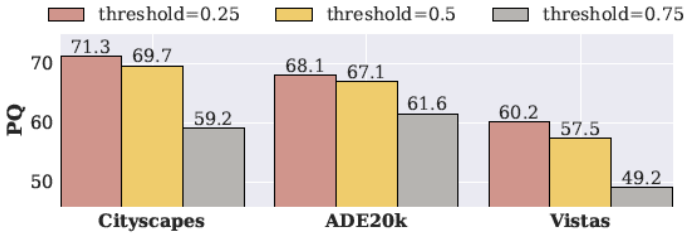

Figure 7: **Human consistency for different IoU thresholds.** The difference in PQ using a matching threshold of 0.25 *vs.* 0.5 is relatively small. For IoU of 0.25 matching is obtained by solving a maximum weighted bipartite matching problem. For a threshold greater than 0.5 the matching is unique and much easier to obtain.

- SQ $vs.$ RQ balance
  - $RQ_\alpha = \frac{|TP|}{|TP|+\alpha|FP|+\alpha|FN|}$ (default $\alpha = 0.5$)
  - default α 가 SQ와 RQ 사이에 균형을 맞춰 줌
  - 상황에 따라 변경해주면 될 듯

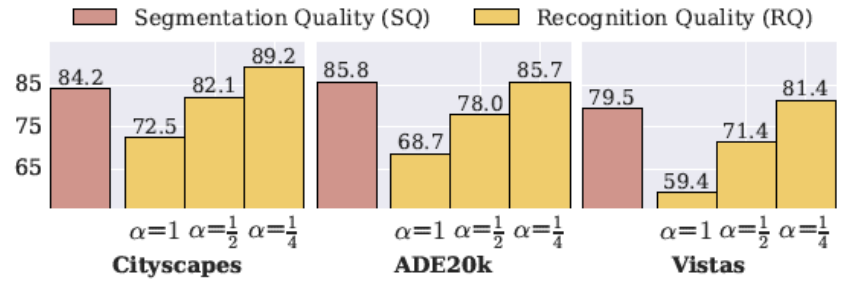

Figure 8: **SQ *vs.* RQ** for different α, see (3). Lowering α reduces the penalty of unmatched segments and thus increases the reported RQ (SQ is not affected). We use α of 0.5 throughout but by tuning α one can balance the influence of SQ and RQ in the final metric.

Machine Performance Baseline

- **세 가지 관점**
  - **How do heuristic combinations** of top-performing instance and semantic segmentation systems perform on panoptic segmentation?
  - **How does PQ compare to existing metrics** like AP and IoU?
  - **How do the machine results compare to the human results** that we presented previously?

Machine Performance Baseline

- **Algorithms and data**
  - **각 dataset에서 algorithm output을 얻음**
    - *Cityscapes* : masks generated by PSPNet and Mask R-CNN
    - *ADE20k* : the winners of both the semantic and instance seg. in the 2017 Places Challenge
    - *Vistas* : 주최측으로 부터 1k testset에 대한 semantic & instance seg. 결과를 제공받음

Machine Performance Baseline

- **Instance segmentation**
  - Overlapping segment를 얻음 → 해결
    - Confidence score 순으로 sorting

      → 낮은 score는 제거
    - 가장 confident한 것부터 시작해서 반복
      - **Remove pixels** which have been assigned to previous segments
      - Segment의 충분한 부분이 남아있으면 non-overlapping, 그렇지 않으면 전체 segment 제거

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

Table 3: **Machine results on instance segmentation** (stuff classes ignored). Non-overlapping predictions are obtained using the proposed heuristic. AP^{NO} is AP of the non-overlapping predictions. As expected, removing overlaps harms AP as detectors benefit from predicting multiple overlapping hypotheses. Methods with better AP also have better AP^{NO} and likewise improved PQ.

Machine Performance Baseline

- **Semantic segmentation**
  - No overlapping → 바로 PQ 계산 가능
    - Multi-scale : skip connection (?)

Cityscapes	IoU	PQ^{St}	SQ^{St}	RQ^{St}
PSPNet multi-scale [54]	**80.6**	**66.6**	**82.2**	**79.3**
PSPNet single-scale [54]	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

Table 4: **Machine results on semantic segmentation** (thing classes ignored). Methods with better mean IoU also show better PQ results. Note that G-RMI has quite low PQ. We found this is because it hallucinates many small patches of classes not present in an image. While this only slightly affects IoU which counts *pixel* errors it severely degrades PQ which counts *instance* errors.

Machine Performance Baseline

- **Panoptic segmentation**
  - Instance + Semantic for resolving any overlap between thing and stuff

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

Table 5: **Panoptic *vs.* independent predictions.** The 'machine-separate' rows show PQ of semantic and instance segmentation methods computed independently (see also Tables 3 and 4). For 'machine-panoptic', we merge the non-overlapping thing and stuff predictions obtained from state-of-the-art methods into a true panoptic segmentation of the image. Due to the merging heuristic used, PQ^{Th} stays the same while PQ^{St} is slightly degraded.

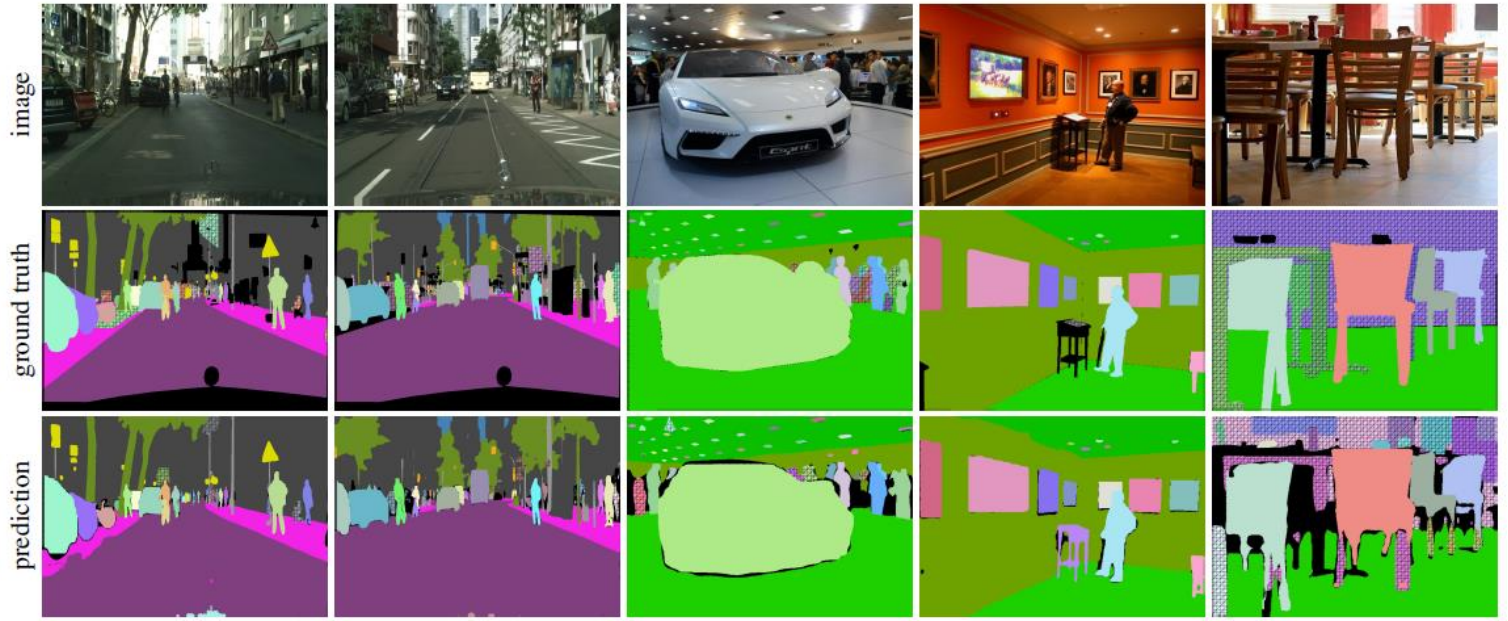

Figure 9: **Panoptic segmentation results** on Cityscapes (left two) and ADE20k (right three). Predictions are based on the merged outputs of state-of-the-art instance and semantic segmentation algorithms (see Tables 3 and 4). Colors for matched segments (IoU>0.5) match (crosshatch pattern indicates unmatched regions and black indicates unlabeled regions). Best viewed in color and with zoom.

Machine Performance Baseline

- Human $vs.$ machine panoptic segmentation

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

Table 6: **Human *vs.* machine performance.** On each of the considered datasets human consistency is much higher than machine performance (approximate comparison, see text for details). This is especially true for RQ, while SQ is closer. The gap is largest on ADE20k and smallest on Cityscapes. Note that as only a small set of human annotations is available, we use bootstrapping and show the the 5^{th} and 95^{th} percentiles error ranges for human results.

Future of Panoptic Segmentation

- **Stuff와 thing을 동시에 설명하는 end-to-end model**
  - Instance segmentation에서 non-overlapping을 위한 실험들이 진행됨
- **Some form of higher-level 'reasoning' may be beneficial**
  - Extend learnable NMS to PS

Future of Panoptic Segmentation

- **Stuff와 thing을 동시에 설명하는 end-to-end model**
  - Instance segmentation에서 non-overlapping을 위한 실험들이 진행됨
- **Some form of higher-level 'reasoning' may be beneficial**
  - Extend learnable NMS to PS

UPSNet (in CVPR 2019)

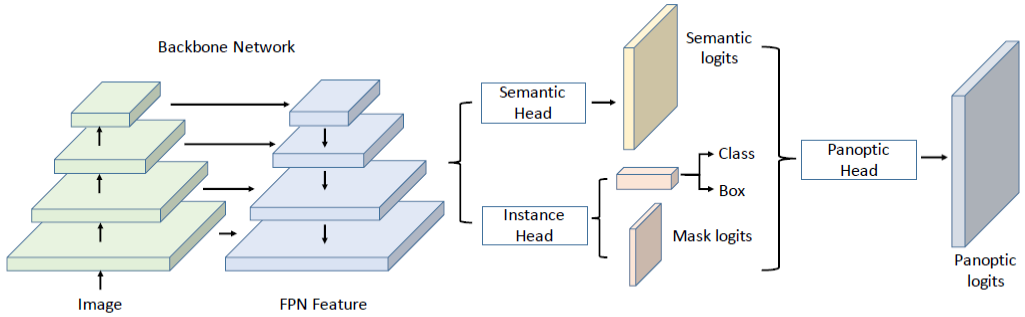

Figure 1: Overall architecture of our UPSNet.

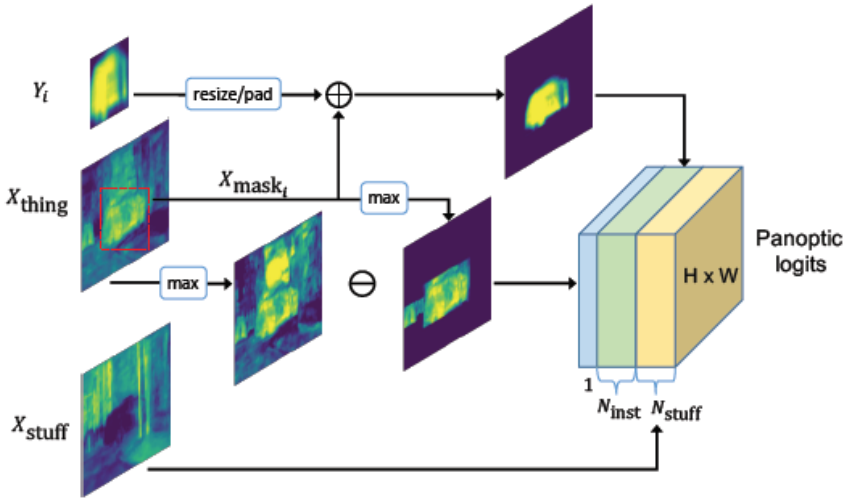

Figure 3: Architecture of our panoptic segmentation head.

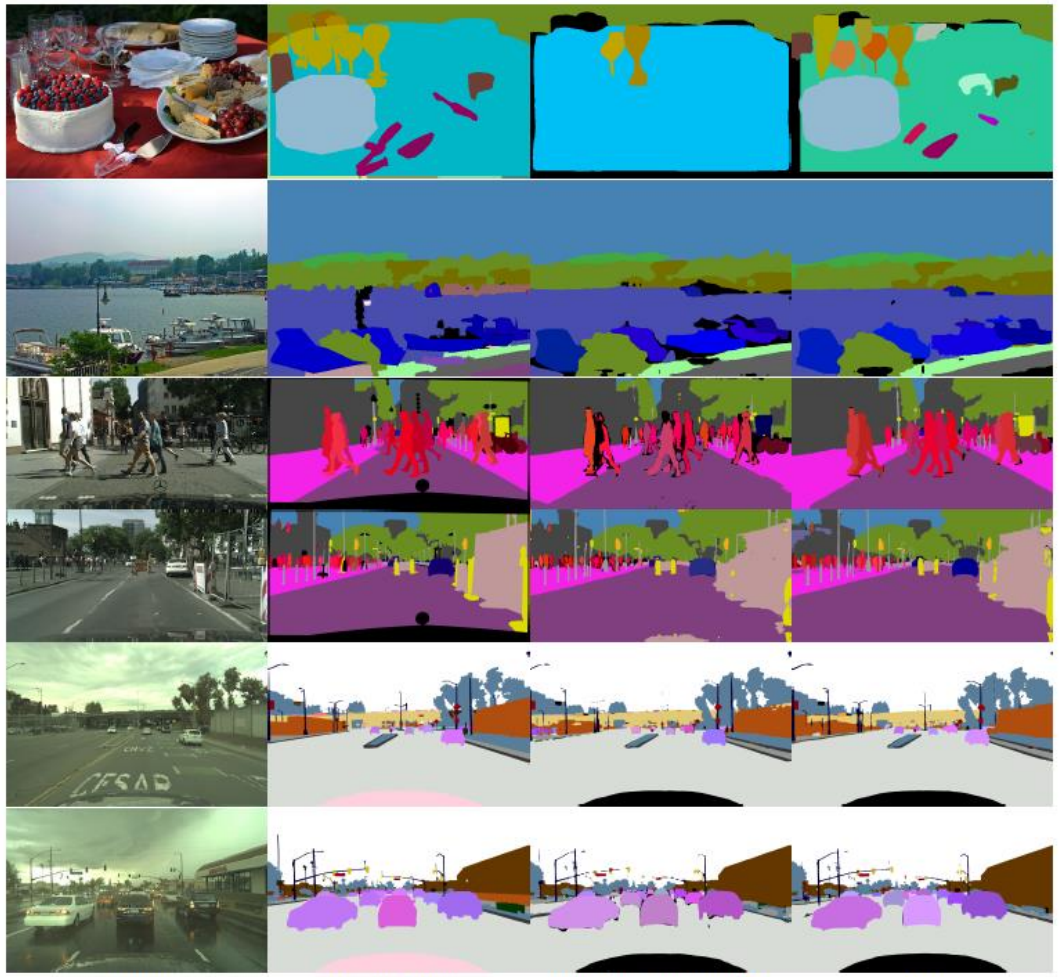

Image Ground truth Combined method [17] Ours

Figure 4: Visual examples of panoptic segmentation. 1-st and 2-nd rows are from COCO. 3-rd and 4-th rows are from Cityscapes. 5-th and 6-th rows are from our internal dataset.

감 사 합 니 다