

Improving Panoptic Segmentation at All Scales

Supplementary Material

Lorenzo Porzi, Samuel Rota Bulò, Peter Kotschieder
Facebook

{porzi, rotabulo, pkotschieder}@fb.com

Abstract

This document provides the following additional contributions to our CVPR2021 submission:

- In Sec. **A**, we provide proof of results related to the paper and the optimization algorithm for our new loss.
- In Sec. **B** we provide pseudo-algorithms for solving the optimization problems described in the main paper.
- In Sec. **C** we discuss the panoptic evaluation metrics used in the main paper.
- In Sec. **D** we provide a detailed breakdown of the training hyper-parameters used in our experiments.
- In Sec. **E** we provide additional ablation results related to ISUS.
- In Sec. **F** we show qualitative results for all datasets we consider in our experimental evaluation.

A. Proof of Results

Let $\omega_0 = b_1 - a_1$ and let $\xi(\omega)$ be the objective of (O_1) with first-order derivative

$$\xi'(\omega) = \frac{1}{2} \ell'_\beta \left(\frac{\omega - \hat{\omega}}{2} \right) + \frac{1}{\omega} \ell'_\beta (\log(\omega) - \log(\omega_P)). \quad (1)$$

The first-order derivative of the Huber loss is given by $\ell'_\beta(x) = \max(\min(\beta^{-1}x, 1), -1)$. We assume that $\omega_P > 0$ and $\omega_0 > 0$.

Proposition 1. *If a strictly feasible local solution (δ^*, ω^*) of (O_2) exists, then $\delta^* = \delta_P$ and $\omega^* = \omega_P$.*

Proof. Let $\delta_\lambda = \lambda \delta_P + (1 - \lambda) \delta^*$ and $\omega_\lambda = \lambda \omega_P + (1 - \lambda) \omega^*$. By contradiction, assume that a strictly feasible local solution (δ^*, ω^*) exists such that $(\delta^*, \omega^*) \neq (\delta_P, \omega_P)$. Then,

we expect $\frac{d}{d\lambda} \varphi(\delta_\lambda, \omega_\lambda) \Big|_{\lambda=0} = 0$, where $\varphi(\delta, \omega)$ denotes the objective of (O_2) . However

$$\begin{aligned} & \frac{d}{d\lambda} \varphi(\delta_\lambda, \omega_\lambda) \Big|_{\lambda=0} \\ &= \frac{d}{d\lambda} \ell_\beta(\delta_\lambda - \delta_P) \Big|_{\lambda=0} + \frac{d}{d\lambda} \ell_\beta(\log(\omega_\lambda) - \log(\omega_P)) \Big|_{\lambda=0} \\ &= (\delta_P - \delta^*) \ell'_\beta(\delta^* - \delta_P) + \frac{\omega_P - \omega^*}{\omega^*} \ell'_\beta(\log(\omega^*) - \log(\omega_P)) \end{aligned}$$

is negative because $\ell'_\beta(x) < 0$ if $x < 0$ and $\ell'_\beta(x) > 0$ if $x > 0$, the logarithm is an ordering-preserving mapping and $\omega^* > 0$. This yields a contradiction thus proving the result. \square

Proposition 2. $\xi'(\omega) < 0$ for all $0 < \omega < \min\{\hat{\omega}, \omega_P\}$ and $\xi'(\omega) > 0$ for all $\omega > \max\{\hat{\omega}, \omega_P\}$.

Proof. For $0 < \omega < \hat{\omega}$ we have that $\ell'_\beta\left(\frac{\omega - \hat{\omega}}{2}\right) < 0$ and for $0 < \omega < \omega_P$ we have that $\ell'_\beta(\log(\omega) - \log(\omega_P)) < 0$. Accordingly, $\xi'(\omega) < 0$ for $0 < \omega < \min\{\hat{\omega}, \omega_P\}$.

Similarly, for $\omega > \hat{\omega}$, we have that $\ell'_\beta\left(\frac{\omega - \hat{\omega}}{2}\right) > 0$ and for $\omega > \omega_P$ we have that $\ell'_\beta(\log(\omega) - \log(\omega_P)) > 0$. Accordingly, $\xi'(\omega) > 0$ for $\omega > \max\{\hat{\omega}, \omega_P\}$. \square

Proposition 3. *If $\max\{\hat{\omega}, \omega_P\} \leq \omega_0$ then ω_0 is the solution to (O_1) .*

Proof. By Prop. 2, $\xi'(\omega) > 0$ for all $\omega > \max\{\hat{\omega}, \omega_P\}$. Accordingly, the same holds true for all $\omega \geq \omega_0$, which implies that $\xi(\omega_0)$ yields the lowest feasible objective value. \square

Proposition 4. *A solution to (O_1) exists in $[\max\{\omega_0, \min\{\hat{\omega}, \omega_P\}\}, \max\{\hat{\omega}, \omega_P\}]$ if $\omega_0 \leq \max\{\hat{\omega}, \omega_P\}$.*

Proof. A feasible solution $\omega < \max\{\omega_0, \min\{\hat{\omega}, \omega_P\}\}$ exists only if $\omega_0 \leq \omega < \min\{\hat{\omega}, \omega_P\}$. If this is the case, $\xi'(\omega) < 0$ holds in the latter interval by Prop. 2. Accordingly, for $\omega \leq \min\{\hat{\omega}, \omega_P\}$ the best objective is attained at $\min\{\hat{\omega}, \omega_P\}$. Similarly by Prop. 2, $\xi'(\omega) > 0$ if $\omega > \max\{\hat{\omega}, \omega_P\}$ and, therefore, for $\omega \geq \max\{\hat{\omega}, \omega_P\}$ the

best objective is attained at $\max\{\hat{\omega}, \omega_p\}$. Hence, a solution to (O_1) exists in the required interval. \square

Proposition 5. *If $\max\{\omega_0, \hat{\omega}\} < \omega_p$ then a solution to (O_1) exists in $[\max\{\omega_0, \hat{\omega}\}, \omega_p]$ and there ξ' is strictly increasing.*

Proof. For all $\hat{\omega} \leq \omega < \omega'$ we have that $0 \leq \ell'_\beta(\frac{\omega-\hat{\omega}}{2}) \leq \ell'_\beta(\frac{\omega'-\hat{\omega}}{2})$. Moreover, for all $0 < \omega < \omega' \leq \omega_p$, both $\ell'_\beta(\log(\omega) - \log(\omega_p)) \leq \ell'_\beta(\log(\omega') - \log(\omega_p)) \leq 0$ and $\frac{1}{\omega} > \frac{1}{\omega'} > 0$ hold, which imply $\frac{1}{\omega} \ell'_\beta(\log(\omega) - \log(\omega_p)) < \frac{1}{\omega'} \ell'_\beta(\log(\omega') - \log(\omega_p)) \leq 0$. It follows that $\xi'(\omega) < \xi'(\omega')$ holds in the required interval. \square

Proposition 6. $\eta(\omega) = \frac{1}{\omega} \ell'_\beta(\log(\omega) - \log(\omega_p))$ is

- strictly increasing in $(0, e^{\min\{\beta, 1\}}\omega_p]$, and
- strictly decreasing for $\omega \geq e\omega_p$.

Proof. $\eta(\omega)$ is strictly increasing for $0 < \omega < e^{-\beta}\omega_p$ because in this case $\eta(\omega) = -\frac{1}{\omega}$ and strictly decreasing for $\omega > e\omega_p$ because in this case $\eta(\omega) = \frac{1}{\omega}$. For $e^{-\beta}\omega_p \leq \omega \leq e\omega_p$ we have $\eta(\omega) = \frac{1}{\omega\beta}(\log(\omega) - \log(\omega_p))$ and

$$\eta'(\omega) = \frac{1}{\omega^2\beta} [1 - \log(\omega) + \log(\omega_p)].$$

Since $\eta'(e^{-\beta}\omega_p) > 0$, $\eta'(e\omega_p) < 0$ and $\eta'(\omega) = 0$ only at $\omega = e\omega_p$, it follows by continuity of η' that $\eta'(\omega) > 0$ for $\omega < e\omega_p$ and $\eta'(\omega) < 0$ for $\omega > e\omega_p$. Accordingly, $\eta(\omega)$ is strictly increasing in $[e^{-\beta}\omega_p, e^{\min\{\beta, 1\}}\omega_p]$. Since the same holds for $0 < \omega < e^{-\beta}\omega_p$ as shown before, by continuity of η , we conclude that $\eta(\omega)$ is strictly increasing along the whole interval $(0, e^{\min\{\beta, 1\}}\omega_p]$. Similarly, we have that $\eta(\omega)$ is strictly decreasing in $[e\omega_p, e\beta\omega_p]$ and for $\omega > e\beta\omega_p$ as shown before. Hence, by continuity of η , we conclude that $\eta(\omega)$ is strictly decreasing for $\omega \geq e\omega_p$. \square

Proposition 7. *If $\max\{\omega_0, \omega_p\} < \hat{\omega}$ then a solution to (O_1) is either ω_0 or $\hat{\omega}$, or it belongs to one of the following intervals:*

- (i) $J_1 = [\max\{\omega_0, \omega_p\}, \min\{e^{\min\{\beta, 1\}}\omega_p, \hat{\omega}\}]$,
- (ii) $J_2 = [\max\{\omega_0, 2\sqrt{\beta}, \hat{\omega} - 2\beta, e\beta\omega_p\}, \hat{\omega}]$,
- (iii) $J_3 = [\max\{\omega_0, \hat{\omega} - 2\beta, e\omega_p\}, \min\{e\beta\omega_p, \hat{\omega}\}]$ if $\hat{\omega} \leq 4\sqrt{2}$
- (iv) $J_4 = [\max\{\omega_0, \hat{\omega} - 2\beta, e\omega_p\}, \min\{e\beta\omega_p, \hat{\omega}, \nu_1\}]$ if $\hat{\omega} > 4\sqrt{2}$,
- (v) $J_5 = [\max\{\omega_0, \hat{\omega} - 2\beta, e\omega_p, \nu_2\}, \min\{e\beta\omega_p, \hat{\omega}\}]$ if $\hat{\omega} > 4\sqrt{2}$,

where $\nu_{1,2} = \frac{\hat{\omega}}{4} \left(1 \pm \sqrt{1 - \frac{32}{\hat{\omega}^2}}\right)$.

Moreover, ξ' is strictly increasing in (i)-(ii) and $\sigma(\omega) = \omega\xi'(\omega)$ is strictly increasing in (iii) – (v).

Proof. By Prop. 4 a solution to (O_1) exists in $I = [\max\{\omega_0, \omega_p\}, \hat{\omega}]$. We partition I into sections where ξ' or σ are either strictly increasing or strictly decreasing. We work by cases:

- J_1 . In this interval $\ell'_\beta(\frac{\omega-\hat{\omega}}{2})$ is increasing in ω and η is strictly increasing by Prop. 6. Hence, $\xi'(\omega) = \frac{1}{2}\ell'_\beta(\frac{\omega-\hat{\omega}}{2}) + \eta(\omega)$ is strictly increasing as well.
- $[\max\{\omega_0, e\omega_p\}, \hat{\omega} - 2\beta]$. In this interval $\ell'_\beta(\frac{\omega-\hat{\omega}}{2})$ is constant and η is strictly decreasing by Prop. 6. Hence, ξ' is strictly decreasing as well.
- $[\max\{\omega_0, \hat{\omega} - 2\beta, e\beta\omega_p\}, \hat{\omega}]$. In this interval, $\xi'(\omega) = \frac{1}{4\beta}(\omega - \hat{\omega}) + \frac{1}{\omega}$ and $\xi''(\omega) = 0$ holds only in the feasible point $\omega = 2\sqrt{\beta}$, while $\xi''(\omega) < 0$ for $0 < \omega < 2\sqrt{\beta}$ and $\xi''(\omega) > 0$ for $\omega > 2\sqrt{\beta}$. Accordingly $\xi'(\omega)$ is strictly decreasing in the interval $[\max\{\omega_0, \hat{\omega} - 2\beta, e\beta\omega_p\}, \min\{2\sqrt{\beta}, \hat{\omega}\}]$ and strictly increasing in the interval J_2 .
- J_3 . In this interval, $\xi'(\omega) = \frac{1}{4\beta}(\omega - \hat{\omega}) + \frac{1}{\omega\beta}(\log(\omega) - \log(\omega_p))$ and by setting $\sigma'(\omega) = 0$ we find at most two solutions, namely $\nu_{1,2}$, which are distinct and real for $\hat{\omega} > 4\sqrt{2}$. Both solutions might potentially belong to the interval under consideration. The sign of $\sigma'(\omega)$ is negative for $\nu_1 < \omega < \nu_2$ and positive for $\omega < \nu_1$ and $\omega > \nu_2$. Accordingly $\sigma(\omega)$ is strictly increasing in the intervals J_4 and J_5 , while it is strictly decreasing in the interval $[\max\{\omega_0, \hat{\omega} - 2\beta, e\omega_p, \nu_1\}, \min\{e\beta\omega_p, \hat{\omega}, \nu_2\}]$. If $\hat{\omega} \leq 4\sqrt{2}$ then $\sigma'(\omega) \geq 0$ in the whole interval J_3 , with equality only if $\hat{\omega} = 4\sqrt{2}$ and $\omega = \nu_1 = \nu_2$. Accordingly, if $\hat{\omega} \leq 4\sqrt{2}$ we have that $\sigma(\omega)$ is strictly increasing in J_3 .

Since $\sigma(\omega)$ and $\xi'(\omega)$ share the same sign, given an interval J where ξ' or σ is strictly decreasing, we have one of the following cases: a) ξ' is strictly positive, b) ξ' is strictly negative or c) ξ' transitions once from a positive to a negative sign. In all three cases, a solution to (O_1) cannot exist in the interior of J but can be at most at one endpoint of J . For the same reason, no solution can be at the junction of two intervals where ξ' or σ are strictly decreasing. Hence, the endpoint has to be either an endpoint of I or be in common with an interval where either ξ' or σ are strictly increasing, which proves the result. \square

B. Optimization Algorithms

In this section we provide the optimization algorithms used to solve (O_2) and (O_1) , which exploit the theoretical results given in Sec. A.

Algorithm 1 Solves the optimization problem (O_1) .

```

1: function SOLVE_ $O_1(\omega_p, \hat{\omega}, a_1, b_1)$ 
2:    $\omega_0 = b_1 - a_1$ 
3:    $S = \{\omega_0\}$   $\triangleright$  Used to collect potential solutions
4:   if  $\max\{\omega_0, \hat{\omega}\} < \omega_p$  then  $\triangleright$  Prop. 5
5:     return FIND_MIN( $[\max\{\omega_0, \hat{\omega}\}, \omega_p], \xi'$ )
6:   else if  $\max\{\omega_0, \omega_p\} < \hat{\omega}$  then  $\triangleright$  Prop. 7
7:      $S = S \cup \{\hat{\omega}, \text{FIND\_MIN}(J_1, \xi'), \text{FIND\_MIN}(J_2, \xi')\}$ 
8:     if  $\hat{\omega} \leq 4\sqrt{2}$  then
9:        $S = S \cup \{\text{FIND\_MIN}(J_3, \sigma)\}$ 
10:    else
11:       $S = S \cup \{\text{FIND\_MIN}(J_4, \sigma), \text{FIND\_MIN}(J_5, \sigma)\}$ 
12:    end if
13:  end if
14:  return  $\arg \min_{\omega \in S} \xi(\omega)$ 
15: end function

```

Algorithm 2 Solves the optimization problem (O_2) .

```

1: function SOLVE_ $O_2(\delta_p, \omega_p, a_2, b_2)$ 
2:    $\hat{\omega}_1 = 2(\delta_p - a_2)$ 
3:    $\hat{\omega}_2 = 2(b_2 - \delta_p)$ 
4:   if  $\omega_p \geq \max\{\hat{\omega}_1, \hat{\omega}_2\}$  then
5:     return  $(\delta_p, \omega_p)$ 
6:   end if
7:    $\omega_1 = \text{SOLVE\_}O_1(\omega_p, \hat{\omega}_1, a_2, b_2)$ 
8:    $\omega_2 = \text{SOLVE\_}O_1(\omega_p, \hat{\omega}_2, a_2, b_2)$ 
9:   if  $\xi(\omega_1) \leq \xi(\omega_2)$  then
10:    return  $(a_2 + \frac{\omega_1}{2}, \omega_1)$ 
11:   else
12:    return  $(b_2 - \frac{\omega_2}{2}, \omega_2)$ 
13:   end if
14: end function

```

Algorithm for (O_2) . Alg. 2 provides a solution to the optimization problem (O_2) . The idea of the algorithm is sketched also in Sec. 2.4 of the main paper. The global, unconstrained solution to (O_2) is attained at (δ_p, ω_p) . Accordingly, if this solution is feasible it is also the global solution to the constrained version of the problem (line 5). If it's not feasible, we have two options, the solution is in the interior of the feasible set, or at the boundary. However, by Prop. 1 the former case is not possible, because the only solution would be the one we excluded already, namely (δ_p, ω_p) . Hence, the solution has to lie at the boundary of the feasible set. Since we have only two constraints, we can apply a brute force approach, and explore two cases where we assume that the solution activates the first constraint (line 7) or the second one (line 8). In both cases, we boil down to solving an instance of the optimization problem (O_1) where $\hat{\omega} = 2(\delta_p - a_2)$ and $\hat{\omega} = 2(b_2 - \delta_p)$, respectively. The solution to each of those problems, denoted by ω_1 and ω_2 in the

Algorithm 3 Finds the minimum of an objective ξ in a given interval $[u, v]$ by leveraging an increasing, continuous function φ , whose sign agrees with the sign of ξ' .

```

1: function FIND_MIN( $[u, v], \varphi$ )
2:   if  $\varphi(u) \geq 0$  then
3:     return  $u$ 
4:   else if  $\varphi(v) \leq 0$  then
5:     return  $v$ 
6:   else
7:      $m = \frac{u+v}{2}$ 
8:     if  $v - u < \epsilon$  then  $\triangleright \epsilon$  is a tolerance
9:       return  $m$ 
10:    else if  $\varphi(m) \geq 0$  then
11:      return FIND_MIN( $[u, m], \varphi$ )
12:    else
13:      return FIND_MIN( $[m, v], \varphi$ )
14:    end if
15:  end if
16: end function

```

algorithm, is given by applying Alg. 1, which is discussed later. Among those two solutions, we retain the one minimizing the objective of (O_1) , where the objective is denoted by ξ in the algorithm. If ω_1 is the best one then the solution to (O_2) is given by (δ_1, ω_1) in line 10, where $\delta_1 = a_2 + \frac{\omega_1}{2}$ is obtained by substituting ω_1 in the first constraint. Otherwise, the solution is given by (δ_2, ω_2) in line 12, where $\delta_2 = b_2 - \frac{\omega_2}{2}$ is obtained similarly from the second constraint.

Algorithm for (O_1) . Alg. 1 provides a solution to the optimization problem (O_1) . According to Prop. 3, if $\max\{\hat{\omega}, \omega_p\} \leq \omega_0$ then ω_0 is the solution. Indeed, in this case we return ω_0 in line 14 since it is the only element of S . If condition in line 4 is hit, then by Prop. 5 we can search a solution in the interval $[\max\{\omega_0, \hat{\omega}\}, \omega_p]$ by leveraging the monotonicity of ξ' . We do so by exploiting Alg. 3 in line 5, which will be discussed later. If condition in line 6 is hit instead, according to Prop. 7, we need to search for the best solution within the intervals J_i with $i \in \{1, \dots, 5\}$ eventually satisfying the given conditions. Moreover, we need to include in the pool also ω_0 and $\hat{\omega}$. The search over each interval J_i is performed via Alg. 3, by leveraging the monotonicity of ξ' or σ . All potential solutions are collected into S and the best one in terms of the objective is retained in line 14.

Alg. 3. Finds the minimum of an objective ξ in a given interval $[u, v]$ by leveraging an increasing, continuous function φ , whose sign agrees with the sign of ξ' . This can be done by searching the element in $[u, v]$ that is the closest one to a zero of φ . Since the function is increasing, if $\varphi(u)$

is non-negative then the closest element to a zero is u , while if $\varphi(v)$ is non-positive then the closest element is v . Otherwise, we perform a dichotomic search on the half-interval where we have discording signs of φ at the extremes until we reach the zero with sufficient accuracy.

C. Evaluation Metrics

Panoptic Quality (PQ), originally described in [1], is the most commonly adopted metric to evaluate panoptic segmentation results. We report it together with semantic Intersection over Union (mIoU) and mask mean Average Precision (mAP), in order to detailedly measure our network’s segmentation and detection performance, respectively. Some recent works [2, 3] have proposed alternatives to PQ aimed at highlighting different aspects of the panoptic predictions, or overcoming potential pitfalls of PQ. Note that we denote by PQ^{th} and PQ^{st} the PQ scores computed only on thing and stuff classes, respectively.¹

Parsing Covering. PQ assigns equal importance to all image segments, a choice which is not always desirable, *e.g.* autonomous driving systems might care more about objects closer to the vehicle, and thus appear bigger in the image, than far away ones. Motivated by this observation, [3] proposed Parsing Covering (PC) as an alternative panoptic metric that weights image segments in proportion to their areas. Since our CABB loss focuses on improving detection results of large objects, PC helps highlighting its impact.

PQ^\dagger . Porzi *et al.* [2] discussed a potential limitation of PQ, as it handles all classes in a uniform way, imposing a hard 0.5 threshold on IoUs of both things and stuff. While this is strictly necessary to obtain a unique matching between thing segments and their respective ground truth, it can result in strong over-penalization of stuff segments. To solve this, they propose PQ^\dagger as a direct modification of PQ which avoids the thresholding issue, giving a more faithful representation of the quality of stuff predictions.

D. Training hyper-parameters

All our networks are trained using stochastic gradient descent with momentum 0.9 and weight decay 10^{-4} . The training schedule starts with a warm-up phase, where the learning rate is increased linearly from 0 to a value lr_0 in the first 200 training steps. Then, the learning rate follows a linear decay schedule given by $lr_i = lr_0(1 - \frac{i}{\#steps})$, where lr_i is the value at training step i . In all of our experiments we augment the data with random horizontal flipping, and in those involving ISUS we fix the maximum “things” scale

augmentation range to $r_{\text{th}} = [0.25, 4]$. The scale augmentation range used in CUS always matches the r_{st} of the corresponding ISUS experiments on the same dataset. In the following we list the dataset-specific hyper-parameters. Note that all schedules used for a particular dataset result in approximately the same number of training iterations.

Mapillary Vistas. All MVD experiments use a “stuff” scale augmentation range of $r_{\text{st}} = [0.8, 1.25]$. When utilizing full images we set $s_0 = 1344$, $lr_0 = 0.02$, and we train for 75 epochs on batches including a single image per GPU. In all other experiments we set $s_0 = 2400$, $lr_0 = 0.04$, take crops of size 1024×1024 , and train for 300 epochs using batches of 4 crops per GPU.

Indian Driving Dataset. In the IDD experiments we fix $s_0 = 1080$ and $r_{\text{st}} = [0.5, 2]$. We train for 75 epochs with batch size of 1 per GPU and $lr_0 = 0.02$ when using full images, and for 600 epochs with batch size of 8 per GPU, $lr_0 = 0.08$ and crop size 512×512 when using crops.

Cityscapes. Finally, in the Cityscapes experiments we pre-train our networks on Mapillary Vistas, and fix $s_0 = 1024$ and $r_{\text{st}} = [0.5, 2]$. When using full images, we train for 20 epochs with batch size of 1 per GPU and $lr_0 = 0.01$. When using crops, we train for 150 epochs with batch size of 8 per GPU, $lr_0 = 0.04$ and crop size 512×512 .

E. Additional ISUS ablations

In order to validate the efficacy of ISUS, we perform an additional ablation experiment where we train our CROP network variant (with CUS) using standard scale augmentation in the range $[0.25, 4]$. Note that this is the same range as the r_{th} used in the ISUS experiments. The aim here is to verify whether the instance-aware scale sampling in ISUS has any impact on detection compared to a uniform sampling in the same range. When training on MVD, we obtain the following results: $PQ^{\text{th}}=42.3$, $mAP=22.8$. Compare these to the corresponding CROP + ISUS results: $PQ^{\text{th}}=43.1$, $mAP=23.0$.

F. Qualitative Results

In the following we visualize sample outputs of our best performing CROP + CABB + ISUS networks on Mapillary Vistas (Fig. 1), Cityscapes (Fig. 2) and the Indian Driving Dataset (Fig. 3).

References

[1] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *(CVPR)*, pages 9404–9413, 2019. 4

¹A similar notation is also used for PC.



Figure 1: Sample outputs of CROP + CABB + ISUS on Mapillary Vistas. Best viewed on screen.

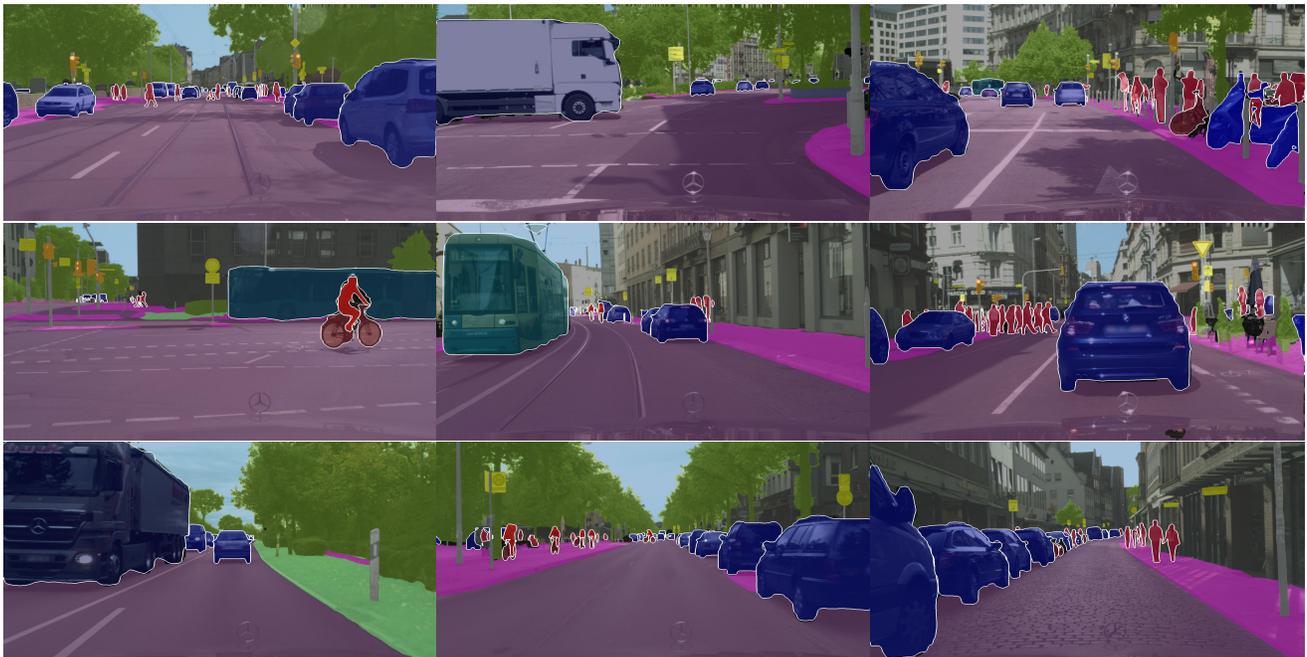


Figure 2: Sample outputs of CROP + CABB + ISUS on Cityscapes. Best viewed on screen.

[2] Lorenzo Porzi, Samuel Rota Bulò, Aleksander Colovic, and Peter Kotschieder. Seamless scene segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pat-*

tern Recognition, 2019. 4

[3] Tien-Ju Yang, Maxwell D. Collins, Yukun Zhu, Jyh-Jing Hwang, Ting Liu, Xiao Zhang, Vivienne Sze, George Papan-



Figure 3: Sample outputs of CROP + CABB + ISUS on the Indian Driving Dataset. Best viewed on screen.

dreou, and Liang-Chieh Chen. Deeperlab: Single-shot image parser. *CoRR*, abs/1902.05093, 2019. 4