Building Detail-Sensitive Semantic Segmentation Networks with Polynomial Pooling

Supplementary File

Zhen Wei, Jingyi Zhang, Li Liu, Fan Zhu, Fumin Shen, Yi Zhou, Si Liu
Yao Sun, Ling Shao

1. Proofs for the Propositions in Section 3.3

Proposition 1: *The p pooling is equivalent to average pooling when* $\alpha = 0$.

Proof. In the forward process, denote N to be the number of total elements in a pooling region and we have

$$y_j = \frac{\sum_{x_i \in R_j} x_i^{\alpha+1}}{\sum_{x_i \in R_j} x_i^{\alpha}} = \frac{\sum_{x_i \in R_j} x_i^{1}}{\sum_{x_i \in R_j} x_i^{0}} = \frac{1}{N} \sum_{x_i \in R_j} x_i \ .$$

While in the backward process, considering the gradient flow in one pooling region, namely to disentangle the summation over index j, then

$$\frac{\partial E}{\partial x_i} = \frac{(\alpha+1)x_i^{\alpha} \sum_{x_i \in R_j} x_i^{\alpha} - \alpha x_i^{\alpha-1} \sum_{x_i \in R_j} x_i^{\alpha+1}}{(\sum_{x_i \in R_j} x_i^{\alpha})^2} \delta_j^{l+1} = \frac{x_i^0 \sum_{x_i \in R_j} x_i^0}{(\sum_{x_i \in R_j} x_i^0)^2} \delta_j^{l+1} = \frac{N}{N^2} \delta_j^{l+1} = \frac{1}{N} \delta_j^{l+1} \ . \square$$

Proposition 2: *The p pooling performs as max pooling when* $\alpha \to +\infty$.

Proof. In j-th pooling region, given the greatest element x_{max} and the other elements $x_k \in R_j - \{x_{max}\}$, then in the forward process we have,

$$\lim_{\alpha \to +\infty} y_j = \frac{\sum_{x_i \in R_j} x_i^{\alpha+1}}{\sum_{x_i \in R_j} x_i^{\alpha}} = \frac{x_{max}^{\alpha+1}}{x_{max}^{\alpha}} = x_{max} \ \ ,$$

In the backward process, for each element,

$$\lim_{\alpha \to +\infty} \frac{\partial E}{\partial x_i} = \frac{(\alpha+1)x_i^{\alpha} \sum_{x_i \in R_j} x_i^{\alpha} - \alpha x_i^{\alpha-1} \sum_{x_i \in R_j} x_i^{\alpha+1}}{(\sum_{x_i \in R_j} x_i^{\alpha})^2} \delta_j^{l+1} = \frac{(\alpha+1)x_i^{\alpha} x_{max}^{\alpha} - \alpha x_i^{\alpha-1} x_{max}^{\alpha+1}}{x_{max}^{2\alpha}} \delta_j^{l+1} \ ,$$

If $x_i = x_{max}$, then

$$\lim_{\alpha \to +\infty} \frac{\partial E}{\partial x_{max}} = \frac{(\alpha+1)x_{max}^{2\alpha} - \alpha x_{max}^{2\alpha}}{x_{max}^{2\alpha}} \delta_j^{l+1} = \delta_j^{l+1} \ ,$$

Otherwise,

$$\lim_{\alpha \to +\infty} \frac{\partial E}{\partial x_k} = \frac{(\alpha+1)x_k^{\alpha} - \alpha x_k^{\alpha-1} x_{max}}{x_{max}^{\alpha}} \delta_j^{l+1} < \frac{(\alpha+1)x_k^{\alpha} - \alpha x_k^{\alpha}}{x_{max}^{\alpha}} \delta_j^{l+1} = (\frac{x_k}{x_{max}})^{\alpha} \delta_j^{l+1} = 0 \ ,$$

$$\begin{aligned} \lim_{\alpha \to +\infty} \frac{(\alpha+1)x_k^{\alpha} - \alpha x_k^{\alpha-1} x_{max}}{x_{max}^{\alpha}} \delta_j^{l+1} > \frac{(\alpha+1)x_k^{\alpha-1} - \alpha x_k^{\alpha-1} x_{max}}{x_{max}^{\alpha}} \delta_j^{l+1} &= (\frac{x_k}{x_{max}})^{\alpha-1} (\frac{\alpha+1-x_{max}}{x_{max}}) \delta_j^{l+1} \\ &= (\frac{x_k}{x_{max}})^{\alpha-1} (\frac{\alpha}{x_{max}}) \delta_j^{l+1} \ , \end{aligned}$$

By using L'Hospital's rule, we have

$$\lim_{\alpha \to +\infty} (\frac{x_k}{x_{max}})^{\alpha-1} (\frac{\alpha}{x_{max}}) \delta_j^{l+1} = (\frac{\alpha}{x_{max}}) \delta_j^{l+1} / (\frac{x_k}{x_{max}})^{1-\alpha} = (\frac{1}{x_{max}}) \delta_j^{l+1} / (\frac{x_k}{(1-\alpha)x_{max}})^{-\alpha} = 0 \ .$$

As $0 < \partial E / \partial x_k < 0$, so $\partial E / \partial x_k = 0$. $\square$

2. Details of the Backward Process of L_p Norm Pooling

As we claimed in Section 3.6, the backward process of L_p norm pooling is problematic. The reasons are two folds. First, the gradients respect to input data is not equal to the gradient in average ($p = 1$) or max pooling ($p \to +\infty$). Second, the gradient will explode when $p = 0$, 1 or $p \to +\infty$, making end-to-end optimization difficult. Here we provide our mathematical evidences for these claims.

Given the notations defined in Section 3.1, the forward process of L_p norm is

$$f_{L_p}(R_j) = (\sum_{x_i \in R_j} x_i^p)^{1/p} \quad . \tag{1}$$

Then the gradients for input data and p are

$$\frac{\partial E}{\partial x_i} = \sum_j \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial x_i} = \sum_j x_i^{p-1}(\sum_{x_i \in R_j} x_i^p)^{\frac{p}{1-p}}\delta_j^{l+1} \quad , \quad \frac{\partial E}{\partial p} = (-\frac{1}{p^2} ln(\sum_{x_i \in R_j} x_i^p) + \frac{x_i^{p-1}}{\sum_{x_i \in R_j} x_i^p})(\sum_{x_i \in R_j} x_i^p)^{\frac{1}{p}}\delta_j^{l+1} \quad . \tag{2}$$

It can be easily observed that: (a) numerical issues will occur in $\partial E/\partial p$ when $p = 0$ and the result will explode when $p \to +\infty$; (b) numerical issues will occur in $\partial E/\partial x_i$ when $p = 1$ and the result will explode when $p \to +\infty$. Therefore, the inequality of $\partial E/\partial x_i$ between L_p norm and average/max poolings is corroborated because $\partial E/\partial x_i$ cannot be calculated in the boundary conditions of L_p norm.

3. Full Quantitative Results of the Comparison on VOC Dataset in Table 1

In Section 4.1, we provide quantitative results of baseline models and our models in terms of their overall performances on VOC dataset. Here we demonstrate detailed results at category level.

Method	bg	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
Max	87.7	59.7	25.0	62.6	49.1	55.2	70.5	**66.1**	68.7	21.3	51.6	40.6	60.5	48.4	57.3	70.6	33.6	53.6	36.7	62.9	48.2	53.8
Average	85.9	57.7	24.3	61.9	47.6	54.1	66.0	64.7	67.3	20.5	50.5	39.9	60.0	47.6	56.5	69.8	32.6	51.5	35.3	61.4	46.9	52.5
Strided-C	86.1	54.0	19.9	59.2	45.7	53.9	60.2	63.8	65.5	20.1	49.2	39.7	58.7	46.8	56.0	67.1	31.5	52.1	34.9	61.0	44.4	50.9
Gated	86.8	58.9	24.8	62.1	48.2	54.8	67.1	64.9	67.5	20.7	51.4	40.8	59.8	47.9	57.7	69.7	34.2	52.8	36.1	62.5	47.4	53.1
DPP	87.9	60.2	26.3	63.3	48.9	55.0	70.9	65.7	67.9	21.9	51.9	41.0	60.7	47.5	58.7	70.8	33.9	53.4	35.9	63.5	48.6	54.0
$\mathcal{P}$-*pooling* (our method)	**88.3**	**60.8**	**28.8**	**63.9**	**49.9**	**55.7**	**71.1**	66.0	**69.2**	**25.2**	**52.8**	**43.1**	**61.1**	**48.9**	**60.3**	**71.4**	**36.1**	**54.8**	**37.5**	**64.0**	**49.1**	**55.1**

Table 1. Quantitative comparisons of VGG-based baseline models and the model w/ $\mathcal{P}$-*pooling* w.r.t. mIoU on the validation set of PASCAL VOC 2012.

Method	bg	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
Max	89.9	71.7	30.1	69.1	55.7	62.8	77.7	**72.4**	75.1	25.0	55.1	32.7	66.5	56.3	66.8	74.6	46.3	62.2	37.1	70.7	65.5	60.2
Average	88.7	70.8	29.5	69.6	54.8	62.1	77.1	70.4	74.0	24.9	54.9	32.9	65.4	54.7	65.0	74.9	45.9	61.4	36.8	70.9	65.1	59.5
Strided-C	90.4	71.2	30.0	68.9	56.1	63.3	78.4	71.9	74.8	25.3	55.3	33.2	66.8	57.1	66.5	75.0	46.7	61.9	37.1	71.0	66.3	60.3
Gated	89.6	71.0	30.4	67.9	55.5	62.5	77.4	71.2	73.4	24.6	54.8	32.9	65.0	55.4	65.9	74.1	45.2	61.5	36.5	69.4	64.8	59.5
DPP	88.9	71.4	30.5	70.1	54.7	63.3	76.9	72.1	74.5	26.1	56.3	32.6	65.5	56.5	67.0	73.7	46.9	62.3	37.2	70.1	65.7	60.1
$\mathcal{P}$-*pooling* (our method)	**90.5**	**72.1**	**32.3**	**70.8**	**56.5**	**64.1**	**78.6**	72.3	**75.9**	**26.7**	**58.0**	**34.5**	**67.0**	**57.3**	**68.1**	**75.3**	**48.9**	**62.7**	**38.3**	**71.8**	**66.6**	**61.3**

Table 2. Quantitative comparisons of ResNet-based baseline models and the model w/ $\mathcal{P}$-*pooling* w.r.t. mIoU on the validation set of PASCAL VOC 2012.

4. Segmentation Results

In this section, we provide additional comparisons on segmentation results on PASCAL VOC 2012, Cityscapes and ADE20K datasets, respectively. Models with $\mathcal{P}$-*pooling* are more sensitive to details, enabling the models to recover more small structures and suppress false positives near object boundaries.

Figure 1. Comparisons on segmentation results of the VGG-based baseline model (w/ max pooling) and the proposed model (w/ $\mathcal{P}$*-pooling*) on VOC dataset.

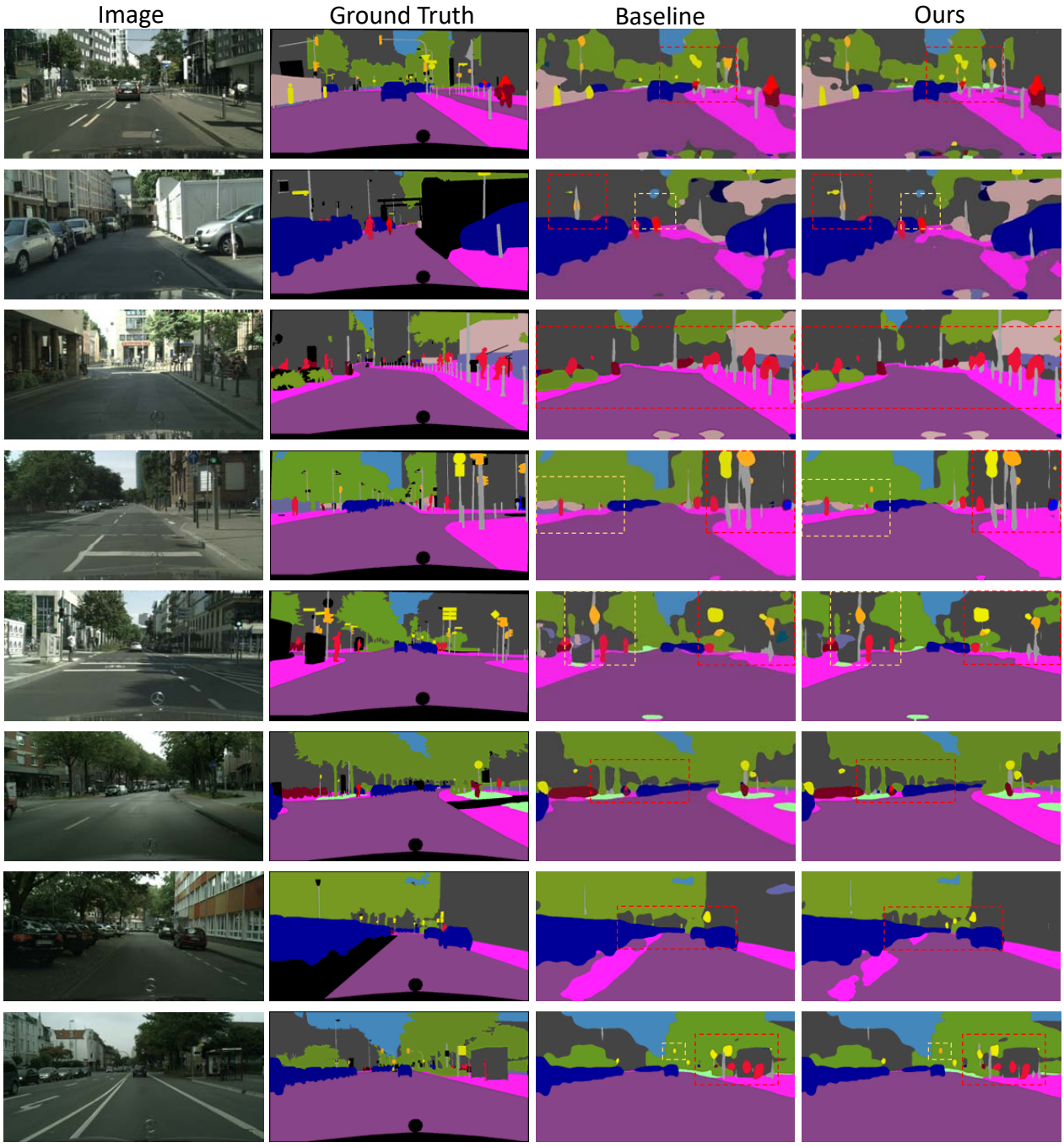

Figure 2. Comparisons on segmentation results of the Inceptionv2-based baseline model (w/ max pooling) and the proposed model (w/ $\mathcal{P}$*-pooling*) on Cityscapes dataset.

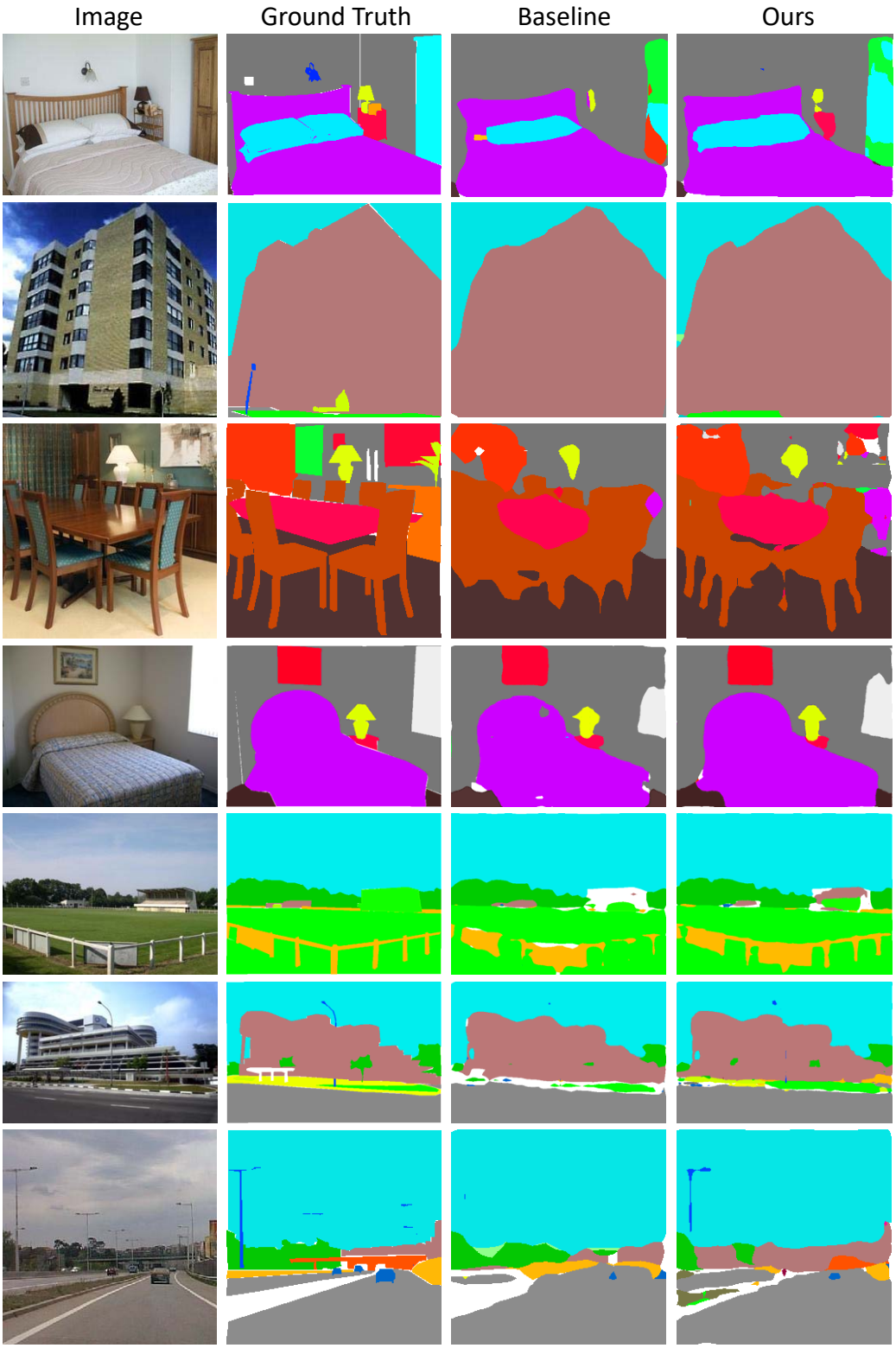

Figure 3. Comparisons on segmentation results of the Inceptionv2-based baseline model (w/ max pooling) and the proposed model (w/ $\mathcal{P}$*-pooling*) on ADE20K dataset.