Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

Supplementary material of Learning to Trace: Expressive Line Drawing Generation from Photographs

N. Inoue[1†], D. Ito[2], N. Xu[2], J. Yang[2], B. Price[2] and T. Yamasaki[1]

[1]The University of Tokyo, Japan [2]Adobe Research, U.S.

Table 1: *The architecture for the generator G. Up-sampling is done using nearest neighbours.*

Layer type	Kernel	Strides	Output size
input			$3 \times H \times W$
ResNet50 (∼conv3_4)			$512 \times H/8 \times W/8$
spatial dropout			$512 \times H/8 \times W/8$
up-sampling			$512 \times H/4 \times W/4$
convolution	3×3	1×1	$512 \times H/4 \times W/4$
convolution	3×3	1×1	$256 \times H/4 \times W/4$
convolution	3×3	1×1	$128 \times H/4 \times W/4$
up-sampling			$128 \times H/2 \times W/2$
convolution	3×3	1×1	$128 \times H/2 \times W/2$
convolution	3×3	1×1	$64 \times H/2 \times W/2$
convolution	3×3	1×1	$32 \times H/2 \times W/2$
up-sampling			$32 \times H \times W$
convolution	3×3	1×1	$32 \times H \times W$
convolution	3×3	1×1	$16 \times H \times W$
convolution	3×3	1×1	$1 \times H \times W$

1. Detail of G and R

We show detailed configuration of our proposed generator G and restorer R in Table 1 and Table 2, respectively. In the last convolutional layer, a Sigmoid layer is employed to normalize the output to the range of [0.0, 1.0]. A 1×1 zero padding is employed for each convolutional layer to maintain the resolution same. Batch normalization (BN) [IS15] is applied after each convolutional layer, followed by a Rectified Linear Unit (ReLU) [NH10]. Spatial Dropout [TGJ*15] is also employed before the first upsampling layer.

2. Additional Results

We show the additional results for the comparison among our model and comparable approaches for *manga BG* in Fig. 1 and Fig. 2. We can see that `Canny` and `Photoshop` is sometimes sensitive to global illumination, tends to produce texture-like lines, and sometimes ignore almost all the lines in the shadow. `Pix2pixHD` can handle these problems while there are too short lines, noises,

Table 2: *The architecture for the restorer R. Up-sampling is done using nearest neighbours.*

Layer type	Kernel	Strides	Output size
input			$3 \times H \times W$
convolution	5×5	2×2	$16 \times H/2 \times W/2$
convolution	3×3	2×2	$32 \times H/4 \times W/4$
convolution	3×3	1×1	$64 \times H/4 \times W/4$
convolution	3×3	2×2	$128 \times H/8 \times W/8$
convolution	3×3	1×1	$256 \times H/8 \times W/8$
convolution	3×3	1×1	$256 \times H/8 \times W/8$
convolution	3×3	1×1	$128 \times H/8 \times W/8$
convolution	3×3	1×1	$64 \times H/8 \times W/8$
spatial dropout			$64 \times H/8 \times W/8$
up-sampling			$64 \times H/4 \times W/4$
convolution	3×3	1×1	$64 \times H/4 \times W/4$
convolution	3×3	1×1	$32 \times H/4 \times W/4$
up-sampling			$32 \times H/2 \times W/2$
convolution	3×3	1×1	$32 \times H/2 \times W/2$
convolution	3×3	1×1	$16 \times H/2 \times W/2$
up-sampling			$16 \times H \times W$
convolution	3×3	1×1	$16 \times H \times W$
convolution	3×3	1×1	$8 \times H \times W$
convolution	3×3	1×1	$1 \times H \times W$

and lines with inconsistent intensity. On the other hand, our model produces clean and expressive line drawing images without any post-processing. We also show the comparison for *face/body* in Fig. 3.

3. Post-processing for LPCB

As a learnable CNN for edge detection, we tested LPCB [DSL*18]. Since the result of LPCB is still very blurry, we post-processed the result by binarization and morphological line thinning. We show the results with and without post-processing for *face/body* and *manga BG* in Fig. 4. Without post-processing, the lines are far from those in line drawing images.

References

[Ado] ADOBE SYSTEMS INC. *Adobe Photoshop CC*. https://www.adobe.com/products/photoshop.html 3–5.

[Can86] CANNY, JOHN. "A computational approach to edge detection". *IEEE TPAMI* 6 (1986), 679–698. DOI: 10.1109/TPAMI.1986.4767851 3–5.

[DSL*18] DENG, RUOXI, SHEN, CHUNHUA, LIU, SHENGJUN, et al. "Learning to predict crisp boundaries". *Proc. ECCV*. 2018, 562–578. DOI: 10.1007/978-3-030-01231-1_35 1, 3–6.

[IS15] IOFFE, SERGEY and SZEGEDY, CHRISTIAN. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". *Proc. ICML*. 2015, 448–456 1.

[NH10] NAIR, VINOD and HINTON, GEOFFREY E. "Rectified linear units improve restricted boltzmann machines". *Proc. ICML*. 2010, 807–814 1.

[TGJ*15] TOMPSON, JONATHAN, GOROSHIN, ROSS, JAIN, ARJUN, et al. "Efficient object localization using convolutional networks". *Proc. CVPR*. 2015, 648–656 1.

[WLZ*18] WANG, TING-CHUN, LIU, MING-YU, ZHU, JUN-YAN, et al. "High-resolution image synthesis and semantic manipulation with conditional gans". *Proc. CVPR*. 2018, 8798–8807. DOI: 10.1109/CVPR.2018.00917 3–5.

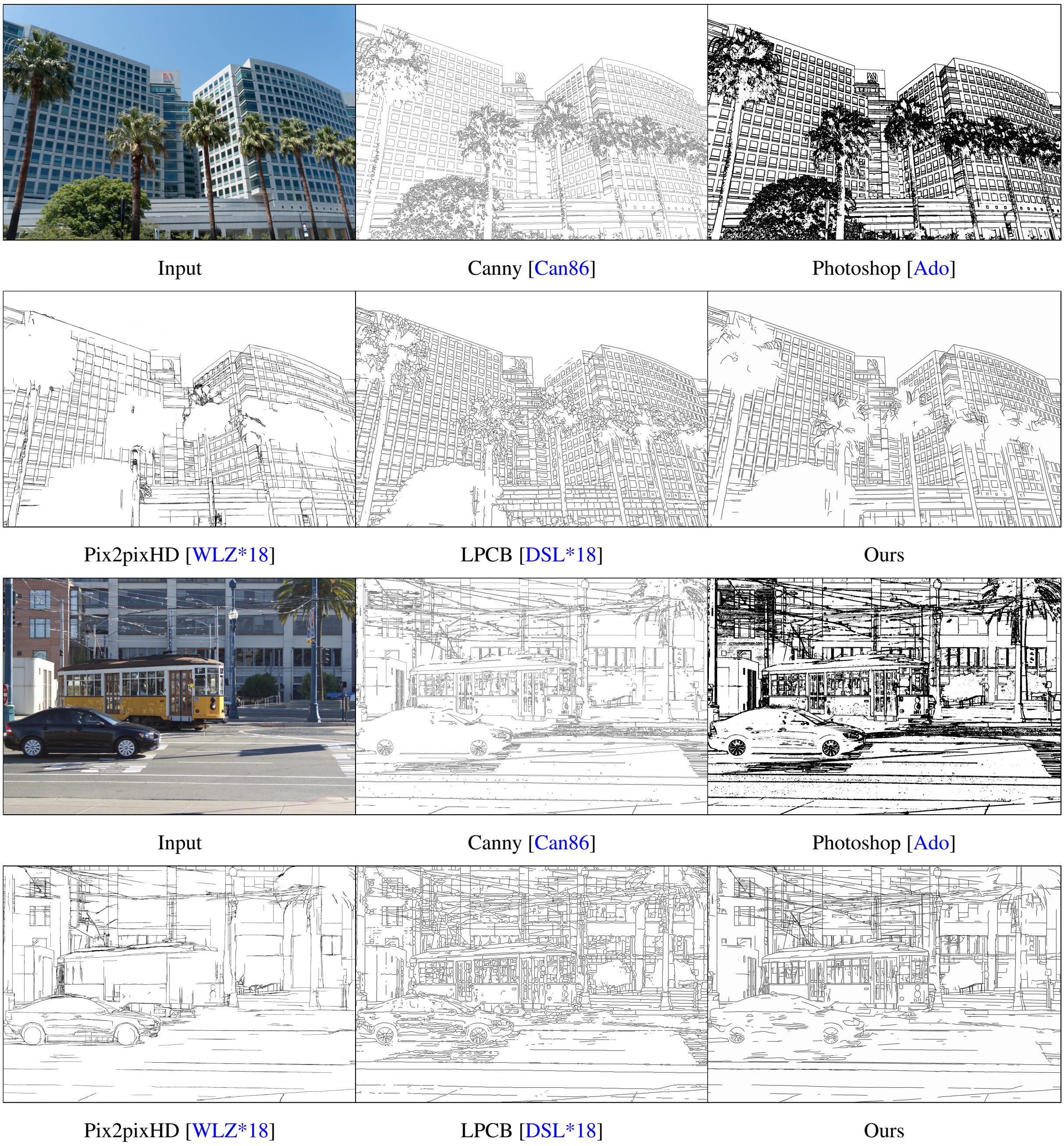

Figure 1: *Comparison with the comparable approaches for face/body. Note that no pre-processing and post-processing is applied. We can see that our approach outperforms the other approaches regarding cleanness and expressiveness. (best viewed in color and with zoom)*

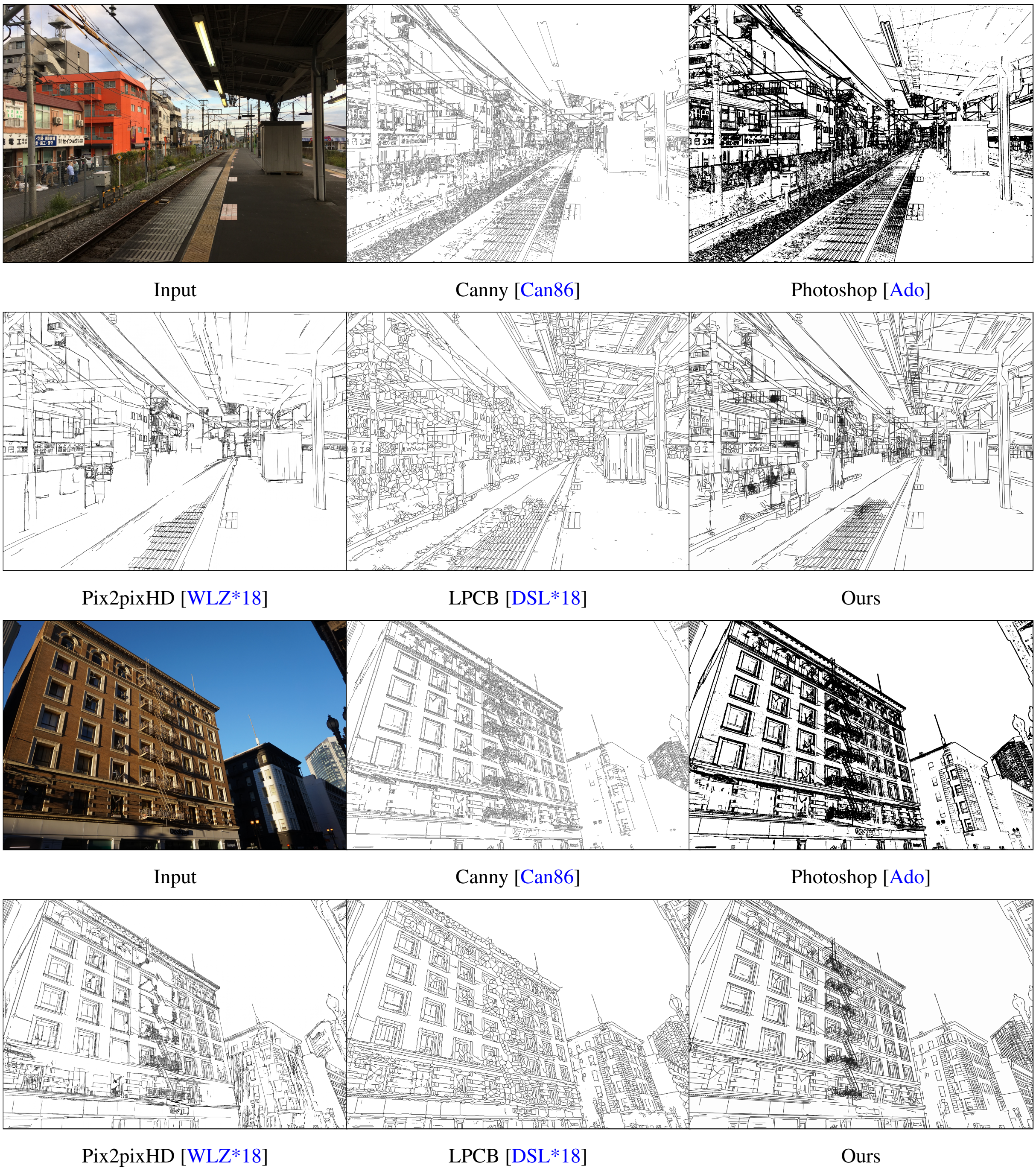

Figure 2: *Comparison with the comparable approaches for face/body. Note that no pre-processing and post-processing is applied. We can see that our approach outperforms the other approaches regarding cleanness and expressiveness. (best viewed in color and with zoom)*

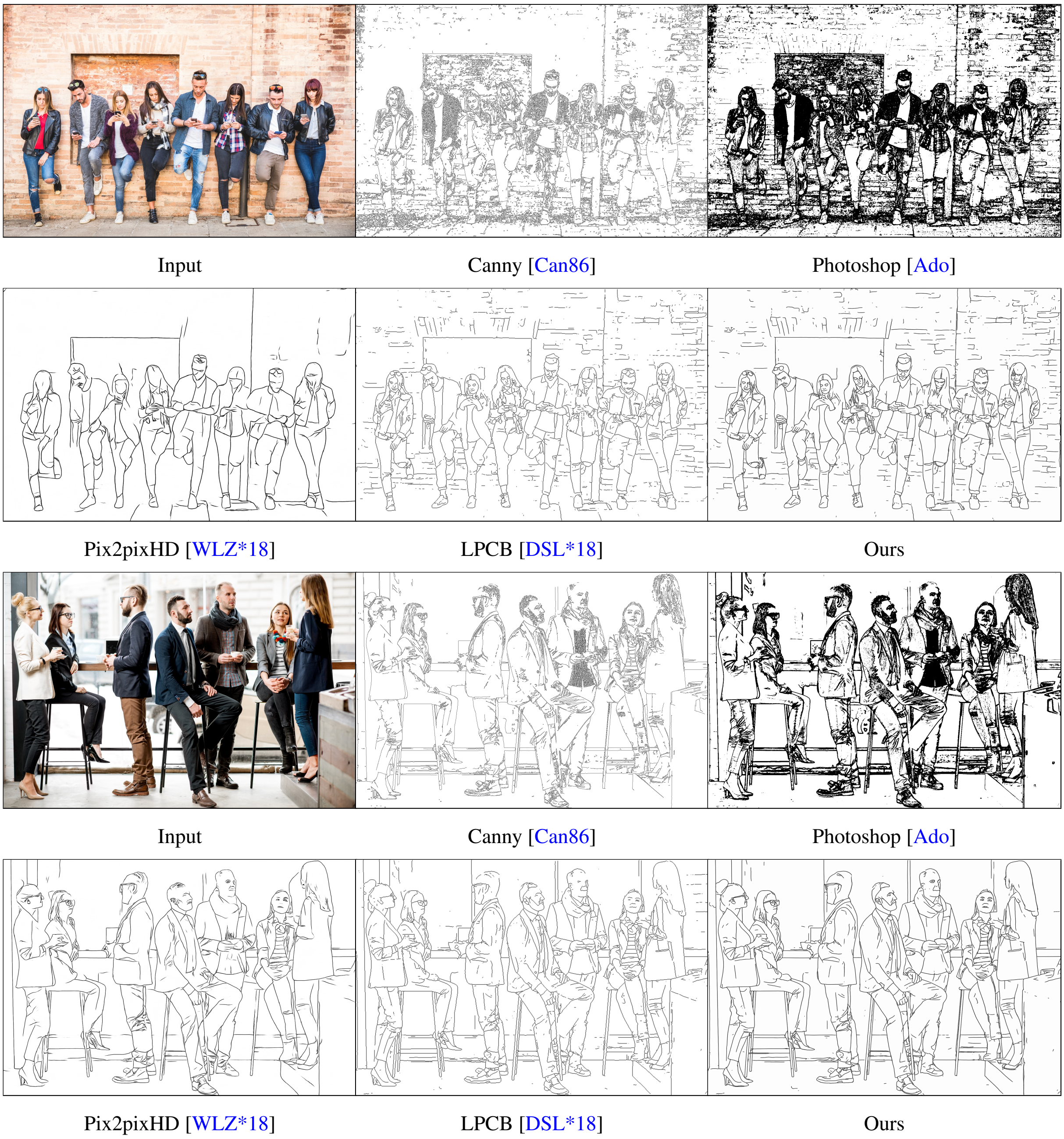

Figure 3: *Comparison with the comparable approaches for face/body. Note that no pre-processing and post-processing is applied. We can see that our approach outperforms the other approaches regarding cleanness and expressiveness. The photographs in the top and bottom are from Mirko - stock.adobe.com and rh2010 - stock.adobe.com,respectively. (best viewed in color and with zoom)*

Input w/o post-processing w/ post-processing

Figure 4: *The result of LPCB [DSL*18] with and without post-processing for face/body (in the first row) and for manga BG (in the second row). Without post-processing, the lines are far from those in line drawing images. The photograph in the first row is from wayne fleshman (Public Domain) (best viewed in color and with zoom)*