Supplementary Materials of Learning to Shadow Hand-drawn Sketches

Qingyuan Zheng*
University of Maryland,
Baltimore County
qing3@umbc.edu

Zhuoru Li*
Project HAT
hatsuame@gmail.com

Adam Bargteil
University of Maryland,
Baltimore County
adamb@umbc.edu

1. Lighting Directions

We found '810' numbering scheme to be more intuitive for the users than the other two methods ($[-1,0,1]^3$ or an integer between 1 and 26).

Label	Direction	Position
001	rear center	[0,0,1]
002	front center	[0,0,-1]
110	center top, front lighting	[0,1,-1]
120	center top, side lighting	[0,1,0]
130	center top, back lighting	[0,1,1]
210	upper right, front lighting	[1,1,-1]
220	upper right, side lighting	[1,1,0]
230	upper right, back lighting	[1,1,1]
310	center right, front lighting	[1,0,-1]
320	center right, side lighting	[1,0,0]
330	center right, back lighting	[1,0,1]
410	lower right, front lighting	[1,-1,-1]
420	lower right, side lighting	[1,-1,0]
430	lower right, back lighting	[1,-1,1]
510	bottom, front lighting	[0,-1,-1]
520	bottom, side lighting	[0,-1,0]
530	bottom, back lighting	[0,-1,1]
610	lower left, front lighting	[-1,-1,-1]
620	lower left, side lighting	[-1,-1,0]
630	lower left, back lighting	[-1,-1,1]
710	center left, front lighting	[-1,0,-1]
720	center left, side lighting	[-1,0,0]
730	center left, back lighting	[-1,0,1]
810	upper left, front lighting	[-1,1,-1]
820	upper left, side lighting	[-1,1,0]
830	upper left, back lighting	[-1,1,1]

Table 1: A lookup table of our 26 lighting direction labels, the actual lighting directions, and $[-1,0,1]^3$ style positions in programming.

*Equal contribution

2. Network Architecture

More details of network architecture are in Table 2, 3, 4, 5, 6, 7, 8 and Figure 1, 2 .

'ResiBlock': Residual Blocks. 'ShapeNet': the encoder of Generator. 'RenderNet': the decoder of Generator.

The first stage of RenderNet does not contain UpResiBlock nor SE block, and has four ResiBlocks. The second to the last stage of RenderNet does not contain SelfAttention. The last stage of RenderNet consists of just three ResiBlocks.

Layer	Filter	Output Size
Linear	128	128
Tanh()	-	128

Table 2: **Light Position embedding**

F(x)
BatchNorm
LeakyReLU()
Conv2D(1×1)
BatchNorm
LeakyRuLU()
Conv2D(3×3)
BatchNorm
LeakyReLU()
Conv2D(1×1)
Shortcut Branch
Conv2D(1×1)

Table 3: **ResiBlock**

F(x)
BatchNorm
LeakyReLU()
Conv2D(1 $\times$ 1)
BatchNorm
LeakyRuLU()
Conv2D(3 $\times$ 3, strides=2)
BatchNorm
LeakyReLU()
Conv2D(1 $\times$ 1)
Shortcut Branch
Conv2D(1 $\times$ 1, strides=2)

Table 4: **DownResiBlock (Downscale ResiBlock)**

F(x)
BatchNorm
LeakyReLU()
Conv2D(1 $\times$ 1)
BatchNorm
LeakyRuLU()
SubPixelConv2D(3 $\times$ 3, strides=2)
BatchNorm
LeakyReLU()
Conv2D(1 $\times$ 1)
Shortcut Branch
SubPixelConv2D(1 $\times$ 1, strides=2)

Table 5: **UpResiBlock (Upscale ResiBlock)**. Dropout(0.1) was added after the residual addition.

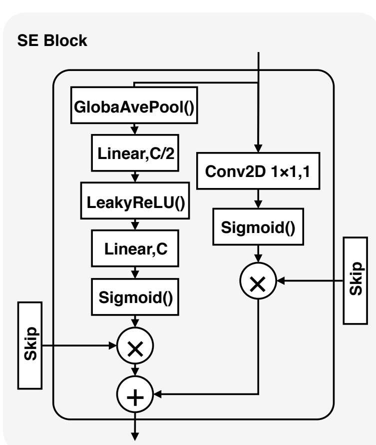


Figure 1: **SE(SE Block)**. 'C' is filter size.

Layer	Filter	Output Size
Concat(Coord)	-	$3 \times 320 \times 320$
ResiBlock	8, 8, 32	$32 \times 320 \times 320$
ResiBlock	8, 8, 32	$32 \times 320 \times 320$
DownResiBlock	16, 16, 64	$64 \times 160 \times 160$
ResiBlock	16, 16, 64	$64 \times 160 \times 160$
ResiBlock	16, 16, 64	$64 \times 160 \times 160$
DownResiBlock	32, 32, 128	$128 \times 80 \times 80$
ResiBlock	32, 32, 128	$128 \times 80 \times 80$
ResiBlock	32, 32, 128	$128 \times 80 \times 80$
DownResiBlock	64, 64, 256	$256 \times 40 \times 40$
ResiBlock	64, 64, 256	$256 \times 40 \times 40$
ResiBlock	64, 64, 256	$256 \times 40 \times 40$
DownResiBlock	64, 64, 256	$256 \times 20 \times 20$
ResiBlock	64, 64, 256	$256 \times 20 \times 20$
ResiBlock	64, 64, 256	$256 \times 20 \times 20$
DownResiBlock	128, 128, 512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$

Table 6: **ShapeNet**

Layer	Filter	Output Size
Concat(Pos, Coord)	-	$6 \times 320 \times 320$
DownResiBlock	8, 8, 32	$32 \times 160 \times 160$
ResiBlock	8, 8, 32	$32 \times 160 \times 160$
DownResiBlock	16, 16, 64	$64 \times 80 \times 80$
ResiBlock	16, 16, 64	$64 \times 80 \times 80$
DownResiBlock	32, 32, 128	$128 \times 40 \times 40$
ResiBlock	32, 32, 128	$128 \times 40 \times 40$
SelfAttention	-	$128 \times 40 \times 40$
DownResiBlock	64, 64, 256	$256 \times 20 \times 20$
ResiBlock	64, 64, 256	$256 \times 20 \times 20$
SelfAttention	-	$256 \times 20 \times 20$
DownResiBlock	128, 128, 512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$
GlobalAvgPool()	-	512
Dropout(0.3)	-	512
Linear	256	256
Linear	1	1
Sigmoid()	-	1

Table 7: **Discriminator**

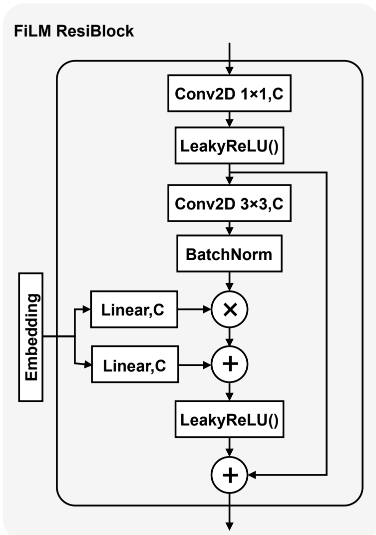


Figure 2: **FiLMResiBlock**[4]. 'C' is filter size.

3. Compositing sketches and shadows

Our result images (I) are composited by a simple weighted sum of the output shadow (S) and original line drawing (L)

$$I = 0.2S + 0.8L, \tag{1}$$

where all images are grayscale in $[0, 1]$.

In our training process, both line drawings and shadows are processed. The line drawings are inverted, $L' = 1 - L$, to achieve white lines on a black background. The shadow images are inverted, then scaled and shifted to the interval $[-1, 1]$, $S' = (1 - S) \times 2 - 1$. The inverse transform is applied to output from the Generator before compositing as described above.

Additionally, simply concatenating the line drawing and shadow for input to the Discriminator produced poor results; instead we composite these images with another weighted sum,

$$I' = L' + 0.25(S' + 1). \tag{2}$$

This compositing is applied to both ground truth and Generator shadows.

4. More Results

Figure 3, 4 : zoom in the Fig3 in our main body of paper. Figure 5, 6, 7, 8, 9 : more comparisons with related work. Figure 10, 11, 12 : our results in more lighting directions. Figure 13, 14, 15: examples of our shadowing system applied to artistic line drawings. Figure 16 : our results with and without pre-processing, and the robustness of our results in the wild. Figure 17: generalization ability. Figure 18 : Failure cases.

Layer	Filter	Output Size
Concat(Coord)	-	$514 \times 10 \times 10$
FiLMResiBlock(x, e)	512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$
ResiBlock	128, 128, 512	$512 \times 10 \times 10$
SelfAttention	-	$512 \times 10 \times 10$
UpResiBlock	64, 64, 256	$256 \times 20 \times 20$
Concat(SE(x, s), Coord)	-	$514 \times 20 \times 20$
FiLMResiBlock(x, e)	256	$256 \times 20 \times 20$
ResiBlock	64, 64, 256	$256 \times 20 \times 20$
ResiBlock	64, 64, 256	$256 \times 20 \times 20$
SelfAttention	-	$256 \times 20 \times 20$
UpResiBlock	64, 64, 256	$256 \times 40 \times 40$
Concat(SE(x, s), Coord)	-	$514 \times 40 \times 40$
FiLMResiBlock(x, e)	256	$256 \times 40 \times 40$
ResiBlock	64, 64, 256	$256 \times 40 \times 40$
ResiBlock	64, 64, 256	$256 \times 40 \times 40$
SelfAttention	-	$256 \times 40 \times 40$
UpResiBlock	32, 32, 128	$128 \times 80 \times 80$
Concat(SE(x, s), Coord)	-	$258 \times 80 \times 80$
FiLMResiBlock(x, e)	128	$128 \times 80 \times 80$
ResiBlock	32, 32, 128	$128 \times 80 \times 80$
ResiBlock	32, 32, 128	$128 \times 80 \times 80$
SelfAttention	-	$128 \times 80 \times 80$
UpResiBlock	16, 16, 64	$64 \times 160 \times 160$
Concat(SE(x, s), Coord)	-	$130 \times 160 \times 160$
FiLMResiBlock(x, e)	64	$64 \times 160 \times 160$
ResiBlock	16, 16, 64	$64 \times 160 \times 160$
ResiBlock	16, 16, 64	$64 \times 160 \times 160$
SelfAttention	-	$64 \times 160 \times 160$
UpResiBlock	8, 8, 32	$32 \times 320 \times 320$
Concat(SE(x, s), Coord)	-	$66 \times 320 \times 320$
FiLMResiBlock(x, e)	32	$32 \times 320 \times 320$
ResiBlock	8, 8, 32	$32 \times 320 \times 320$
ResiBlock	8, 8, 32	$32 \times 320 \times 320$
ResiBlock	4, 4, 16	$16 \times 320 \times 320$
ResiBlock	4, 4, 16	$16 \times 320 \times 320$
ResiBlock	4, 4, 16	$16 \times 320 \times 320$
Conv2D(1×1)	1	$1 \times 320 \times 320$
Tanh()	-	$1 \times 320 \times 320$

Table 8: **RenderNet**. 'e' is lighting position embedding. 's' is skip connection from ShapeNet.

Table 9 and 10 shows the statistically significance report of our user study in Likert scores. We deploy levene's test for the equality of variance, and Fisher's Least Significant Difference (LSD) for the analysis of variance.

	GT	Our	[3]	[5]	pix2pix	U-net
Mean	6.37	6.70	5.78	3.35	3.78	3.06
SD	1.55	1.33	1.27	1.30	1.31	2.11
Mean	6.50	6.66	5.69	3.44	3.91	3.03
SD	1.40	1.15	1.02	1.12	1.10	2.10

Table 9: Mean and standard deviation of user study (on Likert score). Top: total results. Bottom: results of people with drawing experience.

	MeanDiff	**T-value**	**P-value**	**Alpha**
Our / GT	0.33	1.18	0.24	0.05
[3] / GT	-0.59	-2.16	0.03	0.05
[3] / Our	-0.92	-3.34	$9.23E^{-4}$	0.05
[5] / GT	-3.02	-10.98	$2.41E^{-24}$	0.05
[5] / Our	-3.35	-12.16	$1.11E^{-28}$	0.05
[5] / [3]	-2.43	-8.82	$5.21E^{-17}$	0.05
pix2pix / GT	-2.60	-9.45	$4.83E^{-19}$	0.05
pix2pix / Our	-2.93	-10.63	$4.36E^{-23}$	0.05
pix2pix / [3]	-2.00	-7.29	$2.07E^{-12}$	0.05
pix2pix / [5]	0.42	1.53	0.126	0.05
U-net / GT	-3.31	-12.02	$3.76E^{-28}$	0.05
U-net / Our	-3.63	-13.20	$1.22E^{-32}$	0.05
U-net / [3]	-2.71	-9.87	$1.98E^{-20}$	0.05
U-net / [5]	-0.29	-1.04	0.30	0.05
U-net/pix2pix	-0.71	-2.57	0.01	0.05
Our / GT	0.16	0.52	0.60	0.05
[3] / GT	-0.82	-2.67	$8.07E^{-3}$	0.05
[3] / Our	-0.98	-3.19	$1.62E^{-3}$	0.05
[5] / GT	-3.06	-10.02	$6.79E^{-20}$	0.05
[5] / Our	-3.22	-10.54	$1.67E^{-21}$	0.05
[5] / [3]	-2.25	-7.34	$3.33E^{-12}$	0.05
pix2pix / GT	-2.60	-8.49	$2.35E^{-15}$	0.05
pix2pix / Our	-2.75	-9.01	$7.39E^{-17}$	0.05
pix2pix / [3]	-1.78	-5.82	$1.91E^{-8}$	0.05
pix2pix / [5]	0.47	1.53	0.128	0.05
U-net / GT	-3.48	-11.37	$3.87E^{-24}$	0.05
U-net / Our	-3.63	-11.89	$8.28E^{-26}$	0.05
U-net / [3]	-2.65	-8.70	$6.06E^{-16}$	0.05
U-net / [5]	-0.41	-1.35	0.178	0.05
U-net/pix2pix	-0.87	-2.88	$4.40E^{-3}$	0.05

Table 10: Statistical significance report of user study (on Likert score). Top: total results. Bottom: results of people with drawing experience.

5. Dataset Samples

Figure 19 shows {sketch, light direction, shadow, mask} sample pairs from our dataset. Our dataset comprises 1,160 sketch/shadow pairs and includes a variety of lighting directions and subjects. Specifically, 372 front-lighting, 506 side-lighting, 111 back-lighting, 85 center-back, and 86 center-front. With regard to subjects there are 867 single-person, 56 multi-person, 177 body-part, and 60 mecha.

References

[1] Zhengyan Gao, Taizan Yonetsuji, Tatsuya Takamura, Toru Matsuoka, and Jason Naradowsky. Automatic illumination effects for 2d characters. In *NIPS Workshop on Machine Learning for Creativity and Design*, 2018.

[2] Todd Goodwin, Ian Vollick, and Aaron Hertzmann. Isophote distance: a shading approach to artistic stroke thickness. In *Proceedings of the 5th international symposium on Non-photorealistic animation and rendering*, pages 53–62. ACM, 2007.

[3] Matis Hudon, Mairéad Grogan, Rafael Pagés, and Aljoša Smolić. Deep normal estimation for automatic shading of hand-drawn characters. In *European Conference on Computer Vision*, pages 246–262. Springer, 2018.

[4] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[5] Wanchao Su, Dong Du, Xin Yang, Shizhe Zhou, and Hongbo Fu. Interactive sketch-based normal map generation with deep neural networks. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1(1):22, 2018.

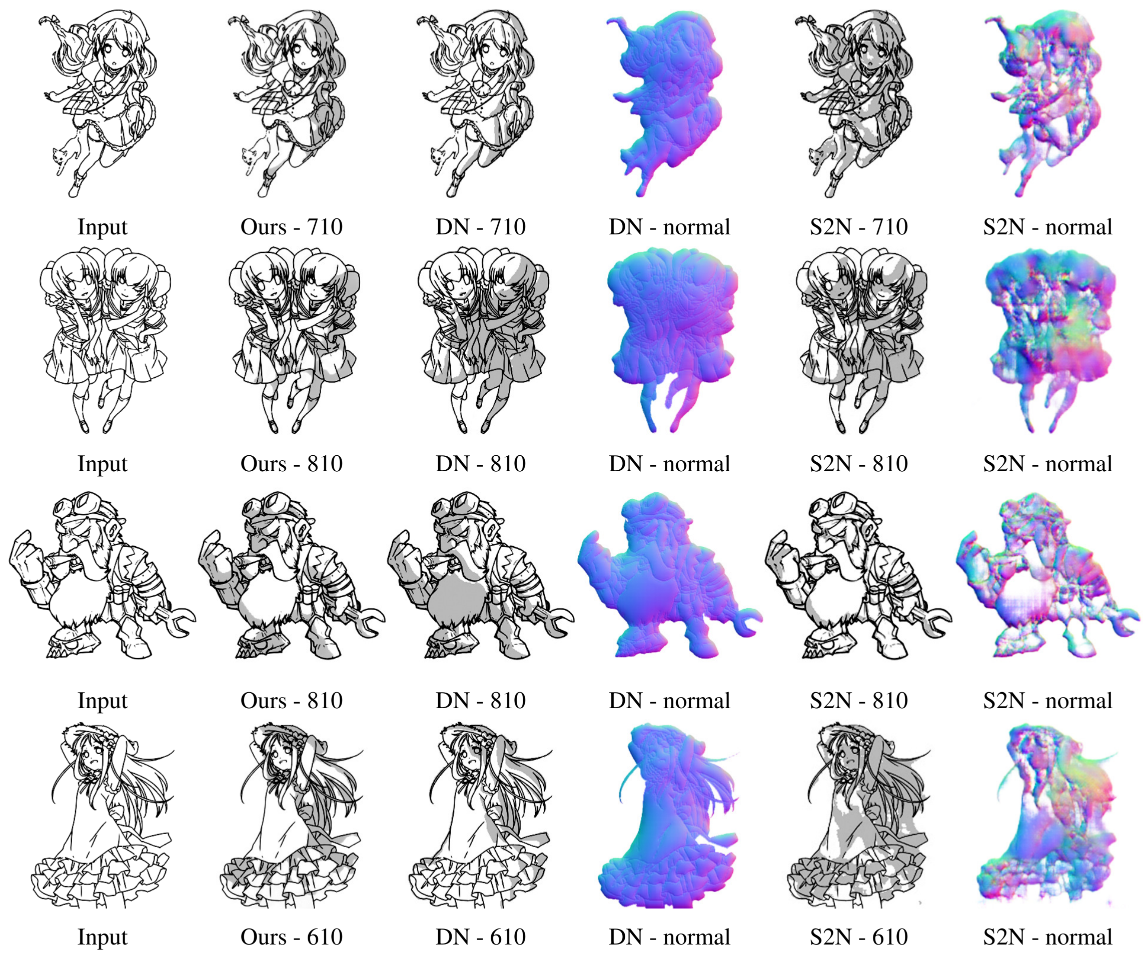


Figure 3: Comparisons with previous work DeepNormal (DN) [3] and Sketch2Normal (S2N) [5] in front lighting. (Part 1)

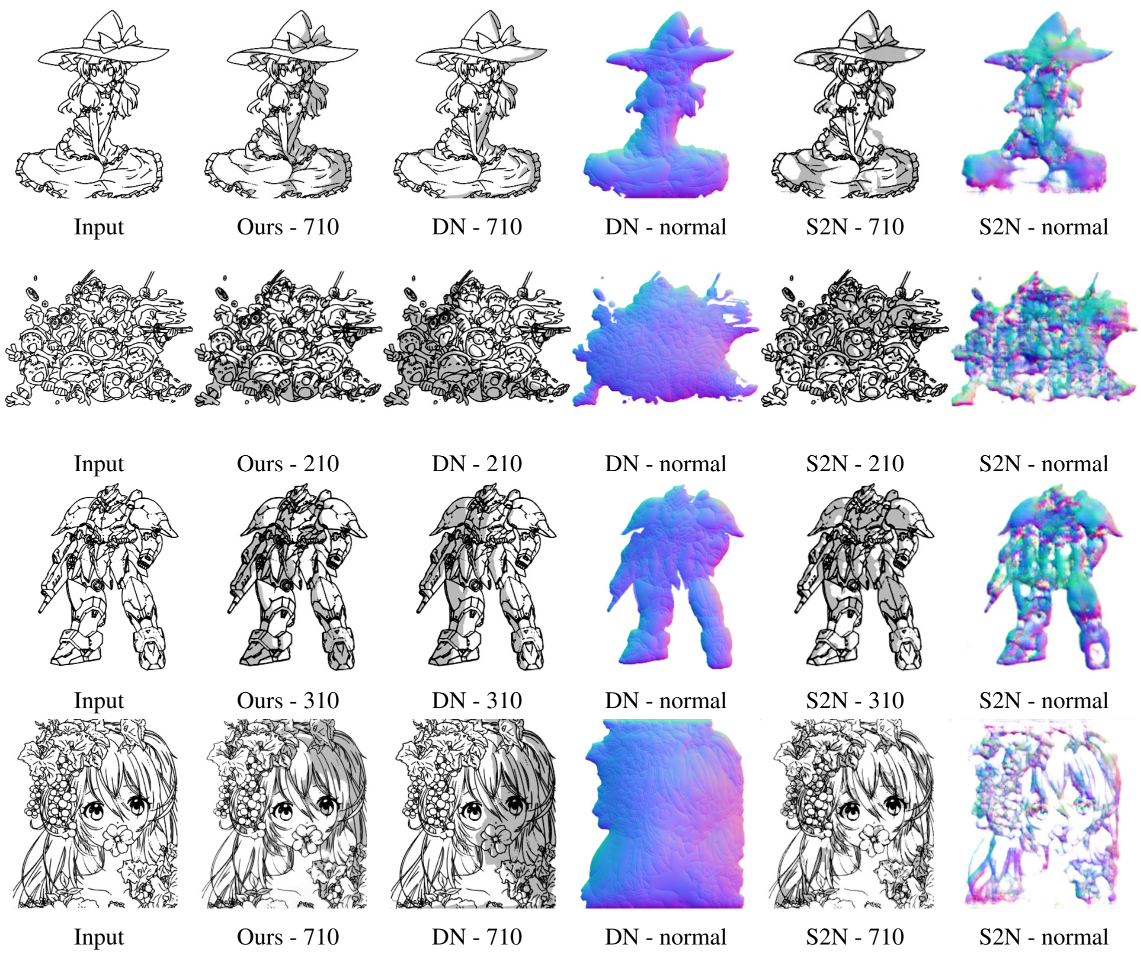


Figure 4: Comparisons with previous work DeepNormal (DN) [3] and Sketch2Normal (S2N) [5] in front lighting. (Part 2)

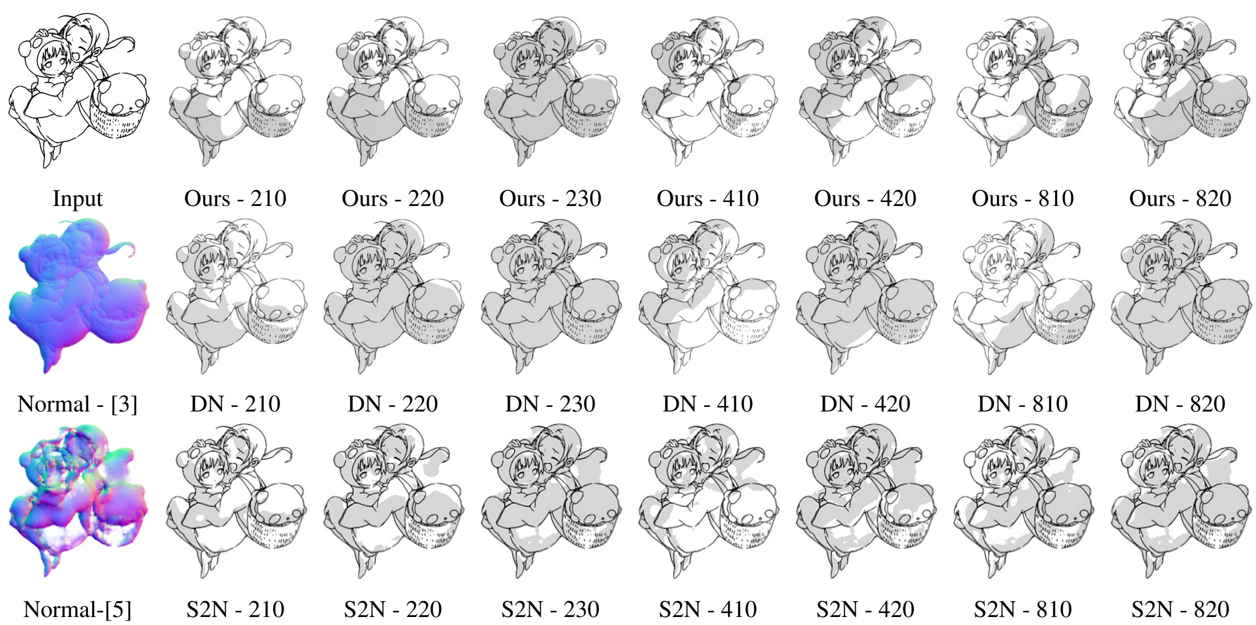


Figure 5: Comparisons with previous work DeepNormal (DN) [3] and Sketch2Normal (S2N) [5] when the light source changes depth. First row is ours. The second row is DeepNormal's. The third row is Sketch2Normal's.

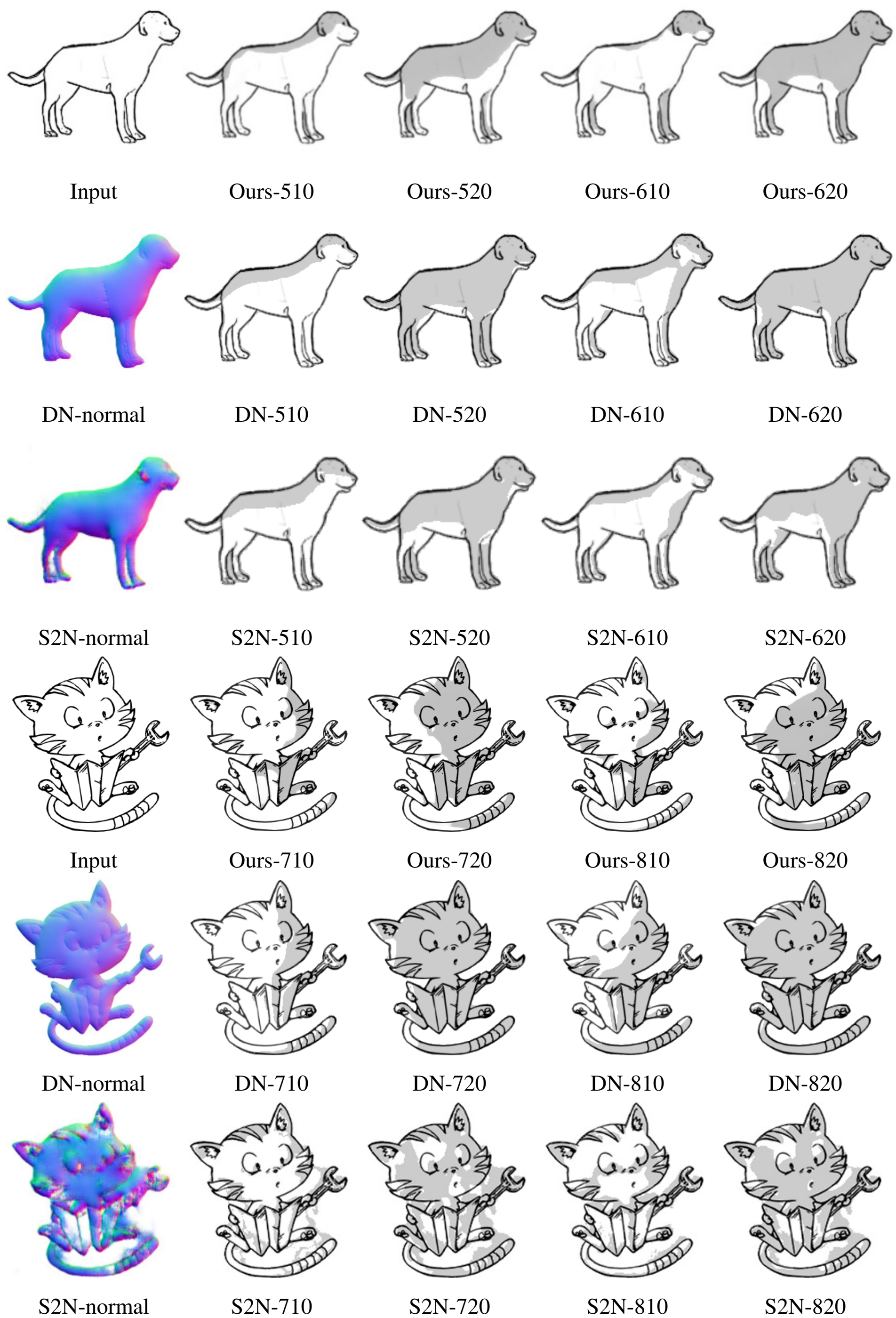


Figure 6: Comparisons with previous work DeepNormal (DN) [3] and Sketch2Normal (S2N) [5] using the line drawings from their papers. Dog image is from Sketch2Normal [5]. Cat image is from DeepNormal [3].

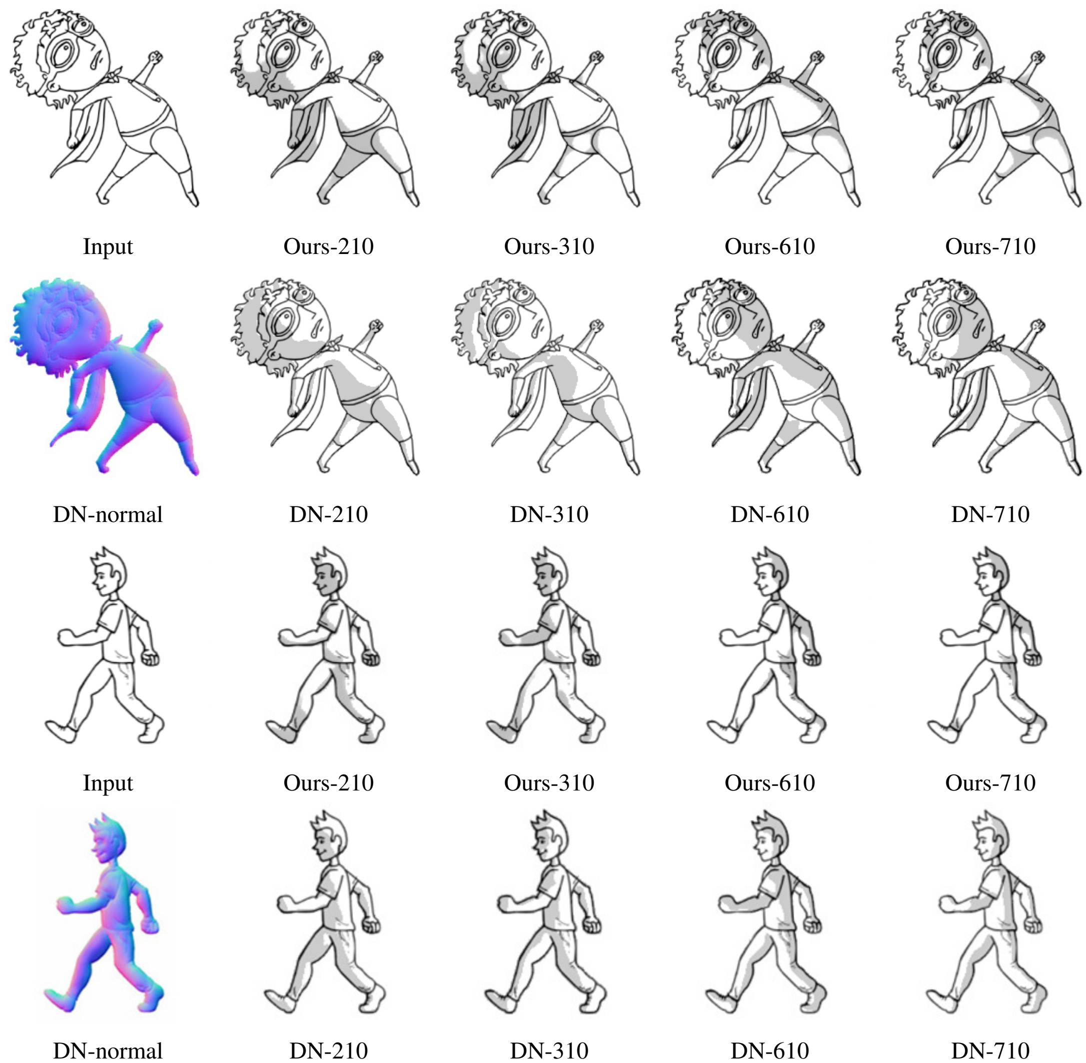


Figure 7: Comparisons with DeepNormal (DN) [3] using the line drawings and normal maps in the paper.

(a) colorized (b) lines (c) shadow - 610 (d) shadow - 710 (e) shadow - 810

(f) colorized (g) lines (h) shadow - 210 (i) shadow - 320 (j) shadow - 820

(k) colorized (l) lines (m) shadow - 220 (n) shadow - 320 (o) shadow - 710

(p) colorized (q) lines (r) shadow - 210 (s) shadow - 320 (t) shadow - 820

Figure 8: Our results using [1]'s images. [1] solves similar problems, inputting colorized images to predict the normal maps midway then generate the binary shadows. The colorized images (a), (f), (k), (p) are from [1]. (b), (g), (l), (q) are the lines that we subtract from the colorized images. Our system uses (b), (g), (l), (q) as the inputs to predict the pure shadows, then composite the shadows with the colorized images. For each line drawing, we show our results in three lighting directions. Please refer to [1] for the their results in similar lighting directions.

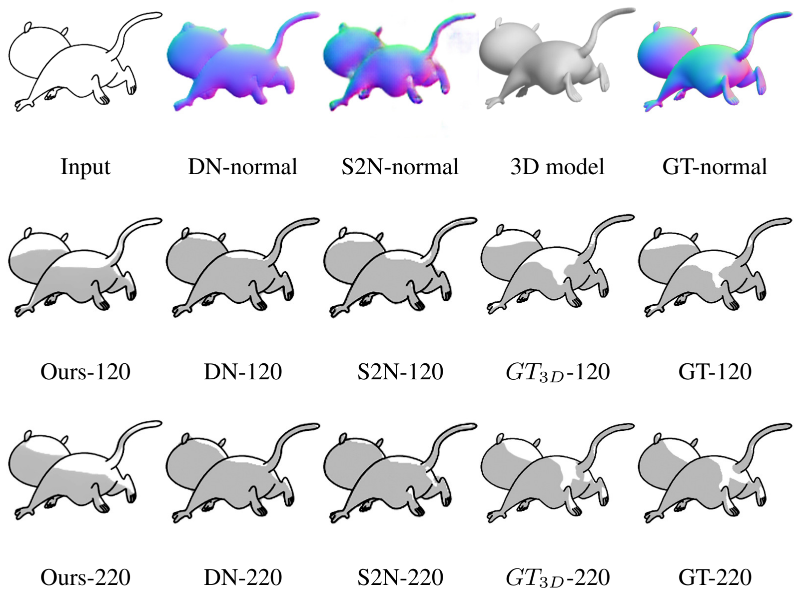


Figure 9: Comparisons on 3D test model [2] with DeepNormal (DN) [3], Sketch2Normal (S2N) [5] and Ground Truth (GT). GT_{3D}: rendered directly from 3D model with commercial software. GT: rendered from its normal map. All of the normal maps, including ground truth, are rendered with the same settings as the paper (use the renderer scripts provided by DeepNormal and threshold the continuous shadings at 0.5). Along with the bunny 3D test model in paper, our shadows are most close to the ground truth.

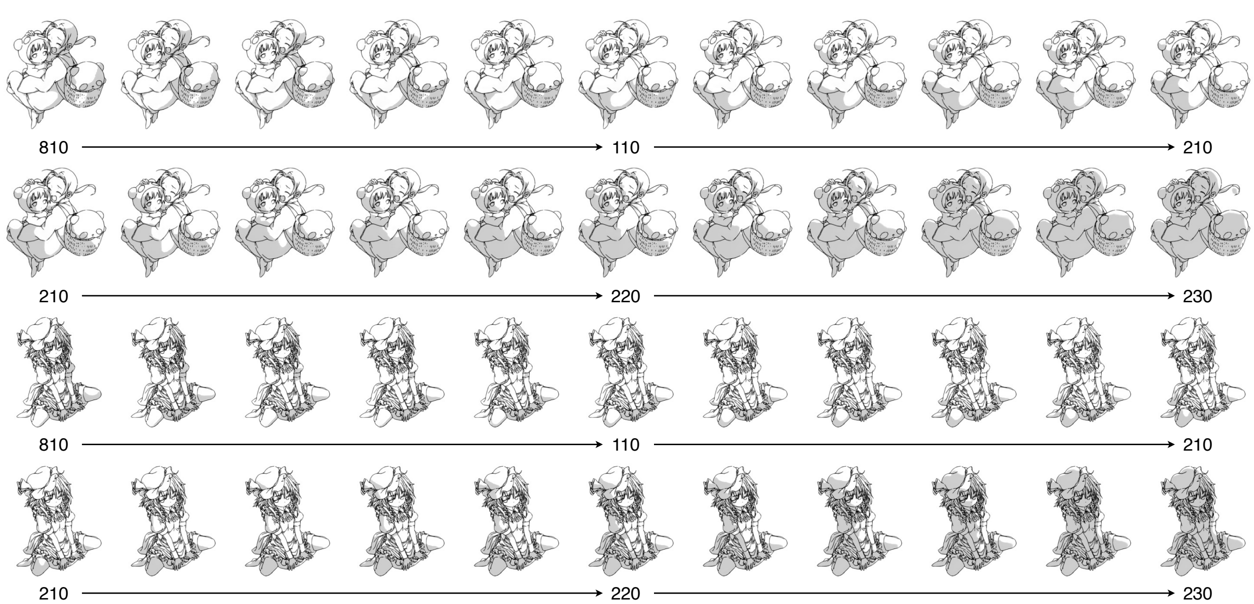


Figure 10: Examples of our continuous shadows between the discrete lighting source.

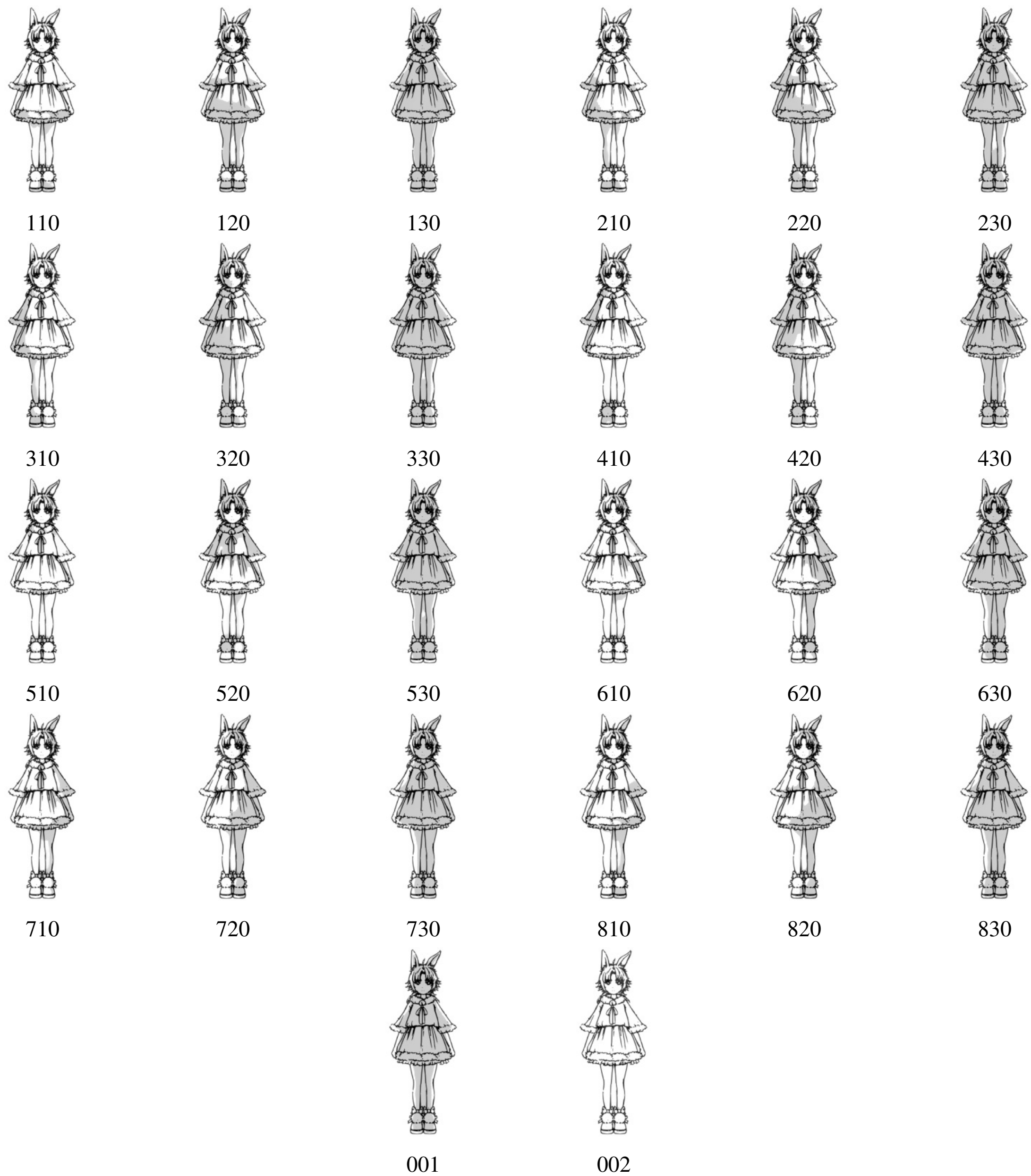


Figure 11: Evaluations of our work in all 26 directions.

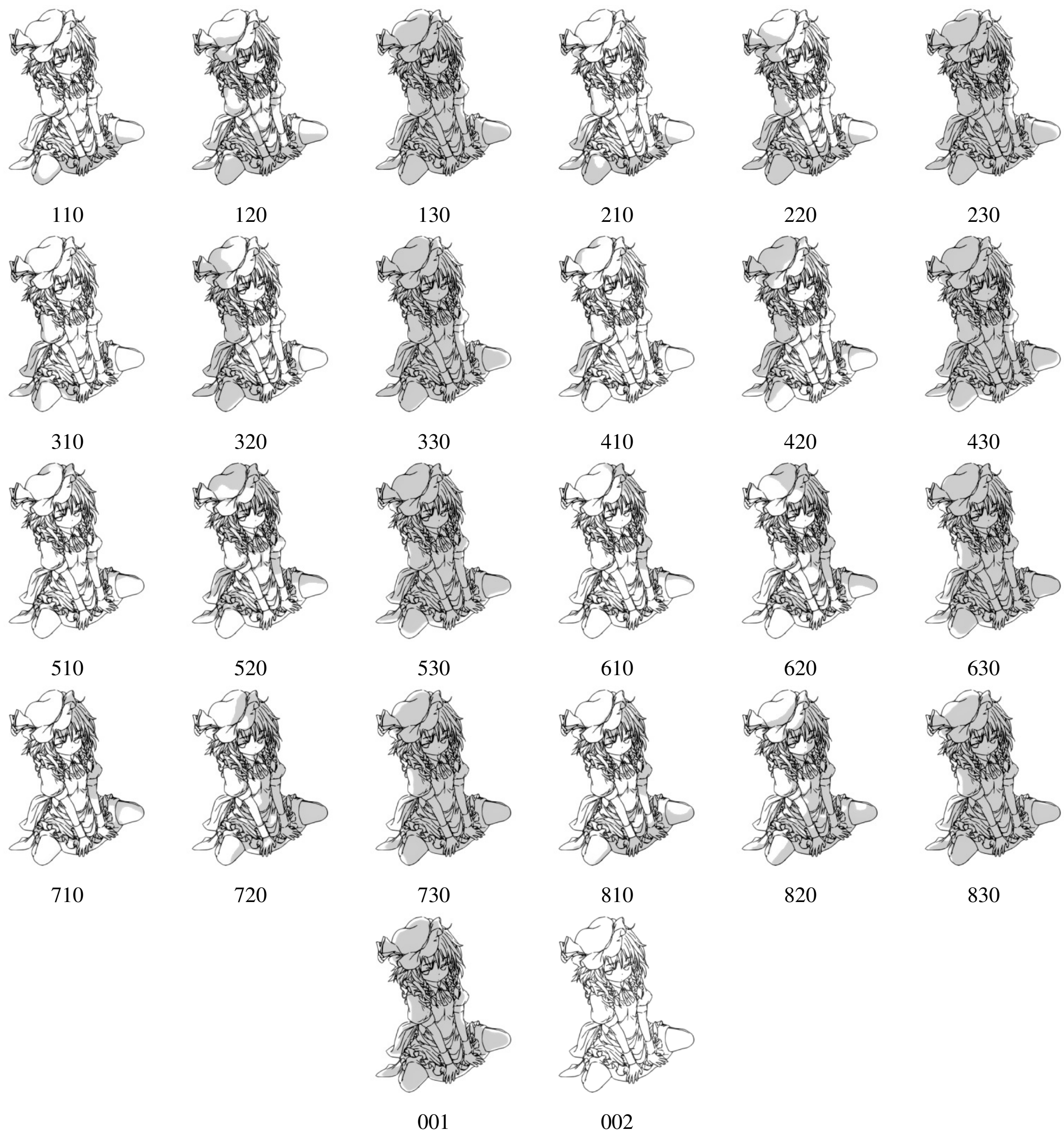


Figure 12: Evaluations of our work in all 26 directions.

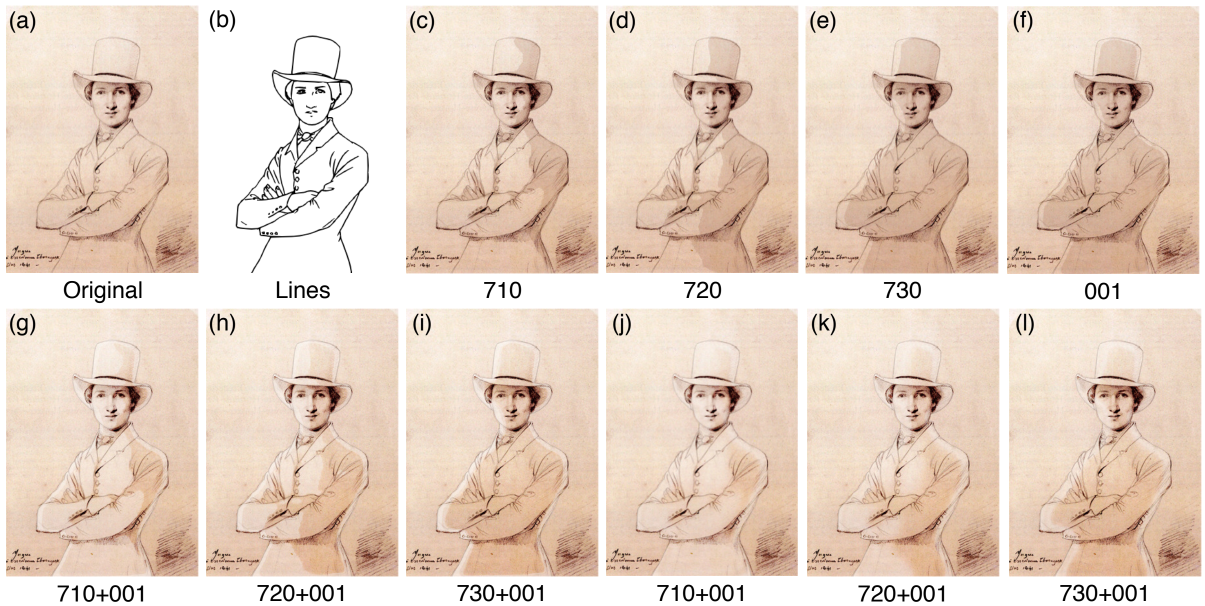


Figure 13: Examples of our shadowing system applying to artistic line drawing (Antoine Thomeguex, drawn by Jean Auguste Dominique Ingres. Public domain.). (a): original sketch. (b): extract lines from (a). (c)-(f): binary shadows in 710, 720, 730 and 001 lighting directions. (g)-(i): composites of binary shadows in dual lighting directions. (j)-(l): soft shadows in dual lighting directions. The results show that our shadowing system can give artists hints or a starting point to study shadows in different lighting sources.

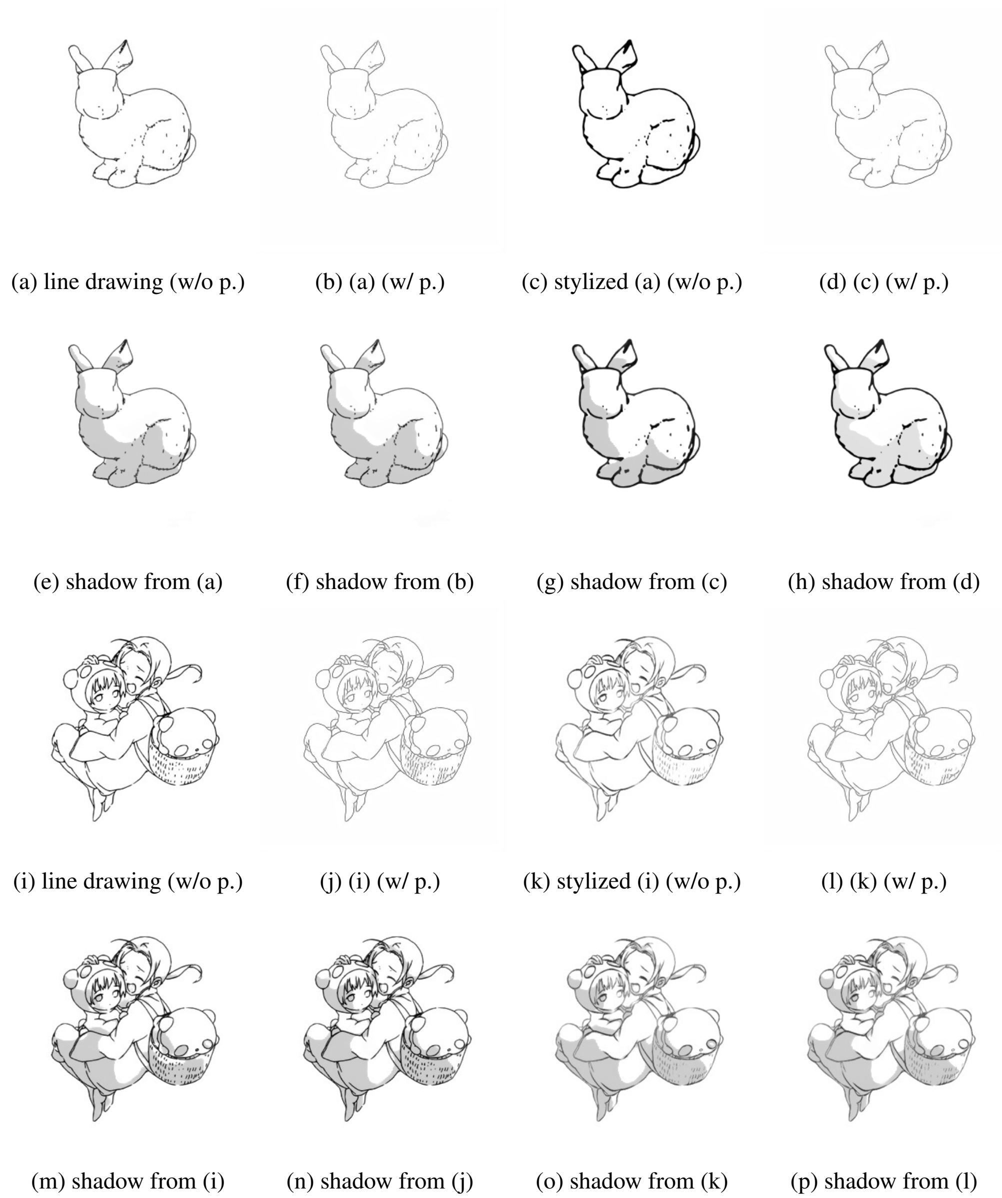
Figure 14: Examples of our shadowing system applying to Ukiyo-e (Kabuki Actor Segawa Kikunoj III as the Shirabyshi Hisakata Disguised as Yamato Manzai, by Toshusai Sharaku. Public domain.). Composite our shadows with pure colorized artwork.

(a) Original (b) w/o shadows (c) Lines (d) 510 (e) 520 (f) 530

(g) 210 (h) 220 (i) 230 (j) 810 (k) 820 (l) 830

(m) 110 (n) 120 (o) 130 (p) 710 (q) 720 (r) 730

Figure 15: Examples of our shadowing system applying to poster (Jardin de Paris, Fłte de Nuit Bal, illustrated by Jules Cheret. Public domain.). (a): original poster. (b): remove the shadows in the human. (c): extract line drawing from (a). (d)-(r): composites our shadows in various lighting directions with (b). Assuming the artists draw artwork with digital tools, they can rapidly try different shadows with our shadowing system.



(a) line drawing (w/o p.) (b) (a) (w/ p.) (c) stylized (a) (w/o p.) (d) (c) (w/ p.)

(e) shadow from (a) (f) shadow from (b) (g) shadow from (c) (h) shadow from (d)

(i) line drawing (w/o p.) (j) (i) (w/ p.) (k) stylized (i) (w/o p.) (l) (k) (w/ p.)

(m) shadow from (i) (n) shadow from (j) (o) shadow from (k) (p) shadow from (l)

Figure 16: The comparisons of our shadowing system with and without pre-processing (denoted as w/ p. and w/o p.). (a), (c), (i), (k) are line drawings without our pre-processing. (b), (d), (j), (l) are line drawings after our pre-processing. We test the robustness of our pre-processing system with stylized lines (c) and (k) which have different line width, line transparency, and line strokes.

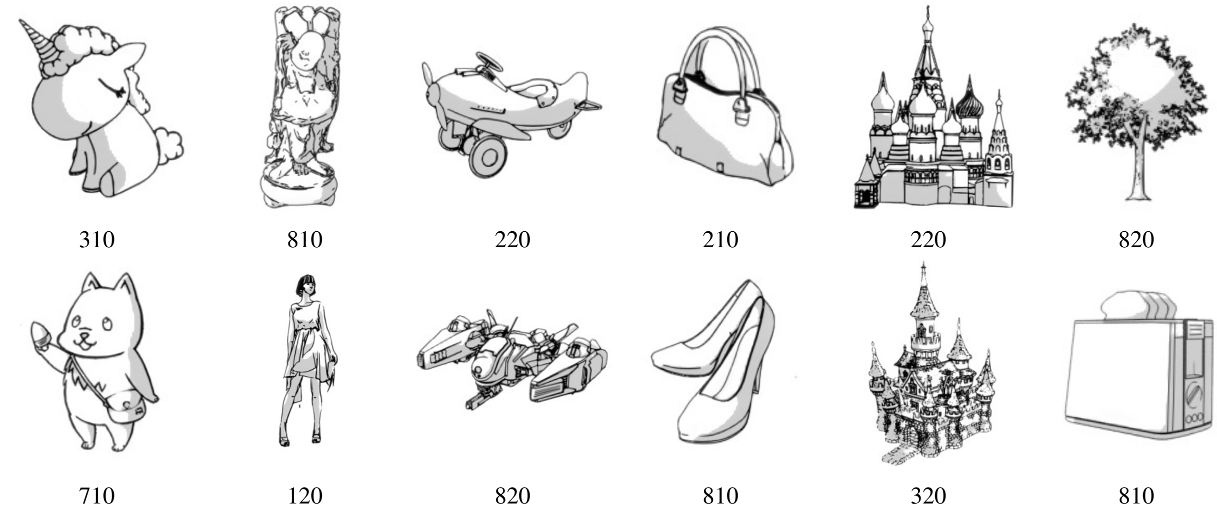


Figure 17: Evaluations on various categories of sketch (e.g. sculpture, bags, shoes, toys, sketchy cloth, buildings and etc.). This demonstrates that our work has generalization ability.

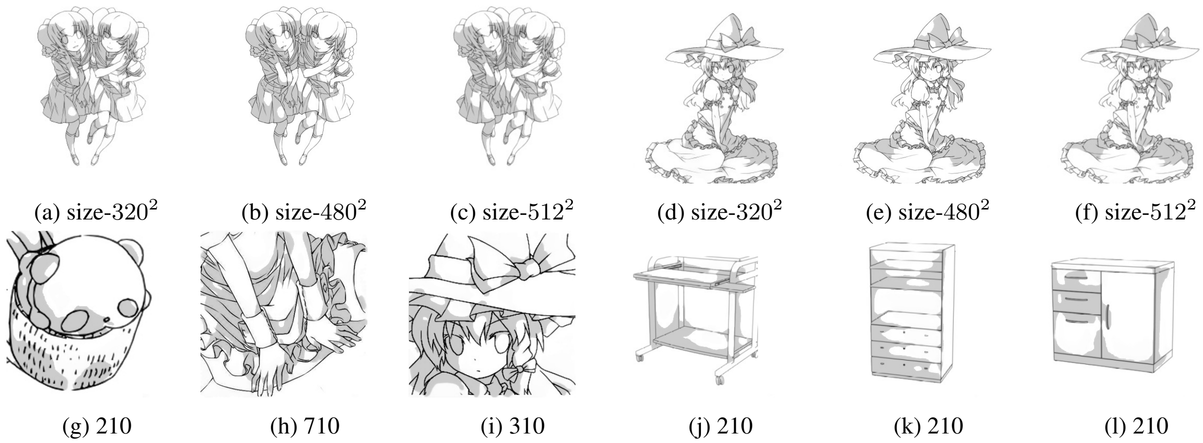


Figure 18: Limitation examples of the Future Work section. (a)-(c) and (d)-(f): invariant performance of shadows in different input size under the same lighting direction. (g)-(i): results of the local parts of line drawings being inputted. (j)-(l): unrealistic shadows in complex hard surface object.

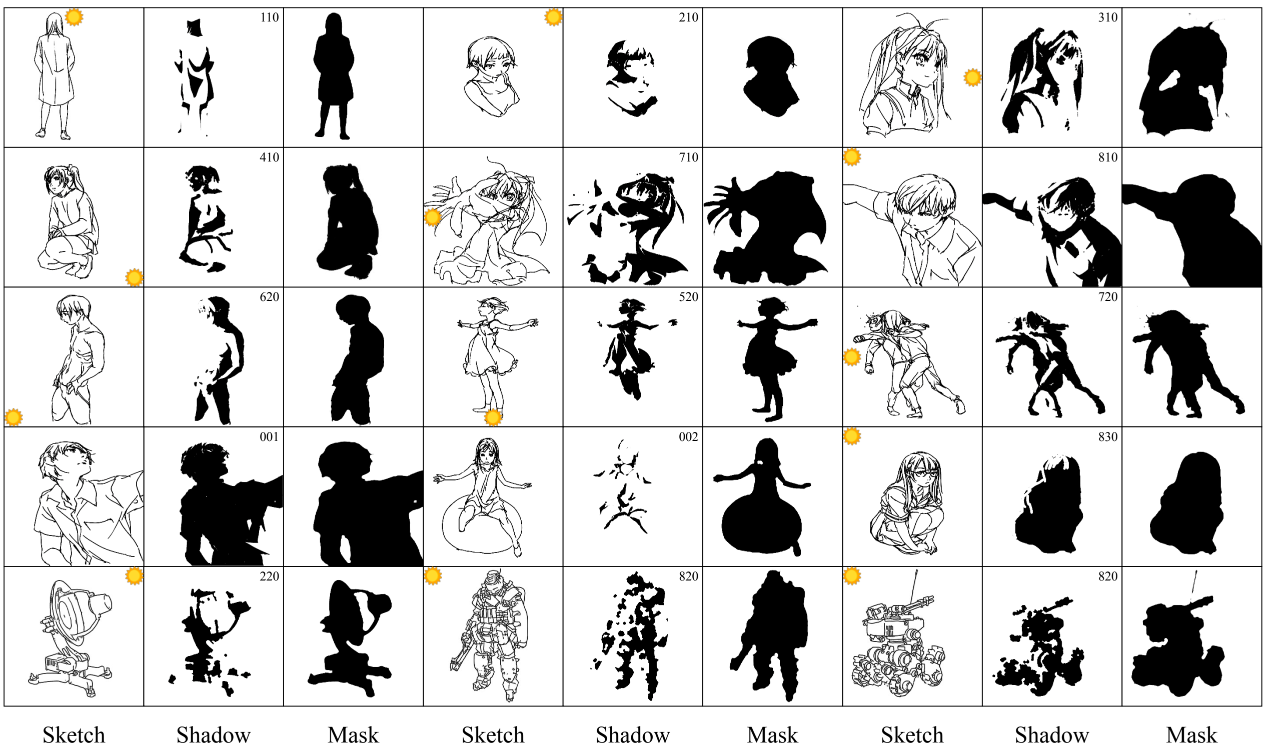


Figure 19: Sketch/shadow/mask pairs from our dataset. Our dataset contains alpha masks for the line drawings, but we did not need to use these masks in this paper.