

Supplementary Material of Pixel-Level Domain Transfer

Donggeun Yoo¹, Namil Kim¹, Sunggyun Park¹, Anthony S. Paek², In So Kweon¹

¹KAIST, Daejeon, South Korea.

²Lunit Inc., Seoul, South Korea.

{dgyoo,nikim}@rcv.kaist.ac.kr

{sunggyun,iskweon}@kaist.ac.kr

apaek@lunit.io

In this supplementary material of [1], we provide more details about our dataset and the user study, and conduct the image retrieval. We also show results obtained from the *inverse setting of domains*; products as a source domain and fashion models as a target domain. This material is composed as follows.

Section 1. Detail of LookBook.	2
Section 2. All Images and Scores in the User Study.	3
Section 3. Image Retrieval.	8
Section 4. From Products to Fashion Model Images.	11

1 Detail of LookBook

ROI crop All images collected from on-line shopping malls have been cropped to be mostly occupied by fashion models or products. For the fashion model images, we employ Faster R-CNN [2] to initially detect fashion models from all images. All detection boxes are cropped and manually screened to reject false positives or inaccurate boxes. For product images, we use Canny edge detector [3] to detect products because all product images have a clean background with a solid color (e.g. white).

Product categories All products in LookBook belong to the category “top” including the dresses. The following 17 sub-categories are included in LookBook. We could not provide the number of samples per sub-category because the shopping malls have different definitions and divisions on the sub-categories. It is further required to re-define sub-categories and label all images with them for more advanced domain transfer researches.

Midi dress, mini dress, coat, jacket, fur jacket, padded jacket, hooded jacket, jumper, cardigan, knitwear, blouse, shirt, sleeveless tee, short sleeve tee, long sleeve tee, hoody, vest.
--

Statistics We provide more details of image statistics in LookBook.

Item	Value
Total # of images	84,748
# of product images	9,732
# of fashion model images	75,016
Avg. # of fashion model images per product	7.71 ± 4.71
Min. # of fashion model images per product	1
Max. # of fashion model images per product	55
Avg. (width, height) of product images	(378.85 ± 178.33 , 493.94 ± 199.47)
Avg. (width, height) of fashion model images	(350.31 ± 138.58 , 762.91 ± 233.65)

2 All Images and Scores in the User Study

In this section, we provide all images used in the user study, and their average scores. Note that all of 100 source images used for the user study are *randomly chosen* from the test set of LookBook. “RF” and “MSE” indicate “Converter+RF-Discriminator” and “Converter+MSE”, respectively. “Real”, “Att.” and “Cat.” indicate the realistic score, attribute score and the category score, respectively.

Source	RF	MSE	Ours	Source	RF	MSE	Ours	Source	RF	MSE	Ours
Real	0.18	0.42	0.90	Real	0.60	0.10	0.78	Real	0.32	0.24	0.94
Att.	0.16	0.80	0.72	Att.	0.64	0.56	0.84	Att.	0.00	0.88	0.96
Cat.	0.02	0.74	0.68	Cat.	0.18	0.48	0.76	Cat.	0.00	0.48	1.00
Real	0.82	0.16	0.52	Real	0.34	0.42	0.74	Real	0.36	0.16	0.98
Att.	0.48	0.44	0.16	Att.	0.56	0.72	0.12	Att.	0.28	0.20	0.88
Cat.	0.24	0.28	0.90	Cat.	0.14	0.86	0.42	Cat.	0.10	0.36	0.94
Real	0.22	0.44	0.84	Real	0.34	0.34	0.82	Real	0.36	0.24	0.90
Att.	0.16	0.56	0.28	Att.	0.00	0.80	0.64	Att.	0.28	0.88	0.32
Cat.	0.00	0.86	0.56	Cat.	0.02	0.86	0.52	Cat.	0.00	0.74	0.68
Real	0.66	0.28	0.56	Real	0.60	0.14	0.76	Real	0.40	0.16	0.94
Att.	0.28	0.68	0.72	Att.	0.44	0.36	0.88	Att.	0.28	0.68	0.72
Cat.	0.00	0.60	0.82	Cat.	0.06	0.42	0.94	Cat.	0.00	0.60	0.82
Real	0.22	0.34	0.94	Real	0.68	0.22	0.60	Real	0.24	0.36	0.90
Att.	0.16	0.12	0.52	Att.	0.44	0.16	0.16	Att.	0.00	0.36	0.64
Cat.	0.22	0.34	0.86	Cat.	0.48	0.22	0.74	Cat.	0.00	0.64	0.78
Real	0.28	0.32	0.90	Real	0.32	0.18	1.00	Real	0.24	0.36	0.90
Att.	0.16	0.72	0.64	Att.	0.28	0.20	0.88	Att.	0.04	0.80	0.52
Cat.	0.00	0.68	0.74	Cat.	0.02	0.44	0.94	Cat.	0.00	0.86	0.56

Continued (1/5)

Source	RF	MSE	Ours	Source	RF	MSE	Ours	Source	RF	MSE	Ours
Real	0.28	0.26	0.94	Real	0.22	0.32	0.98	Real	0.44	0.26	0.78
Att.	0.48	0.72	0.84	Att.	0.36	0.64	1.00	Att.	0.00	0.80	0.88
Cat.	0.02	0.60	0.78	Cat.	0.00	0.52	0.90	Cat.	0.00	0.58	0.84
Real	0.24	0.32	0.94	Real	0.32	0.24	0.94	Real	0.34	0.34	0.76
Att.	0.00	0.44	0.80	Att.	0.16	0.72	0.84	Att.	0.44	0.84	0.12
Cat.	0.00	0.48	0.94	Cat.	0.00	0.68	0.74	Cat.	0.06	0.82	0.56
Real	0.44	0.28	0.76	Real	0.08	0.58	0.84	Real	0.74	0.16	0.60
Att.	0.28	0.88	0.64	Att.	0.20	0.84	0.52	Att.	0.48	0.20	0.36
Cat.	0.00	0.66	0.76	Cat.	0.00	0.86	0.56	Cat.	0.10	0.52	0.78
Real	0.68	0.36	0.44	Real	0.66	0.18	0.66	Real	0.24	0.34	0.92
Att.	0.72	0.72	0.56	Att.	0.00	0.20	0.80	Att.	0.20	0.52	0.68
Cat.	0.50	0.58	0.34	Cat.	0.06	0.44	0.92	Cat.	0.02	0.48	0.92
Real	0.44	0.14	0.92	Real	0.48	0.34	0.68	Real	0.36	0.32	0.82
Att.	0.48	0.36	0.84	Att.	0.00	0.96	0.84	Att.	0.16	0.16	0.16
Cat.	0.34	0.42	0.66	Cat.	0.00	0.98	0.56	Cat.	0.08	0.64	0.72
Real	0.40	0.24	0.86	Real	0.28	0.40	0.82	Real	0.34	0.32	0.84
Att.	0.00	0.80	0.72	Att.	0.04	0.96	0.04	Att.	0.00	0.88	0.88
Cat.	0.00	0.68	0.74	Cat.	0.06	0.94	0.42	Cat.	0.00	0.58	0.84
Real	0.48	0.28	0.74	Real	0.32	0.24	0.94	Real	0.24	0.26	1.00
Att.	0.12	0.44	0.36	Att.	0.00	0.68	0.56	Att.	0.28	0.68	0.80
Cat.	0.02	0.72	0.68	Cat.	0.00	0.66	0.76	Cat.	0.00	0.56	0.86
Real	0.56	0.22	0.74	Real	0.68	0.18	0.64	Real	0.40	0.24	0.86
Att.	0.20	0.72	0.72	Att.	0.00	0.12	0.88	Att.	0.36	0.20	0.68
Cat.	0.02	0.50	0.90	Cat.	0.02	0.50	0.90	Cat.	0.02	0.56	0.84

Continued (2/5)

Source	RF	MSE	Ours	Source	RF	MSE	Ours	Source	RF	MSE	Ours
Real	0.32	0.24	0.94	Real	0.34	0.16	1.00	Real	0.32	0.22	0.98
Att.	0.04	0.68	0.80	Att.	0.04	0.28	0.72	Att.	0.04	0.64	0.72
Cat.	0.00	0.66	0.76	Cat.	0.00	0.52	0.90	Cat.	0.00	0.58	0.84
Real	0.06	0.48	0.98	Real	0.32	0.36	0.82	Real	0.82	0.10	0.58
Att.	0.28	0.32	0.36	Att.	0.16	0.72	0.64	Att.	0.20	0.28	0.84
Cat.	0.00	0.72	0.72	Cat.	0.00	0.74	0.68	Cat.	0.44	0.18	0.78
Real	0.58	0.44	0.48	Real	0.28	0.40	0.82	Real	0.60	0.26	0.64
Att.	0.00	0.68	0.64	Att.	0.52	0.56	0.32	Att.	0.04	0.52	0.80
Cat.	0.00	0.68	0.74	Cat.	0.08	0.66	0.68	Cat.	0.08	0.74	0.60
Real	0.52	0.22	0.76	Real	0.36	0.16	0.98	Real	0.68	0.28	0.52
Att.	0.36	0.64	0.72	Att.	0.16	0.72	0.88	Att.	0.16	0.96	0.28
Cat.	0.00	0.72	0.72	Cat.	0.00	0.56	0.86	Cat.	0.02	0.76	0.64
Real	0.34	0.18	0.98	Real	0.28	0.24	0.98	Real	0.36	0.26	0.86
Att.	0.04	0.80	0.88	Att.	0.12	0.00	0.56	Att.	0.48	0.84	0.96
Cat.	0.02	0.50	0.90	Cat.	0.00	0.52	0.90	Cat.	0.00	0.56	0.98
Real	0.52	0.18	0.78	Real	0.74	0.14	0.64	Real	0.50	0.32	0.68
Att.	0.16	0.72	0.96	Att.	0.48	0.32	0.64	Att.	0.12	0.96	0.72
Cat.	0.00	0.56	0.86	Cat.	0.44	0.34	0.60	Cat.	0.00	0.82	0.58
Real	0.42	0.22	0.86	Real	0.14	0.44	0.92	Real	0.84	0.22	0.44
Att.	0.12	0.56	0.80	Att.	0.00	0.80	0.56	Att.	0.56	0.28	0.12
Cat.	0.00	0.68	0.74	Cat.	0.00	0.74	0.68	Cat.	0.66	0.52	0.24
Real	0.32	0.26	0.92	Real	0.64	0.24	0.64	Real	0.58	0.18	0.74
Att.	0.04	0.56	0.80	Att.	0.68	0.80	0.16	Att.	0.00	0.72	0.84
Cat.	0.06	0.60	0.76	Cat.	0.36	0.50	0.56	Cat.	0.00	0.58	0.84

Continued (3/5)

Source	RF	MSE	Ours	Source	RF	MSE	Ours	Source	RF	MSE	Ours
Real	0.52	0.32	0.66	Real	0.32	0.24	0.94	Real	0.64	0.28	0.58
Att.	0.28	0.12	0.36	Att.	0.36	0.84	0.84	Att.	0.04	0.72	0.72
Cat.	0.36	0.26	0.78	Cat.	0.00	0.58	0.84	Cat.	0.02	0.84	0.56
Real	0.64	0.28	0.58	Real	0.28	0.34	0.86	Real	0.26	0.42	0.82
Att.	0.20	0.44	0.64	Att.	0.12	0.56	0.44	Att.	0.00	0.68	0.88
Cat.	0.02	0.74	0.66	Cat.	0.00	0.64	0.78	Cat.	0.00	0.64	0.78
Real	0.32	0.18	1.00	Real	0.34	0.24	0.92	Real	0.28	0.32	0.90
Att.	0.36	0.52	1.00	Att.	0.04	0.64	0.96	Att.	0.12	0.32	0.44
Cat.	0.00	0.50	0.92	Cat.	0.00	0.50	0.92	Cat.	0.22	0.58	0.64
Real	0.24	0.34	0.92	Real	0.34	0.26	0.90	Real	0.18	0.36	0.94
Att.	0.32	0.56	0.84	Att.	0.00	0.84	0.96	Att.	0.56	0.88	0.88
Cat.	0.00	0.60	0.82	Cat.	0.02	0.50	0.90	Cat.	0.00	0.58	0.84
Real	0.50	0.10	0.90	Real	0.58	0.24	0.68	Real	0.32	0.32	0.86
Att.	0.16	0.32	0.88	Att.	0.12	0.96	0.84	Att.	0.04	0.56	0.44
Cat.	0.06	0.42	0.94	Cat.	0.00	0.68	0.74	Cat.	0.24	0.52	0.66
Real	0.24	0.28	0.98	Real	0.22	0.40	0.90	Real	0.26	0.34	0.90
Att.	0.16	0.16	0.68	Att.	0.20	0.72	0.56	Att.	0.28	0.44	0.84
Cat.	0.00	0.50	0.92	Cat.	0.02	0.72	0.68	Cat.	0.02	0.56	0.84
Real	0.48	0.28	0.74	Real	0.02	0.58	0.90	Real	0.40	0.22	0.90
Att.	0.04	0.80	0.64	Att.	0.00	0.80	1.00	Att.	0.36	0.48	0.96
Cat.	0.10	0.66	0.66	Cat.	0.00	0.56	0.86	Cat.	0.00	0.48	0.94
Real	0.44	0.32	0.74	Real	0.44	0.22	0.84	Real	0.56	0.22	0.74
Att.	0.16	0.68	0.44	Att.	0.32	0.44	0.44	Att.	0.04	0.28	1.00
Cat.	0.02	0.78	0.60	Cat.	0.06	0.64	0.78	Cat.	0.02	0.44	0.94

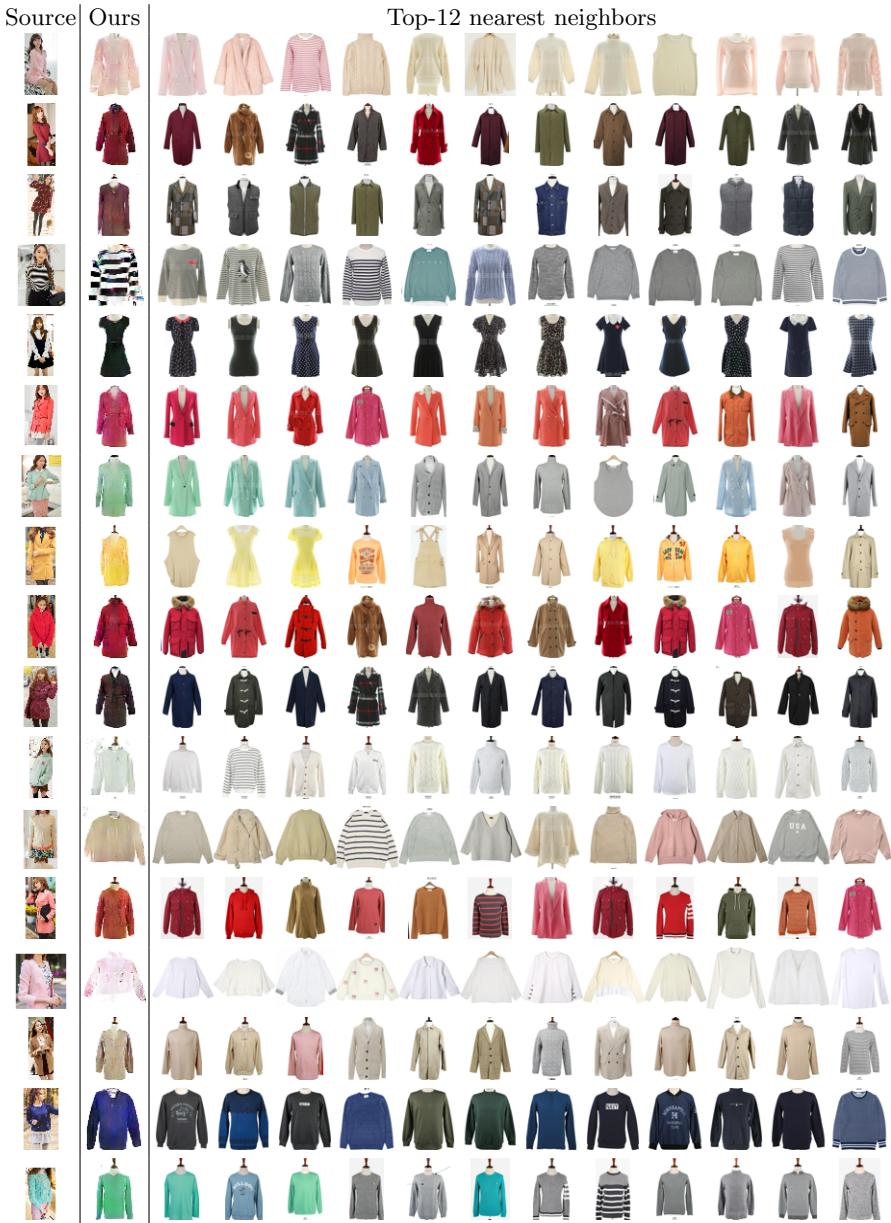
Continued (4/5)

Source	RF	MSE	Ours	Source	RF	MSE	Ours	Source	RF	MSE	Ours
											
Real	0.36	0.28	0.84	Real	0.42	0.28	0.78	Real	0.48	0.26	0.76
Att.	0.04	0.84	0.84	Att.	0.00	0.80	0.64	Att.	0.20	0.68	0.52
Cat.	0.00	0.60	0.82	Cat.	0.00	0.86	0.56	Cat.	0.00	0.60	0.82
											
Real	0.18	0.34	0.98	Real	0.40	0.22	0.90	Real	0.40	0.18	0.92
Att.	0.04	0.72	0.96	Att.	0.28	0.84	0.84	Att.	0.36	0.68	0.96
Cat.	0.00	0.50	0.92	Cat.	0.00	0.50	0.92	Cat.	0.02	0.52	0.86
											
Real	0.44	0.18	0.86	Real	0.48	0.18	0.84	Real	0.34	0.26	0.90
Att.	0.48	0.36	0.96	Att.	0.12	0.68	0.84	Att.	0.36	0.56	0.44
Cat.	0.02	0.48	0.92	Cat.	0.00	0.48	0.94	Cat.	0.02	0.74	0.66
											
Real	0.08	0.50	0.92								
Att.	0.48	0.88	0.80								
Cat.	0.00	0.60	0.82								

End (5/5)

3 Image Retrieval

We show the image retrieval in the down-sampled pixel space $\subset \mathbb{R}^{16 \times 16 \times 3}$ to demonstrate that our converter is *not simply memorizing and copying* the training images. Each of our generated images in Fig. 4 of the submitted paper is a query, and we measure l_2 -distances between the query and all product images.



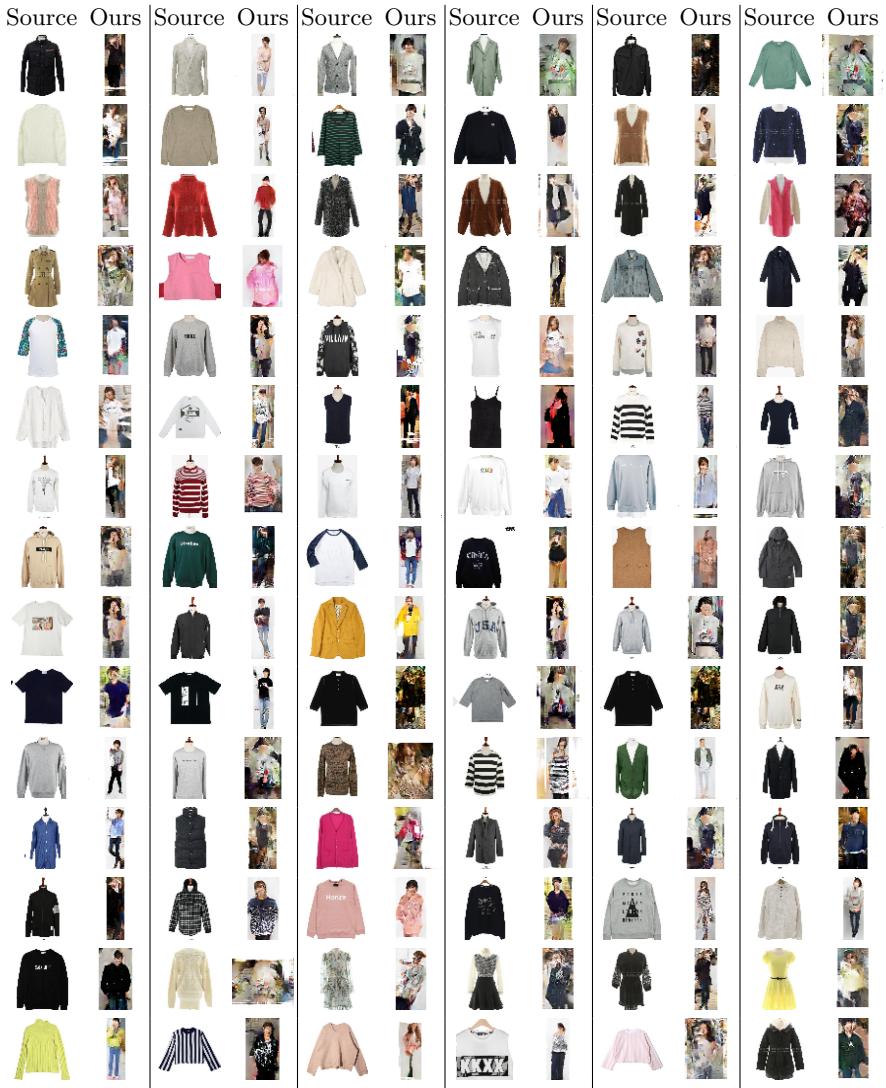


Continued (2/3)

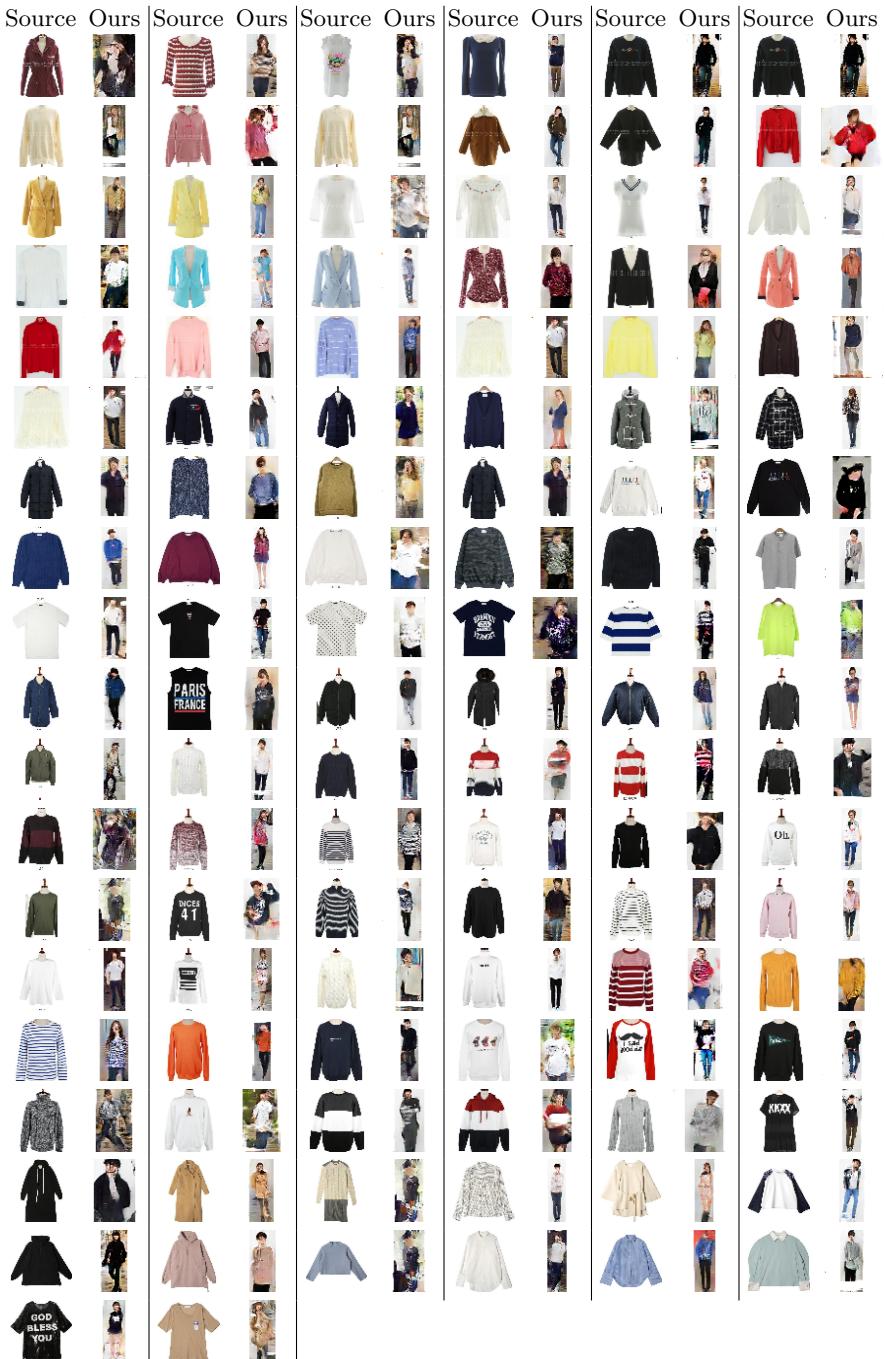


4 From Products to Fashion Model Images

We also provide results of the *inverse setting of domains*, where the source domain has products and the target domain has fashion models. Since generating fashion models is a more complex task, we found that 65 epochs for initial training and 5 more epochs for fine-tuning are required in our experiment. All the other details are exactly identical to those of the original setting. We show 200 examples *randomly chosen* from test set of LookBook as follows.



Continued (1/2)



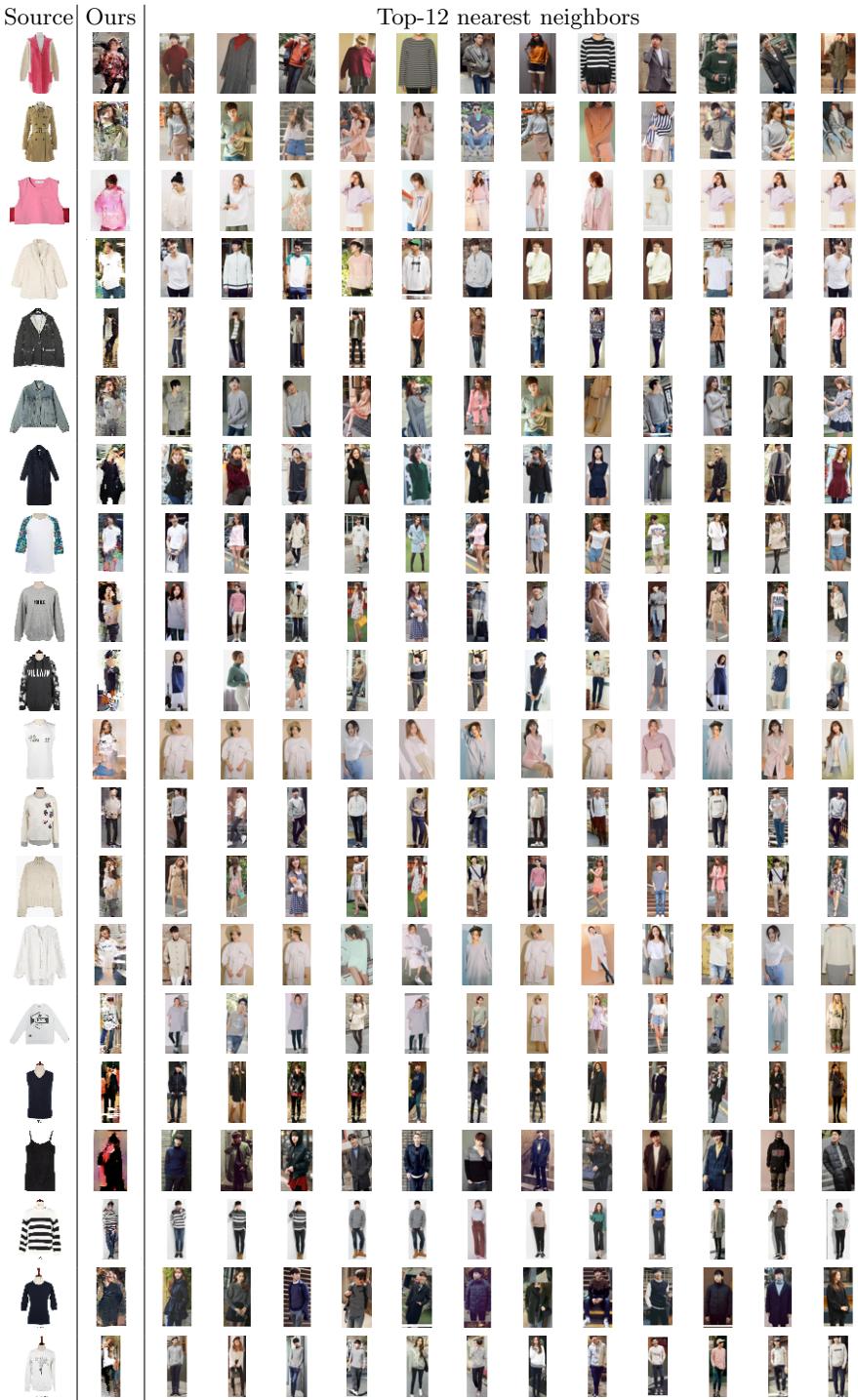
End (2/2)

Here, again we show the image retrieval in the down-sampled pixel space $\subset \mathbb{R}^{16 \times 16 \times 3}$ to demonstrate that our converter is ***not simply memorizing and copying*** the training images. Each of our randomly chosen results is a query, and we measure l_2 -distances between the query and all the fashion model images.



Continued (1/3)

Source | Ours



Continued (2/3)

References

1. Yoo, D., Kim, N., Park, S., Paek, A.S., Kweon, I.S.: Pixel-level domain transfer. In: Proceedings of the 14th European Conference on Computer Vision. (2016)
2. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems. (2015) 91–99
3. Canny, J.: A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6) (1986) 679–698