

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107

# Learning to Cartoonize Using White-box Cartoon Representations

## Supplementary Material

Anonymous CVPR submission

Paper ID 6791

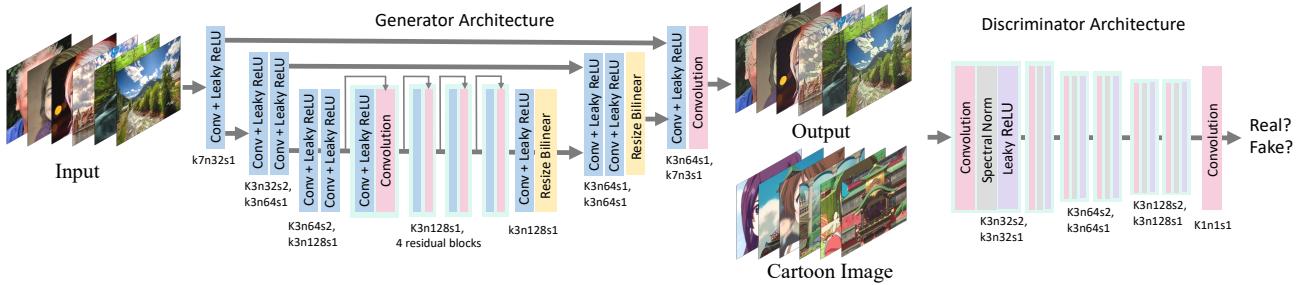


Figure 0. The architecture of generator network and discriminator network.

In this supplementary material, we show more experimental results, including the architecture of generator network and discriminator network, illustration of our method in different scenes with different style, comparison between our results and cartoons in the same scenes, and examples used in the user study.

## 1. Network Architecture

We show the architecture of generator network and discriminator network in Figure 0. The generator network is a fully-convolutional network, it consists of only three kind of layers: convolution, Leaky ReLU (LReLU) [2] and bilinear-resize layers. This enables it to be easily embedded in edge devices such as mobile phones. PatchGAN [1] is adapted in the discriminator network, where the last layer is a convolution layer. Each pixel in the output feature map correspond to a patch in the input image, with the size equals to the perceptive field, and is used to judge whether the patch belongs cartoon images or generated images. Spectral normalization [3] is placed after every convolution layer (except the last one) to enforce Lipschitz constrain on the network and stabilize training.

## 2. Results in Different Use Cases

In Section 4.1 of the main paper, we apply our method on different scenes and show the cartoonized results. Due to the limitation of space, Only 12 pairs of examples with small

resolution are shown. Here we collect images from more use cases with higher resolution, and show the results generated by our method. The content of images includes male celebrities (shown in Figure 1), female celebrities (shown in Figure 2), food (shown in Figure 3), sceneries (shown in Figure 4 and Figure 5), indoors (shown in Figure 6), city views (shown in Figure 7 and Figure 8), and other objects (shown in Figure 9). Overall, the above-shown results demonstrate that our method can generate high-quality cartoonized images, and can be applied on diverse use cases and real-world scenes.

## 3. Comparison with Cartoon Images

We collect several images from Shinkai Makoto's cartoon films and their counterpart real-world photos that are taken in the same scenes. We then apply our method on the collected real-world photos, and show the comparison of our results and cartoon images in Figure 10 and Figure 11. Note that the images are randomly placed without class tags, and are supposed to be distinguished by the readers.

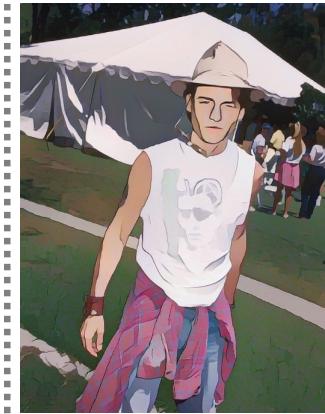
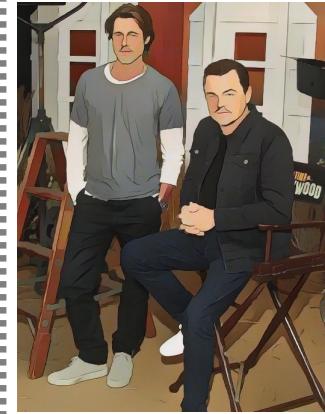
## 4. Images Shown in the User Study

In the user study, we collect 30 images and process them with Fast Neural Style ??, CycleGAN ??, CartoonGAN ?? and our methods. 10 candidate are asked to score each image from 1 to 5, and the results of user study are shown in the Section 4.4 in the main paper. Here, we show all 30

108 groups of images processed by different methods. Note that 162  
109 CartoonGAN has 4 styles, and we show them all in the user 163  
110 study, and calculate the average score of the four styles as 164  
111 the final score for CartoonGAN. 165  
112

## 113 References 167

- 114
- [1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 168  
115 Image-to-image translation with conditional adversarial net- 169  
116 works. In *Proceedings of the IEEE conference on computer 170  
117 vision and pattern recognition*, pages 1125–1134, 2017. 1 171  
118
- [2] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier 172  
119 nonlinearities improve neural network acoustic models. In 173  
120 *Proc. icml*, volume 30, page 3, 2013. 1 174  
121
- [3] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and 175  
122 Yuichi Yoshida. Spectral normalization for generative adver- 176  
123 sarial networks. *arXiv preprint arXiv:1802.05957*, 2018. 1 177  
124
- 125
- 126
- 127
- 128
- 129
- 130
- 131
- 132
- 133
- 134
- 135
- 136
- 137
- 138
- 139
- 140
- 141
- 142
- 143
- 144
- 145
- 146
- 147
- 148
- 149
- 150
- 151
- 152
- 153
- 154
- 155
- 156
- 157
- 158
- 159
- 160
- 161

216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323

Real-world photo

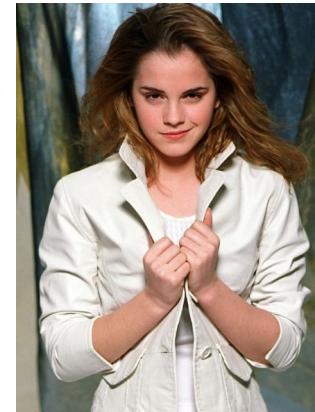
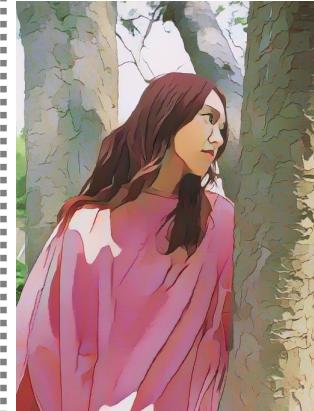
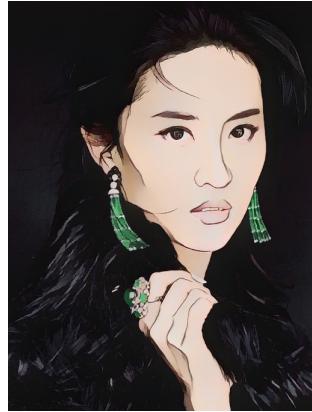
Cartoonized result

Real-world photo

Cartoonized result

Figure 1. Cartoonized male Celebrities.

CVPR 2020 Submission #6791. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431

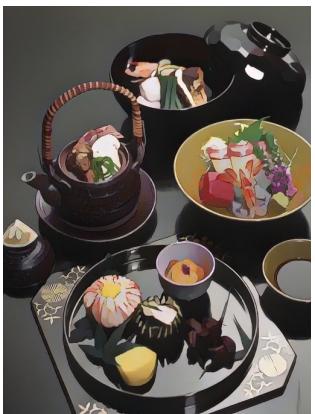
Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

Figure 2. Cartoonized female Celebrities.

432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539

Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

Figure 3. Cartoonized food.

540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593



Real-world photo



Cartoonized result

594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647

Figure 4. Cartoonized scenery.

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

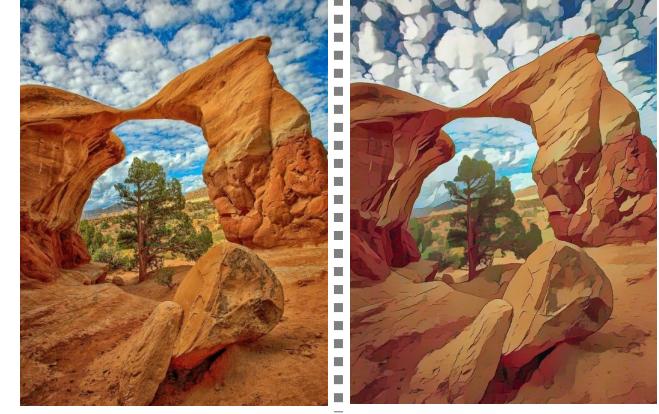
697

698

699

700

701



Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

Figure 5. Cartoonized scenery.

756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863

Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

Figure 6. Cartoonized indoor scenes.

864



865



866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971



Real-world photo

Cartoonized result

Figure 7. Cartoonized city scenes.

972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
10101011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
10231024  
1025

Real-world photo

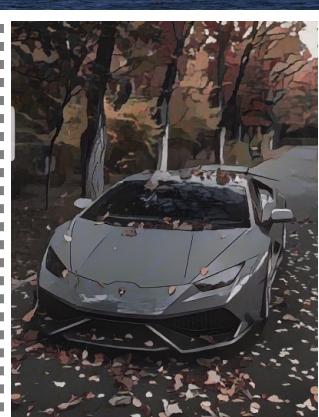
Cartoonized result

Real-world photo

Cartoonized result

Figure 8. Cartoonized city scenes.

1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133



Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187

Figure 9. Cartoonized different objects.

1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200



1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212



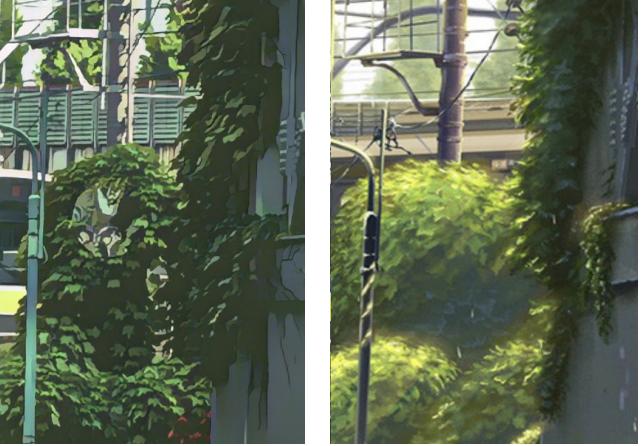
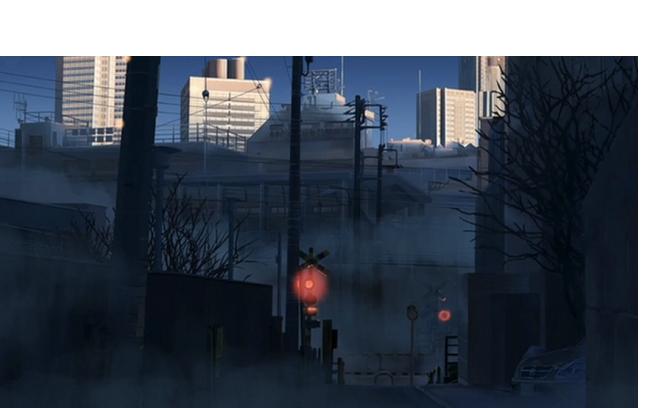
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224



1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238



1239  
1240  
1241



1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254

1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266

1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278

1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295

Figure 10. Cartoons and results of our method in the same scene. They are randomly placed and supposed to be distinguished by the readers.

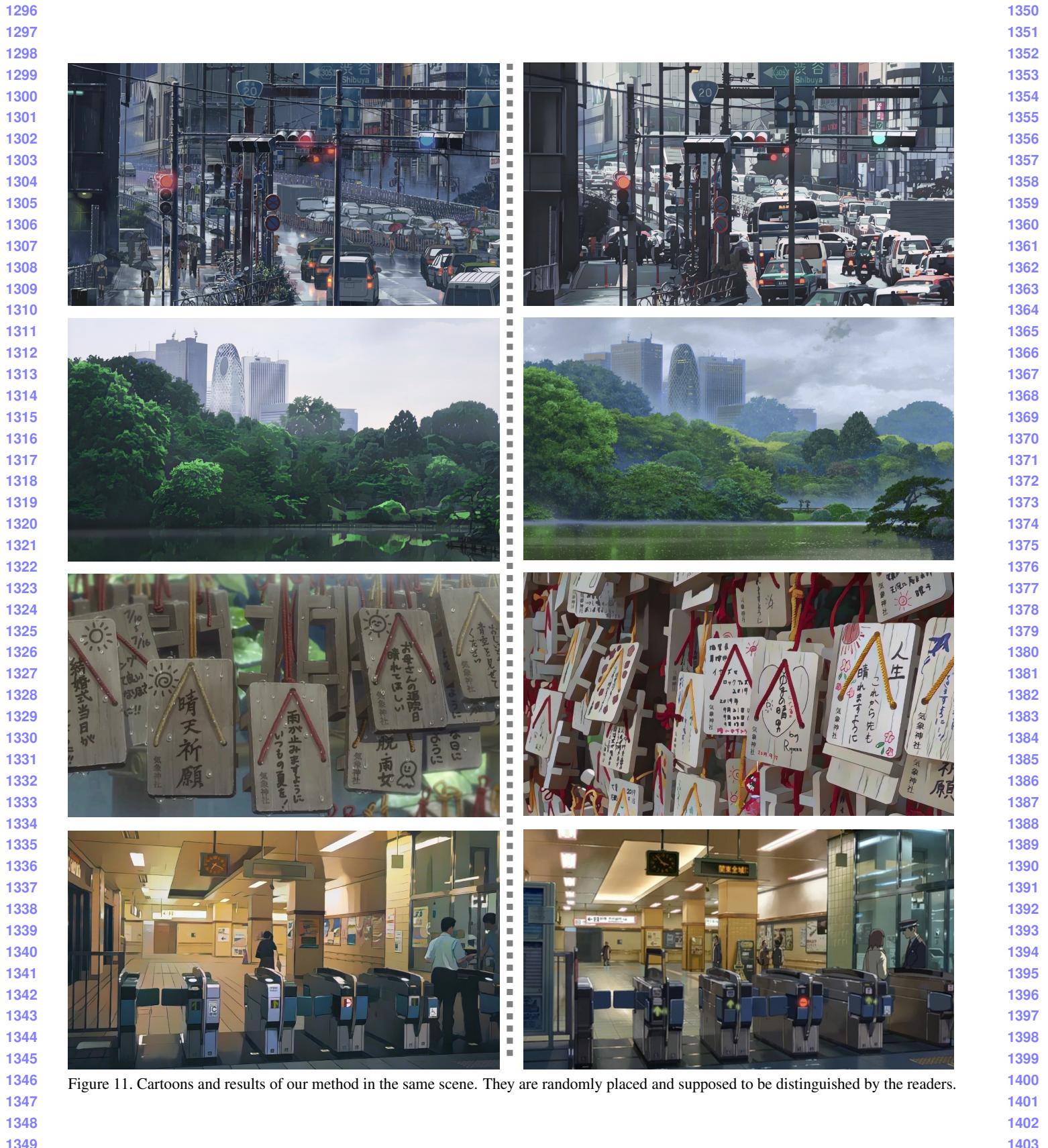


Figure 11. Cartoons and results of our method in the same scene. They are randomly placed and supposed to be distinguished by the readers.

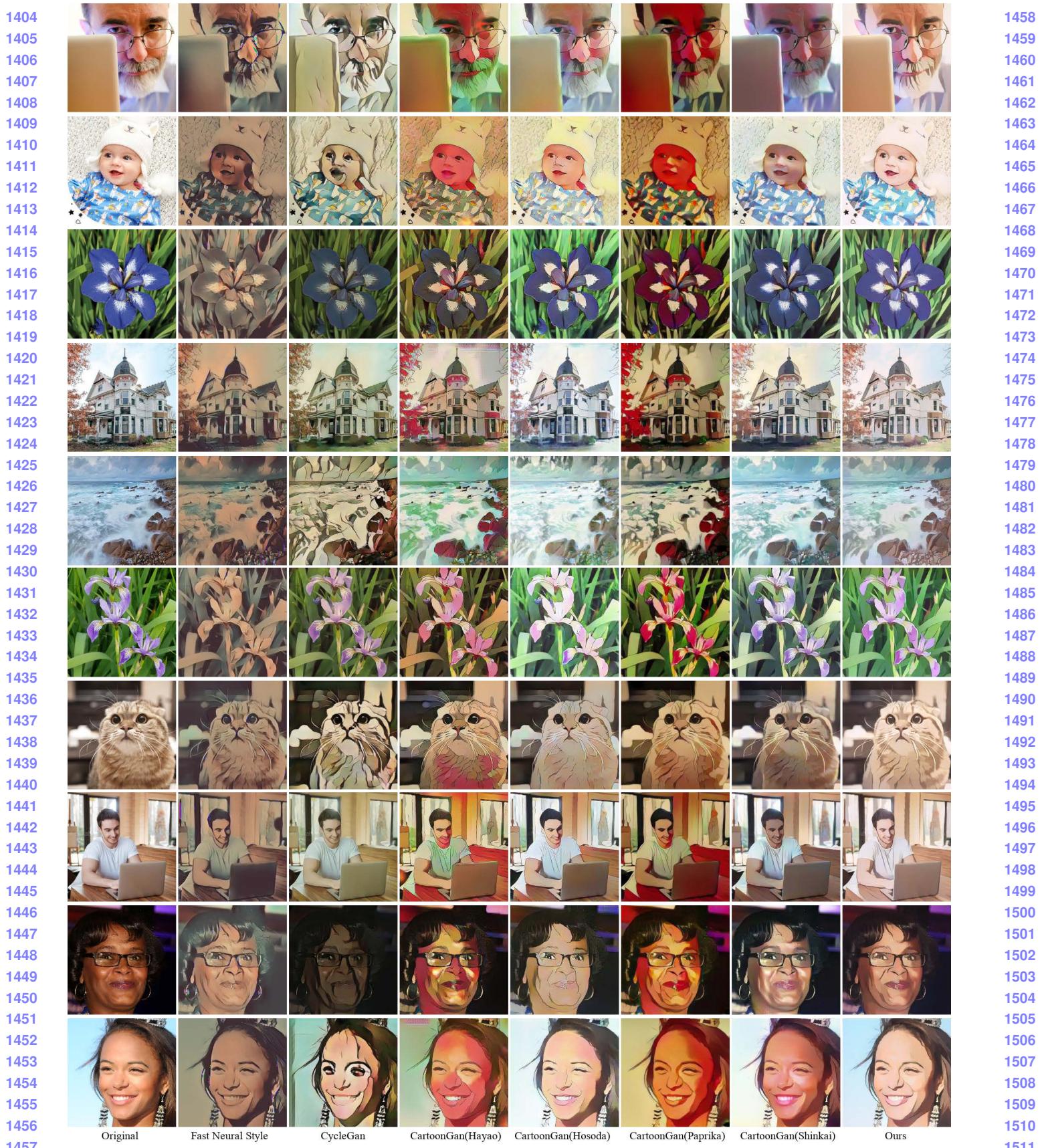


Figure 12. Images shown in the user study.

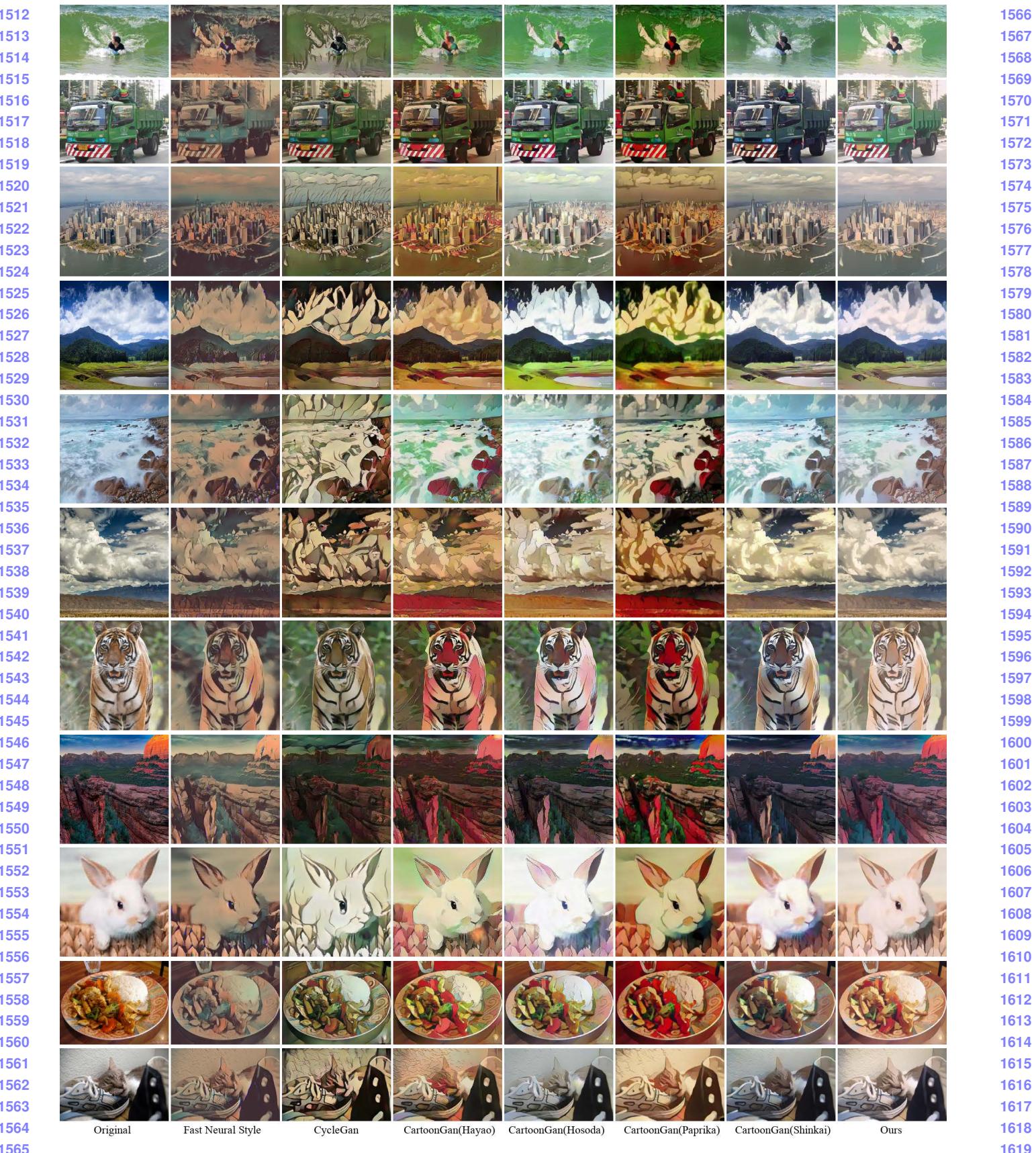


Figure 13. Images shown in the user study.

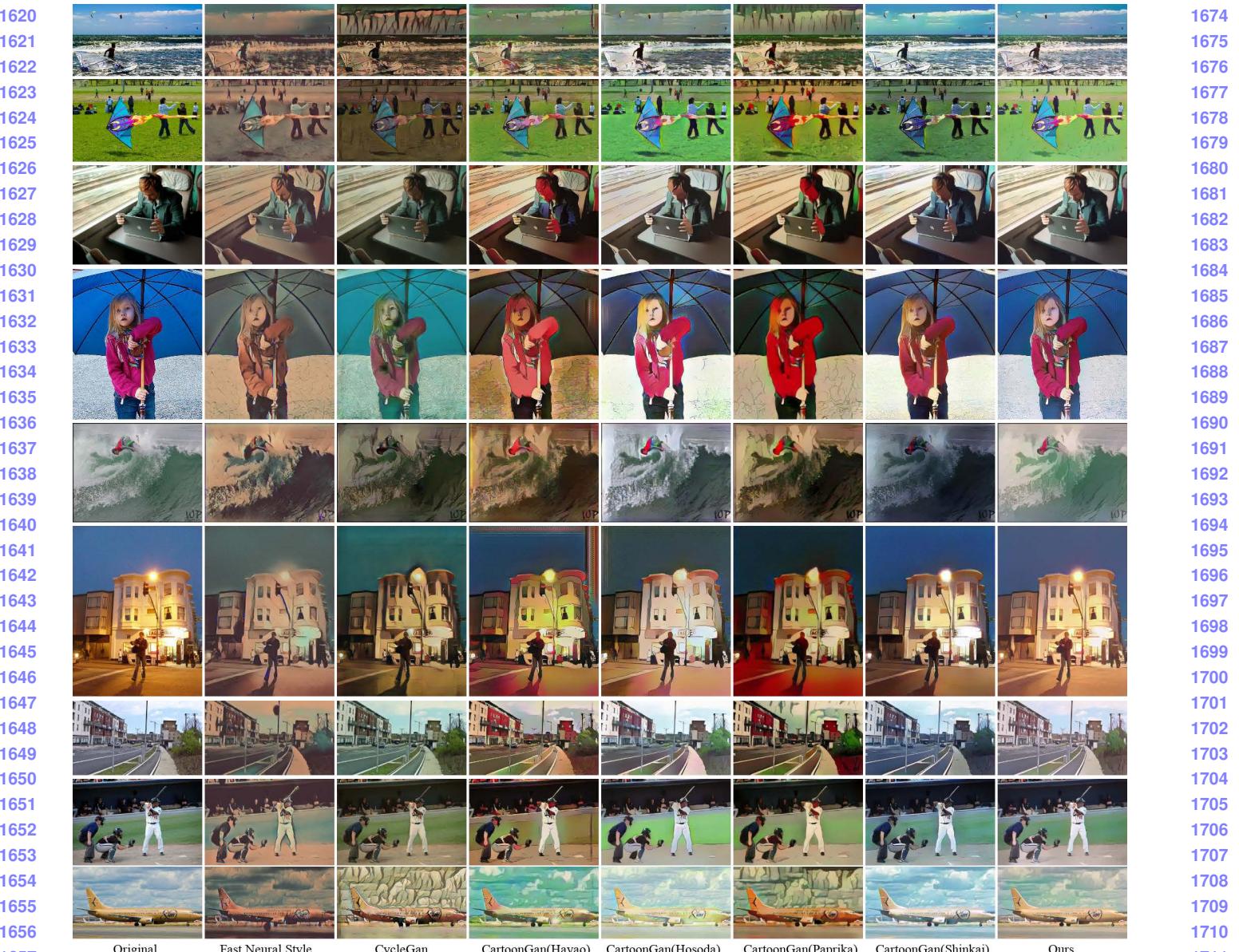


Figure 14. Images shown in the user study.