

000	054
001	055
002	056
003	057
004	058
005	059
006	060
007	061
008	062
009	063
010	064
011	065
012	066
013	067
014	068
015	069
016	070
017	071
018	072
019	073
020	074
021	075
022	076
023	077
024	078
025	079
026	080
027	081
028	082
029	083
030	084
031	085
032	086
033	087
034	088
035	089
036	090
037	091
038	092
039	093
040	094
041	095
042	096
043	097
044	098
045	099
046	100
047	101
048	102
049	103
050	104
051	105
052	106
053	107

Learning to Cartoonize Using White-box Cartoon Representations

Supplementary Material

Anonymous CVPR submission

Paper ID 6791

1. Overview

In this supplementary material, we show more experimental results, including the architecture of generator network and discriminator network, the influence of different loss functions in generative adversarial networks (GANs), illustration of our method in different scenes with different style, and examples used in the user study.

2. Network Architecture

We show the architecture of generator network and discriminator network in Figure 1. The generator network is a fully-convolutional network consisted of only convolution, activation and bilinear-resize layers, which 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 [2] is placed after every convolution layer (except the last one) to enforce the Lipschitz constrain on the network and stabilize training.

3. Influence of Different GAN loss

In our proposed framework, least square GAN (LSGAN) loss [?] is used for adversarial training. We also tested the vanilla GAN loss [?], Wasserstein GAN loss with gradient-penalty [?], and LSGAN loss with spectral norm removed in discriminator. The results of different gan loss are shown in Figure ??.

4. Influence of Different GAN loss

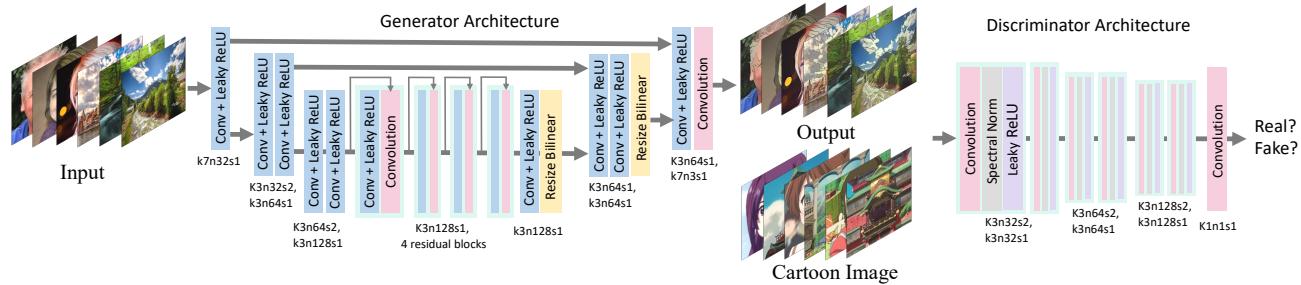
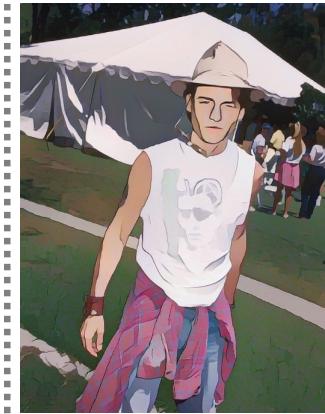
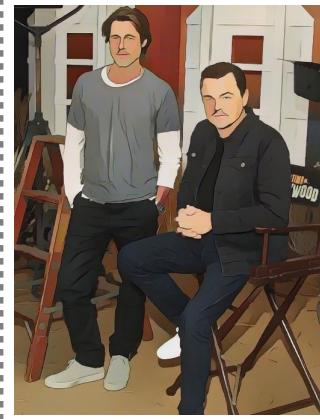
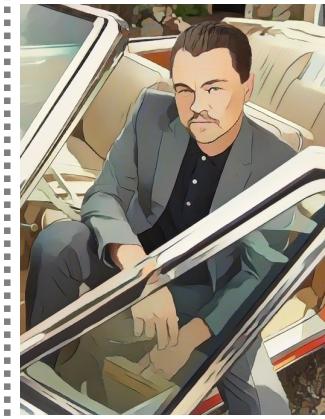


Figure 1. The architecture of generator network and discriminator network.

108	162
109	163
110	164
111	165
112	166
113	167
114	168
115	169
116	170
117	171
118	172
119	173
120	174
121	175
122	176
123	177
124	178
125	179
126	180
127	181
128	182
129	183
130	184
131	185
132	186
133	187
134	188
135	189
136	190
137	191
138	192
139	193
140	194
141	195
142	196
143	197
144	198
145	199
146	200
147	201
148	202
149	203
150	204
151	205
152	206
153	207
154	208
155	209
156	210
157	211
158	212
159	213
160	214
161	215

216
217
218
219
220
221
222
223
224
225
226
227
228270
271
272
273
274
275
276
277
278
279
280
281
282229
230
231
232
233
234
235
236
237
238
239
240
241283
284
285
286
287
288
289
290
291
292
293
294
295242
243
244
245
246
247
248
249
250
251
252
253
254
255296
297
298
299
300
301
302
303
304
305
306
307
308
309256
257
258
259
260
261
262
263
264
265
266
267
268
269310
311
312
313
314
315
316
317
318
319
320
321
322
323

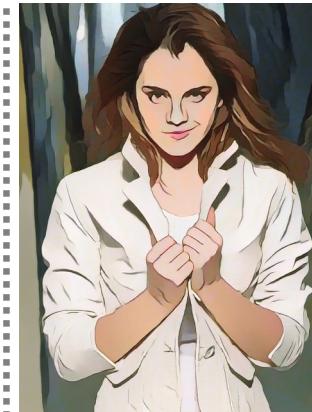
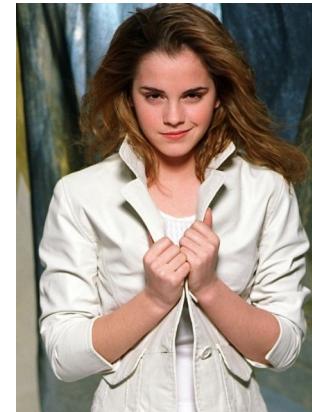
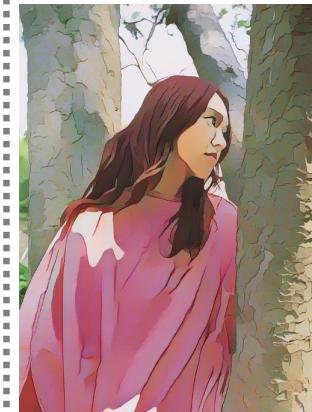
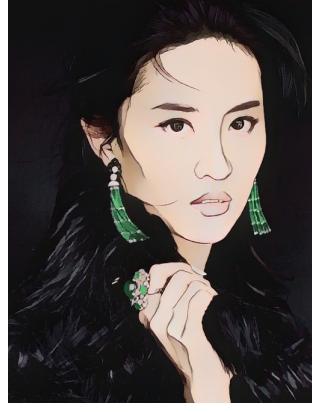
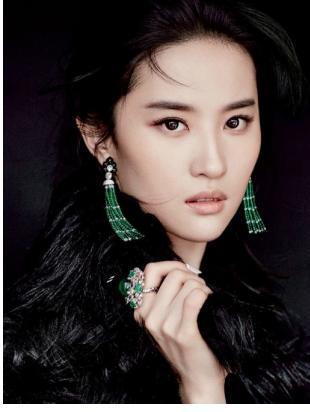
Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

Figure 2. Cartoonized male Celebrities.

324
325
326
327
328
329
330
331
332
333
334
335
336378
379
380
381
382
383
384
385
386
387
388
389
390337
338
339
340
341
342
343
344
345
346
347
348
349391
392
393
394
395
396
397
398
399
400
401
402
403350
351
352
353
354
355
356
357
358
359
360
361
362
363404
405
406
407
408
409
410
411
412
413
414
415
416
417364
365
366
367
368
369
370
371
372
373
374
375
376
377418
419
420
421
422
423
424
425
426
427
428
429
430
431

Real-world photo

Cartoonized result

Real-world photo

Cartoonized result

Figure 3. Cartoonized female Celebrities.

432
433
434
435
436
437
438
439
440
441
442
443
444



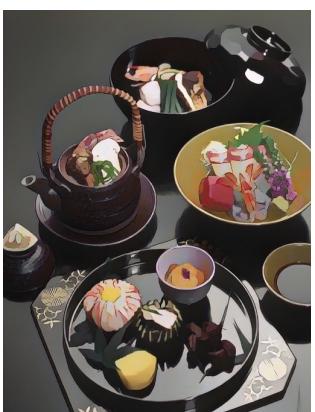
486
487
488
489
490
491
492
493
494
495
496
497
498

445
446
447
448
449
450
451
452
453
454
455
456
457



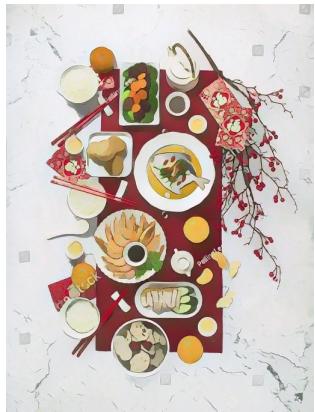
499
500
501
502
503
504
505
506
507
508
509
510
511

458
459
460
461
462
463
464
465
466
467
468
469
470
471



512
513
514
515
516
517
518
519
520
521
522
523
524
525

472
473
474
475
476
477
478
479
480
481
482
483
484
485



526
527
528
529
530
531
532
533
534
535
536
537
538
539

Real-world photo

Cartoonized result

Real-world photo

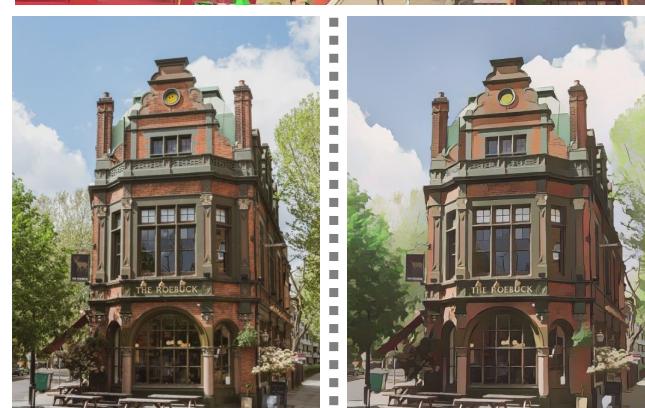
Cartoonized result

Figure 4. Cartoonized food.

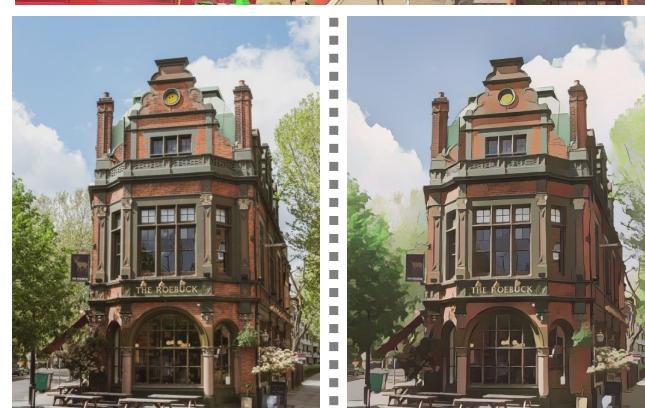
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593



Real-world photo



Cartoonized result



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 5. 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

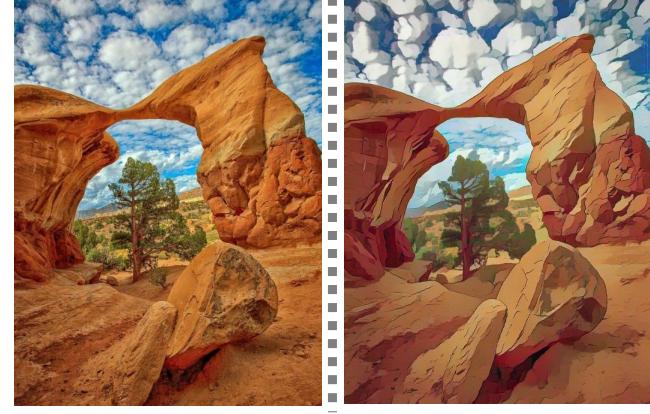
699

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 6. 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 7. 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 8. Cartoonized city scenes.

972
973
974
975
976
977
978
979
980
981
982
983
9841026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038985
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 9. Cartoonized city scenes.

1080	References	1134
1081		1135
1082	[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial net- works. In <i>Proceedings of the IEEE conference on computer vision and pattern recognition</i> , pages 1125–1134, 2017. 1	1136
1083		1137
1084		1138
1085		1139
1086	[2] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adver- sarial networks. <i>arXiv preprint arXiv:1802.05957</i> , 2018. 1	1140
1087		1141
1088		1142
1089		1143
1090		1144
1091		1145
1092		1146
1093		1147
1094		1148
1095		1149
1096		1150
1097		1151
1098		1152
1099		1153
1100		1154
1101		1155
1102		1156
1103		1157
1104		1158
1105		1159
1106		1160
1107		1161
1108		1162
1109		1163
1110		1164
1111		1165
1112		1166
1113		1167
1114		1168
1115		1169
1116		1170
1117		1171
1118		1172
1119		1173
1120		1174
1121		1175
1122		1176
1123		1177
1124		1178
1125		1179
1126		1180
1127		1181
1128		1182
1129		1183
1130		1184
1131		1185
1132		1186
1133		1187