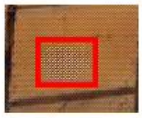
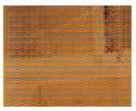
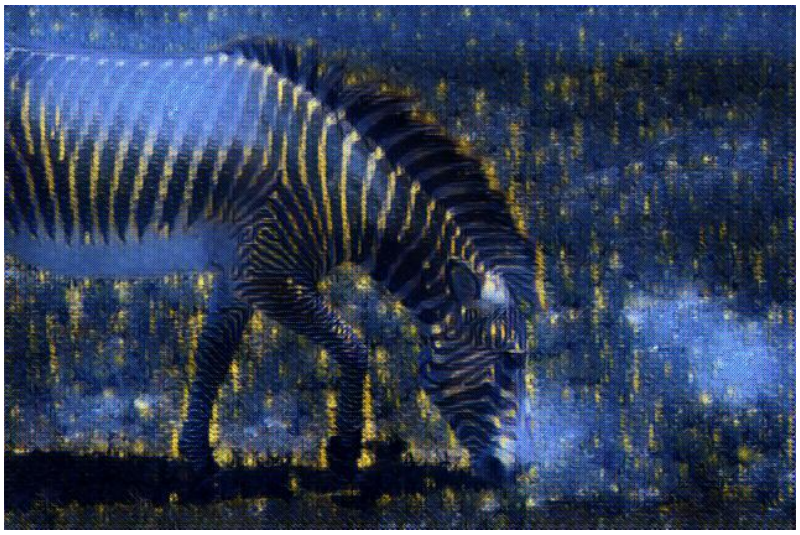
User study on the results of style transfer

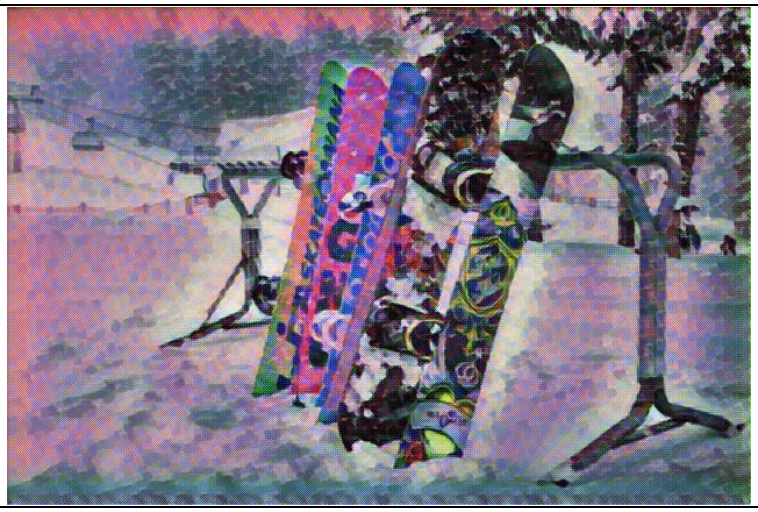
Style transfer is a hot topic in the fields of multimedia, image processing and computer graphics. The ideal style transfer should transfer image with the style of style image while maintaining the content consistency with original image. We have prepared a series of results from baseline method and our method. Please carefully compare the results following specific standards below as your opinions will be used to evaluate the methods. There are four comparison items, each comparison item involves four comparison groups. Thank you for your cooperation.

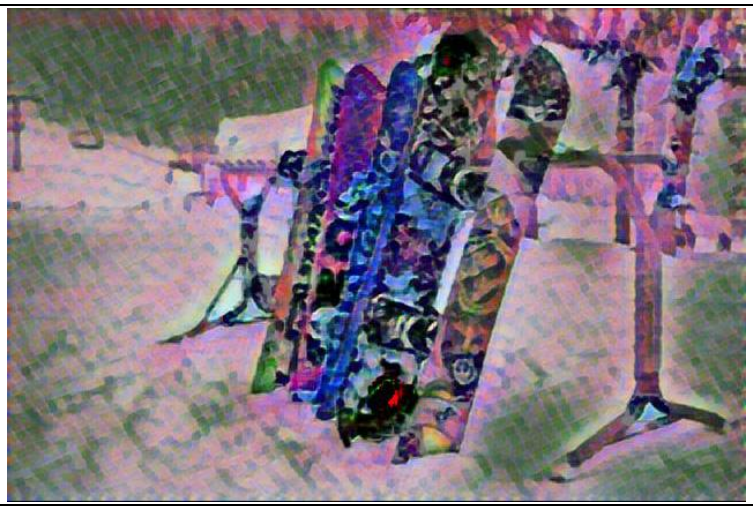
1.Image quality

Compare item	Image quality
Compare standard	Please follow standards below to compare results, and give your preference in the lower right column. Note: When comparing this item, please enlarge the image to compare as the noise will be more obvious after zoomed in Standards: (1) There is no discordant texture (noise) at details. (2) The picture does not have checkerboard effects. (3) There is no line-like noise in the images We have provided some noise examples for your reference, including but not limited to these.

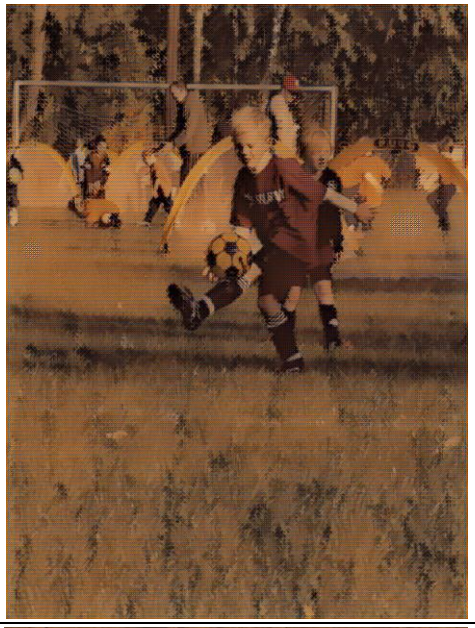
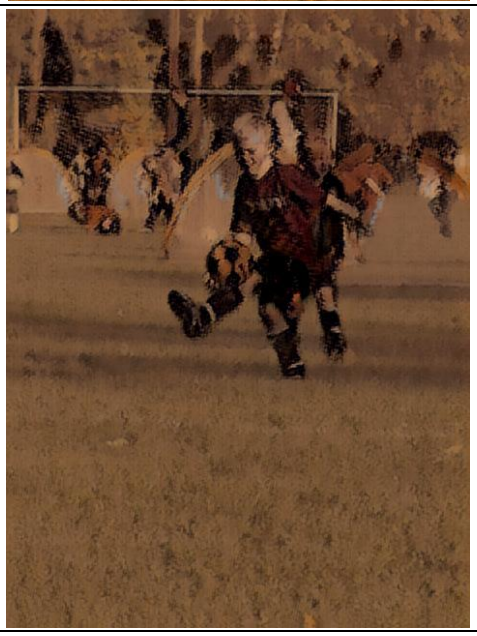
	Mark √ to the prefer ones in the last column	
examples	 checkerboard noise checkerboard noise  global line-like noise	
Group 1		Preference
A		
B		
Group 2		Preference

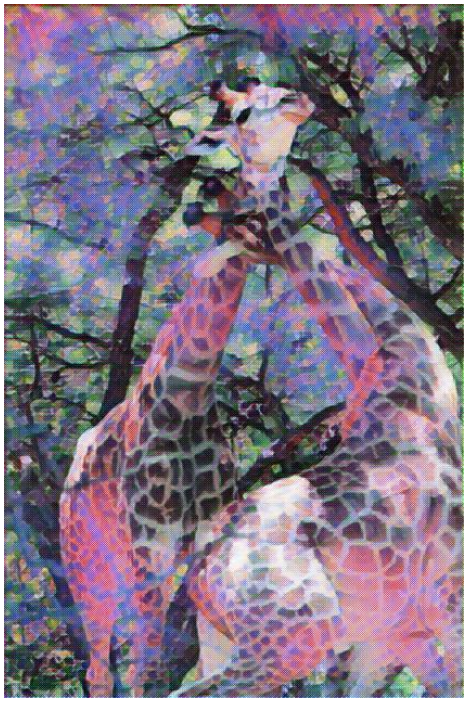
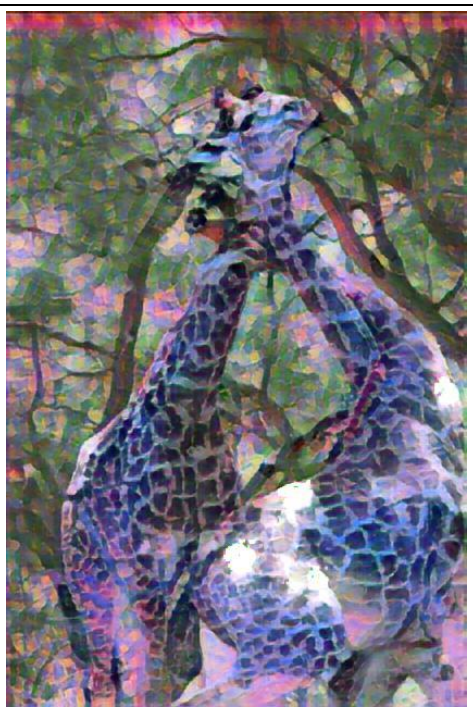
A		
B		
Group 3		Preference

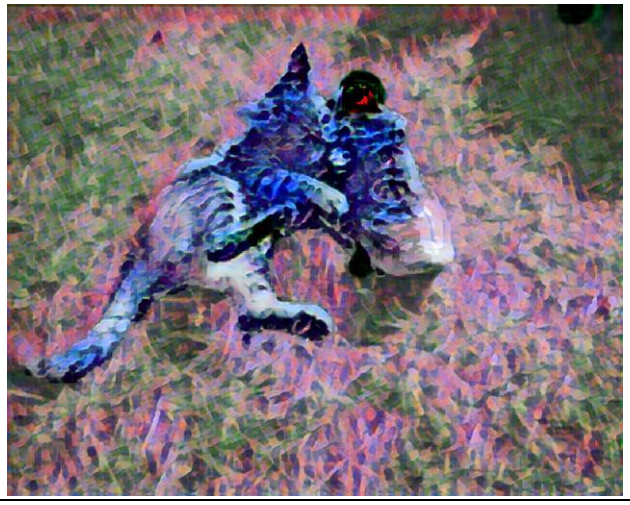
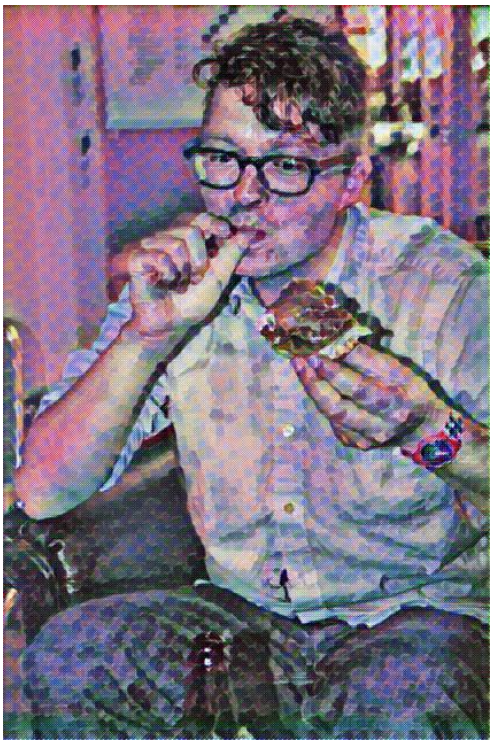
A		
B		
Group 4		Preference
A		

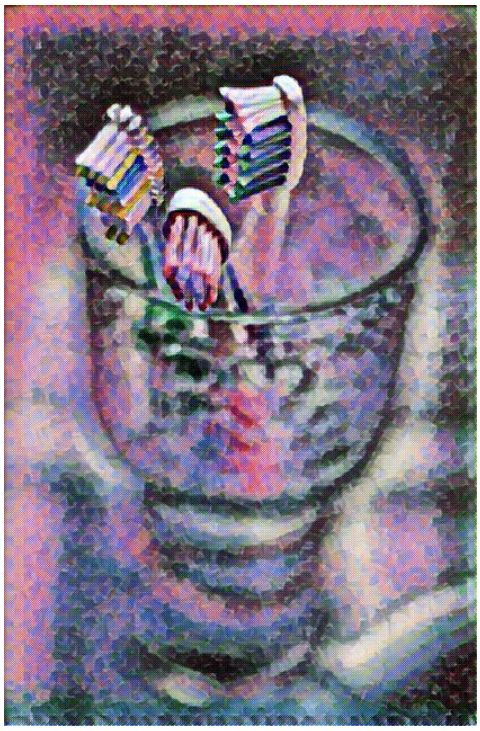
B		
Group 5		Preference
A		
B		
Group 6		Preference

A		
B		
Group 7		Preference

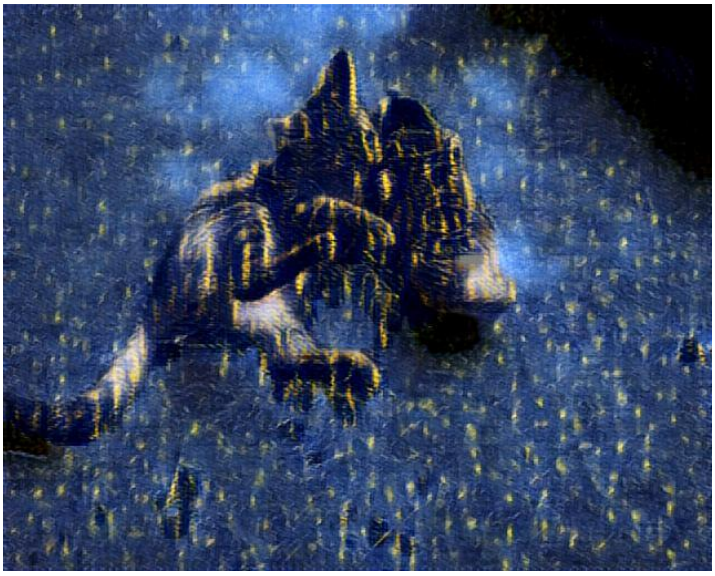
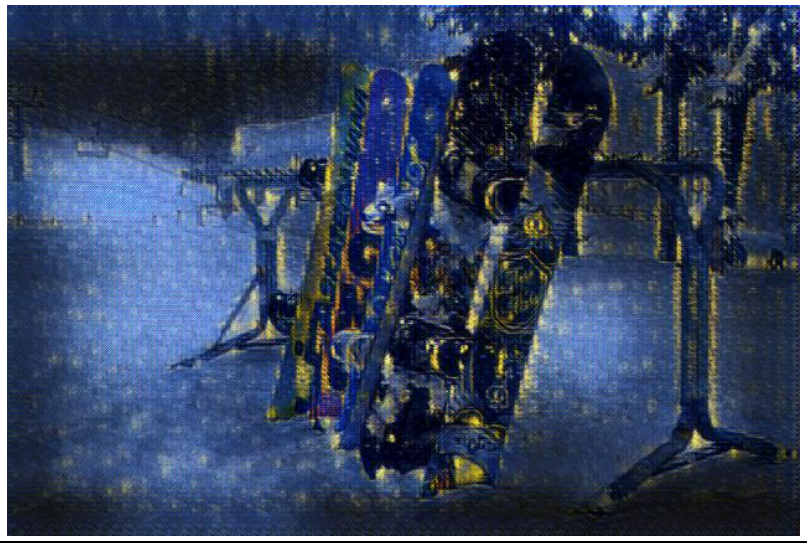
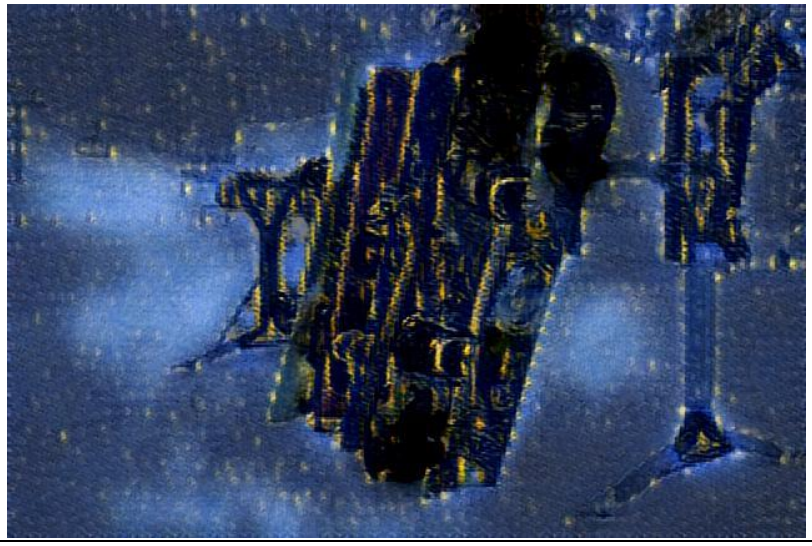
A		
B		
Group 8		Preference

A		
B		
Group 9		Preference

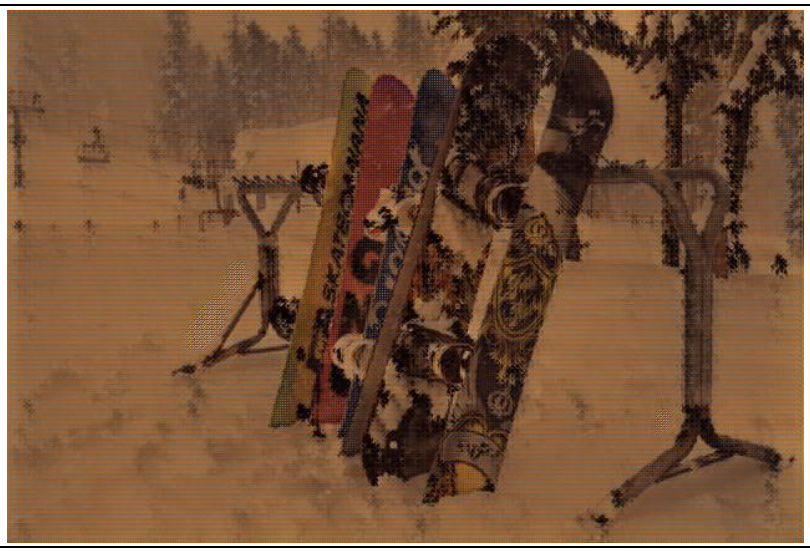
A		
B		
Group 10		Preference
A		

B		
Group 11		Preference
A		

B		
Group 12		Preference
A		

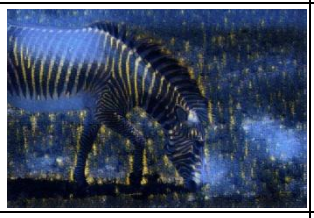
B		
Group 13		Preference
A		
B		
Group 14		Preference

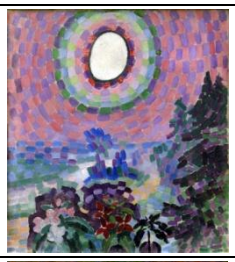
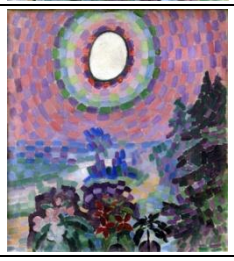
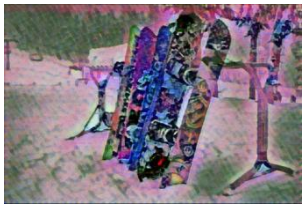
A		
B		
Group 15		Preference

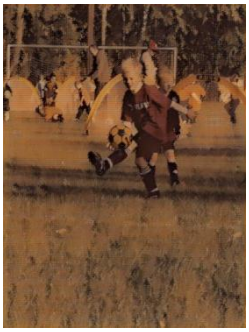
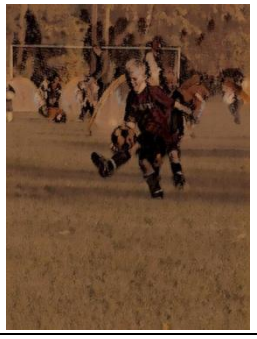
A		
B		

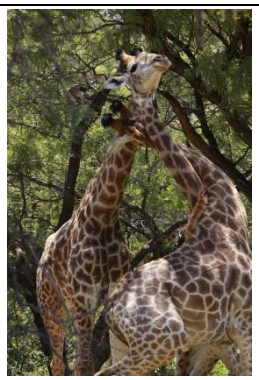
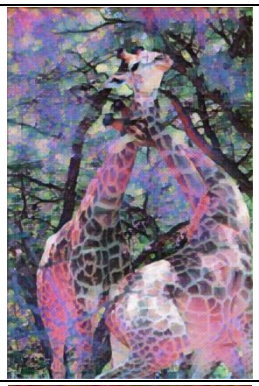
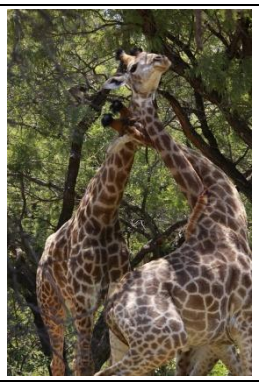
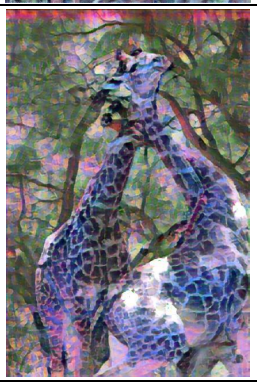
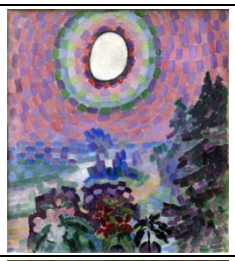
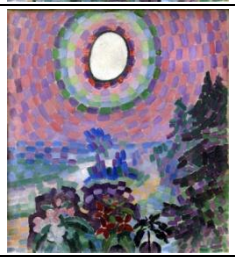
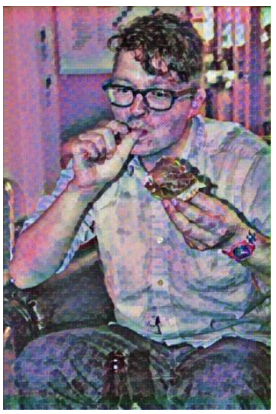
2. Saliency order preservation

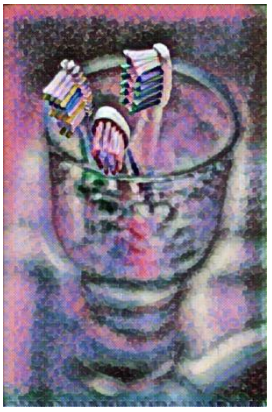
Compare item	Saliency order preservation
Compare standard	Please follow the standards below to compare images in the table, the content images and the stylized image according to the style image, and select the images that you would prefer in each group. Standards: (1) The style images and the stylized images in the table have

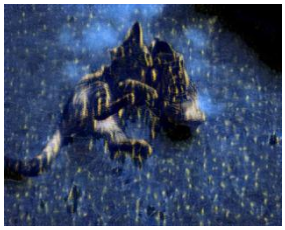
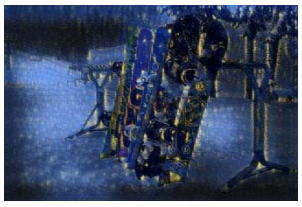
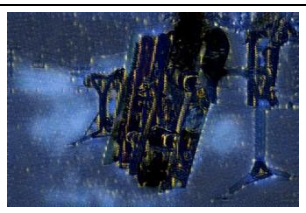
	the same emphasis on important objects. For example, the core color of the style image is transformed to the core elements of the result image. The non-core parts have the colors and textures that correspond to non-core areas in the style image. (2) Core elements can be highlighted by the color and textures distribution. Mark √ to the prefer ones in the last column			
Style image		Content image	Stylized image	
Group 1				Preference
A				
B				
Group 2				Preference
A				
B				
Group 3				Preference

A				
B				
Group 4				Preference
A				
B				
Group 5				Preference
A				
B				
Group 6				Preference

A				
B				
Group 7				Preference
A				
B				
Group 8				Preference

A				
B				
Group 9				Preference
A				
B				
Group 10				Preference
A				

B				
Group 11				Preference
A				
B				
Group 12				Preference
A				

B				
Group 13				Preference
A				
B				
Group 14				Preference
A				
B				
Group 15				Preference
A				

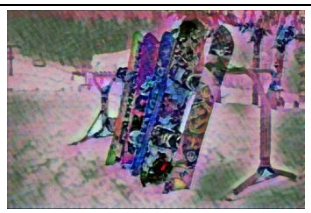
B				

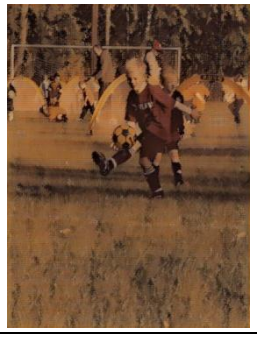
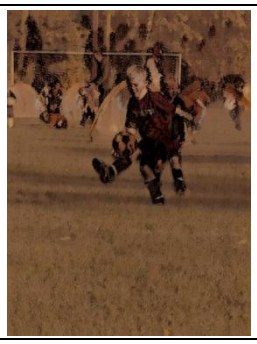
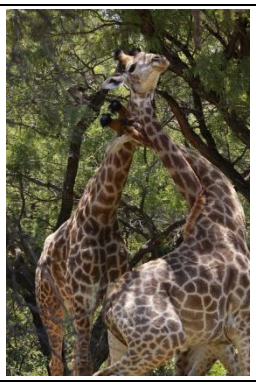
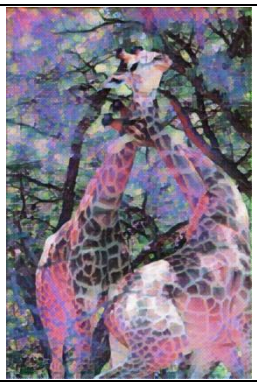
3. Image content preservation

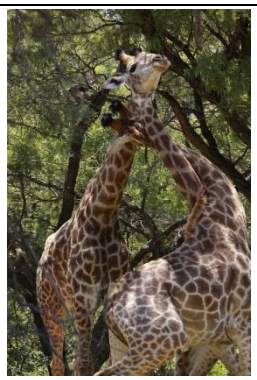
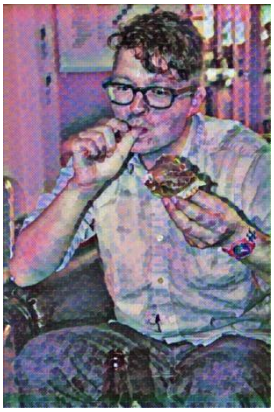
Compare item	Image content preservation
Compare standard	Please compare the content images in the table with the stylized image according to the following standards, and select the result that you would prefer in each group. 1. The content image and the stylized image have the same content according to visual judgement. We provided some examples for reference. Mark √ to the prefer ones in the last column
examples	 This is an example of content being destroyed. The 3rd and 4th images from left to right are generated according to the content of 2nd image. The style image of the 3rd image is the 1st image. However, the human eyes in the 3rd image are enlarged, which makes the portrait looks more like the person in the 1st image, rather than the person in the 2nd image. Although

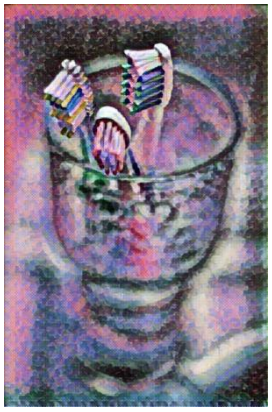
	the 4th image is also generated according to the content of the 2nd image, it is almost impossible to recognize the original content from the 4th image. Therefore, the preference of the 3rd ,4th images will be relatively low and the preference of 4th image should be lower than that of the 3rd image.

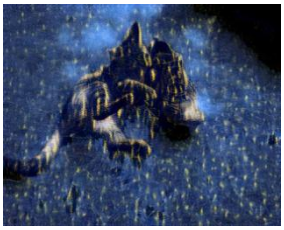
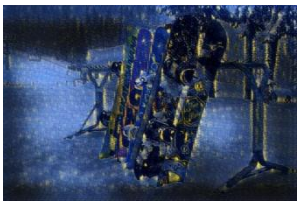
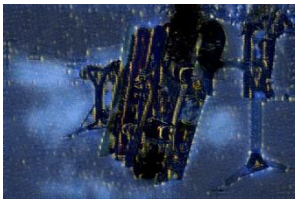
Content image		Stylized image	
Group 1			Preference
A			
B			
Group 2			Preference
A			
B			
Group 3			Preference
A			

B			
Group 4			Preference
A			
B			
Group 5			Preference
A			
B			
Group 6			Preference
A			

B			
Group 7			Preference
A			
B			
Group 8			Preference
A			

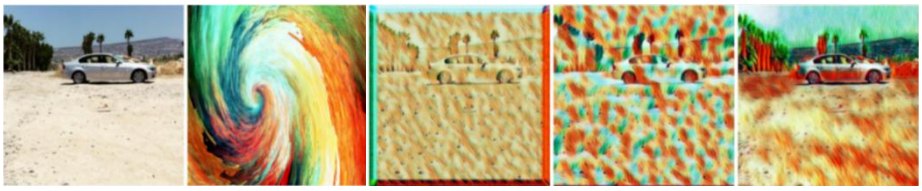
B			
Group 9			Preference
A			
B			
Group 10			Preference
A			

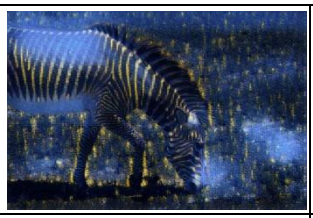
B			
Group 11			Preference
A			
B			
Group 12			Preference
A			

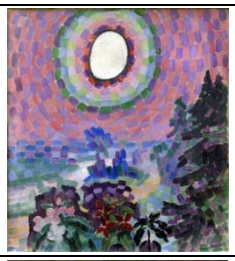
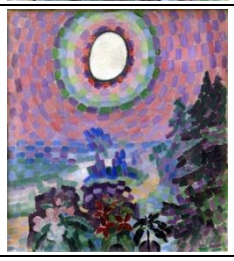
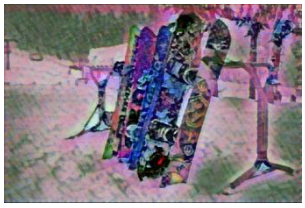
B			
Group 13			Preference
A			
B			
Group 14			Preference
A			
B			
Group 15			Preference
A			

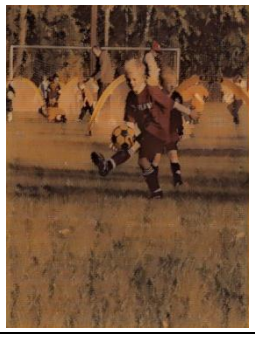
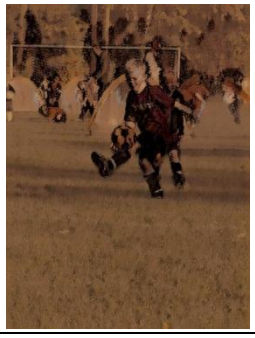
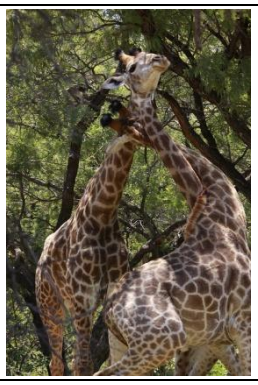
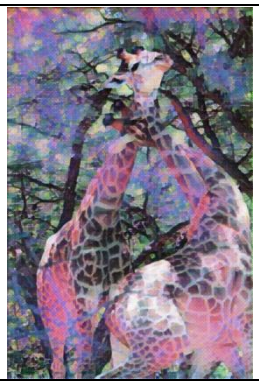
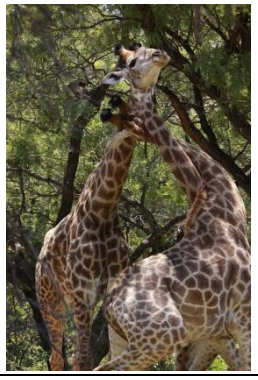
B			

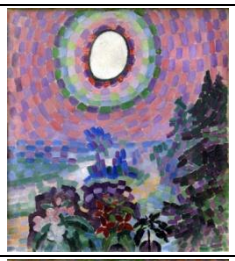
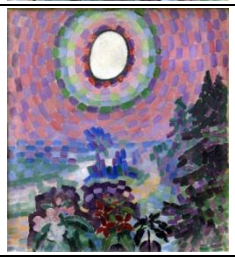
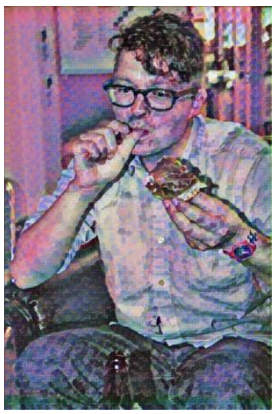
4. Style similarity

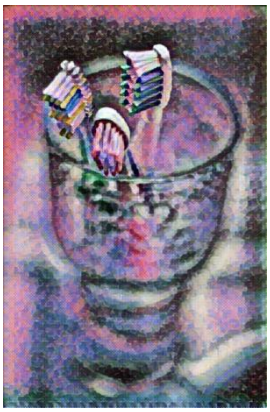
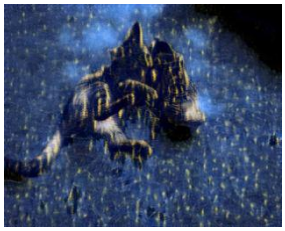
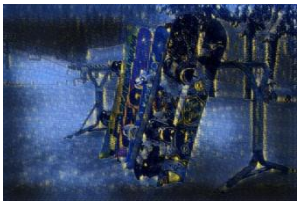
Compare item		Style similarity		
Compare standard	Please follow standards to compare the images in the table, the content images and the stylized images according to the style image, and select the result that you would prefer in each group. 1.The style images and stylized images have the same style. 2.The stylized images and the content images have the same content. 3. The stylized images are visually pleasing. We have provided corresponding examples for reference. Mark √ to the prefer ones in the last column			
examples	 Examples of relatively good and bad results are shown. The 1st image is the content image. The 2nd image is the style image. The 3rd, 4th, and 5th images show examples of poor and visually pleasing stylized results. The corresponding preference should increase in turn.			
Style image		Content image	Stylized image	
Group 1				Preference

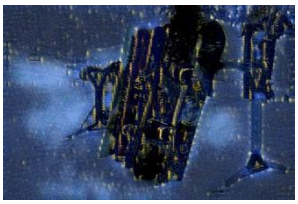
A				
B				
Group 2				Preference
A				
B				
Group 3				Preference
A				
B				
Group 4				Preference

A				
B				
Group 5				Preference
A				
B				
Group 6				Preference
A				
B				

Group 7				Preference
A				
B				
Group 8				Preference
A				
B				
Group 9				Preference

A				
B				
Group 10				Preference
A				
B				
Group 11				Preference

A				
B				
Group 12				Preference
A				
B				
Group 13				Preference
A				

B				
Group 14				Preference
A				
B				
Group 15				Preference
A				
B				

Results

Image quality	
Group 1	Number of Preference
Baseline(A)	5
our method(B)	35
Group 2	
Baseline(A)	10
our method(B)	30
Group 3	
Baseline(A)	17
our method(B)	23
Group 4	
Baseline(A)	16
our method(B)	24
Group 5	
Baseline(A)	11
our method(B)	29
Group 6	
Baseline(A)	9
our method(B)	31
Group 7	

Baseline(A)	21
our method(B)	19
Group 8	
Baseline(A)	9
our method(B)	31
Group 9	
Baseline(A)	20
our method(B)	20
Group 10	
Baseline(A)	23
our method(B)	17
Group 11	
Baseline(A)	17
our method(B)	23
Group 12	
Baseline(A)	25
our method(B)	15
Group 13	
Baseline(A)	20
our method(B)	20
Group 14	
Baseline(A)	21

our method(B)	19
Group 15	
Baseline(A)	9
our method(B)	31
Sum of groups	
Baseline	233
our method	367

Saliency order preservation	
Group 1	Number of Preference
Baseline(A)	13
our method(B)	27
Group 2	
Baseline(A)	14
our method(B)	26
Group 3	
Baseline(A)	10
our method(B)	30
Group 4	
Baseline(A)	25
our method(B)	15
Group 5	

Baseline(A)	10
our method(B)	30
Group 6	
Baseline(A)	9
our method(B)	31
Group 7	
Baseline(A)	19
our method(B)	21
Group 8	
Baseline(A)	10
our method(B)	30
Group 9	
Baseline(A)	27
our method(B)	13
Group 10	
Baseline(A)	20
our method(B)	20
Group 11	
Baseline(A)	16
our method(B)	24
Group 12	
Baseline(A)	17

our method(B)	23
Group 13	
Baseline(A)	29
our method(B)	11
Group 14	
Baseline(A)	24
our method(B)	16
Group 15	
Baseline(A)	23
our method(B)	17
Sum of groups	
Baseline	266
our method	334

Image content preservation	
Group 1	Number of Preference
Baseline(A)	11
our method(B)	29
Group 2	
Baseline(A)	21
our method(B)	19
Group 3	

Baseline(A)	16
our method(B)	24
Group 4	
Baseline(A)	21
our method(B)	19
Group 5	
Baseline(A)	12
our method(B)	28
Group 6	
Baseline(A)	16
our method(B)	24
Group 7	
Baseline(A)	24
our method(B)	16
Group 8	
Baseline(A)	11
our method(B)	29
Group 9	
Baseline(A)	15
our method(B)	25
Group 10	
Baseline(A)	18

our method(B)	22
Group 11	
Baseline(A)	16
our method(B)	24
Group 12	
Baseline(A)	16
our method(B)	24
Group 13	
Baseline(A)	23
our method(B)	17
Group 14	
Baseline(A)	15
our method(B)	25
Group 15	
Baseline(A)	17
our method(B)	23
Sum of groups	
Baseline	252
our method	348

Style similarity	
Group 1	Number of Preference

Baseline(A)	12
our method(B)	28
Group 2	
Baseline(A)	13
our method(B)	27
Group 3	
Baseline(A)	11
our method(B)	29
Group 4	
Baseline(A)	17
our method(B)	23
Group 5	
Baseline(A)	12
our method(B)	28
Group 6	
Baseline(A)	14
our method(B)	26
Group 7	
Baseline(A)	19
our method(B)	21
Group 8	
Baseline(A)	8

our method(B)	32
Group 9	
Baseline(A)	28
our method(B)	12
Group 10	
Baseline(A)	18
our method(B)	22
Group 11	
Baseline(A)	16
our method(B)	24
Group 12	
Baseline(A)	13
our method(B)	27
Group 13	
Baseline(A)	19
our method(B)	21
Group 14	
Baseline(A)	19
our method(B)	21
Group 15	
Baseline(A)	11
our method(B)	29

Sum of groups	
Baseline	230
our method	370

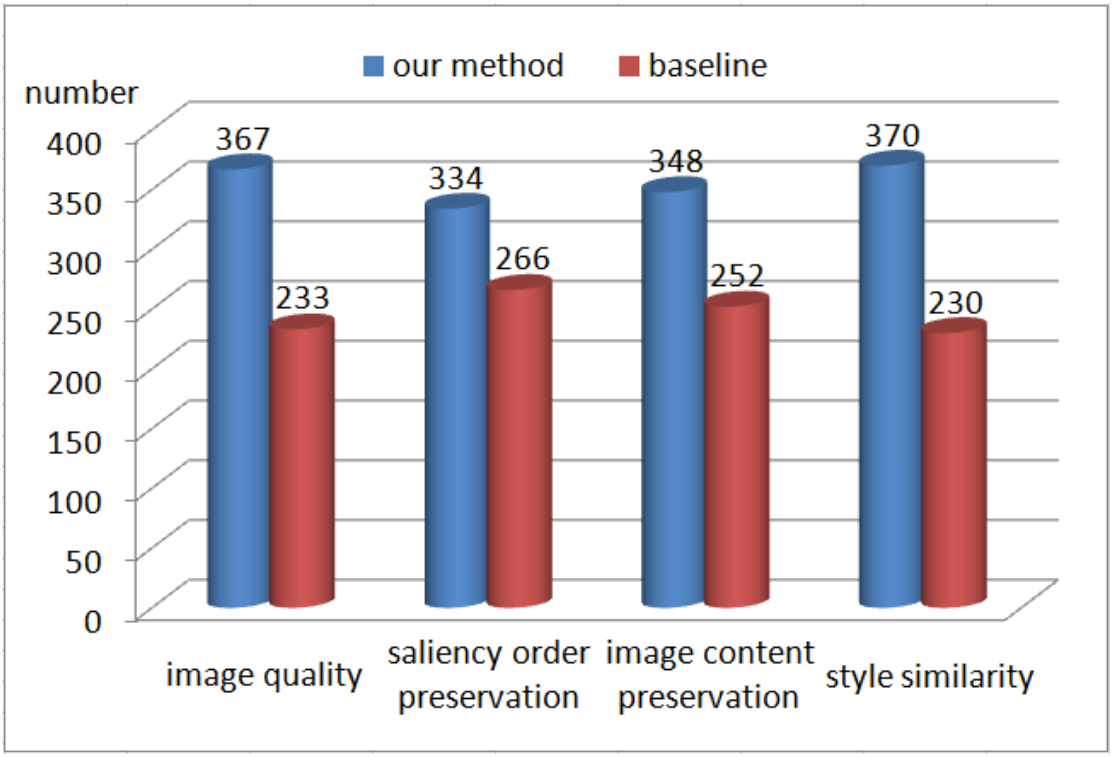


Figure 1

We design a user study to compare our results with baseline results in four standards: image quality, saliency order preservation, image content preservation and style similarity. Each comparison for one standard is composed of four groups of comparison. We use one of our results and one of the baseline results to form a group comparison. All the test images are generated from a sub set of COCOSTUFF dataset, which is not contained in the training set. We require users to decide which one in each group is better according to the corresponding standards. We take results from 40 users. The users are from different

professional fields: Economy(3), Law(3), literature(1), History(1), Medical Science(4), Management(1), Computer(22), Math(2), Astronomy(1), Electronic Information(1), Transportation(1).

We sum up all the groups of results to compare and the results are shown in Figure1. In the standard of image quality, 367(61.2%) votes prefer our results, rather than baseline results. This proves that our method can generate images that have better quality than baseline method. In the standard of saliency order preservation, 334(55.7%) votes prefer our results, rather than baseline results. This proves that our method generates results with stronger saliency preservation, which is the foreground in our method. In the standard of content preservation, 348(58.0%) votes prefer our results, rather than baseline results. This proves that our method did not undermine the content in images, even better than baseline method. In the standard of style similarity, 370(61.7%) votes prefer our results, rather than baseline results. This proves that our method can generate result that looks more like style image than baseline results thank to the attention mechanism. Quantitative Comparison shows that our model is better than baseline method in each standard.