Supplementary Materials

[Details of the cartoon BIQA Dataset]

The first dataset proposed in this paper consists of 200 HQ cartoon images with a resolution of 640×480. As shown in Fig. S-1, these cartoon images contain different kinds of contents, *e.g.*, cartoon characters, buildings, animals, simple patterns, design elements, sceneries. These HQ images are then degraded via 5 types of distortions with two levels for each type of distortion, *i.e.*, JPEG compression, Gaussian noise, salt and pepper noise, Gaussian blur, and AVS2 compression. Hence, there are total 2000 degraded images in this dataset. Examples of different types of distorted images are shown in Fig. S-2.

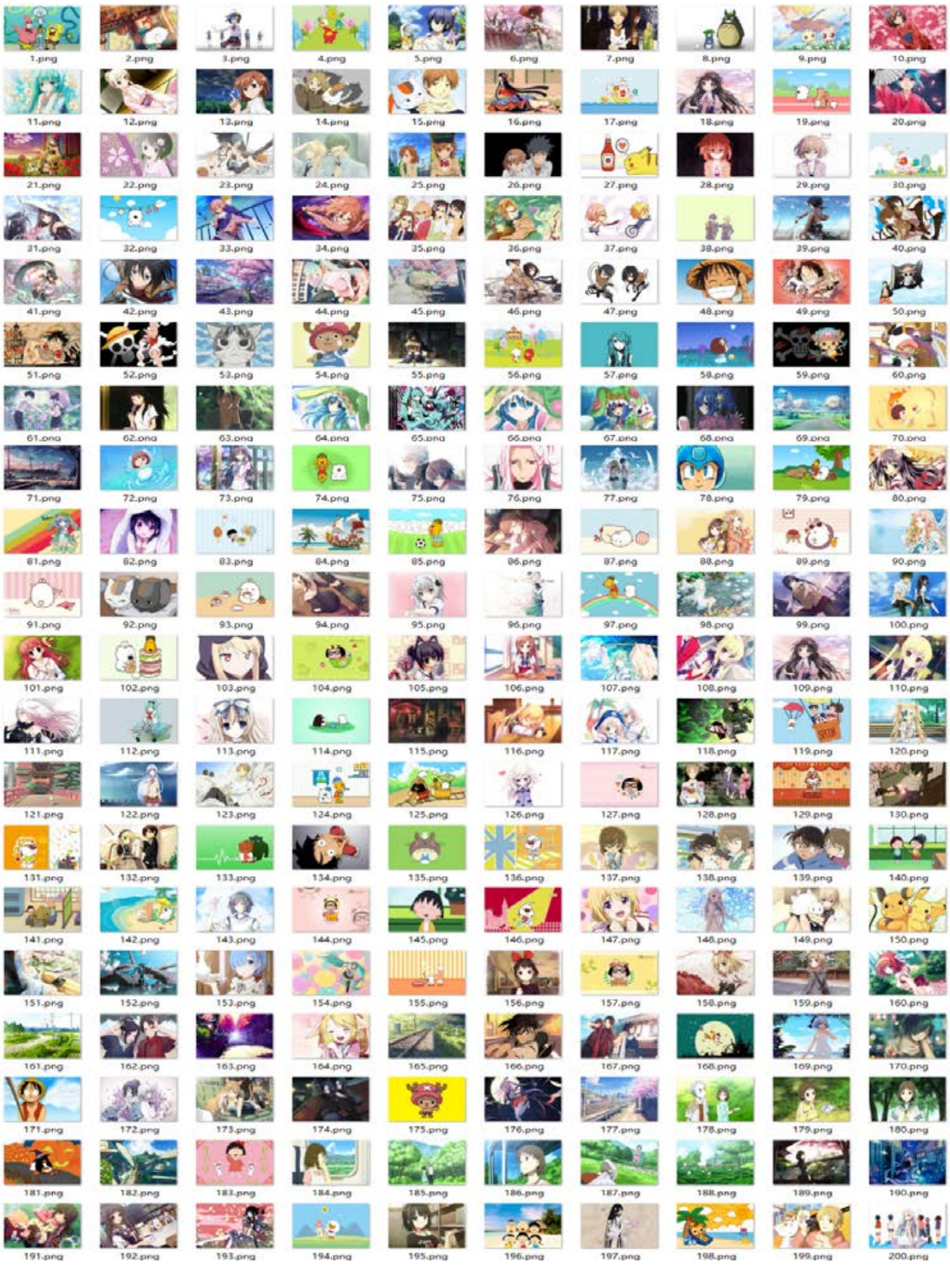


Fig. S-1. Cartoon images in the proposed cartoon BIQA dataset

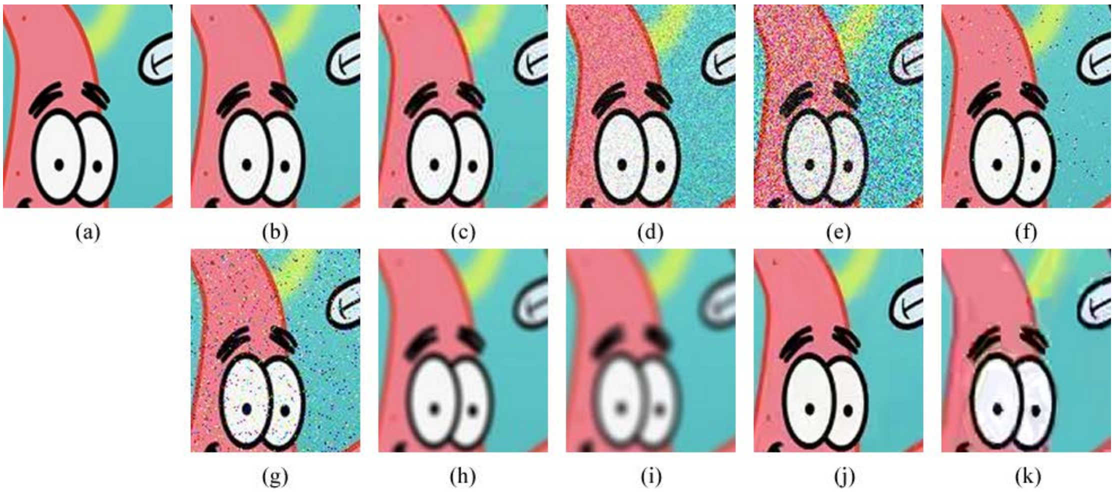


Fig. S-2. Examples of different distortions in the cartoon BIQA dataset, (a) original high quality, (b) JPEG compression, QF: 60, (c) JPEG compression, QF: 30, (d) Gaussian noise, mean 0 and variance 0.01, (e) Gaussian noise, mean 0 and variance 0.05, (f) Salt and pepper noise, density 0.01, (g) Salt and pepper noise, density 0.05, (h) Gaussian blur, standard deviation 3 and filter size 5, (i) Gaussian blur, standard deviation 3 and filter size 8, (j) AVS2 compression, QP:38, (k) AVS2 compression, QP:52.

To obtain the mean opinion score (MOS), 30 viewers were invited to participate in the subjective image quality assessment. Because each participant needs to score as many as 2000 images, a MATLAB interface was provided for participants to facilitate the quality assessment process. As shown in Fig. S-3, a degraded image and the corresponding reference HQ image are simultaneously displayed in the MATLAB interface. The participants subjectively assessed each image with a score, ranging from 0 to 100, by comparing the image with the reference one.

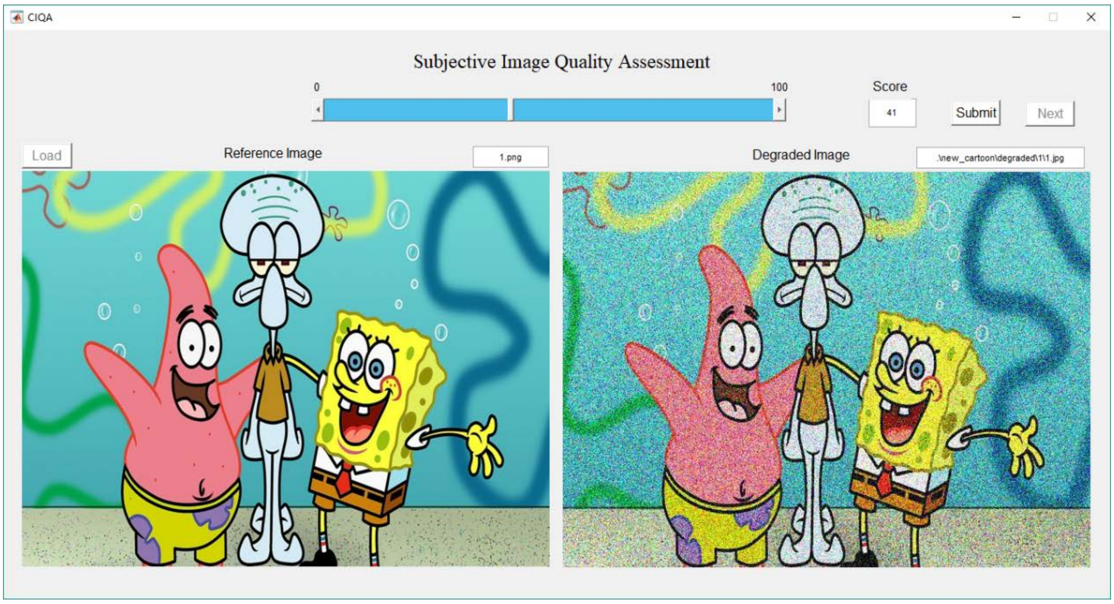


Fig. S-3. MATLAB interface used to subjective image quality assessment

After the subjective image quality assessment, the average score of each image is computed firstly,

$$\bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i$$

where s_i is the subjective score of the *i*-th participant, N (N=30) is the number of participants. To obtain more accurate scores, the standard deviation is computed:

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{(s_i - \bar{s})^2}{(N-1)}}$$

Based on the assumption that the subjective scores follow the normal distribution, a confidence interval is then computed, *i.e.*, $[\bar{s} - \delta, \bar{s} + \delta]$, where $\delta = Z_{\alpha/2}(N-1)\frac{\sigma}{\sqrt{N}}$. In this paper, the confidence level is set as 99.9%, thus $\alpha = 0.001$, $Z_{\alpha/2}(N-1)$ is 3.659. Finally, a score is removed if it is not in the confidence interval, MOS is computed by the average of the scores after removing the outliers.

[Details of the early cartoon testing dataset]

The second dataset contains a variety of real-world early cartoon images collected from the web, and these early cartoons contain real and complex artifacts rather than artificial distortions. In order to obtain corresponding HQ images, these early cartoons are reconstructed via two deep neural networks (DNN), *i.e.*, a deep convolutional neural network (CNN), and a deep residual network (ResNet) architecture[1]. Fig. S-4 shows the cartoon images in this dataset, and Fig. S-5 displays some close-ups to observe the visual quality. It is clear that the reconstructed images are much visually better than original LQ images.

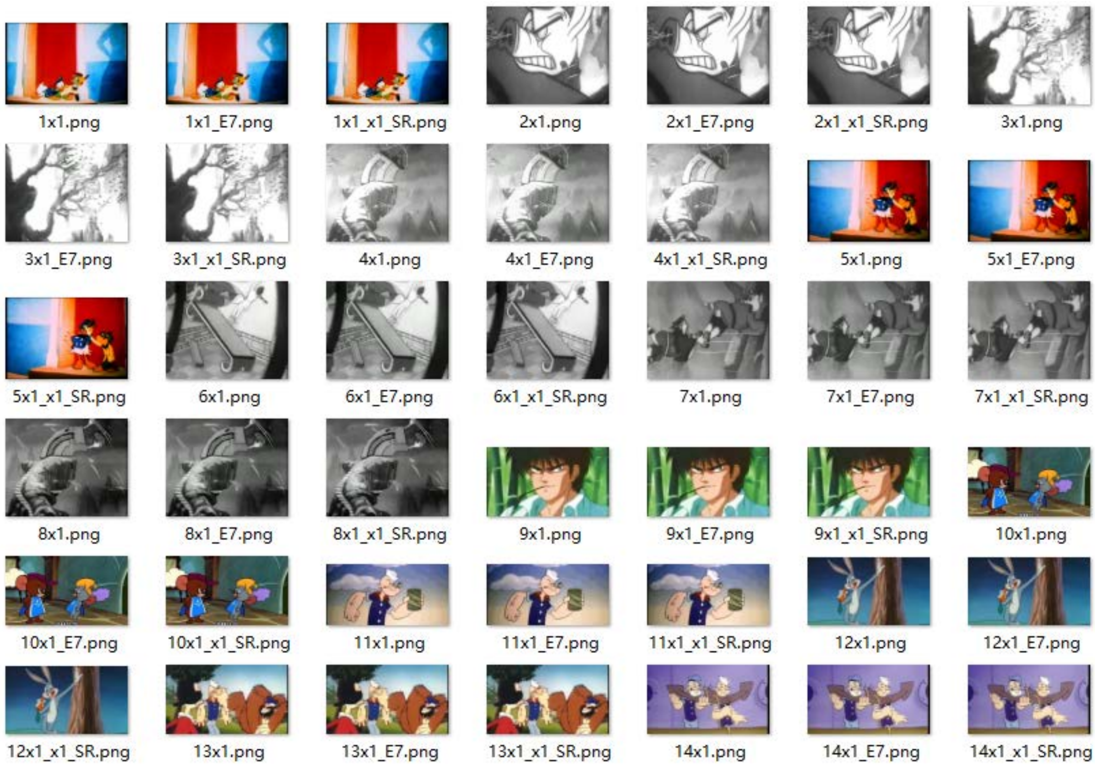


Fig. S-4. Cartoon images in the early cartoon testing dataset.

[1] This DNN-based cartoon image restoration method is another work of ourselves. Here we do not introduce the details of these two networks. By the way, we found that current natural image BIQA methods often fail on cartoon scenarios in our cartoon restoration experiments. This phenomenon forces us to carefully analyze the characteristics of cartoon images and specially design an effective cartoon BIQA method.

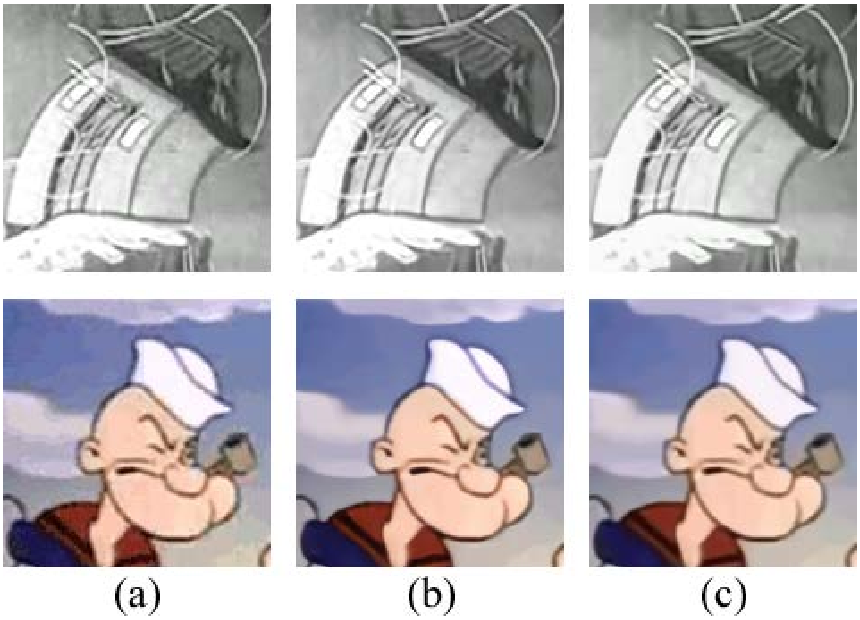
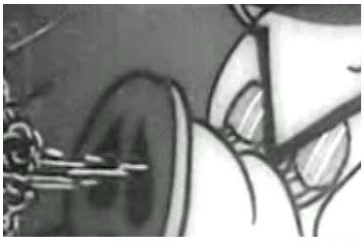


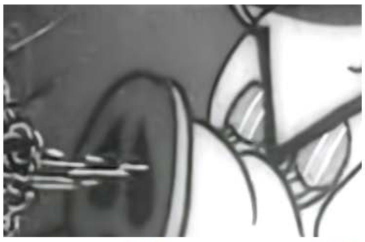
Fig. S-5. Some close-ups of images in the early cartoon testing dataset, (a) denotes the LQ early cartoons; (b) and (c) denote the reconstructed HQ images via different DNN methods.

[More results on early cartoon testing dataset]

More BIQA results of early cartoons are shown in Fig. S-6. Note that the BIQI, ILNIQE, NFERM, BRISQUE and BMPRI have been retrained with the cartoon image dataset. The left column is LQ images, the middle and right columns denote the reconstructed HQ images via two different DNN methods. The HQ reconstructed images obviously have better visual quality. However, these natural image BIQA methods all produce unreasonable scores (which are marked in red), because they often tend to generate higher scores to significant LQ images. In contrast, the proposed method always produces reasonable scores which are highly consistent with human perception. These results demonstrate that the proposed method is more suitable and robust for cartoon scenarios since it is specially designed based on the characteristics of cartoons.



BIQI[21]	0.7393	TCLT[22]	0.5816
ILNIQE[41]	0.7754	NFERM[24]	0.9493
DIIVINE[39]	0.5791	SISBLIM[42]	0.1705
BRISQUE[40]	0.7769	BPRI[27]	0.9854
BMPRI[28]	0.8383	CBIQA(ours)	0.7684



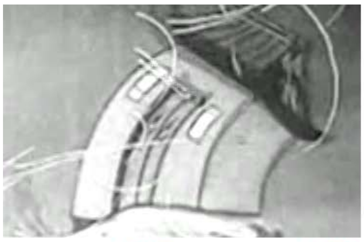
BIQI[21]	0.6167	TCLT[22]	0.4470
ILNIQE[41]	0.7587	NFERM[24]	0.6841
DIIVINE[39]	0.5350	SISBLIM[42]	0.1545
BRISQUE[40]	0.7974	BPRI[27]	0.8619
BMPRI[28]	0.6850	CBIQA(ours)	0.8323

BIQI[21]	0.5855	TCLT[22]	0.3973
ILNIQE[41]	0.7612	NFERM[24]	0.7145
DIIVINE[39]	0.5148	SISBLIM[42]	0.1651
BRISQUE[40]	0.7896	BPRI[27]	0.8860
BMPRI[28]	0.6930	CBIQA(ours)	0.8201

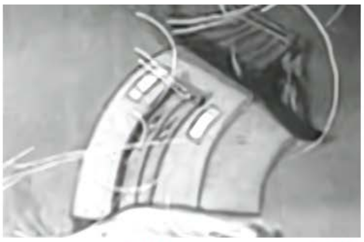
BIQI[21]	0.7463	TCLT[22]	0.5525
ILNIQE[41]	0.7233	NFERM[24]	0.9228
DIIVINE[39]	0.3061	SISBLIM[42]	0.1664
BRISQUE[40]	0.7358	BPRI[27]	0.9859
BMPRI[28]	0.8738	CBIQA(ours)	0.6888

BIQI[21]	0.6510	TCLT[22]	0.5436
ILNIQE[41]	0.7420	NFERM[24]	0.6969
DIIVINE[39]	0.2672	SISBLIM[42]	0.1493
BRISQUE[40]	0.8091	BPRI[27]	0.8505
BMPRI[28]	0.6957	CBIQA(ours)	0.7740

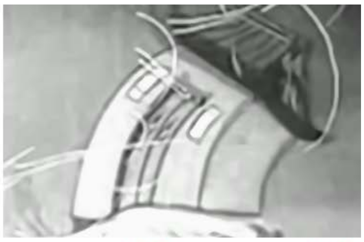
BIQI[21]	0.6183	TCLT[22]	0.4460
ILNIQE[41]	0.7452	NFERM[24]	0.7072
DIIVINE[39]	0.2565	SISBLIM[42]	0.1556
BRISQUE[40]	0.8120	BPRI[27]	0.8538
BMPRI[28]	0.6924	CBIQA(ours)	0.8306



BIQI[21]	0.7657	TCLT[22]	0.6532
ILNIQE[41]	0.7714	NFERM[24]	1.0096
DIIVINE[39]	0.6234	SISBLIM[42]	0.1817
BRISQUE[40]	0.8389	BPRI[27]	0.9888
BMPRI[28]	0.8649	CBIQA(ours)	0.4606



BIQI[21]	0.6604	TCLT[22]	0.5143
ILNIQE[41]	0.7730	NFERM[24]	0.7726
DIIVINE[39]	0.6053	SISBLIM[42]	0.1760
BRISQUE[40]	0.9019	BPRI[27]	0.8791
BMPRI[28]	0.6929	CBIQA(ours)	0.6235



BIQI[21]	0.6787	TCLT[22]	0.4492
ILNIQE[41]	0.7515	NFERM[24]	0.7930
DIIVINE[39]	0.5370	SISBLIM[42]	0.1751
BRISQUE[40]	0.9284	BPRI[27]	0.8944
BMPRI[28]	0.6964	CBIQA(ours)	0.6231

BIQI[21]	0.7536	TCLT[22]	0.8276
ILNIQE[41]	0.6919	NFERM[24]	0.9978
DIIVINE[39]	0.5799	SISBLIM[42]	0.1779
BRISQUE[40]	0.9321	BPRI[27]	0.9946
BMPRI[28]	0.8641	CBIQA(ours)	0.6258

BIQI[21]	0.5544	TCLT[22]	0.5901
ILNIQE[41]	0.6543	NFERM[24]	0.8142
DIIVINE[39]	0.4007	SISBLIM[42]	0.1710
BRISQUE[40]	0.9010	BPRI[27]	0.9245
BMPRI[28]	0.7111	CBIQA(ours)	0.7781

BIQI[21]	0.6806	TCLT[22]	0.5606
ILNIQE[41]	0.5591	NFERM[24]	0.8336
DIIVINE[39]	0.2678	SISBLIM[42]	0.1603
BRISQUE[40]	0.9349	BPRI[27]	0.9308
BMPRI[28]	0.7110	CBIQA(ours)	0.7396

BIQI[21]	0.7447	TCLT[22]	0.4935
ILNIQE[41]	0.7647	NFERM[24]	0.9556
DIIVINE[39]	0.5365	SISBLIM[42]	0.1575
BRISQUE[40]	0.7576	BPRI[27]	0.9816
BMPRI[28]	0.8323	CBIQA(ours)	0.5568

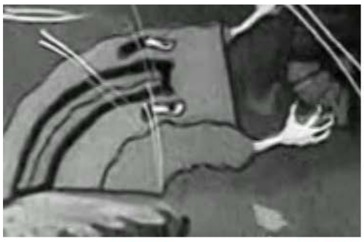
BIQI[21]	0.8530	TCLT[22]	0.4663
ILNIQE[41]	0.7606	NFERM[24]	0.7335
DIIVINE[39]	0.5093	SISBLIM[42]	0.1402
BRISQUE[40]	0.8090	BPRI[27]	0.8512
BMPRI[28]	0.6876	CBIQA(ours)	0.6750

BIQI[21]	0.8287	TCLT[22]	0.4264
ILNIQE[41]	0.7637	NFERM[24]	0.7246
DIIVINE[39]	0.3927	SISBLIM[42]	0.1385
BRISQUE[40]	0.7767	BPRI[27]	0.8621
BMPRI[28]	0.6869	CBIQA(ours)	0.5989

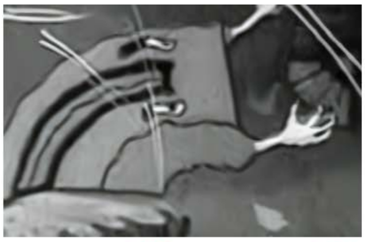
BIQI[21]	0.7184	TCLT[22]	0.6373
ILNIQE[41]	0.7286	NFERM[24]	0.9593
DIIVINE[39]	0.5226	SISBLIM[42]	0.1557
BRISQUE[40]	0.8740	BPRI[27]	0.9843
BMPRI[28]	0.8564	CBIQA(ours)	0.2454

BIQI[21]	0.8809	TCLT[22]	0.4742
ILNIQE[41]	0.7348	NFERM[24]	0.7786
DIIVINE[39]	0.3817	SISBLIM[42]	0.1292
BRISQUE[40]	0.8979	BPRI[27]	0.8890
BMPRI[28]	0.6921	CBIQA(ours)	0.3269

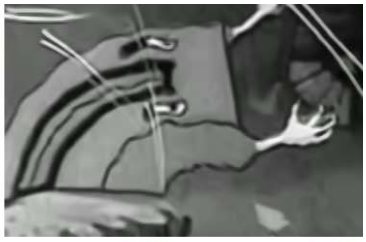
BIQI[21]	0.8726	TCLT[22]	0.4174
ILNIQE[41]	0.7440	NFERM[24]	0.7878
DIIVINE[39]	0.2087	SISBLIM[42]	0.1265
BRISQUE[40]	0.8795	BPRI[27]	0.9038
BMPRI[28]	0.6948	CBIQA(ours)	0.3023



BIQI[21]	0.8017	TCLT[22]	0.6415
ILNIQE[41]	0.7818	NFERM[24]	1.0204
DIIVINE[39]	0.6445	SISBLIM[42]	0.1942
BRISQUE[40]	0.8348	BPRI[27]	0.9886
BMPRI[28]	0.8604	CBIQA(ours)	0.5693



BIQI[21]	0.6257	TCLT[22]	0.5141
ILNIQE[41]	0.7457	NFERM[24]	0.7791
DIIVINE[39]	0.6252	SISBLIM[42]	0.2029
BRISQUE[40]	0.8648	BPRI[27]	0.8765
BMPRI[28]	0.6923	CBIQA(ours)	0.6546



BIQI[21]	0.5979	TCLT[22]	0.4478
ILNIQE[41]	0.7278	NFERM[24]	0.8010
DIIVINE[39]	0.5928	SISBLIM[42]	0.1951
BRISQUE[40]	0.8830	BPRI[27]	0.8922
BMPRI[28]	0.6925	CBIQA(ours)	0.7534

BIQI[21]	0.7430	TCLT[22]	0.7109
ILNIQE[41]	0.7493	NFERM[24]	0.7256
DIIVINE[39]	0.6324	SISBLIM[42]	0.1914
BRISQUE[40]	0.8159	BPRI[27]	0.9673
BMPRI[28]	0.8111	CBIQA(ours)	0.7580

BIQI[21]	0.5511	TCLT[22]	0.6395
ILNIQE[41]	0.7310	NFERM[24]	0.7428
DIIVINE[39]	0.6062	SISBLIM[42]	0.1777
BRISQUE[40]	0.9245	BPRI[27]	0.9025
BMPRI[28]	0.7094	CBIQA(ours)	0.9034

BIQI[21]	0.5981	TCLT[22]	0.5470
ILNIQE[41]	0.7066	NFERM[24]	0.7585
DIIVINE[39]	0.5532	SISBLIM[42]	0.1684
BRISQUE[40]	0.9313	BPRI[27]	0.9229
BMPRI[28]	0.7205	CBIQA(ours)	0.8722

BIQI[21]	0.6751	TCLT[22]	0.7395
ILNIQE[41]	0.7947	NFERM[24]	0.7799
DIIVINE[39]	0.6517	SISBLIM[42]	0.2635
BRISQUE[40]	0.9918	BPRI[27]	0.9665
BMPRI[28]	0.8279	CBIQA(ours)	0.8643

BIQI[21]	0.6030	TCLT[22]	0.7251
ILNIQE[41]	0.7472	NFERM[24]	0.6302
DIIVINE[39]	0.4391	SISBLIM[42]	0.2616
BRISQUE[40]	0.9799	BPRI[27]	0.9501
BMPRI[28]	0.7693	CBIQA(ours)	0.9396

BIQI[21]	0.5927	TCLT[22]	0.7030
ILNIQE[41]	0.7374	NFERM[24]	0.7125
DIIVINE[39]	0.5732	SISBLIM[42]	0.2475
BRISQUE[40]	1.0277	BPRI[27]	0.9510
BMPRI[28]	0.7647	CBIQA(ours)	0.9336

BIQI[21]	0.8149	TCLT[22]	0.7127
ILNIQE[41]	0.7823	NFERM[24]	0.8700
DIIVINE[39]	0.6678	SISBLIM[42]	0.1922
BRISQUE[40]	1.0348	BPRI[27]	0.9722
BMPRI[28]	0.8123	CBIQA(ours)	0.7668

BIQI[21]	0.6634	TCLT[22]	0.6838
ILNIQE[41]	0.7797	NFERM[24]	0.7861
DIIVINE[39]	0.6898	SISBLIM[42]	0.1984
BRISQUE[40]	1.0367	BPRI[27]	0.9400
BMPRI[28]	0.7292	CBIQA(ours)	0.9206

BIQI[21]	0.5984	TCLT[22]	0.5957
ILNIQE[41]	0.7606	NFERM[24]	0.7779
DIIVINE[39]	0.5266	SISBLIM[42]	0.1917
BRISQUE[40]	1.0575	BPRI[27]	0.9453
BMPRI[28]	0.7321	CBIQA(ours)	0.9060

BIQI[21]	0.7482	TCLT[22]	0.7151	BIQI[21]	0.7116	TCLT[22]	0.7107	BIQI[21]	0.6200	TCLT[22]	0.5830
ILNIQE[41]	0.7661	NFERM[24]	0.8222	ILNIQE[41]	0.7665	NFERM[24]	0.7631	ILNIQE[41]	0.7577	NFERM[24]	0.7956
DIIVINE[39]	0.7270	SISBLIM[42]	0.1982	DIIVINE[39]	0.6899	SISBLIM[42]	0.2019	DIIVINE[39]	0.6660	SISBLIM[42]	0.1914
BRISQUE[40]	0.9091	BPRI[27]	0.9834	BRISQUE[40]	0.8441	BPRI[27]	0.9436	BRISQUE[40]	0.8723	BPRI[27]	0.9547
BMPRI[28]	0.7993	CBIQA(ours)	0.4889	BMPRI[28]	0.7385	CBIQA(ours)	0.8270	BMPRI[28]	0.7415	CBIQA(ours)	0.8217

Fig. S-6. BIQA results of early cartoons. The left column denotes the LQ early cartoons; the middle and right columns denote the reconstructed HQ images via different DNN methods (score that is inconsistent with human perception is marked in red.).

[Other implementation details]

In this paper, Chi-square distances between the histograms of a cartoon image and prior models are computed to measure the image quality. In order to normalize the distance to a score between 0 and 1, we experimentally select a modified sigmoid function. The values of original sigmoid function are between 0.5 and 1 when the independent variable is greater than 0, and it is a monotonous increasing function. Hence, we transform sigmoid function to a monotonous decreasing function with the values from 1 to 0, as follows,

$$f(x) = \frac{2e^{-x}}{1 + e^{-x}}$$

The function curve is shown in Fig. S-7. It is obvious that this curvilinear function can distinguish image quality better than linear function, especially when the distances are relatively small. That means this modified sigmoid curve is helpful to assess image quality in some difficult situations.

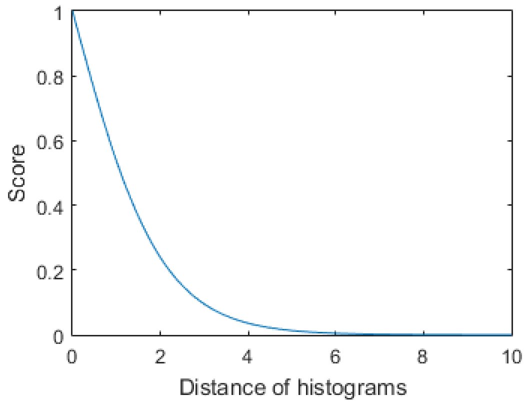


Fig. S-7. Function curve used in our score normalization.