A Reduced-Reference Quality Metric for Screen Content Image

Zhaohui. Che[1], Guangtao. Zhai[1], Ke. Gu[2], Patrick Le. Callet[3]

[1]Shanghai Jiao Tong University, China

[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore

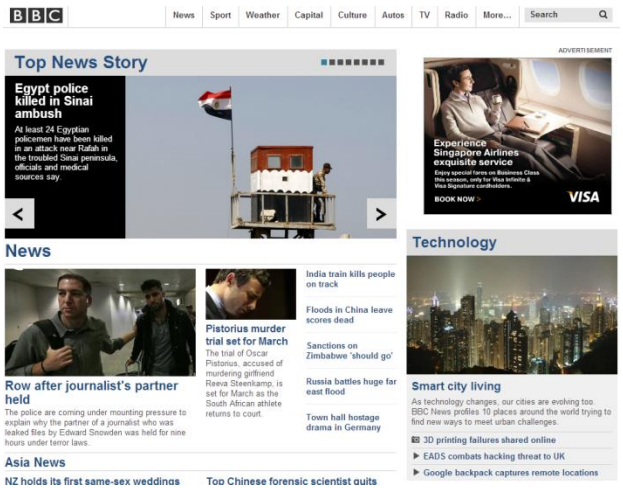
[3]Luman Universite, Universit´ e de Nantes, IRCCyN UMR CNRS 6597, Polytech Nantes, France

September, 2017

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Catalogue

Motivation: Screen Content Image



Pictorial Regions



.....

- Abundant Color Information
- Low-frequency Components
- Texture Details
- Regular Aspect Ratio

Text Regions

Top News Story

Egypt police killed in Sinai ambush

At least 24 Egyptian policemen have been killed in an attack near Rafah in the troubled Sinai peninsula, officials and medical sources say.

Pistorius murder trial set for March

The trial of Oscar Pistorius, accused of murdering girlfriend Reeva Steenkamp, is set for March as the South African athlete returns to court.

.....

- Sharp Edges, Strong Boundaries
- Pixels with High Gradient Values
- A Few Major Colors
- Anomalous Aspect Ratio

Motivation: Sensitivity of Distortion Type

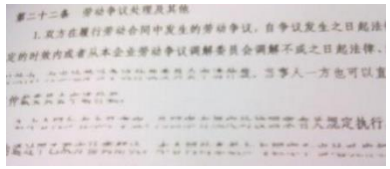
Pictorial Regions

- Noise
- Contrast Degradation
- Compression
- Blockiness
- Blur

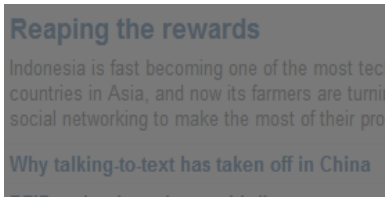
Text Regions

- Blur

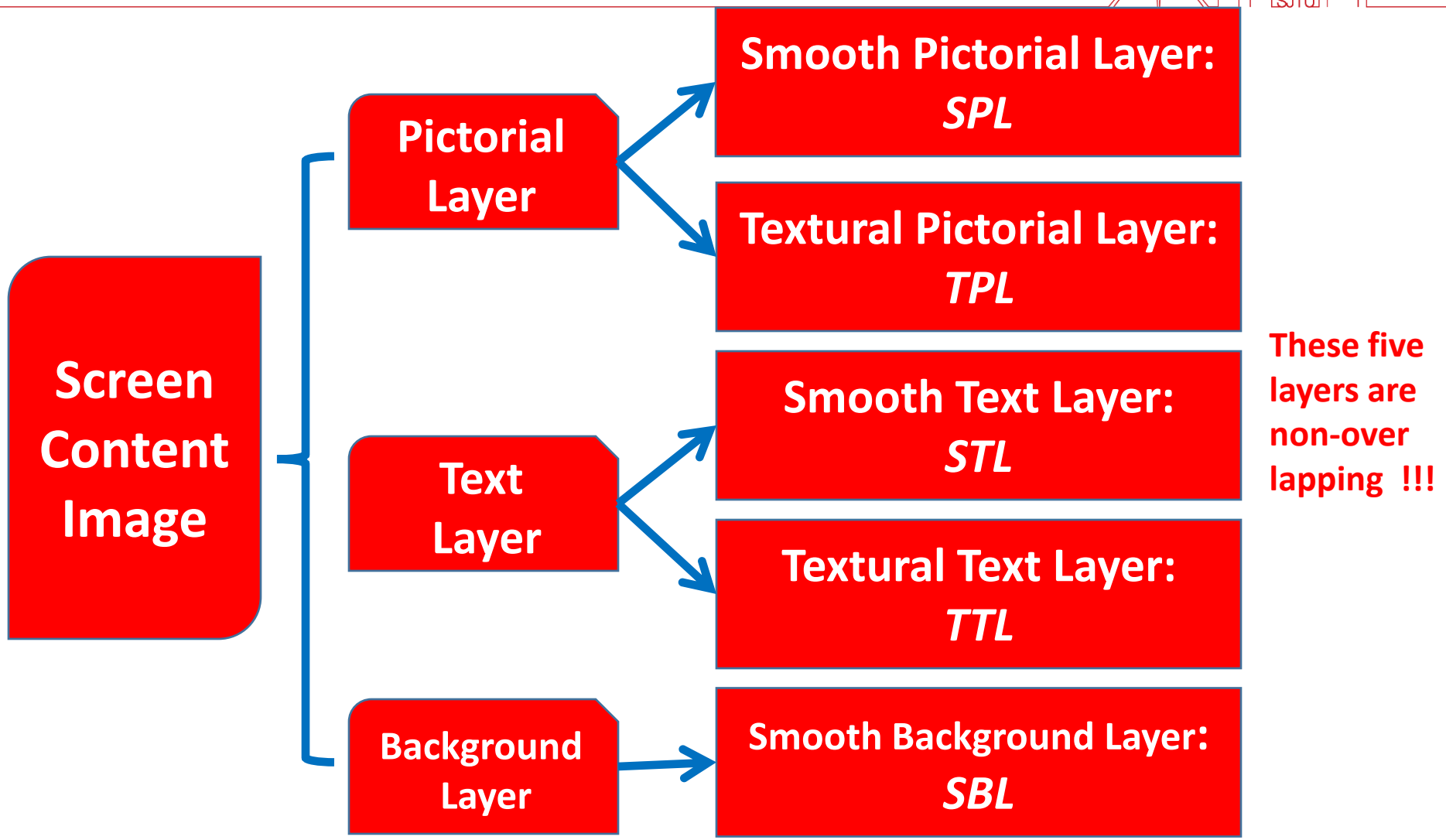
- Integrity Destruction



- Extremely Low Contrast



Layer-based Segmentation Method



Layer-based Segmentation Method

Autoregressive Model (AR)

The AR model specifies that the output depends linearly on its own previous variable value and on a stochastic term.

$$y_i = \boldsymbol{\alpha} \times \boldsymbol{\gamma}^k(y_i) + \varepsilon_i$$

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} ||\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{\alpha}||_2$$

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^Y\boldsymbol{y}$$

The AR model can protect pictorial details well, but it performs poorly on steep edges of text regions.

Guided Image Filter (GIF)

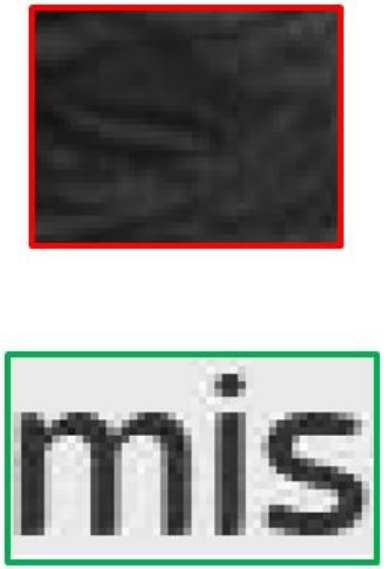
The GIF can generate output according to the guide image. And GIF behaves as an efficient edge-preserving smoothing operator when the guide image is identical to the original input image.

GIF can protect the edges and gradient information of text regions, while the texture information of pictorial regions will be destroyed.

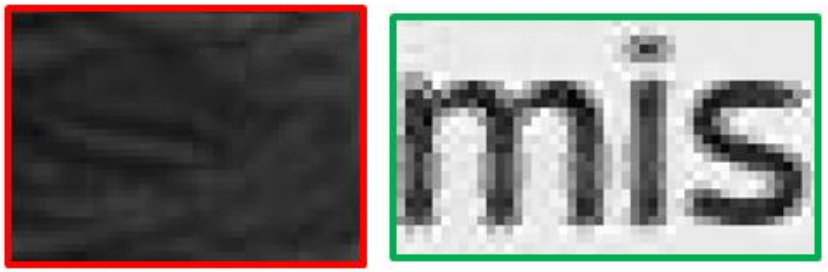
Layer-based Segmentation Method

Original Image

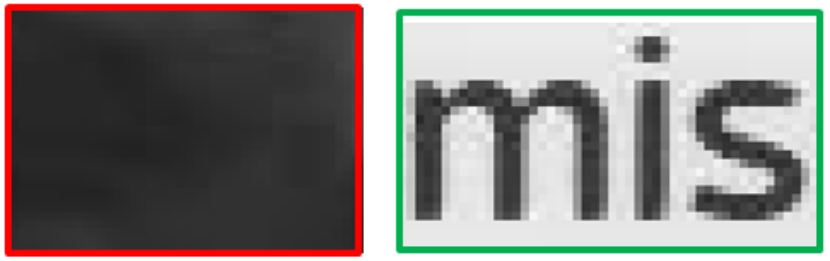
Original Patches



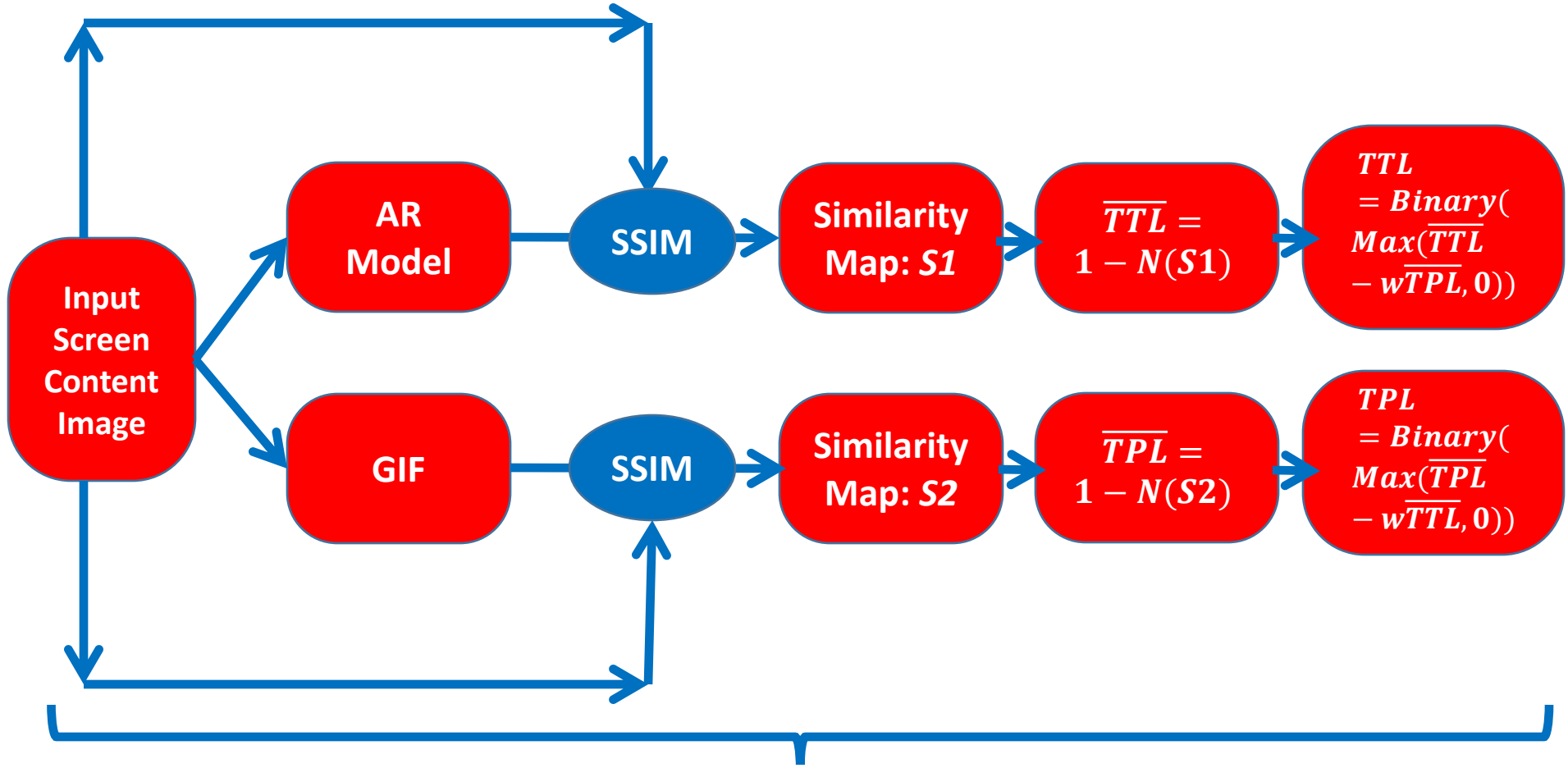
Patches filtered by AR model



Patches filtered by GIF



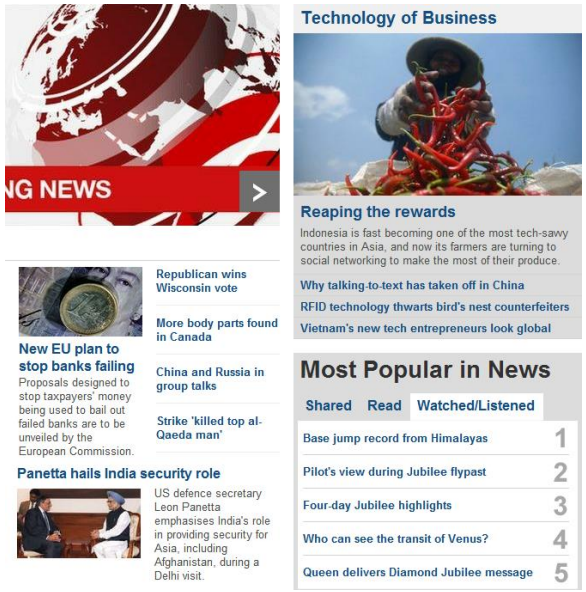
Layer-based Segmentation Method



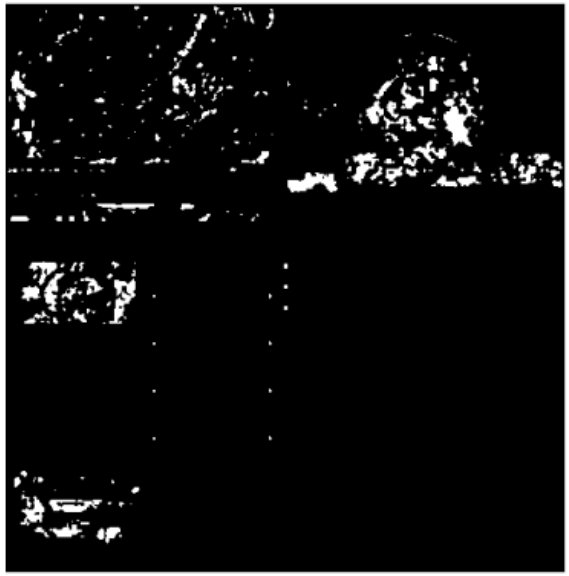
Coarse Segmentation Step

Layer-based Segmentation Method

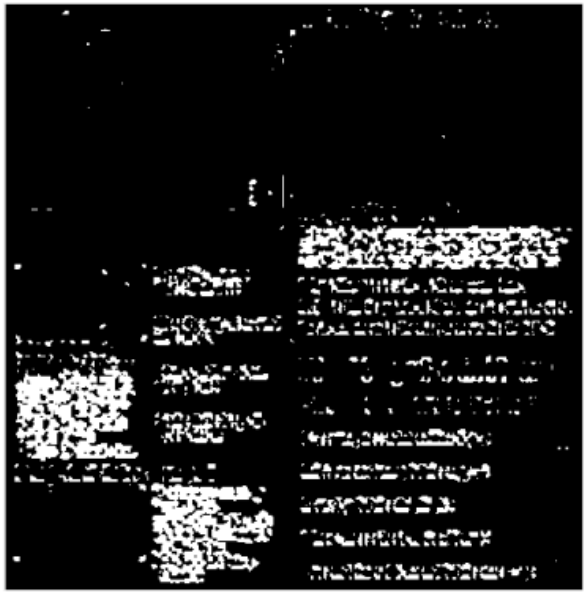
Original Image



TPL



TTL



How about the rest parts?

Such as *SPL, STL, and SBL* ?

- ***SBL:*** Most screen content images have smooth backgrounds in a few base colors. Therefore, we found out the most frequent base colors accounting for at least 20% of all pixels, so that we could extract *SBL* in base colors.
- ***SPL & STL :*** Remaining *SP L* and *ST L* are difficult to differentiate since they have similar small variances.

$$(SPL \cup TPL) \cup STL = Pictorial\ Layer \cup STL = 1 - SBL - TTL$$

Matlab Function : bwareaopen

Eliminate the tiny *STL* using connected region open operator

Quality Metric for Pictorial Layer

Free Energy Principle

Suppose that the internal generative model *g* of human brain is parametric for visual perception, and the perceived scene can be explained by adjusting the parameter vector φ. Given the input visual signal *s*, its surprise (measured by entropy) can be attained by integrating the joint distribution $p(s;\varphi|g)$ over the space of model parameter φ.

$$f(\phi) = -\int q(\phi|s)\log\frac{p(s,\phi)}{q(\phi|s)}\mathrm{d}\phi$$

Free Energy guided Degradation Metric (*FEDM*)

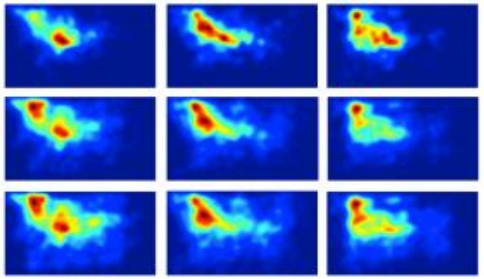
Considering the computational and operational aspects of free energy, we adopted AR model to simulate human brain generative model *g*, so that the quantitative measurement of *FEDM* is defined as entropy of error map $I_{i;\Delta}$ between input image I_i and its AR model filtering result $I_{i;ar}$ ($I_{i;\Delta} = I_i - I_{i;ar}$).

FEDM of the reference image: $$FE_{i,r} = -\sum_k p_k(I_{i,\Delta})\log p_k(I_{i,\Delta})$$

Quality Metric for Pictorial Layer

Human Visual Fixation of Screen Content Image

Shen chenyao and Zhao qi [*Webpage Saliency, ECCV2014*] pointed that human visual fixations usually fall in the top-left region when viewing the screen content images. We proposed the top-left bias pooling strategy to emphasis impact of pictures' locations on ultimate quality score.



Quality Score of Pictorial Layer : $score_p$

$$\begin{cases} score_p = \sum_{i=1}^{N_r} \mu_i |FE_{i,r} - FE_{i,d}| \\ \mu_i = \dfrac{D([X_{i,c}, Y_{i,c}],[1,1])^{-1}}{\sum_{j=1}^{N_r} D([X_{j,c}, Y_{j,c}],[1,1])^{-1}} \end{cases}$$

Where $FE_{i,d}$ is the free energy quality index of the i_{th} picture $I_{i;d}$ located at the distorted screen content image. Specifically, the physical meaning of pooling coefficient μ_i is the Euclidean distance (represented by *D*) between centroid point [$X_{i;c}$, $Y_{i;c}$] of picture $I_{i;d}$ and top-left corner [1, 1] of the distorted image.

Quality Metric for Text Layer

For text region, human visual system is very sensitive to *Blur* and *Extremely Low Contrast*.

(*STL* ∪ *TTL*) = *Text Layer* = 1 – *Pictorial Layer* - *SBL*

Contrast Feature for Text Layer

$$f_1 = \frac{1}{255}\frac{|B_r - B_d|}{|T_d - B_d| + C1}$$

Where B_r, B_d and T_d represent the gray values of reference SCI's background, distorted SCI's background and distorted SCI's text separately. The weighting coefficient 1/255 guarantees that the $f1$ is from 0 to 1. Obviously, higher $|B_r - B_d|$ means severe contrast change distortion, while lower $|T_d - B_d|$ means that the text and background of distorted SCI is in low contrast. Higher $f1$ means that it's difficult for human eyes to distinguish between text and background, i.e. lower quality score.

Quality Metric for Text Layer

Blurness Index for Text Layer

$$f_2 = \frac{|N_{t,r} - N_{t,d}|}{N_{t,r}}$$

We firstly adopted Matlab function *bwareaopen.m* to eliminate the tiny connected regions (noise) from *Text Layer*, then we utilized *bwconncomp.m* to find out the number $N_{t,d}$ of remaining connected regions of *Text Layer*. And the $N_{t,d}$ is the number of connected regions of the reference SCI's text layer.

Final Quality Score

$$\begin{cases} score_p = \sum_{i=1}^{N_r} \mu_i |FE_{i,r} - FE_{i,d}| \\ \mu_i = \frac{D([X_{i,c}, Y_{i,c}], [1,1])^{-1}}{\sum_{j=1}^{N_r} D([X_{j,c}, Y_{j,c}], [1,1])^{-1}} \end{cases}$$

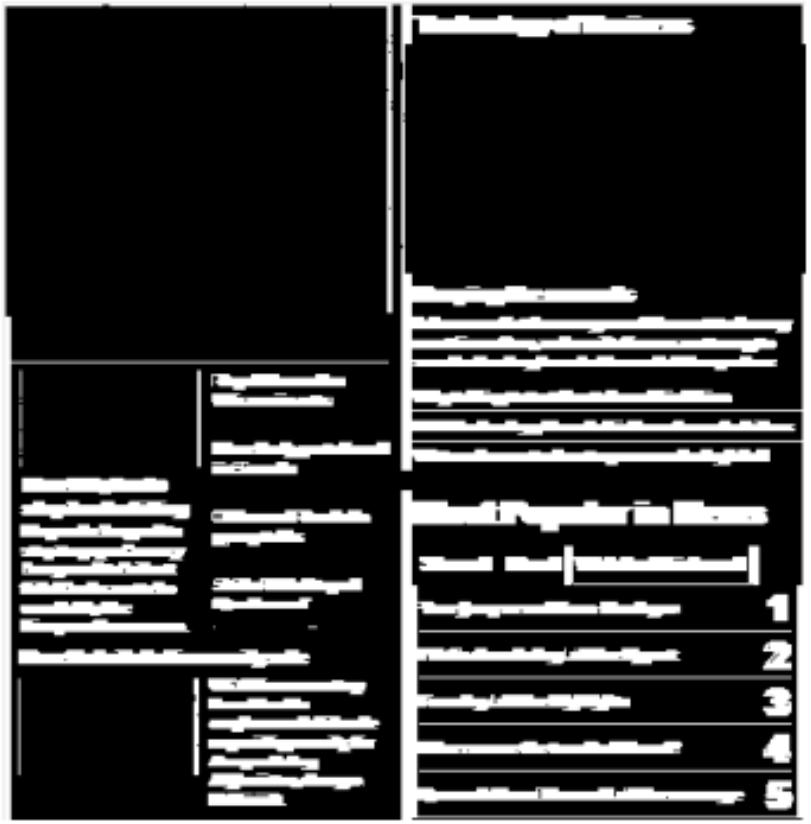
$$score_t = \frac{1}{2} f_1 + \frac{1}{2} f_2$$

$$score = \theta score_p + (1 - \theta) score_t$$

where the weighting coefficient ϑ is the area ration between pictorial layer and the whole screen content image.

Quality Metric for Text Layer

The Text Layer of the reference image, and the $N_{t,r}$ = 211



The Text Layer of the distorted image with motion blur, and the $N_{t,d}$ = 34

Experimental Results

Table 1. Performance over all distortion types

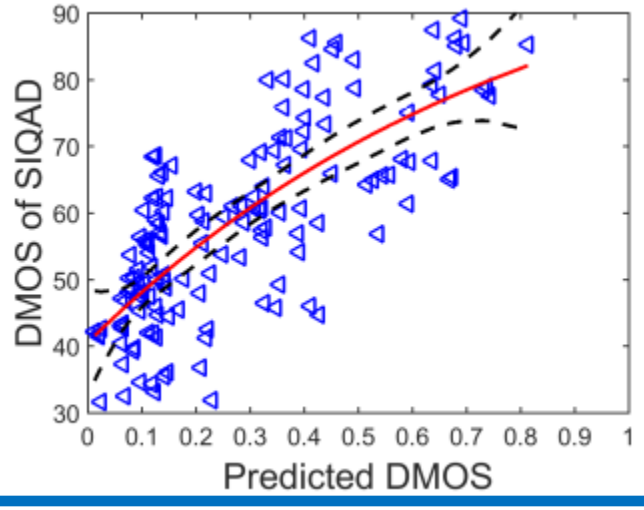
IQA Metrics	PLCC	SROCC	RMSE
SSIM [4]	0.7445	0.7433	9.4713
PSNR	0.5788	0.5539	11.5691
VIF [15]	0.8026	0.7857	8.4642
VSI [16]	0.5403	0.5199	11.9384
FSIM [17]	0.5741	0.5647	11.6164
Q_s [5]	0.8573	0.8456	7.3030
SIQM [6]	0.8518	0.8452	7.4219
SPQA [3]	**0.8631**	**0.8579**	**7.2297**
Proposed	**0.8126**	**0.7962**	**8.2633**

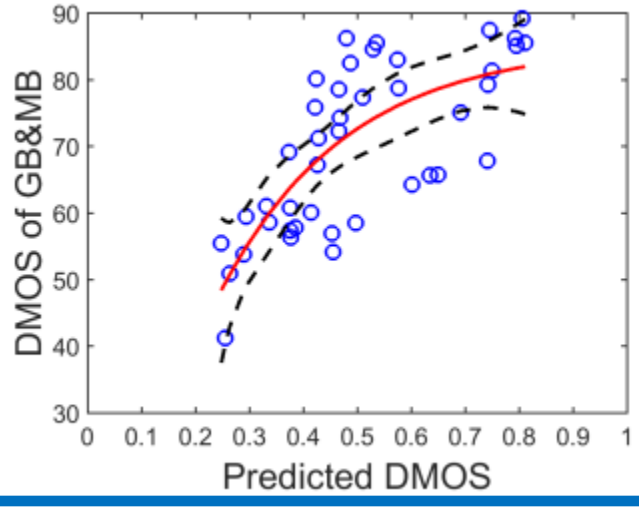
Experimental Results

Table 2. Performance over Gaussian Blur and Motion Blur

IQA Metrics	PLCC	SROCC	RMSE
SSIM [4]	0.8537	0.8481	7.1334
Q_s [5]	**0.8972**	**0.8856**	**6.7335**
SIQM [6]	0.8785	0.8750	6.9241
SPQA [3]	0.8687	0.8636	6.8262
Proposed	**0.8907**	**0.8846**	**6.7638**

Scatter Plots





Thank you!