Visual Saliency Based Image Quality Assessment

Project Presentation | EE698K | 16th April 2018

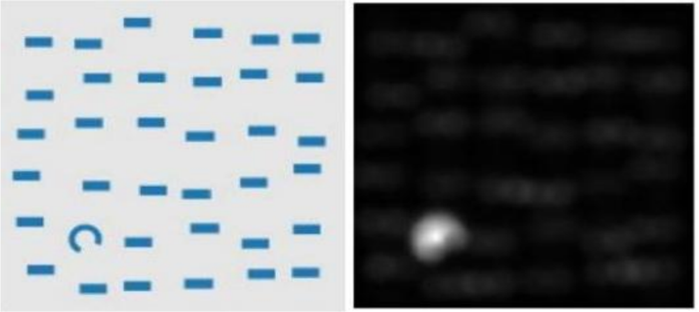
Gaurav Verma & Paridhi Maheshwari
Department of Electrical Engineering
Indian Institute of Technology Kanpur
{gverma, paridhi}@iitk.ac.in

Why we think this is important?

- **Visual Saliency:** The distinct subjective perceptual quality which makes parts of an image stand out from their surroundings and immediately grab our attention.

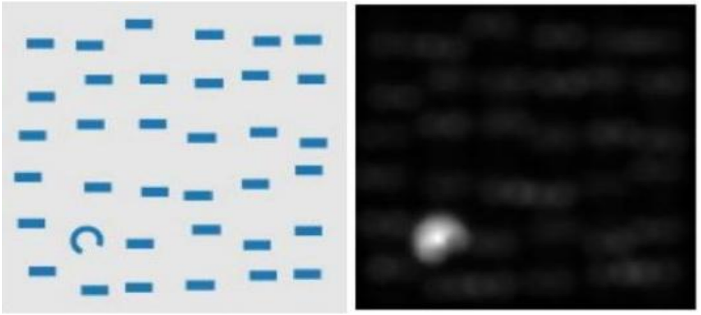
Why we think this is important?

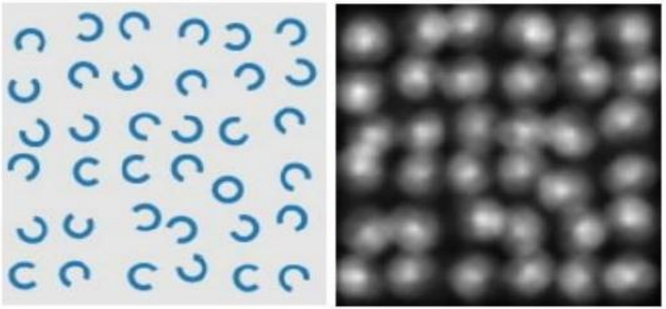
- **Visual Saliency:** The distinct subjective perceptual quality which makes parts of an image stand out from their surroundings and immediately grab our attention.



Why we think this is important?

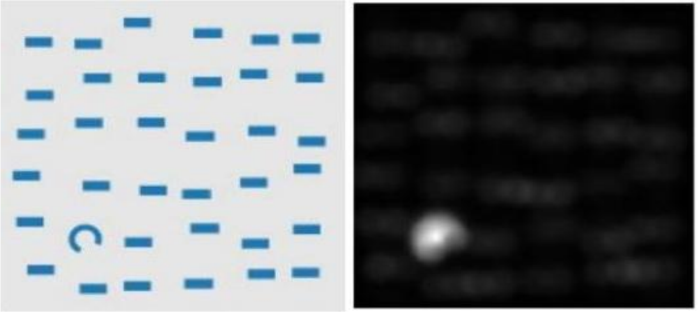
- **Visual Saliency:** The distinct subjective perceptual quality which makes parts of an image stand out from their surroundings and immediately grab our attention.

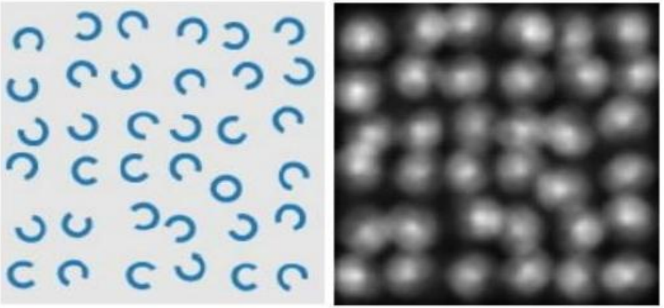




Why we think this is important?

- **Visual Saliency:** The distinct subjective perceptual quality which makes parts of an image stand out from their surroundings and immediately grab our attention.





- Spectral residual approach: analyze the log spectrum of an input image (Hou et al. [1])
- LSTM-based saliency attentive model (Cornia et al. [2])
- Generating saliency maps using a variant of pix2px (*proposed*)

Why we think this is important?

- **Image Quality Assessment:**
  - One of the most fundamental and yet challenging problems
  - IQA algorithms are designed to mimic the subjective judgements of humans
  - PSNR/MSE do not correlate well with human’s subjective fidelity rating
  - Recently several sophisticated IQA models have been introduced (SSIM)

Why we think this is important?

- **Image Quality Assessment:**
  - One of the most fundamental and yet challenging problems
  - IQA algorithms are designed to mimic the subjective judgements of humans
  - PSNR/MSE do not correlate well with human's subjective fidelity rating
  - Recently several sophisticated IQA models have been introduced (SSIM)

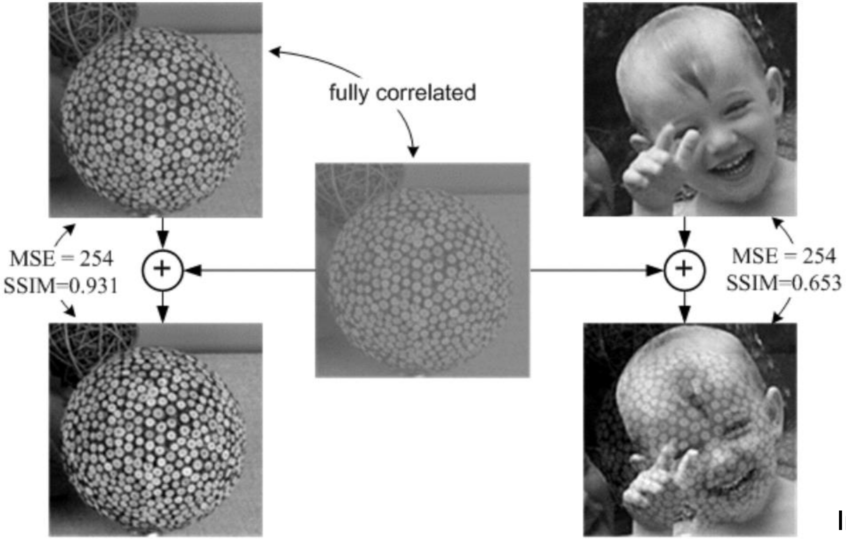


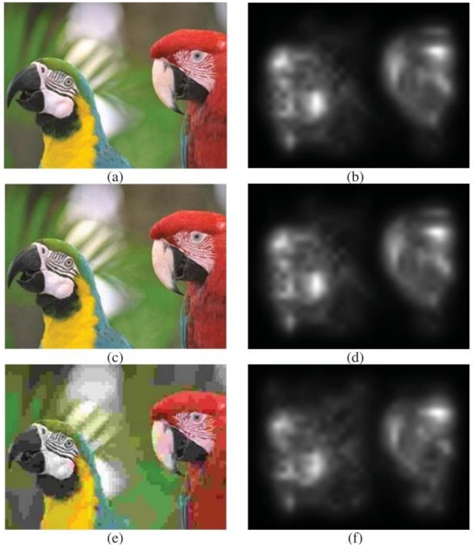
Image from Wang et al., 2009

What are others upto?

- **Image Quality Assessment using Visual Saliency Maps:**
  - Distortion occurring in an area that attracts the viewer’s attention is more annoying than in any other area, and should be weighted accordingly
  - Zhang et al. [3] have suggested that VS values change with distortions
  - They proposed an index that uses VS as a feature to compute the local similarity between the reference image and its distorted version

What are others upto?

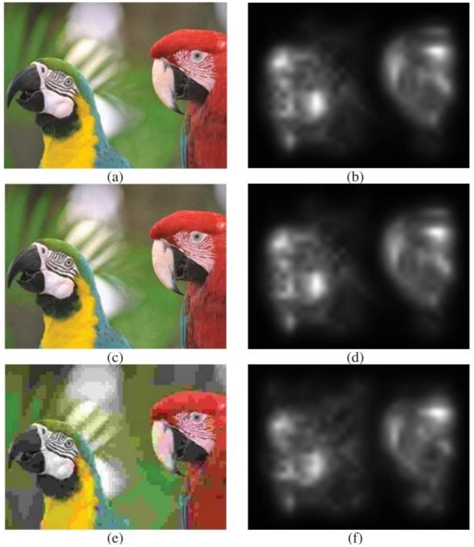
- **Image Quality Assessment using Visual Saliency Maps:**
  - Distortion occurring in an area that attracts the viewer’s attention is more annoying than in any other area, and should be weighted accordingly
  - Zhang et al. [3] have suggested that VS values change with distortions
  - They proposed an index that uses VS as a feature to compute the local similarity between the reference image and its distorted version



What are others upto?

- **Image Quality Assessment using Visual Saliency Maps:**
  - Distortion occurring in an area that attracts the viewer's attention is more annoying than in any other area, and should be weighted accordingly
  - Zhang et al. [3] have suggested that VS values change with distortions
  - They proposed an index that uses VS as a feature to compute the local similarity between the reference image and its distorted version

- **However,**
  - VS maps do not show any significant difference (in terms of mean square error) for the distortion types of contrast reduction (CR) and change of color saturation (CCS).
  - This occurs due to the normalization operations involved in VS computational models.



What we did? And why we did it?

- **A new Fourier-based distance metric**
  ↳ for comparing saliency maps
  ↳ for comparing heat maps

What we did? And why we did it?

- **A new Fourier-based distance metric**
  ↳ for comparing saliency maps
  ↳ for comparing heat maps
- Inputs: Two saliency maps S_1 and S_2
- Proposed Fourier-based distance:

$$d(S_1, S_2) = \sqrt{\sum_k \frac{(\mathfrak{F}(S_1 - S_2)_k)^2}{1 + (2\pi \mid k \mid)^2}}$$

where $k = (k_x, k_y),\ k_x \in \{0, \ldots, N-1\} \text{ and } k_y \in \{0, \ldots, M-1\}$

What we did? And why we did it?

- **A new Fourier-based distance metric**
  ↳ for comparing saliency maps
  ↳ for comparing heat maps
- Inputs: Two saliency maps S_1 and S_2
- Proposed Fourier-based distance:

$$d(S_1, S_2) = \sqrt{\sum_k \frac{(\mathfrak{F}(S_1 - S_2)_k)^2}{1 + (2\pi \mid k \mid)^2}}$$

where $k = (k_x, k_y),\ k_x \in \{0, \ldots, N-1\} \text{ and } k_y \in \{0, \ldots, M-1\}$

- In order to limit the value between 0 and 1, apply the transformation

$$\hat{d}(S_1, S_2) = \frac{d(S_1, S_2)}{1 + d(S_1, S_2)}$$

What we did? And why we did it?

- **A new Fourier-based distance metric**
  ↳ for comparing saliency maps
  ↳ for comparing heat maps
- Inputs: Two saliency maps S_1 and S_2
- Proposed Fourier-based distance:

$$d(S_1, S_2) = \sqrt{\sum_k \frac{(\mathfrak{F}(S_1 - S_2)_k)^2}{1 + (2\pi \mid k \mid)^2}} \quad \Bigg\}$$

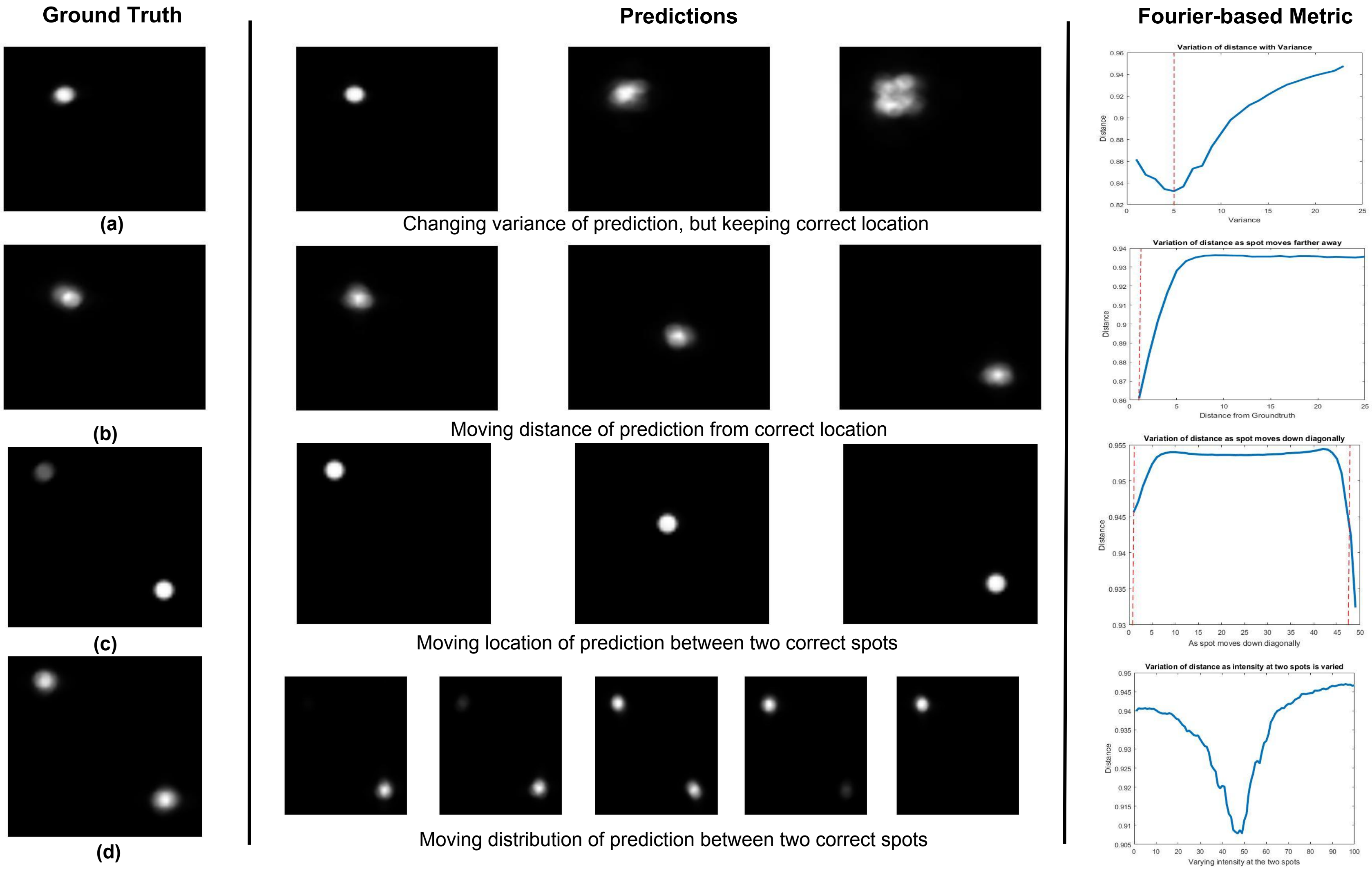
Properties

1. Symmetric ✓
2. Non-negative ✓
3. △ inequality ✓

where $k = (k_x, k_y),\ k_x \in \{0, \ldots, N-1\} \text{ and } k_y \in \{0, \ldots, M-1\}$

- In order to limit the value between 0 and 1, apply the transformation

$$\hat{d}(S_1, S_2) = \frac{d(S_1, S_2)}{1 + d(S_1, S_2)}$$



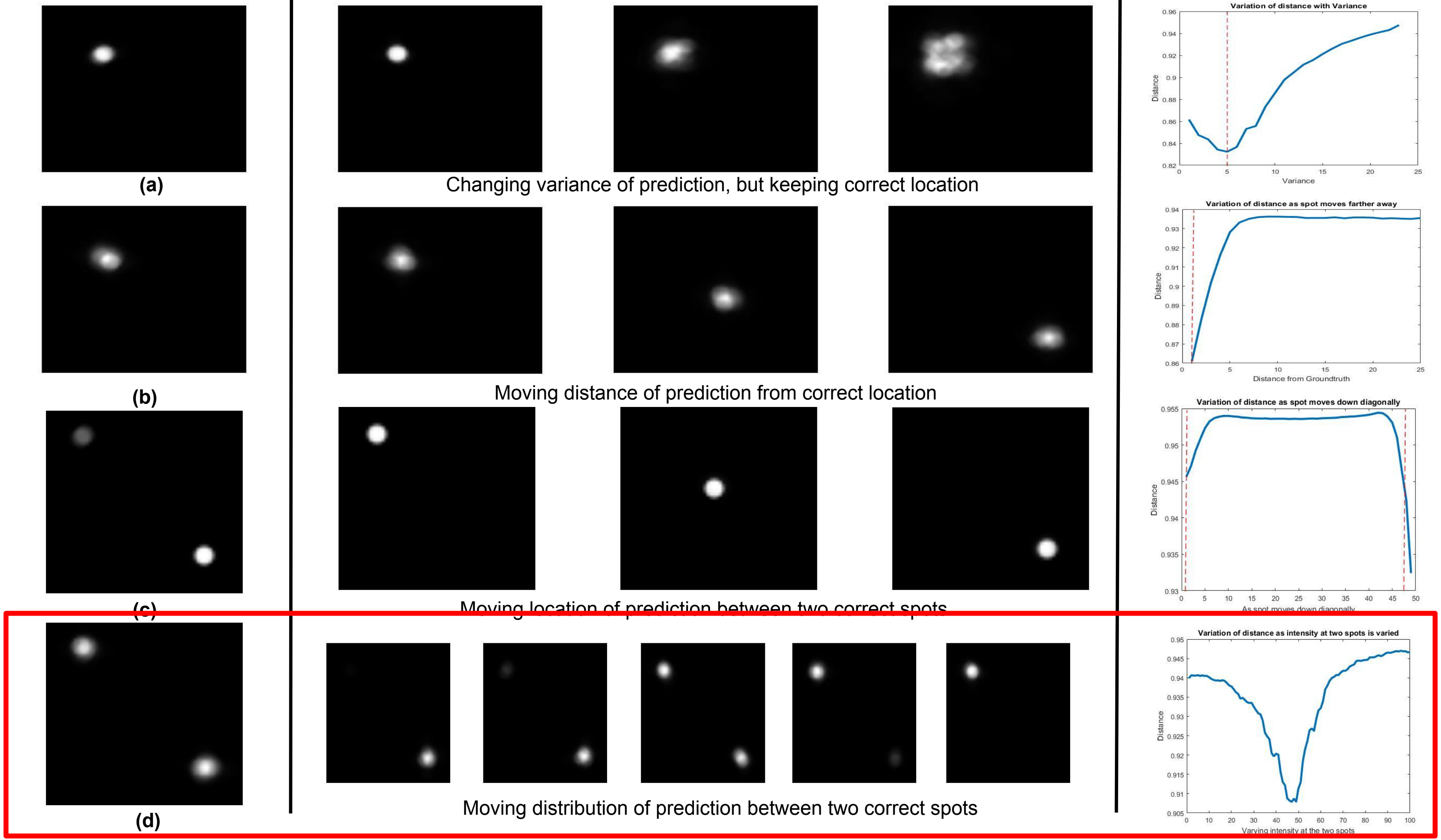
Changing variance of prediction, but keeping correct location

Moving distance of prediction from correct location

Moving location of prediction between two correct spots

Moving distribution of prediction between two correct spots

Ground Truth **Predictions** **Fourier-based Metric**



(a) Changing variance of prediction, but keeping correct location

(b) Moving distance of prediction from correct location

(c) Moving location of prediction between two correct spots

(d) Moving distribution of prediction between two correct spots

What we did? And why we did it?

- **Comparing saliency maps**

What we did? And why we did it?

- **Comparing saliency maps**

- The following metrics were investigated:
  - Area under the ROC Curve (AUC_Borji and AUC_Judd)
  - Pearson’s Correlation Coefficient (CC)
  - Fourier-based Metric (Fourier)
  - Kullback-Leibler Divergence (KL)
  - Mean Square Error (MSE)
  - Normalized Scanpath Saliency (NSS)
  - Similarity of histogram intersection (SIM)

What we did? And why we did it?

- **Comparing saliency maps**

- The following metrics were investigated:
  - Area under the ROC Curve (AUC_Borji and AUC_Judd)
  - Pearson’s Correlation Coefficient (CC)
  - Fourier-based Metric (Fourier) ← *proposed*
  - Kullback-Leibler Divergence (KL)
  - Mean Square Error (MSE) ← *VSI*
  - Normalized Scanpath Saliency (NSS)
  - Similarity of histogram intersection (SIM)

What we did? And why we did it?

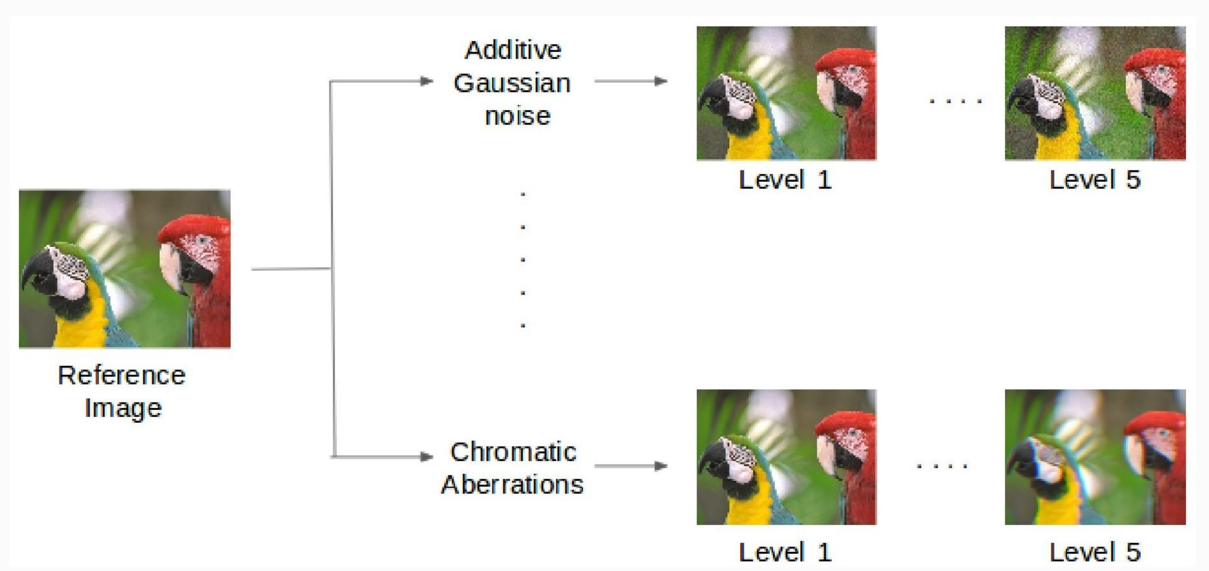
- **Dataset:** TID2013, a comprehensive dataset for IQA research

  There are a total of 25 images and for each reference image, there are 24 distortion types and 5 distortion levels for each distortion type.
  ⇛ 25 samples at a given distortion level and particular distortion type

What we did? And why we did it?

- **Dataset:** TID2013, a comprehensive dataset for IQA research

  There are a total of 25 images and for each reference image, there are 24 distortion types and 5 distortion levels for each distortion type.
  ⇛ 25 samples at a given distortion level and particular distortion type



What we did? And why we did it?

- **Dataset:** TID2013, a comprehensive dataset for IQA research

  There are a total of 25 images and for each reference image, there are 24 distortion types and 5 distortion levels for each distortion type.
  ⇛ 25 samples at a given distortion level and particular distortion type

- **Experiments:**
  Saliency maps were generated using Spectral Residual approach [1] and SAM-VGG [2].
  The metrics described above were computed between every reference image and the corresponding 24x5 distorted images.

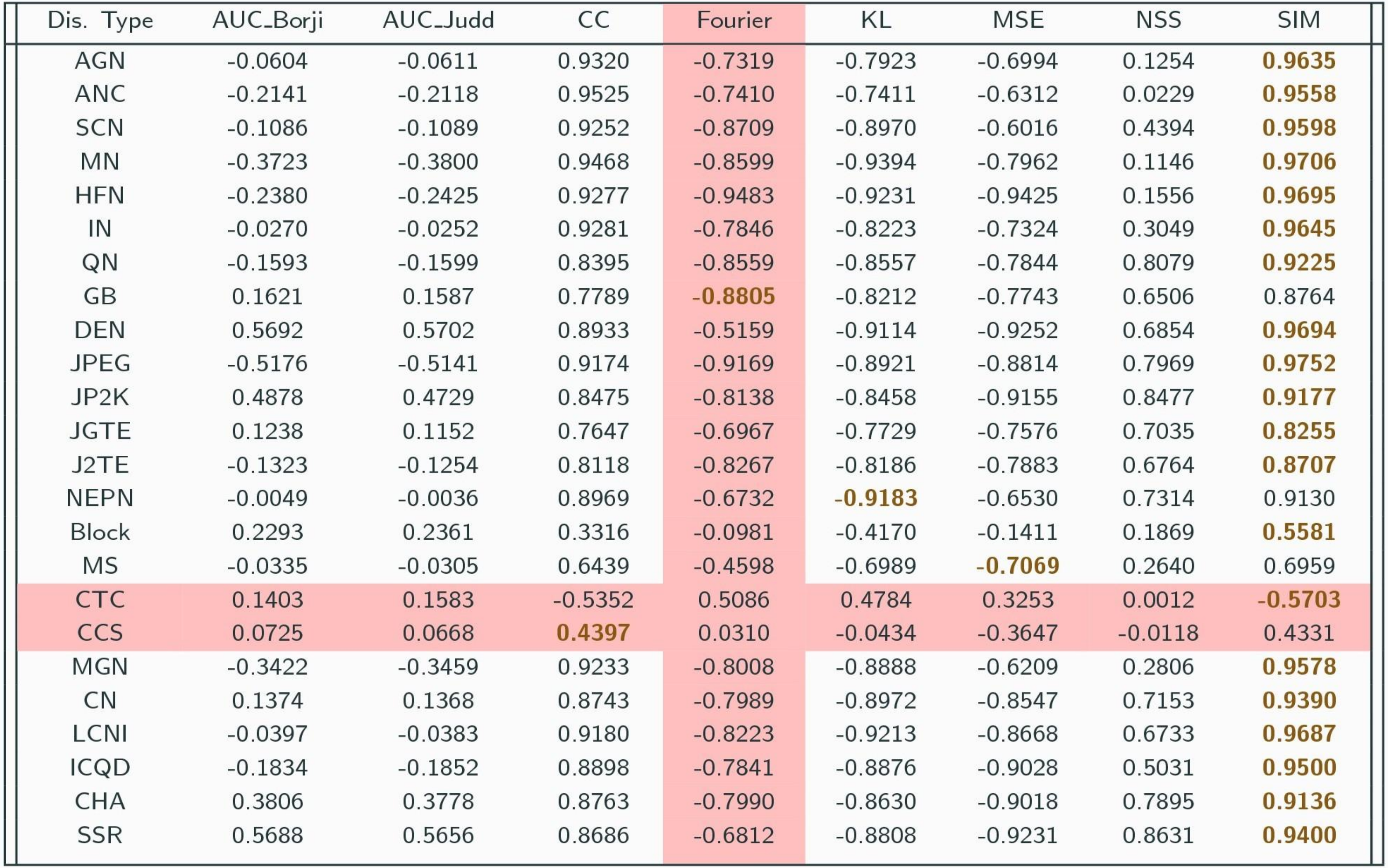
What we did? And why we did it?

- **Dataset:** TID2013, a comprehensive dataset for IQA research

  There are a total of 25 images and for each reference image, there are 24 distortion types and 5 distortion levels for each distortion type.
  ⇛ 25 samples at a given distortion level and particular distortion type

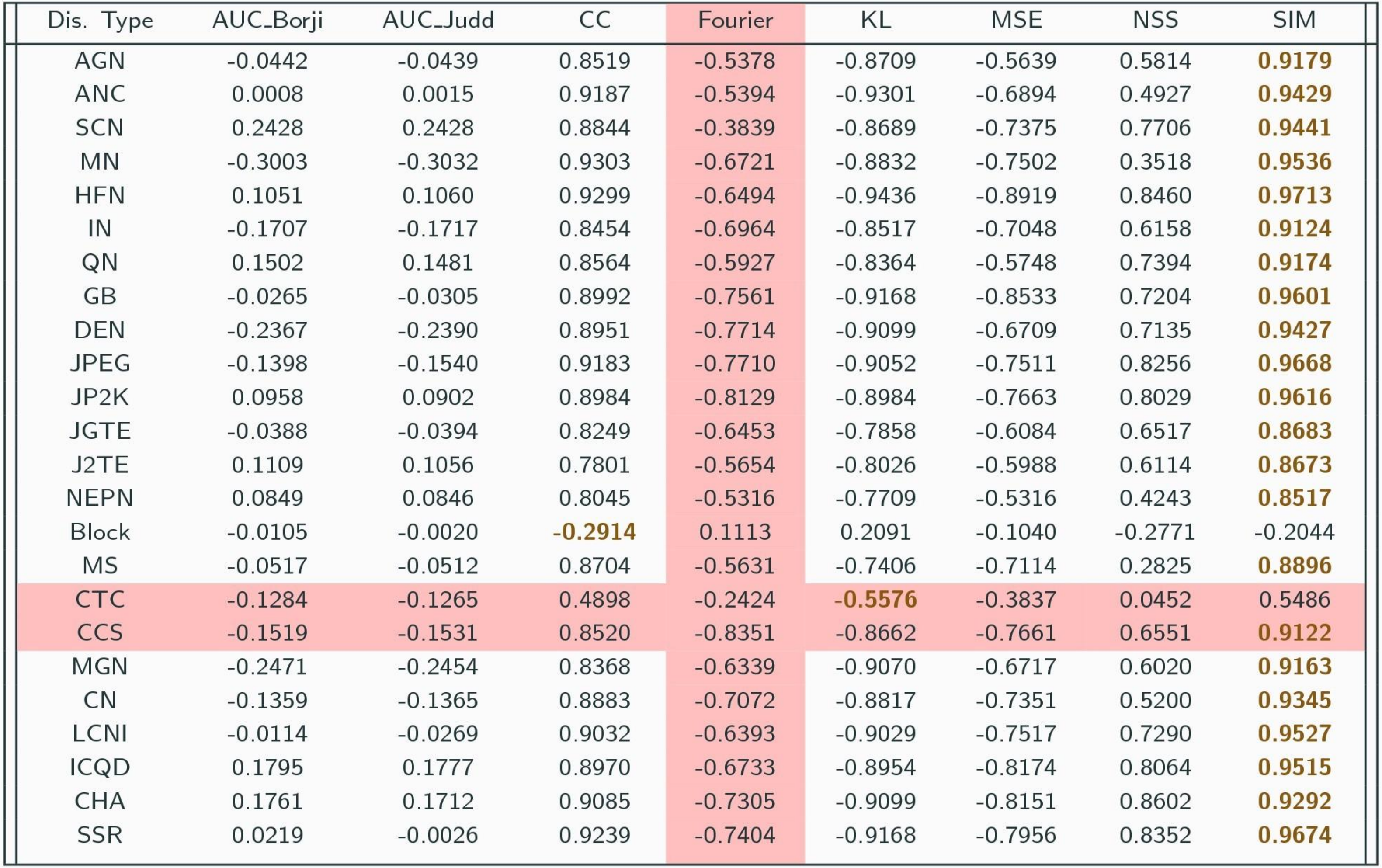
- **Experiments:**
  Saliency maps were generated using Spectral Residual approach [1] and SAM-VGG [2]. The metrics described above were computed between every reference image and the corresponding 24x5 distorted images.

- **Desired Properties**
  - High absolute value of the average correlation with subjective scores
  - Consistency in sign of the correlation values throughout the distortion types



Dis. Type	AUC_Borji	AUC_Judd	CC	Fourier	KL	MSE	NSS	SIM
AGN	-0.0604	-0.0611	0.9320	-0.7319	-0.7923	-0.6994	0.1254	**0.9635**
ANC	-0.2141	-0.2118	0.9525	-0.7410	-0.7411	-0.6312	0.0229	**0.9558**
SCN	-0.1086	-0.1089	0.9252	-0.8709	-0.8970	-0.6016	0.4394	**0.9598**
MN	-0.3723	-0.3800	0.9468	-0.8599	-0.9394	-0.7962	0.1146	**0.9706**
HFN	-0.2380	-0.2425	0.9277	-0.9483	-0.9231	-0.9425	0.1556	**0.9695**
IN	-0.0270	-0.0252	0.9281	-0.7846	-0.8223	-0.7324	0.3049	**0.9645**
QN	-0.1593	-0.1599	0.8395	-0.8559	-0.8557	-0.7844	0.8079	**0.9225**
GB	0.1621	0.1587	0.7789	**-0.8805**	-0.8212	-0.7743	0.6506	0.8764
DEN	0.5692	0.5702	0.8933	-0.5159	-0.9114	-0.9252	0.6854	**0.9694**
JPEG	-0.5176	-0.5141	0.9174	-0.9169	-0.8921	-0.8814	0.7969	**0.9752**
JP2K	0.4878	0.4729	0.8475	-0.8138	-0.8458	-0.9155	0.8477	**0.9177**
JGTE	0.1238	0.1152	0.7647	-0.6967	-0.7729	-0.7576	0.7035	**0.8255**
J2TE	-0.1323	-0.1254	0.8118	-0.8267	-0.8186	-0.7883	0.6764	**0.8707**
NEPN	-0.0049	-0.0036	0.8969	-0.6732	**-0.9183**	-0.6530	0.7314	0.9130
Block	0.2293	0.2361	0.3316	-0.0981	-0.4170	-0.1411	0.1869	**0.5581**
MS	-0.0335	-0.0305	0.6439	-0.4598	-0.6989	**-0.7069**	0.2640	0.6959
CTC	0.1403	0.1583	-0.5352	0.5086	0.4784	0.3253	0.0012	**-0.5703**
CCS	0.0725	0.0668	**0.4397**	0.0310	-0.0434	-0.3647	-0.0118	0.4331
MGN	-0.3422	-0.3459	0.9233	-0.8008	-0.8888	-0.6209	0.2806	**0.9578**
CN	0.1374	0.1368	0.8743	-0.7989	-0.8972	-0.8547	0.7153	**0.9390**
LCNI	-0.0397	-0.0383	0.9180	-0.8223	-0.9213	-0.8668	0.6733	**0.9687**
ICQD	-0.1834	-0.1852	0.8898	-0.7841	-0.8876	-0.9028	0.5031	**0.9500**
CHA	0.3806	0.3778	0.8763	-0.7990	-0.8630	-0.9018	0.7895	**0.9136**
SSR	0.5688	0.5656	0.8686	-0.6812	-0.8808	-0.9231	0.8631	**0.9400**

Table 1: Correlation with Subjective Scores: Spectral Residual



Dis. Type	AUC_Borji	AUC_Judd	CC	Fourier	KL	MSE	NSS	SIM
AGN	-0.0442	-0.0439	0.8519	-0.5378	-0.8709	-0.5639	0.5814	**0.9179**
ANC	0.0008	0.0015	0.9187	-0.5394	-0.9301	-0.6894	0.4927	**0.9429**
SCN	0.2428	0.2428	0.8844	-0.3839	-0.8689	-0.7375	0.7706	**0.9441**
MN	-0.3003	-0.3032	0.9303	-0.6721	-0.8832	-0.7502	0.3518	**0.9536**
HFN	0.1051	0.1060	0.9299	-0.6494	-0.9436	-0.8919	0.8460	**0.9713**
IN	-0.1707	-0.1717	0.8454	-0.6964	-0.8517	-0.7048	0.6158	**0.9124**
QN	0.1502	0.1481	0.8564	-0.5927	-0.8364	-0.5748	0.7394	**0.9174**
GB	-0.0265	-0.0305	0.8992	-0.7561	-0.9168	-0.8533	0.7204	**0.9601**
DEN	-0.2367	-0.2390	0.8951	-0.7714	-0.9099	-0.6709	0.7135	**0.9427**
JPEG	-0.1398	-0.1540	0.9183	-0.7710	-0.9052	-0.7511	0.8256	**0.9668**
JP2K	0.0958	0.0902	0.8984	-0.8129	-0.8984	-0.7663	0.8029	**0.9616**
JGTE	-0.0388	-0.0394	0.8249	-0.6453	-0.7858	-0.6084	0.6517	**0.8683**
J2TE	0.1109	0.1056	0.7801	-0.5654	-0.8026	-0.5988	0.6114	**0.8673**
NEPN	0.0849	0.0846	0.8045	-0.5316	-0.7709	-0.5316	0.4243	**0.8517**
Block	-0.0105	-0.0020	**-0.2914**	0.1113	0.2091	-0.1040	-0.2771	-0.2044
MS	-0.0517	-0.0512	0.8704	-0.5631	-0.7406	-0.7114	0.2825	**0.8896**
CTC	-0.1284	-0.1265	0.4898	-0.2424	**-0.5576**	-0.3837	0.0452	0.5486
CCS	-0.1519	-0.1531	0.8520	-0.8351	-0.8662	-0.7661	0.6551	**0.9122**
MGN	-0.2471	-0.2454	0.8368	-0.6339	-0.9070	-0.6717	0.6020	**0.9163**
CN	-0.1359	-0.1365	0.8883	-0.7072	-0.8817	-0.7351	0.5200	**0.9345**
LCNI	-0.0114	-0.0269	0.9032	-0.6393	-0.9029	-0.7517	0.7290	**0.9527**
ICQD	0.1795	0.1777	0.8970	-0.6733	-0.8954	-0.8174	0.8064	**0.9515**
CHA	0.1761	0.1712	0.9085	-0.7305	-0.9099	-0.8151	0.8602	**0.9292**
SSR	0.0219	-0.0026	0.9239	-0.7404	-0.9168	-0.7956	0.8352	**0.9674**

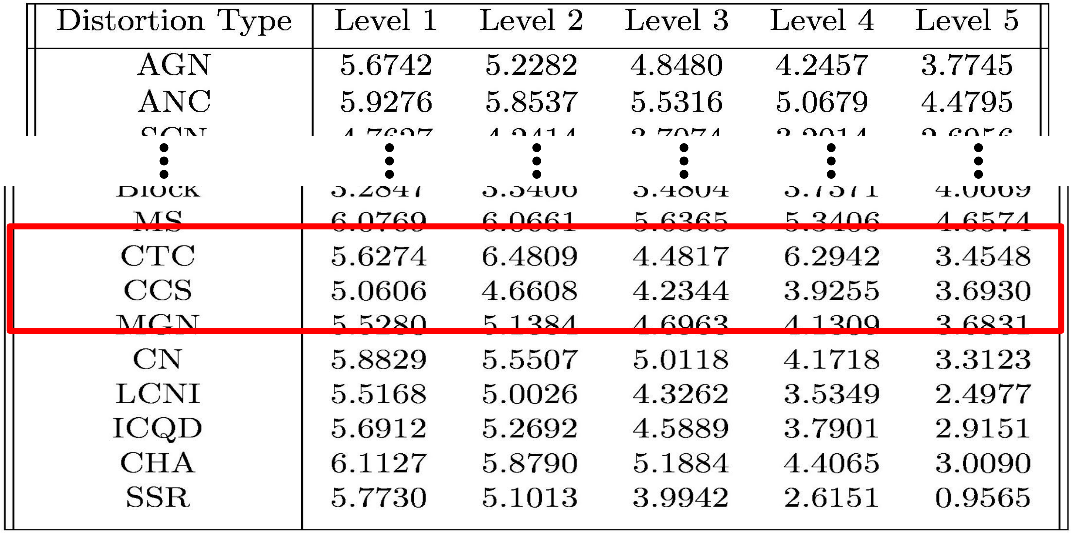
Table 2: Correlation with Subjective Scores: SAM-VGG

Observations:

- SIM shows a better correlation with subjective scores for most distortion types.
- Proposed Fourier-based metric outperforms several of the conventional metrics.
- The subjective scores of the distortion type CTC and CCS are very unpredictable.

Observations:

- SIM shows a better correlation with subjective scores for most distortion types.
- Proposed Fourier-based metric outperforms several of the conventional metrics.
- The subjective scores of the distortion type CTC and CCS are very unpredictable.



Distortion Type	Level 1	Level 2	Level 3	Level 4	Level 5
AGN	5.6742	5.2282	4.8480	4.2457	3.7745
ANC	5.9276	5.8537	5.5316	5.0679	4.4795
SCN	4.7627	4.3414	3.7974	3.3014	2.6956
⋮	⋮	⋮	⋮	⋮	⋮
Block	5.2847	5.3406	5.4804	5.7571	4.0669
MS	6.0769	6.0661	5.6365	5.3406	4.6574
CTC	5.6274	6.4809	4.4817	6.2942	3.4548
CCS	5.0606	4.6608	4.2344	3.9255	3.6930
MGN	5.5280	5.1384	4.6963	4.1309	3.6831
CN	5.8829	5.5507	5.0118	4.1718	3.3123
LCNI	5.5168	5.0026	4.3262	3.5349	2.4977
ICQD	5.6912	5.2692	4.5889	3.7901	2.9151
CHA	6.1127	5.8790	5.1884	4.4065	3.0090
SSR	5.7730	5.1013	3.9942	2.6151	0.9565

Table 4: Subjective Scores (averaged over all images), for 24 distortion types and 5 levels.

Roadblocks and the Way Around:

- In case of contrast reduction, the perceived saliency patterns change only for the first 4 seconds and become constant thereafter [6]

- The TID2013 experiments are not in direct compliance with these. Hence, due to limited availability of time and resources, we faced a serious roadblock to proceed further.

Roadblocks and the Way Around:

- In case of contrast reduction, the perceived saliency patterns change only for the first 4 seconds and become constant thereafter [6]

- The TID2013 experiments are not in direct compliance with these. Hence, due to limited availability of time and resources, we faced a serious roadblock to proceed further.

- However, we realized that Zhang et al. [3] had exploited the gradient maps of the original image to make up for the inability of their saliency maps in capturing contrast change!

Image-to-Image Translation using cGANs [5]

- GANs learn a loss function rather than using an existing one
- The generator G is trained to produce outputs that cannot be distinguished from “real” images by an adversarially trained discriminator, D which is trained to do as well as possible at detecting the generator’s “fakes”.
- Conditional GANs (cGANs) learn a mapping from observed image ***x*** and random noise vector ***z*** to ***y: f(x, z)***

Image-to-Image Translation using cGANs [5]

- GANs learn a loss function rather than using an existing one
- The generator G is trained to produce outputs that cannot be distinguished from “real” images by an adversarially trained discriminator, D which is trained to do as well as possible at detecting the generator’s “fakes”.
- Conditional GANs (cGANs) learn a mapping from observed image ***x*** and random noise vector ***z*** to ***y: f(x, z)***
- Objective of conditional GANs:

$$\mathfrak{L}_{cGAN} = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z)))]$$

Image-to-Image Translation using cGANs [5]

- GANs learn a loss function rather than using an existing one
- The generator G is trained to produce outputs that cannot be distinguished from “real” images by an adversarially trained discriminator, D which is trained to do as well as possible at detecting the generator’s “fakes”.
- Conditional GANs (cGANs) learn a mapping from observed image ***x*** and random noise vector ***z*** to ***y: f(x, z)***
- Objective of conditional GANs:

$$\mathfrak{L}_{cGAN} = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z))]$$

- Beneficial to mix the GAN objective with a more traditional loss, such as L2 distance. The discriminator’s job remains unchanged, but the generator is tasked to not only fool the discriminator but also to be near the ground truth output in an L2 sense.

Image-to-Image Translation using cGANs [5]

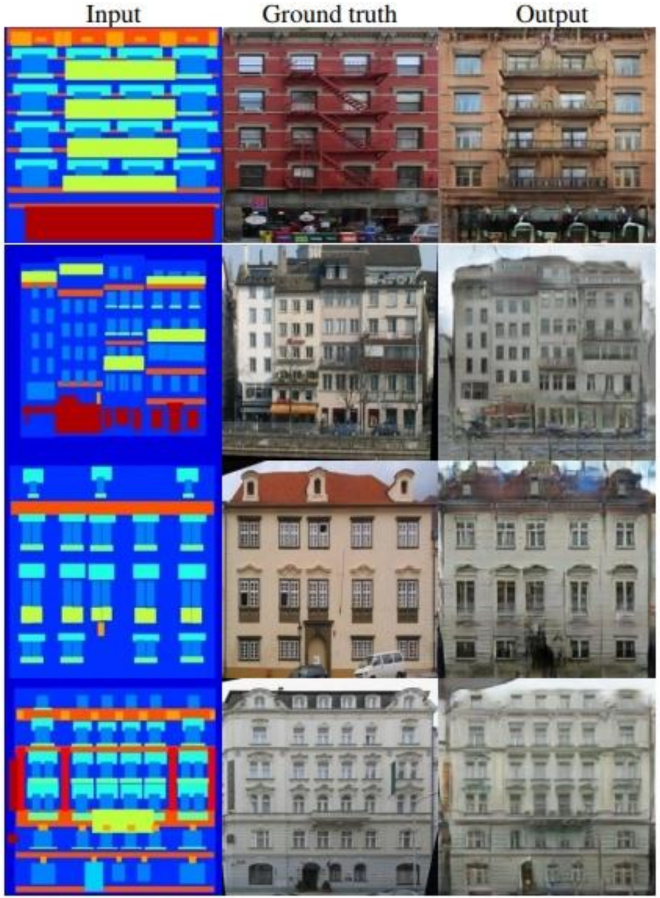
- GANs learn a loss function rather than using an existing one
- The generator G is trained to produce outputs that cannot be distinguished from “real” images by an adversarially trained discriminator, D which is trained to do as well as possible at detecting the generator’s “fakes”.
- Conditional GANs (cGANs) learn a mapping from observed image ***x*** and random noise vector ***z*** to ***y: f(x, z)***
- Objective of conditional GANs:

$$\mathfrak{L}_{cGAN} = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x,G(x,z)))]$$

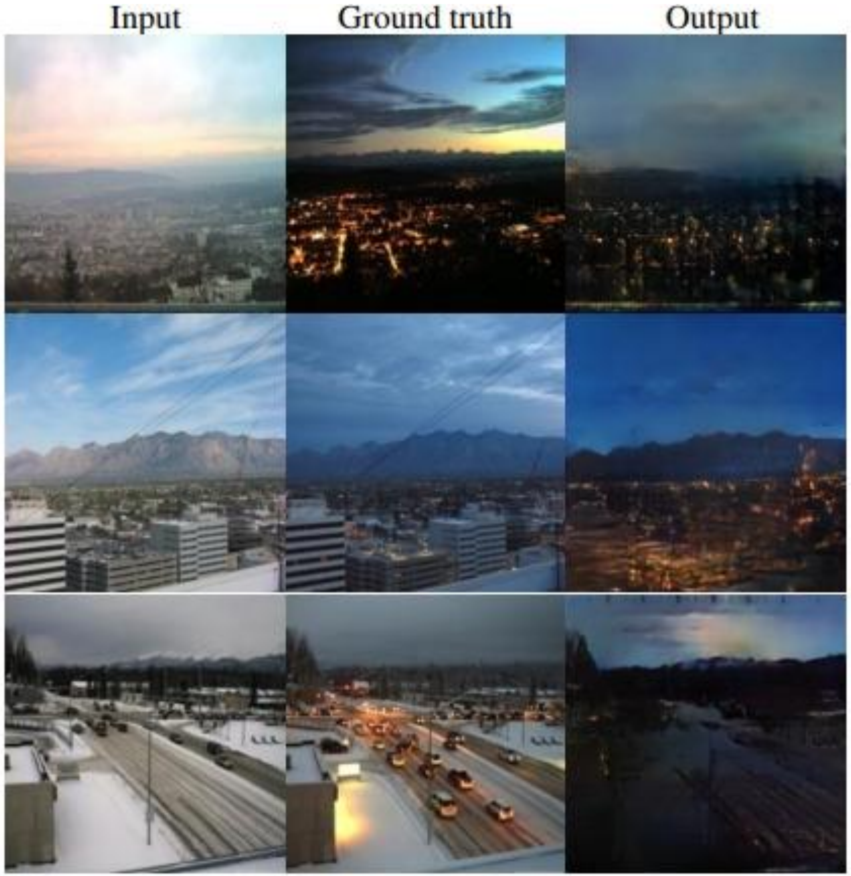
- Beneficial to mix the GAN objective with a more traditional loss, such as L2 distance. The discriminator’s job remains unchanged, but the generator is tasked to not only fool the discriminator but also to be near the ground truth output in an L2 sense.
- However, authors prefer L1 over L2 as it encourages less blurring. Final objective:

$$G^* = \arg\ \min_G\ \max_D\ \mathfrak{L}_{cGAN}(G,D) + \lambda \mathfrak{L}_{L1}(G)$$
$$\text{where, } \mathfrak{L}_{L1}(G) = \mathbb{E}[||\ y - G(x,z)\ ||_1]$$

Image-to-Image Translation using cGANs [5]



Input Ground truth Output

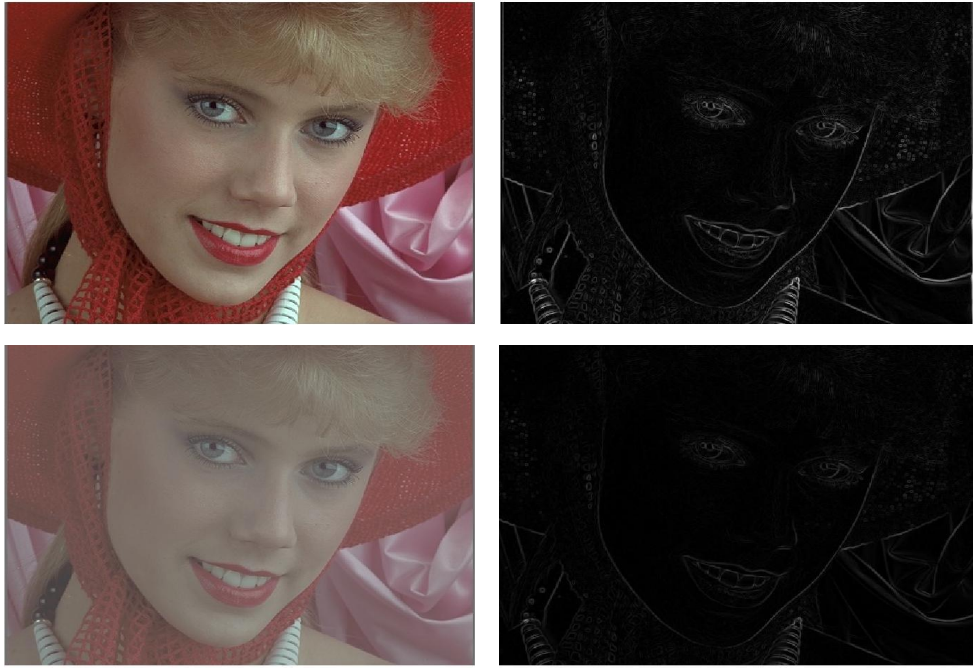


Input Ground truth Output

Image Credits: Isola et al.

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss

- Zhang et al. [3] tried to make up for the ineffectiveness of VS maps in detecting contrast change, by taking the gradient maps of the original images into account
  - But, you still need the original images!



Pix2Pix for Generating Saliency Maps + Edge Preserving Loss

- Zhang et al. [3] tried to make up for the ineffectiveness of VS maps in detecting contrast change, by taking the gradient maps of the original images into account
  - But, you still need the original images!
- We too modified the final objective to take into account the gradient, while generating saliency maps

~~$G^* = \arg\min_G \max_D \mathfrak{L}_{cGAN}(G, D) + \lambda \mathfrak{L}_{L1}(G)$~~

$$G^* = \arg\min_G \max_D \mathfrak{L}_{cGAN}(G, D) + \lambda_1 \mathfrak{L}_{L1}(G) + \underline{\lambda_2 \mathfrak{L}_{EPL}(G)}$$

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss

- Zhang et al. [3] tried to make up for the ineffectiveness of VS maps in detecting contrast change, by taking the gradient maps of the original images into account
  - But, you still need the original images!
- We too modified the final objective to take into account the gradient, while generating saliency maps

~~$G^* = \arg\min_G \max_D \mathfrak{L}_{cGAN}(G, D) + \lambda \mathfrak{L}_{L1}(G)$~~

$$G^* = \arg\min_G \max_D \mathfrak{L}_{cGAN}(G, D) + \lambda_1 \mathfrak{L}_{L1}(G) + \underline{\lambda_2 \mathfrak{L}_{EPL}(G)}$$

- $\lambda_2 \mathfrak{L}_{EPL}(G)$ denotes the binary cross-entropy loss between gradient maps of the original image and its saliency map. The gradients were treated as 1 or 0, based on a threshold.
- The VS model learns to change the maps as edges in original images change.

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss

- Zhang et al. [3] tried to make up for the ineffectiveness of VS maps in detecting contrast change, by taking the gradient maps of the original images into account
    - But, you still need the original images!
- We too modified the final objective to take into account the gradient, while generating saliency maps

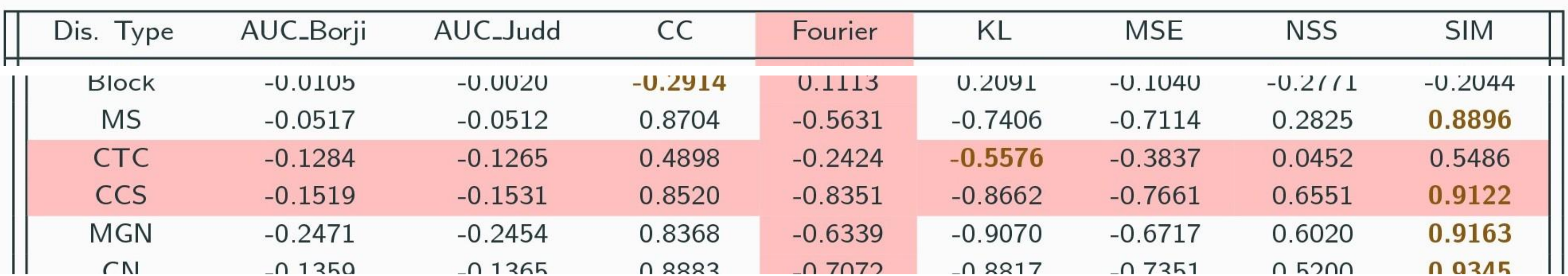
~~$G^* = \arg\min_G \max_D \mathfrak{L}_{cGAN}(G, D) + \lambda \mathfrak{L}_{L1}(G)$~~

$$G^* = \arg\min_G \max_D \mathfrak{L}_{cGAN}(G, D) + \lambda_1 \mathfrak{L}_{L1}(G) + \underline{\lambda_2 \mathfrak{L}_{EPL}(G)}$$

- $\lambda_2 \mathfrak{L}_{EPL}(G)$ denotes the binary cross-entropy loss between gradient maps of the original image and its saliency map. The gradients were treated as 1 or 0, based on a threshold.
- The VS model learns to change the maps as edges in original images change.

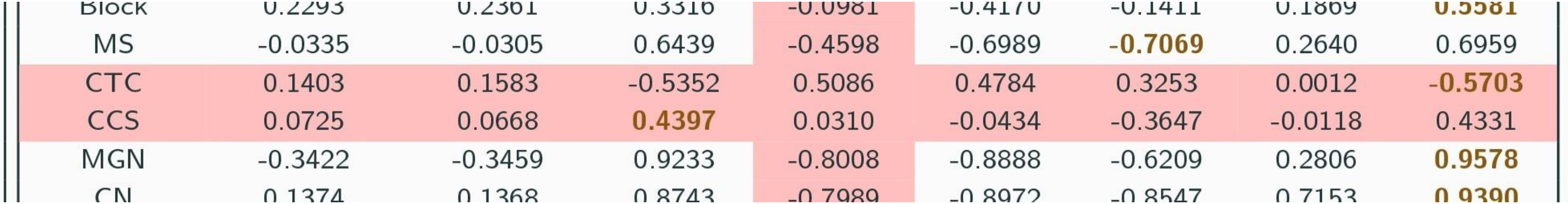
- *Does it generate saliency maps that are good for image quality assessment?*

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss



Dis. Type	AUC_Borji	AUC_Judd	CC	Fourier	KL	MSE	NSS	SIM
Block	-0.0105	-0.0020	**-0.2914**	0.1113	0.2091	-0.1040	-0.2771	-0.2044
MS	-0.0517	-0.0512	0.8704	-0.5631	-0.7406	-0.7114	0.2825	**0.8896**
CTC	-0.1284	-0.1265	0.4898	-0.2424	**-0.5576**	-0.3837	0.0452	0.5486
CCS	-0.1519	-0.1531	0.8520	-0.8351	-0.8662	-0.7661	0.6551	**0.9122**
MGN	-0.2471	-0.2454	0.8368	-0.6339	-0.9070	-0.6717	0.6020	**0.9163**
CN	-0.1359	-0.1365	0.8883	-0.7072	-0.8817	-0.7351	0.5200	**0.9345**

Table 1: Correlation with Subjective Scores: Spectral Residual

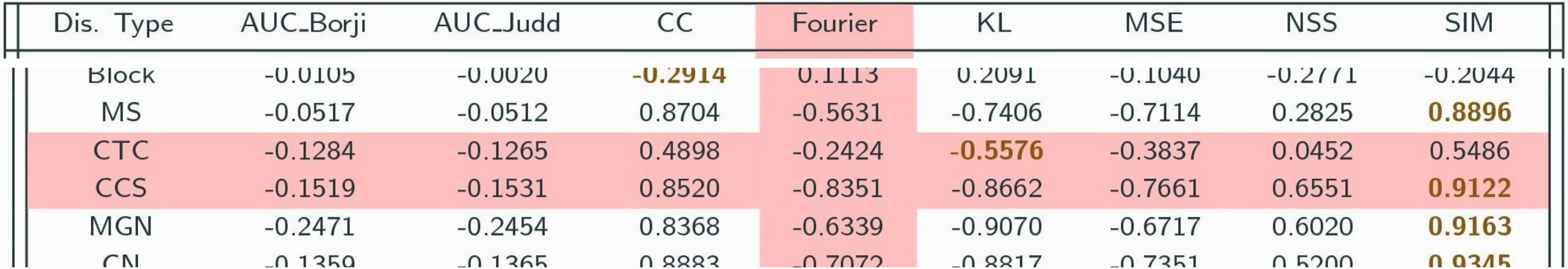


Block	0.2293	0.2361	0.3316	-0.0981	-0.4170	-0.1411	0.1869	**0.5581**
MS	-0.0335	-0.0305	0.6439	-0.4598	-0.6989	**-0.7069**	0.2640	0.6959
CTC	0.1403	0.1583	-0.5352	0.5086	0.4784	0.3253	0.0012	**-0.5703**
CCS	0.0725	0.0668	**0.4397**	0.0310	-0.0434	-0.3647	-0.0118	0.4331
MGN	-0.3422	-0.3459	0.9233	-0.8008	-0.8888	-0.6209	0.2806	**0.9578**
CN	0.1374	0.1368	0.8743	-0.7989	-0.8972	-0.8547	0.7153	**0.9390**

Table 2: Correlation with Subjective Scores: SAM-VGG

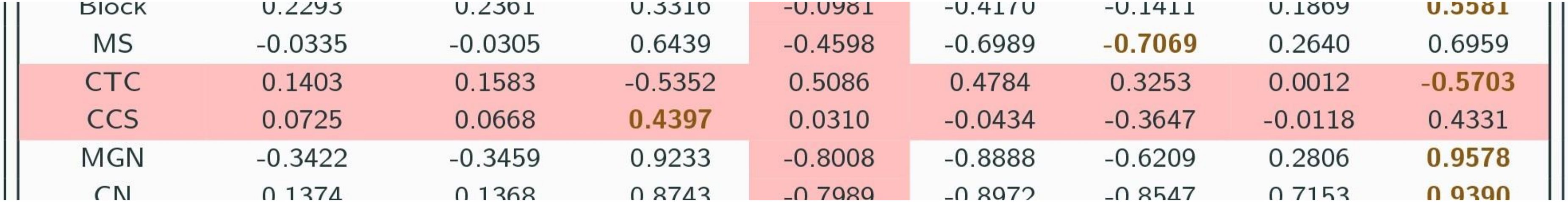
- *Does it generate saliency maps that are good for image quality assessment?* **Yes! ***

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss



Dis. Type	AUC_Borji	AUC_Judd	CC	Fourier	KL	MSE	NSS	SIM
Block	-0.0105	-0.0020	**-0.2914**	0.1113	0.2091	-0.1040	-0.2771	-0.2044
MS	-0.0517	-0.0512	0.8704	-0.5631	-0.7406	-0.7114	0.2825	**0.8896**
CTC	-0.1284	-0.1265	0.4898	-0.2424	**-0.5576**	-0.3837	0.0452	0.5486
CCS	-0.1519	-0.1531	0.8520	-0.8351	-0.8662	-0.7661	0.6551	**0.9122**
MGN	-0.2471	-0.2454	0.8368	-0.6339	-0.9070	-0.6717	0.6020	**0.9163**
CN	-0.1359	-0.1365	0.8883	-0.7072	-0.8817	-0.7351	0.5200	**0.9345**

Table 1: Correlation with Subjective Scores: Spectral Residual



Block	0.2293	0.2361	0.3316	-0.0981	-0.4170	-0.1411	0.1869	**0.5581**
MS	-0.0335	-0.0305	0.6439	-0.4598	-0.6989	**-0.7069**	0.2640	0.6959
CTC	0.1403	0.1583	-0.5352	0.5086	0.4784	0.3253	0.0012	**-0.5703**
CCS	0.0725	0.0668	**0.4397**	0.0310	-0.0434	-0.3647	-0.0118	0.4331
MGN	-0.3422	-0.3459	0.9233	-0.8008	-0.8888	-0.6209	0.2806	**0.9578**
CN	0.1374	0.1368	0.8743	-0.7989	-0.8972	-0.8547	0.7153	**0.9390**

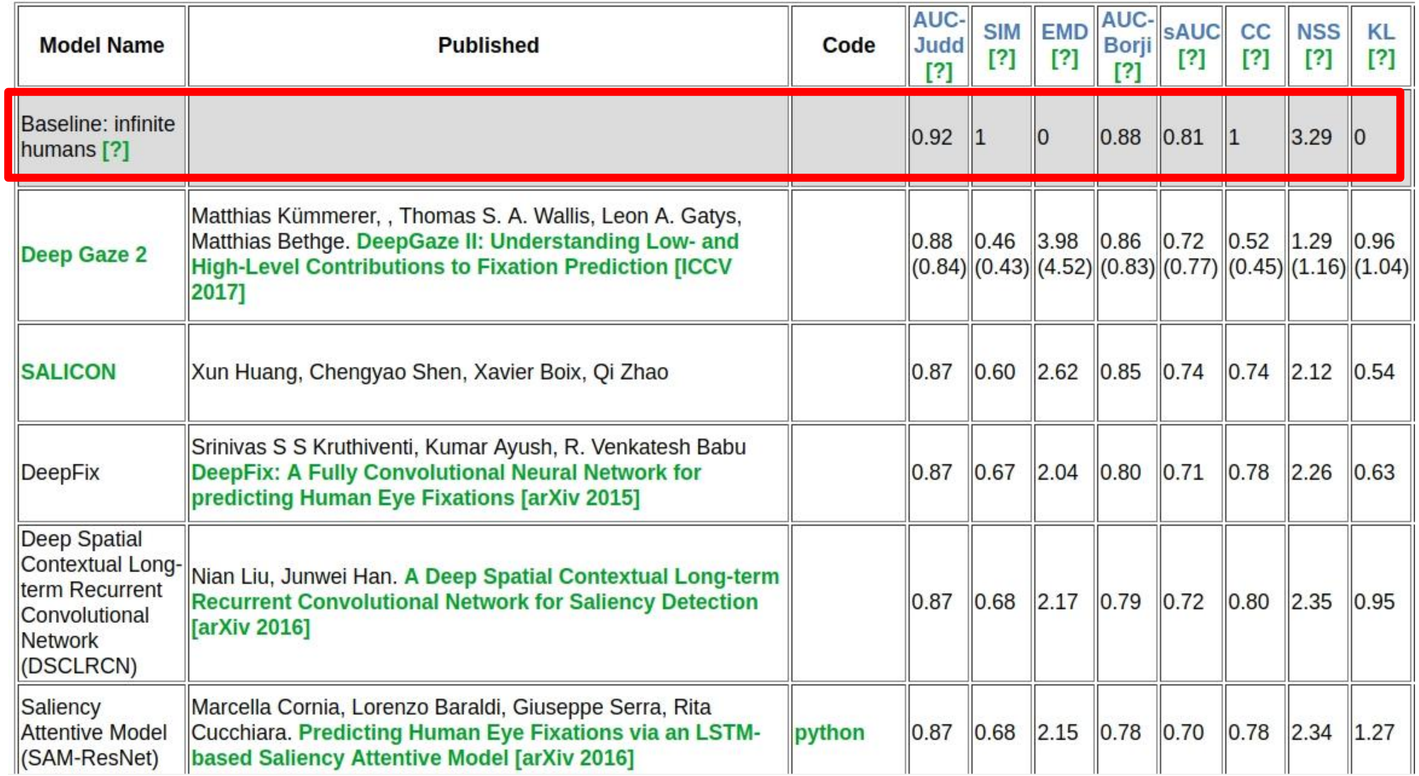
Table 2: Correlation with Subjective Scores: SAM-VGG

CTC	**-0.4289**	**0.0368**	**0.6827**	**-0.2191**	**-0.3642**	**-0.3768**	**0.4571**	**0.3061**
CCS	**-0.2293**	**0.0429**	**0.5321**	**0.0468**	**-0.1279**	**-0.2987**	**0.2842**	**0.6821**

Table 3: Correlation with Subjective Scores: pix2pix + EPL (*proposed*)

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss

- But, *are the generated saliency maps good for general purposes?* **Yes!**

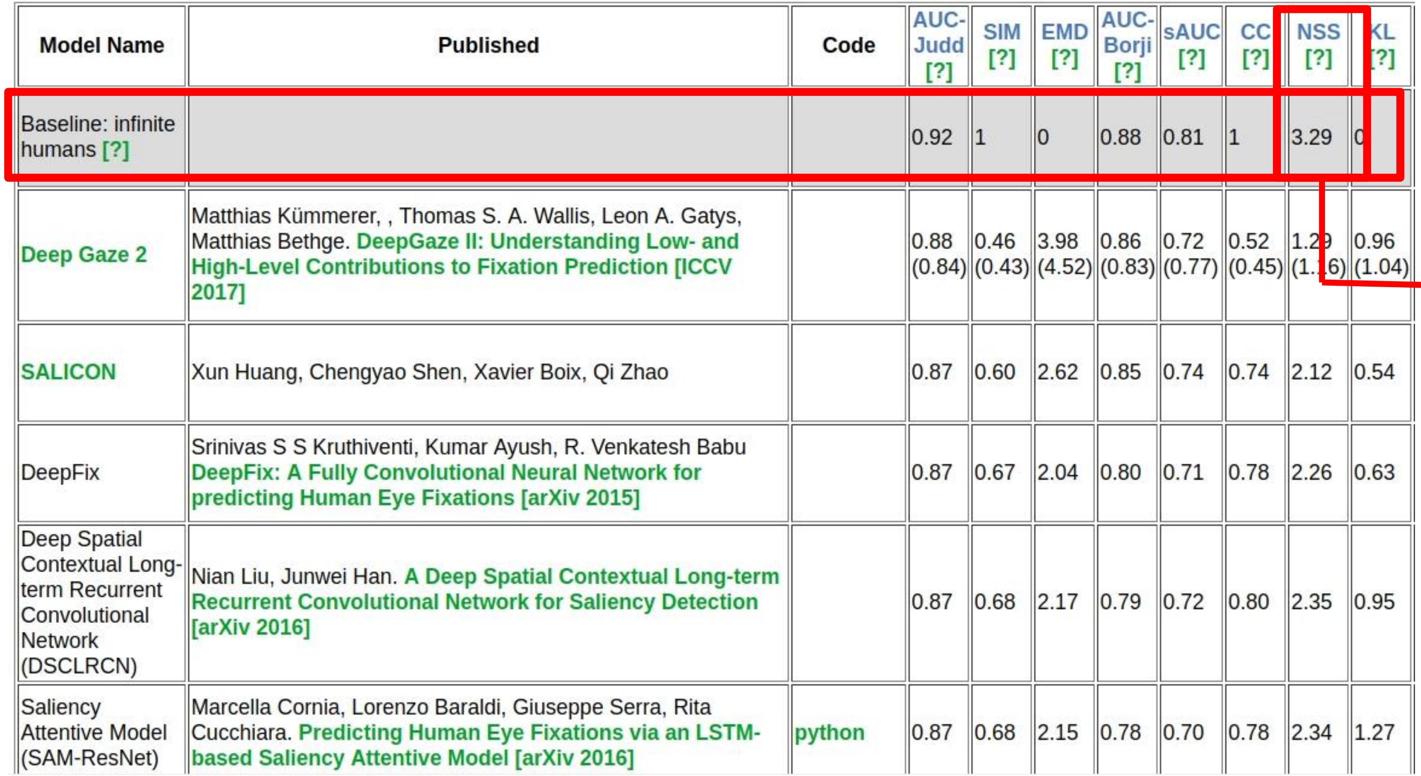


Model Name	Published	Code	AUC-Judd [?]	SIM [?]	EMD [?]	AUC-Borji [?]	sAUC [?]	CC [?]	NSS [?]	KL [?]
Baseline: infinite humans [?]			0.92	1	0	0.88	0.81	1	3.29	0
Deep Gaze 2	Matthias Kümmerer, , Thomas S. A. Wallis, Leon A. Gatys, Matthias Bethge. **DeepGaze II: Understanding Low- and High-Level Contributions to Fixation Prediction [ICCV 2017]**		0.88 (0.84)	0.46 (0.43)	3.98 (4.52)	0.86 (0.83)	0.72 (0.77)	0.52 (0.45)	1.29 (1.16)	0.96 (1.04)
SALICON	Xun Huang, Chengyao Shen, Xavier Boix, Qi Zhao		0.87	0.60	2.62	0.85	0.74	0.74	2.12	0.54
DeepFix	Srinivas S S Kruthiventi, Kumar Ayush, R. Venkatesh Babu **DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations [arXiv 2015]**		0.87	0.67	2.04	0.80	0.71	0.78	2.26	0.63
Deep Spatial Contextual Long-term Recurrent Convolutional Network (DSCLRCN)	Nian Liu, Junwei Han. **A Deep Spatial Contextual Long-term Recurrent Convolutional Network for Saliency Detection [arXiv 2016]**		0.87	0.68	2.17	0.79	0.72	0.80	2.35	0.95
Saliency Attentive Model (SAM-ResNet)	Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, Rita Cucchiara. **Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model [arXiv 2016]**	**python**	0.87	0.68	2.15	0.78	0.70	0.78	2.34	1.27

Image Credits:
http://saliency.mit.edu

Pix2Pix for Generating Saliency Maps + Edge Preserving Loss

- But, *are the generated saliency maps good for general purposes?* **Yes!**



Model Name	Published	Code	AUC-Judd [?]	SIM [?]	EMD [?]	AUC-Borji [?]	sAUC [?]	CC [?]	NSS [?]	KL [?]
Baseline: infinite humans [?]			0.92	1	0	0.88	0.81	1	3.29	0
Deep Gaze 2	Matthias Kümmerer, , Thomas S. A. Wallis, Leon A. Gatys, Matthias Bethge. DeepGaze II: Understanding Low- and High-Level Contributions to Fixation Prediction [ICCV 2017]		0.88 (0.84)	0.46 (0.43)	3.98 (4.52)	0.86 (0.83)	0.72 (0.77)	0.52 (0.45)	1.29 (1.16)	0.96 (1.04)
SALICON	Xun Huang, Chengyao Shen, Xavier Boix, Qi Zhao		0.87	0.60	2.62	0.85	0.74	0.74	2.12	0.54
DeepFix	Srinivas S S Kruthiventi, Kumar Ayush, R. Venkatesh Babu DeepFix: A Fully Convolutional Neural Network for predicting Human Eye Fixations [arXiv 2015]		0.87	0.67	2.04	0.80	0.71	0.78	2.26	0.63
Deep Spatial Contextual Long-term Recurrent Convolutional Network (DSCLRCN)	Nian Liu, Junwei Han. A Deep Spatial Contextual Long-term Recurrent Convolutional Network for Saliency Detection [arXiv 2016]		0.87	0.68	2.17	0.79	0.72	0.80	2.35	0.95
Saliency Attentive Model (SAM-ResNet)	Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, Rita Cucchiara. Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model [arXiv 2016]	python	0.87	0.68	2.15	0.78	0.70	0.78	2.34	1.27

NSS: 2.31

KL: 0.67

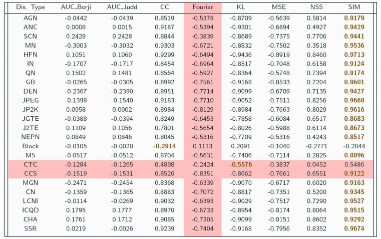
Image Credits:
http://saliency.mit.edu

Summary

$$d(S_1, S_2) = \sqrt{\sum_k \frac{(\mathfrak{F}(S_1 - S_2)_k)^2}{1 + (2\pi \mid k \mid)^2}}$$

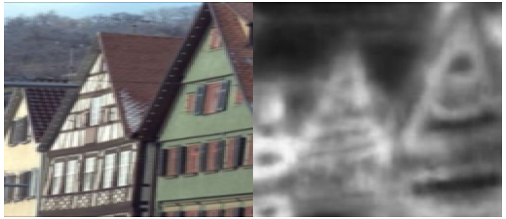
$$\hat{d}(S_1, S_2) = \frac{d(S_1, S_2)}{1 + d(S_1, S_2)}$$

A new Fourier-based distance metric for comparing visual saliency maps



Dis. Type	AUC_Borji	AUC_Judd	CC	Fourier	KL	MSE	NSS	SIM
AGN	-0.0442	-0.0439	0.8519	-0.5378	-0.8709	-0.5639	0.5814	**0.9179**
ANC	0.0008	0.0015	0.9187	-0.5394	-0.9301	-0.6894	0.4927	**0.9429**
SCN	0.2428	0.2428	0.8844	-0.3839	-0.8689	-0.7375	0.7706	**0.9441**
MN	-0.3003	-0.3032	0.9303	-0.6721	-0.8832	-0.7502	0.3518	**0.9536**
HFN	0.1051	0.1060	0.9299	-0.6494	-0.9436	-0.8919	0.8460	**0.9713**
IN	-0.1707	-0.1717	0.8454	-0.6964	-0.8517	-0.7048	0.6158	**0.9124**
QN	0.1502	0.1481	0.8564	-0.5927	-0.8364	-0.5748	0.7394	**0.9174**
GB	-0.0265	-0.0305	0.8992	-0.7561	-0.9168	-0.8533	0.7204	**0.9601**
DEN	-0.2367	-0.2390	0.8951	-0.7714	-0.9099	-0.6709	0.7135	**0.9427**
JPEG	-0.1398	-0.1540	0.9183	-0.7710	-0.9052	-0.7511	0.8256	**0.9668**
JP2K	0.0958	0.0902	0.8984	-0.8129	-0.8984	-0.7663	0.8029	**0.9616**
JGTE	-0.0388	-0.0394	0.8249	-0.6453	-0.7858	-0.6084	0.6517	**0.8683**
J2TE	0.1109	0.1056	0.7801	-0.5654	-0.8026	-0.5988	0.6114	**0.8673**
NEPN	0.0849	0.0846	0.8045	-0.5316	-0.7709	-0.5316	0.4243	**0.8517**
Block	-0.0105	-0.0020	**-0.2914**	0.1113	0.2091	-0.1040	-0.2771	-0.2044
MS	-0.0517	-0.0512	0.8704	-0.5631	-0.7406	-0.7114	0.2825	**0.8896**
CTC	-0.1284	-0.1265	0.4898	-0.2424	**-0.5576**	-0.3837	0.0452	0.5486
CCS	-0.1519	-0.1531	0.8520	-0.8351	-0.8662	-0.7661	0.6551	**0.9122**
MGN	-0.2471	-0.2454	0.8368	-0.6339	-0.9070	-0.6717	0.6020	**0.9163**
CN	-0.1359	-0.1365	0.8883	-0.7072	-0.8817	-0.7351	0.5200	**0.9345**
LCNI	-0.0114	-0.0269	0.9032	-0.6393	-0.9029	-0.7517	0.7290	**0.9527**
ICQD	0.1795	0.1777	0.8970	-0.6733	-0.8954	-0.8174	0.8064	**0.9515**
CHA	0.1761	0.1712	0.9085	-0.7305	-0.9099	-0.8151	0.8602	**0.9292**
SSR	0.0219	-0.0026	0.9239	-0.7404	-0.9168	-0.7956	0.8352	**0.9674**

Extensive experimentation with existing saliency maps comparison metrics



Proposed a model to generate saliency maps that can be for image quality assessment

Code Contributions

Our code:

1. Generation of saliency maps using Spectral-Residual method
2. All experiments pertaining to Fourier based metric
3. Evaluation and comparison of different metrics
4. Modification to the pix2pix code (including edge preserving loss)

Off the shelf codes:

1. SAM-VCG Saliency map generation code: [Link]
2. Saliency metric codes: [Link]
3. Pix2pix code: [Link]

References

- [1] Hou, Xiaodi, and Liqing Zhang. "Saliency detection: A spectral residual approach." *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007.

- [2] Cornia, Marcella, et al. "Predicting human eye fixations via an LSTM-based saliency attentive model." *arXiv preprint arXiv:1611.09571* (2016).

- [3] Zhang, Lin, Ying Shen, and Hongyu Li. "VSI: A visual saliency-induced index for perceptual image quality assessment." *IEEE Transactions on Image Processing* 23.10 (2014): 4270-4281.

- [4] Bylinskii, Zoya, et al. "What do different evaluation metrics tell us about saliency models?." *arXiv preprint arXiv:1604.03605*(2016).

- [5] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *arXiv preprint* (2017).

- [6] Einhäuser, Wolfgang, and Peter König. "Does luminance‐contrast contribute to a saliency map for overt visual attention?." *European Journal of Neuroscience* 17.5 (2003): 1089-1097.