

# Deep Learning VS. Traditional Algorithms For Saliency Prediction Of Distorted Images

Xin Zhao<sup>1</sup> (Presenter), Hanhe Lin<sup>2</sup>, Pengfei Guo<sup>3</sup>, Dietmar Saupe<sup>2</sup> and Hantao Liu<sup>1</sup>

<sup>1</sup>School of Computer Science and Informatics, Cardiff University, United Kingdom

<sup>2</sup>Department of Computer and Information Science, University of Konstanz, Germany

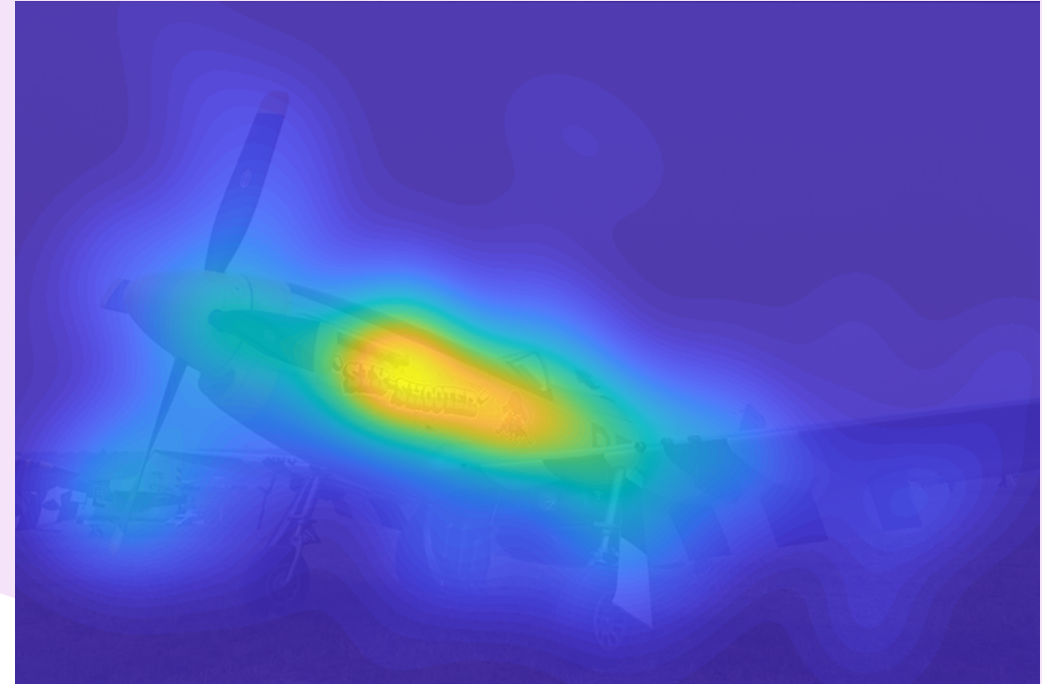
<sup>3</sup>School of Computational Science, Zhongkai University of Agriculture and Engineering, China



# What is the saliency map for this image?



Plane  
Image type: original  
(High Quality)

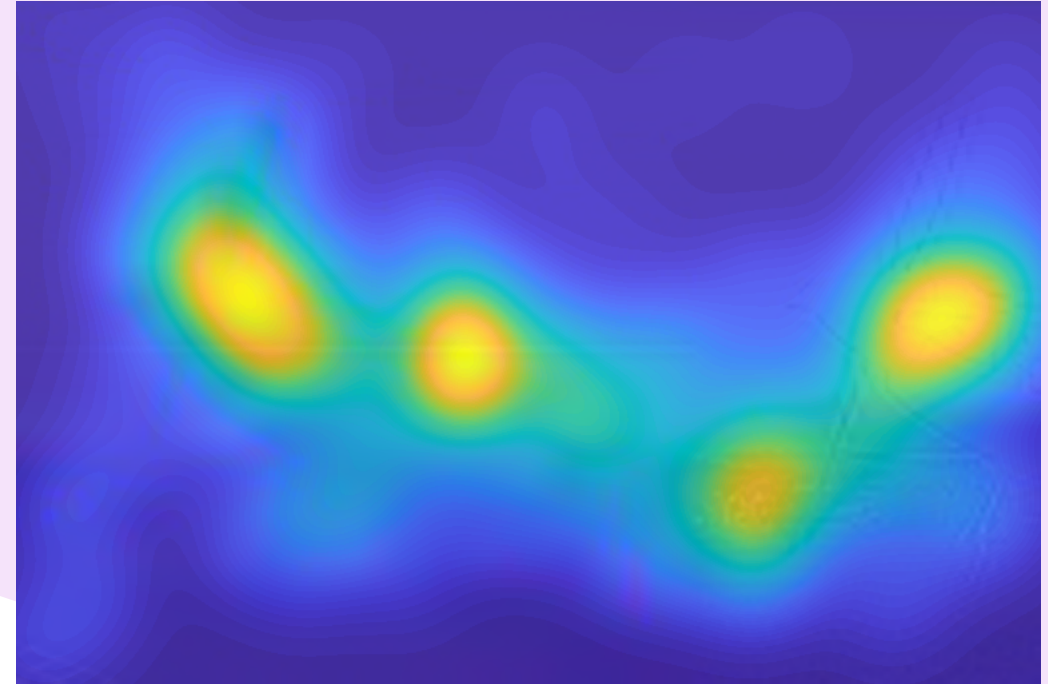


Plane  
Image type: saliency map

# How about this image?



Plane  
Distortion type: Fast Fading  
Distortion level: 3  
(Low quality)



Plane  
Image type: saliency map

Original



(a)

High quality



(b)

Medium quality

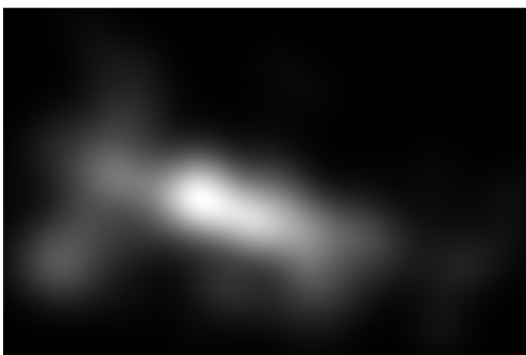


(c)

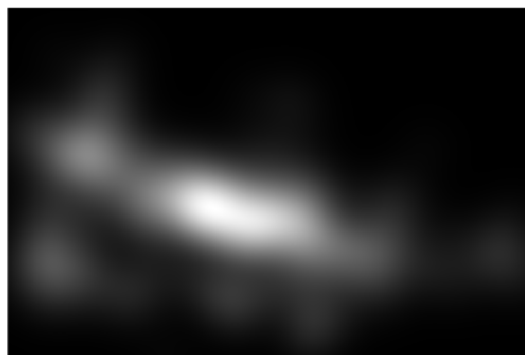
Low quality



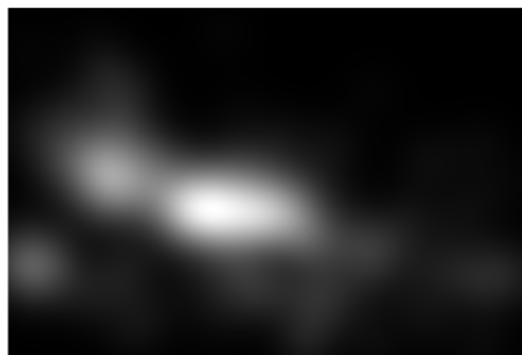
(d)



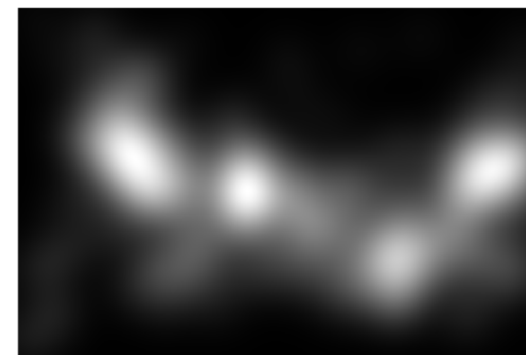
Reference



DMSS=36



DMSS=42



DMSS=64

# What questions bring out this paper?

- Do **image quality** affect the **saliency prediction**
- Whether and to what extent state-of-the-art methods are beneficial for saliency prediction of **distorted** images
- Will the ability of deep learning and traditional algorithms be different in predicting saliency, based on an **IQA-aware saliency dataset** (SIQ288)

# Contributions

- In this paper, we carry out an evaluation of state-of-the-art saliency models, including **5 deep learning models** and **5 traditional models** by using an **IQA-aware** saliency benchmark, i.e. the SIQ288 database.
- Building on the results of our analyses and cross-comparisons, we offer **guidelines** for choosing saliency models and approaches for IQA applications.

# IQA-aware saliency benchmark, SIQ288

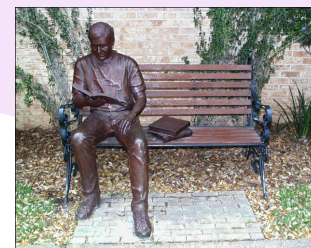
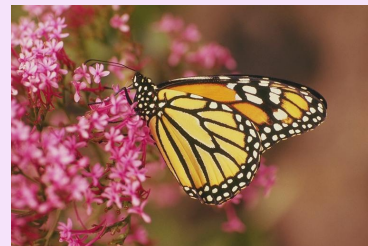
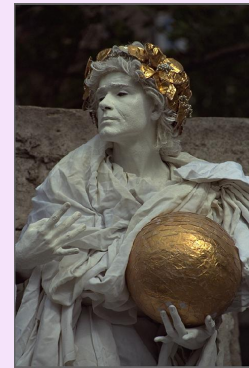
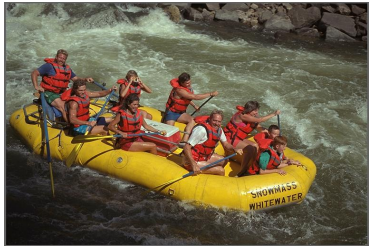
- Consists total of 288 images
  - **18** pristine images
  - Each pristine image has **3** level of distortion (i.e. **Low, Medium and High distortion**)
  - and **5** different types of distortion (i.e. Fast Fading (**FF**), Gaussian Blur (**GBLUR**), JPEG Compression (**JPEG**), JPEG2000 Compression (**JP2K**), and White Noise (**WN**)).
- Saliency maps were obtained via eye-tracking of 160 human observers under totally lab-controlled environment.

Ref: W. Zhang, A. Borji, Z. Wang, P. L. Callet, and H. Liu, "The Application of Visual Saliency Models in Objective Image Quality Assessment: A Statistical Evaluation," IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 6, pp. 1266–1278, Jun 2016.

W. Zhang, Y. Tian, X. Zha, and H. Liu, "Benchmark in state-of-the-art visual saliency models for image quality assessment," in IEEE International Conference on Acoustics, Speech and Signal Processing, May 2016, vol. 2016-May, pp. 1090–1094.

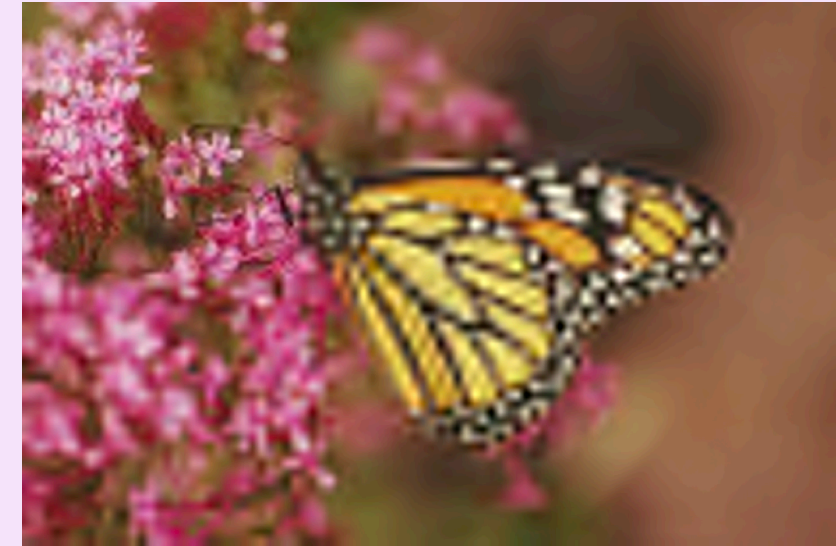
W. Zhang and H. Liu, "Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications," IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2424–2437, May 2017.

# SIQ288-examples-prestine images





# SIQ288-examples-3 distortion levels

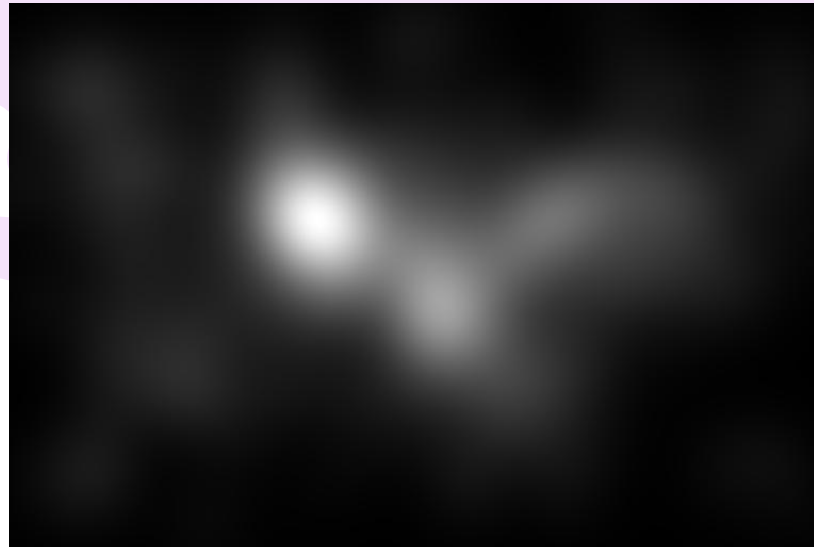
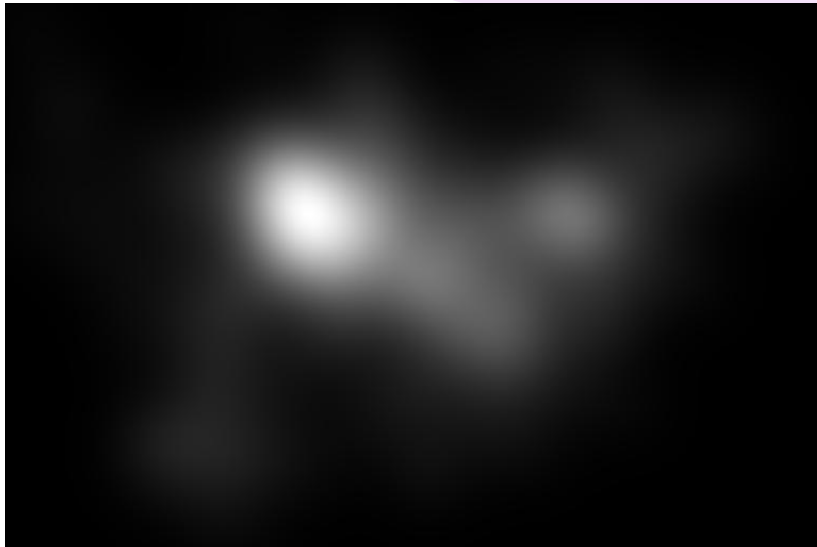


Ref: W. Zhang, A. Borji, Z. Wang, P. L. Callet, and H. Liu, "The Application of Visual Saliency Models in Objective Image Quality Assessment: A Statistical Evaluation," IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 6, pp. 1266–1278, Jun 2016.

W. Zhang, Y. Tian, X. Zha, and H. Liu, "Benchmark in state-of-the-art visual saliency models for image quality assessment," in IEEE International Conference on Acoustics, Speech and Signal Processing, May 2016, vol. 2016-May, pp. 1090–1094.

W. Zhang and H. Liu, "Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications," IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2424–2437, May 2017.

# SIQ288-examples-3 distortion levels

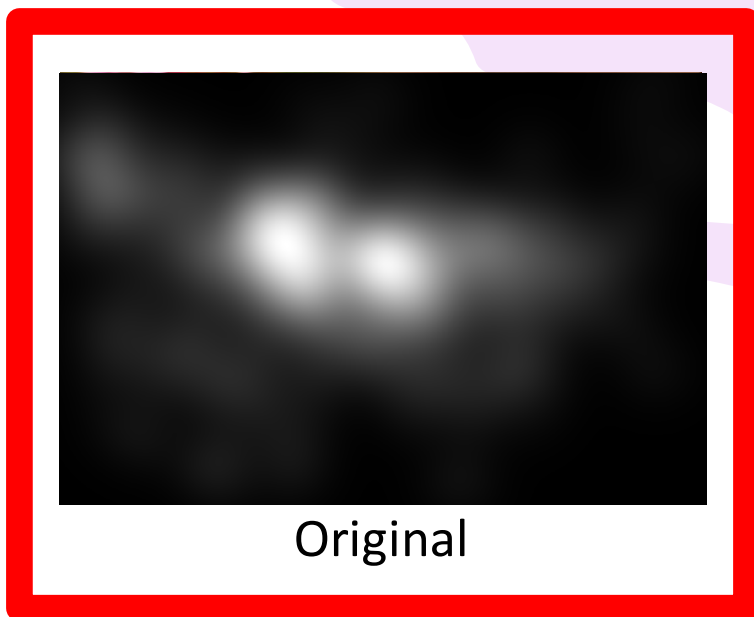


Ref: W. Zhang, A. Borji, Z. Wang, P. L. Callet, and H. Liu, "The Application of Visual Saliency Models in Objective Image Quality Assessment: A Statistical Evaluation," IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 6, pp. 1266–1278, Jun 2016.

W. Zhang, Y. Tian, X. Zha, and H. Liu, "Benchmark in state-of-the-art visual saliency models for image quality assessment," in IEEE International Conference on Acoustics, Speech and Signal Processing, May 2016, vol. 2016-May, pp. 1090–1094.

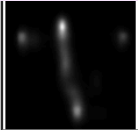
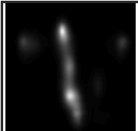
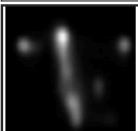
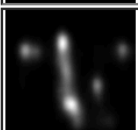
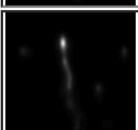
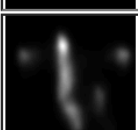
W. Zhang and H. Liu, "Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications," IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2424–2437, May 2017.

# SIQ288-examples

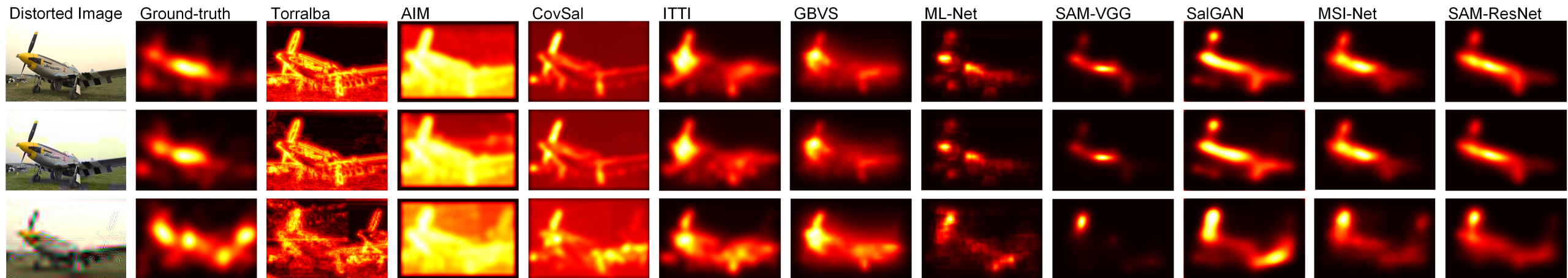


Ref: W. Zhang, A. Borji, Z. Wang, P. L. Callet, and H. Liu, "The Application of Visual Saliency Models in Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 1266–1278, Jun 2018.  
W. Zhang, Y. Tian, X. Zha, and H. Liu, "Benchmark in state-of-the-art visual saliency models for image quality assessment," *IEEE Transactions on Image Processing*, May 2016, vol. 2016-May, pp. 1090–1094.  
W. Zhang and H. Liu, "Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research," *IEEE Transactions on Image Processing*, pp. 2424–2437, May 2017.

# Visual Saliency Models

Saliency Attentive Model (SAM-ResNet)	Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, Rita Cucchiara. <a href="#">Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model [IEEE TIP 2018]</a>	python	0.87	0.68	2.15	0.78	0.70	0.78	2.34	1.27	first tested: 10/30/2016 last tested: 03/03/2017 maps from authors	
Saliency Attentive Model (SAM-VGG)	Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, Rita Cucchiara. <a href="#">Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model [IEEE TIP 2018]</a>	python	0.87	0.67	2.14	0.78	0.71	0.77	2.30	1.13	first tested: 10/30/2016 last tested: 03/03/2017 maps from authors	
DenseSal	Taiki Oyama, Takao Yamanaka. <a href="#">Influence of Image Classification Accuracy on Saliency Map Estimation [CAAI Transactions on Intelligence Technology, 2018]</a>		0.87	0.67	1.99	0.81	0.72	0.79	2.25	0.48	first tested: 14/06/2017 last tested: 14/06/2017 maps from authors	
DPNSal	Taiki Oyama, Takao Yamanaka. <a href="#">Influence of Image Classification Accuracy on Saliency Map Estimation [CAAI Transactions on Intelligence Technology, 2018]</a>		0.87	0.69	2.05	0.80	0.74	0.82	2.41	0.91	first tested: 19/04/2018 last tested: 19/04/2018 maps from authors	
CEDNS	Chunhuan Lin, Fei Qi, Guangming Shi, Hao Li		0.87	0.64	2.23	0.74	0.69	0.75	2.43	0.63	first tested: 24/06/2018 last tested: 24/06/2018 maps from authors	
MSI-Net	Alexander Kroner, Mario Senden, Kurt Driessens, Rainer Goebel. <a href="#">Contextual Encoder-Decoder Network for Visual Saliency Prediction [arXiv 2019]</a>	Python	0.87	0.68	1.99	0.82	0.72	0.79	2.27	0.66	first tested: 06/12/2018 last tested: 06/12/2018 maps from authors	

- Attention-based on information maximisation (AIM)

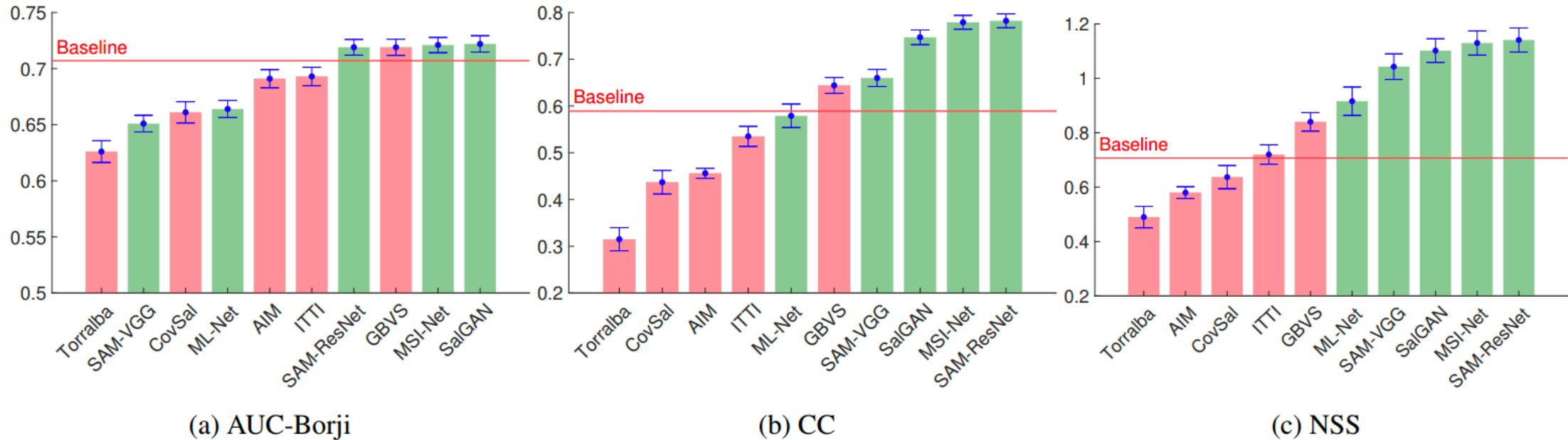


# Evaluation metrics

**3 commonly used** saliency metrics from different aspects

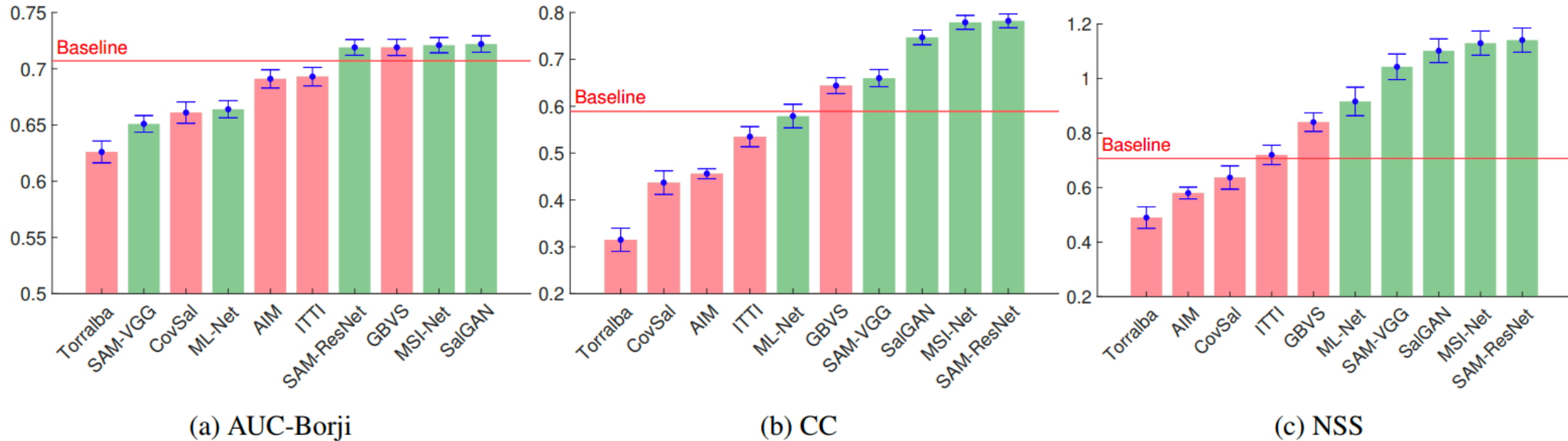
- **Value-based** metrics NSS  
(Normalised Scanpath Saliency metric)
- **Location-based** metrics AUC-Borji  
(Area under the curve-AUC Borji)
- **Distribution-based** metrics CC  
(Pearson Linear Correlation Coefficient)

# Overall results



- **Baseline:** indicates the performance of a 'base' saliency model that is computed by stretching a symmetric Gaussian to fit the aspect ratio of a given image, under the assumption that the **centre of the image is most salient**.
- **Only one** traditional model, **GBVS**, performing above the baseline, others are dominated by the deep learning models

# Overall results

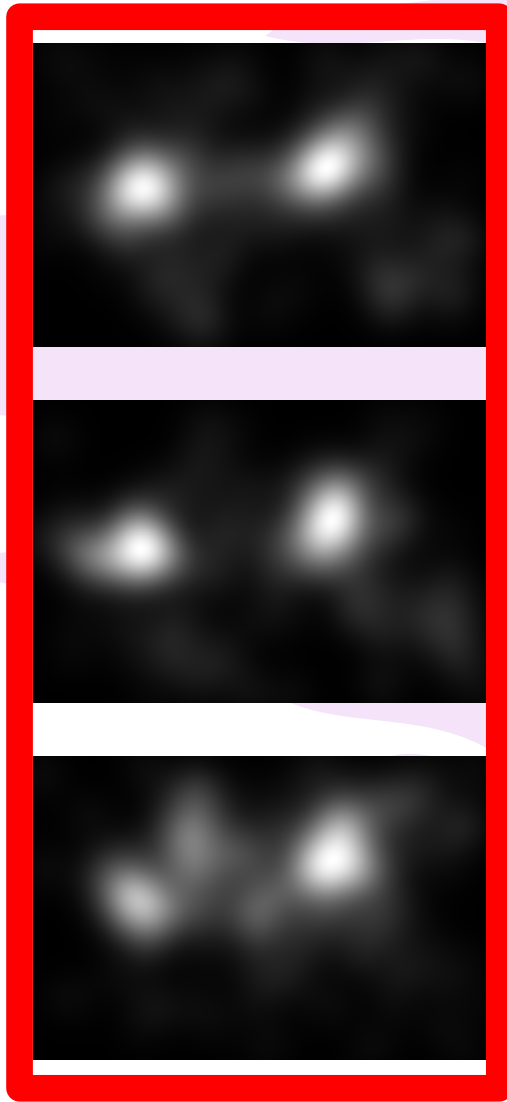


- SAM-ResNet, MSI-Net and SalGAN are **consistently ranked higher than other models**.
- In order to verify whether the difference in performance between traditional and deep learning models is statistically significant, hypothesis testing is performed on the AUC-Borji, NSS, and CC data. The results show that in all cases, **the deep learning models are statistically significantly better than traditional models**.

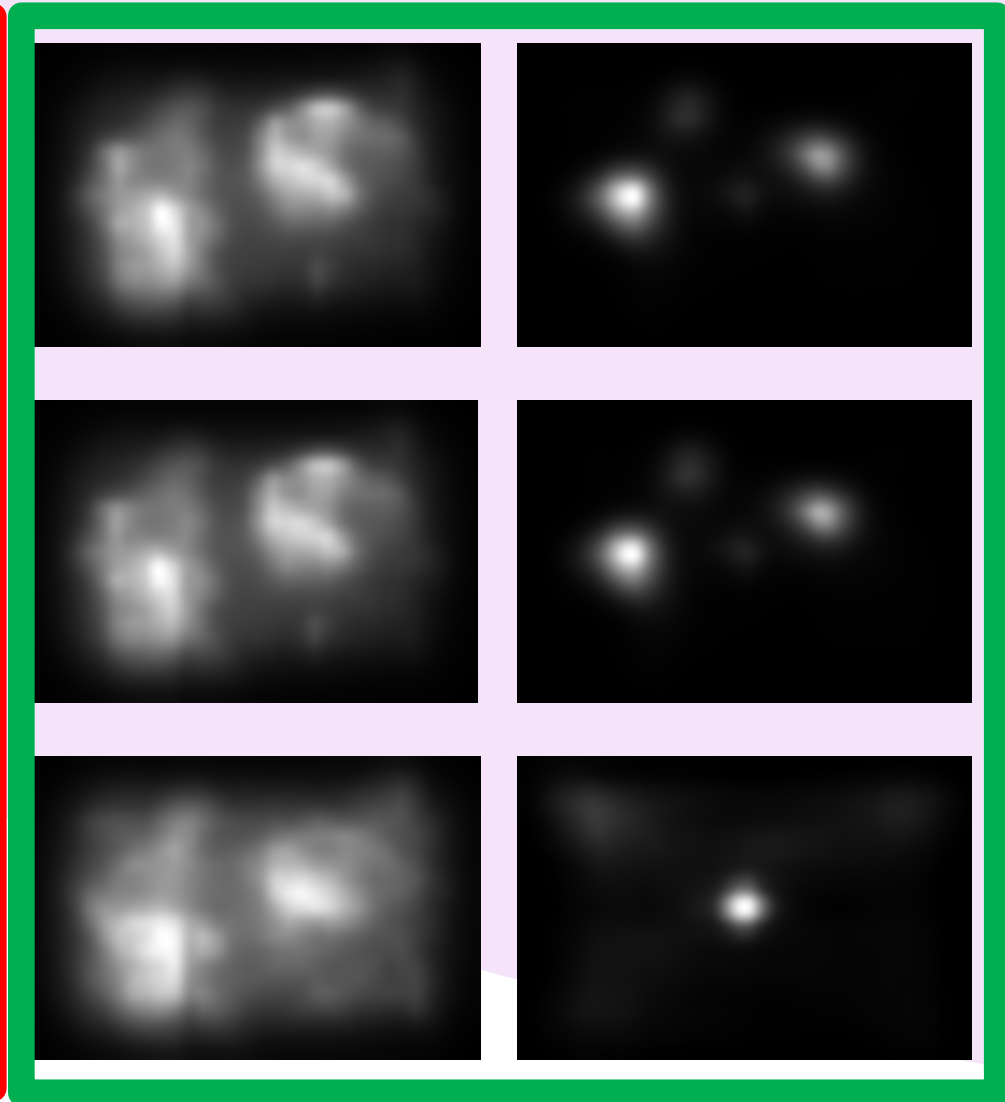




Parrorts  
Distortion type:  
white noise



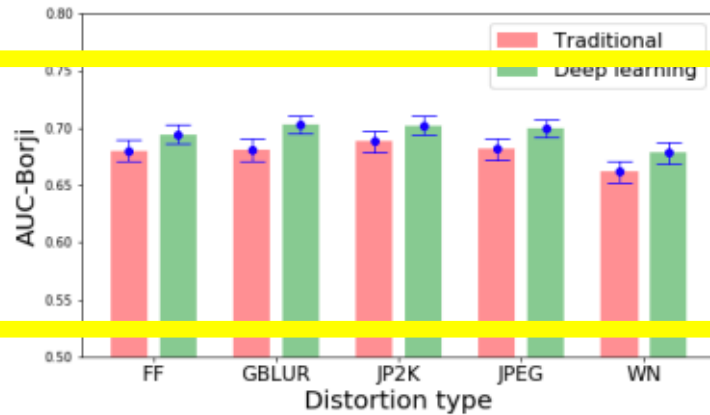
**Ground truth**  
saliency map



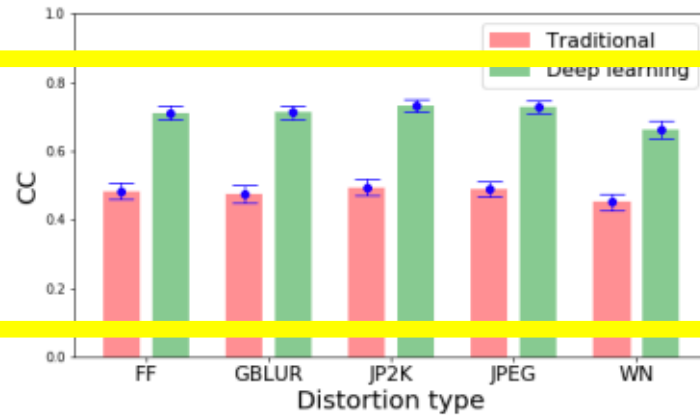
**Traditional**  
prediction:  
GBVS

**Deep learning**  
prediction:  
SAM-Vgg

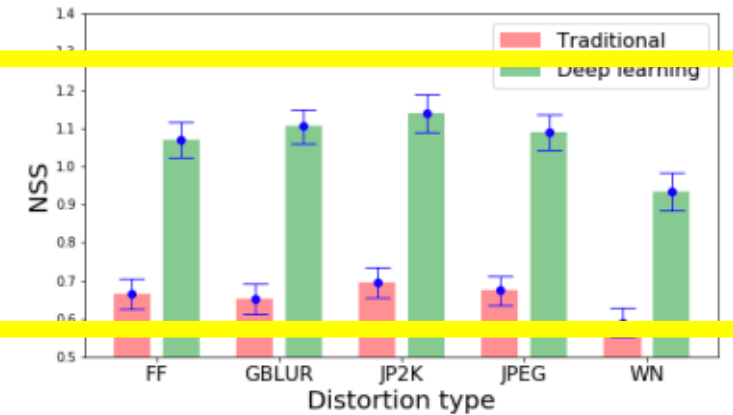
# Results: Impact of distortion types



(a) AUC-Borji



(b) CC



(c) NSS

- It clearly indicates that deep learning models consistently **outperform** traditional models **for all distortion types**.
- an independent samples t-test for each comparison, and the results show that for each of the 15 cases (i.e. 5 types  $\times$  3 evaluation metrics) the **difference is statistically significant** ( $p < 0.05$ ).

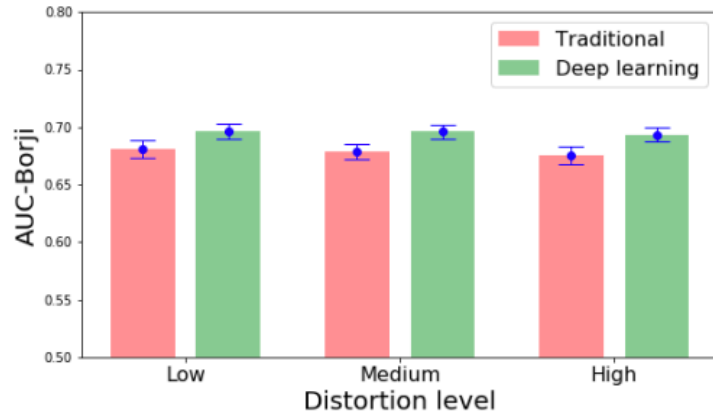
# Results: Impact of distortion types

**Table 1:** Performance of individual saliency models measured by AUC-Borji, CC and NSS, for different distortion types.

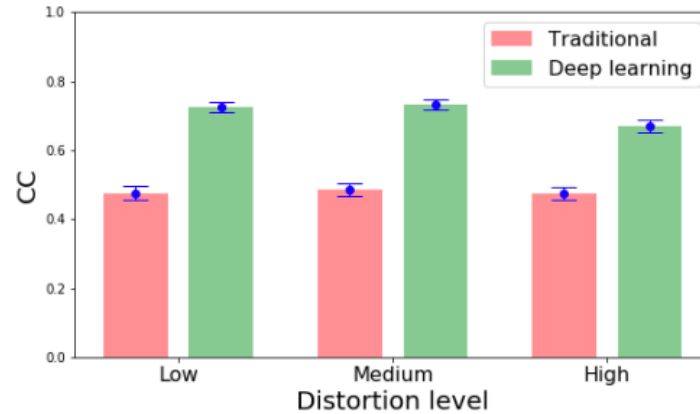
	FF			GBLUR			JP2K			JPEG			WN		
	NSS ↑	CC ↑	AUC ↑	NSS ↑	CC ↑	AUC ↑	NSS ↑	CC ↑	AUC ↑	NSS ↑	CC ↑	AUC ↑	NSS ↑	CC ↑	AUC ↑
SAM-VGG	1.06	0.67	0.66	1.09	0.67	0.66	1.12	0.68	0.66	1.06	0.68	0.65	0.89	0.60	0.62
SAM-ResNet	<b>1.15</b>	<b>0.79</b>	0.72	<b>1.18</b>	<b>0.78</b>	0.72	<b>1.21</b>	0.80	<b>0.73</b>	<b>1.15</b>	<b>0.79</b>	<b>0.72</b>	1.02	<b>0.75</b>	<b>0.71</b>
ML-Net	0.93	0.59	0.67	0.97	0.60	0.67	0.99	0.61	0.67	0.97	0.62	0.68	0.71	0.47	0.63
SalGAN	1.08	0.74	<b>0.73</b>	1.11	0.74	<b>0.73</b>	1.16	0.75	0.72	1.12	0.76	<b>0.72</b>	<b>1.04</b>	<b>0.75</b>	<b>0.71</b>
MSI-Net	1.13	0.77	<b>0.73</b>	1.17	<b>0.78</b>	<b>0.73</b>	<b>1.21</b>	<b>0.81</b>	<b>0.73</b>	1.14	<b>0.79</b>	<b>0.72</b>	1.01	0.74	<b>0.71</b>
Torralba	0.49	0.32	0.62	0.45	0.29	0.62	0.55	0.35	0.64	0.52	0.33	0.63	0.44	0.29	0.61
ITTI	0.75	0.55	0.71	0.77	0.56	0.71	0.75	0.55	0.70	0.72	0.54	0.69	0.62	0.48	0.67
GBVS	0.86	0.65	<b>0.73</b>	0.87	0.65	<b>0.73</b>	0.85	0.64	0.72	0.84	0.64	<b>0.72</b>	0.81	0.65	<b>0.71</b>
CovSal	0.63	0.43	0.66	0.58	0.42	0.66	0.71	0.47	0.67	0.71	0.47	0.67	0.55	0.40	0.64
AIM	0.58	0.46	0.70	0.59	0.46	0.70	0.60	0.46	0.70	0.59	0.47	0.70	0.53	0.44	0.67

- From the results that for the deep learning models, their performance on the **WN** distortion is **relatively lower than other distortion types**.

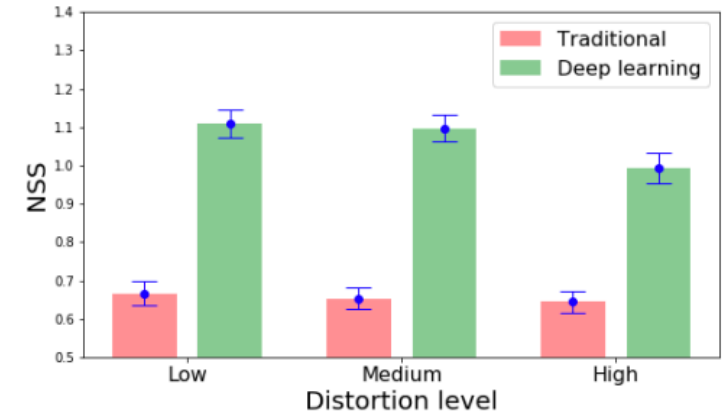
# Results: Impact of distortion level



(a) AUC-Borji



(b) CC



(c) NSS

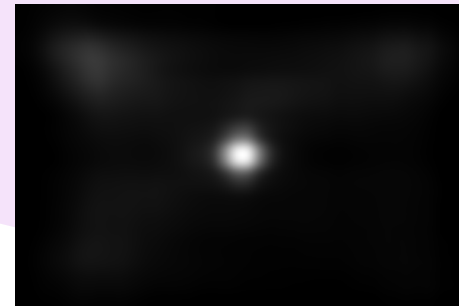
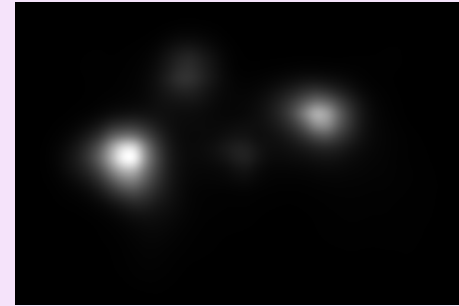
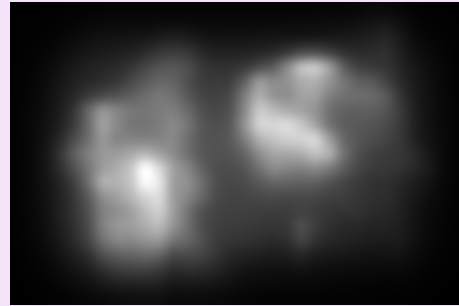
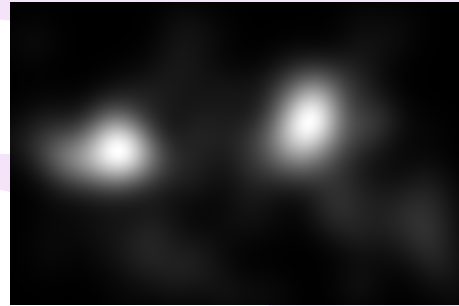
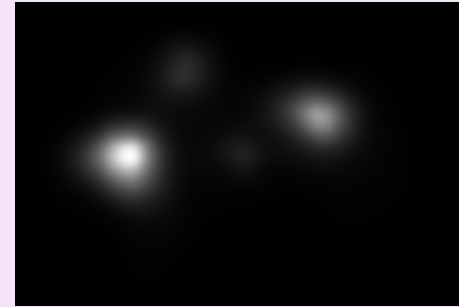
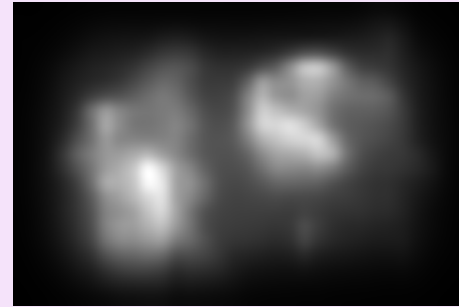
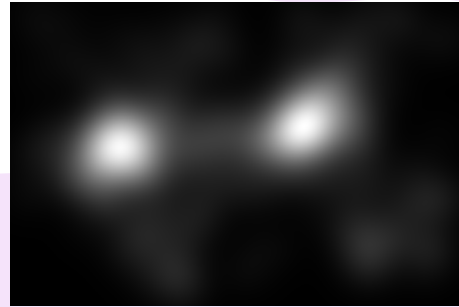
- All table shows the average performance of traditional vs deep learning models for different levels of distortion.
- The results of hypothesis testing show that for each of the 9 cases (i.e. 3 levels × 3 evaluation metrics) the performance of the deep learning models is **statistically significantly better** than the traditional models.

# Results: Impact of distortion levels

**Table 2:** Performance of individual saliency models measured by AUC-Borji, CC and NSS, for different distortion levels. (i.e., Low, Medium and High distortion)

	AUC-Borji			CC			NSS		
	Low↑	Medium↑	High↑	Low↑	Medium↑	High↑	Low↑	Medium↑	High↑
SAM-VGG	0.66	0.65	0.64	0.67	0.68	0.63	1.08	1.07	0.98
SAM-ResNet	<b>0.72</b>	<b>0.72</b>	0.72	<b>0.79</b>	<b>0.80</b>	<b>0.75</b>	<b>1.17</b>	<b>1.16</b>	<b>1.09</b>
ML-Net	0.67	0.66	0.66	0.62	0.61	0.50	1.00	0.97	0.78
SalGAN	<b>0.72</b>	<b>0.72</b>	0.72	0.75	0.77	0.72	1.13	1.13	1.06
MSI-Net	<b>0.72</b>	<b>0.72</b>	0.72	<b>0.79</b>	<b>0.80</b>	0.74	<b>1.17</b>	<b>1.16</b>	1.06
Torralba	0.63	0.63	0.62	0.33	0.33	0.29	0.53	0.49	0.45
ITTI	0.69	0.69	0.70	0.52	0.54	0.55	0.71	0.72	0.74
GBVS	0.71	<b>0.72</b>	<b>0.73</b>	0.62	0.65	0.67	0.82	0.83	0.88
CovSal	0.67	0.66	0.65	0.45	0.44	0.41	0.69	0.64	0.58
AIM	0.70	0.69	0.69	0.46	0.47	0.44	0.59	0.58	0.57

- **Deep learning** models are promising, but they show relatively **low** performance in handling **highly distorted images** compared to images of low and medium levels of distortion.



Parrorts  
Distortion type:  
white noise

**Ground truth**  
saliency map[

**Traditional**  
prediction:  
GBVS

**Deep learning**  
prediction:  
SAM-Vgg

# Conclusion

In this paper, we conducted statistical analyses to evaluate the performance of deep learning versus traditional models for saliency prediction of distorted images. Obviously, deep learning models significantly outperform traditional models.

In addition, we found that model performance tends to depend on the type and level of image distortion. Future work could focus on improving deep learning models for challenging cases, e.g. white noise distortion or highly distorted images.

# Reflection on learning

1. Why deep learning model has a better prediction performance? and even it is better but not good enough.
2. Some traditional models still have a comparable predicting performance. Why? Will traditional models' aspects could lead a better learning ?
3. Can we build a network to predicting images with different image quality? and how can we generalise a network in predicting images will all kinds of qualities?



*THANKS*

# Q&A