Interpretation, and Applications of Gradients

Part 4: Gradients as Expectancy-Mismatch

Objectives

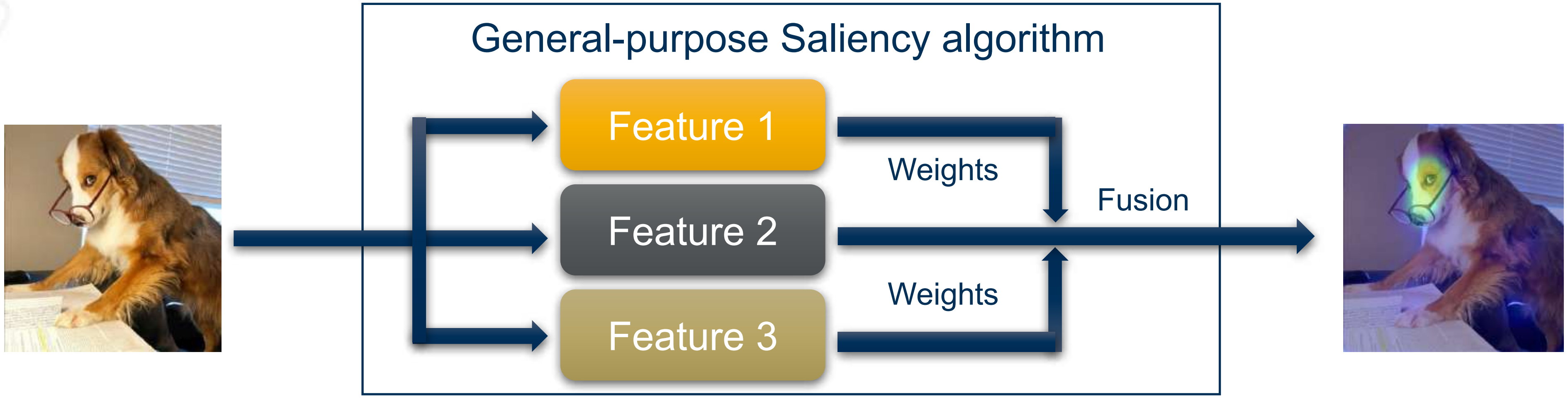
Objectives in Part IV

Case Study: Expectancy-Mismatch

- Interpret gradients as Expectancy-Mismatch
  - Define expectancy-mismatch utilizing saliency
  - Demonstrate counterfactual manifolds as expectancy-mismatch
- Human Visual Saliency
- Image Quality Assessment

Saliency

Saliency in Literature



Bottom-Up Saliency : Innovation is in designing features and fusion

Top-Down Saliency : Innovation is in designing weights

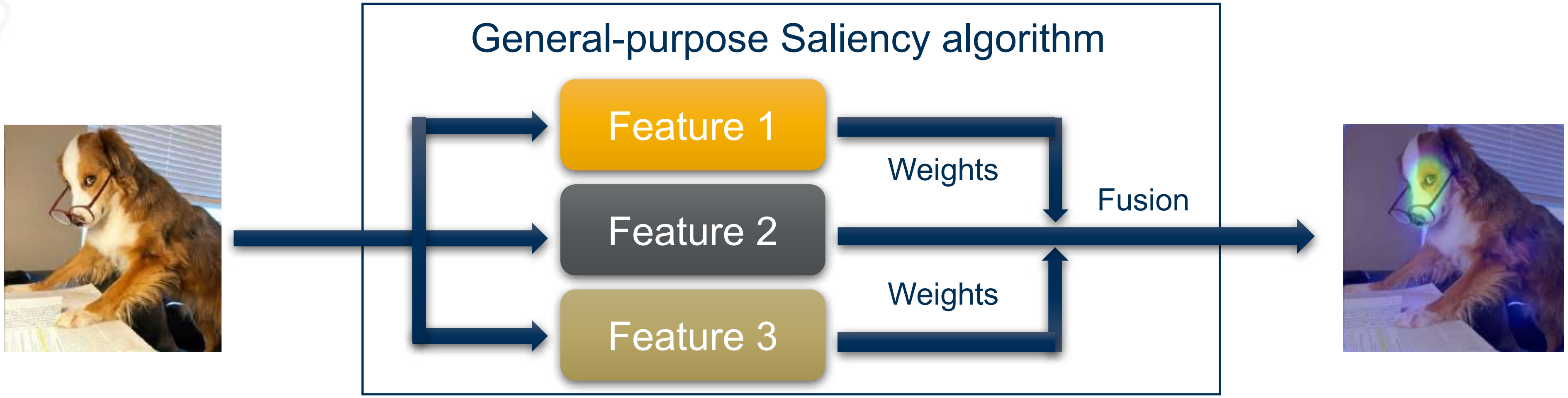
Color, Intensity, Orientation [1]

Faces, text, object detectors [1]

[1] Judd, Tilke, Frédo Durand, and Antonio Torralba. "A benchmark of computational models of saliency to predict human fixations." (2012).

Saliency

Our Goal: Introduce Implicit Saliency in Neural Networks



Bottom-Up Saliency : Innovation is in designing features and fusion

Top-Down Saliency : Innovation is in designing weights

Color, Intensity, Orientation [1]

Faces, text, object detectors [1]

Features that are new and unexpected (novel) in a scene are salient

[1] Judd, Tilke, Frédo Durand, and Antonio Torralba. "A benchmark of computational models of saliency to predict human fixations." (2012).

Expectancy-Mismatch

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

At Inference, construct local contrastive manifolds

Change in Network Parameters: Expectancy-Mismatch when presented with novel data!

We demonstrate on two applications:

1. Human Visual Saliency
2. Image Quality Assessment

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

At Inference, construct local contrastive manifolds

Change in Network Parameters: Expectancy-Mismatch when presented with novel data!

We demonstrate on two applications:

1. Human Visual Saliency
2. Image Quality Assessment

frontiers in Neuroscience

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

Mohit Prabhushankar, PhD
Postdoc

Ghassan AlRegib, PhD
Professor

SCAN ME

30th ICIP 2023 Kuala Lumpur

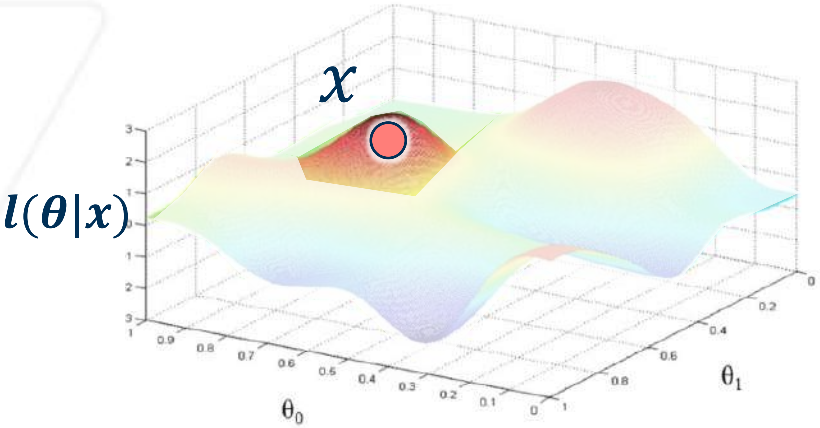
OLIVES @GeorgiaTech

Georgia Tech

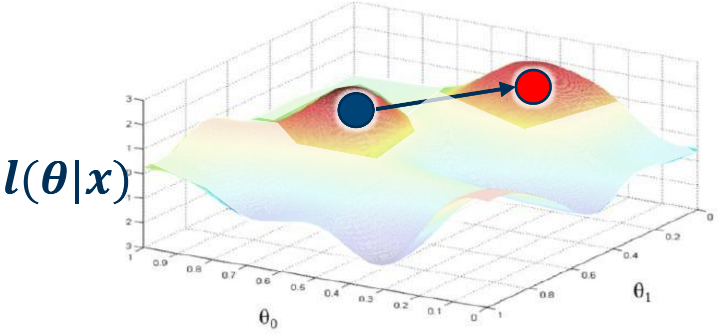
Expectancy-Mismatch

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

Similar to introspective learning!

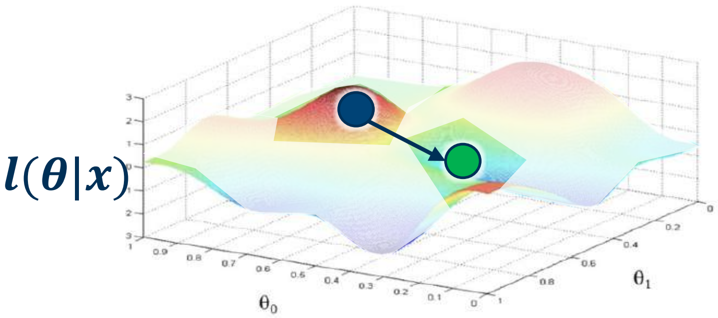


Contrast class 1



.
.
.

Contrast class N

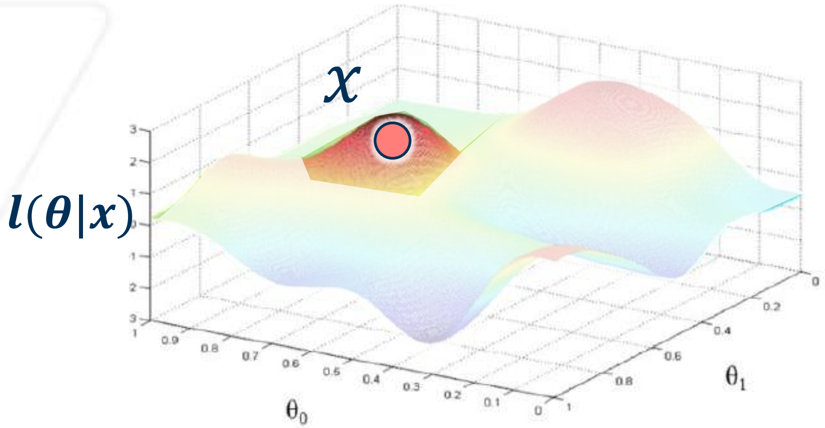


Mean of projected gradients is the expectancy!

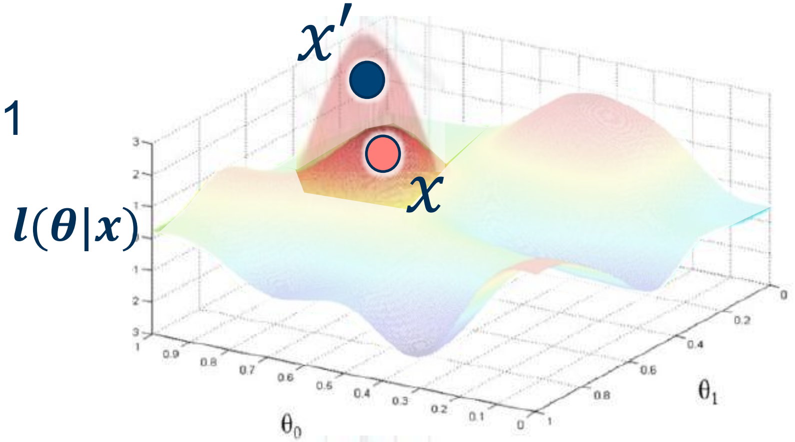
Expectancy-Mismatch

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

Similar to introspective learning!

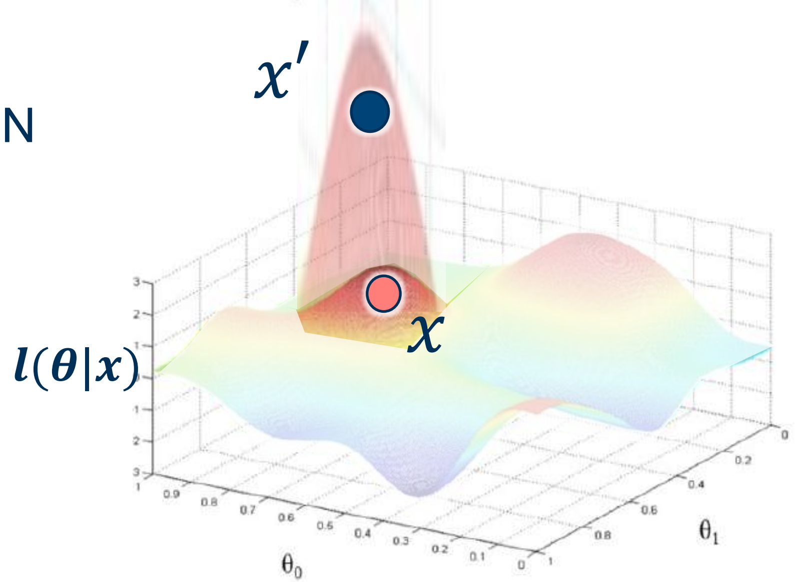


Contrast class 1



.
.
.

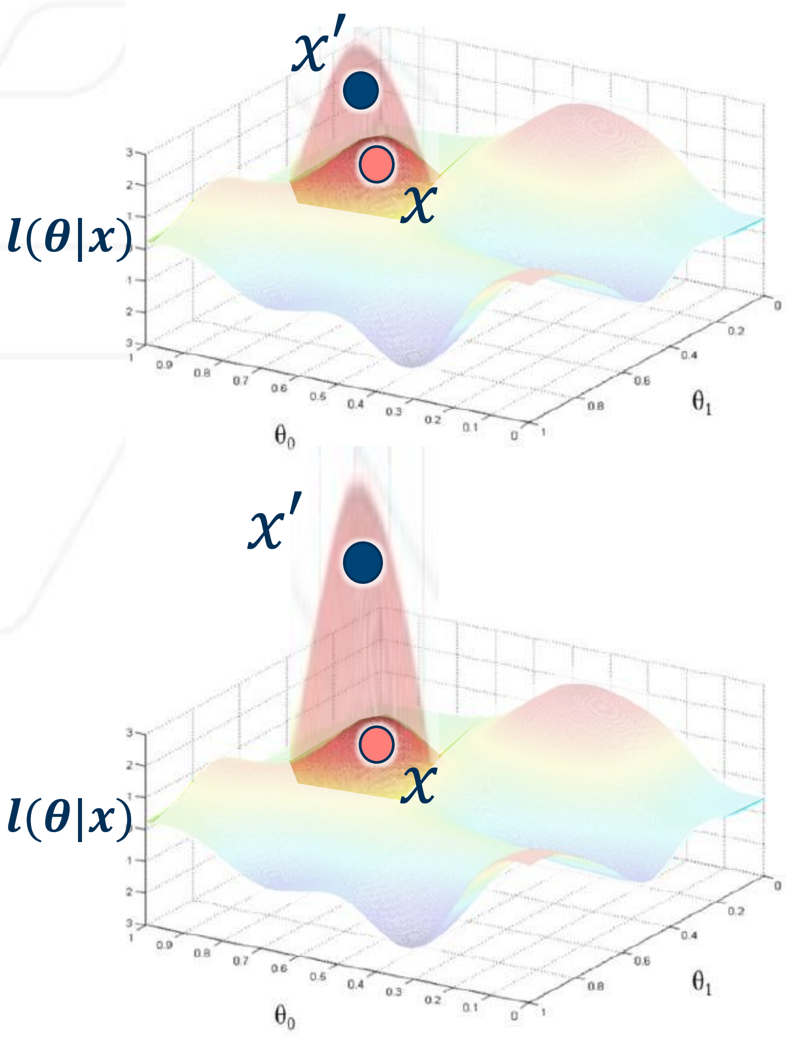
Contrast class N



Variance of gradients is the mismatch!

Expectancy-Mismatch

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

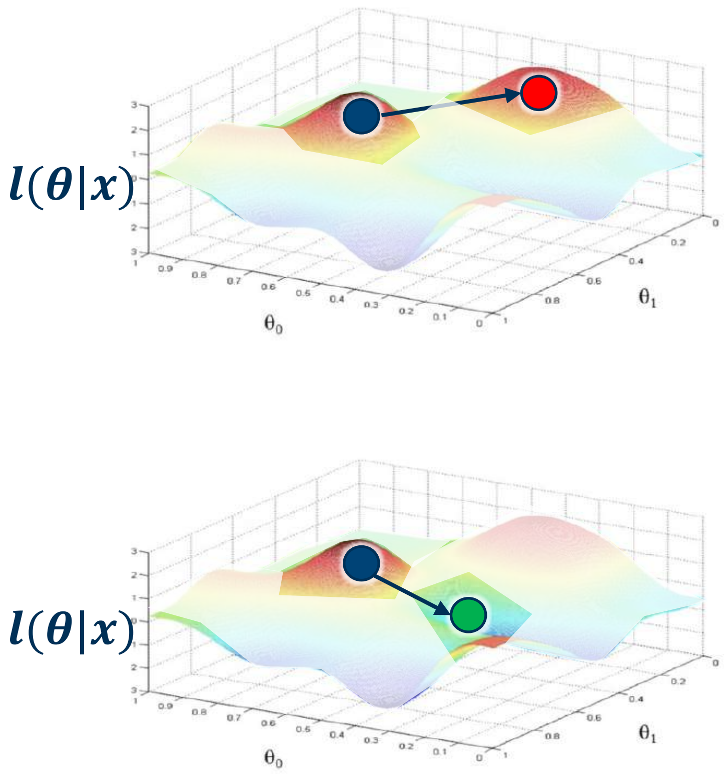


Variance of gradients is the mismatch!

Point-wise Multiplication

Mean of projected gradients is the expectancy!

Saliency Map

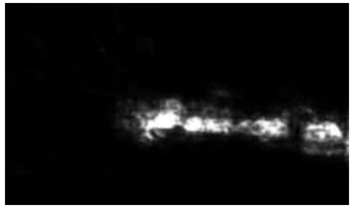


Expectancy-Mismatch

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

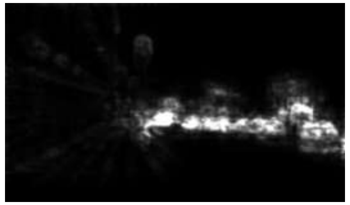
Similar to introspective learning!

Wrong class 1

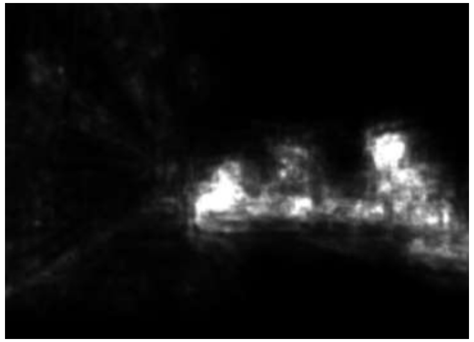


.
.
.

Wrong class N



Saliency Map

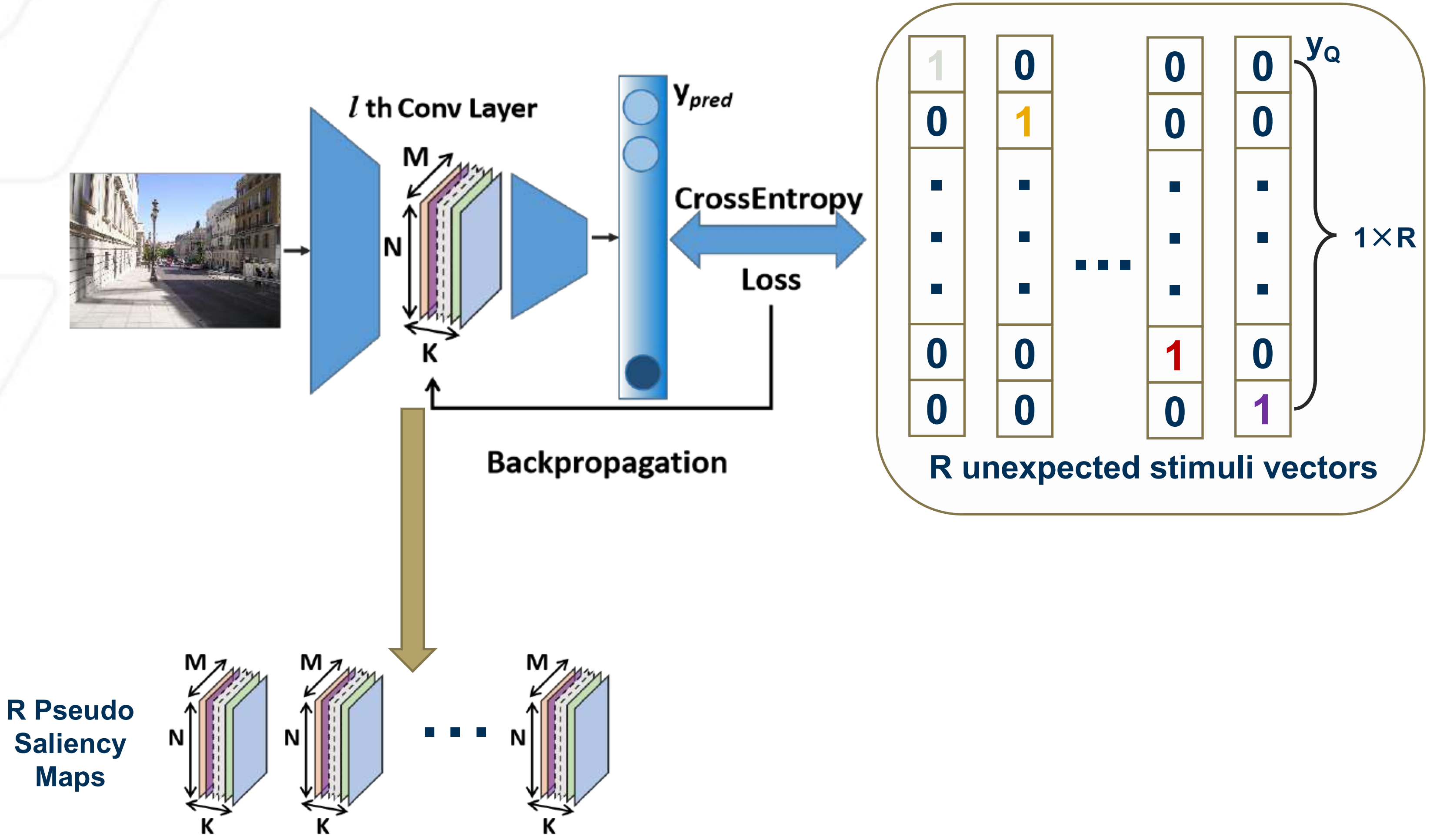


Gradients in the k^{th} layer: Pseudo-saliency maps

Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

cSaliency

Deriving Gradient-based Implicit Saliency



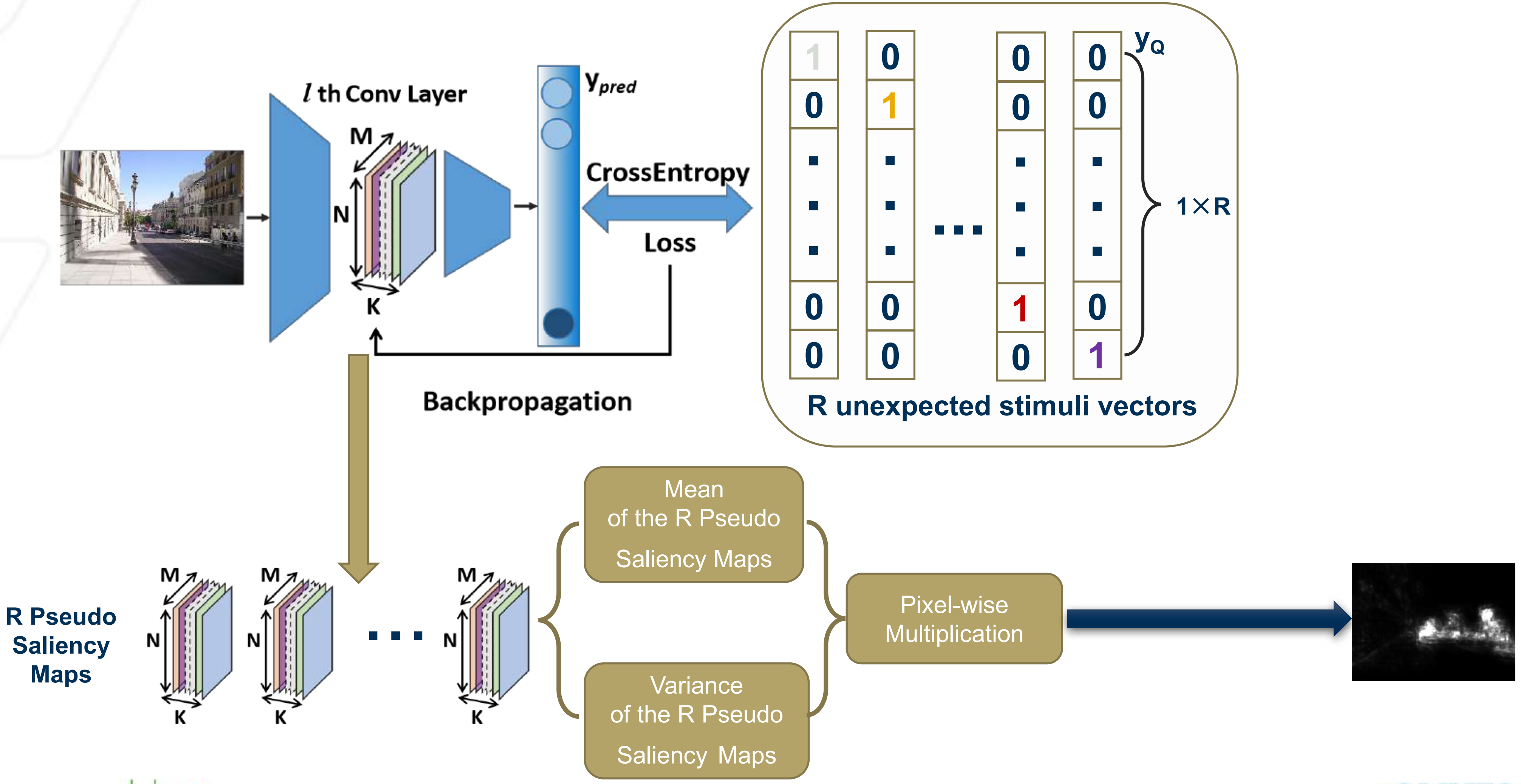
Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

Implicit Saliency

Deriving Gradient-based Implicit Saliency

SCAN ME

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

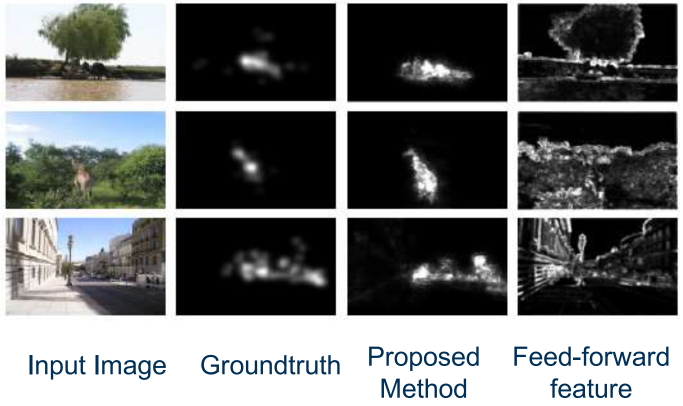


Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

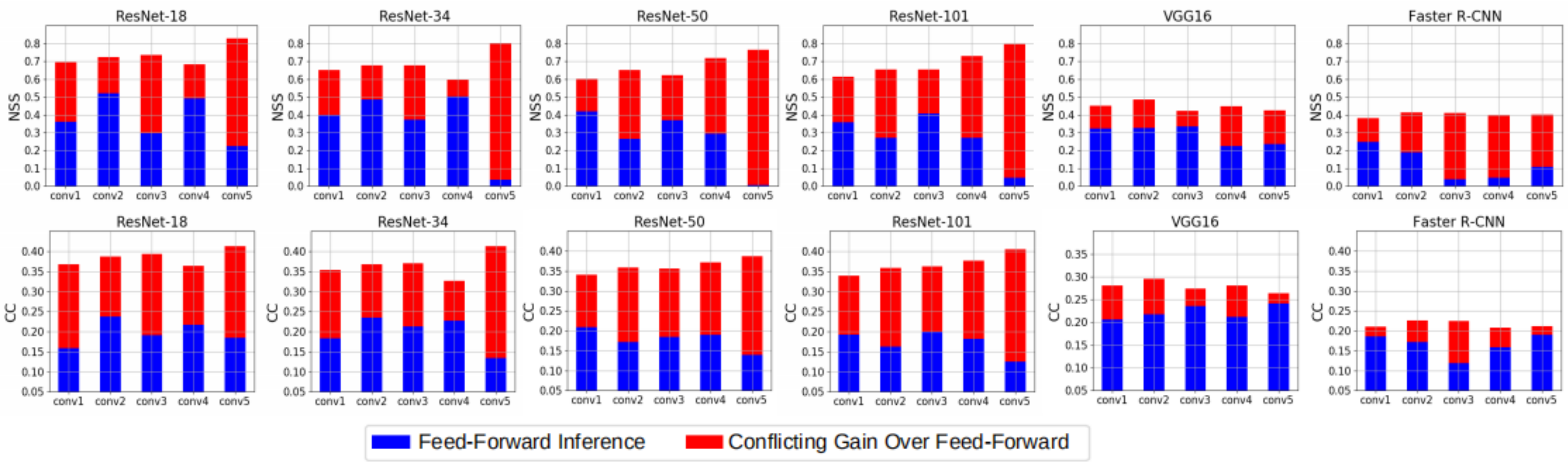
Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

Contrastive saliency is correlated with attention more than its Feed-Forward counterpart



- Feed-forward expectation features:
  - Edges and textures
  - Without specific localization
- Proposed expectation-mismatch Saliency:
  - Localized saliency maps
  - Highly correlated with ground truth



Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

Implicit Saliency

Experiments

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

Contrastive Saliency outperforms explanation methods like GradCAM and Guided Backprop

	NSS				CC			
Networks	ResNet-18	ResNet-34	ResNet-50	ResNet-101	ResNet-18	ResNet-34	ResNet-50	ResNet-101
GradCam	0.7657	0.7545	0.7203	0.7335	0.3496	0.3396	0.3190	0.3210
GBP	0.3862	0.4191	0.3898	0.3415	0.2474	0.2453	0.2443	0.2233
Contrastive Saliency	**0.8274**	**0.8018**	**0.7659**	**0.7981**	**0.4132**	**0.4112**	**0.3868**	**0.4051**

Input Image

GradCam

Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

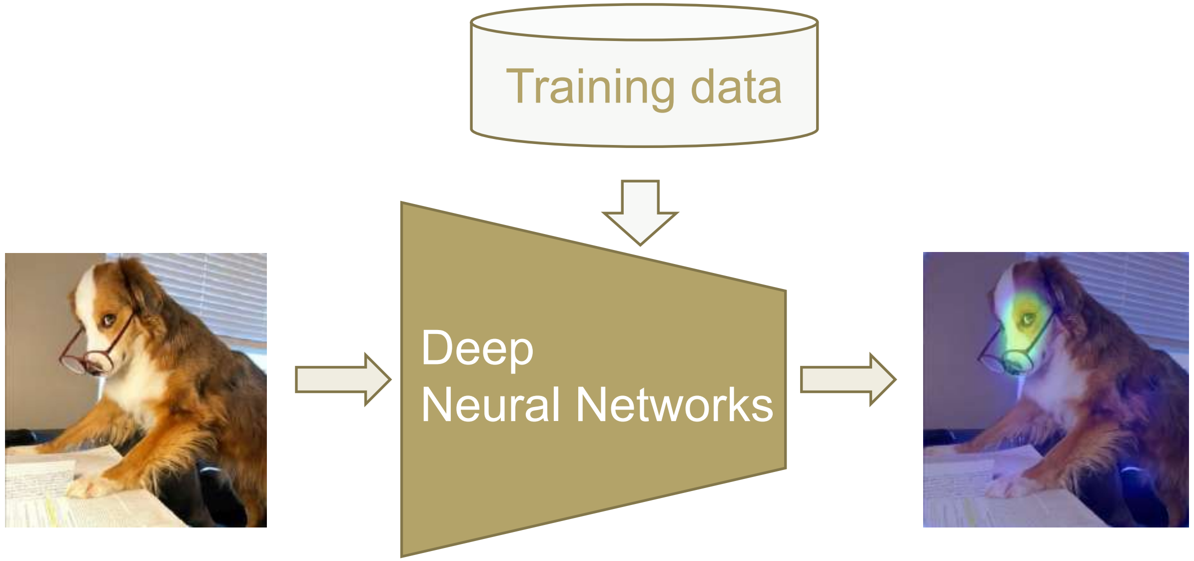
Implicit Saliency

Experiments

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

Compare performance of unsupervised Contrastive Saliency model against existing saliency models



Existing Learning based methods

Contrastive Saliency is unsupervised!

Saliency Models	Training data
SalGan	SALICON
ML-Net	SALICON
DeepGazeII	SALICON
ShallowDeep	SALICON/iSUN

Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

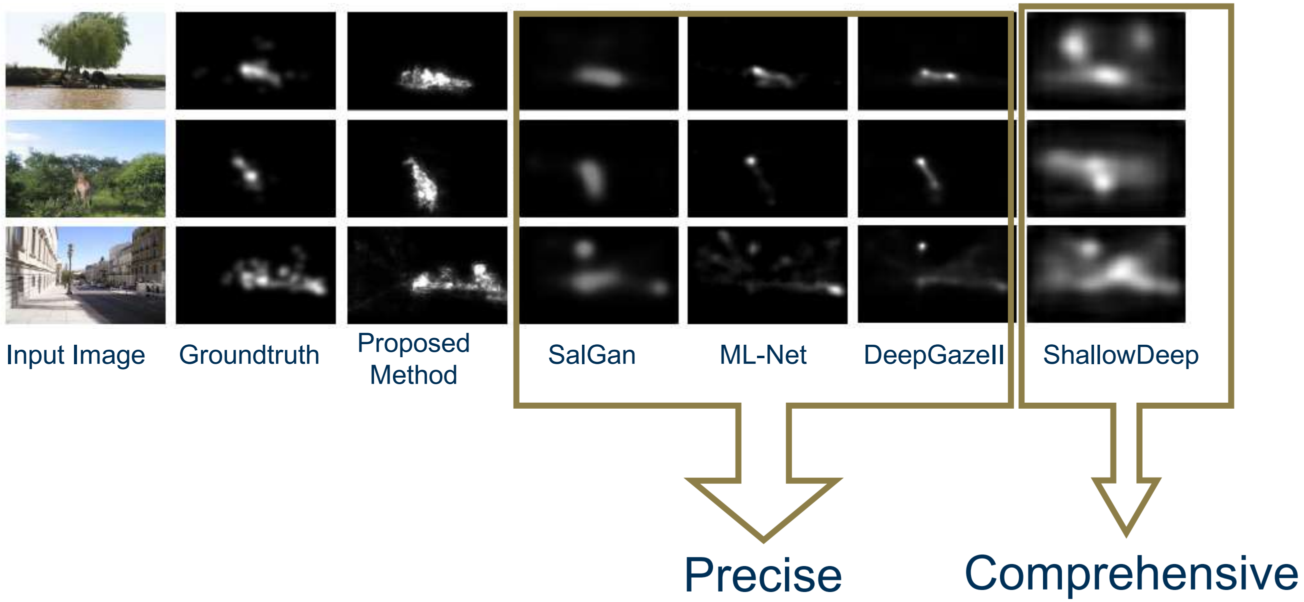
Implicit Saliency

Experiments

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

Compare performance of unsupervised Contrastive Saliency model against existing saliency models



NSS					CC				
Sal Gan	Deep GazeII	ML Net	Shallow Deep	**Contrastive Saliency**	Sal Gan	Deep GazeII	ML Net	Shallow Deep	**Contrastive Saliency**
0.8977	0.6214	0.5431	**0.9306**	0.7981	**0.6280**	0.5927	0.4481	0.5120	0.4051

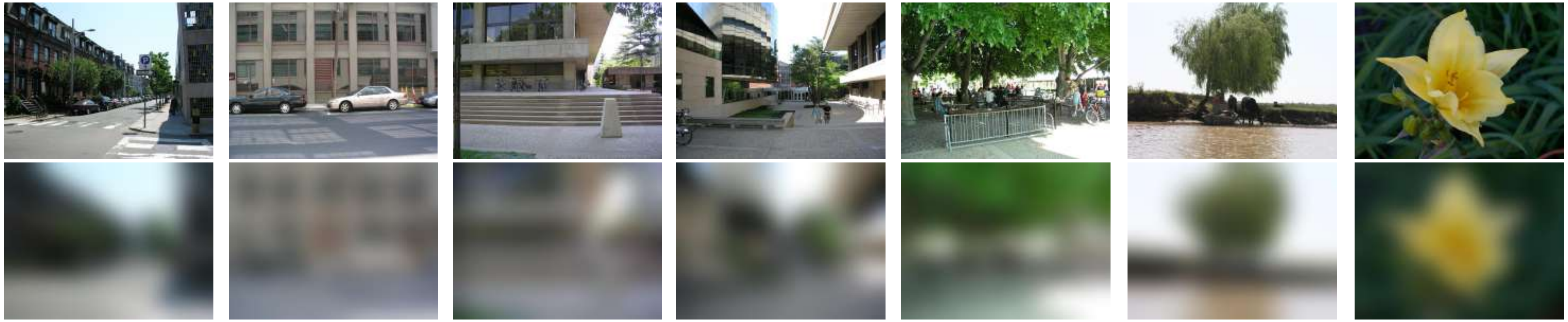
Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

Implicit Saliency
Robustness Analysis

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

Contrastive Saliency drops the least performance with noise added



	NSS					CC				
Gaussian Blur	Sal Gan	Deep GazeII	ML Net	Shallow Deep	**Contrastive Saliency**	Sal Gan	Deep GazeII	ML Net	Shallow Deep	**Contrastive Saliency**
$r = 0$	0.8977	0.6214	0.5431	**0.9306**	0.7981	**0.6280**	0.5927	0.4481	0.5120	0.4051
$r = 50$	↓ 0.2239	↓ 0.3436	↓ 0.2484	↓ 0.2025	↓ **0.1793**	↓ 0.2731	↓ 0.3954	↓ 0.2940	↓ 0.1840	↓ **0.1432**

Sun, Yutong, Mohit Prabhushankar, and Ghassan AlRegib. "Implicit saliency in deep neural networks." *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020.

Expectancy-Mismatch

Our Goal: Introduce Expectancy-Mismatch in Neural Networks

At Inference, construct local contrastive manifolds

Change in Network Parameters: Expectancy-Mismatch when presented with novel data!

We demonstrate on two applications:

1. Human Visual Saliency
2. Image Quality Assessment

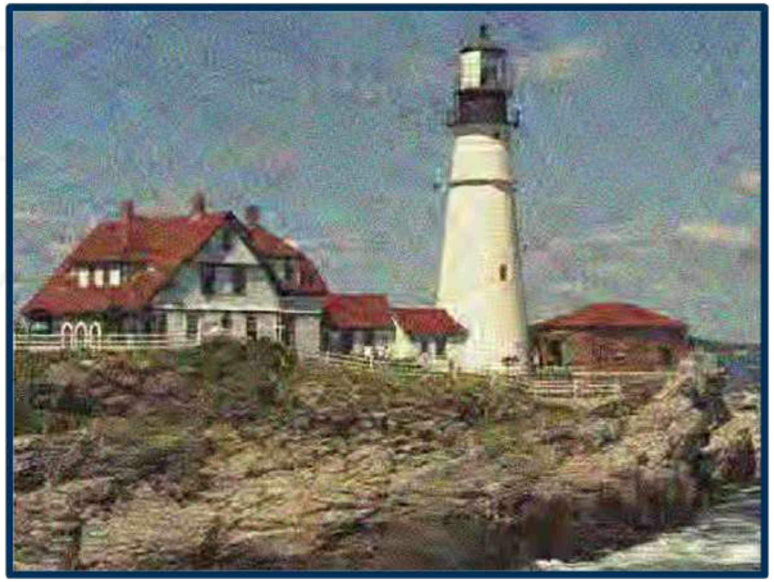
Image Quality Assessment

What is IQA?

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

IQA is the objective Assessment of Subjective Quality



Lighthouse image with level 5 lossy compression from TID 2013 dataset

Image Quality Assessment Algorithm : DIQaM [1]

The given image is somewhat OK quality

Score : 0.58

Bad Quality

Good Quality

[1] Bosse S, Maniry D, Müller K R, et al. Deep neural networks for no-reference and full-reference image quality assessment. IEEE Transactions on Image Processing, 2018, 27(1): 206-219.

Image Quality Assessment

Expectancy-Mismatch in Dataset Construction

Expectancy-Mismatch arises during Dataset Construction

- Subjects are shown a reference image in a controlled setting
- Based on the reference image, they are asked to pick one of the images on the top that differs least from the reference image
- Reference image sets the expectancy
- The task of subjectively picking the least mis-matched image is IQA

This requires **Fine-grained** Analysis!

[1] Ponomarenko, Nikolay, et al. "Image database TID2013: Peculiarities, results and perspectives." *Signal processing: Image communication* 30 (2015): 57-77

Image Quality Assessment

Expectancy-Mismatch in Dataset Construction

Expectancy-Mismatch arises during Dataset Construction

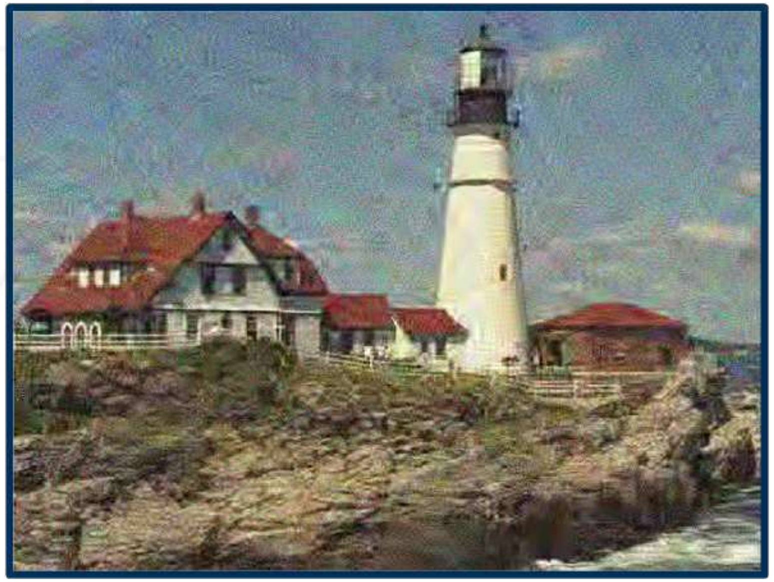
This requires **Fine-grained** Analysis on the part of the subjects!

Our Goal: To determine if a trained IQA detector understands the fine-grained nature of expectancy-mismatch in quality

[1] Ponomarenko, Nikolay, et al. "Image database TID2013: Peculiarities, results and perspectives." *Signal processing: Image communication* 30 (2015): 57-77

GradCAM in IQA

GradCAM explanation for Why 0.58?



Lighthouse image with level 5 lossy compression from TID 2013 dataset

The given image is somewhat OK quality

DIQaM : 0.58

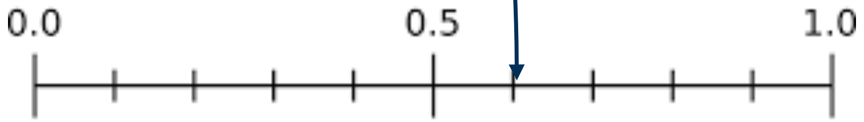
Grad-CAM

Why 0.58?

Add heatmap
Explain blue
Yellow, red, green

Bad Quality

Good Quality



Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Image Quality Assessment

GradCAM in IQA

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

GradCAM explanation may not be useful for fine-grained analysis

Grad-CAM explanation tells us that the quality score was decided based on all parts of the image and specifically based on the base of the lighthouse

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

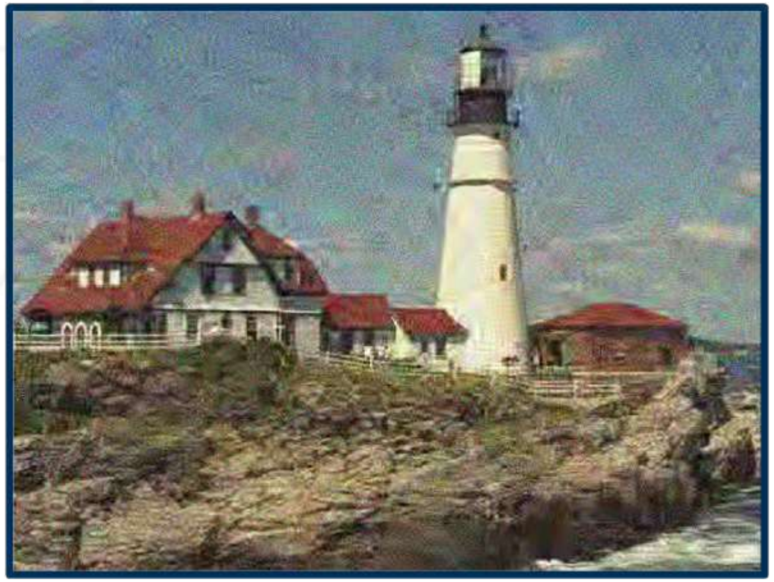
Image Quality Assessment

ContrastCAM in IQA

SCAN ME

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

All the distortions in the foreground prevent a quality score of 1



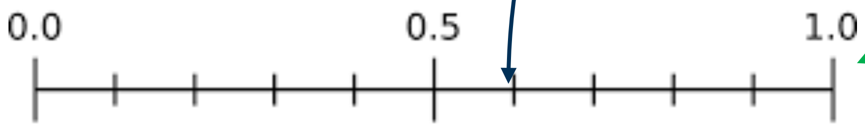
Lighthouse image with level 5 lossy compression from TID 2013 dataset

DIQaM : 0.58

Contrastive explanation

Why 0.58, rather than 1?

Bad Quality



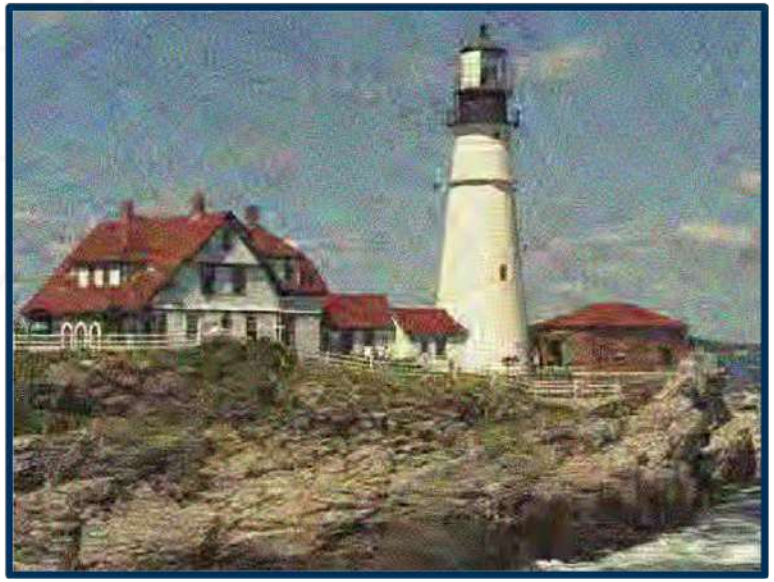
Good Quality

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

The distortions on the lighthouse and houses prevent a higher score of 0.75



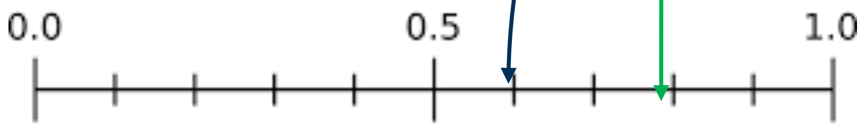
Lighthouse image with level 5 lossy compression from TID 2013 dataset

DIQaM : 0.58

Contrastive explanation

Why 0.58, rather than 0.75?

Bad Quality



Good Quality

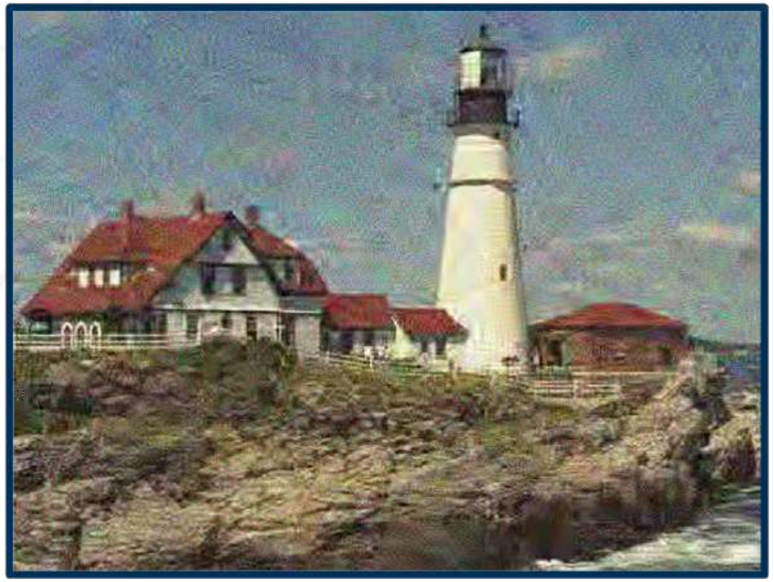
Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

ContrastCAM in IQA

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

The quality of the lighthouse and sky is better than a score of 0.5



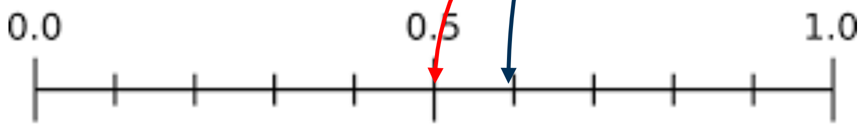
Lighthouse image with level 5 lossy compression from TID 2013 dataset

DIQaM : 0.58

Contrastive explanation

Why 0.58, rather than 0.5?

Bad Quality



Good Quality

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

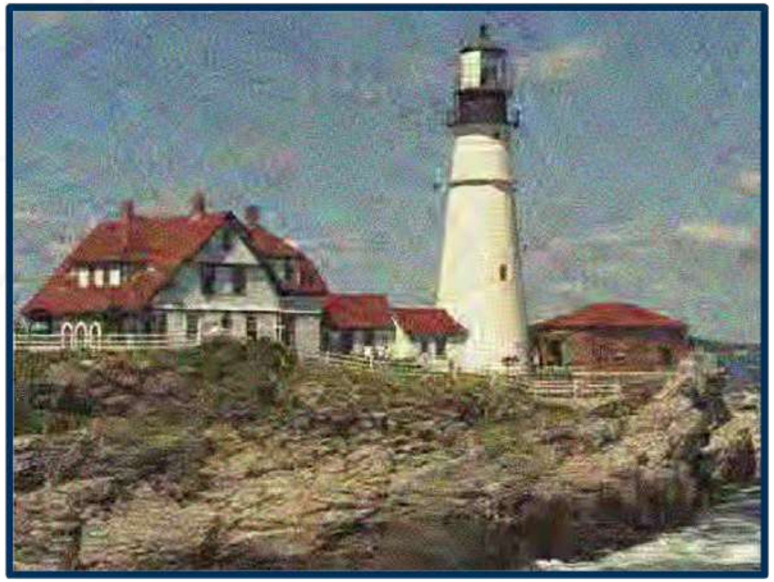
Image Quality Assessment

ContrastCAM in IQA

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

The sky, lighthouse, and cliff merit a quality higher than 0.25



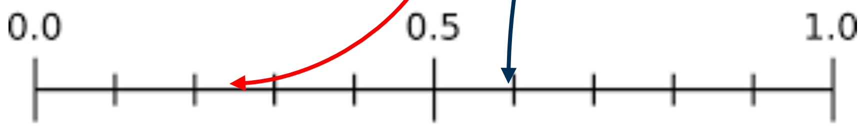
DIQaM : 0.58

Contrastive explanation

Why 0.58, rather than 0.25?

Lighthouse image with level 5 lossy compression from TID 2013 dataset

Bad Quality



Good Quality

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

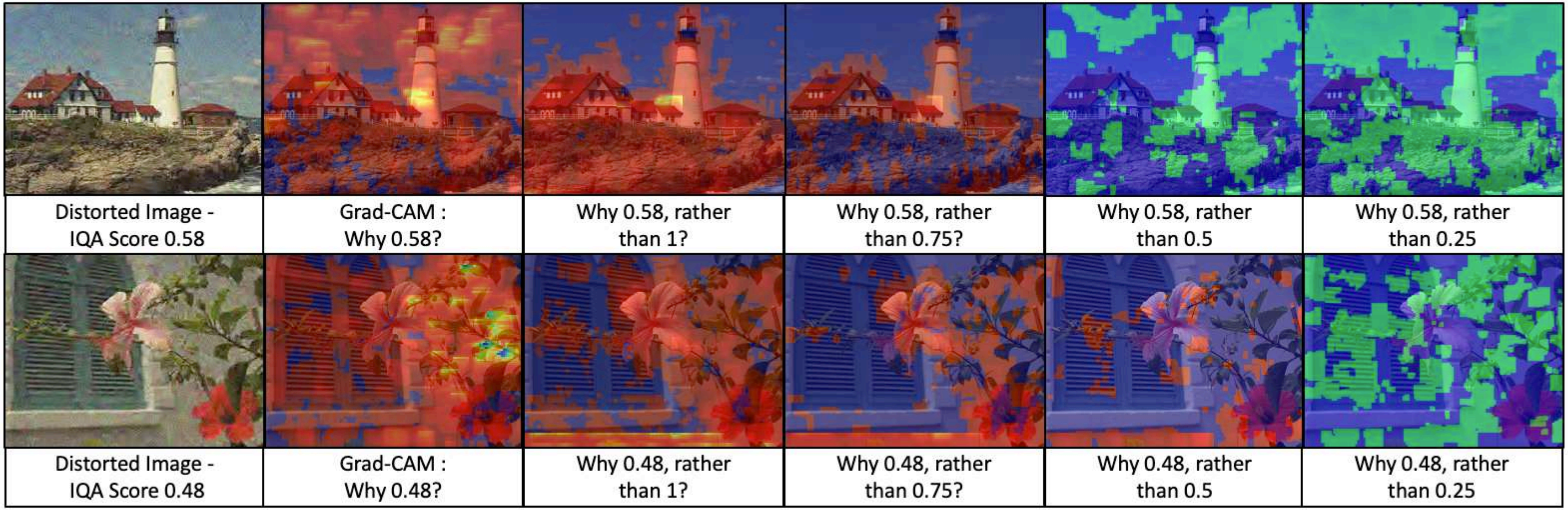
Image Quality Assessment

ContrastCAM in IQA

Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks

SCAN ME

Contrastive IQA elicits the fine-grained decisions made by the network



Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Objectives

Takeaways from Part IV

- Part I: Gradients in Neural Networks
- Part 2: Gradients as Information
- Part 3: Gradients as Uncertainty
- **Part 4: Gradients as Expectancy-Mismatch**
    - Presented a case study of utilizing both the contrastive manifolds and manifold traversal perspectives
    - Human Visual Saliency is a by-product of expectancy-mismatch
    - Neural networks that have never explicitly learned human salient regions have implicitly been trained to use them in tasks
    - Using Contrastive explanations in IQA provides a fine-grained analysis of neural network's perception of quality
- Part 5: Conclusion and Future Directions