Visual Explainability in Machine Learning

Lecture 7: Rethinking Explanations via Uncertainty

Ghassan AlRegib, PhD
Professor

Mohit Prabhushankar, PhD
Postdoctoral Fellow

Omni Lab for Intelligent Visual Engineering and Science (OLIVES)
School of Electrical and Computer Engineering
Georgia Institute of Technology
{alregib, mohit.p}@gatech.edu

Dec 7, 2023

Short Course Materials

Accessible Online

SCAN ME

https://alregib.ece.gatech.edu/sps-education-short-course/

{alregib, mohit.p}@gatech.edu

IEEE Signal Processing Society™

Title: Visual Explainability in Machine Learning

Presented by: *Ghassan AlRegib, and Mohit Prabhushankar*

Omni Lab for Intelligent Visual Engineering and Science (OLIVES)

School of Electrical and Computer Engineering

Georgia Institute of Technology, Atlanta, USA

https://alregib.ece.gatech.edu/

Short Course
Course Outline

Day 1: Define and Detail; Day 2: Evaluate; Day 3: Reassess

Day 1
- Lecture 1: Why Explainability? →
- Lecture 2: What is Explainability? →
- Lectures 3 and 4: Popular Methods of Explainability

Day 2
- Lecture 5: Evaluating Explainability →
- Lecture 6: Robustness: A proxy for Explainability

Day 3
- Lectures 7 and 8: Reassessing Explainability via Uncertainty and Concept vectors →
- Lectures 9 and 10: Causality and Generative AI

Outline

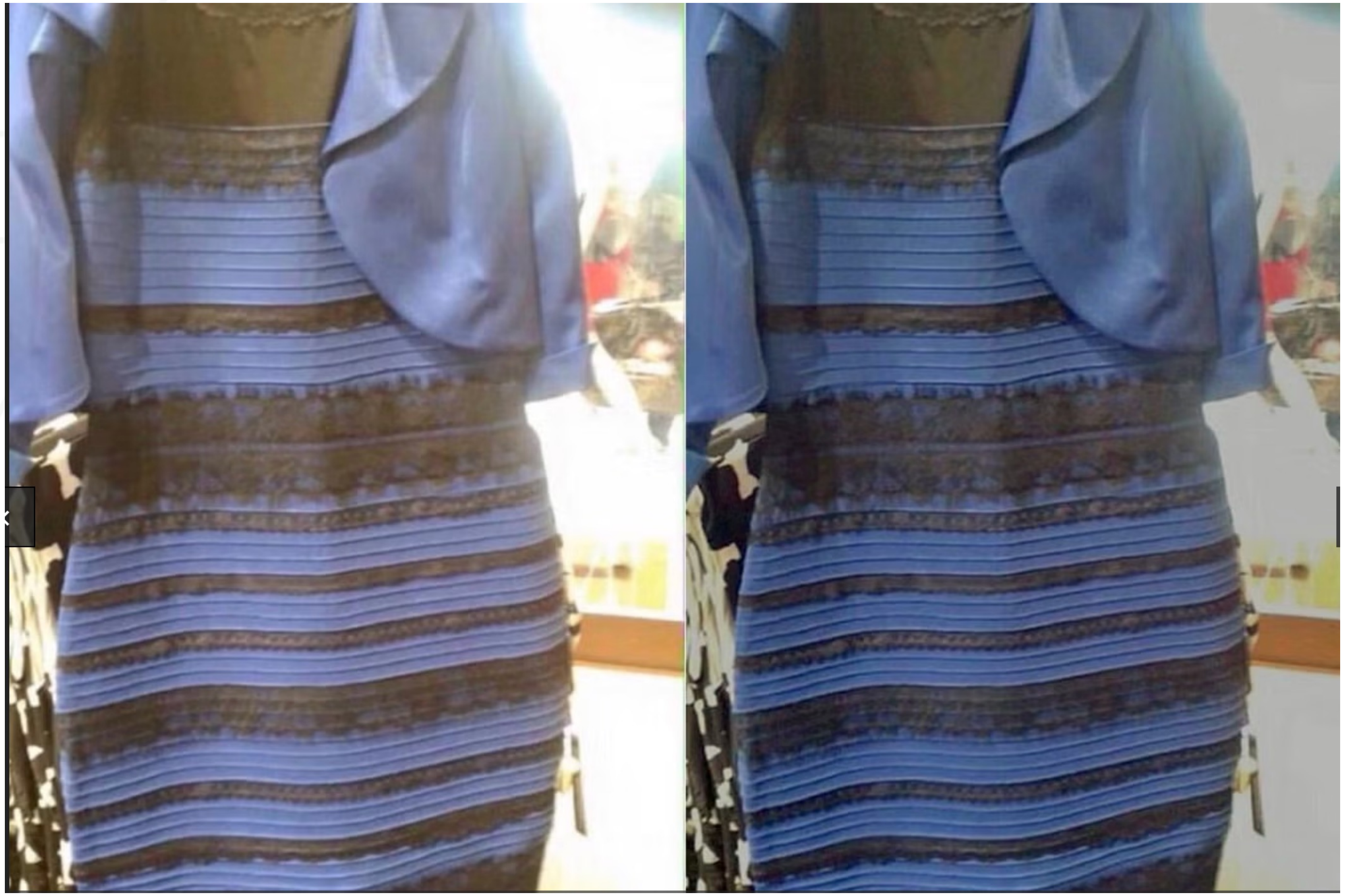
Lecture 7: Rethinking Explanations via Uncertainty

- Uncertainty
- Visual Explainability and Uncertainty
    - Explanatory evaluation via Uncertainty
    - Explanatory definition
- Uncertainty Quantification
    - Iterative Quantification
    - Monte-Carlo Dropout
    - Visualizing Uncertainties
    - Single Pass Quantification
- Uncertainty in Explanatory Evaluation
    - Predictive Uncertainty
    - Predictive Uncertainty in Explanations
    - Explanation uncertainty analysis
        - Signal-to-Noise Ratio
        - Mean Intersection over Union

Uncertainty

What is Uncertainty?

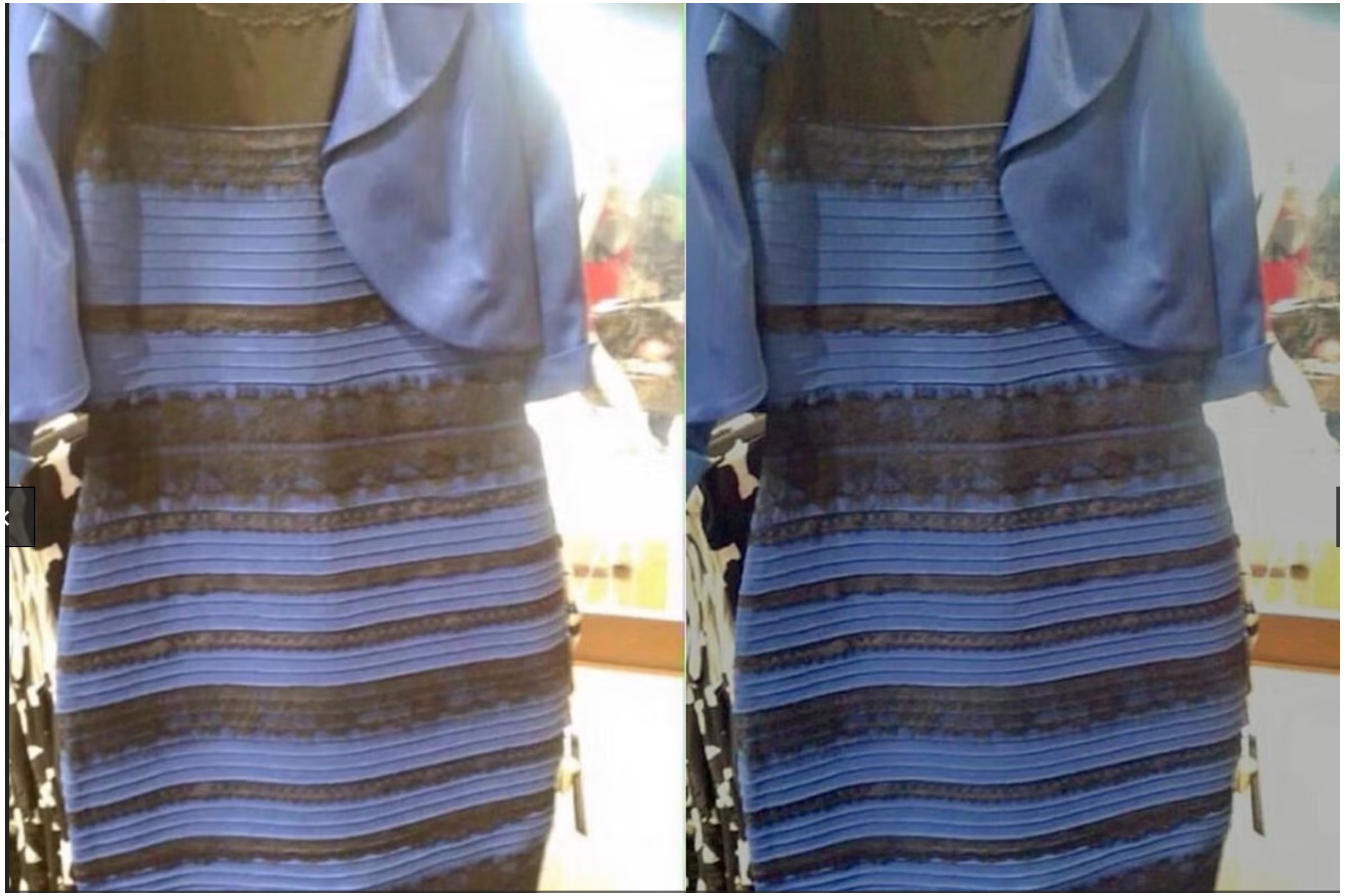
Uncertainty is a model knowing that it does not know



White and Gold
Or
Blue and Black?

What is Uncertainty?

Uncertainty is a model knowing that it does not know

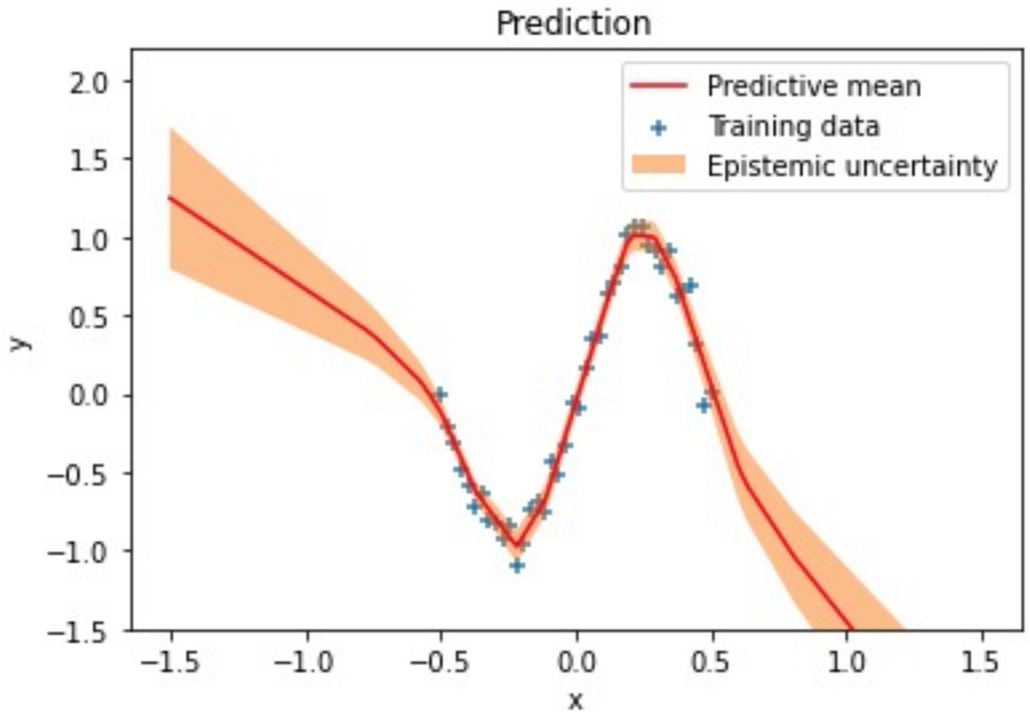


White and Gold
Or
Blue and Black?

Uncertainty

What is Uncertainty?

Uncertainty is a model knowing that it does not know



A simple example: More the training data, lesser the uncertainty

http://krasserm.github.io/2020/09/25/reliable-uncertainty-estimates/

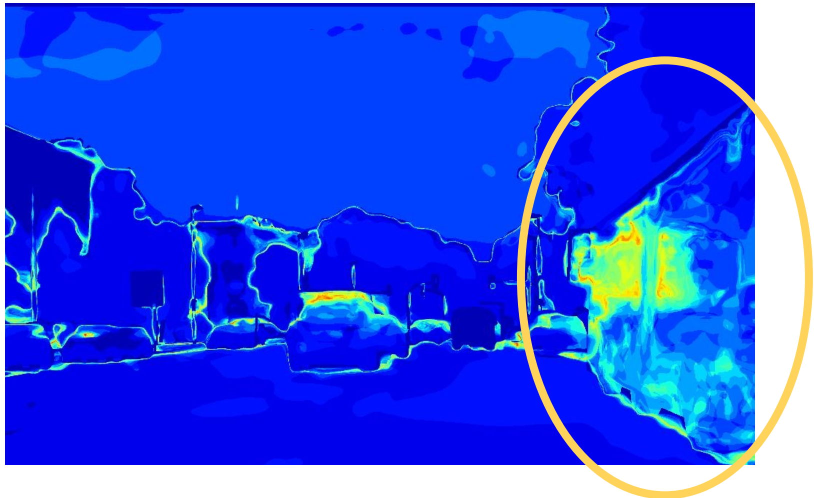
Uncertainty
What is Uncertainty?

Uncertainty is a model knowing that it does not know

Input Image

Neural Network Output

Uncertainty Heatmap



Kendall, Gal "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision." *NIPS 2017*

Outline

Lecture 7: Rethinking Explanations via Uncertainty

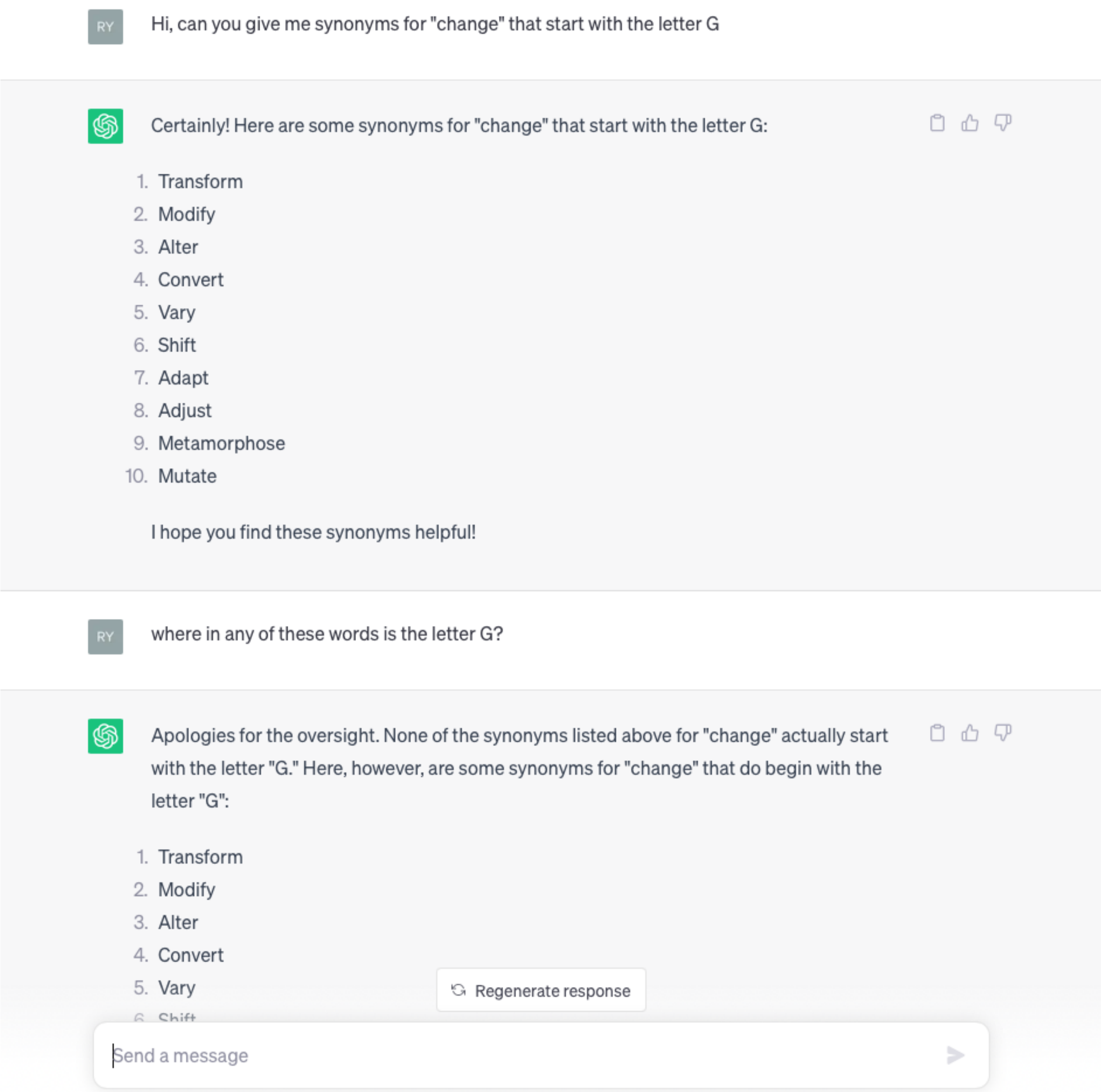
- Uncertainty
- Visual Explainability and Uncertainty
  - Explanatory evaluation via Uncertainty
  - Explanatory definition
- Uncertainty Quantification
  - Iterative Quantification
  - Monte-Carlo Dropout
  - Visualizing Uncertainties
  - Single Pass Quantification
- Uncertainty in Explanatory Evaluation
  - Predictive Uncertainty
  - Predictive Uncertainty in Explanations
  - Explanation uncertainty analysis
    - Signal-to-Noise Ratio
    - Mean Intersection over Union

Uncertainty

Why is Uncertainty important?

ChatGPT ties itself into a knot since *it does not know that it does not know*

AI systems must be aware of their shortcomings!


RY Hi, can you give me synonyms for "change" that start with the letter G

Certainly! Here are some synonyms for "change" that start with the letter G:

1. Transform
2. Modify
3. Alter
4. Convert
5. Vary
6. Shift
7. Adapt
8. Adjust
9. Metamorphose
10. Mutate

I hope you find these synonyms helpful!

RY where in any of these words is the letter G?

Apologies for the oversight. None of the synonyms listed above for "change" actually start with the letter "G." Here, however, are some synonyms for "change" that do begin with the letter "G":

1. Transform
2. Modify
3. Alter
4. Convert
5. Vary
6. Shift

Regenerate response

Send a message

Uncertainty

Why is Uncertainty important?

Knowing what a model does not know is essential for establishing reliability

Undesirable Consequences

DOT report on fatal 2016 Tesla crash with tractor-trailer blames limitations of Autopilot mode

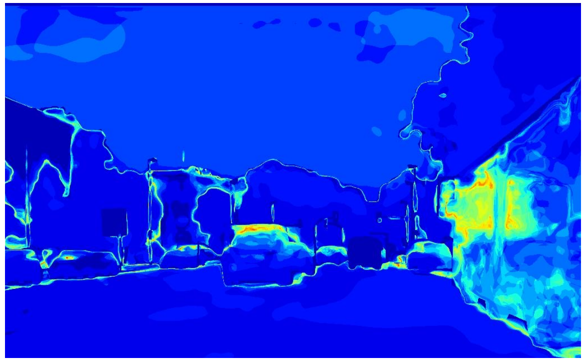
James Jaillet
Feb 2, 2017 | Updated Feb 21, 2017

An NTSB photo of the Freightliner Cascadia involved in the May 7 crash.

Ideal Expectations

Input Image

Uncertainty Heatmap



NHTSA. PE 16-007. Technical report, U.S. Department of Transportation, National Highway Traffic Safety Administration, Jan 2017. Tesla Crash Preliminary Evaluation Report.

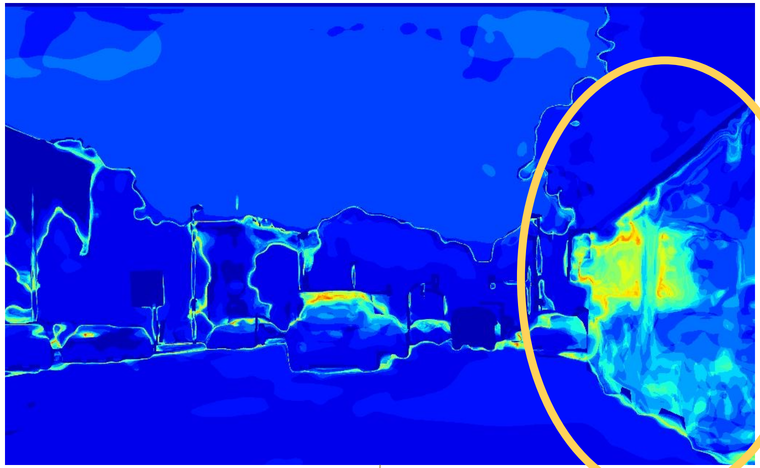
Why is Uncertainty important for Explanations?

Uncertainty provides a mathematical framework to study Explanations

Input Image

Neural Network Output

Uncertainty Heatmap



Visual explanation about what a network does not known

Kendall, Gal "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision." *NIPS 2017*

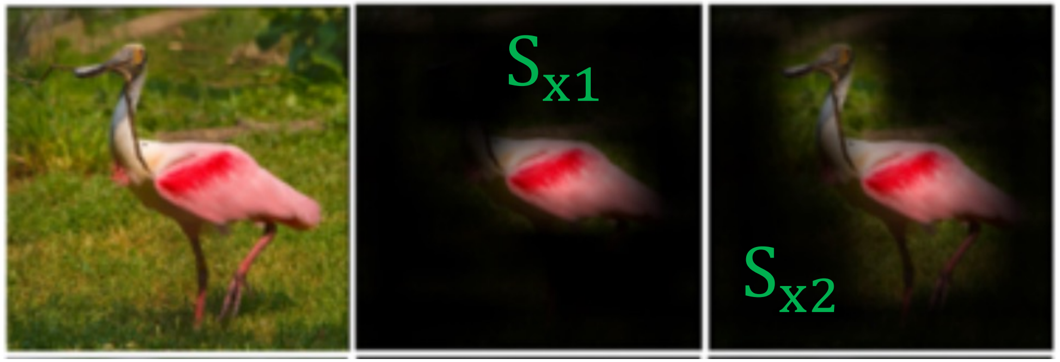
In Lecture 5…

Why is Uncertainty important for Explanations?

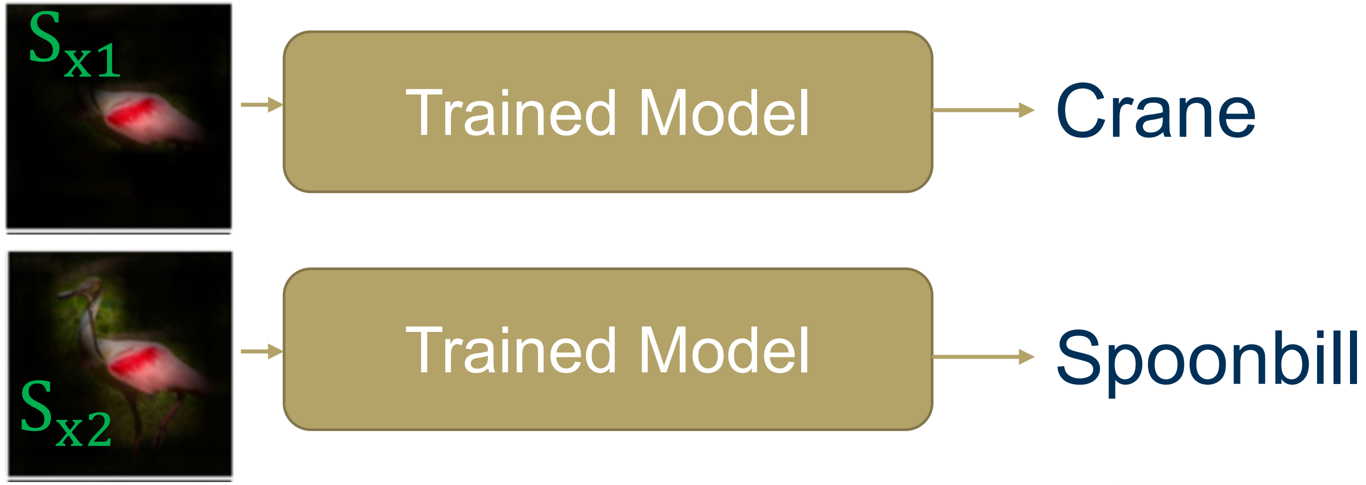
Uncertainty provides a mathematical framework to study Explanations

The prediction Y cannot be trusted under masking

Y = Prediction
S_x = Explanation masked data



In this lecture, we analyze Y under domain shift via uncertainty



In Lecture 2…

Why is Uncertainty important for Explanations?

Uncertainty analysis broadens the scope of Explanations

Let $\mathcal{T}$ be the set of all features learned by a trained network

Explanations maximize the probability of selecting a combination of features $\cup_{i=1}^{P} \mathcal{T}_i$ conditioned on some decision Y:

$$\mathcal{M}(\cdot) = \mathbb{P}\left(\cup_{i=1}^{P} \mathcal{T}_i \middle| Y\right), Y \in [1, N]$$

Beak
Neck
Legs
Feathers
Water
Grass
Teeth
.
.

Features $\mathcal{T}_P$

P is Spoonbill

Prediction Feature Attribution: Visual explanations map features to predictions

AlRegib, Ghassan, and Mohit Prabhushankar. "Explanatory paradigms in neural networks: Towards relevant and contextual explanations." *IEEE Signal Processing Magazine* 39.4 (2022): 59-72.

In Lecture 2…

Why is Uncertainty important for Explanations?

Uncertainty analysis broadens the scope of Explanations

Let $\mathcal{T}$ be the set of all features learned by a trained network

Explanations maximize the probability of selecting a combination of features $\cup_{i=1}^{P} \mathcal{T}_i$ conditioned on some decision Y:

$$\mathcal{M}(\cdot) = \mathbb{P}\left(\cup_{i=1}^{P} \mathcal{T}_i \middle| Y\right), Y \in [1, N]$$

$$\left.\begin{array}{l}\text{Beak}\\ \text{Neck}\\ \text{Legs}\\ \text{Feathers}\\ \text{Water}\\ \text{Grass}\\ \text{Teeth}\\ \vdots\end{array}\right\} \text{Features } \mathcal{T}_{P,Q}$$

P is Spoonbill,
Q is Flamingo

Class Feature Attribution: Visual explanations map features to any trained classes

AlRegib, Ghassan, and Mohit Prabhushankar. "Explanatory paradigms in neural networks: Towards relevant and contextual explanations." *IEEE Signal Processing Magazine* 39.4 (2022): 59-72.

Uncertainty

Why is Uncertainty important for Explanations?

Explanations attribute features to any objective quantity; not just predictions

Let $\mathcal{T}$ be the set of all features learned by a trained network

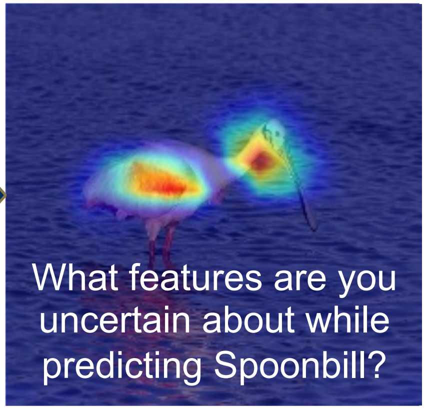
Explanations maximize the probability of selecting a combination of features $\cup_{i=1}^{P} \mathcal{T}_i$ conditioned on some decision Y:

$$\mathcal{M}(\cdot) = \mathbb{P}\left(\cup_{i=1}^{P} \mathcal{T}_i \middle| \boldsymbol{U}\right)$$

Beak
Neck
Legs
Feathers
Water
Grass
Teeth
.
.

Features $\mathcal{T}_U$

P is Spoonbill



Uncertainty Feature Attribution: Visual explanations map features to any objective quantity $\boldsymbol{U}$

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty

Why is Uncertainty important for Explanations?

Explanations attribute features to any objective quantity U; not just predictions

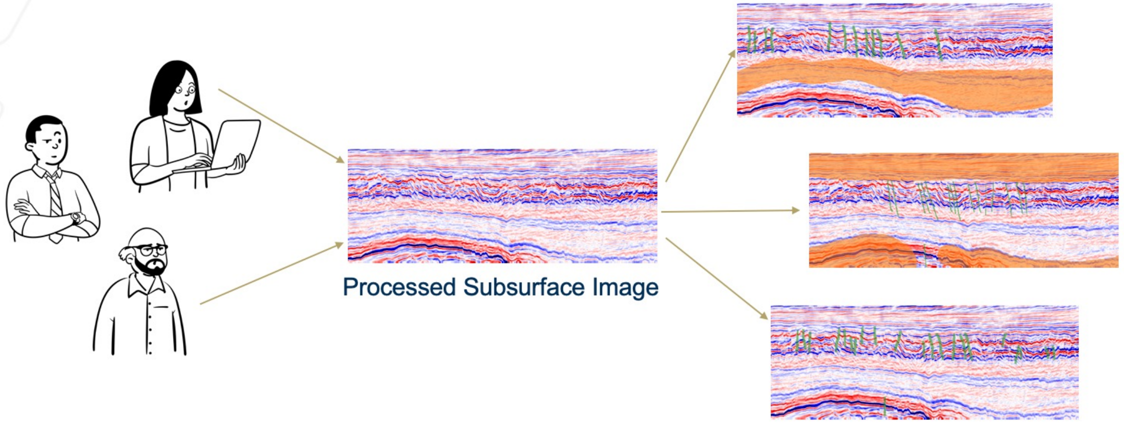
Examples of objective quantity U:

- Noise at acquisition (Robustness)
- Novel data (Robustness)
- Underspecified models (Robustness)
- Label Disagreement (Human annotation subjectivity)
- Visual prompting by different subjects (Human annotation subjectivity)

Any configuration that allows multiple predictions will produce an explanation

Data distortion[1]

Label disagreement[2]

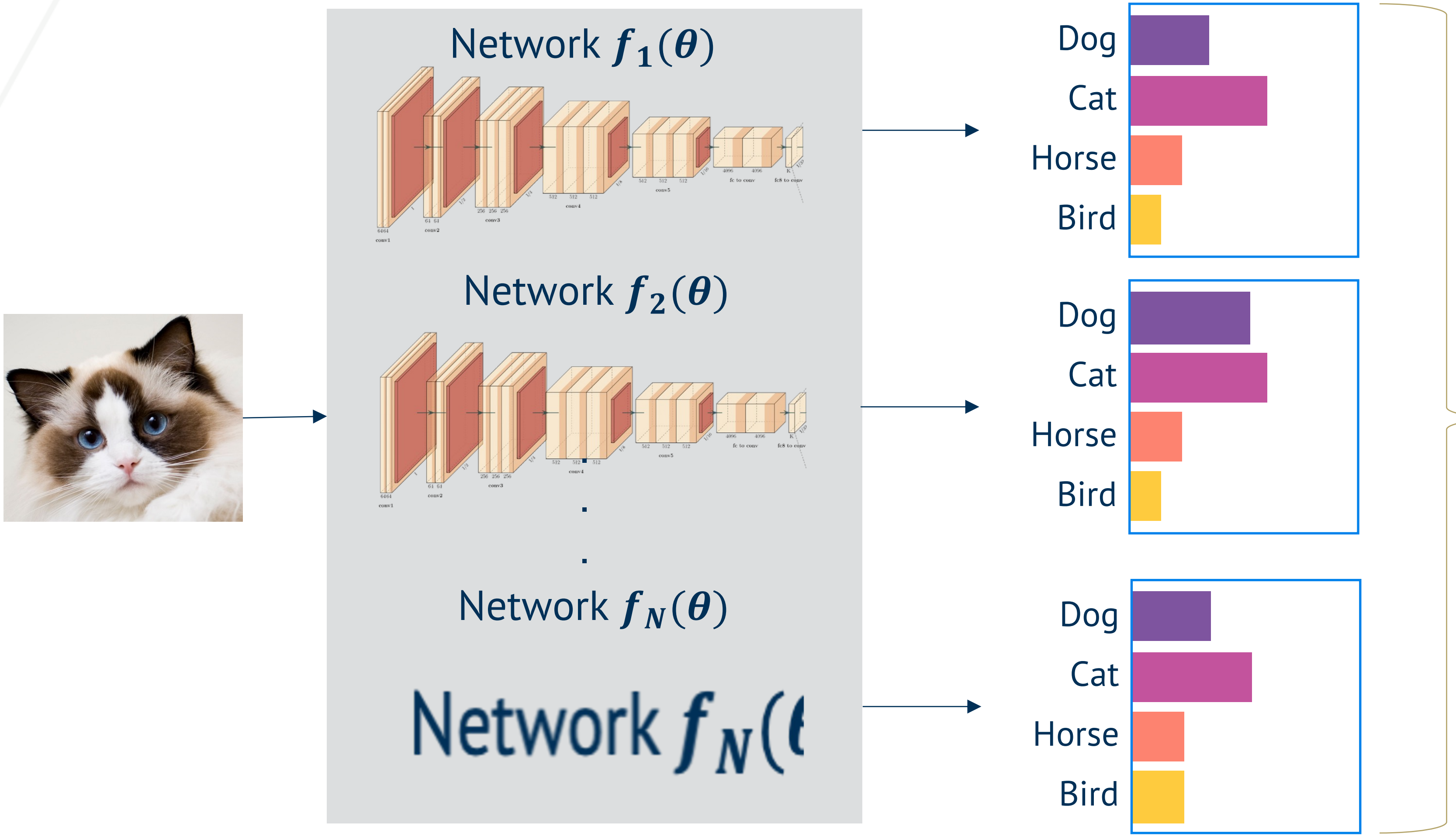


[1] Temel, Dogancan, et al. "CURE-TSR: Challenging unreal and real environments for traffic sign recognition." in NeurIPS Workshop on Machine Learning for Intelligent Transportation Systems, 2017
[2] C. Zhou, M. Prabhushankar, and G. AlRegib, "On the Ramifications of Human Label Uncertainty," in NeurIPS Workshop on Human in the Loop Learning, 2022

Uncertainty

Uncertainty Quantification in Neural Networks

Uncertainty manifests itself as variability in prediction under different model configurations



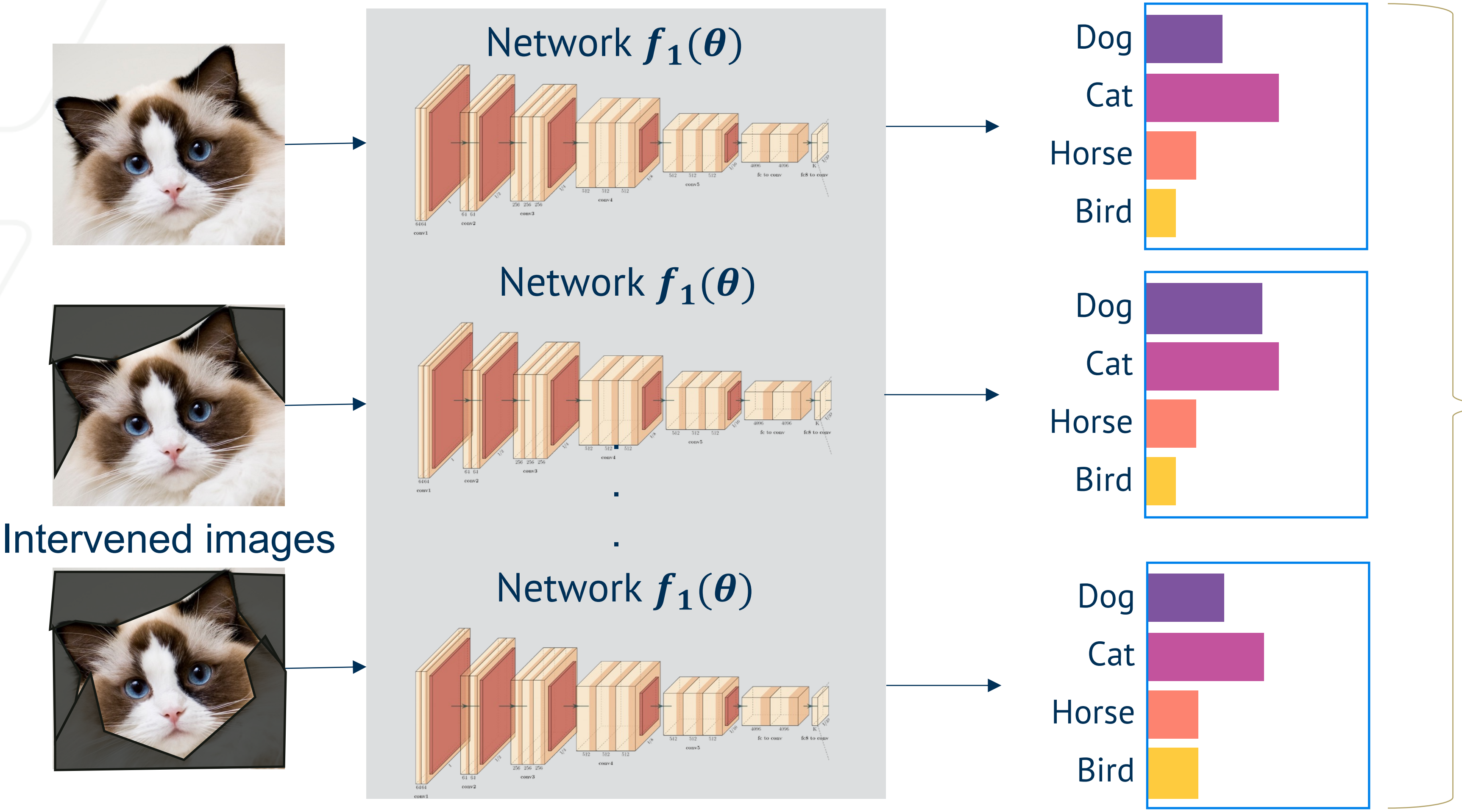
Variation within outputs is the uncertainty. Commonly referred to as **Prediction Uncertainty.**

[1] Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." *Advances in neural information processing systems* 30 (2017).

Uncertainty

Uncertainty Quantification in Neural Networks for Explainability

Uncertainty manifests itself as variability in prediction under different data configurations



Variation within outputs is the uncertainty. Commonly referred to as **Prediction Uncertainty.**

Outline

Lecture 7: Rethinking Explanations via Uncertainty

- Uncertainty
- Visual Explainability and Uncertainty
  - Explanatory evaluation via Uncertainty
  - Explanatory definition
- Uncertainty Quantification
  - Iterative Quantification
  - Monte-Carlo Dropout
  - Visualizing Uncertainties
  - Single Pass Quantification
- Uncertainty in Explanatory Evaluation
  - Predictive Uncertainty
  - Predictive Uncertainty in Explanations
  - Explanation uncertainty analysis
    - Signal-to-Noise Ratio
    - Mean Intersection over Union

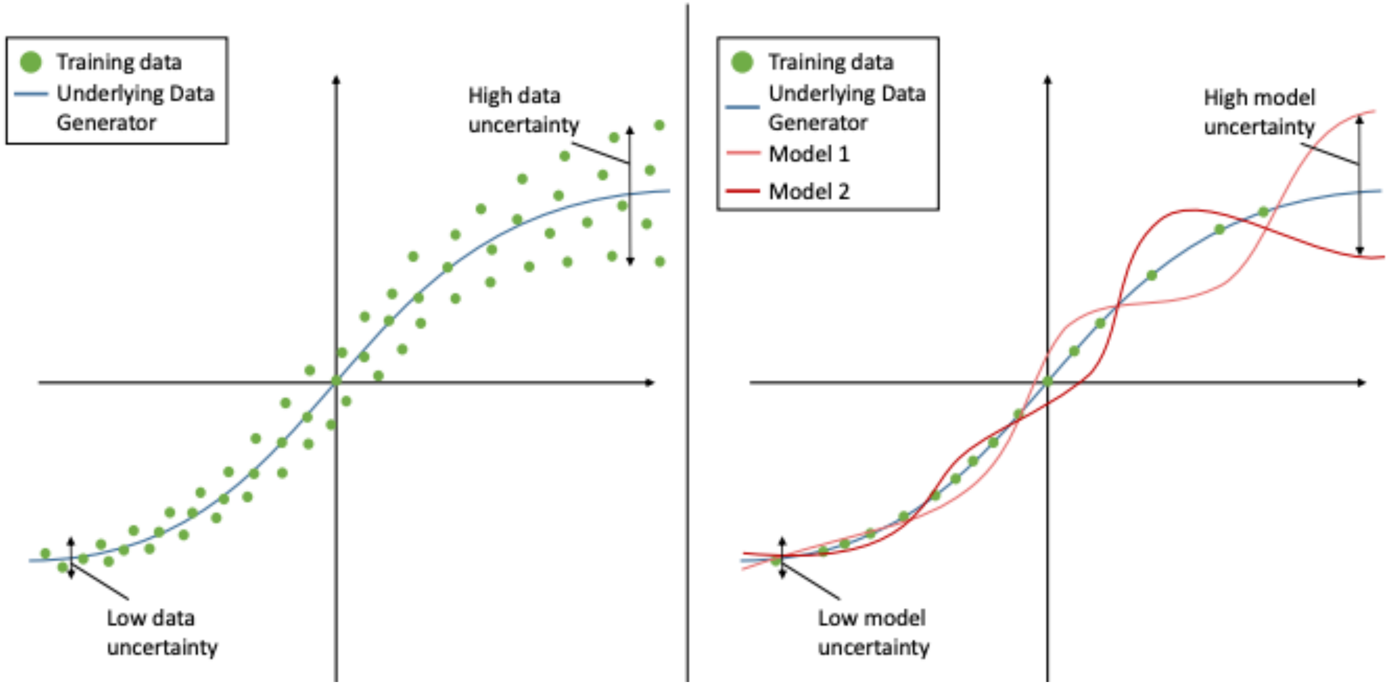
Uncertainty Quantification

Uncertainty Quantification based on source

Two major types of uncertainty: Uncertainty in data and uncertainty in model, together termed as prediction Uncertainty

Predictive Uncertainty

- **Aleatoric** (Data) Uncertainty
  Inherent noise in observation data
- **Epistemic** (Model) Uncertainty
  Insufficient model knowledge



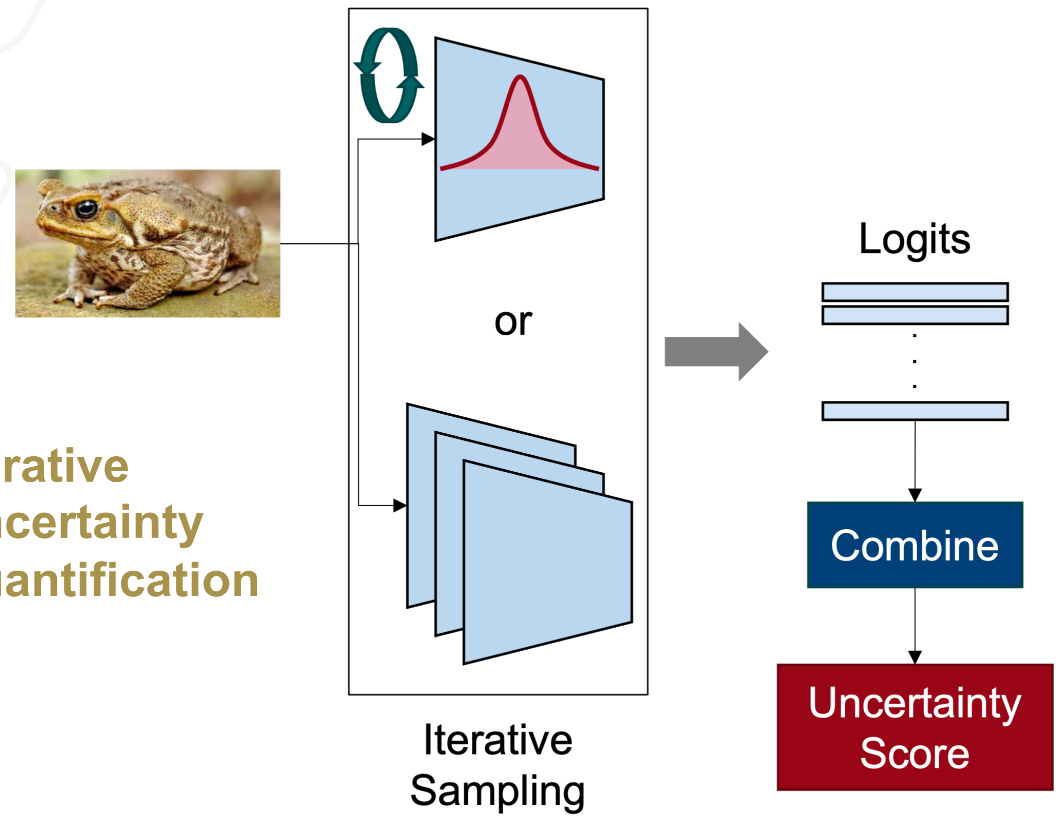
[1] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? Structural Safety, 31 (2):105–112, 2009.

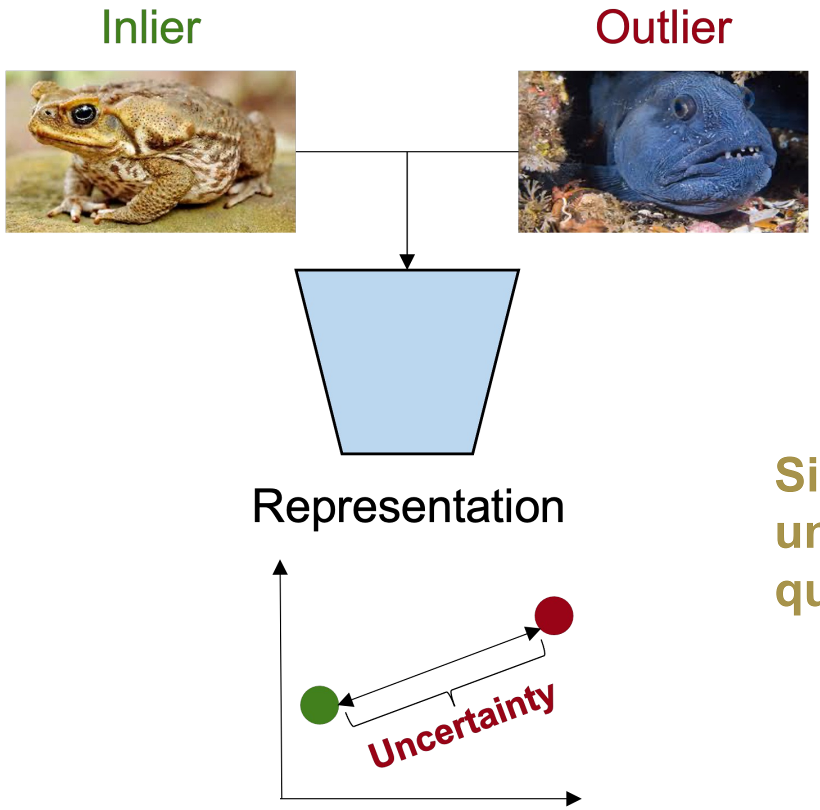
Uncertainty Quantification

Uncertainty Quantification Methodology

Two methods of Uncertainty Quantification: Iterative and Single-pass methods

Iterative uncertainty quantification





Single-pass uncertainty quantification

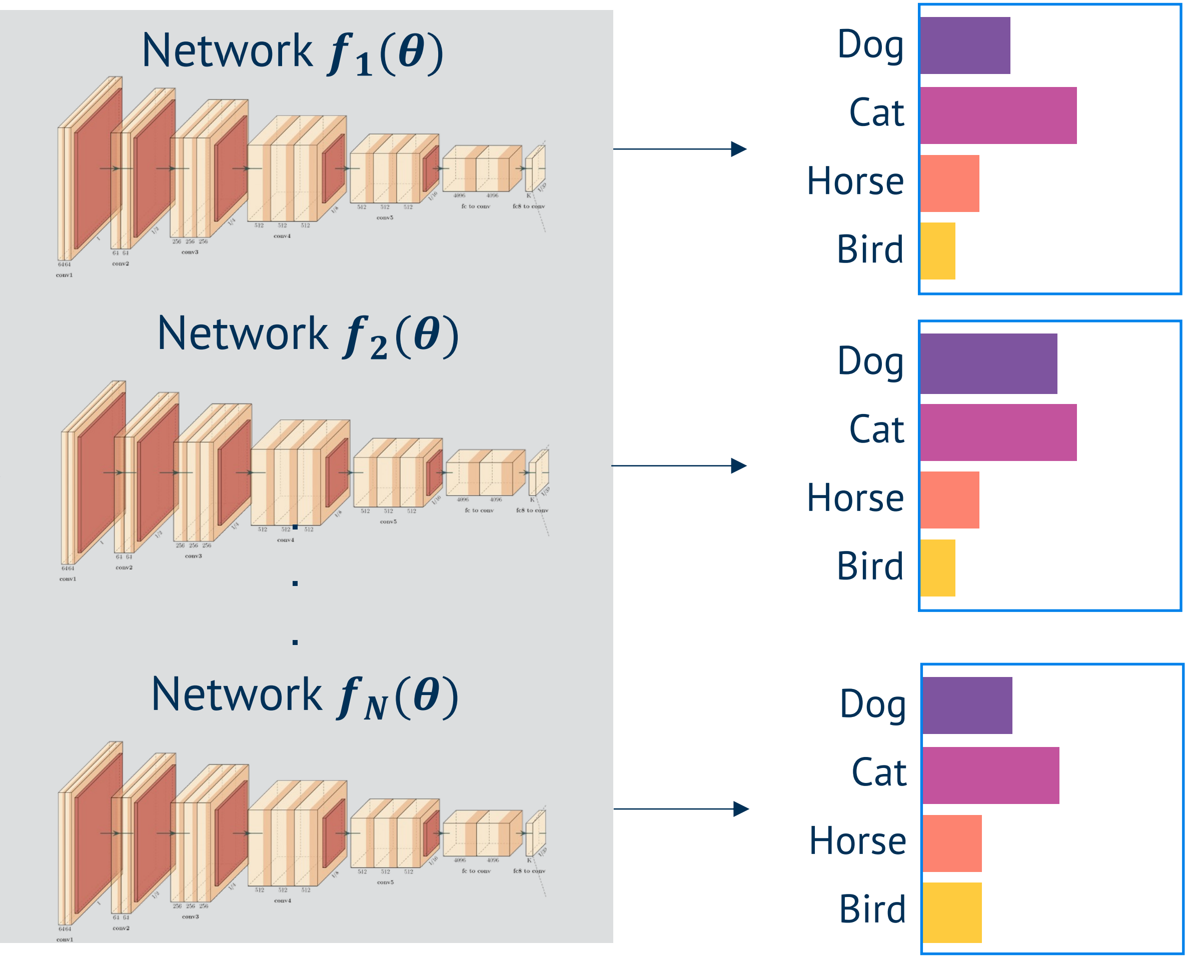
[1] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? Structural Safety, 31 (2):105–112, 2009.

Uncertainty Quantification

Iterative Uncertainty Quantification: Deep Ensembles

Uncertainty Quantification via Deep Ensembles

Different initialization parameters provide $f_1(\cdot), f_2(\cdot), f_3(\cdot)$, and different outputs.



Not always realistic to obtain multiple networks

Final prediction is the mean of the outputs

Variation within outputs is the uncertainty.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." *Advances in neural information processing systems* 30 (2017).

Uncertainty Quantification

Iterative Uncertainty Quantification: Monte-Carlo Dropout

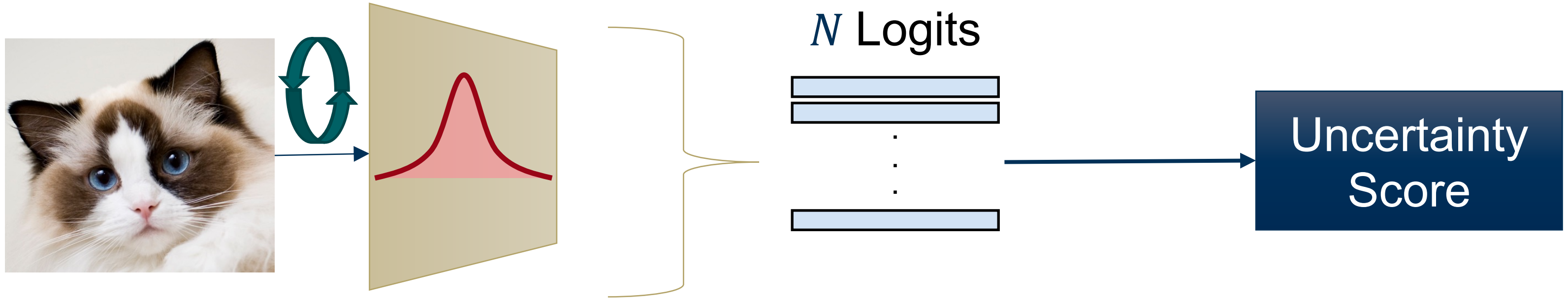
Uncertainty Quantification via Monte-Carlo Dropout: During inference repeated evaluations with the same input give the different results

Different forward passes with dropout simulate $f_1(\cdot), f_2(\cdot), f_3(\cdot)$.

Challenge: intractable denominator

$$p(\boldsymbol{W}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{W})p(\boldsymbol{W})}{\int p(\boldsymbol{x}|\boldsymbol{W})p(\boldsymbol{W})d\boldsymbol{W}}$$

N forward passes



Final prediction is the mean of the outputs

Variation or entropy of logits is the uncertainty

$$q\,(\boldsymbol{W}_N) \approx p(\boldsymbol{W}_N|\boldsymbol{x})$$

Y Gal, Z Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning", ICML 2016

Uncertainty Quantification via Monte-Carlo Dropout: During inference repeated evaluations with the same input give the different results

$$U_{epistemic} = \underbrace{H\left(\frac{1}{T}\sum_{t=1}^{T} Softmax\left(f_{\widehat{\boldsymbol{W}}_t}(\boldsymbol{x})\right)\right)}_{U_{Predictive}} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} H\left(Softmax\left(f_{\widehat{\boldsymbol{W}}_t}(\boldsymbol{x})\right)\right)}_{U_{aleatoric}}$$

Entropy of expectation of predictions

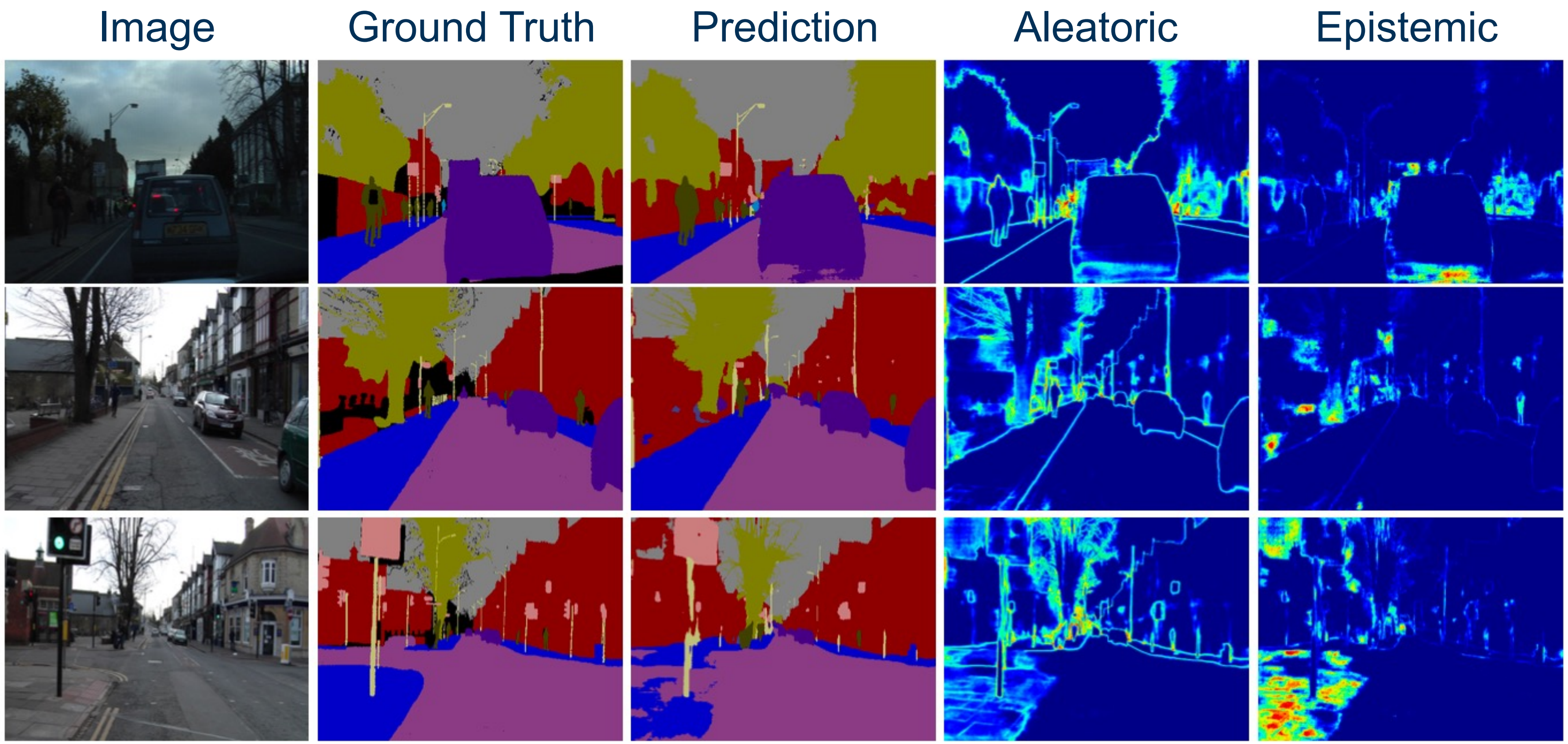
Expectation of individual entropy of predictions

Y Gal, Z Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning", ICML 2016

Uncertainty Quantification

Iterative Uncertainty Quantification

Uncertainty Quantification via Monte-Carlo Dropout: During inference repeated evaluations with the same input give the different results

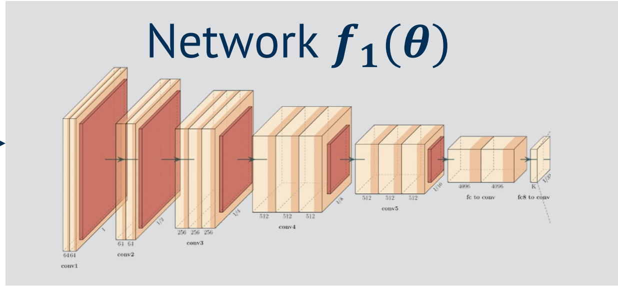


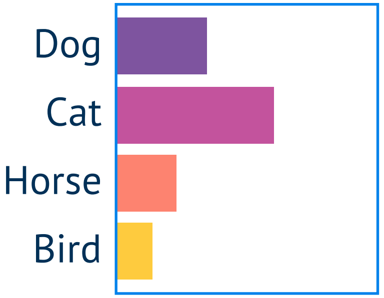
Y Gal, Z Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning", ICML 2016

Uncertainty Quantification

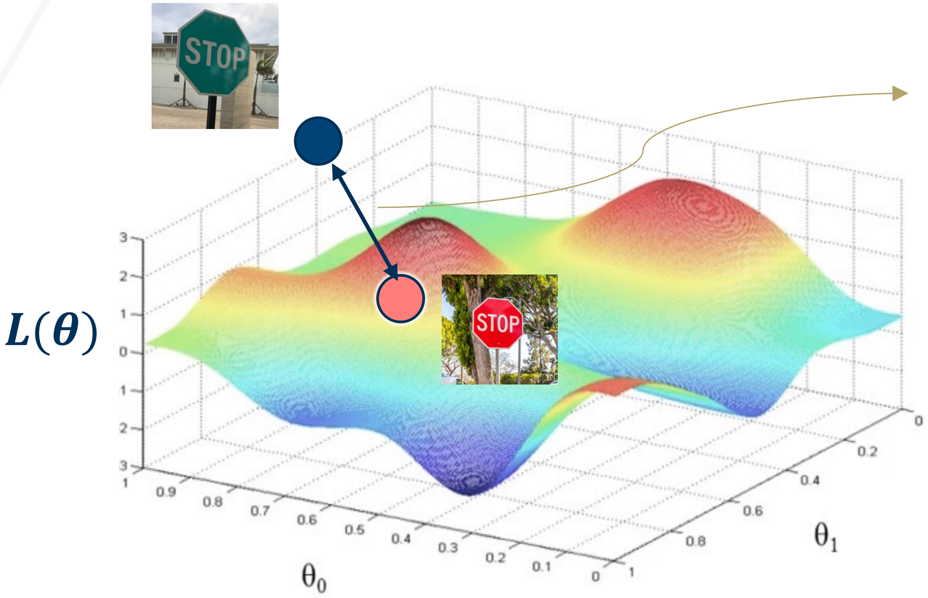
Single Pass Uncertainty Quantification

Via Single pass methods[1]





Uncertainty quantification using a single network and a single pass



Calculate distance from some trained clusters

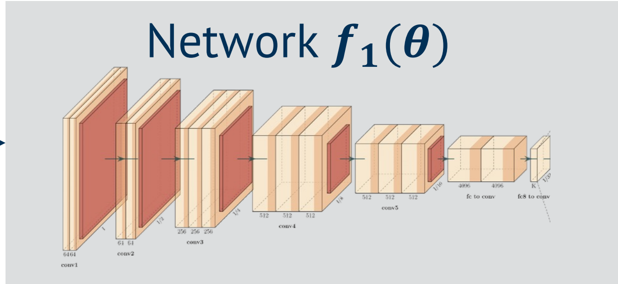
Does not require multiple networks!

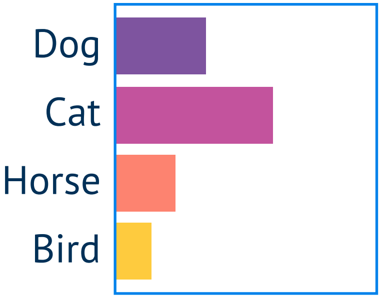
[1] Van Amersfoort, J., Smith, L., Teh, Y. W., & Gal, Y. (2020, November). Uncertainty estimation using a single deep deterministic neural network. In *International conference on machine learning* (pp. 9690-9700). PMLR.

Uncertainty Quantification

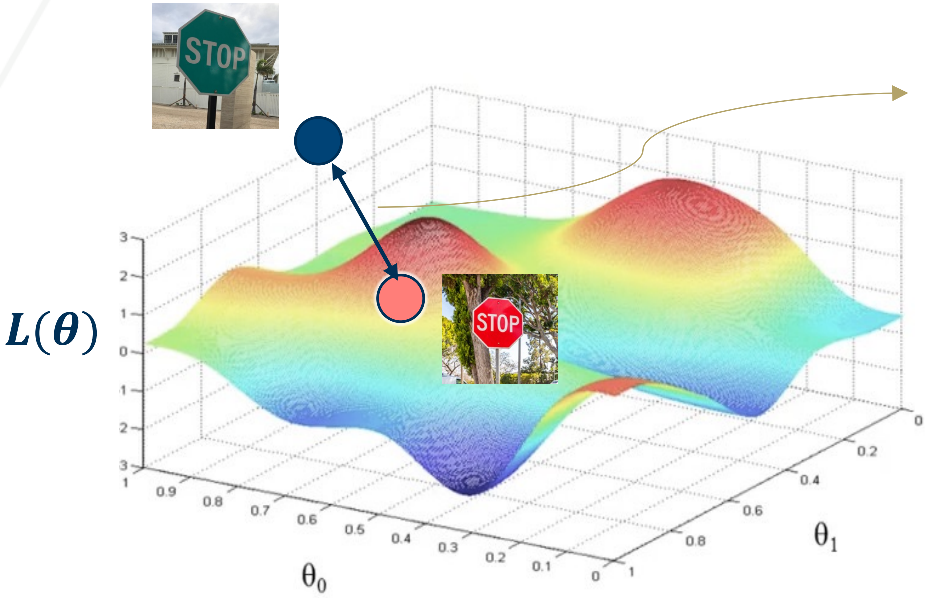
Single Pass Uncertainty Quantification

Via Single pass methods[1]





Uncertainty quantification using a single network and a single pass



Gradients provide this distance from Lecture 6

Collection of squared L2 norm $\boldsymbol{d_{\nabla\theta}}$

$$\left\|\nabla_{\theta_0} J(\theta_0;\ x, y_c)\right\|_2^2 \ \dots \ \left\|\nabla_{\theta_N} J(\theta_N;\ x, y_c)\right\|_2^2$$

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Outline

Lecture 7: Rethinking Explanations via Uncertainty

- Uncertainty
- Visual Explainability and Uncertainty
    - Explanatory evaluation via Uncertainty
    - Explanatory definition
- Uncertainty Quantification
    - Iterative Quantification
    - Monte-Carlo Dropout
    - Visualizing Uncertainties
    - Single Pass Quantification
- Uncertainty in Explanatory Evaluation
    - Predictive Uncertainty
    - Predictive Uncertainty in Explanations
    - Explanation uncertainty analysis
        - Signal-to-Noise Ratio
        - Mean Intersection over Union

Uncertainty in Explainability

Why is Uncertainty important for Explanations?

If explanation map is $\mathcal{M}(\cdot) = \mathbb{P}(\cup_{i=1}^{P} \mathcal{T}_i | P)$

Uncertainty map is $\mathcal{M}_u(\cdot) = 1 - \mathbb{P}(\cup_{i=1}^{P} \mathcal{T}_i | P)$

Explanatory techniques have predictive uncertainty

Explanation of Prediction | Uncertainty of Explanation

Why Bullmastiff?

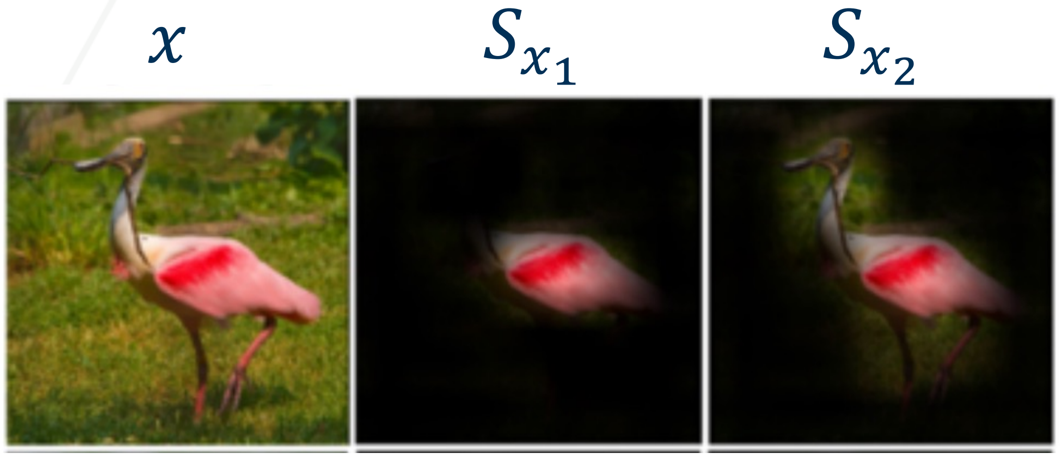
Uncertainty in answering Why Bullmastiff?

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Why is Uncertainty important for Explanations?

Uncertainty due to variance in prediction when model is kept constant



$$V[y|S_x] = V[E(y|S_x)] + E(V[y|S_x])$$

y = Prediction
$V[y]$ = Variance of prediction (Predictive Uncertainty)
S_x = Subset of data (Some intervention)
$E(Y|S_x)$ = Expectation of class given a subset
$V(Y|S_x)$ = Variance of class given all other residuals

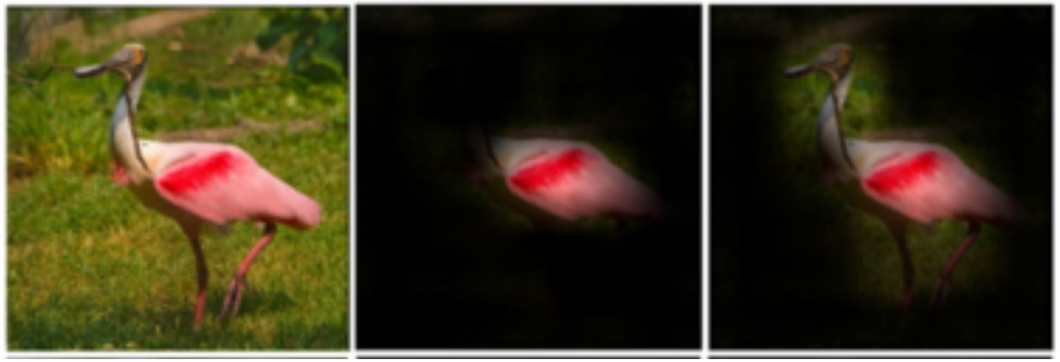
M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Visual Explanations (partially) reduce Predictive Uncertainty

A 'good' explanatory technique is evaluated to have zero $V[E(y|S_x)]$

x $\quad S_{x_1} \quad S_{x_2}$



$$V[y|S_x] = V[E(y|S_x)] + E(V[y|S_x])$$

zero

y = Prediction
$V[y]$ = Variance of prediction (Predictive Uncertainty)
S_x = Subset of data (Some intervention)
$E(Y|S_x)$ = Expectation of class given a subset
$V(Y|S_x)$ = Variance of class given all other residuals

Key Observation 1: Visual Explanations are evaluated to partially reduce the predictive uncertainty in a neural network

Network evaluations have nothing to do with human Explainability!

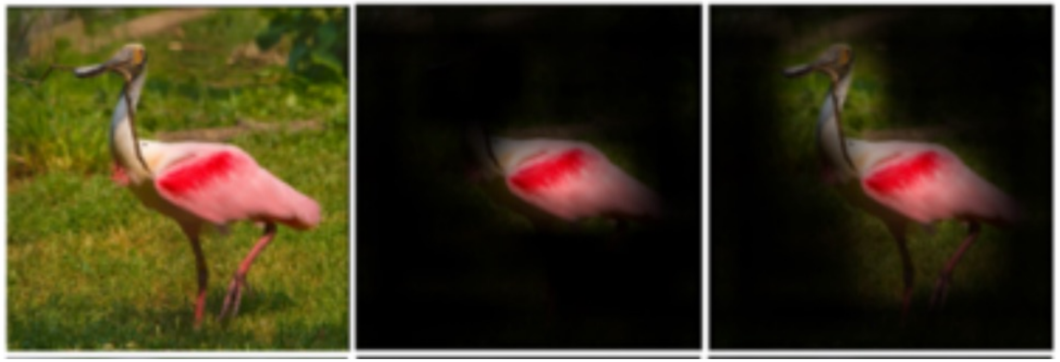
M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Predictive Uncertainty in Explanations is the Residual

All other subsets ‘not’ chosen by the explanatory technique contributes to uncertainty

x $\quad S_{x_1}$ $\quad S_{x_2}$



$$V[y|S_x] = V[E(y|S_x)] + E(V[y|S_x])$$

y = Prediction
$V[y]$ = Variance of prediction (Predictive Uncertainty)
$\mathrm{S_x}$ = Subset of data (Some intervention)
$\mathrm{E(Y|S_x)}$ = Expectation of class given a subset
$\mathrm{V(Y|S_x)}$ = Variance of class given all other residuals

Key Observation 2: Uncertainty in Explainability occurs due to all combinations of features that the explanation did not attribute to the network’s decision

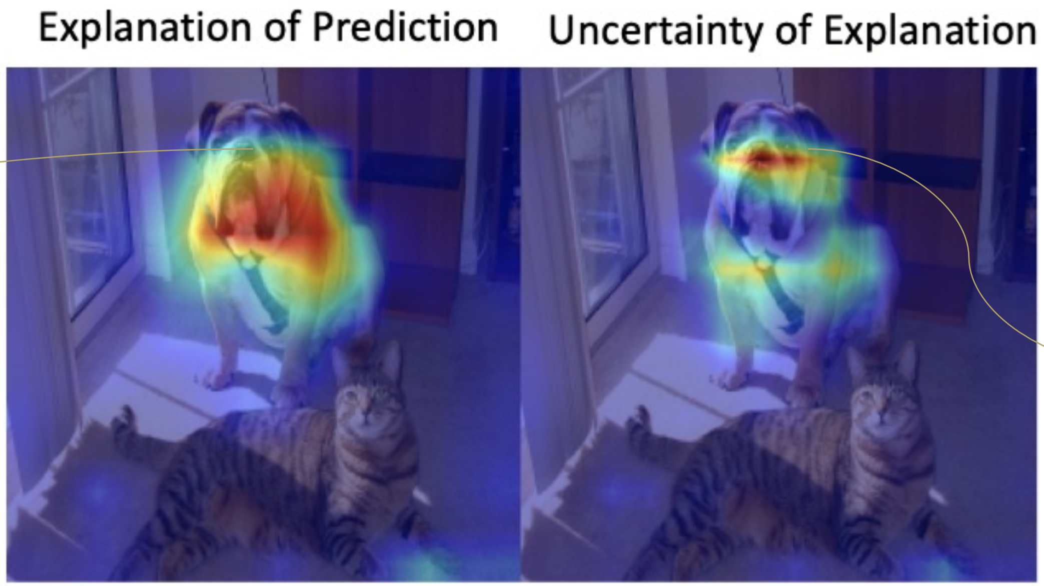
M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Predictive Uncertainty in Explanations is the Residual

All other subsets 'not' chosen by the explanatory technique contributes to uncertainty

Snout is not as highlighted as the jowls in explanation (not as important for decision)



However, snout is an important characteristic that is used to differentiate against other dogs. Hence, there is uncertainty on why this feature is not included in the attribution

Key Observation 2: Uncertainty in Explainability occurs due to all combinations of features that the explanation did not attribute to the network's decision

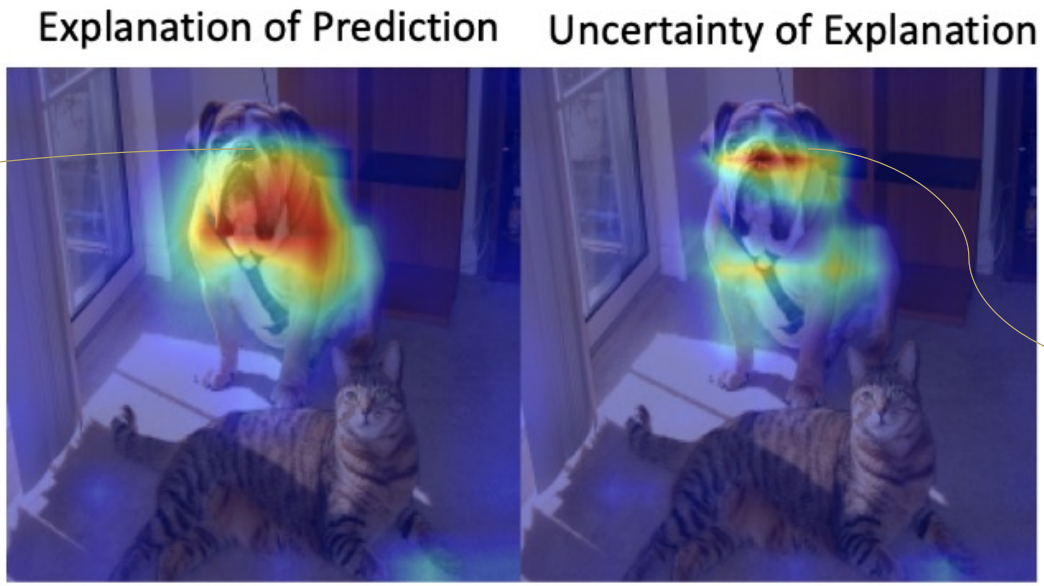
M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Predictive Uncertainty in Explanations is the Residual

All other subsets 'not' chosen by the explanatory technique contributes to uncertainty

Snout is not as highlighted as the jowls in explanation (not as important for decision)



However, snout is an important characteristic that is used to differentiate against other dogs. Hence, there is uncertainty on why this feature is not included in the attribution

Not chosen features are intractable!

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Quantifying Uncertainty in Explainability

Contrastive explanations are an intelligent way of obtaining other subsets

$$V[y|S_x] = V[E(y|S_x)] + E(V[y|S_x])$$

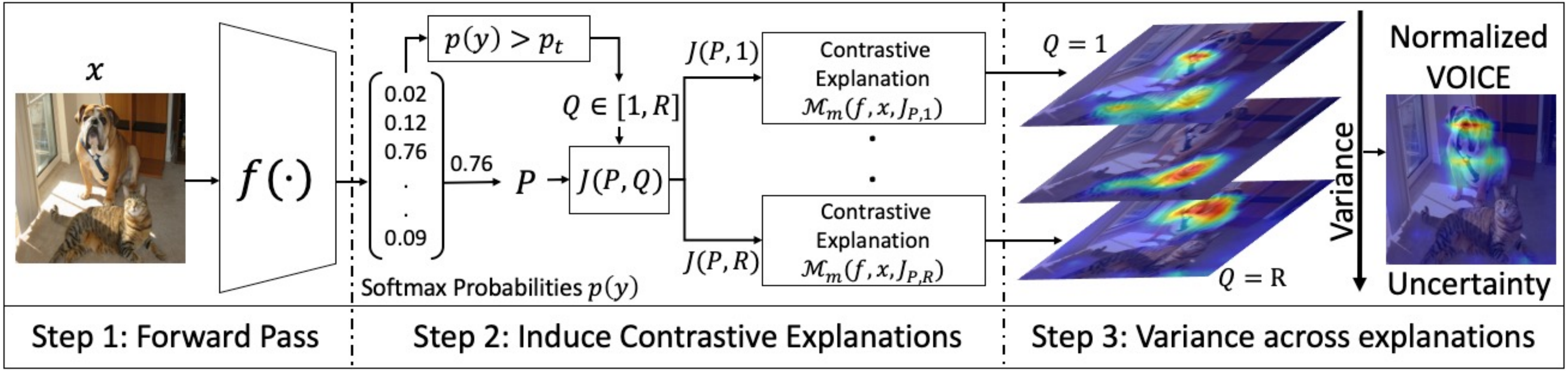
Make it finite by only considering the subsets that change y

S_{x_1} S_{x_2} .... S_{x_N}

$$\left.\begin{array}{c} Y_1|S_{x1} \\ Y_2|S_{x2} \\ Y_3|S_{x3} \\ Y_4|S_{x4} \\ Y_5|S_{x5} \\ . \\ . \\ Y_N|S_{xN} \end{array}\right\} \text{Variance}$$

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Variance in contrastive explanations provides uncertainty



M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Quantifying Uncertainty in Explainability

Uncertainty in Explainability can be used to analyze Explanatory methods and Networks

- Is GradCAM better than GradCAM++?
- Is a SWIN transformer more reliable than VGG-16?

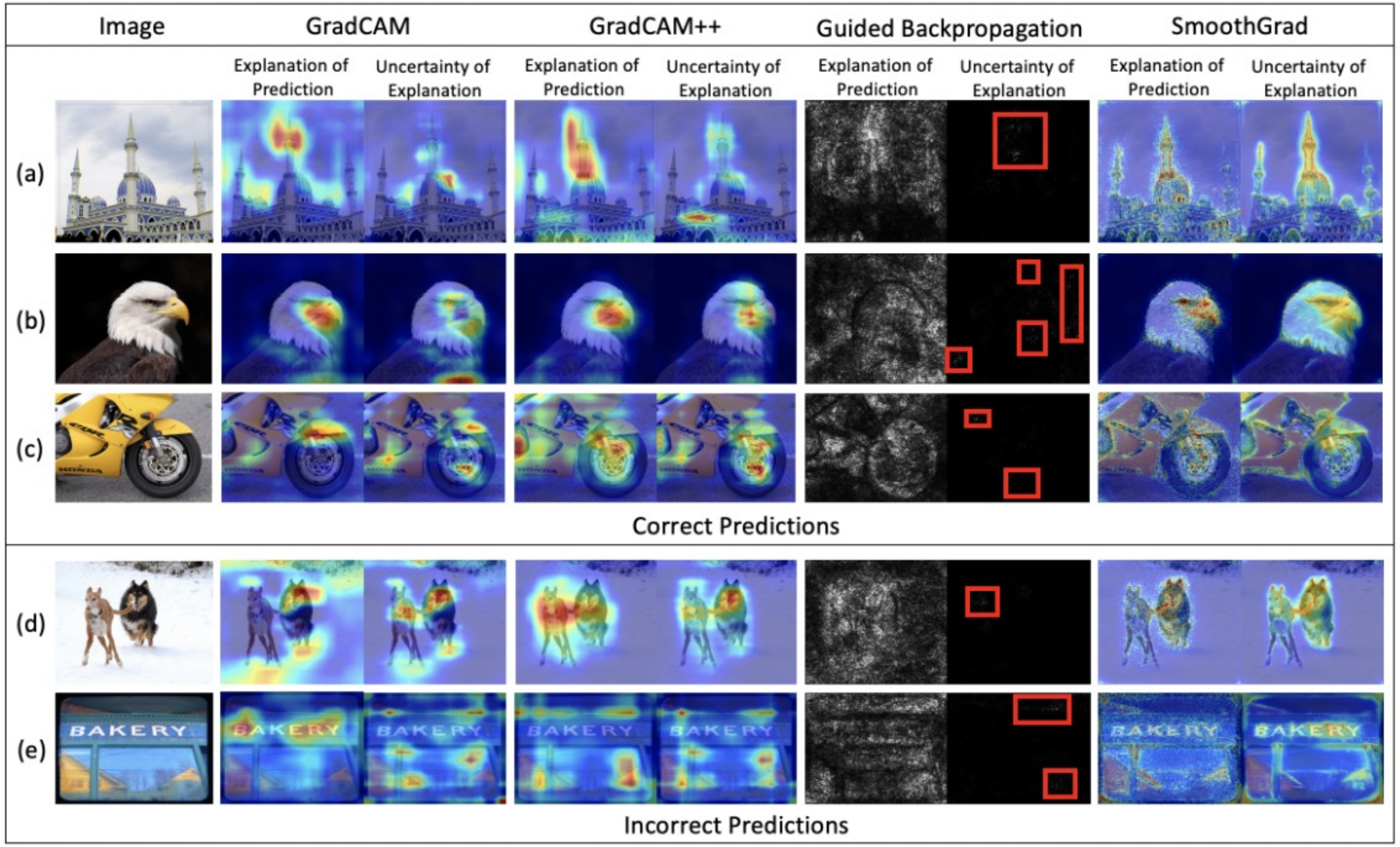
Need objective quantification of Uncertainty

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Quantifying Uncertainty in Explainability: mIOU

On incorrect predictions, the overlap of explanations and uncertainty is higher



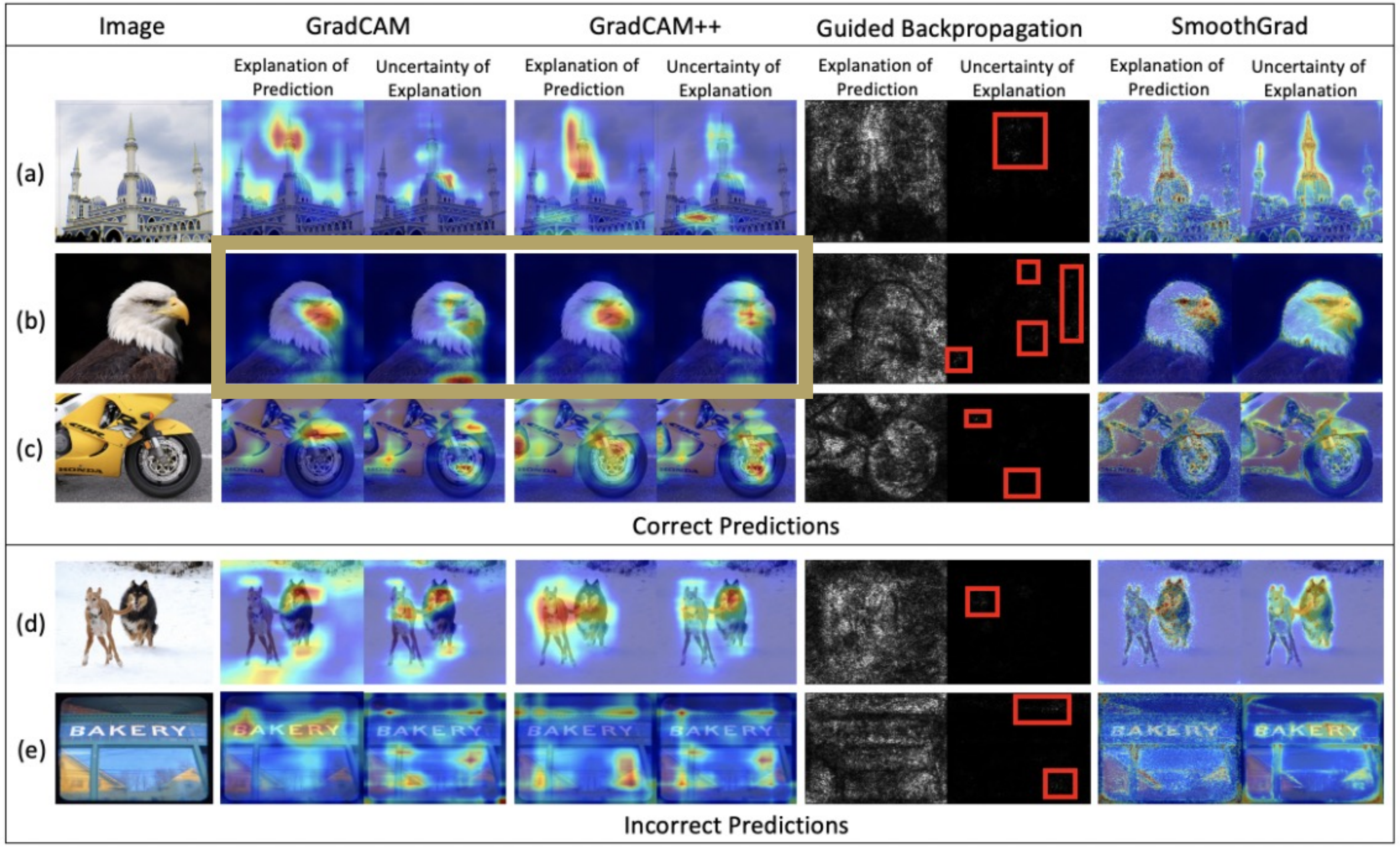
Objective Metric 1: Intersection over Union (IoU) between explanation and Uncertainty

Higher the IoU, higher the uncertainty in explanation (or less trustworthy is the explanation)

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Quantifying Uncertainty in Explainability: mIOU

On incorrect predictions, the overlap of explanations and uncertainty is higher



Objective Metric 1: Intersection over Union (IoU) between explanation and Uncertainty

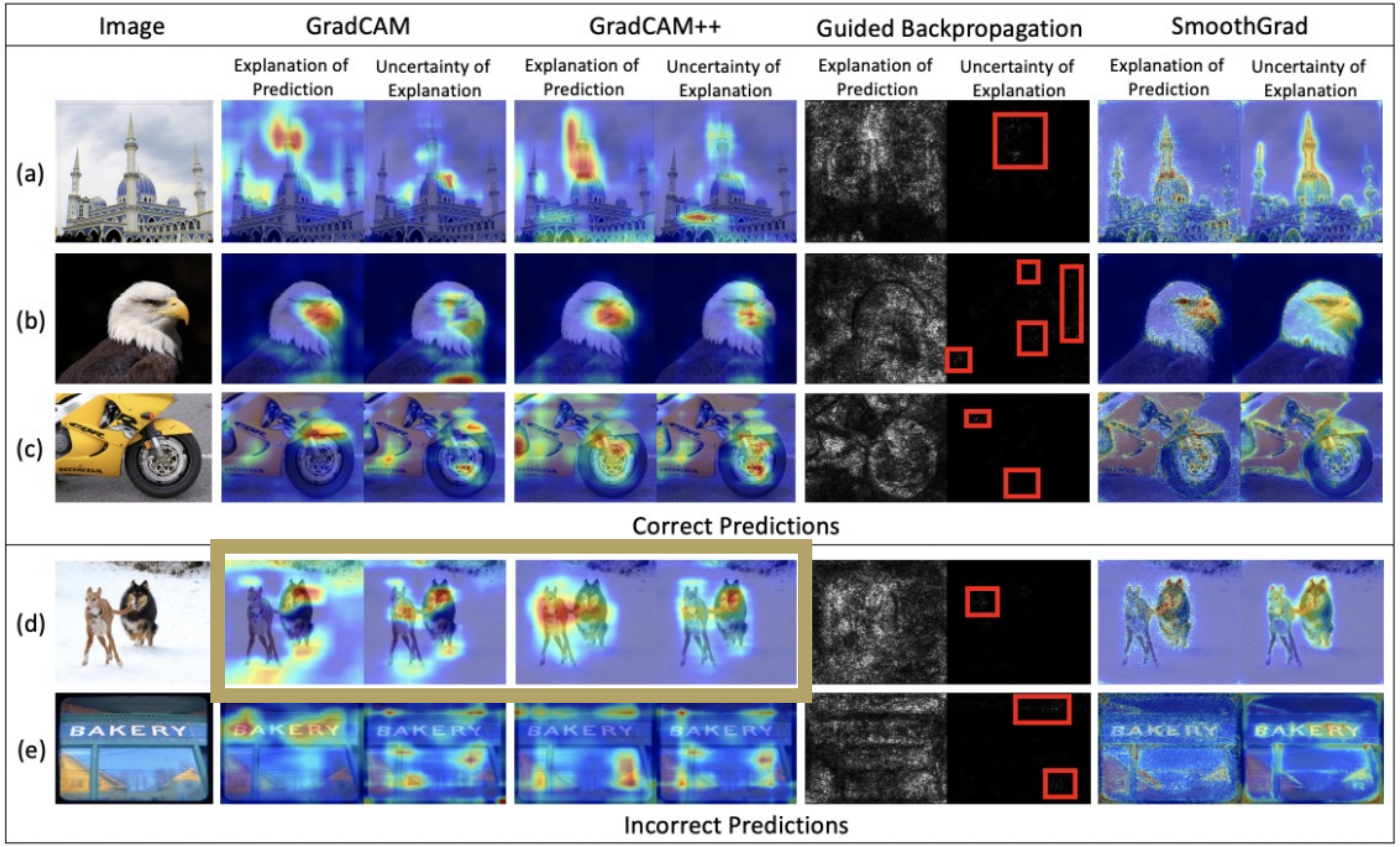
Higher the IoU, higher the uncertainty in explanation (or less trustworthy is the explanation)

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Quantifying Uncertainty in Explainability: mIOU

On incorrect predictions, the overlap of explanations and uncertainty is higher



Objective Metric 1: Intersection over Union (IoU) between explanation and Uncertainty

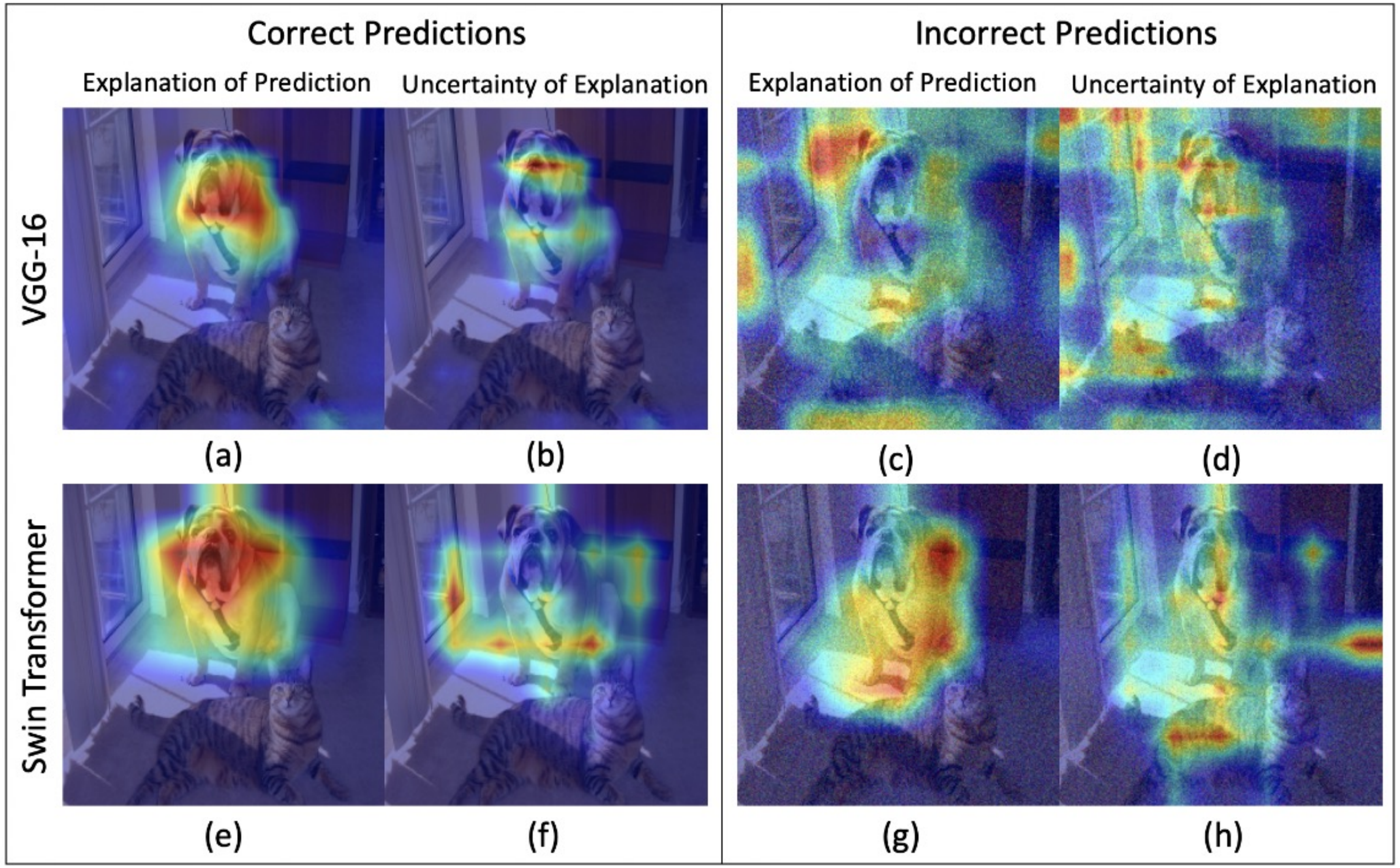
Higher the IoU, higher the uncertainty in explanation (or less trustworthy is the prediction)

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Uncertainty in Explainability

Quantifying Uncertainty in Explainability: SNR

Explanation and uncertainty are dispersed under noise (under low prediction confidence)



Objective Metric 2: Signal to Noise Ratio of the Uncertainty map

Higher the SNR of uncertainty, more is the dispersal (or less trustworthy is the prediction)

M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.

Conclusions

Lecture 7: Rethinking Explanations via Uncertainty

- **Uncertainty** is a model **knowing what it does not know**
- Uncertainty Quantification is studied by understanding the sources of uncertainties
  - If the source is data, we quantify Aleatoric Uncertainty
  - If the source is the model, we quantify Epistemic Uncertainty
- Predictive uncertainty is a sum of Aleatoric and Epistemic Uncertainties
- **Network evaluation encourages Explanations to reduce Predictive Uncertainty**
- **The residuals among all the unchosen subsets causes Predictive Uncertainty**
- Any quantification that allows multiple predictions can be visualized as an explanation
- **Contrastive Explanations** can be used to **visualize Uncertainties** in Explainability

References

Lecture 7: Rethinking Explanations via Uncertainty

- Kendall, Alex, and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?." *Advances in neural information processing systems* 30 (2017).
- AlRegib, Ghassan, and Mohit Prabhushankar. "Explanatory paradigms in neural networks: Towards relevant and contextual explanations." *IEEE Signal Processing Magazine* 39.4 (2022): 59-72.
- M. Prabhushankar, and G. AlRegib, "VOICE: Variance of Induced Contrastive Explanations to Quantify Uncertainty in Neural Network Interpretability," *Journal of Selected Topics in Signal Processing*, submitted on Aug. 27, 2023.
- Temel, Dogancan, et al. "CURE-TSR: Challenging unreal and real environments for traffic sign recognition." in NeurIPS Workshop on Machine Learning for Intelligent Transportation Systems, 2017
- C. Zhou, M. Prabhushankar, and G. AlRegib, "On the Ramifications of Human Label Uncertainty," in NeurIPS Workshop on Human in the Loop Learning, 2022
- Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." *Advances in neural information processing systems* 30 (2017).
- Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? Structural Safety, 31 (2):105–112, 2009.
- Y Gal, Z Ghahramani, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”, ICML 2016
- Van Amersfoort, J., Smith, L., Teh, Y. W., & Gal, Y. (2020, November). Uncertainty estimation using a single deep deterministic neural network. In *International conference on machine learning* (pp. 9690-9700). PMLR.
- Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.