JHU vision lab

Characterizing Semantic Information Content in Data

Aditya Chattopadhyay, Benjamin Haeffele, Donald Geman, René Vidal

MINDS Retreat – March 21st 2022

Why Semantic Information?

Images, Text, Hyperspectral

Goal: develop information theoretic framework for quantifying the semantic information content in multimodal data for a task

Data	Representation	Task
X	$\longrightarrow Z \longrightarrow$	Y

Classification, Detection, Scene Parsing

Approach: learn minimal representation of the data that is sufficient for the task

- Why do we want to quantify semantic information?
  - Such a measure could help in learning more interpretable representations of data.
  - Semantics could help characterize the complexity of a learning task, so as to compare one task with another.
  - Such a measure could help assess which data modalities are most important/relevant/informative for a task.
- Classical notions of information are insufficient: task-agnostic.
- **Proposed approach**: learn ``semantic” representation for a task.

Prior Approach: MSI Representations

- **Prior Work**: representations are functions of the data that are
  - Sufficient: as informative as the data
  - Invariant: discount the effect of uninformative data transformations
  - Minimal: “simpler” than the data, ideally minimal
  - Disentangled?

- Trade-off minimality and sufficiency by minimizing information bottleneck [Tishby-Bialek-Pereira ’99]: but doesn’t generalize

$$\min_{q(z|x)} \mathcal{L} \doteq H_{p,q}(y|z) + \beta I(z;x)$$

S. Soatto and A. Chiuso. Visual representations: Defining properties and deep approximations. ICLR, 2016.
A. Achille and S. Soatto, On the Emergence of Invariance and Disentangling in Deep Representations; JMLR. ArXiv:1706.01350
A. Achille & S. Soatto, Information Dropout, PAMI 2018.

Proposed Approach: Semantic Representation

Images, Text, Hyperspectral

Goal: develop information theoretic framework for quantifying the semantic information content in multimodal data for a task

Data | Representation | Task

$$X \longrightarrow Z \longrightarrow Y$$

Classification, Detection, Scene Parsing

Approach: learn minimal representation of the data that is sufficient for the task

- **Minimal, Sufficient, Invariant Representations**: functions of the data (features) that are informative for a class of tasks.

- **Latent Representations**: variables are latent, not necessarily interpretable in human language, i.e., not semantic.

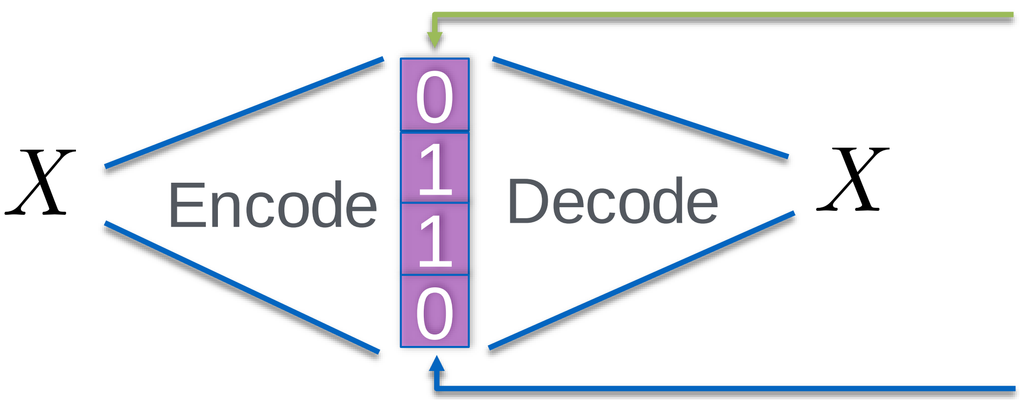
- **Semantic Representations**: representations that depend on a semantic vocabulary that is relevant for the task.

S. Soatto and A. Chiuso. Visual representations: Defining properties and deep approximations. In International Conference on Learning Representations, 2016.

An Interpretation of Shannon's Entropy

- We denote input random variable as X and the output random variable as Y. The joint $P_{X,Y}$ implicitly defines the learning task.

- If the task was transmission/compression, so that $Y = X$



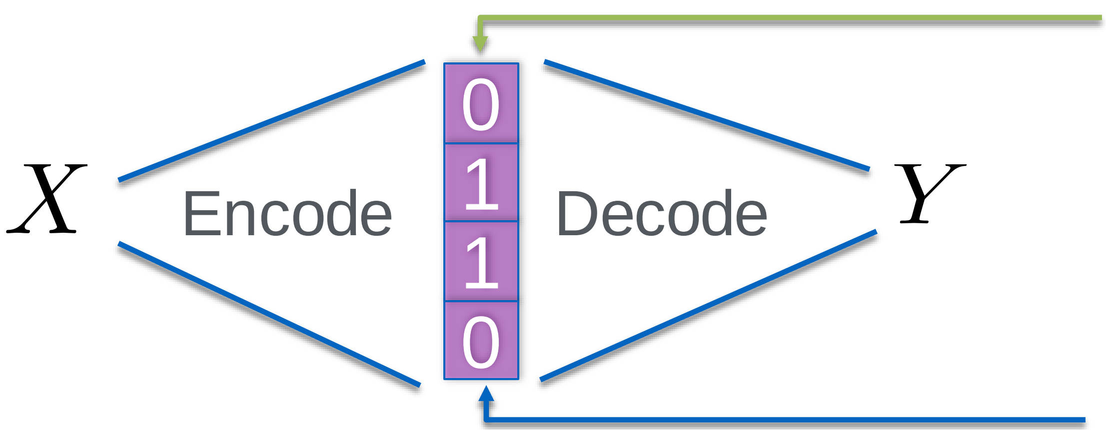
Each bit can be seen as an answer to a binary question on X.

$\mathbb{E}[|Code_Q^E(X)|] \geq \lceil \mathrm{Entropy}(X) \rceil$

- Query set Q is the set of all possible binary functions defined on X. E refers to the coding strategy/encoder.

From Entropy to Semantic Entropy

- Replace “bits” by (task-dependent) elementary atoms of semantic information – “semantic bits” (s-bits) from a user-defined query set Q.



Each s-bit can be seen as an answer $q(X)$ to a question $q \in Q$

$$\mathbb{E}[|Code_Q^E(X)|] \geq SE_Q(X;Y)$$

- Generalize Entropy to Semantic Entropy in X for Y.

$$SE_Q(X;Y) := \min_{\text{Coding Strategy E}} \mathbb{E}[|Code_Q^E(X)|]$$

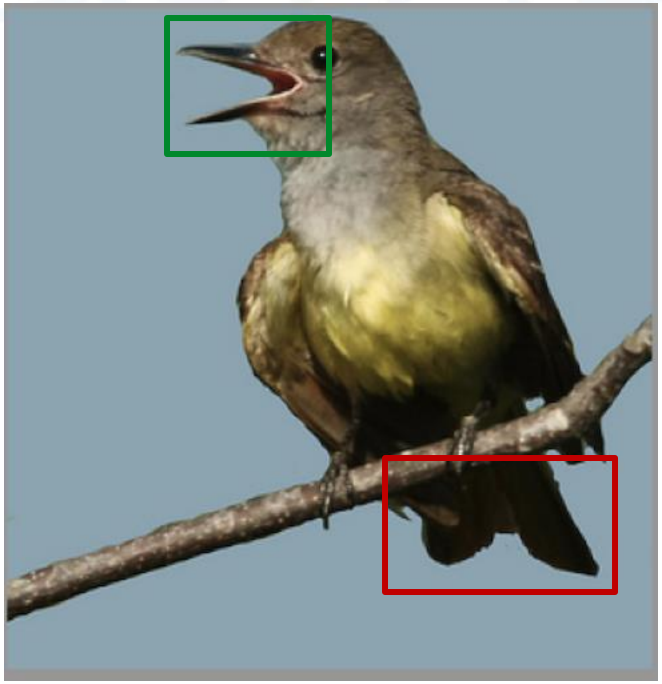
$$\text{s.t. } p(y \mid Code_Q^E(x)) = p(y \mid x) \ \forall x, y$$

Potential Query Sets

1. Q: Is there a person in the blue region? A: yes
2. Q: Is there a unique person in the blue region? A: yes
   (Label this person 1)
3. Q: Is person 1 carrying something? A: yes
4. Q: Is person 1 female? A: yes
5. Q: Is person 1 walking on a sidewalk? A: yes
6. Q: Is person 1 interacting with any other object? A: no

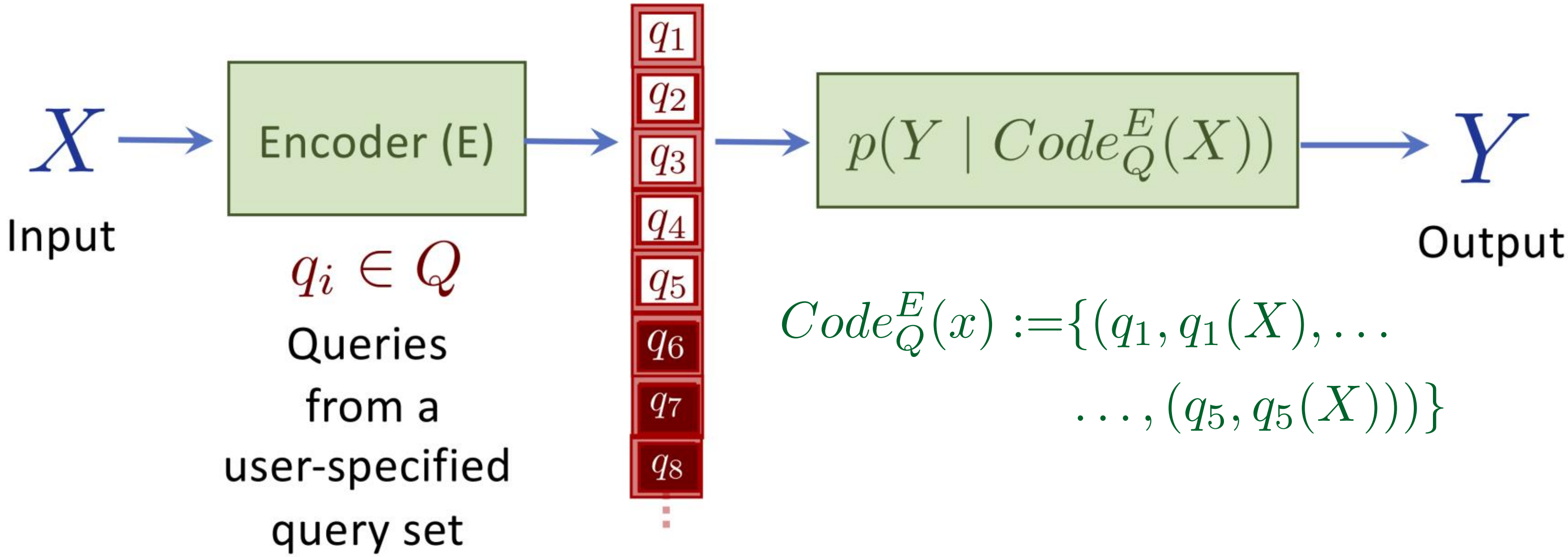
:

Visual Semantic Information



Salient parts of the image

Defintion: Semantic Entropy



Semantic Entropy: minimum number of queries about X from query set Q whose answers are sufficient to predict Y

Minimal: $\min_E \mathbb{E}_X [|Code_Q^E(X)|] =: SE_Q(X;Y)$

Sufficient: s.t. $p(y \mid Code_Q^E(x)) = p(y \mid x) \; \forall x \in \mathcal{X}, y \in \mathcal{Y}$

Approximating Semantic Entropy using IP

- Computing $SE_Q(X;Y)$ is generally intractable.
- **Information Pursuit (IP)**: greedy strategy where the encoder chooses queries sequentially in order of information gain.

Definition: IP Encoder

Queries are chosen according to observed x.

- First query: $q_1 = \arg\max_{q \in Q} I(q(X); Y)$
- Next query: $q_{k+1} = \arg\max_{q \in Q} I(q(X); Y \mid q_{1:k}(x))$
- Termination: $q_{L+1} = q_{STOP}$ if $\max_{q \in Q} I(q(X); Y \mid q_{1:L}(x)) = 0$

$q_{1:k}(x)$ is the event that contains all realizations of X that agree on the first k query-answers for x.

- **Theorem**: If Y is a discrete-valued function of X and Q is the set of all binary queries on X, $SE_Q^{IP}(X;Y) \leq SE(X;Y) + 1$.

Computing Mutual Information is Intractable

- Selecting the first query requires computing $I(q(X);Y)$
  - Need a joint distribution of $q(X)$ and Y.

History

- Later queries require computing $I(q(X);Y \mid q_{1:k}(x))$
  - Need a joint distribution of $(q(X),Y)$ given History.
  - As histories get longer, we run out of samples that match History.

- The above two problems need to be solved $\forall q \in Q$, which scales linearly with the number of queries.

- What do we assume to make computation tractable?

Queries are Independent Given Nuisances

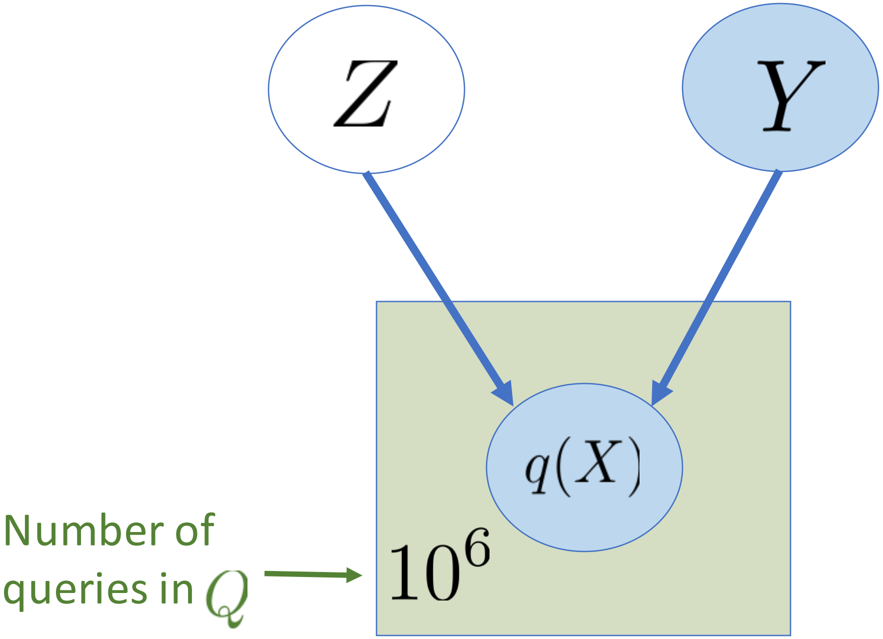
- **Assumption**: query answers are conditionally independent given target variable Y and "some" latent nuisance variable Z

$$p(Q(X), Z, Y) = \prod_q p(q(X) \mid Z, Y) p(Z) p(Y)$$

$$Q(X) = \{q(x) : q \in Q\}$$

- Reasonable assumption unless queries are causally related.

- **Examples**:
  - Z = pose and lighting conditions.
  - Z = phonemes in speech.



Learn a Generative Model for IP

- We learn this joint distribution from data using a VAE.

$$p(Q(X), Z, Y) = \prod_q p(q(X) \mid Z, Y) p(Z) p(Y)$$

- Assuming conditional independence makes estimating

$I(q(X); Y \mid q_{1:k}(x))$ tractable using MCMC sampling.

JHU vision lab

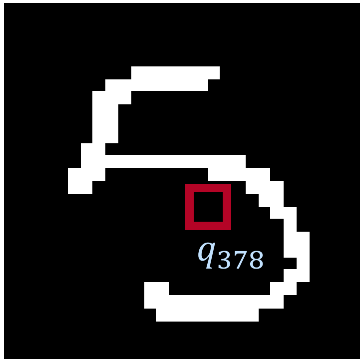
Semantic Information Pursuit for Binary Image Classification

Aditya Chattopadhyay, Donald Geman, Benjamin Haeffele, René Vidal

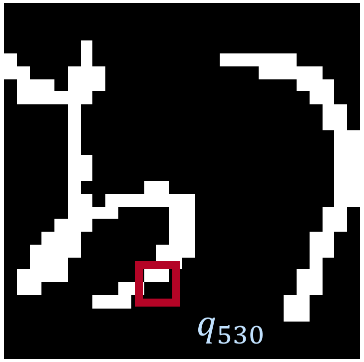
Mathematical Institute for Data Science
Johns Hopkins University

IP for binary image classification

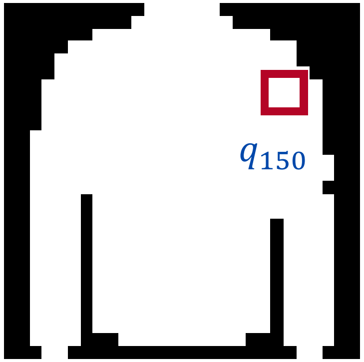
- Task is image classification.
- Queries q_i: "What are the image intensities at the i^{th} patch?"



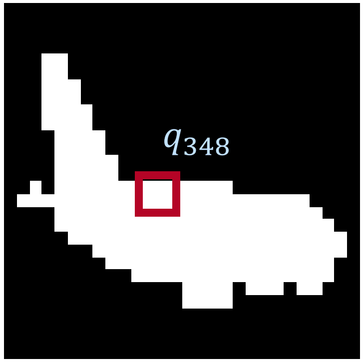
MNIST



KMNIST

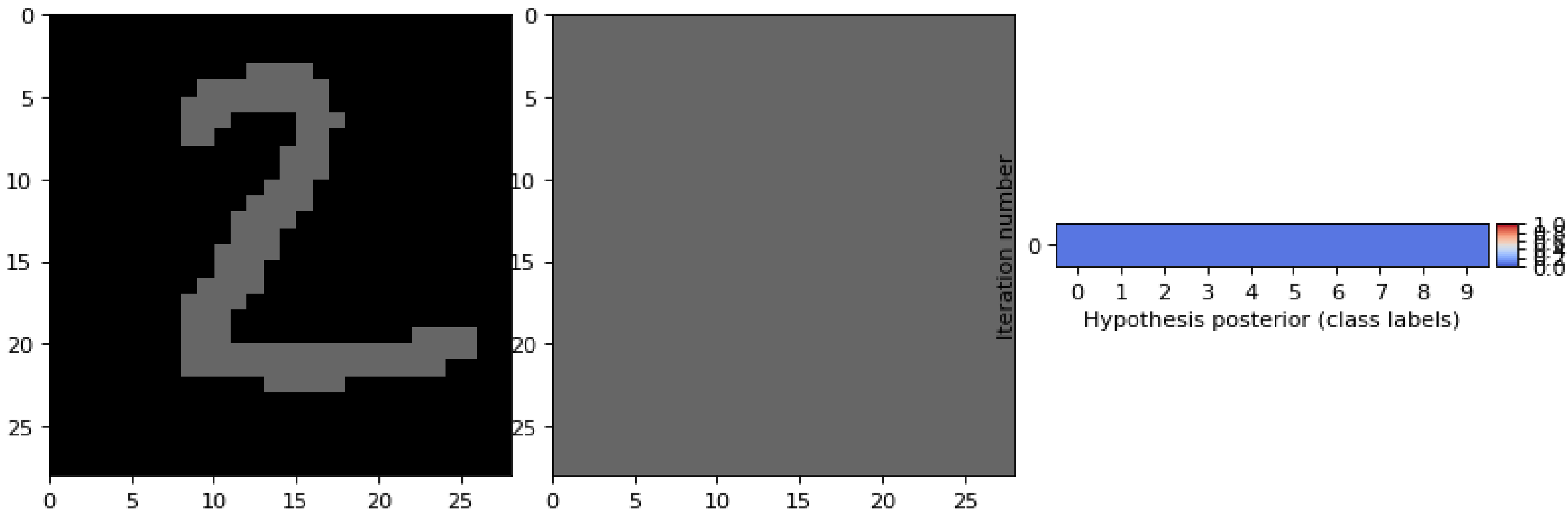


Fashion MNIST

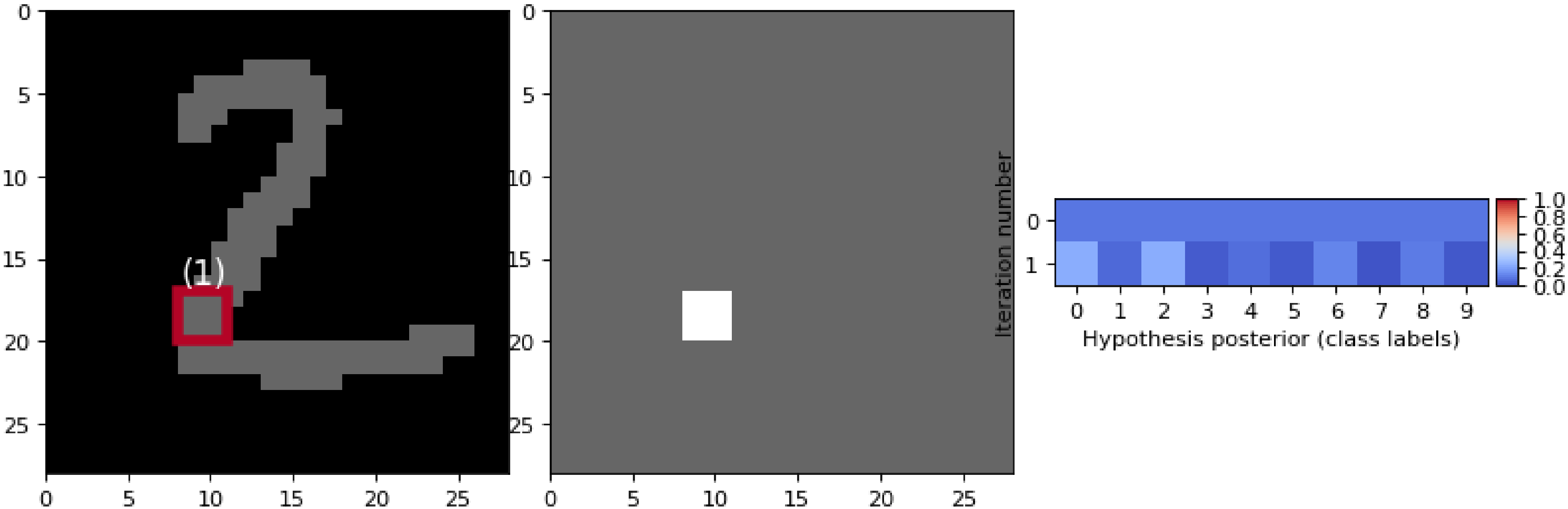


Caltech Silhouettes

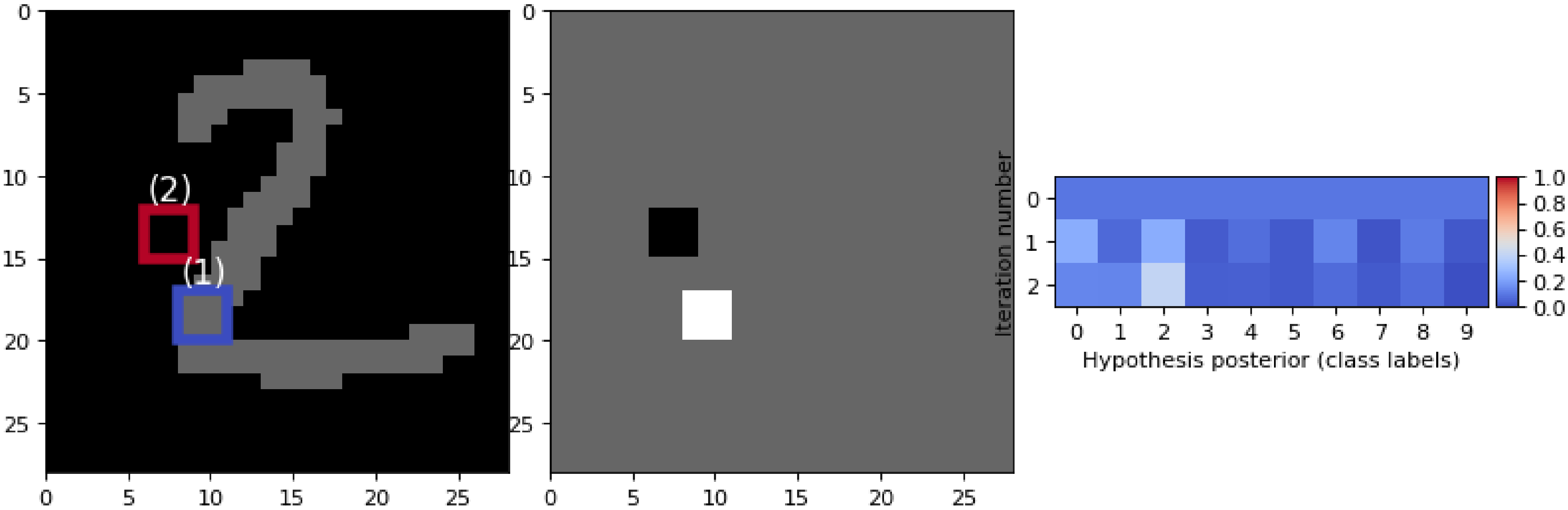
Experiments: IP in action (Iteration 0)



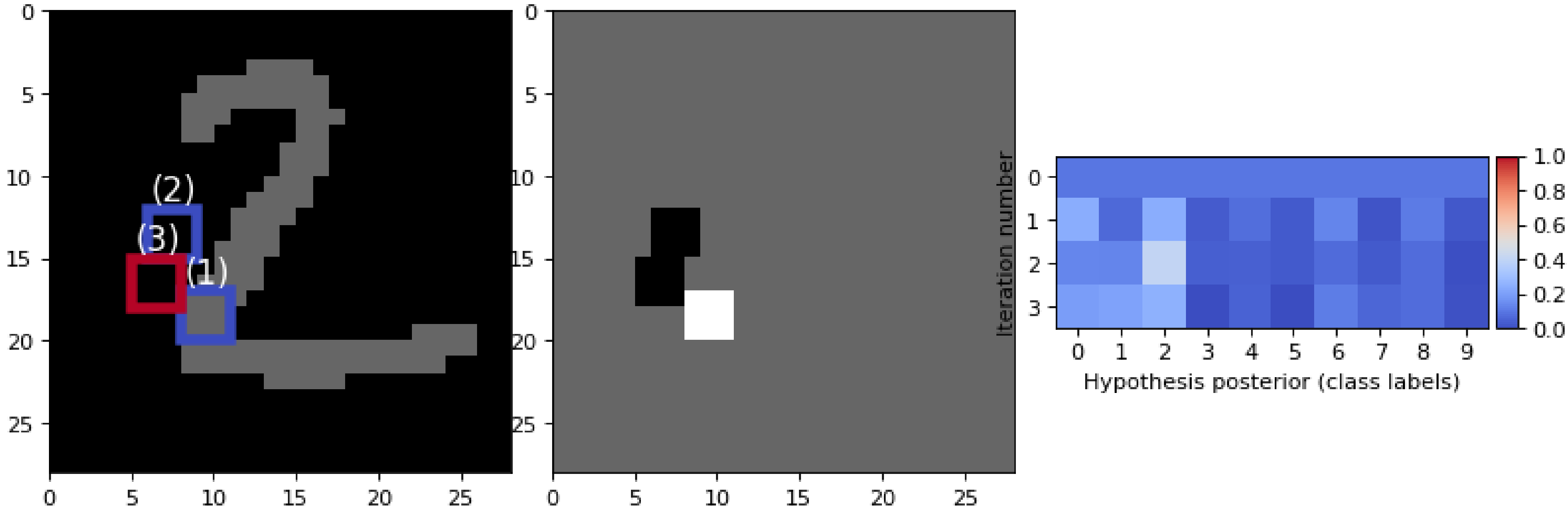
Experiments: IP in action (Iteration 1)



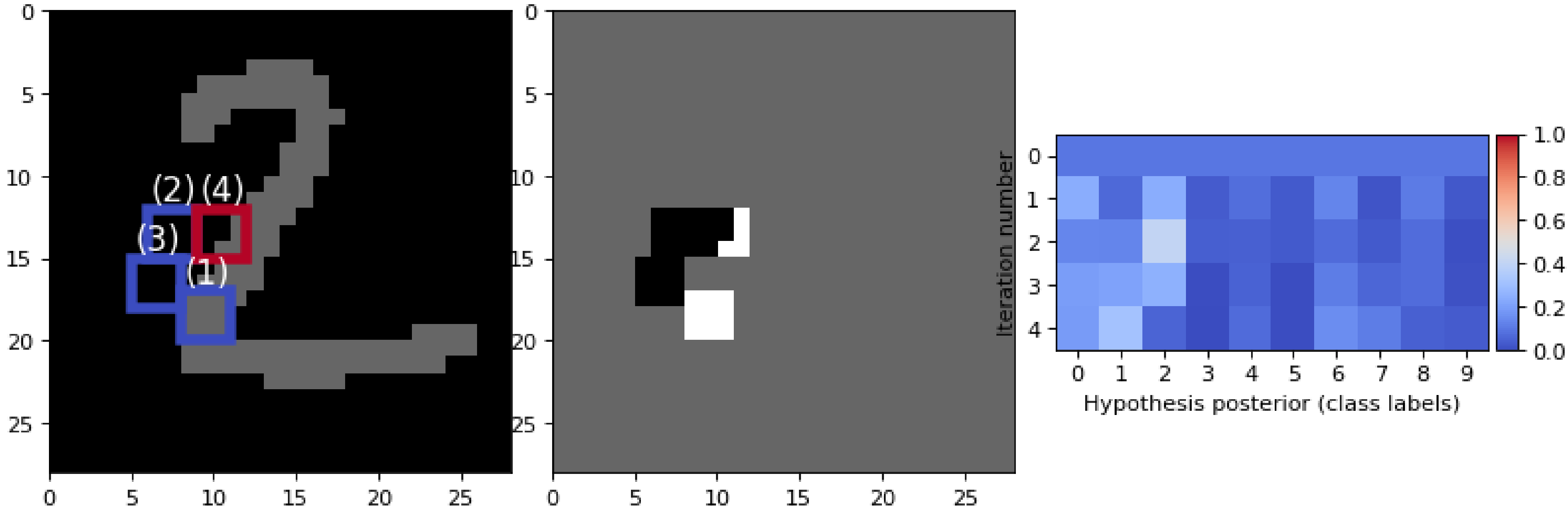
Experiments: IP in action (Iteration 2)



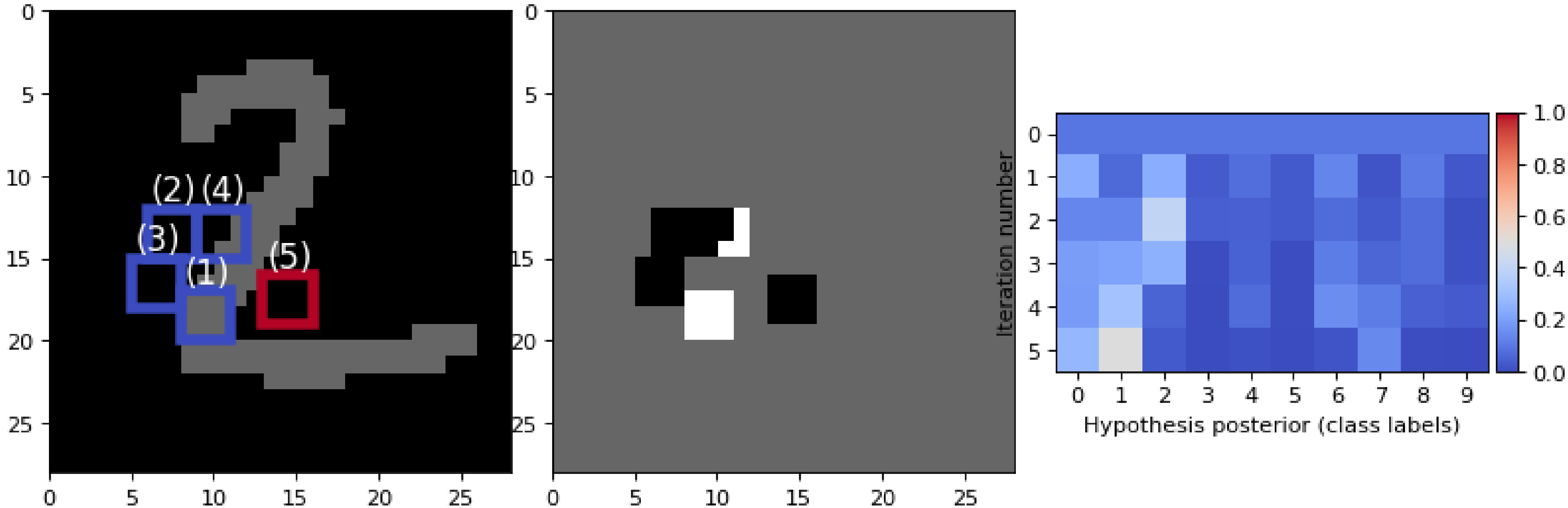
Experiments: IP in action (Iteration 3)



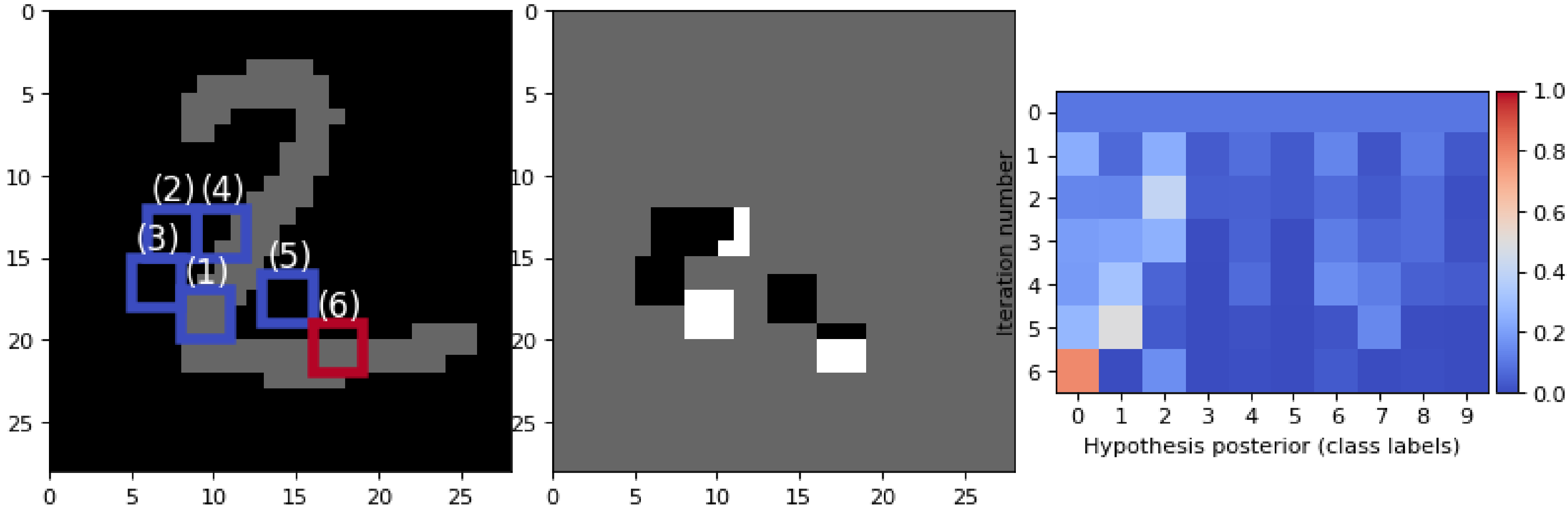
Experiments: IP in action (Iteration 4)



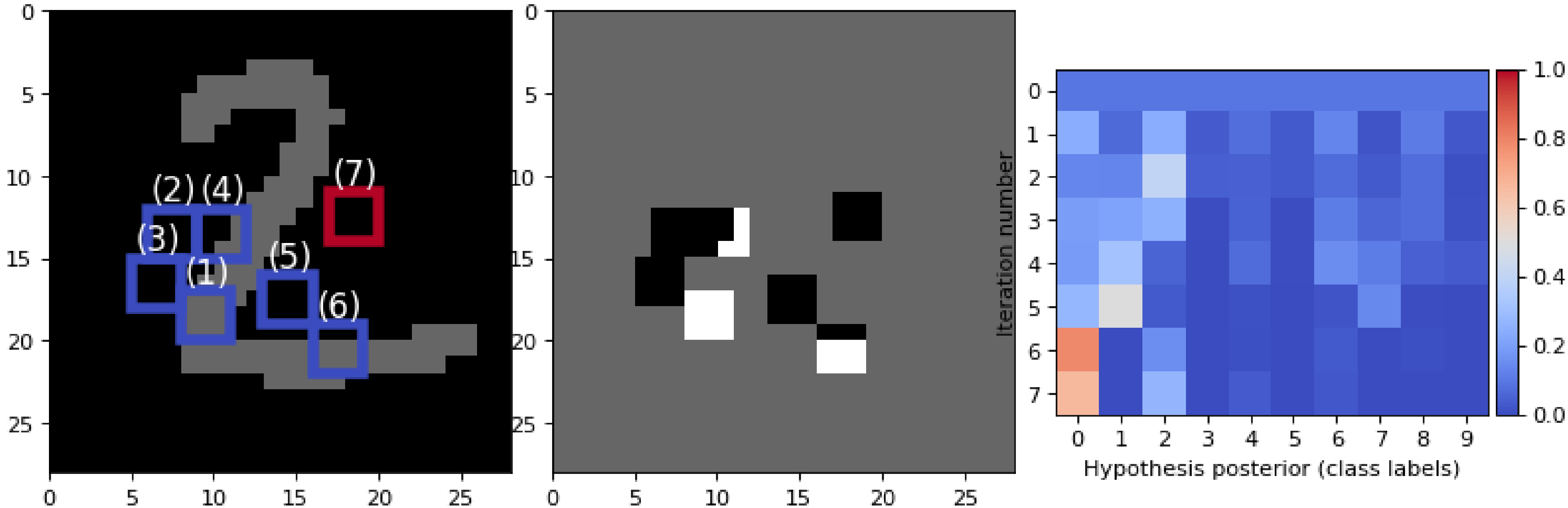
Experiments: IP in action (Iteration 5)



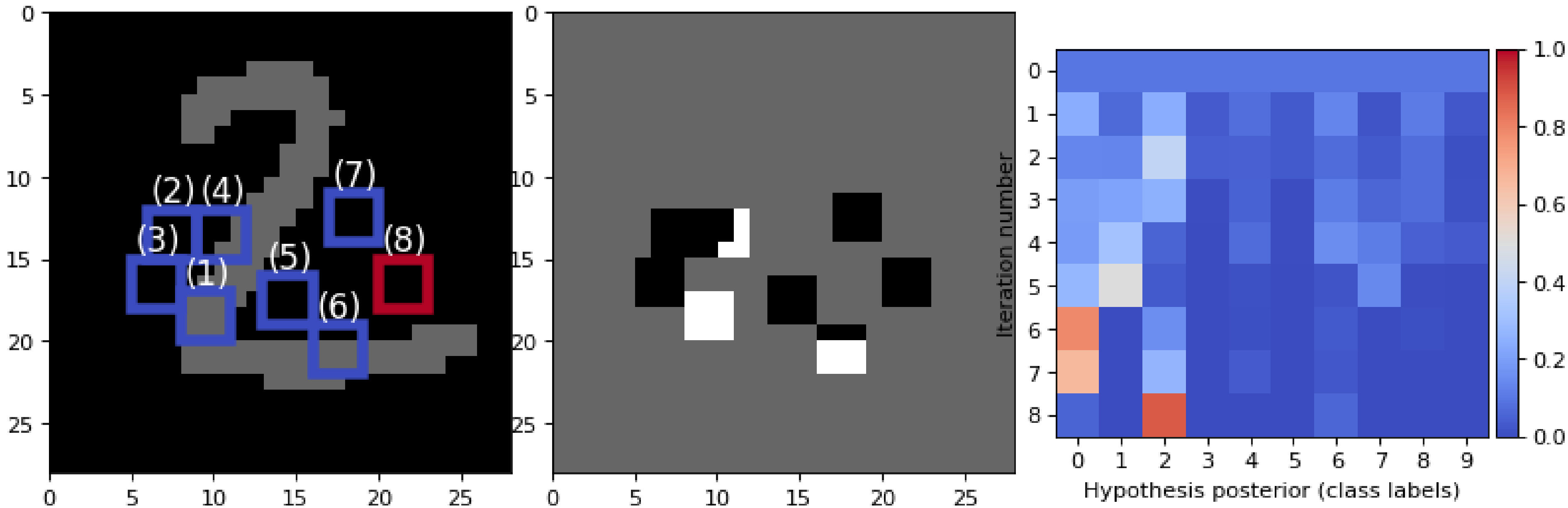
Experiments: IP in action (Iteration 6)



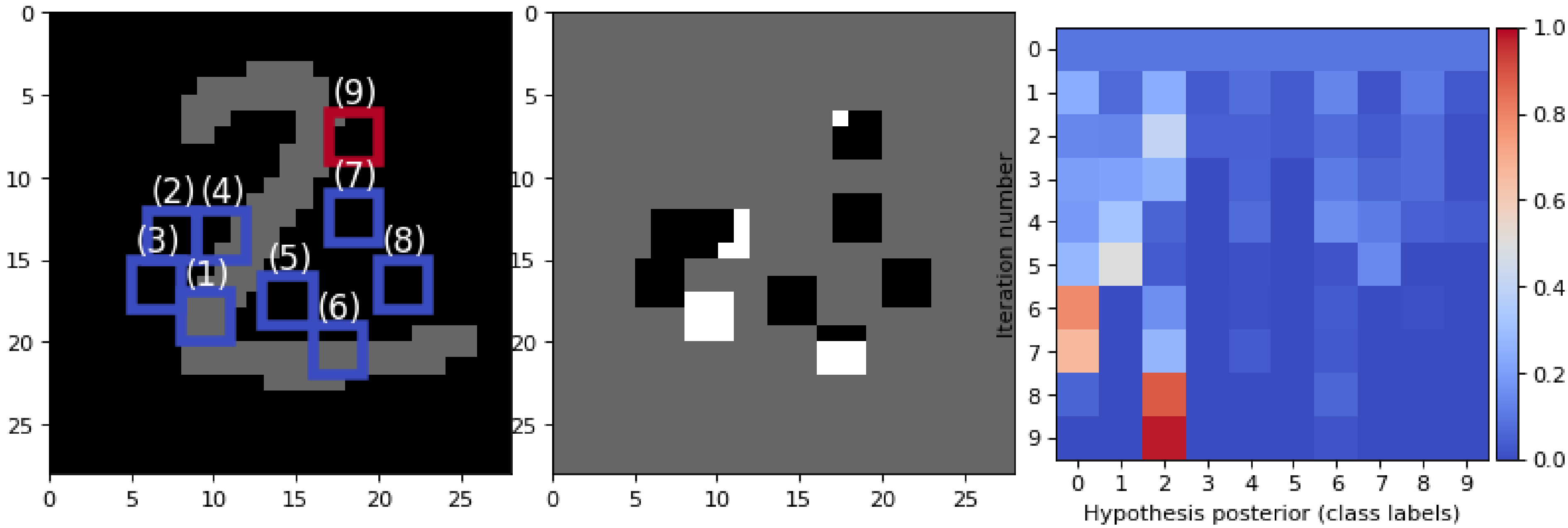
Experiments: IP in action (Iteration 7)



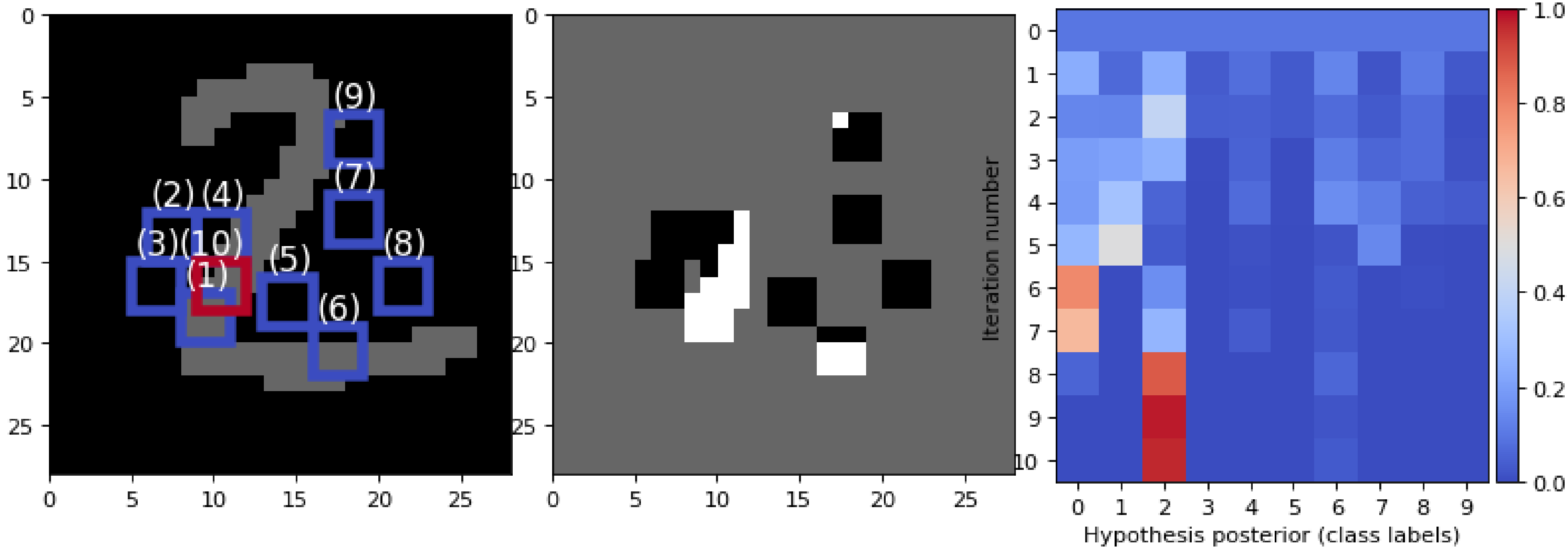
Experiments: IP in action (Iteration 8)



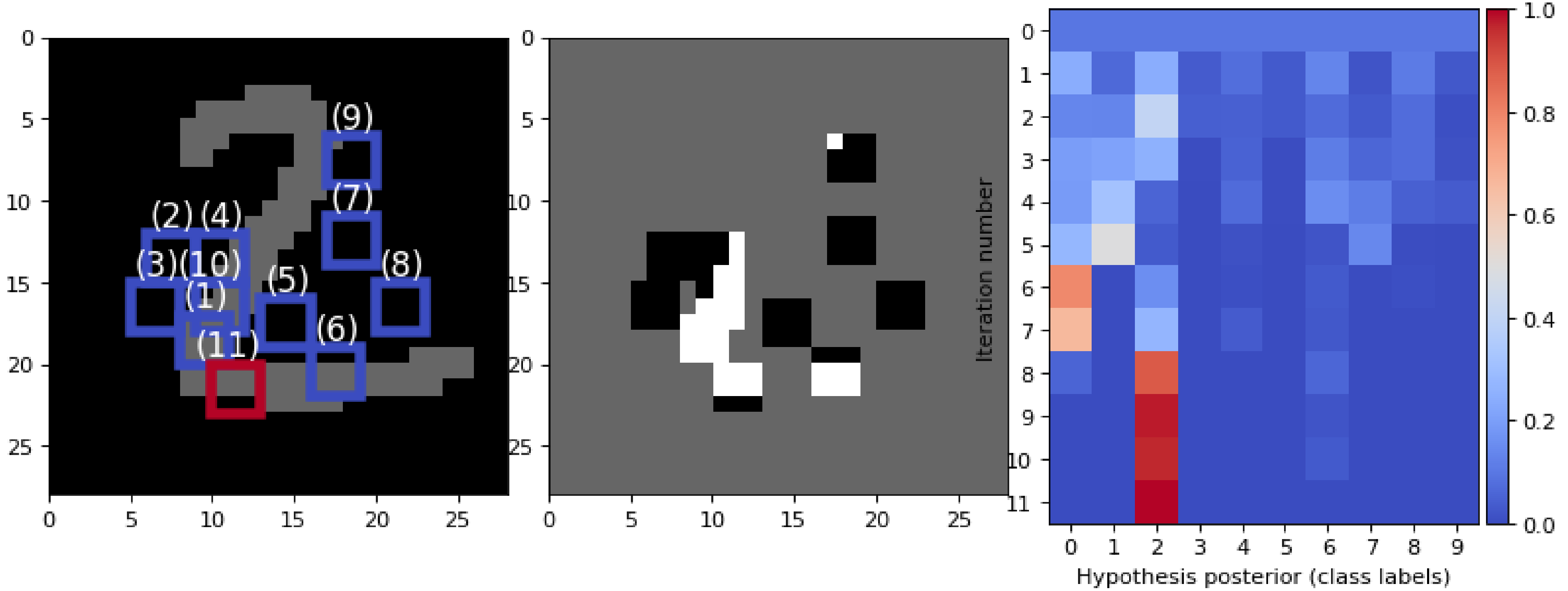
Experiments: IP in action (Iteration 9)



Experiments: IP in action (Iteration 10)



Experiments: IP in action (Iteration 11)



Experiments: Binary Image Classification

- Semantic Entropy correlates with task complexity.

	MNIST	K-MNIST	Fashion-MNIST	Caltech Silhouettes
$SE_Q(X;Y)$ (approx.)	11.54	26.89	40.60	61.58
CNN Test Accuracy	99.15	95.1	86.96	65.15

Figure 2. *The results conform with intuition of more complex datasets having higher semantic entropy. For instance, Caltech Silhouettes, a dataset of binarized images of 101 classes from the Caltech dataset is obviously semantically more complex than handwritten digits in the MNIST dataset.*

JHU vision lab

IP for Interpretable Decision-Making

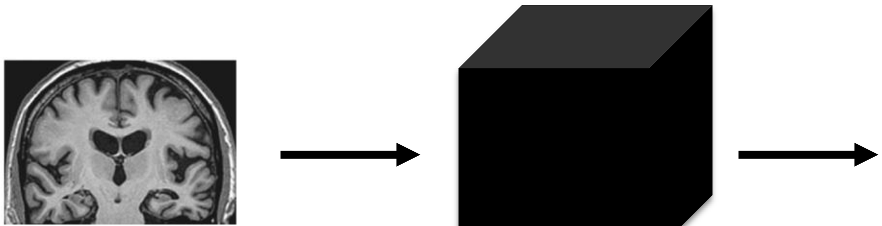
Aditya Chattopadhyay, Stewart Slocum, Benjamin Haeffele, Donald Geman, René Vidal

Annual MURI Review,
December 1-2, 2021

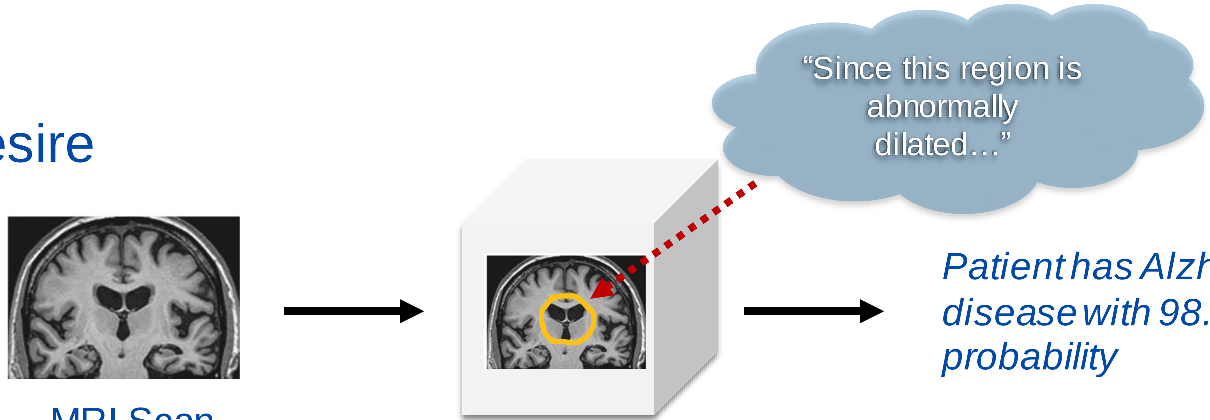
Work supported by the Army Research Office (ARO) under W911NF-17-1-0304. This is part of the collaboration between US DOD, UK MOD and UK EPSRC under the Multidisciplinary University Research Initiative (MURI).

Interpretability Crisis

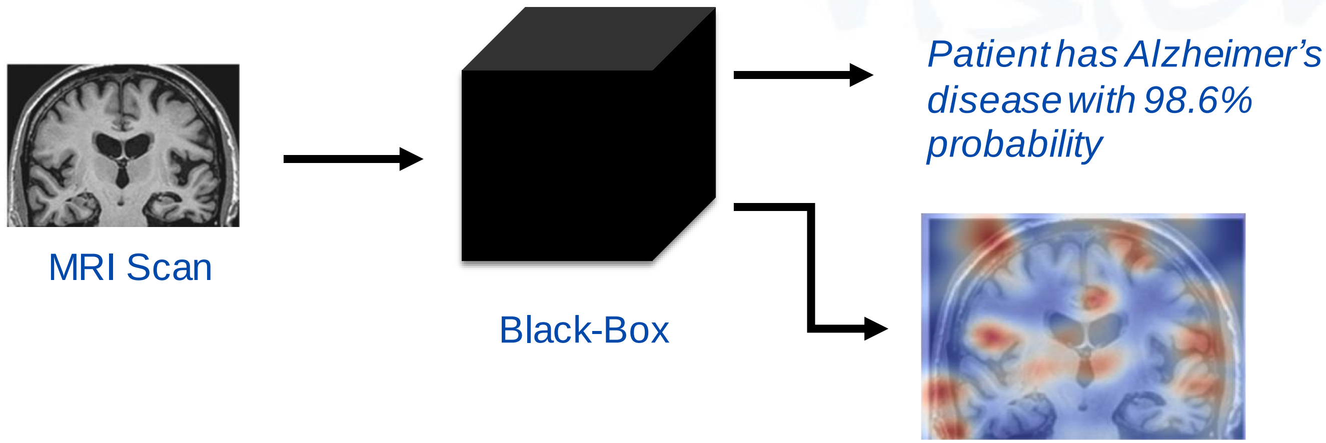
Reality



Desire



Post-Hoc interpretability: The norm



- Current trend is to interpret black-box models post-hoc.
- **The Good:** No need to retrain model, accuracy maintained.
- **The Bad:**
  - Explanations generated are unreliable; not faithful to the model it tries to explain.[1]
  - Salient parts of image might not be most informative to end-users.[2]

1. Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. *Advances in neural information processing systems*, *31*.
2. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215.

Explainable by Design

- Need a new framework for learning that is “explainable by design”.

- ``Explainable'' entails a description in words, symbols or patterns of the reasoning leading to the decision.

- Useful explanations are often domain and task-dependent.

- One way to capture this is by specifying a query set Q.
  - Set of user-defined functions of data, each interpretable to the user.

Explainable by Design (cont.)

- Given Q, efficiently compose queries to explain predictions concisely.

Input image obs

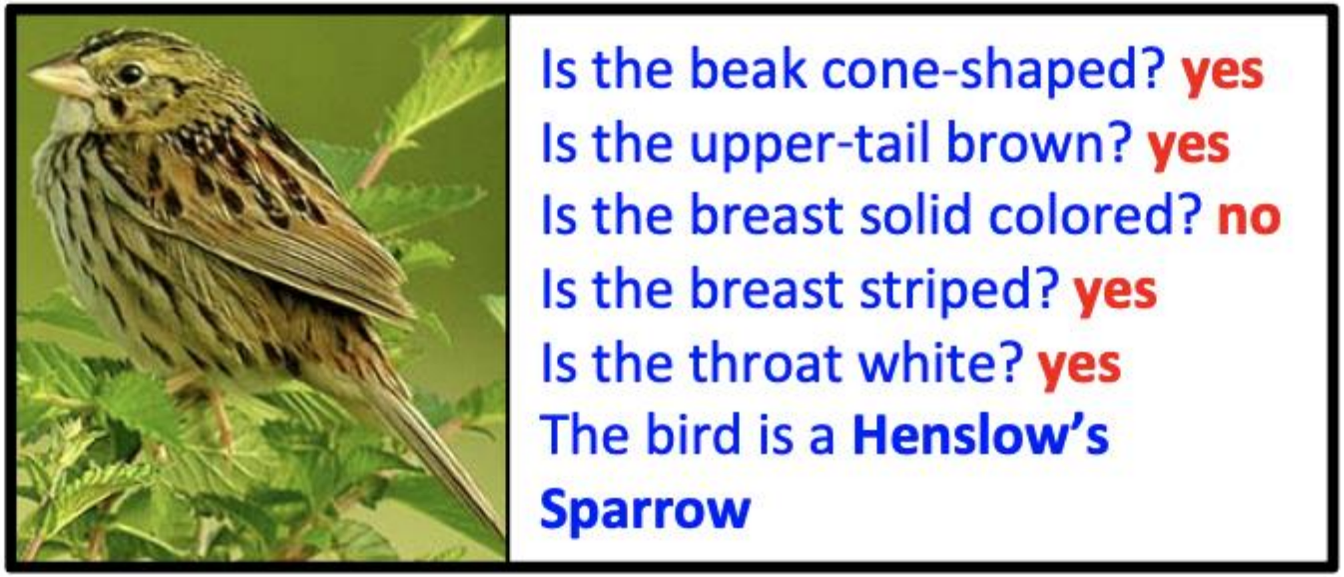
Composing explainable queries

1. Has shape perching-like? **Yes**
2. Has bill shape all-purpose? **Yes**
3. Has belly color yellow? **Yes**
4. Has upperparts color yellow? **No**
5. Has throat color yellow? **No**
6. Has breast color black? **Yes**
7. Has belly color olive? **Yes**

Predicted bird species

Green Jay

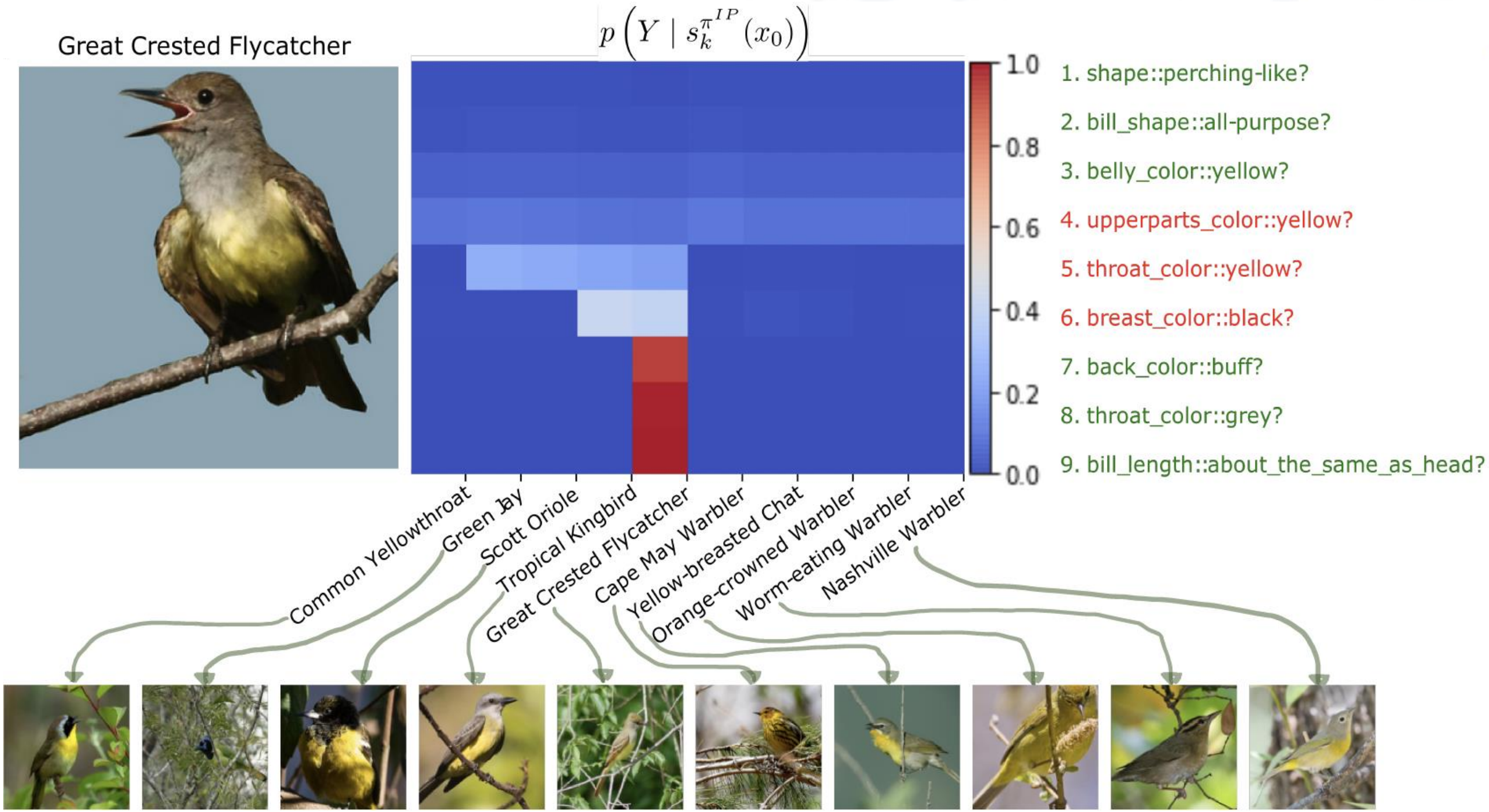
CUB Birds Attribute Dataset



Caltech-UCSD Bird Species Classification

- 11,788 images, 200 bird species
- 312 binary attributes per image
- Difficult, fine-grained task, too hard for non-experts
- **Query set:** One binary query for each attribute.

IP in action

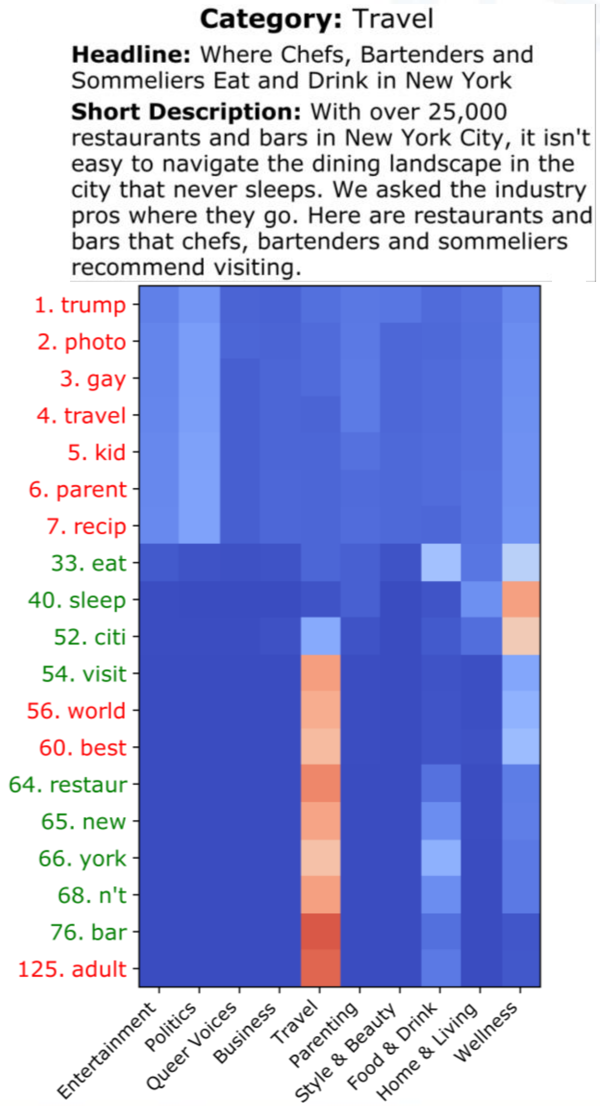


Great Crested Flycatcher
$p\left(Y \mid s_k^{\pi^{IP}}(x_0)\right)$
1.0
0.8
0.6
0.4
0.2
0.0
1. shape::perching-like?
2. bill_shape::all-purpose?
3. belly_color::yellow?
4. upperparts_color::yellow?
5. throat_color::yellow?
6. breast_color::black?
7. back_color::buff?
8. throat_color::grey?
9. bill_length::about_the_same_as_head?
Common Yellowthroat
Green Jay
Scott Oriole
Tropical Kingbird
Great Crested Flycatcher
Cape May Warbler
Yellow-breasted Chat
Orange-crowned Warbler
Worm-eating Warbler
Nashville Warbler

HuffPost dataset

- Dataset: Huffington Post news headlines, $132K$ samples
- Task is to identify topics of newspaper articles from headlines
- Total of 10 topics (e.g. Entertainment, Politics, Food & Drink)
- Given a vocabulary of possible words in the headline, query q_i: “Is the i^{th} word present in the headline?”

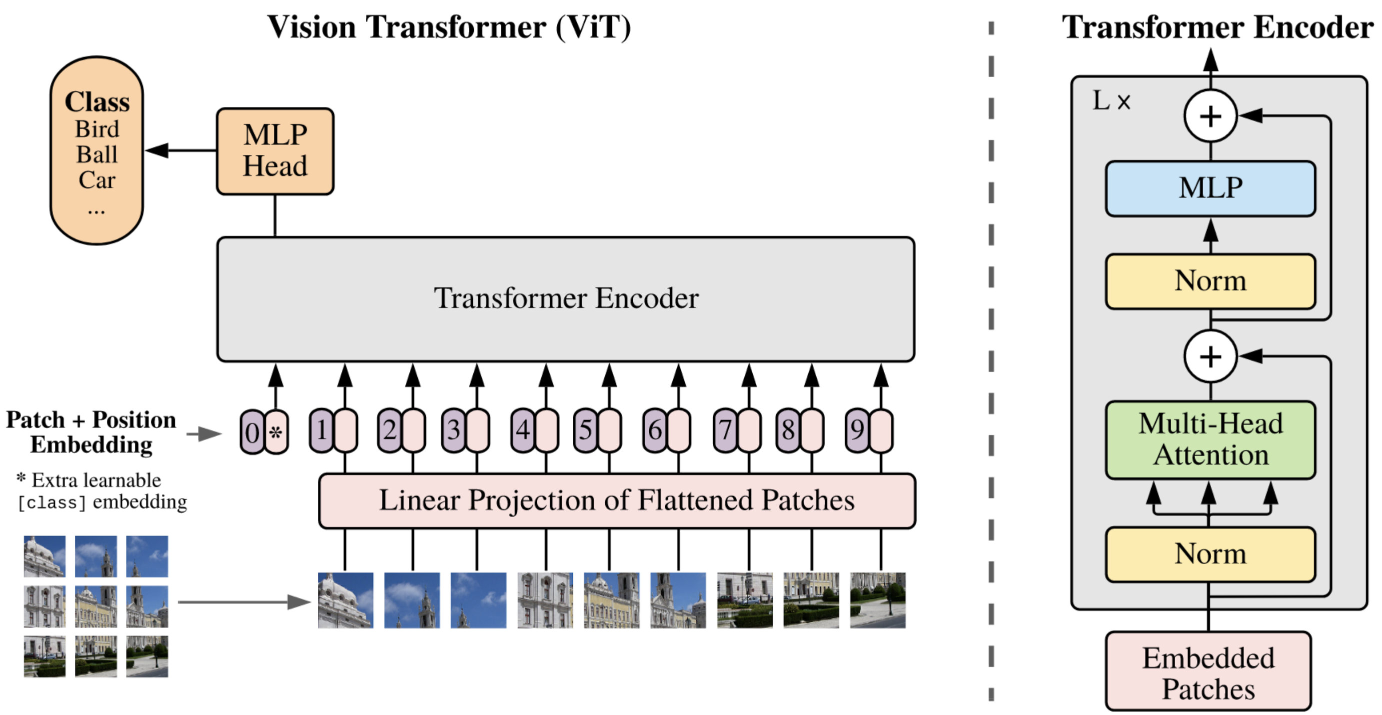
IP in action



Category: Travel
Headline: Where Chefs, Bartenders and Sommeliers Eat and Drink in New York
Short Description: With over 25,000 restaurants and bars in New York City, it isn't easy to navigate the dining landscape in the city that never sleeps. We asked the industry pros where they go. Here are restaurants and bars that chefs, bartenders and sommeliers recommend visiting.
1. trump
2. photo
3. gay
4. travel
5. kid
6. parent
7. recip
33. eat
40. sleep
52. citi
54. visit
56. world
60. best
64. restaur
65. new
66. york
68. n't
76. bar
125. adult
Entertainment
Politics
Queer Voices
Business
Travel
Parenting
Style & Beauty
Food & Drink
Home & Living
Wellness

Future Directions

- Explore more scalable algorithms.

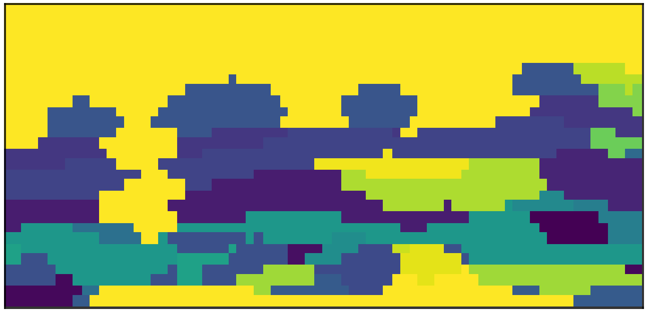


Future Directions

- Explore more semantic tasks

Original Image

Segmentation Label



IP Unsupervised Segmentation

More Information,

Research supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DOI/IBC) contract number D17PC00345.

Vision Lab @ JHU
http://www.vision.jhu.edu

Center for Imaging Science @ JHU
http://www.cis.jhu.edu

Mathematical Institute for Data Science @ JHU
http://www.minds.jhu.edu

Thank You!

JOHNS HOPKINS
MATHEMATICAL INSTITUTE
for DATA SCIENCE