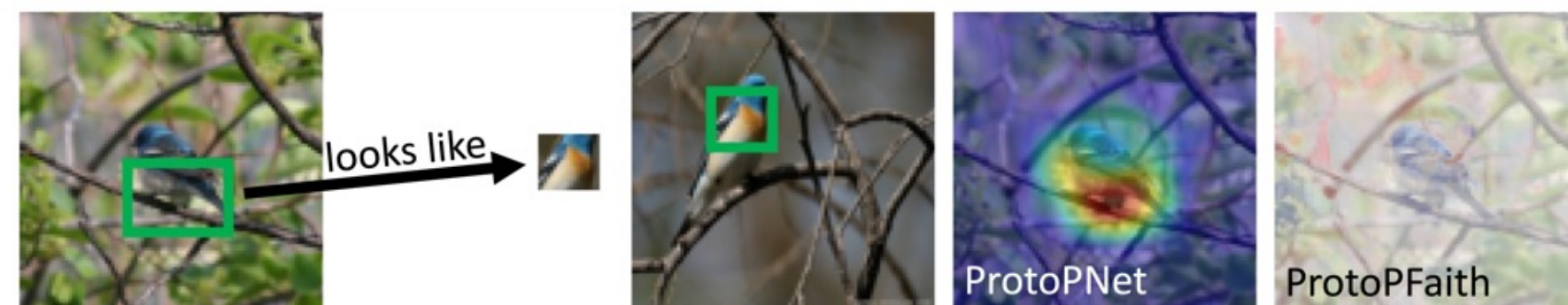


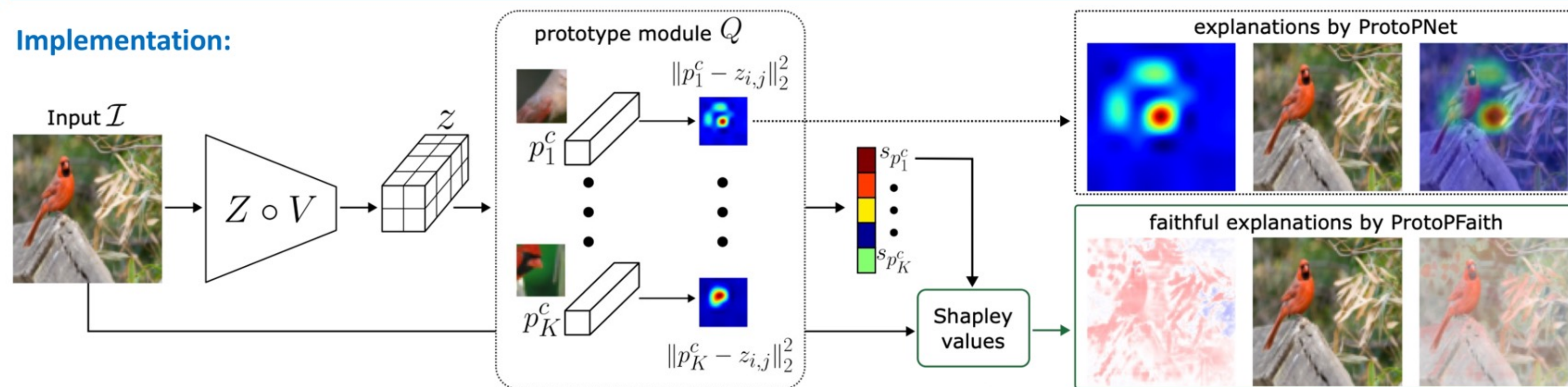
Keep The Faith:

Faithful Explanations in Convolutional Neural Networks for Case-Based Reasoning

Case-Based Reasoning:



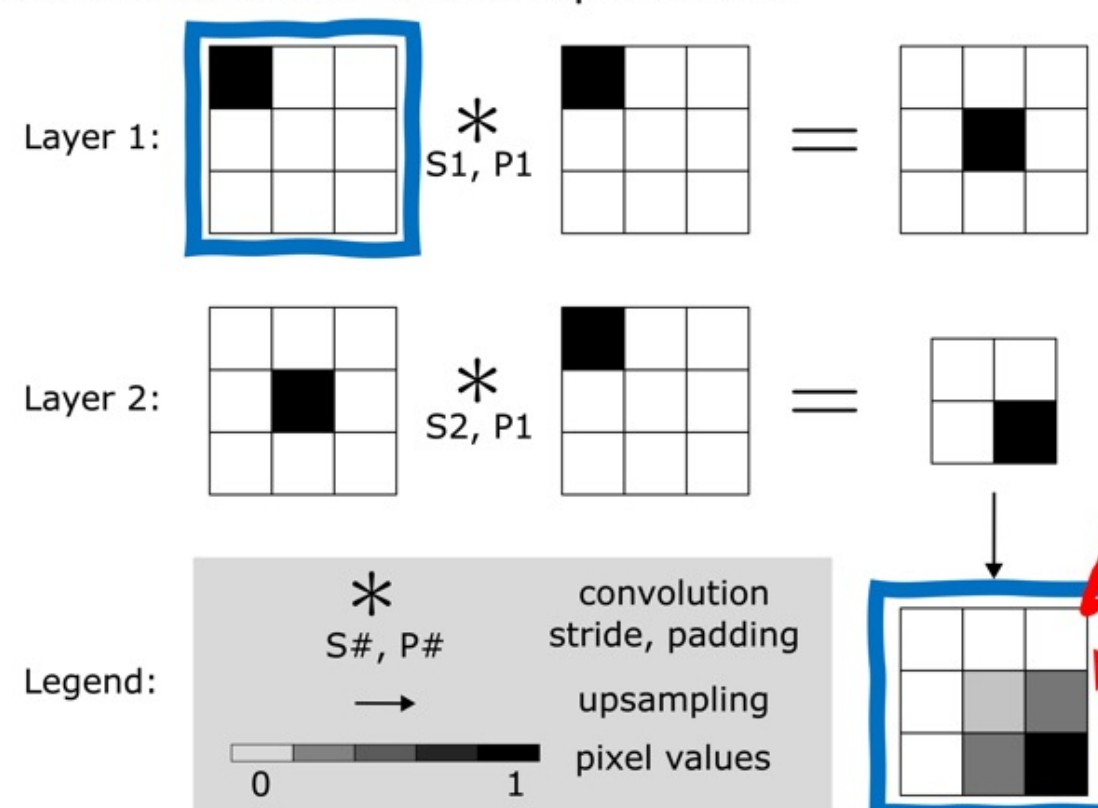
Implementation:



XAI unveiled

Motivation:

- Assumption in ProtoPNet-like [1] architectures: spatial dependency between latent features and input domain



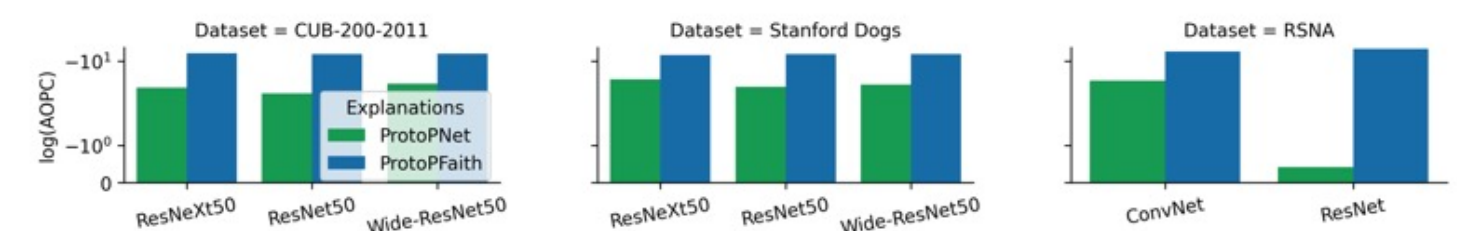
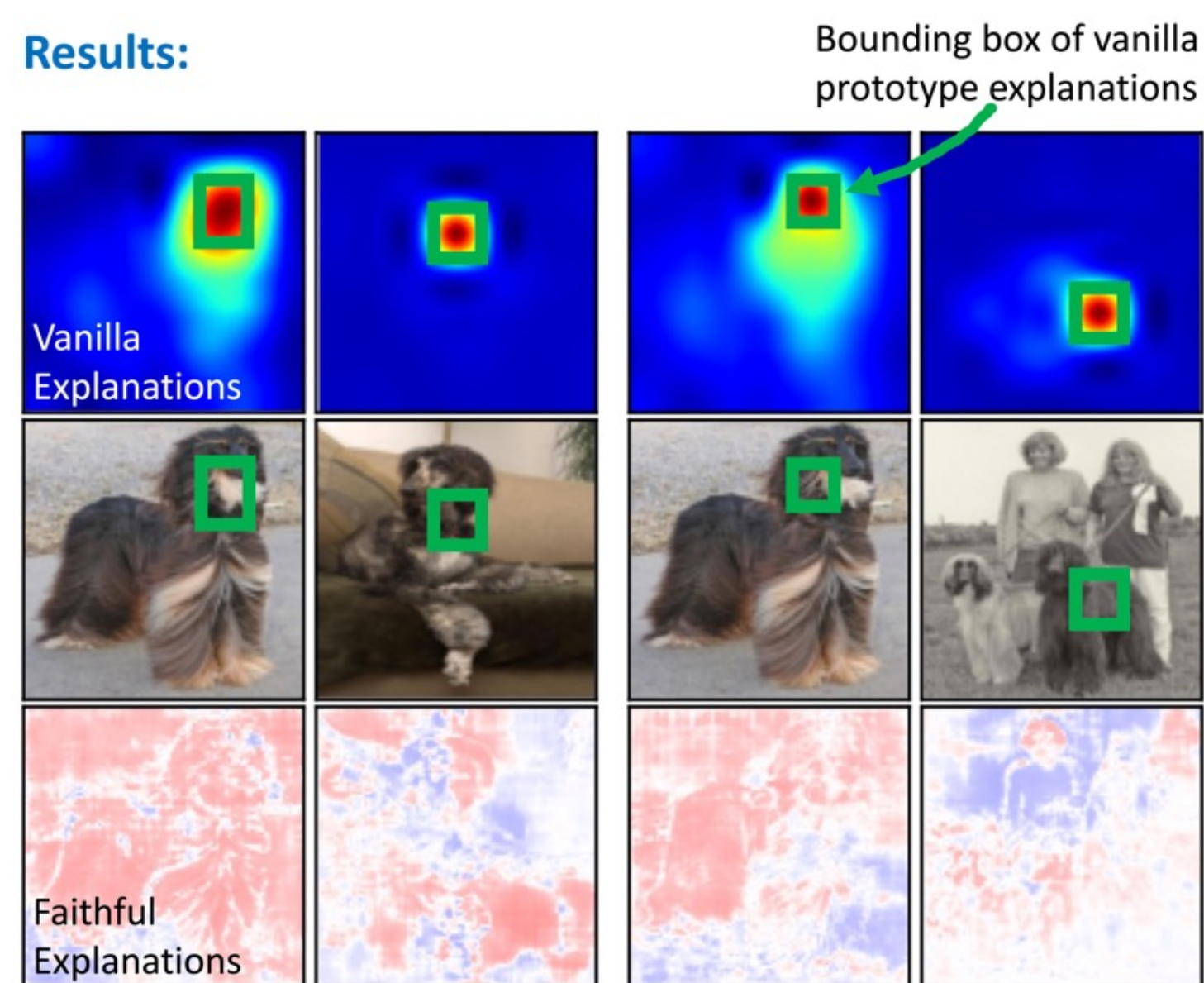
Method:

- Convert *trained* ProtoPNet into Lightweight Probabilistic Neural Network [2]
- Extract explanations following DASP [3] over similarity scores s
- Explanations are based on Shapley values, which satisfy all axioms that we define to be required for *faithfulness*
- Extraction of vanilla explanations still possible for the same model

Requirements:

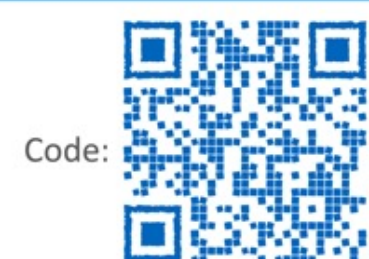
- Closed-form solution for propagation of normal distributions through *all* layers

Results:



Key Takeaways:

- Theoretical violations manifest in experimental results
- Findings generalize to other implementations of case-based reasoning, e.g. ProtoTrees [4] and XProtoNet [5]
- Faithful explanations are difficult to interpret



Solutions now available