AISE: Adaptive Input Sampling for Explanation of Black-box Models

Evgeny Tsykunov, Wonju Lee, Minje Park
Intel Corporation

Introduction

AISE formulates saliency map generation as a kernel density estimation (KDE) problem, and adaptively sample input masks using a derivative-free optimizer to maximize mask saliency score. This adaptive sampling mechanism significantly improves the efficiency of input mask generation and thus increases convergence speed.

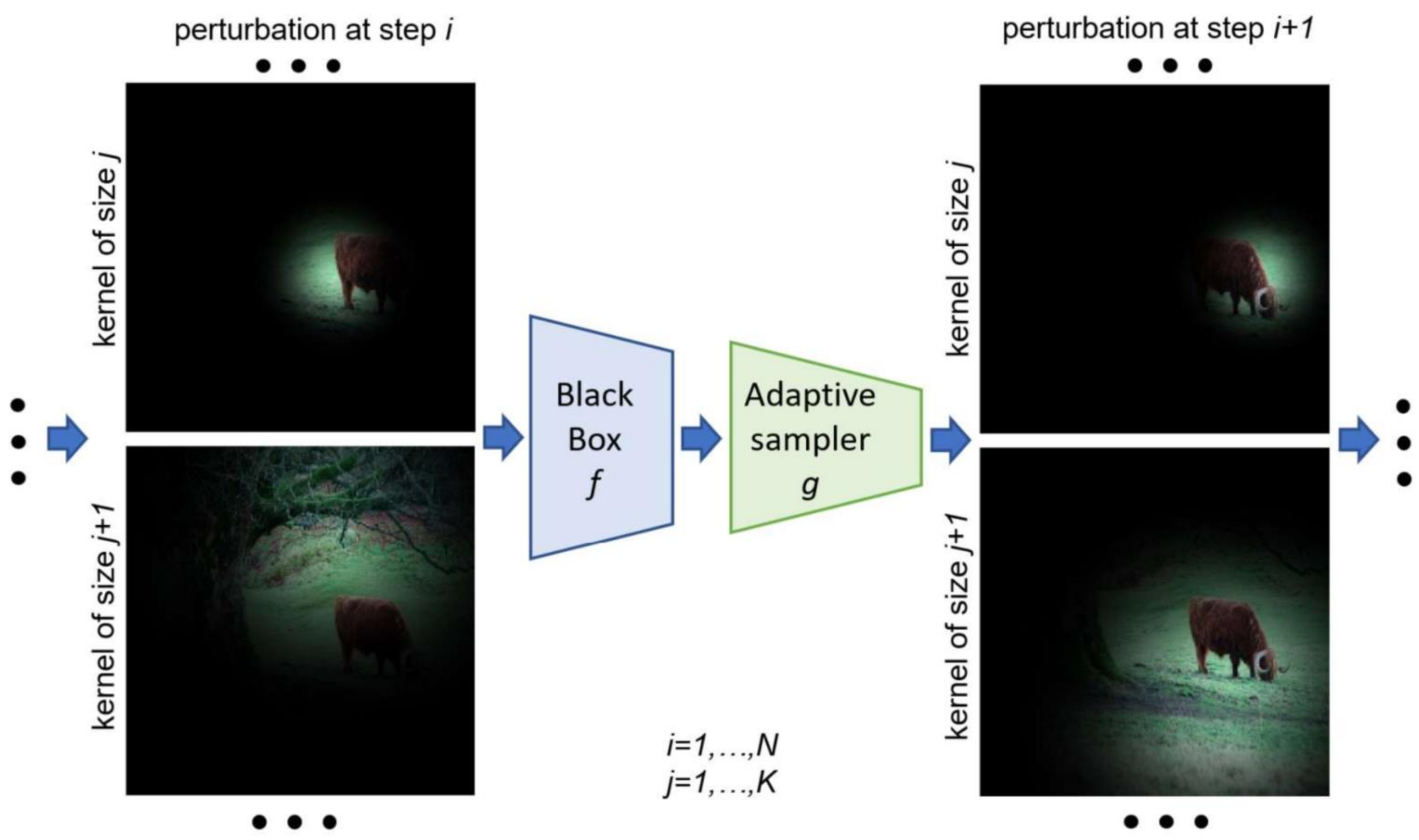
Adaptively sampled kernel locations (red points) guided by the global optimizer and their kernel density estimation (KDE) output. For illustrative purposes, we use 100 samples with a fixed kernel width.

Saliency map generation

We formulate saliency map generation as a kernel density estimation (KDE) problem with a gradient-free adaptive sampling of mask.
KDE requires a good and representative set of samples (addressed by global optimizer) with a proper choice of its kernel width (addressed by multiple widths).

Adaptive sampling

Saliency map landscape can take an arbitrary form with many local optima. we use global (gradient-free) optimization methods that sample representative regions of the input image using mask while optimizing (maximizing) the score.

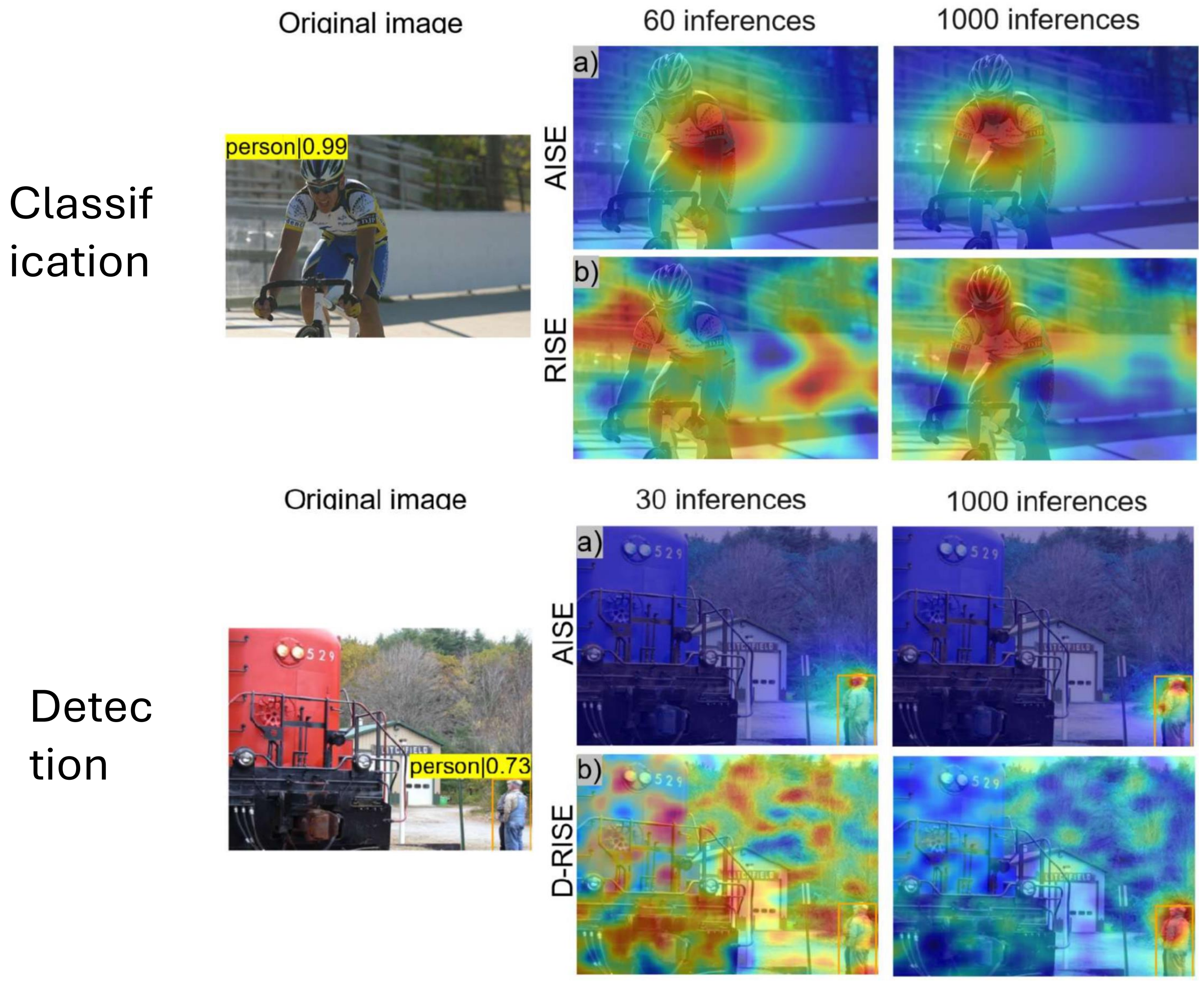


Kernel density estimation

Saliency map is formulated as a non-parametric probability distribution of importance (saliency) among features. We propose to use KDE for the estimation of probability density function – saliency map. We propose simultaneously perturbing the image with kernels of different widths and then aggregating the results

Qualitative comparison

Comparison between random and adaptive mask sampling. In many cases AISE can generate a decent saliency map even after 30 iterations. RISE converges much slower and requires thousands of iterations to reach the same quality.



Quantitative comparison

Localization metric. AISE, having significantly less amount of compute, achieves and in most cases overcomes popular state-of-the-art RISE [1] and others.

	Pointing game			
	VOC07		COCO14	
Method	VGG16	ResNet50	VGG16	ResNet50
Gradient-based				
Gradient	86.0	78.0	54.4	46.2
Guided Backprop.	88.2	85.3	57.4	56.4
GradCAM	94.5	**95.2**	**70.4**	**72.7**
Extremal Perturb.	**95.1**	93.0	66.4	66.6
Gradient-free				
RISE-8000	94.8	90.6	**68.3**	68.4
SHAP-8000	92.4	93.7	64.7	**69.5**
AISE-LIPO-300	95.0	93.6	66.9	66.5
AISE-DIRECT-300	94.9	94.3	66.8	66.8
AISE-LIPO-700	94.6	93.8	67.3	68.8
AISE-DIRECT-700	**95.3**	**93.9**	68.0	68.7

Conclusion

AISE is designed to be task-agnostic and can be applied to a wide range of classification and object detection architectures. We show that AISE achieves state-of-the-art results, having just 300 model inferences or less.

References

1. Vitali Petsiuk, Abir Das, and Kate Saenko. RISE: Randomized input sampling for explanation of black-box models. 2018.