Towards Generating Human-Centered Saliency Maps without Sacrificing Accuracy

Vivek Aswal*, Gore Kao*, Seo Young Kim*, and Katelyn Morrison*
Carnegie Mellon University

NeuroVision Workshop 2022

* Authors contributed equally

Research Question:
What techniques will cause SOTA object detection models to generate saliency maps that are similar to human eye-fixations while maintaining SOTA accuracy?
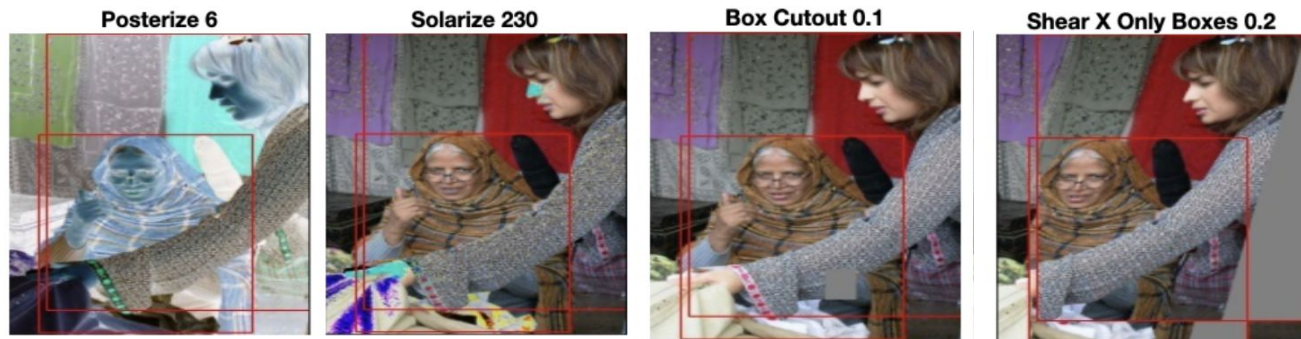
Proposed Solution:
Data Augmentation
- Selective Erasing
- Selective Inpainting
- Non-trivial transformation

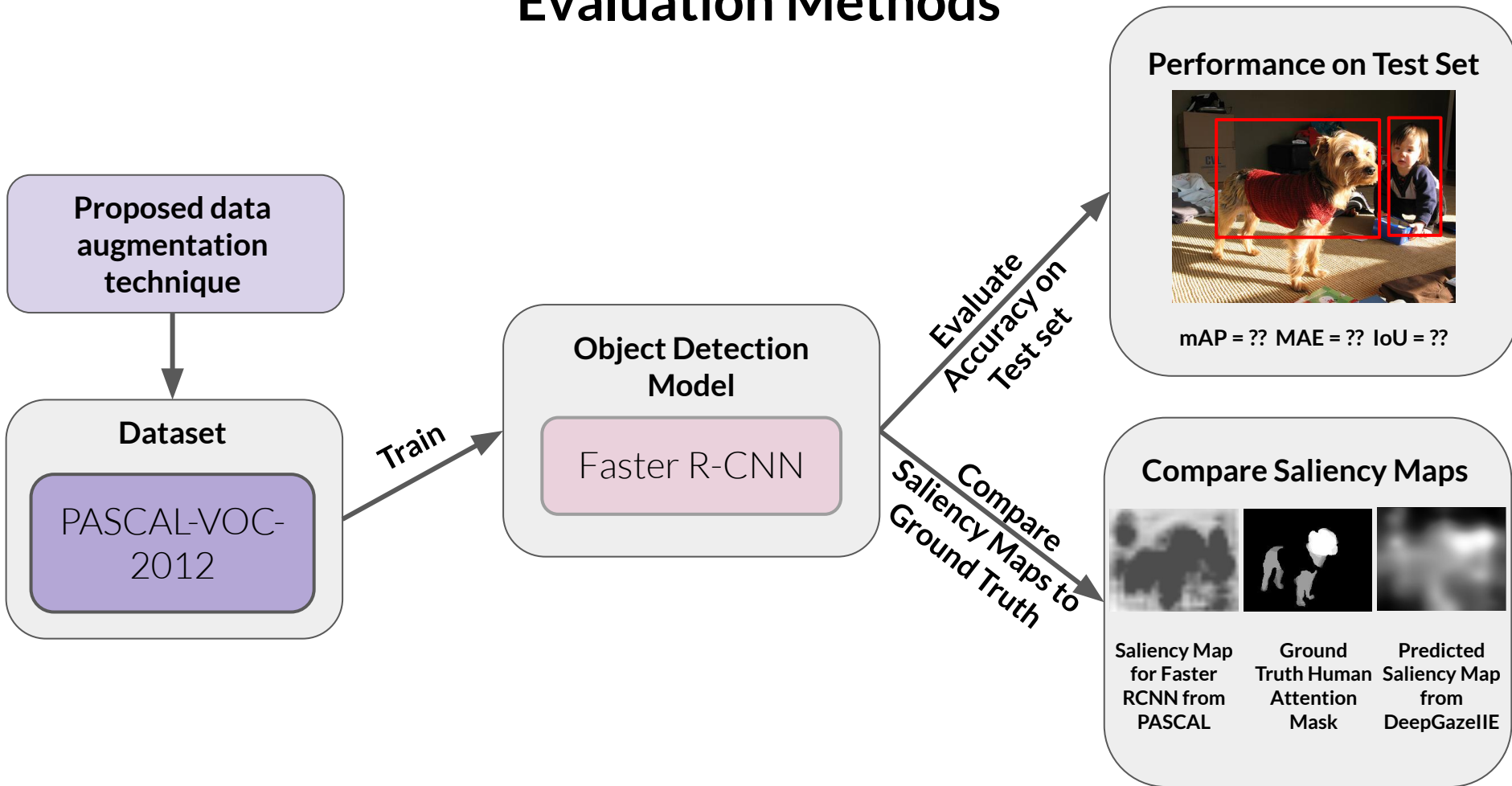# Augmented Dataset Examples

Selective Erasing & Inpainting

Original   Faster R-CNN Saliency Map   DeepGazeIIE Saliency Map   Selectively Erased

Non-trivial Transformations

Posterize 6   Solarize 230   Box Cutout 0.1   Shear X Only Boxes 0.2

Evaluation Methods

Proposed data augmentation technique

Dataset

PASCAL-VOC-2012

Train

Object Detection Model

Faster R-CNN

Evaluate Accuracy on Test set

Compare Saliency Maps to Ground Truth

Performance on Test Set

mAP = ?? MAE = ?? IoU = ??

Compare Saliency Maps

Saliency Map for Faster RCNN from PASCAL

Ground Truth Human Attention Mask

Predicted Saliency Map from DeepGazeIIE

Results

Augmentation	mAP	Compared to Predicted Eye-Fixations		Compared to Human Attention Masks	
		mAE	IoU	mAE	IoU
Selective Erasing	0.754	0.1560	**0.1878**	**0.1561**	0.1878
Selective Inpainting	0.763	**0.1552**	0.1863	0.1572	0.1863
Non-Trivial Transformation	0.781	0.1581	0.1762	0.1600	0.2676
Original	**0.787**	0.1575	0.1823	0.1583	**0.2688**

Towards Generating Human-Centered Saliency Maps without Sacrificing Accuracy

Vivek Aswal*, Gore Kao*, Seo Young Kim*, and Katelyn Morrison*
Carnegie Mellon University

NeuroVision Workshop 2022

* Authors contributed equally