FEW-SHOT SALIENCY

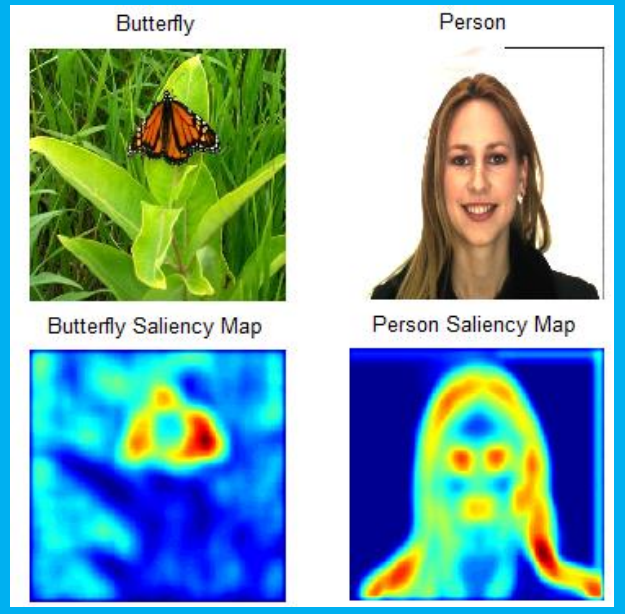
Albert Ding

What is Saliency?

- Computer vision task to generate saliency map
- Finds visually distinctive areas to human eyes
- Ground truth can be generated by an eye tracker

Examples of Saliency

Binary Saliency Map



Regular Saliency Map

Few-Shot Learning

- Theory of human-like learning based on information distance metric conditioned on a set of unlabelled samples.

- Implemented by hierarchical VAE for image classification.

- Bits back paper explains how to use a VAE to compress

Framework Visualization

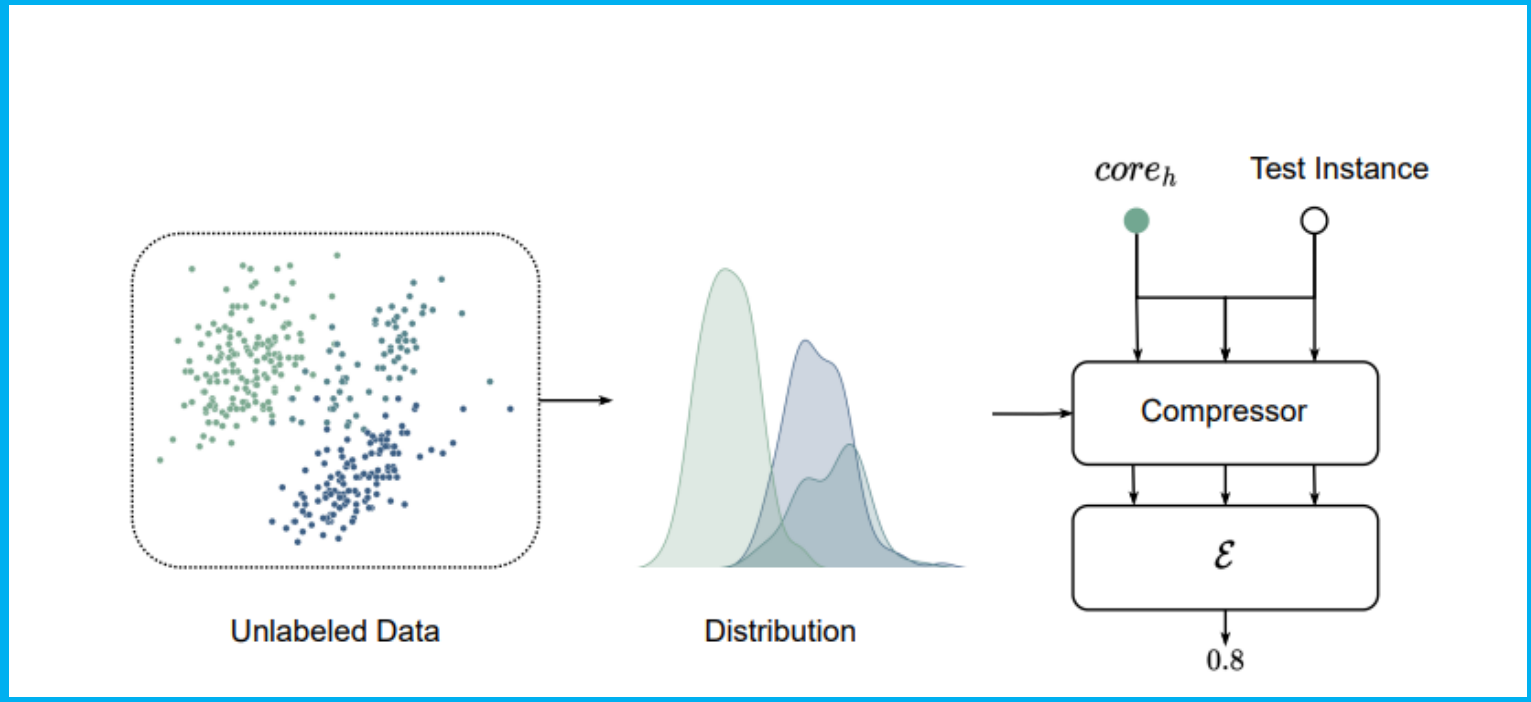


Image from Jiang, et al., A Theory of Human-Like Few-Shot Learning, (January, 2023)

Motivation

- Ground truth maps are expensive for the saliency task (pixel-level labelling)
- Compare with image classification datasets (image-level labelling)
- Training done with only a few labelled samples (few-shot learning) becomes desirable

Motivation

	MNIST	KMNIST	FashionMNIST	STL-10	CIFAR-10
SVM	69.4±2.2	40.3±3.6	67.1±2.1	21.3±2.8	21.1±1.9
CNN	72.4±3.5	41.2±1.9	67.4±1.9	24.8±1.5	23.4±2.9
VGG	69.4±5.7	36.4±4.7	62.8±4.1	20.6±2.0	22.2±1.6
ViT (disc)	58.8±4.6	35.8±4.1	61.5±2.2	24.2±2.5	22.3±1.8
Latent	73.6±3.1	48.1±3.3	69.5±3.5	31.5±3.7	22.2±1.6
Ours	77.6±0.4	55.4±4.3	74.1±3.2	39.6±3.1	35.3±2.9

Table 1: 5-shot image classification accuracy on five datasets.

Classification accuracy improvement results reported from the paper

Main Idea

- Attempt to produce saliency maps with only a few fully-labelled pixel-level ground truth training samples

- Approach: Follow the framework described in the paper. Approximate the information distance metric as well as use a set of unlabelled examples

Related Work

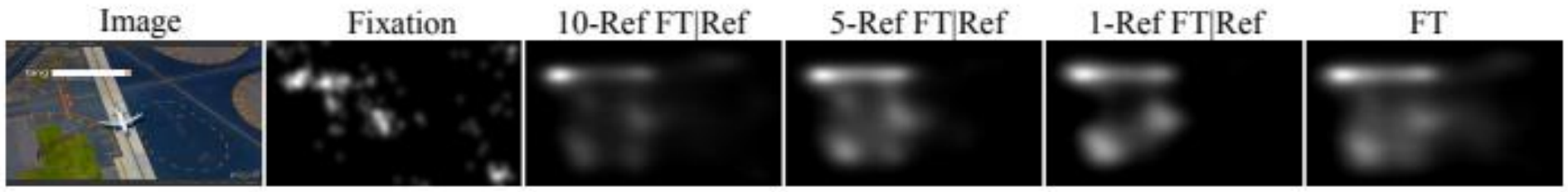
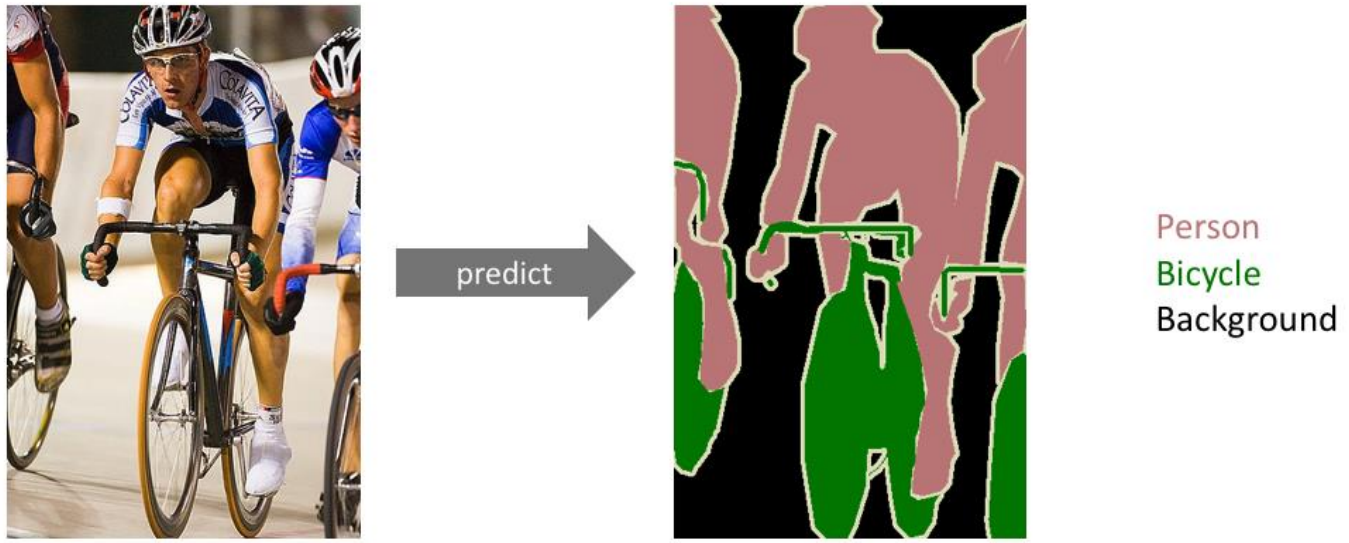


Figure 7: Qualitative results w.r.t. different n

Y. Luo, et al., n-Reference Transfer Learning for Saliency Prediction, (July, 2020)

- Describes a method using a pre-trained network's knowledge of saliency prediction

Possible Extension



Can the proposed framework from the paper also be applied to the computer vision task of semantic segmentation?