Learning Discriminative Representations to Interpret Image Recognition Models Thèse de Doctorat

Felipe Torres Figueroa

École Centrale de Marseille

Laboratoire d'Informatique et de Systèmes (LIS)

Marseille, September 2024







Felipe Torres Figueroa ECM

- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



Introduction

- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



Introduction

00000

My go to exercise is running, but...



References

00000

My go to exercise is running, but...

I think my running shoes are getting worn



My go to exercise is running, but...

I think my running shoes are getting worn

I want a replacement, but I know about machines, not shoes!



My go to exercise is running, but...

I think my running shoes are getting worn



I want a replacement, but I know about machines, not shoes!



still, I know my phone can identify my current shoes



My go to exercise is running, but...

I think my running shoes are getting worn



I want a replacement, but I know about machines, not shoes!



still, I know my phone can identify my current shoes

and obtain a new pair of the shoes Llike



The Nike Free RN Distance 2



Felipe Torres Figueroa ECM

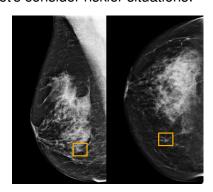
Now let's consider riskier situations:



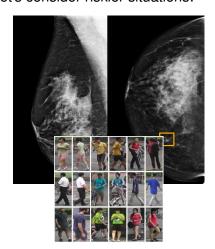
Introduction

00000

Now let's consider riskier situations:



Now let's consider riskier situations:



Now let's consider riskier situations:



Introduction

00000

Motivation: Straight to the point



■ How do we know **how** safe a system is?

ECM

Motivation: Straight to the point

Introduction

00000



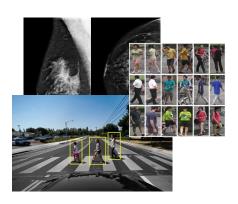
- How do we know how safe a system is?
- How do we know how a system works?

Felipe Torres Figueroa ECM

Motivation: Straight to the point

Introduction

00000



- How do we know how safe a system is?
- How do we know how a system works?
- If a system fails, who is accountable?

Felipe Torres Figueroa ECM

Let's slow for a bit and go step by step:

Computation, Computer Vision and AI

[1]

Introduction

00000



- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



This is a text in second frame. For the sake of showing an example.

■ Text visible on slide 1



Introduction

- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



- Introduction
- 1 Background
- 2 Opti-CAM: Optimizing saliency maps for interpretability
- 3 CA-Stream: Attention-based pooling for interpretable image recognition
- 4 A learning paradigm for interpretable gradients
- References



References I



Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery." *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

