



Interactivity x Explainability:

Toward Understanding How Interactivity Can Improve Computer Vision Explanations

Indu Panigrahi, Sunnie S. Y. Kim*, Amna Liaqat*, Rohan Jinturkar, Olga Russakovsky, Ruth Fong, Parastoo Abtahi Department of Computer Science, Princeton University, Princeton, NJ, USA

Motivation & Goal

Computer vision explanations have been developed to help users understand why such models make their decisions. Traditionally, these explanations have been **static**.

Problems with Static Explanations:

- Information Overload: the amount of information can be overwhelming
- **Semantic-Pixel Gap**: users have difficulty connecting pixel regions to the represented semantic object
- Lack of Exploration: users can't explore the underlying causal relationships



Interactivity as a Customizable Solution:

- Interfaces for inspecting models and datasets at a macroscopic level
- Textual and tabular data at the explanation level for individual datapoints

Research Questions: How do end-users...

- 1. leverage interactivity to understand the information conveyed by computer vision explanations?
- 2. perceive interactive computer vision explanations?



For each of 3 explanation types (rows), we tested 3 interactive mechanisms (columns), chosen for the 3 problems with static explanations:

- Filtering: increase or decrease the amount of information
- **Overlays**: hover over the explanation and a visual dot or textual label is supplied
- **Counterfactuals**: edit the image and observe how the prediction and explanation change

Using a birding dataset¹, we created mock-ups of the 12 explanations and conducted a withinsubjects study with 24 participants of varying AI and domain expertise. Participants verbalized their thoughts while completing a task (identifying the 3 most & least important bird parts).

Qualitative Findings

We created a codebook from the audio and video recordings of the studies and performed a Reflexive Thematic Analysis².

Participants appreciate interactive mechanisms that augment the



Quantitative Findings

00

TO I

explanation without changing the underlying explanation.

"I like being able to progressively add and take away information as [I] see fit" (P19, Filtering); "I feel like the way [Overlays are] presented makes it so that you can take it in easier" (P2, Overlays)

Participants find interactive mechanisms that alter the underlying explanation overwhelming.

"Too many different combinations" (P6); "It was a little daunting to see basically everything change" (P3)

Although participants find Counterfactuals overwhelming, they utilize **Counterfactuals to resolve confusion around static presentations** by inducing systematic changes in model predictions and explanations. "It was helpful to be able to change the color of [a bird part] and see how that affected the bar graph...that sort of helped me understand what the [strikethrough] is for" (P17)



0

Participants leverage Counterfactuals to explore a range of

explanations to better understand the AI model more broadly. Appreciated the ability to "change things around [and] see other

Key Takeaways & Future Work

While interactivity helps address the problems with static computer vision explanations, it also introduces new challenges.

Participants felt that Filtering and Overlays allowed them to quickly focus on information of interest.

"It produces a ranking of the similarity scores that I don't need to think about myself" (P6, Filtering); "faster because I could hover over it and it could tell me what the parts were" (P21, Overlays)

- We provide design recommendations to mitigate such challenges.
- Our study is a preliminary step toward evaluating interactive lacksquarecomputer vision explanations and thus has several limitations.
- More work is needed to test a wider range of mechanisms and examine the potential risks of interactive explanations.

Recommendations

- Avoid interdependent input controls
- Constrain the input and output space
- Design an optimal static default view



Exercise and Health, 2019.

[1] Wah et al. "The Caltech-UCSD Birds-200-2011 Dataset". California Institute of Technology, 2011. [2] Braun & Clarke. "Reflecting on reflexive thematic analysis". Qualitative Research in Sport,

References & Acknowledgements





We acknowledge support from the NSF Graduate Research Fellowship Program (SK), the Princeton SEAS Howard B. Wentz, Jr. Junior Faculty Award (OR), the Princeton SEAS Innovation Fund (RF), and Open Philanthropy (RF, OR).

