VLSlice: Interactive Vision-and-Language Slice Discovery

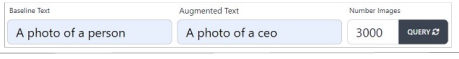
Eric Slyman Oregon State University
Minsuk Kahng Google Research
Stefan Lee Oregon State University

Google Research | Oregon State University | Project Page

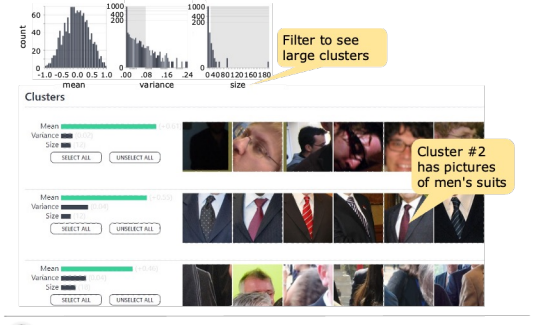
Example Workflow

Does CLIP think CEOs wear men's suits?

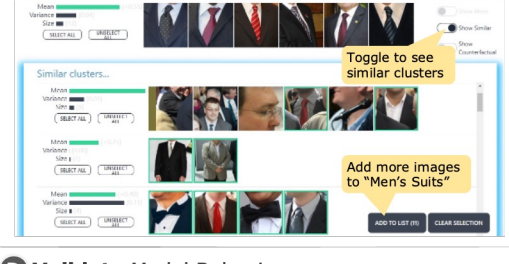
A Query Subjects and Bias Dimension



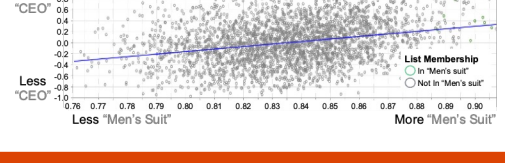
B Explore Visiolinguistic Clusters



C Refine Clusters by Gathering Supporting Examples



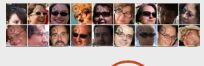
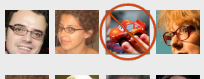
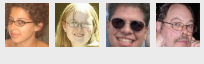
D Validate Model Behavior



***VL*Slice** is an interactive system enabling user-guided discovery of Vision-and-Language Slices, coherent representation-level subgroups with consistent visiolinguistic behavior, from unlabeled image sets. Slices can help identify problematic behaviors and biases learned by web-scale pretrained models. *VL*Slice supports users in discovering and refining slices along arbitrary bias dimensions while leveraging those slices to assist them in validating model behavior.

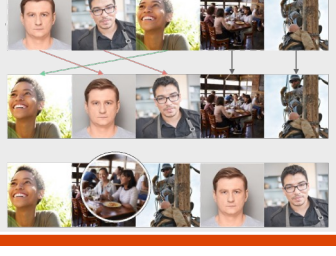
Quality slices are difficult to find

Slices must satisfy several properties to be indicative of model behavior and sufficient for further analysis. We design *VL*Slice to help people more easily discover slices with underlying "good" image sets that meet the following properties:

Large		Contains many samples
Coherent		Coherent and well-defined
Representative		Broadly representative of the true underlying concept
Systemic	 85%↑	Systemic relationship with language

Clustering as a slice bootstrap

*VL*Slice presents users with visiolinguistic clusters capturing image similarity and relationship with a textual attribute (C_a) to use as a basis for slice creation. We propose ΔC as a metric to better capture text attribute relationships than a standard similarity score by comparing with a baseline (C_b) caption.



C_b – "A photo of a person"

C_a – "A photo of a happy person"

ΔC

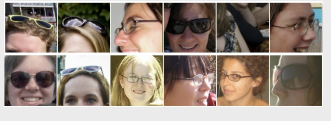
Users iteratively improve slices

Users can leverage recommendation tools to refine discovered slices. New samples trigger recomputation, cyclically improving the tools output as users improve the quality of their slices.

Slice: Glasses ($\Delta C > 0$)

Similar Cluster ($\Delta C > 0$)

Counterfactual Clusters ($\Delta C < 0$)



User study results (n=22)

Participants find *larger* (#Img.) and more *coherent* (F1) slices with **VLSlice** when compared against **ListSort**, a control interface representative of common manual workflows.

	ListSort				VLSlice			
	Slices	#Img.	F1	Missed	Slices	#Img.	F1	Missed
Person/CEO	4.3	107	.42	90	4.6	141	.59	54
House/Nice	3.6	87	.20	41	5.1	211	.46	70

Sample slices captured by participants

"masculine glasses" ($\Delta C > 0$)

"people of color" ($\Delta C < 0$)