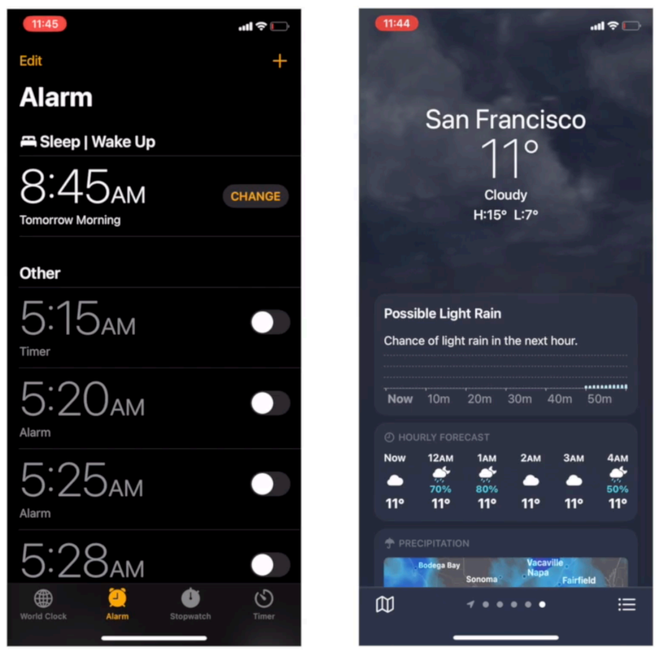
Towards Never-ending Learning of User Interfaces

Jason Wu, Rebecca Krosnick, Eldon Schoop, Amanda Swearngin,
Jeffrey P. Bigham, Jeffrey Nichols

Can models learn to understand UIs by directly interacting with them?

Machines that visually understand user interfaces (UIs) are useful for accessibility, automation, and design, but they currently depend heavily on human annotation. Human annotation is both costly and surprisingly error-prone for tasks such as "tappability" prediction. We introduce the *Never-ending UI Learner,* a system which automatically installs and interacts with real apps to train machine learning models and improve itself over time.

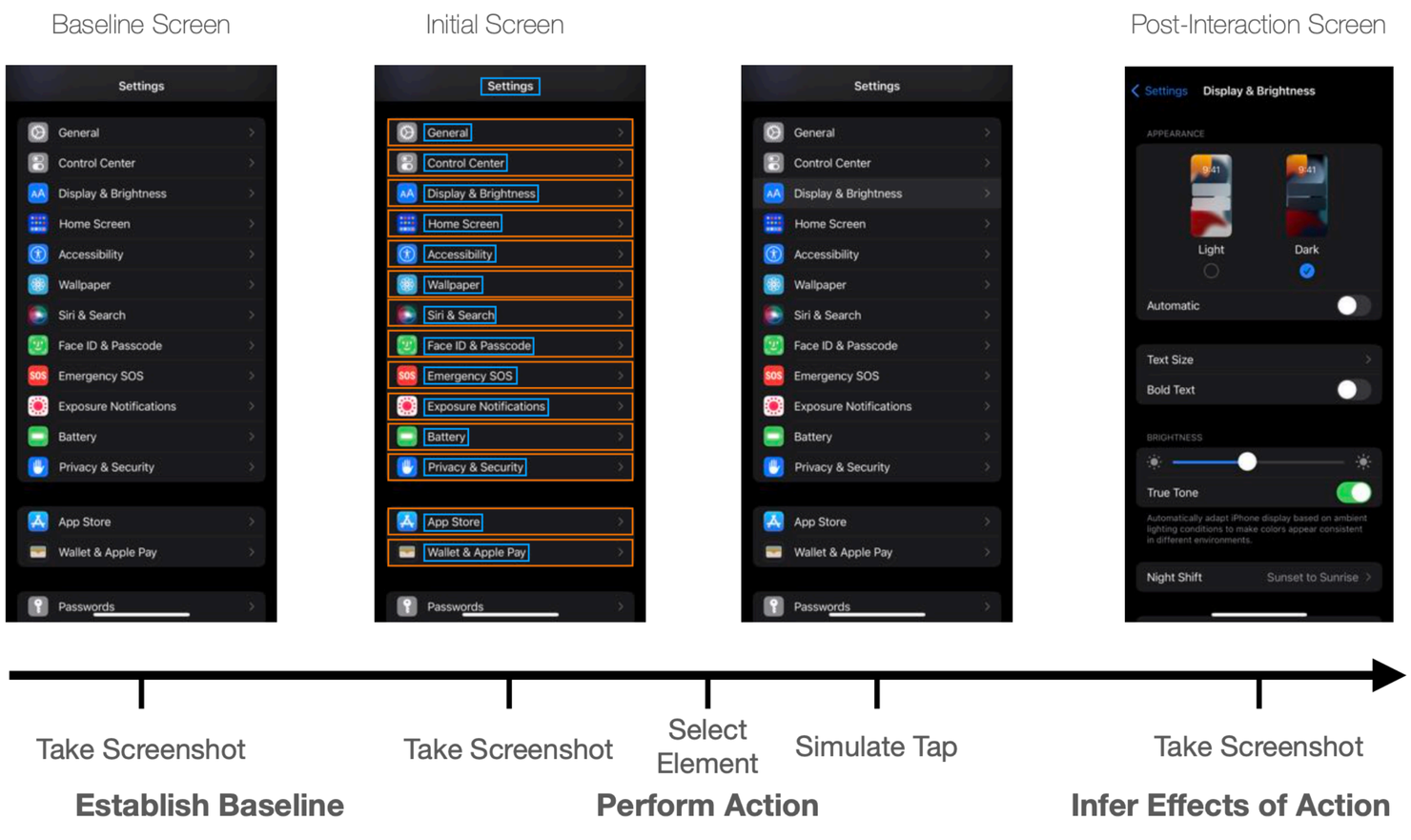
How would you interact with the UI elements in these screenshots?



Which are tappable?

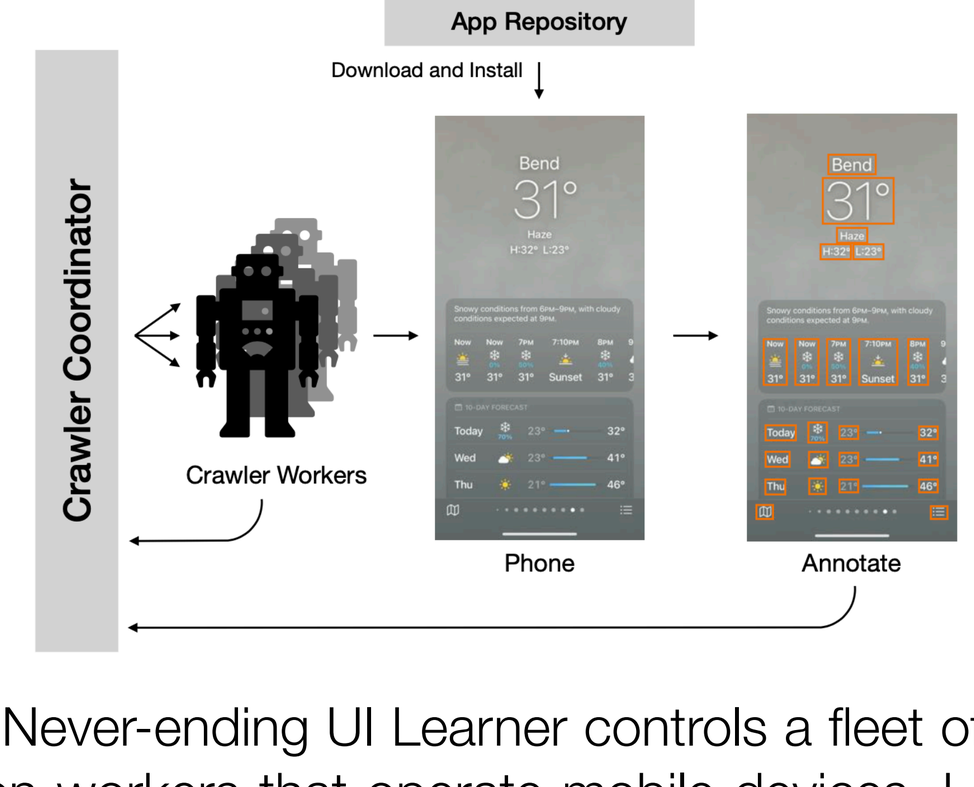
Which are draggable?

UI semantics can be inferred by observing the effects of interaction



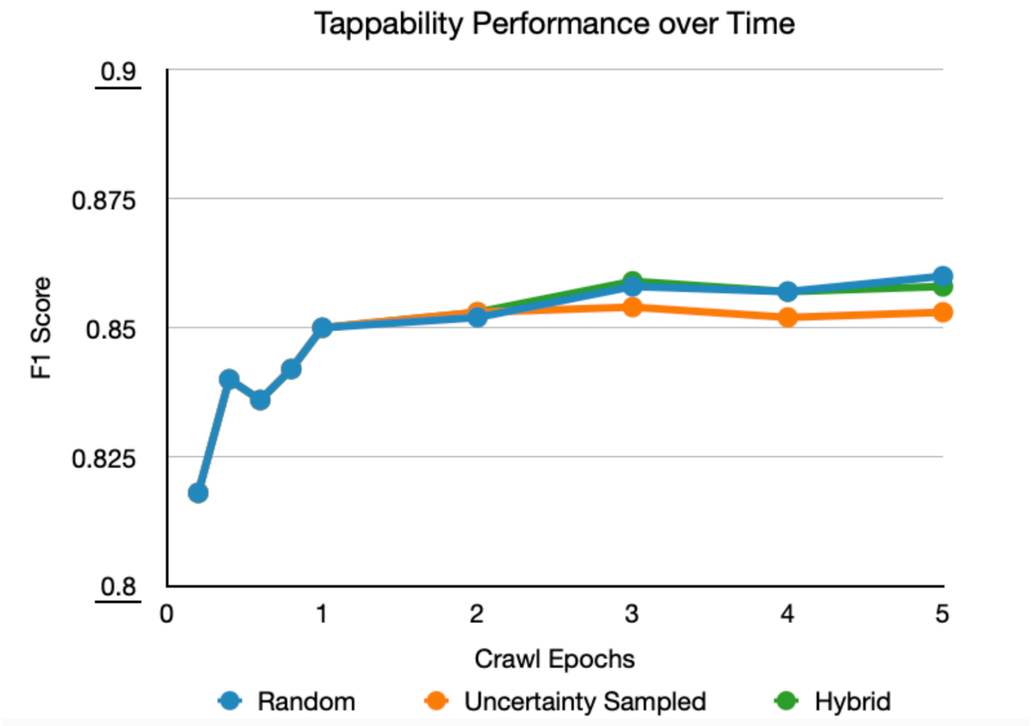
A random UI element is selected then programmatically tapped. Screenshots taken before and after the tap are compared using a *screen similarity model*. If the tap led to a new screen or displayed other visual indications of state change, it is considered tappable. Labels are used to train a "tappability" model.

Never-ending UI Learner



The Never-ending UI Learner controls a fleet of ML-driven workers that operate mobile devices. Labels collected from the workers are used by the coordinator to train and update worker models.

Experimental Results



The Never-ending UI Learner spent 5,000 device-hours crawling a list of 6,000 apps using a variety of policies. Performance improved over time and led to a final model with F1=0.86 without human labels.