



Generator implicitly learns "**ingredients**" of natural scenes.

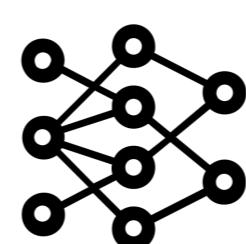


Which ingredients make a scene?

- **visual features**
- **scene grammar features**



How can we **represent** a scene's "grammar"?

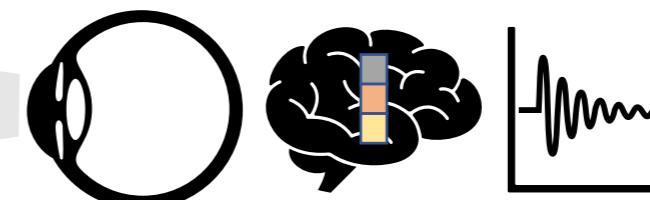


Training graph-based **computational models** to encode scene grammar representations.



How does scene grammar shape **neural object representations**?

visual system



behavior

- **When** can we read out phrase structure?
- **Which features** characterise shared representations?