

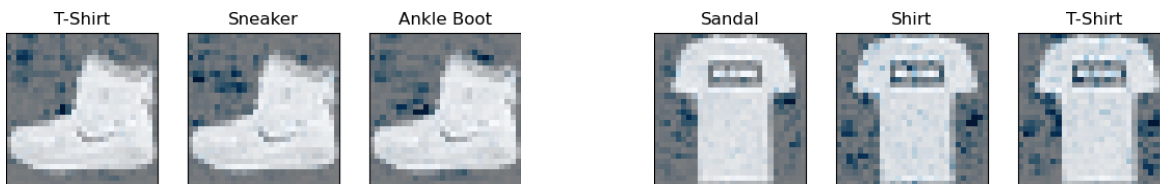
Explaining Deep Networks

Caspar Schneider, Sinan Kuscü, Sinan Harms

September 2022

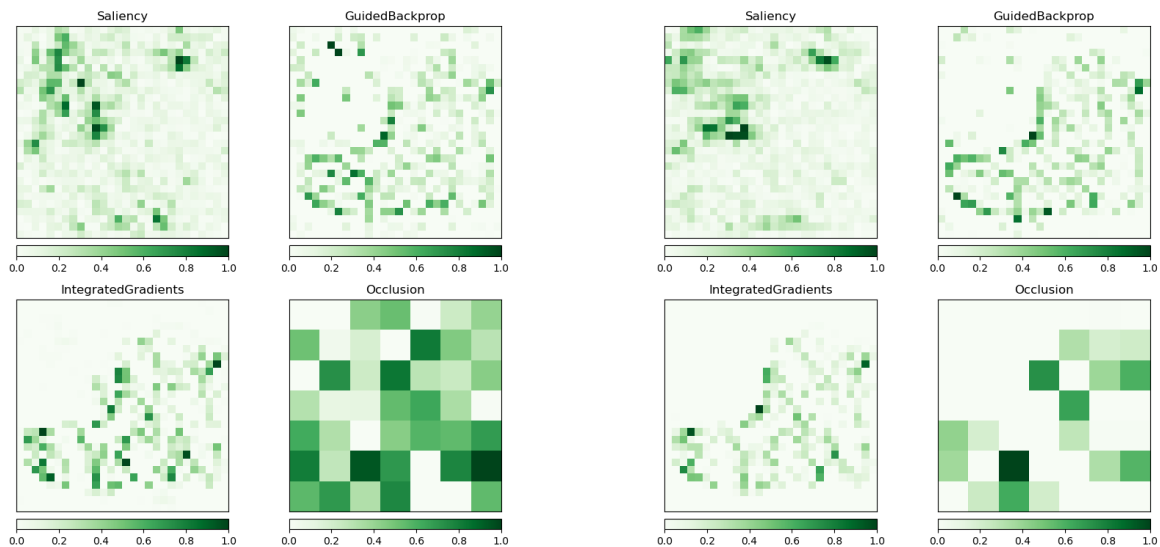
PART II - PRACTICE

Task 1



- expected saliency map to focus more in edges for most likely classification, since training a convolutional network
- for t-shirt can observe that saliency has attention on the area around the short sleeves when comparing to normal shirt

Task 2



- left side attacked, right side original
- GuidedBackprop and IntegratedGradients after attack show more noise around the edges
- saliency map gets less focussed on certain areas compared to original
- occlusion shows the biggest shift → more distribution across whole image