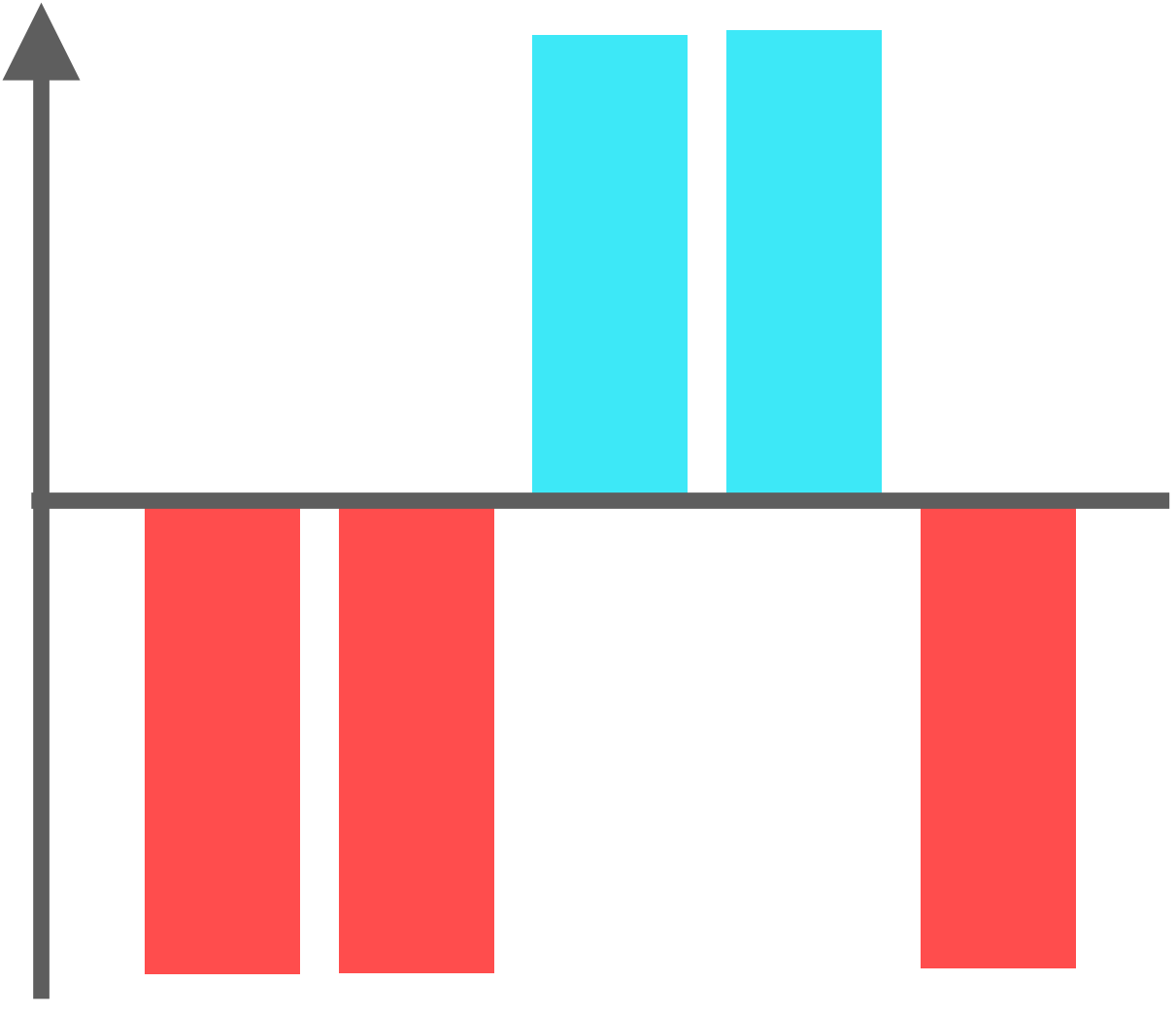
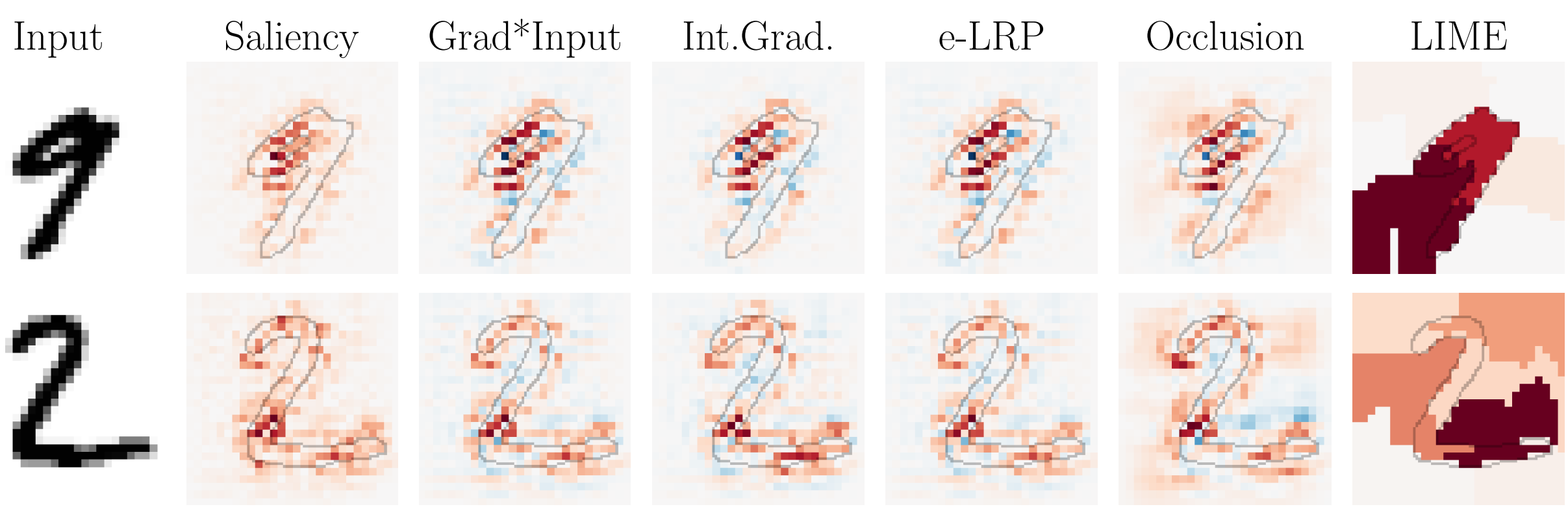
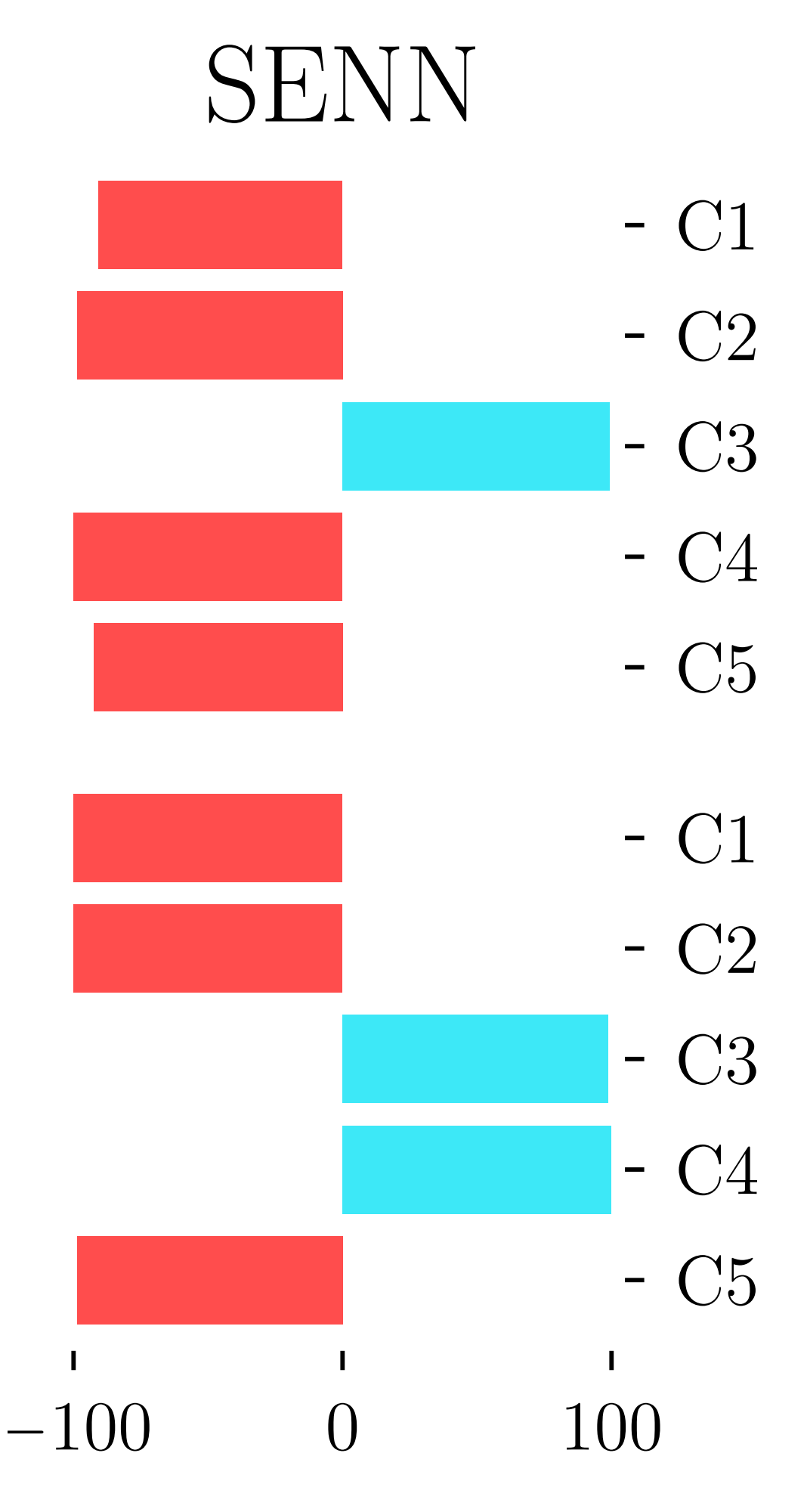
Self-Explaining Neural Networks (SENN)

[redacted]

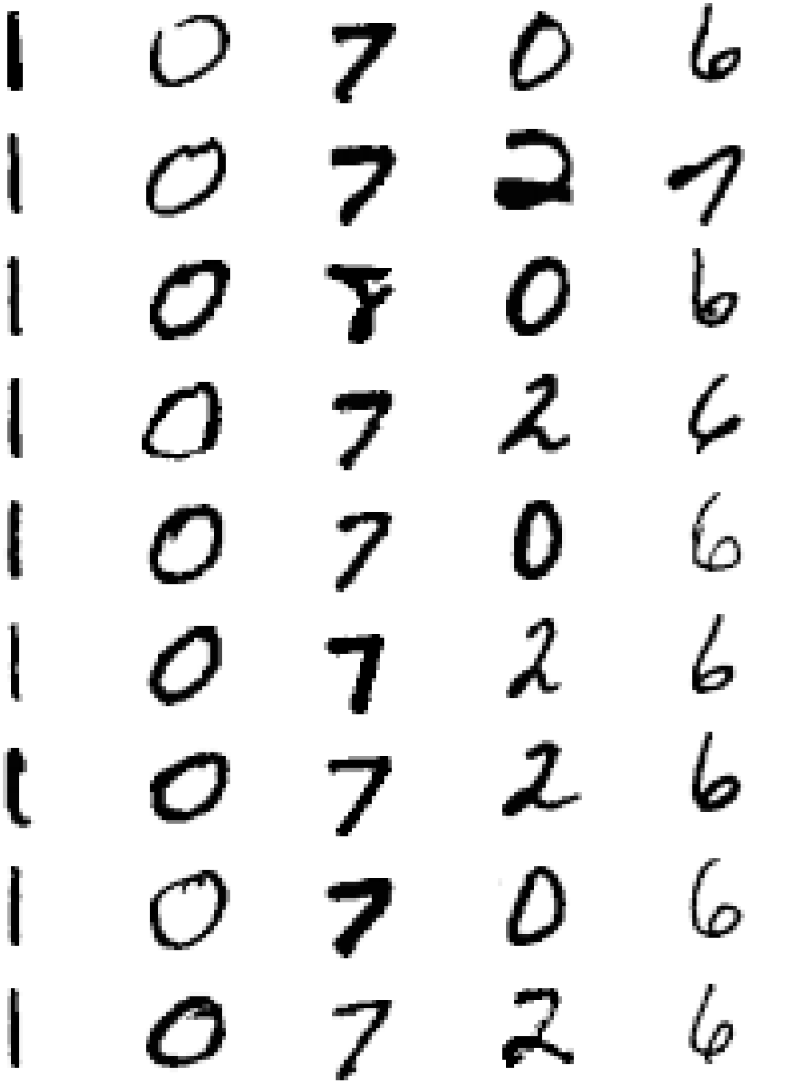
98

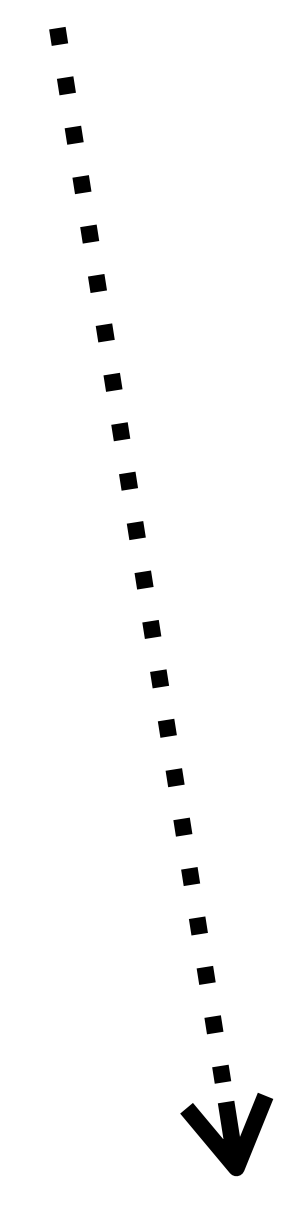


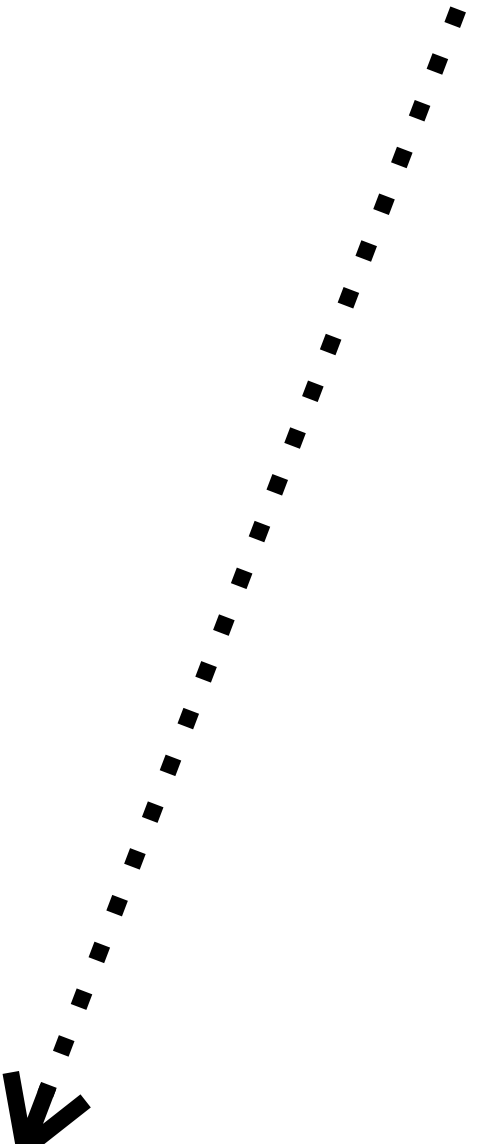

Input
Saliency
Grad*Input
Int.Grad.
e-LRP
Occlusion
LIME

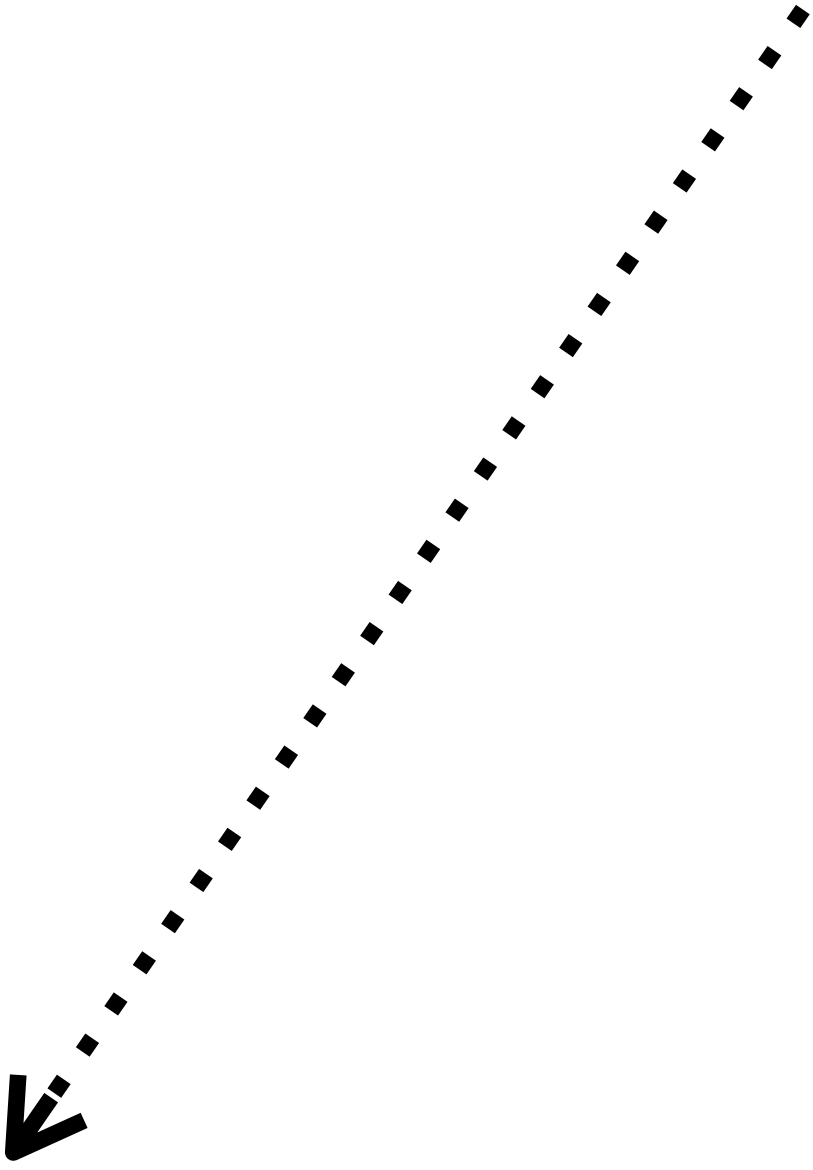

SENN
C1
C2
C3
C4
C5
C1
C2
C3
C4
C5
−100
0
100

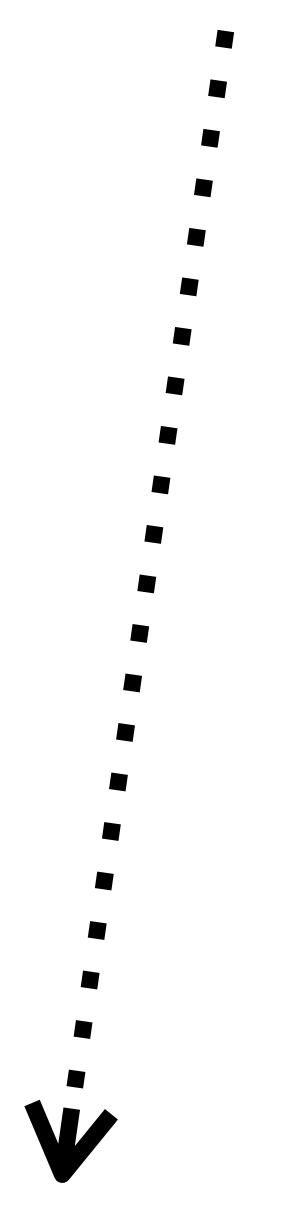
Concept 1 Concept 2 Concept 3 Concept 4 Concept 5

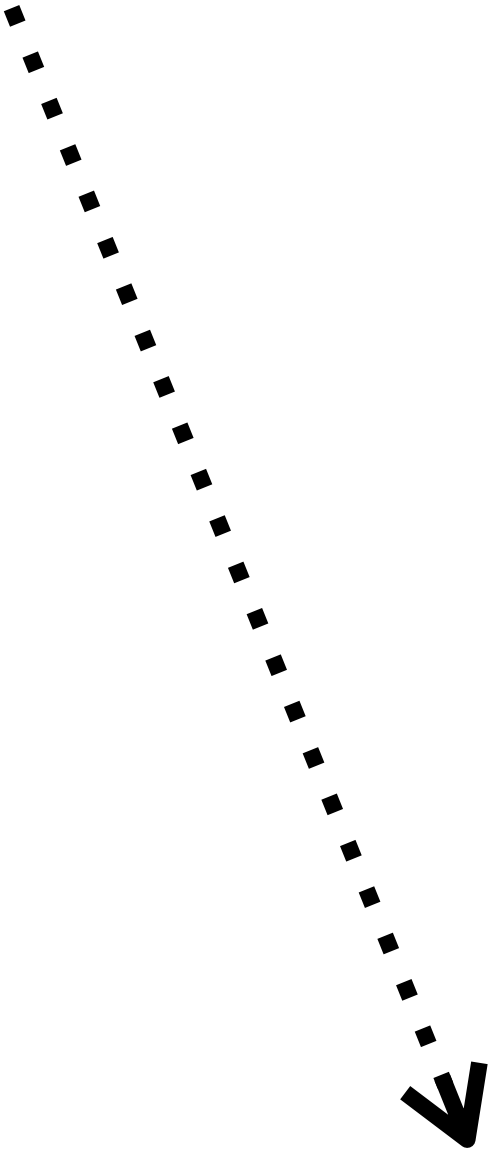




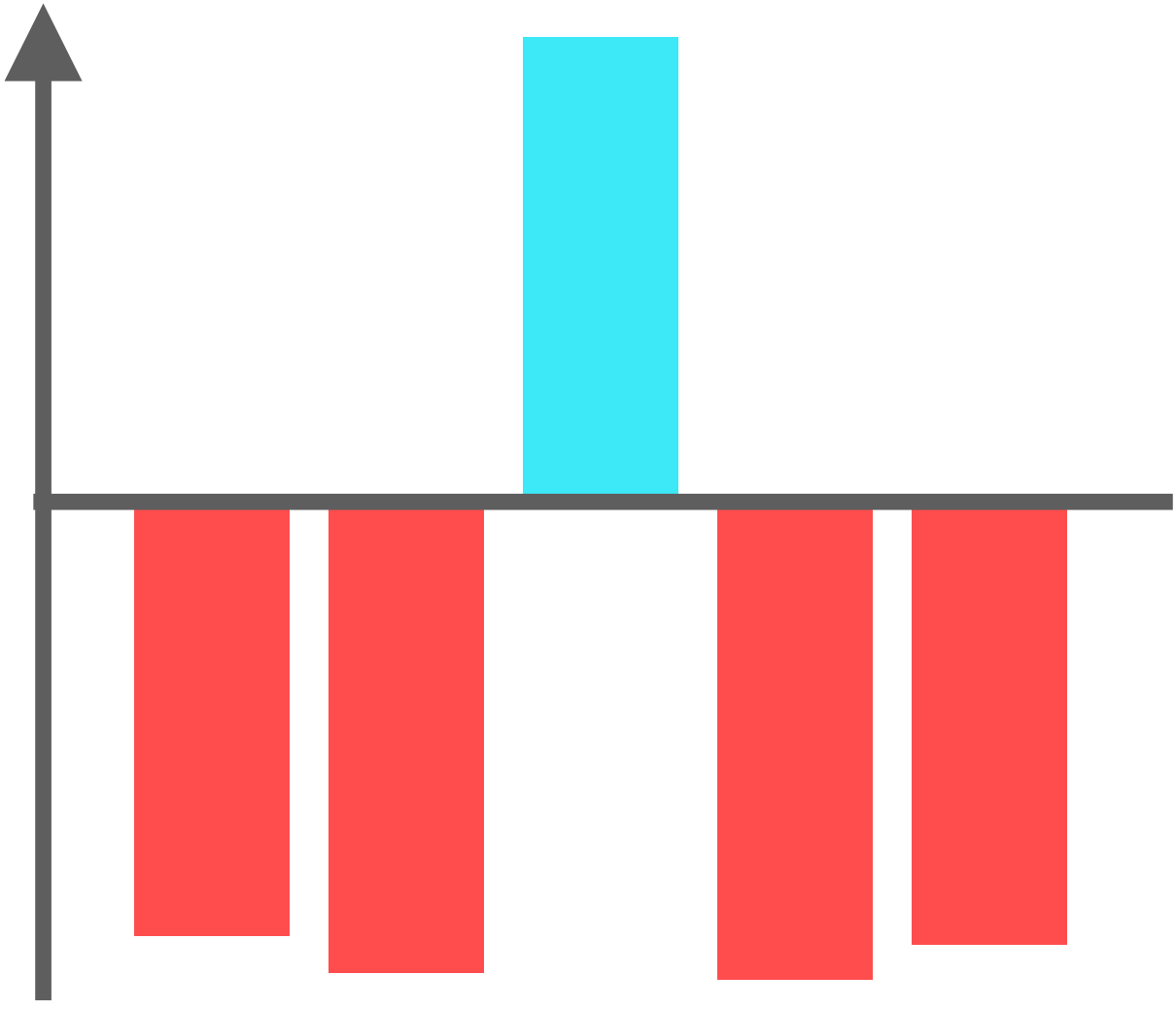




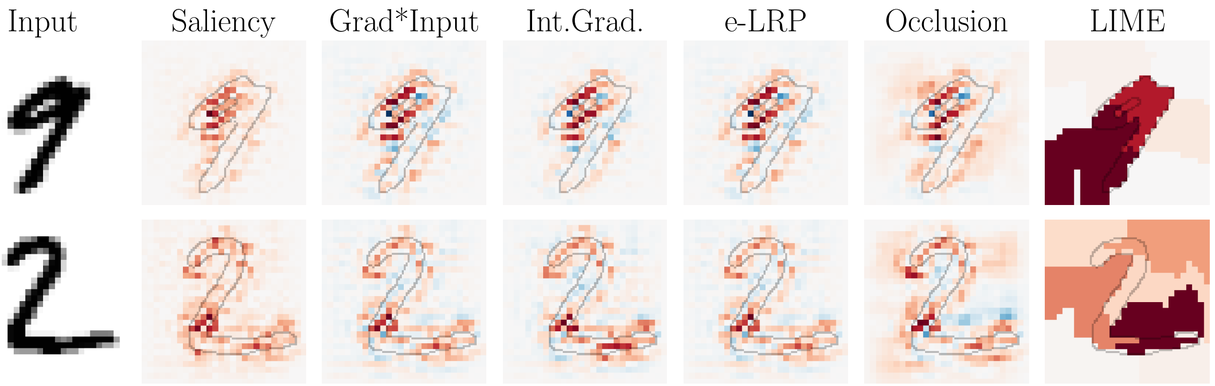


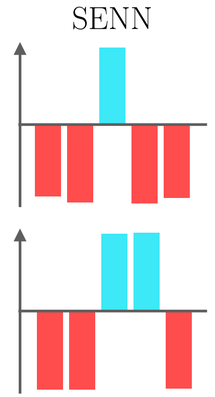


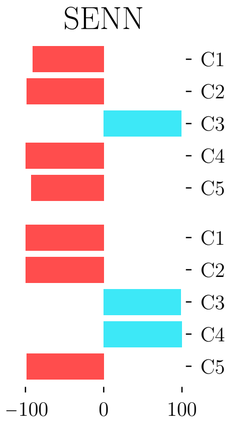
SENN

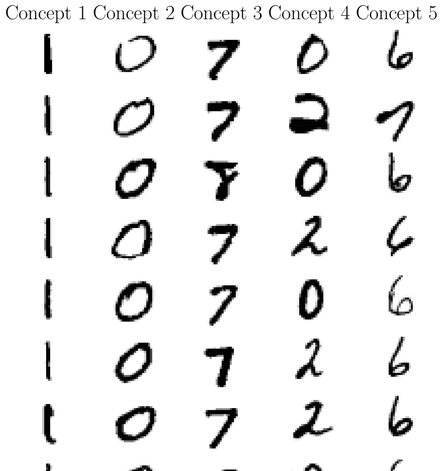
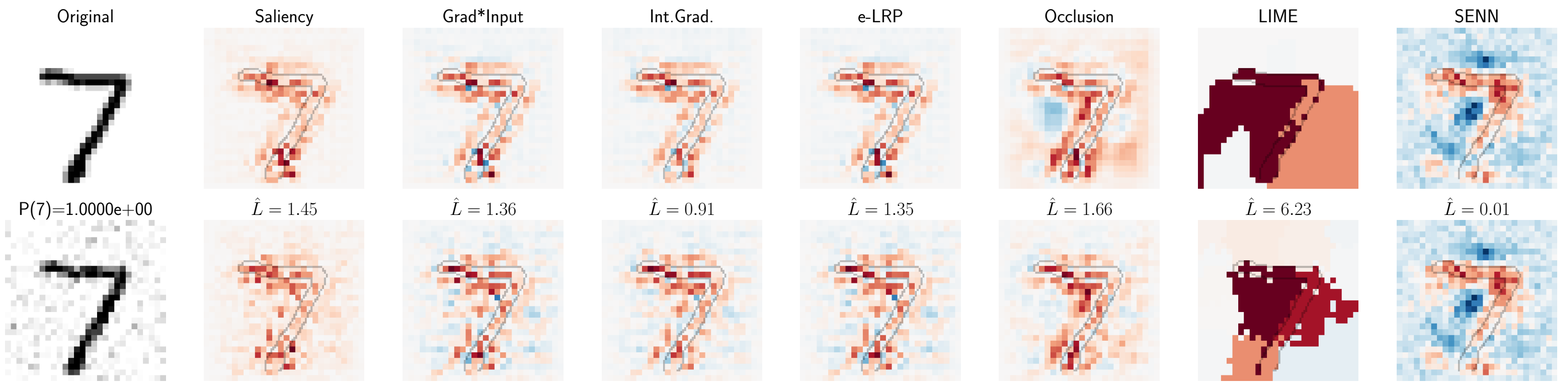


Self-Explaining Neural Networks (SENN)









SENN
Robustness