Rebuttal

March 2024

1 Rebuttal

W-1 : SOP and IntGrad are able to find most number of clusters.

Attr. Method	**Ratio Voids**	**Ratio Clusters**
lime	0.35	0.0004
shap	0.34	0.0001
rise	0.28	0.0
intgrad	0.47	0.0037
gradcam	0.33	0.0001
archipelago	0.22	0.0
sop	0.22	0.0032

Table 1: SOP and IntGrad are able to find most number of clusters.

W-3 : MoRF and LeRF

Attr. Method	**MoRF (low)**	**LeRF (high)**
sop	0.44	0.53
lime	0.13	0.68
shap	0.20	0.62
rise	0.35	0.66
intgrad	0.53	0.51
gradcam	0.35	0.59
archipelago	0.32	0.49
fullgrad	0.33	0.62

Table 2: MoRF and LeRF comparison among attribution methods.

Q3 : Fidelity for ImageNet

Attr. Method	Fidelity Soft (low: better)
lime	0.29
shap	0.059
rise	6.77
intgrad	7.30
gradcam	21.43
archipelago	21.41
fullgrad	21.42
sop	0.0

Table 3: Fidelity for ImageNet across various attribution methods.

Q4 : Fidelity for SST

Attr. Method	Fidelity Soft (low: better)
lime	3.63
shap	0.29
rise	0.04
intgrad	5.94
archipelago	2.83
pls	2.27
fresh	0.0
sop	-2e-10

Table 4: Fidelity for SST evaluated across various attribution methods.