Supplemental Materials: Ensembled Cold-Diffusion Restorations for Unsupervised Anomaly Detection

Table 1: Ablations on AG procedure. AP test set results using $AS_{ensemble}$ anomaly score. Mean and std. dev. on 4 seeds.

AG	ATLAS-ex	BraTS-T1	BraTS-T2
Baseline	0.47 ± 0.01	0.48 ± 0.02	0.79 ± 0.02
FPI	0.44 ± 0.02	0.46 ± 0.02	0.69 ± 0.06
Bias w/o tissue masking	0.43 ± 0.03	0.45 ± 0.04	0.80 ± 0.01

Table 2: AP test set results based on step size for $AS_{ensemble}$. Mean and std. dev. on 4 seeds.

step size	ATLAS-ex	BraTS-T1	BraTS-T2
10	0.48 ± 0.01	0.49 ± 0.02	0.80 ± 0.02
20	0.48 ± 0.01	0.48 ± 0.02	0.80 ± 0.02
25	0.48 ± 0.01	0.48 ± 0.02	0.79 ± 0.02
33	0.48 ± 0.01	0.48 ± 0.02	0.79 ± 0.02
50	0.48 ± 0.01	0.46 ± 0.02	0.67 ± 0.02

Table 3: AP results for ATLAS ($N = 655$) and subsets excluding artefacts (ATLAS-exArt $N = 631$) and excluding artefacts and skull-stripping issues (ATLAS-ex $N = 571$). List of cases in each subset is available in implementation. Skull-stripping is performed using routines from [10]. Mean and std. dev. on 4 seeds.

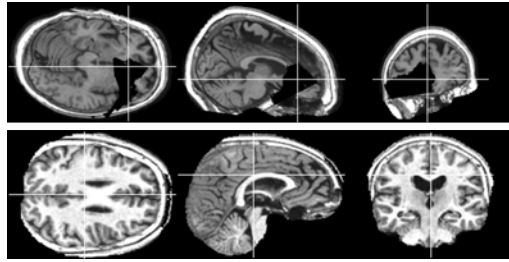| ATLAS	ATLAS-exArt	ATLAS-ex
0.33 ± 0.02	0.44 ± 0.01	0.47 ± 0.01

Fig. 1: Top: Example of ATLAS-T1 MR image with artefact. Bottom: example of ATLAS-T1 MR image with incorrect skull-stripping.
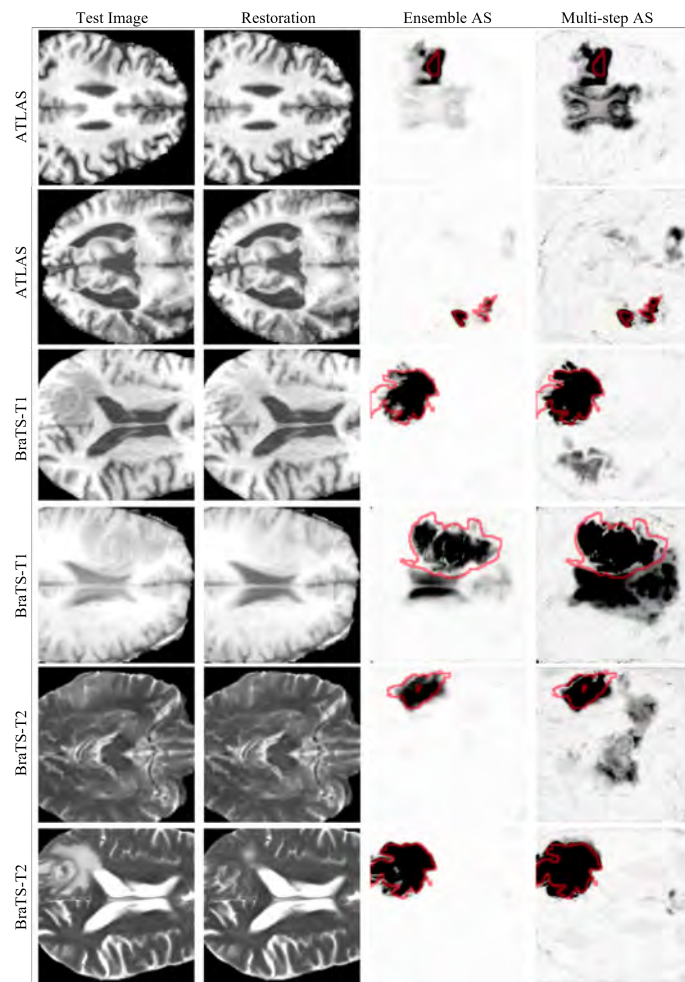

Fig. 2: Examples of restorations, Ensembled and Multi-step AS vs Ground Truth (red outline)