

We include additional model EfficientAD and dataset MVTec. The results are shown in Table 1. Fastflow and EfficientAD achieve a 73.38% average improvement across the four metrics on MVTec. For EfficientAD on VisA, our formal property guidance results in an 56.11% improvement.

Model	Class	$M_s(\downarrow)$		$M_d(\downarrow)$		$M_\omega(\downarrow)$		$M_{1-\omega}(\downarrow)$	
		Baseline	Guided	Baseline	Guided	Baseline	Guided	Baseline	Guided
Fastflow (MVTec)	cable	-12.56 ± 0.14	-13.18 ± 0.18	171.91 ± 13.59	4.19 ± 0.01	-0.006 ± 0.004	-0.007 ± 0.003	0.02 ± 0.01	-0.01 ± 0.006
	capsule	-13.73 ± 0.37	-15.21 ± 0.33	87.01 ± 11.04	4.21 ± 0.01	-0.004 ± 0.004	-0.006 ± 0.003	0.04 ± 0.01	-0.01 ± 0.006
	carpet	-18.93 ± 0.37	-19.47 ± 0.74	170.25 ± 8.92	4.21 ± 0.01	-0.014 ± 0.005	-0.014 ± 0.006	0.00 ± 0.02	-0.01 ± 0.011
	hazelnut	-10.35 ± 0.24	-12.94 ± 0.22	105.00 ± 9.12	4.21 ± 0.01	-0.003 ± 0.004	-0.011 ± 0.004	0.09 ± 0.01	-0.02 ± 0.005
	metal nut	-8.96 ± 0.32	-11.07 ± 0.38	157.62 ± 10.27	4.21 ± 0.01	-0.001 ± 0.005	-0.007 ± 0.002	0.07 ± 0.03	-0.02 ± 0.004
	transistor	-12.15 ± 0.16	-13.81 ± 0.17	265.05 ± 75.78	7.04 ± 1.96	-0.006 ± 0.002	-0.006 ± 0.001	0.02 ± 0.01	-0.01 ± 0.002
	$\Delta(\uparrow)$	+13.47%		+96.82%		+155.56%		+112.63%	
EfficientAD (MVTec)	cable	6.61 ± 0.96	6.27 ± 0.63	379.84 ± 54.00	9.15 ± 1.08	-0.101 ± 0.052	-0.138 ± 0.026	0.34 ± 0.06	-0.03 ± 0.009
	capsule	141.28 ± 3.57	142.71 ± 2.68	223.80 ± 27.91	10.59 ± 1.18	-0.383 ± 0.396	-0.140 ± 0.223	-0.28 ± 0.34	-0.07 ± 0.088
	carpet	6.57 ± 3.61	0.70 ± 0.03	413.55 ± 54.08	10.16 ± 1.38	0.344 ± 0.258	-0.011 ± 0.004	2.23 ± 1.58	-0.00 ± 0.001
	hazelnut	95.76 ± 9.10	49.35 ± 17.89	244.18 ± 26.76	9.72 ± 0.95	2.348 ± 1.048	-0.875 ± 0.787	21.10 ± 2.31	0.04 ± 0.065
	metal nut	3.52 ± 0.09	2.58 ± 0.13	422.73 ± 50.43	10.40 ± 1.27	0.021 ± 0.035	-0.027 ± 0.036	0.63 ± 0.06	-0.00 ± 0.002
	transistor	37.15 ± 3.98	45.77 ± 11.95	272.63 ± 61.83	44.72 ± 140.20	-1.385 ± 0.708	-0.403 ± 0.662	-1.86 ± 2.31	-0.62 ± 2.089
	$\Delta(\uparrow)$	+24.24%		+94.59%		+61.89%		+27.83%	
EfficientAD (VisA)	candle	308.09 ± 38.03	384.16 ± 58.56	228.59 ± 46.47	12.03 ± 2.17	-6.315 ± 3.751	-0.178 ± 0.210	-4.38 ± 6.87	-0.20 ± 0.206
	cashew	375.33 ± 22.36	362.85 ± 9.31	505.46 ± 75.41	17.67 ± 2.59	-2.328 ± 1.293	-2.578 ± 0.829	0.42 ± 0.69	0.12 ± 0.157
	pipe fryum	109.77 ± 22.28	113.82 ± 6.06	119.54 ± 37.03	8.42 ± 0.01	0.424 ± 1.979	-0.111 ± 1.233	0.33 ± 0.89	-0.27 ± 0.263
	pcb1	26.20 ± 0.48	16.10 ± 0.31	218.47 ± 16.85	8.37 ± 0.01	0.363 ± 0.061	0.010 ± 0.040	0.33 ± 0.09	-0.00 ± 0.003
	pcb2	10.74 ± 0.82	9.17 ± 1.68	497.95 ± 68.94	8.41 ± 0.00	0.150 ± 0.241	-0.031 ± 0.035	3.90 ± 0.11	-0.00 ± 0.003
	$\Delta(\uparrow)$	+5.62%		+95.74%		+51.53%		+71.56%	

Table 1: Comparison of baseline and guided performance across four metrics for MVTec dataset categories with EfficientAD and FastFlow. Δ is the mean improvement percentage of the guided result from baseline.

We introduce Llama-2-7b as a new time-series anomaly detector. We additionally introduce two new time-series datasets in WADI and HAI, and present some preliminary results in Table 2. Our formal property guidance consistently yields enhancements across the four metrics. Llama-2-7b exhibits an average improvement of 67.17% in formal metrics, while GPT-2 demonstrates an increase of 53.90%.

Model	Dataset	$M_s(\downarrow)$		$M_d(\downarrow)$		$M_\omega(\downarrow)$		$M_{1-\omega}(\downarrow)$	
		Baseline	Guided	Baseline	Guided	Baseline	Guided	Baseline	Guided
Llama2	SWaT	0.83 ± 0.05	0.59 ± 0.04	3.05 ± 0.61	7.25 ± 2.50	0.084 ± 0.026	-0.026 ± 0.008	0.19 ± 0.02	0.05 ± 0.012
	WADI	0.97 ± 0.00	0.26 ± 0.01	0.98 ± 0.01	0.00 ± 0.00	0.087 ± 0.005	0.000 ± 0.000	0.61 ± 0.01	0.00 ± 0.000
	HAI	0.92 ± 0.00	0.58 ± 0.01	0.75 ± 0.00	0.00 ± 0.00	0.166 ± 0.008	0.000 ± 0.000	0.18 ± 0.01	0.00 ± 0.000
	$\Delta(\uparrow)$	+46.36%		+20.77%		+110.32%		+91.23%	
GPT-2	SWaT	0.68 ± 0.06	0.57 ± 0.05	12.81 ± 15.82	16.28 ± 16.30	-0.029 ± 0.041	-0.094 ± 0.038	0.13 ± 0.04	0.09 ± 0.031
	WADI	0.71 ± 0.04	0.32 ± 0.00	34.45 ± 0.60	42.15 ± 0.94	0.018 ± 0.003	-0.019 ± 0.002	0.43 ± 0.04	0.07 ± 0.002
	HAI	0.92 ± 0.02	0.61 ± 0.11	3.69 ± 9.72	5.67 ± 18.80	0.136 ± 0.050	-0.005 ± 0.018	0.21 ± 0.03	0.04 ± 0.127
	$\Delta(\uparrow)$	+34.93%		-34.37%		+177.79%		+37.24%	

Table 2: Comparison of baseline and guided performance across four metrics for SWaT, WADI, and HAI dataset categories with Llama2 and GPT2 model. Δ is the mean improvement percentage of the guided result from baseline.

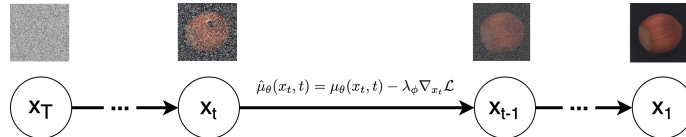


Figure 1: Update to Figure 3 that illustrates how property guidance operates during the reverse diffusion steps. The formal property loss guides the generation at each step from random noise at x_T to clear repair at x_1 .