Energy Transformer

Benjamin Hoover*, Yuchen Liang*, Bao Pham*, Rameswar Panda, Hendrik Strobelt,
Duen Horng Chau, Mohammed J. Zaki, Dmitry Krotov

*Authors contributed equally

IBM Research

Rensselaer

Georgia Tech

Energy Transformer combines 3 paradigms

Transformer

Evolves tokens using attention

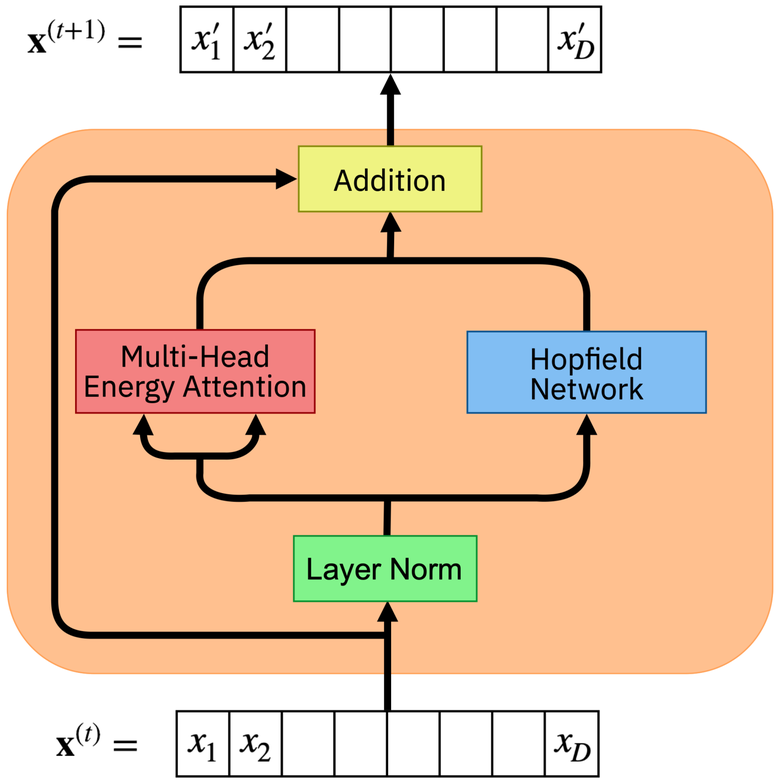
Energy-Based Model

Inference minimizes a computed energy

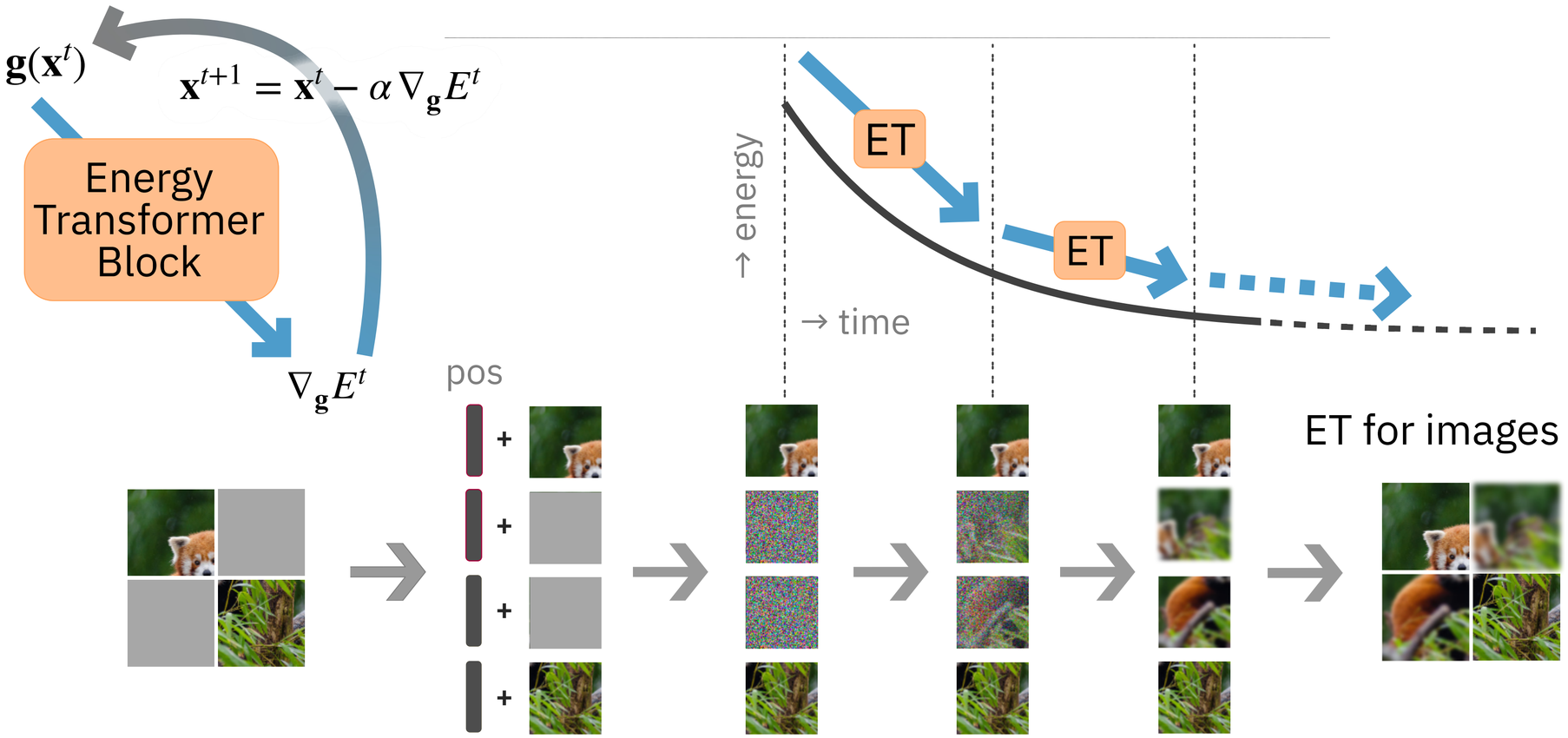
Associative Memory

An dynamical attractor system where inference is pattern completion

✅ Energy transformer block



Energy Transformer in practice



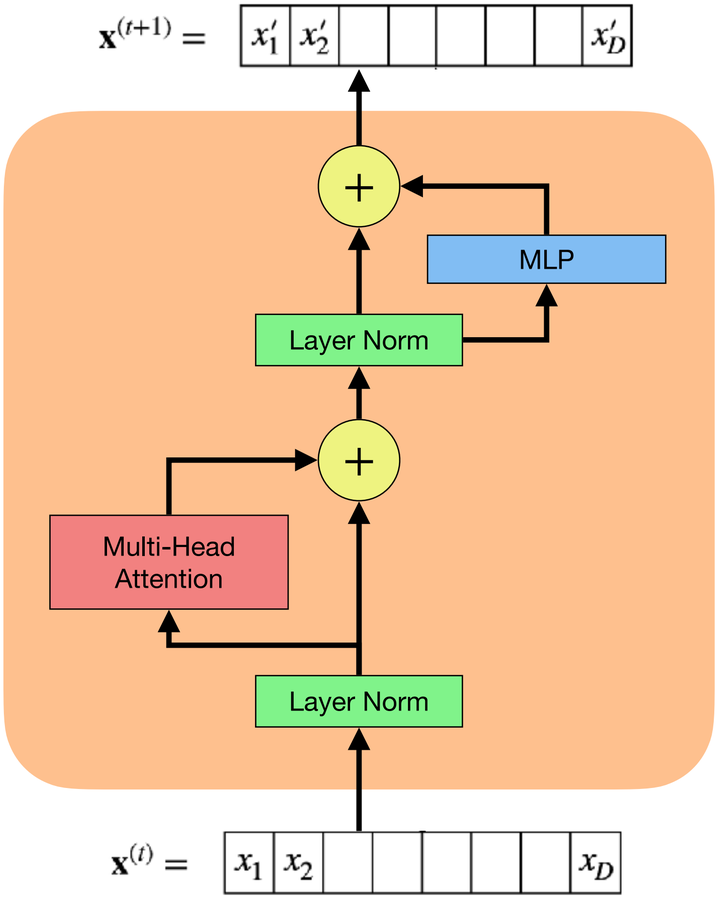
IBM Research

Rensselaer

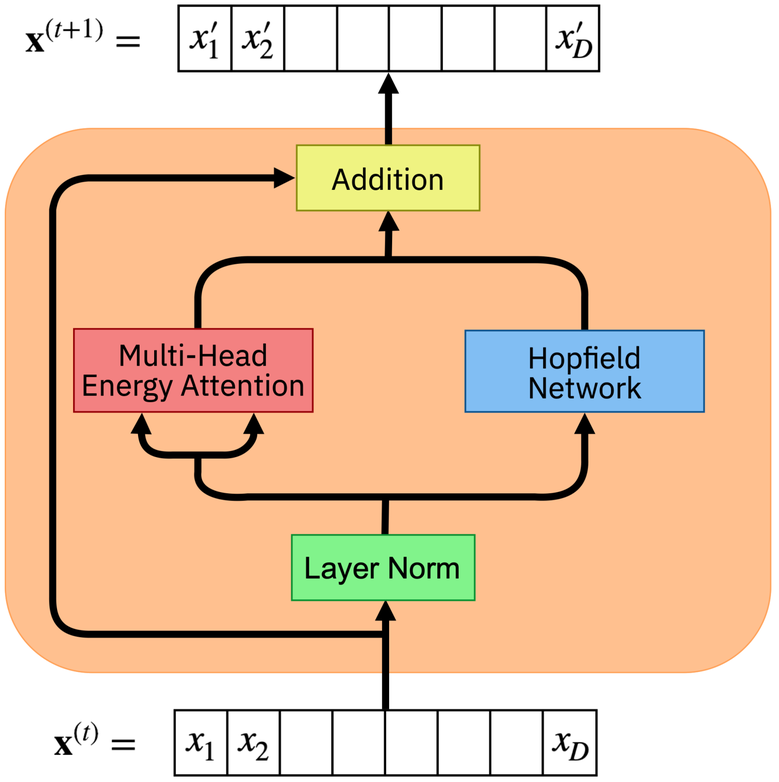
Georgia Tech

Conventional Transformers vs. Energy Transformer

❌ Conventional transformer block



✅ Energy transformer block



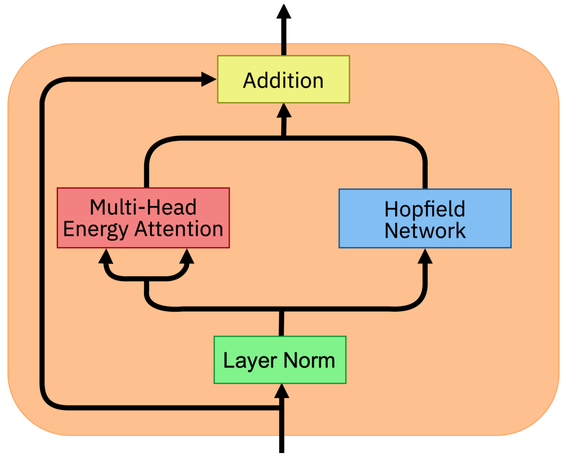
Attention energy is **low when keys are aligned with queries** (and vice versa)

$$E^{\text{ATT}} = -\frac{1}{\beta}\sum_{h}\sum_{C}\log\left(\sum_{B\neq C}\exp\left(\beta\sum_{\alpha}K_{\alpha hB}\,Q_{\alpha hC}\right)\right)$$

Hopfield Network energy is **low when tokens look like memories**

$$E^{\text{HN}} = -\sum_{B=1}^{N}\sum_{\mu=1}^{K}G\Big(\sum_{j=1}^{D}\xi_{\mu j}\,g_{jB}\Big)$$

Energy transformer block

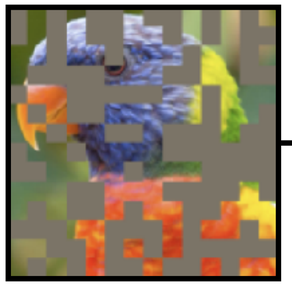


Energy Transformer: **make both sub-blocks happy**

$$E = E^{\text{ATT}} + E^{\text{HN}}$$

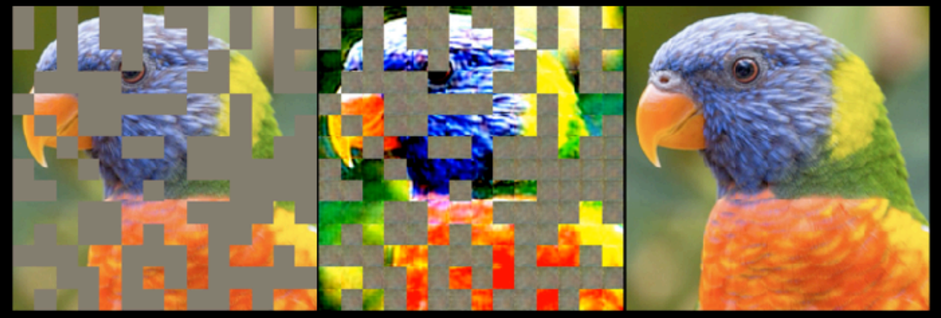
$$\tau\frac{dx_{iA}}{dt} = -\frac{\partial E}{\partial g_{iA}}$$

High energy



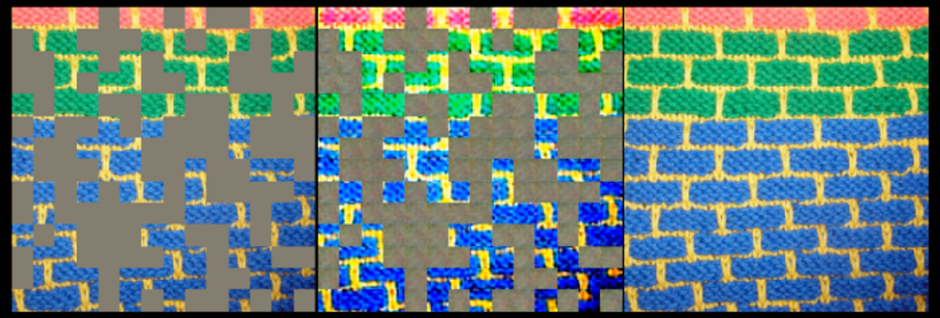
Recurrent ET block

Low energy

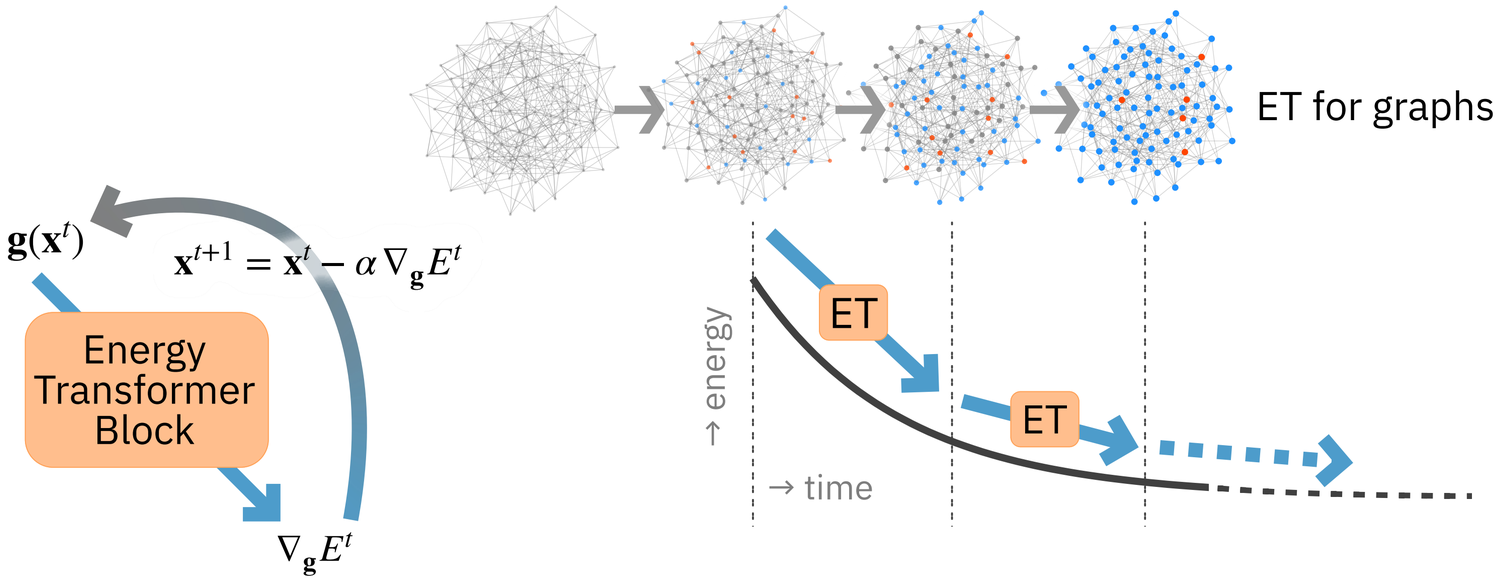






Anomaly Detection



	Datasets	Split	GraphConsis	CAREGNN	PC-GNN	BWGNN	MLP	GT	ET (Ours)
Macro-F1	Yelp	1%	$56.8_{\pm 2.8}$	$62.1_{\pm 1.3}$	$59.8_{\pm 1.4}$	$61.1_{\pm 0.4}$	$53.9_{\pm 0.2}$	$61.7_{\pm 0.4}$	$\mathbf{63.0}_{\pm \mathbf{0.6}}$
		40%	$58.7_{\pm 2.0}$	$63.3_{\pm 0.9}$	$63.0_{\pm 2.3}$	$71.0_{\pm 0.9}$	$57.5_{\pm 0.8}$	$68.7_{\pm 0.4}$	$\mathbf{71.5}_{\pm \mathbf{0.1}}$
	Amazon	1%	$68.5_{\pm 3.4}$	$68.7_{\pm 1.6}$	$79.8_{\pm 5.6}$	$\mathbf{90.9}_{\pm \mathbf{0.7}}$	$74.6_{\pm 1.2}$	$88.6_{\pm 0.5}$	$89.3_{\pm 0.7}$
		40%	$75.1_{\pm 3.2}$	$86.3_{\pm 1.7}$	$89.5_{\pm 0.7}$	$92.2_{\pm 0.4}$	$79.1_{\pm 1.2}$	$91.7_{\pm 0.8}$	$\mathbf{92.8}_{\pm \mathbf{0.3}}$
	T-Finance	1%	71.7	73.3	62.0	84.8	61.0	81.5	$\mathbf{85.1}_{\pm \mathbf{1.0}}$
		40%	73.4	77.5	63.1	86.8	70.5	83.6	$\mathbf{88.2}_{\pm \mathbf{1.0}}$
	T-Social	1%	52.4	55.8	51.1	75.9	50.0	64.3	$\mathbf{79.1}_{\pm \mathbf{0.7}}$
		40%	56.5	56.2	52.1	**83.9**	50.3	68.2	$\mathbf{83.5}_{\pm \mathbf{0.4}}$
AUC	Yelp	1%	$66.4_{\pm 3.4}$	$75.0_{\pm 3.8}$	$\mathbf{75.4}_{\pm \mathbf{0.9}}$	$72.0_{\pm 0.5}$	$59.8_{\pm 0.4}$	$72.5_{\pm 0.6}$	$73.2_{\pm 0.8}$
		40%	$69.8_{\pm 3.0}$	$76.1_{\pm 2.9}$	$79.8_{\pm 0.1}$	$84.0_{\pm 0.9}$	$66.5_{\pm 1.0}$	$81.9_{\pm 0.5}$	$\mathbf{84.9}_{\pm \mathbf{0.3}}$
	Amazon	1%	$74.1_{\pm 3.5}$	$88.6_{\pm 3.5}$	$90.4_{\pm 2.0}$	$89.4_{\pm 0.3}$	$83.6_{\pm 1.7}$	$89.0_{\pm 1.2}$	$\mathbf{91.9}_{\pm \mathbf{1.0}}$
		40%	$87.4_{\pm 3.3}$	$90.5_{\pm 1.6}$	$95.8_{\pm 0.1}$	$98.0_{\pm 0.4}$	$89.8_{\pm 1.0}$	$95.4_{\pm 0.6}$	$\mathbf{97.3}_{\pm \mathbf{0.4}}$
	T-Finance	1%	90.2	90.5	90.7	91.1	82.9	90.0	$\mathbf{92.8}_{\pm \mathbf{1.1}}$
		40%	91.4	92.1	91.2	94.3	87.1	88.2	$\mathbf{95.0}_{\pm \mathbf{3.0}}$
	T-Social	1%	65.2	71.2	59.8	88.0	56.3	81.4	$\mathbf{91.9}_{\pm \mathbf{0.6}}$
		40%	71.2	71.8	68.4	**95.2**	56.9	82.5	$93.9_{\pm 0.2}$

Graph Classification

Method	Dataset							
	PROTEINS	NCI1	NCI109	DD	ENZYMES	MUTAG	MUTAGENICITY	FRANKENSTEIN
WKPI (kmeans)	$78.5_{\pm 0.4}$ ▼(6.4)	$87.5_{\pm 0.5}$	$85.9_{\pm 0.4}$ ▼(1.5)	$82.0_{\pm 0.5}$ ▼(13.7)	-	$85.8_{\pm 2.5}$ ▼(14.2)	-	-
WKPI (kcenters)	$75.2_{\pm 0.4}$ ▼(9.7)	$84.5_{\pm 0.5}$ ▼(3.0)	$87.4_{\pm 0.3}$	$80.3_{\pm 0.4}$ ▼(15.4)	-	$88.3_{\pm 2.6}$ ▼(11.7)	-	-
Spec-GN	-	$84.8_{\pm 1.6}$ ▼(2.7)	$83.6_{\pm 0.8}$ ▼(3.8)	-	$72.5_{\pm 5.8}$ ▼(5.9)	-	-	-
Norm-GN	-	$84.9_{\pm 1.7}$ ▼(2.6)	$83.5_{\pm 1.3}$ ▼(3.9)	-	$73.3_{\pm 8.0}$ ▼(5.1)	-	-	-
GWL-WL	$75.8_{\pm 0.6}$ ▼(9.1)	-	-	-	$71.3_{\pm 1.1}$ ▼(7.1)	-	-	$78.9_{\pm 0.3}$
HGP-SL	$84.9_{\pm 1.6}$	$78.5_{\pm 0.8}$ ▼(9.1)	$80.7_{\pm 1.2}$ ▼(6.7)	$81.0_{\pm 1.3}$ ▼(14.7)	$68.8_{\pm 2.1}$ ▼(9.6)	-	$82.2_{\pm 0.6}$	-
DSGCN	$77.3_{\pm 0.4}$ ▼(7.6)	-	-	-	$78.4_{\pm 0.6}$	-	-	-
U2GNN	$80.0_{\pm 3.2}$ ▼(4.9)	-	-	$95.7_{\pm 1.9}$	-	$88.5_{\pm 7.1}$ ▼(11.5)	-	-
NDP	$73.4_{\pm 3.1}$ ▼(11.5)	$74.2_{\pm 1.7}$ ▼(13.3)	-	$72.8_{\pm 5.4}$ ▼(22.9)	$44.5_{\pm 7.4}$ ▼(34.9)	$87.9_{\pm 5.7}$ ▼(12.1)	$77.9_{\pm 1.4}$ ▼(4.3)	-
ASAP	$74.2_{\pm 0.8}$ ▼(10.7)	$71.5_{\pm 0.4}$ ▼(16.0)	$70.1_{\pm 0.6}$ ▼(17.3)	$76.9_{\pm 0.7}$ ▼(18.8)	-	-	-	$66.3_{\pm 0.5}$ ▼(12.6)
EvoG	-	-	-	-	55.7 ▼(22.7)	**100.0**	-	-
ET (Ours)	$\mathbf{90.3}_{\pm 0.7}$ ▲(5.4)	$\mathbf{90.1}_{\pm 0.1}$ ▲(2.6)	$\mathbf{90.5}_{\pm 0.1}$ ▲(3.1)	$\mathbf{95.9}_{\pm 0.8}$ ▲(0.2)	**99.8** ▲(21.4)	$96.6_{\pm 0.2}$ ▼(3.4)	$\mathbf{98.7}_{\pm 0.1}$ ▲(16.5)	$\mathbf{99.8}_{\pm 0.1}$ ▲(20.9)