



中国科学院大学

University of Chinese Academy of Sciences

# Transformer-based model for symbolic regression via joint supervised learning

Wenqiang Li

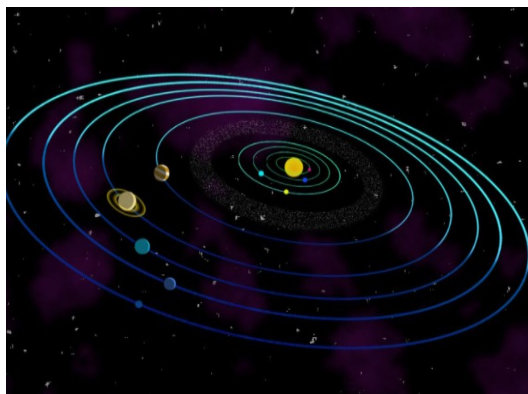
Institute of Semiconductors, Chinese Academy of Sciences



# What is Symbolic Regression?

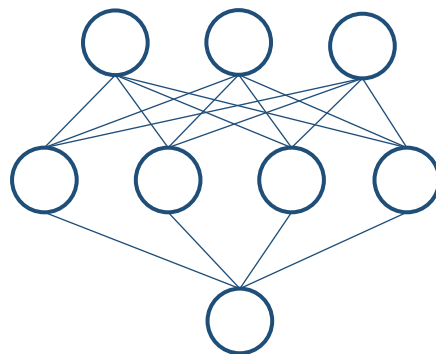
Given a dataset  $(X, y)$ , where each feature  $X_i \in \mathbb{R}^n$  and target  $y_i \in \mathbb{R}$ , the goal of symbolic regression is to identify a function  $f$  (i.e.,  $y \approx f(X): \mathbb{R}^n \rightarrow \mathbb{R}$ ) that best fits the dataset, where the functional form of  $f$  is a short closed-form mathematical expression.

The law of  
universal gravitation



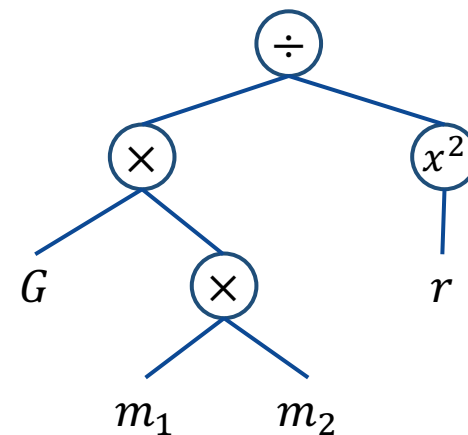
Deep Learning

$$F = \text{Relu}(W \text{Relu}(WX \dots) + b)$$



Symbolic Regression

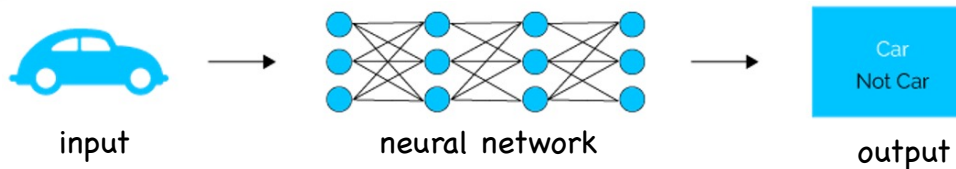
$$F = \frac{Gm_1m_2}{r^2}$$





# Deep Learning Vs Symbolic Regression

## Deep Learning



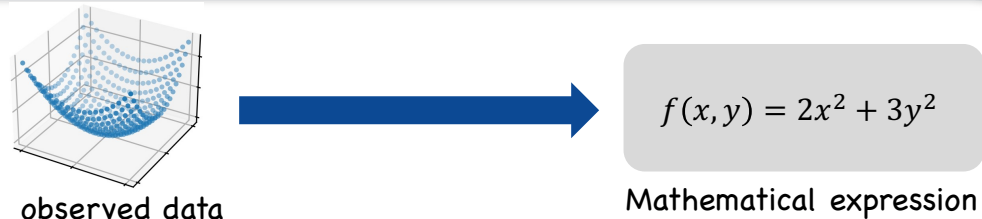
**DALL·E 2**  
DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.

**ChatGPT**

Powerful fitting capability



## Symbolic Regression



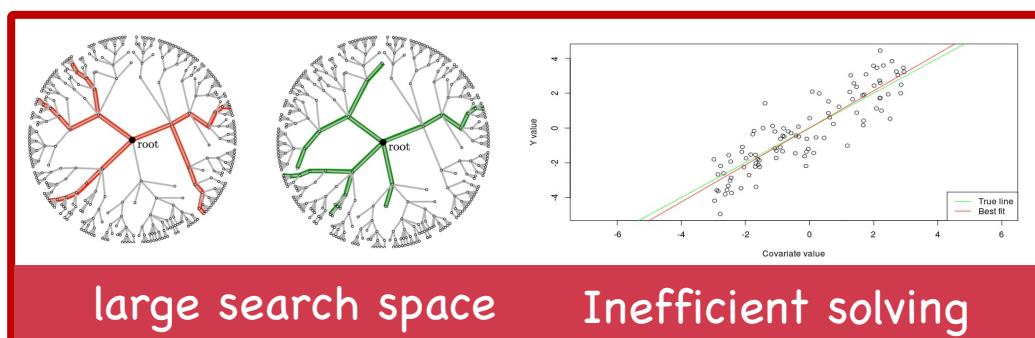
Interpretable      Good generalization

*Feynman*  
THE FEYNMAN LECTURES ON PHYSICS

$$L = \frac{h\omega^3}{\pi^2 c^2 (e^{h\omega/k_b T} - 1)}$$

$$F = \frac{Gm_1 m_2}{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

$$r = \frac{a(1 - e^2)}{1 + e \cos(\theta_1 - \theta_2)}$$

$$\frac{d\sigma}{d \cos \theta} = \frac{\pi \alpha^2 h^2}{m^2 c^2} \left( \frac{\omega'}{\omega} \right)^2 \left( \frac{\omega'}{\omega} + \frac{\omega}{\omega'} - \sin^2 \theta \right)$$


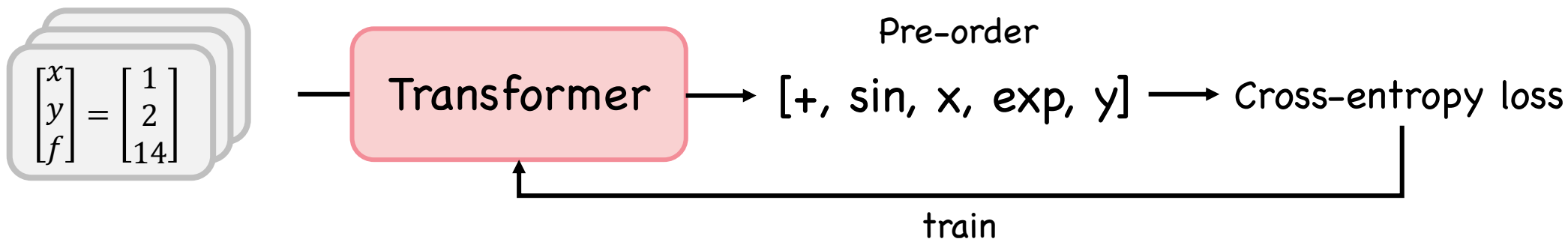
# Deep Symbolic Regression

## Typical approach

1. Encode data points
2. Predict the pre-order traversal
3. Compute cross-entropy loss

SymbolicGPT [Valipour et al., 2021]

NeSymRes [Biggio et al., 2021]



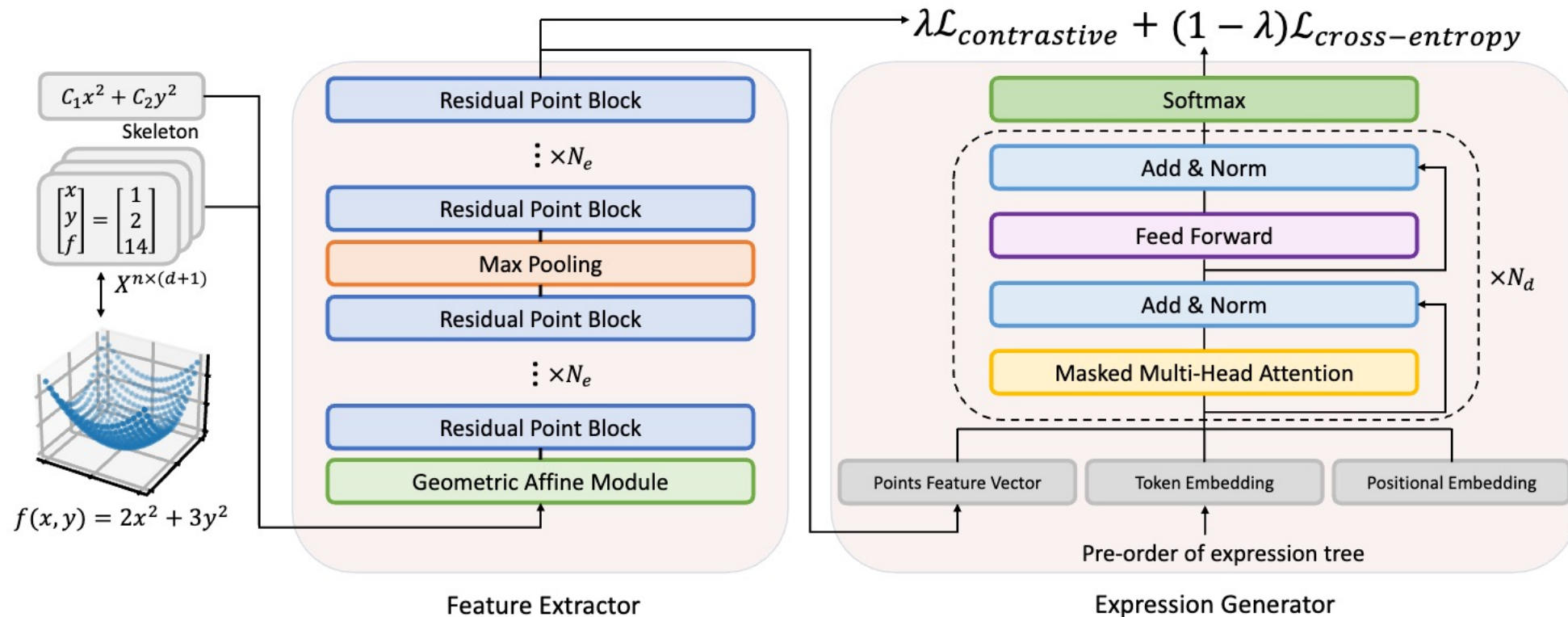
## Shortcomings:

👎 Low-quality feature extraction from data points

👎 Skeletons provide ill-defined supervision



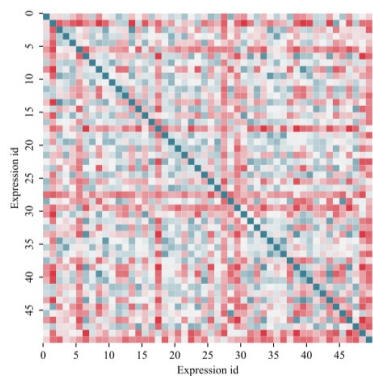
# Our Approach



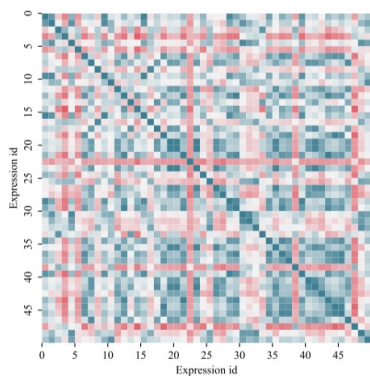
## Advantages :

- 👍 Use a pure residual MLP feature extractor for extracting rich features of data targeting SR tasks, which aids the expression generator in producing more correct expression skeletons
- 👍 Train with a joint supervised learning mechanism combining supervised contrastive learning (CL), which alleviates the ill-posed problem effectively

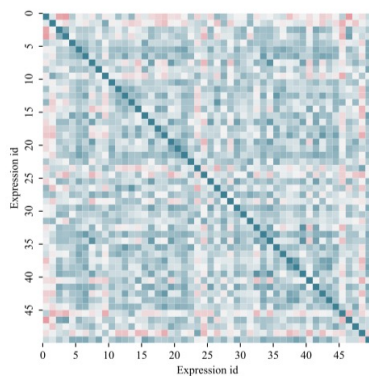
# Performance



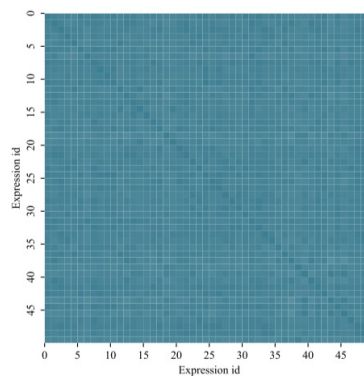
(a) SymbolicGPT



(b) NeSymReS



(c) Ours w/o CL



(d) Ours w/ CL

Method	Recovery rate (%)
SymbolicGPT	$50.3 \pm 1.7$
NeSymReS	$63.4 \pm 1.1$
Ours w/o CL	$69.7 \pm 0.9$
Ours w/ CL	<b><math>75.2 \pm 1.3</math></b>

Recovery rate of eq. skeleton

$$1.2 \sin(x_1) + 1.5 \cos(x_2) + 0.6 \rightarrow \text{data1}$$

$$2.2 \sin(x_1) + 1.2 \cos(x_2) + 0.1 \rightarrow \text{data2}$$

$$3.6 \sin(x_1) + 0.9 \cos(x_2) + 1.2 \rightarrow \text{data3}$$

⋮

Eq. with same skeleton

Our Model

$$a \sin(x_1) + b \cos(x_2) + c$$

Pred. correct skeleton 😊

BFGS

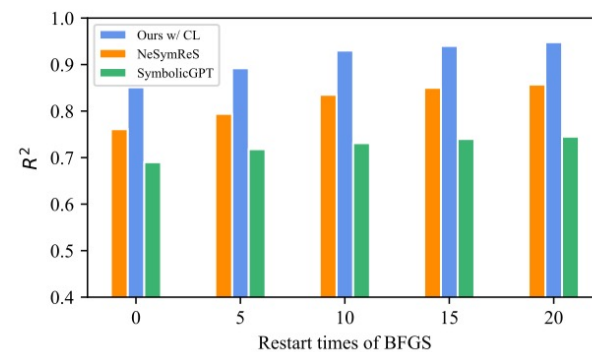
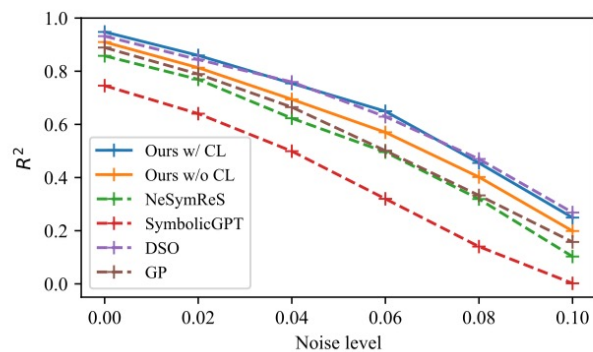
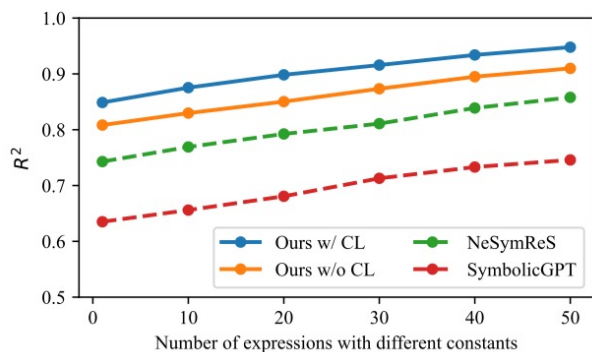


中国科学院大学

University of Chinese Academy of Sciences

# Performance

	Ours	SymbolicGPT	NeSymReS	DSO	GP
Benchmark	$R^2 \uparrow$	$R^2 \uparrow$	$R^2 \uparrow$	$R^2 \uparrow$	$R^2 \uparrow$
Nguyen	<b>0.99999</b>	0.64394	0.97538	0.99489	0.89019
Constant	<b>0.99998</b>	0.69433	0.84935	0.99927	0.90842
Keijzer	<b>0.98320</b>	0.59457	0.97500	0.96928	0.90082
R	<b>0.99999</b>	0.71093	0.99993	0.97298	0.83198
AI-Feynman	<b>0.99999</b>	0.64682	<b>0.99999</b>	<b>0.99999</b>	0.92242
SSDNC	<b>0.94782</b>	0.74585	0.85792	0.93198	0.88913
Overall avg.	<b>0.98850</b>	0.67274	0.94292	0.97806	0.89049







中国科学院大学

University of Chinese Academy of Sciences

Thank you !