Jye Sawtell-Rickson
m1361019@cgu.edu.tw

Chang Gung University

13th March, 2026

LoopViT: Scaling Visual ARC with Looped Transformers

HKUST, CASIA, UC Santa Cruz

Outline

m13610190@cgu.edu.tw

Outline

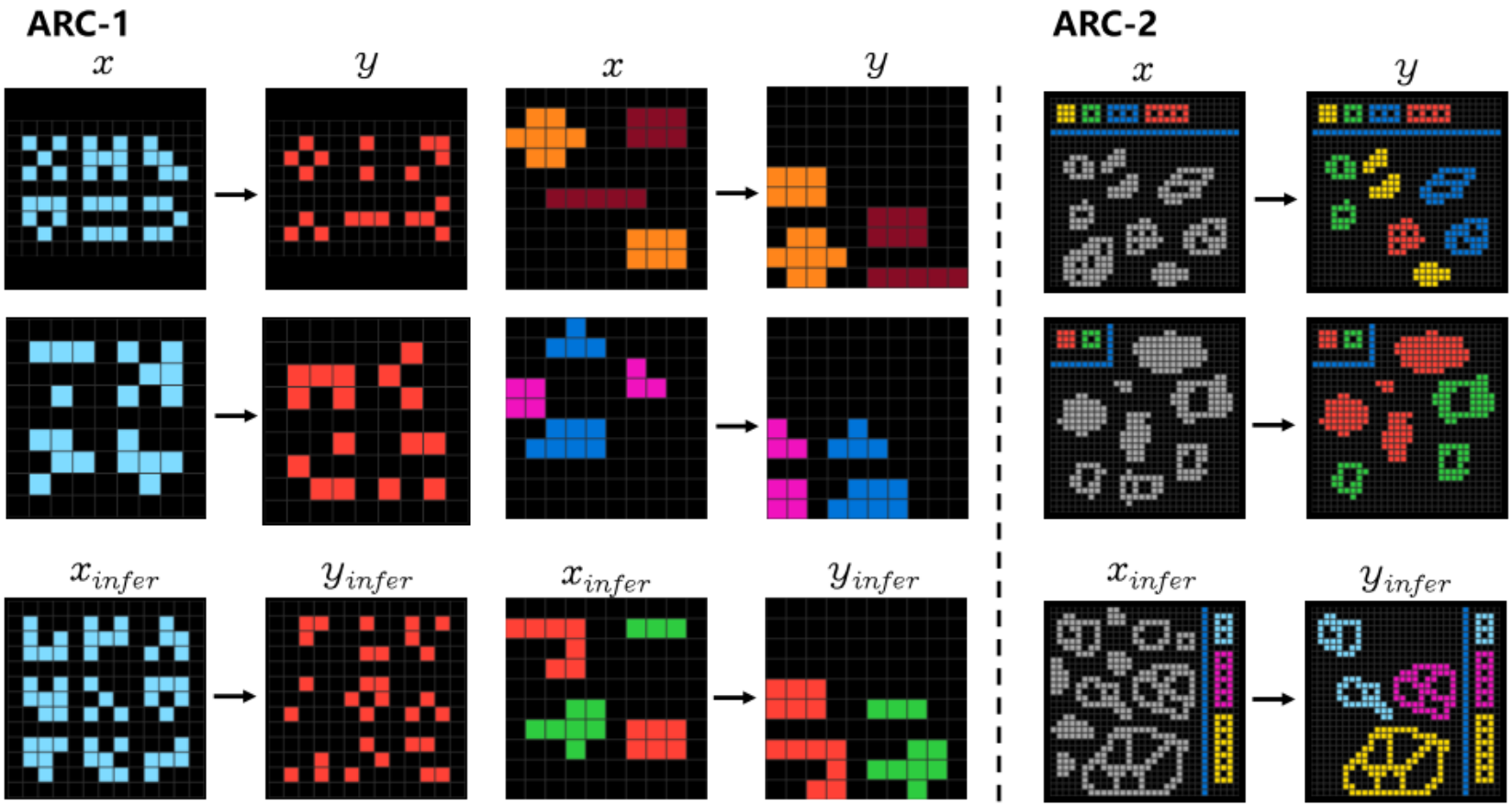
m13610190@cgu.edu.tw

Problem Definition: ARC-AGI

- **The Core Problem:** Abstract visual inductive reasoning.
- **The Benchmark:** Abstraction and Reasoning Corpus (ARC-AGI).
- **Input:** 2–4 "demonstration" pairs of 2D visual grids (integers mapped to colors) + 1 novel test input.
- **Output:** The generated 2D grid for the test input.
- **Objective:** Deduce the hidden algorithmic rule (rotation, physics, mapping) from the demonstrations and apply it perfectly.

m13610190@cgu.edu.tw

ARC-AGI Examples



m1361019@cgu.edu.tw

Related Work & Limitations

Many approaches to solve this problem ignore the fact the visual inductive bias.

- LLMs are the most common example.
- With no visual prior, these models 'think' in one dimension, despite it being a 2D problem.

Some visual models exist, but they are single-pass feed-forward networks.

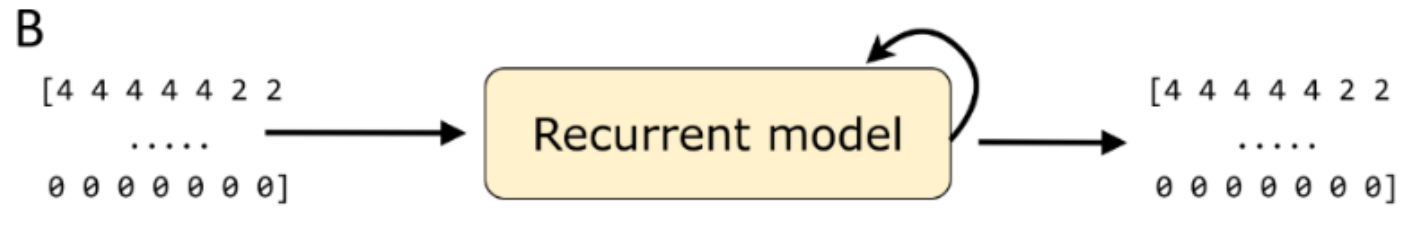
- Visual ARC ensemble achieved ∼60.4% using 73M parameters.
- Computational depth is strictly bound to parameter size. More thinking requires more layers and memory.
- Feed-forward networks lack the iterative, algorithmic loop humans use to form, simulate, and test hypotheses.

m13610190@cgu.edu.tw

Weight-Tied Recurrence

Recurrent models have shown clear benefits from Deep Equilibrium models to TRM by increasing the depth of computation. Decouples reasoning depth from model capacity. Why is this good?

- With a fixed paramater budget, can get much wider and more expressive layers (though this isn't as true for inference cost)
- Learning an algorithm vs. a pipeline, e.g. CNNs different stages of functions
- Forces strong regularisation
- Enables early stopping which is not possible in a standard transformer



m1361019@cgu.edu.tw

Outline

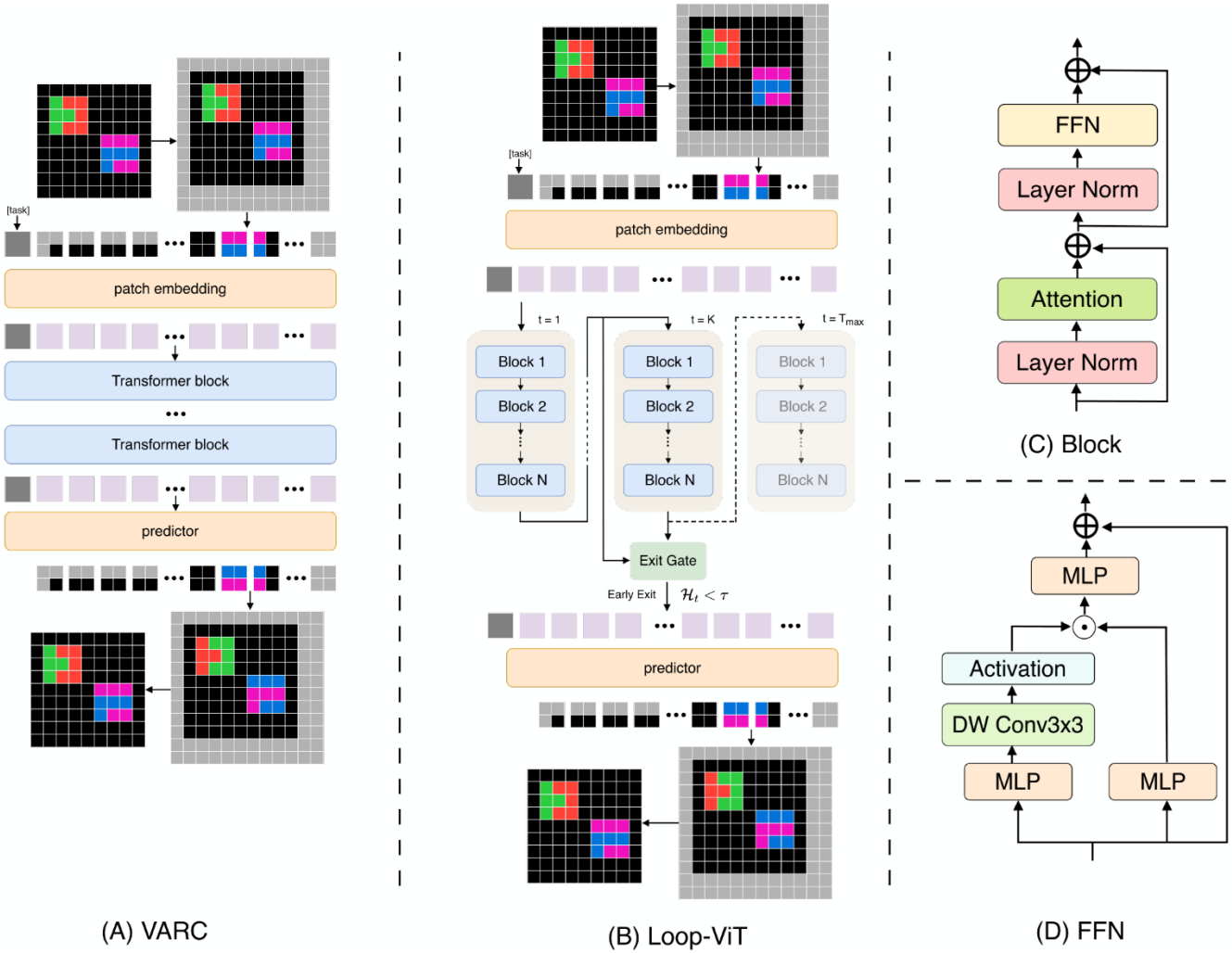
m13610190@cgu.edu.tw

The Innovation: LoopViT Architecture

- **Weight-Tied Hybrid Block:** Combines local convolutions (spatial reasoning) with global self-attention (semantic patterns).
- **Parameter-Free Dynamic Exit:** Measures predictive entropy at each step. Halts inference when the internal state "crystallizes" (uncertainty drops below a threshold).

m13610190@cgu.edu.tw

LoopViT Architecture



m1361019@cgu.edu.tw

Early Stopping Mechanism

Parameter-free Dynamic Exit mechanism:

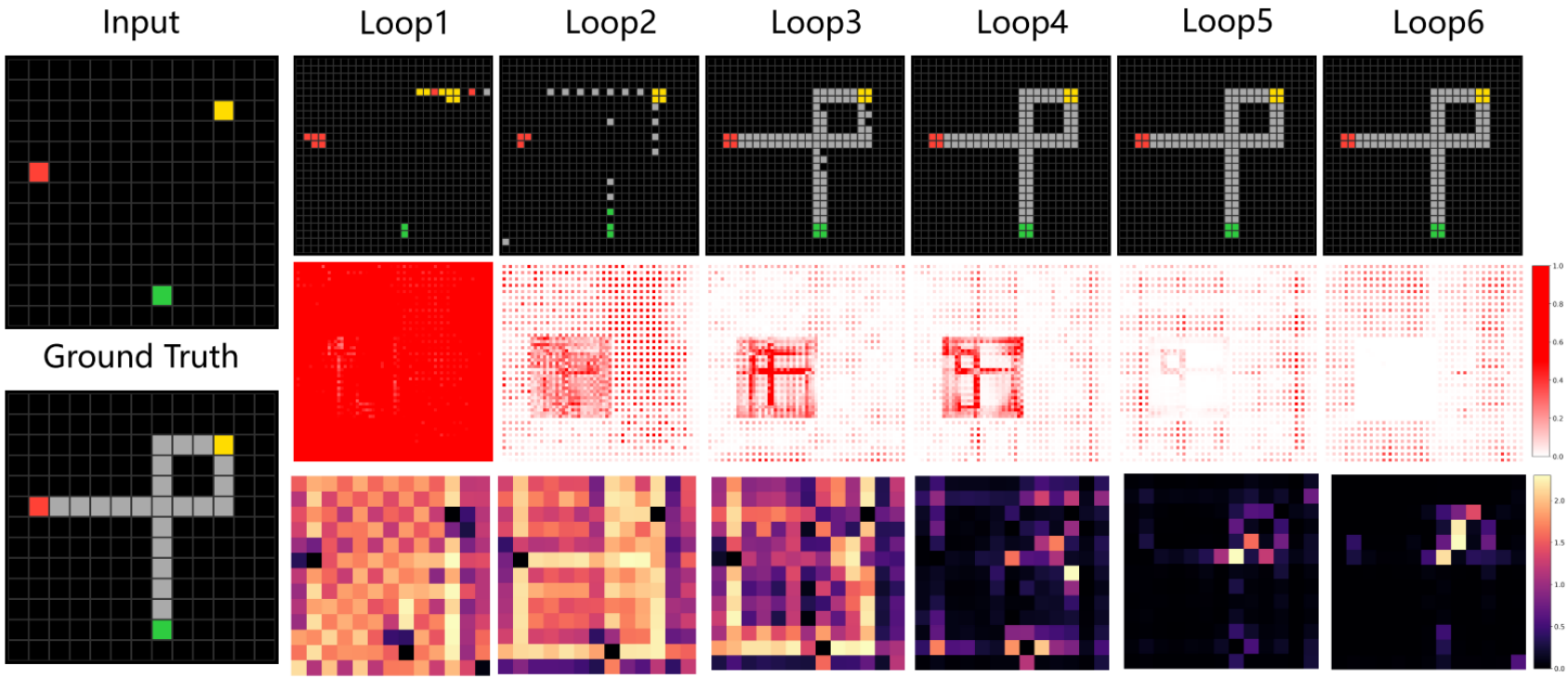
- At each loop, the model evaluates its 'predictive entropy', essentially measuring how uncertain it is about its current answer. Uses the average pixelwise Shannon entropy:

$$\mathcal{H}_t = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} P_{t,i}(c)\log P_{t,i}(c)$$

- Looping continues until its internal state moves into a low-uncertainty attractor. Once the entropy drops below a certain threshold, the model halts inference and outputs the answer.
- Easier tasks need less compute as entropy stabilises quicker, saving compute.

m13610190@cgu.edu.tw

Iterative Refinement Example



m13610190@cgu.edu.tw

Outline

m13610190@cgu.edu.tw

Key Results

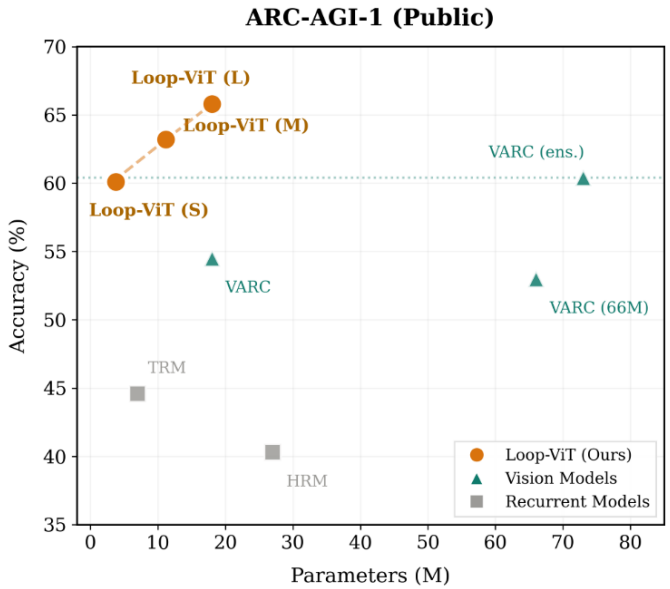
- Highly convincing empirical results on a benchmark designed to resist memorization.
- Shows promising scaling.

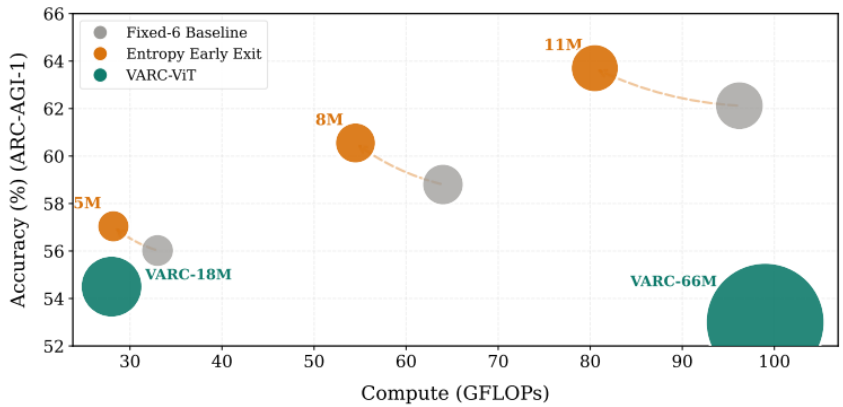
Table: Comparison of Model Performance on the ARC-AGI-1 Benchmark

Model	Architecture	Parameters	Accuracy
Human Average	Biological	N/A	60.2%
VARC	Feed-Forward	73 M	60.4%
LoopViT (Tiny)	**Recurrent**	**3.8 M**	**60.1%**
LoopViT (Standard)	**Recurrent**	**18 M**	**65.8%**

m13610190@cgu.edu.tw

Results Figures





m13610190@cgu.edu.tw

Outline

m13610190@cgu.edu.tw

Known Limitations

- **Confident Failure:** If the model hallucinates a wrong rule early, its entropy may still collapse, causing a premature exit and a highly confident wrong answer.
- **Limits of Pure Vision:** Inherently struggles with tasks requiring strict mathematical, counting, or deep symbolic logic that pixel convolutions cannot easily capture.
- **Peer Review:** Awaiting formal conference peer review.
- **Reproducibility:** Code/data are public. The 3.8M-18M parameter size makes local reproduction theoretically possible, but they also trained on 8 GPUs...

m13610190@cgu.edu.tw

Ideas for Improvement

- **Reduce Cost and Size:** apply INT8 quantization to the recurring block to slash memory bandwidth during iterative loops.
- **Sampling and Search:** rather than using just a single 'train of thought', sample from the model outputs and perform a search process across the different steps, ideally leading to more attractor states.
- **Good Latent Representation:** utilise in neuro-symbolic methods as a good latent representation of the state used to predict a function.

m13610190@cgu.edu.tw

Summary: LoopViT

- **The Problem:** Feed-forward networks inefficiently scale parameter size to increase reasoning depth on ARC-AGI.
- **The Solution:** LoopViT decouples depth from capacity using *weight-tied recurrence* and an entropy-based *dynamic exit*.
- **The Result:** A 18M parameter model (65.8%) outperforms massive 73M parameter ensembles (∼60.0%).
- **The Future:** The next frontier lies in marrying this continuous iterative perception with strict symbolic verification or search.

Thank You! — Questions?

m13610190@cgu.edu.tw