

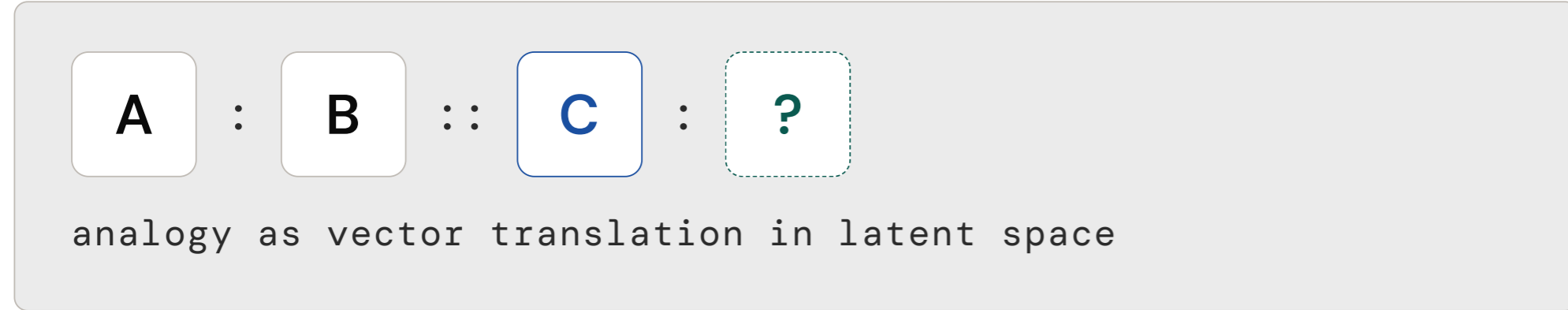
On the Geometry of Analogical Reasoning in Latent Space

Oleg Dats · Ukrainian Catholic University, Lviv · dats.pn@ucu.edu.ua

INTRODUCTION

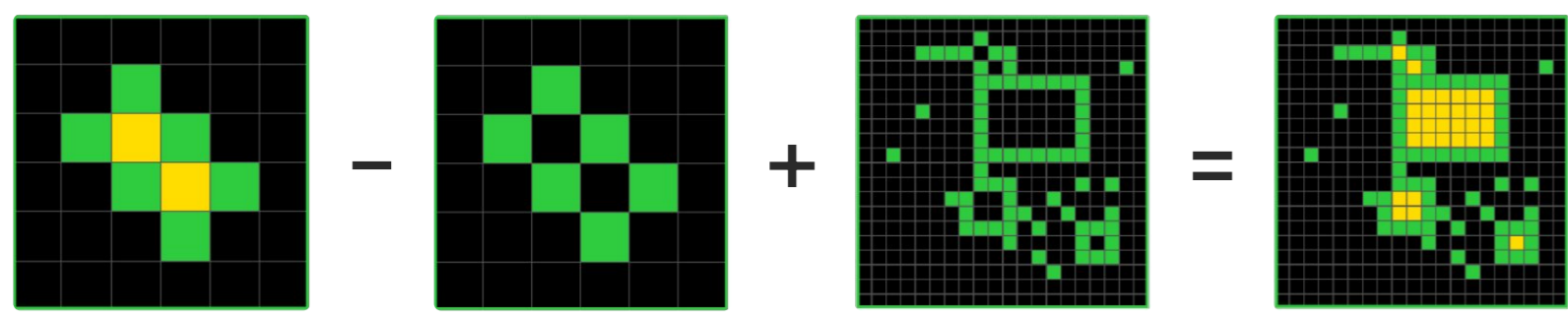
Can Analogical Reasoning be realised as an explicit geometric primitive?

Modern transformers encode relations *implicitly* across attention weights, making them hard to isolate, analyse, or reuse.



king - man + woman \approx queen

ARC BENCHMARK EXAMPLES



METHOD

Analogical reasoning = estimating a shared translation \hat{x} .

PER-PAIR DIFFERENCES

$$\delta_i = \text{mean}(b_i - a_i)$$

CLOSED-FORM LEAST-SQUARES ESTIMATE

$$\hat{x} = (1/n) \sum_i \delta_i$$

APPLY AND TRAIN

$$L = \text{CE}(\text{decode}(a_i + \hat{x}), b_i)$$

*No auxiliary latent losses

GPT BASELINE

$$O(n^2)$$

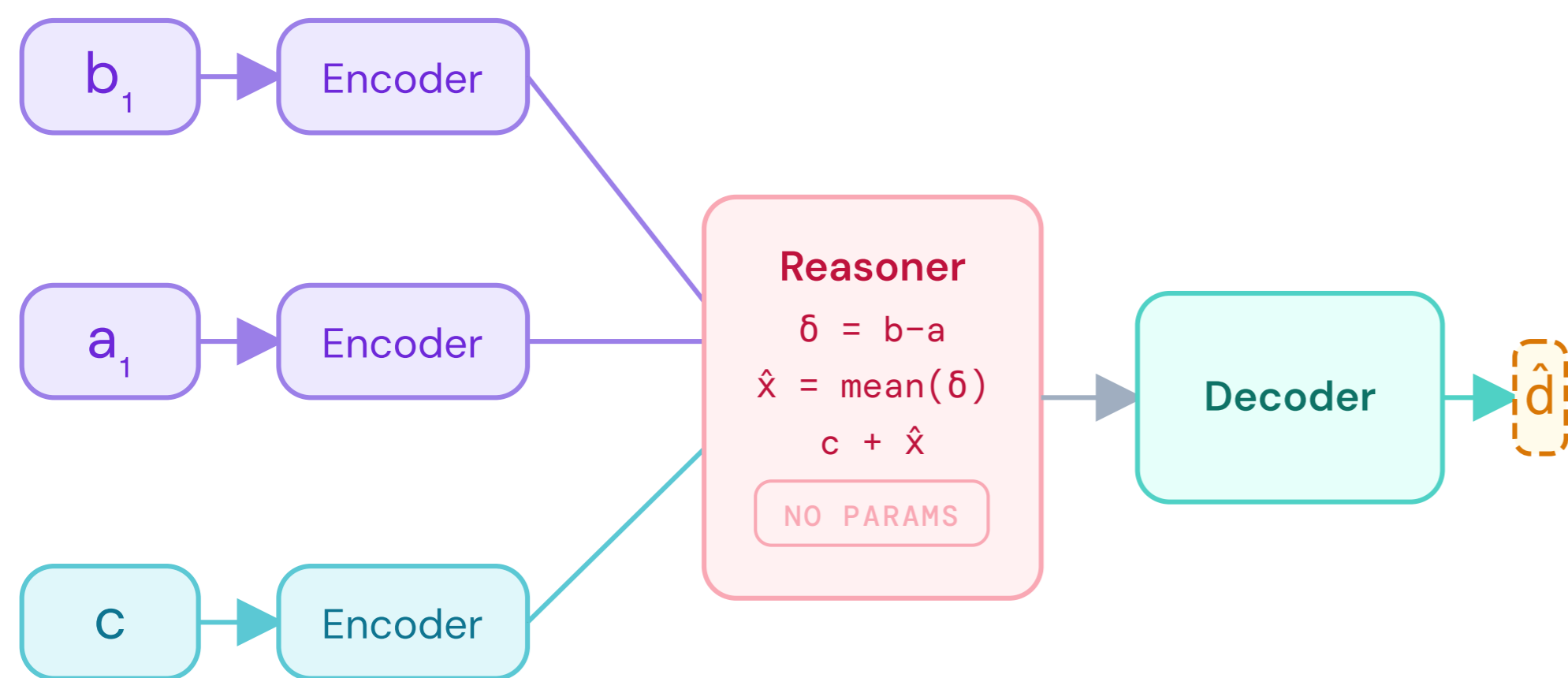
pairwise attention

ERD (OURS)

$$O(n)$$

independent encoding

ERD ARCHITECTURE



BERT-style
no pretraining

least-squares
estimator

GPT-style
cross-attn on grid
position

Each grid encodes independently (parallel). The reasoner aggregates into a single \hat{x} . The decoder never sees example pairs. All reasoning happens in the translation.

RESULTS - ARC BENCHMARK

All models trained from scratch on RE-ARC, evaluated on 400 ARC test tasks.

MODEL	STANDARD	+ TTT	COMPUTE
GPT baseline	3%	19%	~4x
ERD (ours)	6%	26%	1x base
ERD 512d/12L	8%	33%	~4x

2x
more tasks (standard)

1.4x
more tasks (TTT)

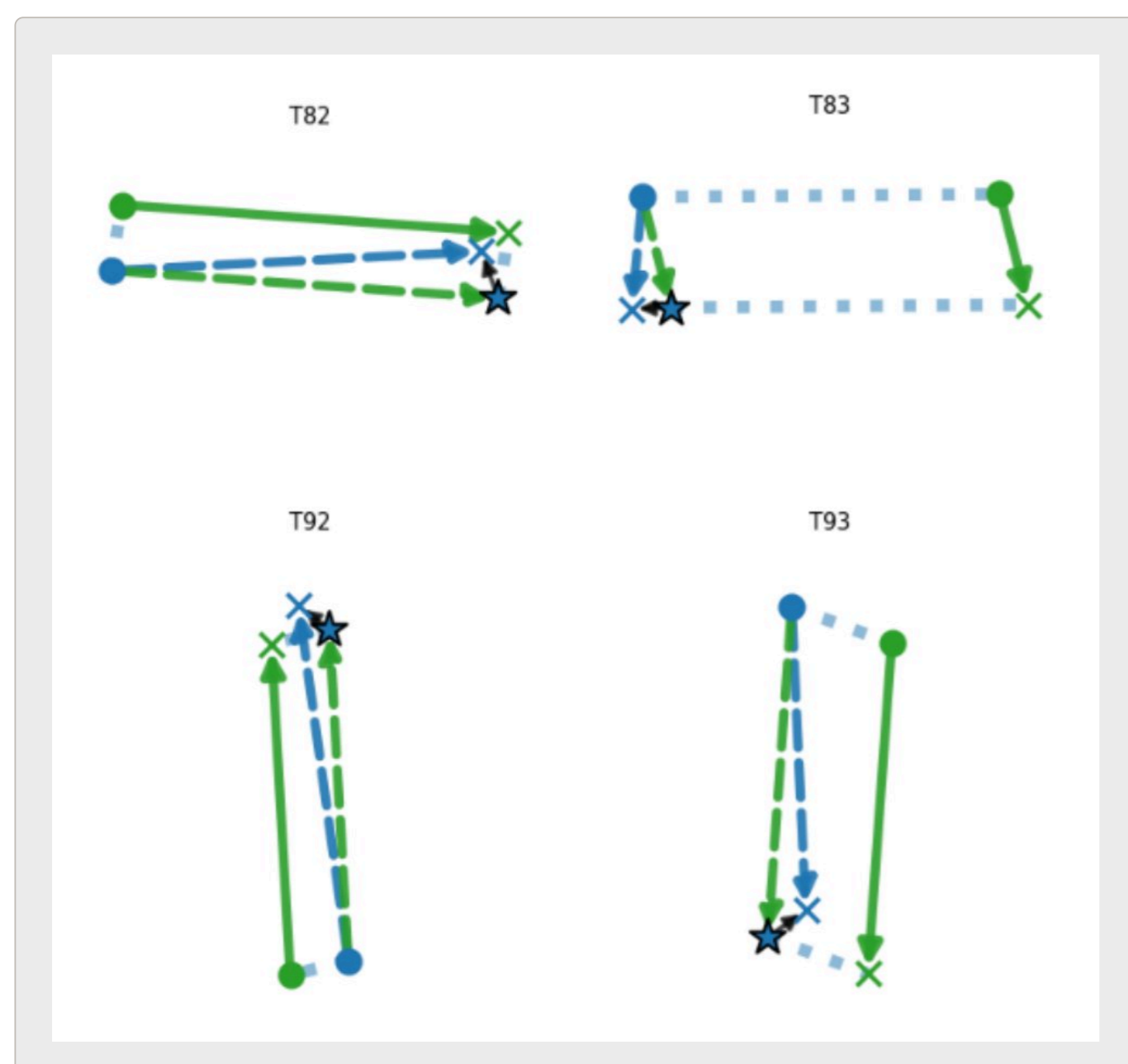
4x
less compute

Ablation: Global task translation is the key component.

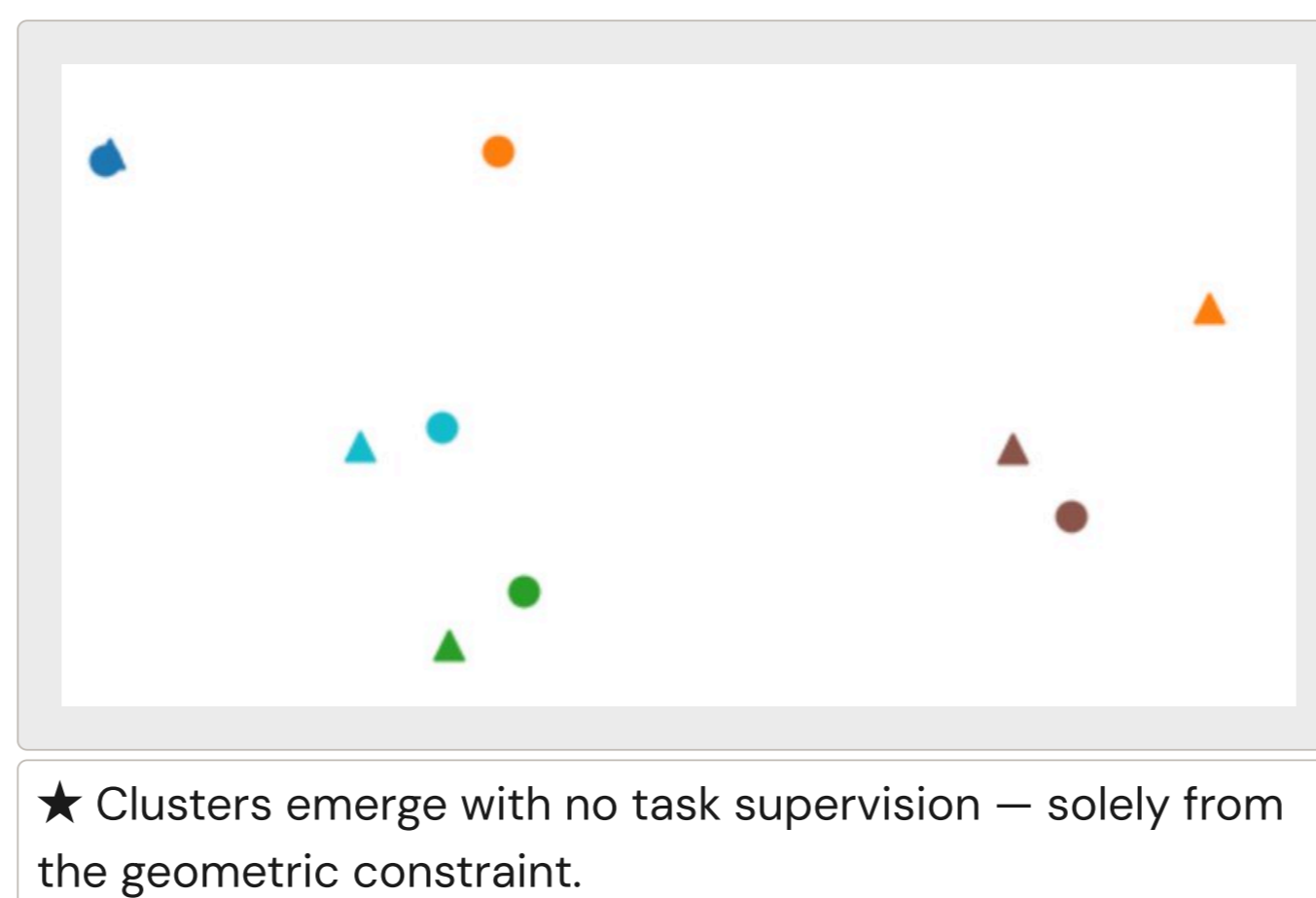
	Standard	+TTT
token-wise + 1D	3%	15.5%
+ global mean	5%	24.5%
+ 2D pos (ERD)	6%	26%

LATENT SPACE GEOMETRY

(A) PARALLELOGRAM STRUCTURE $B - A = D - C$



(B) PCA OF Δ VECTORS BY TASK



(C) ALIGNMENT METRICS

SAME-TASK COS SIM

0.92

DIFF-TASK COS SIM

0.01