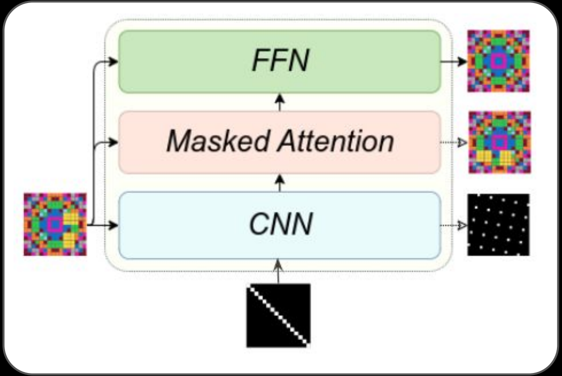
Infusing Lattice Symmetry Priors in Neural Networks for Sample-Efficient Abstract Geometric Reasoning

Speaker: Mattia Atzeni



Authors

Mattia Atzeni, *EPFL*

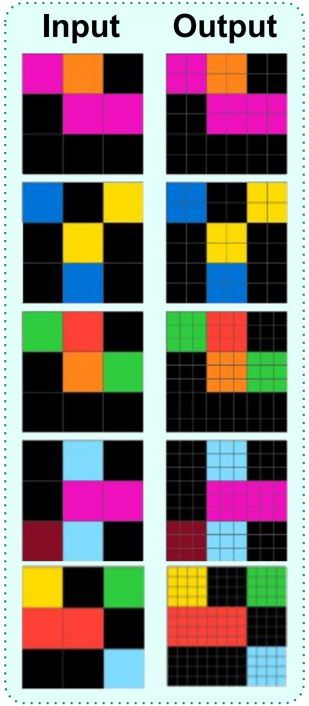
Mrinmaya Sachan, *ETH Zurich*

Andreas Loukas, *Prescient Design, gRED, Roche*

EPFL ETH Roche

Abstraction and Reasoning Corpus (ARC)

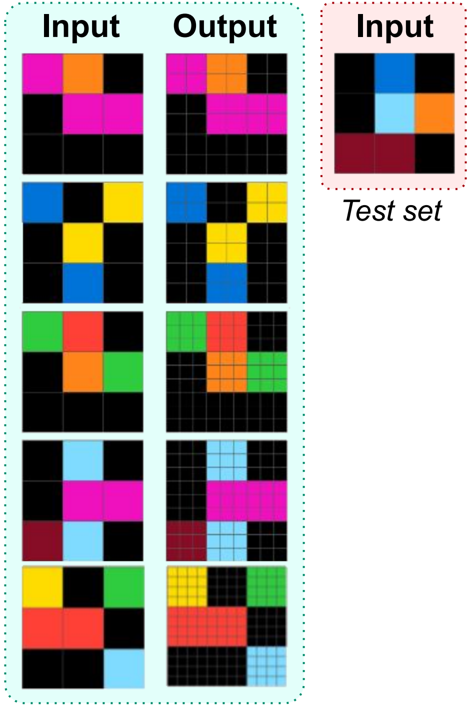
ICML
International Conference
On Machine Learning



Training set

Chollet 2019, *On the measure of intelligence*.

Abstraction and Reasoning Corpus (ARC)

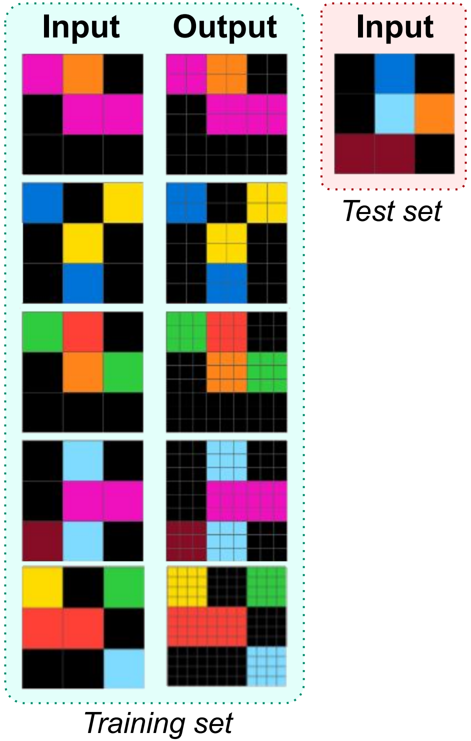


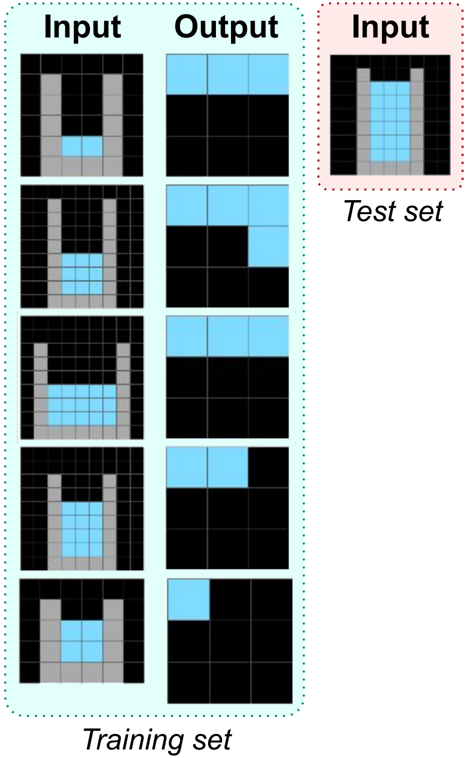
Test set

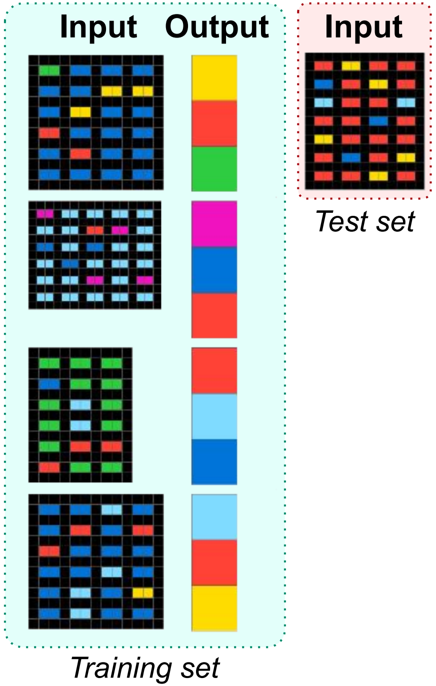
Training set

Chollet 2019, *On the measure of intelligence*.

Abstraction and Reasoning Corpus (ARC)







Chollet 2019, *On the measure of intelligence*.

Core Knowledge Priors

ARC: intelligence is the ability to acquire a new skill efficiently from few examples.

Tasks built upon a set of priors close to innate human priors.

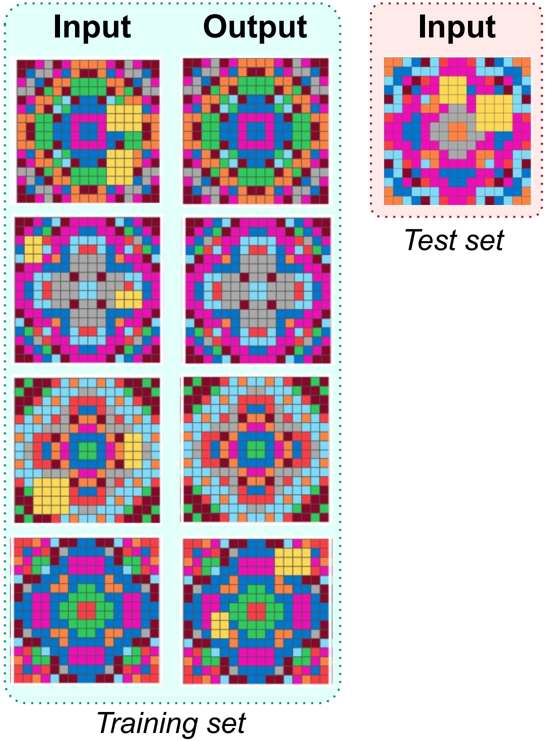
Known as *Core Knowledge* priors in *developmental science* and grouped into:

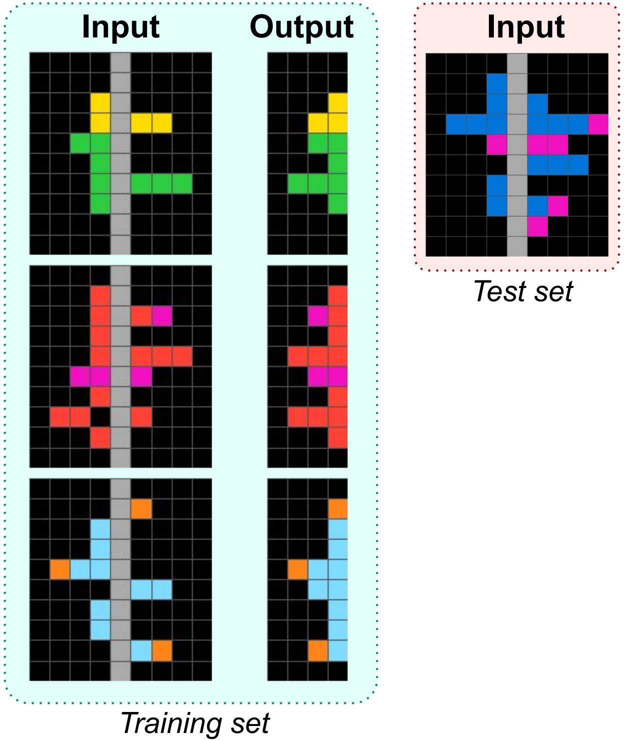
- Geometry and topology
- Objectness
- Math and counting
- Goal directedness

This work: how to infuse geometry and topology priors in transformers

Spelke and Kinzler 2007. *Core knowledge*. Developmental Science.
Chollet 2019, *On the measure of intelligence*.

Geometry and Topology Priors





EPFL ETH Roche

Group-Action Learning

ICML International Conference On Machine Learning

We regard a task as an instantiation of the following general problem:

$$y = f(\mathrm{g} \circ x, x), \quad \mathrm{g} = g(x) \in \mathsf{G}$$

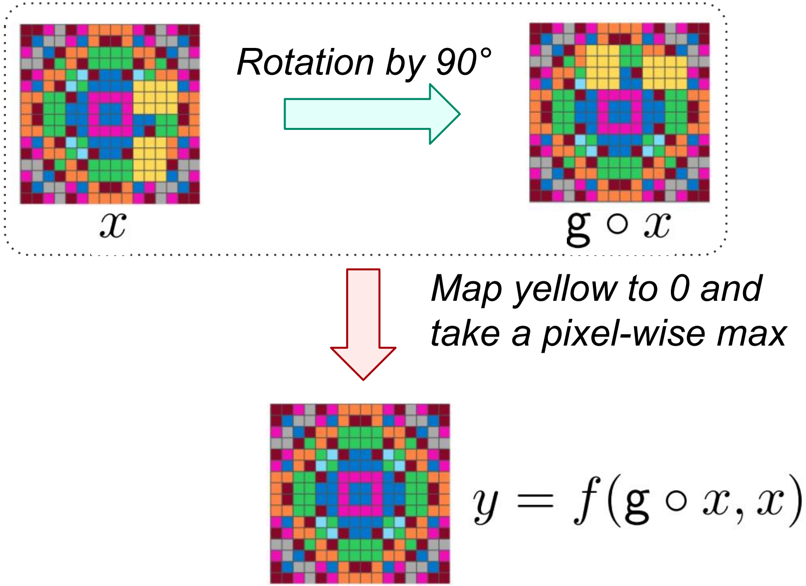
where:

- x and y are input and output examples
- f and g are learnable functions
- g is a group action from group G
- $\mathrm{g} \circ x$ denotes the application of action g to x

Group-Action Learning

As we focus on geometry and topology priors, we consider the group of **lattice transformations**, which includes the following operations:

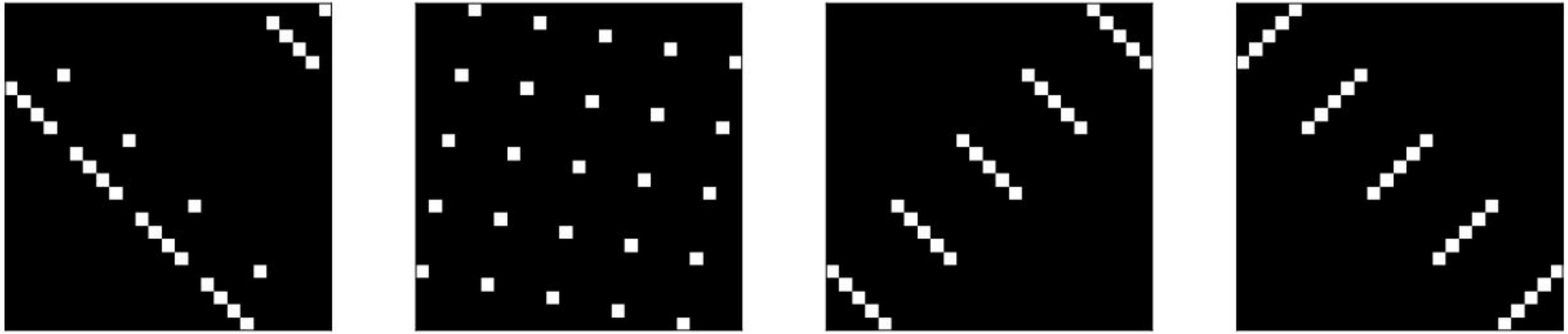
- **Translation**
- **Rotation**
- **Reflection**
- **Scaling**



Attention Masks for Lattice Transformations

In theory, self-attention can learn to express any lattice transformation, but it requires a significant amount of training data to do so.

Theorem 1. Any lattice symmetry action can be expressed by a binary attention mask.

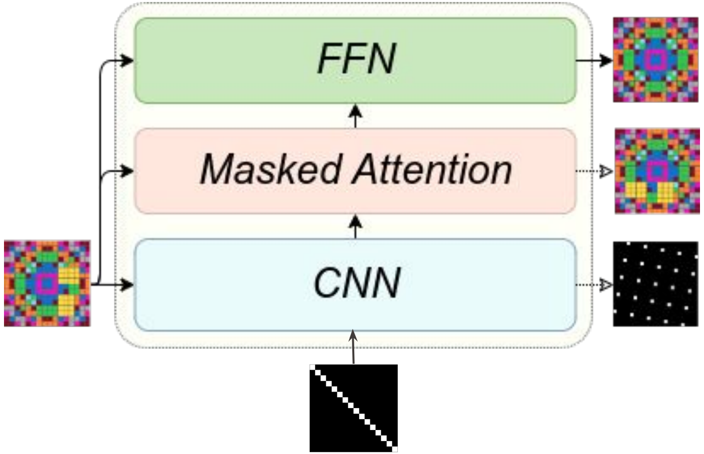


Translation by (1, 1) *Rotation by 90°* *Vertical reflection* *Horizontal reflection*

LatFormer

Theorem 2. These masks can be obtained by applying convolutions on the identity matrix.

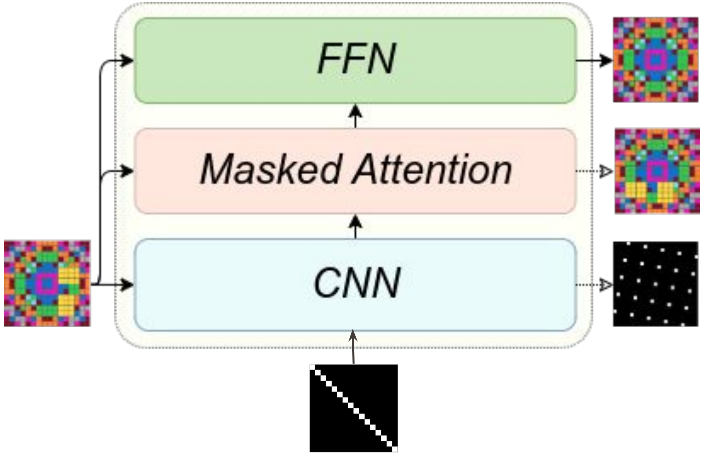
Our approach: use a CNN to generate attention masks to help the standard transformer layer learn lattice transformations.



LatFormer

Theorem 2. These masks can be obtained by applying convolutions on the identity matrix.

Our approach: use a CNN to generate attention masks to help the standard transformer layer learn lattice transformations.



$$A = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right) \odot M$$

$$\mathrm{MaskedAttention}(Q, K, V; M) = \frac{A}{A \cdot \mathbf{1}_{n_K} \mathbf{1}_{n_K}^\top} V$$

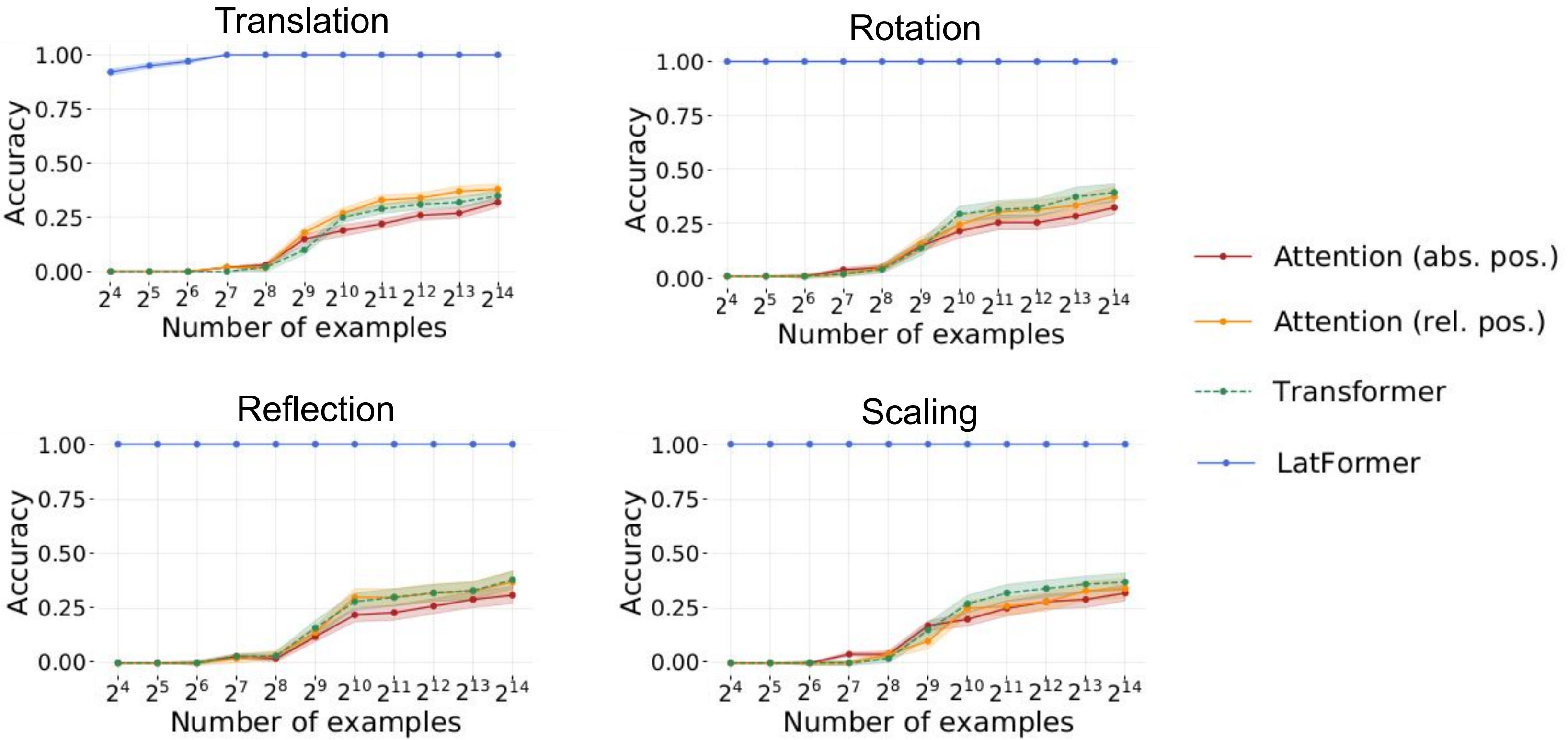
EPFL

ETH

Roche

LatFormer learns lattice group actions sample efficiently

ICML
International Conference
On Machine Learning



EPFL
ETH
Roche

LatFormer on ARC tasks requiring geometric priors

ICML
International Conference On Machine Learning

We evaluated our method on the ARC tasks requiring geometric priors.

	Translate	Rotate + Translate	Reflect + Translate	Scale + Translate
CNN	0.019	0.000	0.000	0.000
Attention (abs. pos.)	0.019	0.000	0.023	0.000
Attention (rel. pos.)	0.019	0.000	0.023	0.000
PixelCNN	0.019	0.000	0.000	0.000
Transformer	0.038	0.000	0.045	0.000
Differentiable Neural Computer	0.038	0.000	0.045	0.000
Transformer + data augmentation	-	0.200	0.184	0.091
LatFormer	0.365	**0.800**	0.591	0.545
Search over hand-crafted DSL	**0.673**	0.400	**0.614**	**0.727**

LatFormer is the only neural network that approaches the accuracy of SotA methods like searching over an hand-crafted domain-specific language (DSL).

EPFL ETH Roche

Comparison with neural program synthesis on LARC

International Conference On Machine Learning

LARC also provides natural language descriptions of the ARC tasks.

	Translate	Rotate + Translate	Reflect + Translate	Scale + Translate
LARC (IO)	0.17	0.00	0.42	**1.00**
LARC (IO + NL)	0.17	0.00	0.42	**1.00**
LARC (IO + NL pseudo)	0.25	0.00	0.42	**1.00**
LatFormer	**0.33**	**1.00**	0.50	**1.00**
LatFormer + NL	**0.33**	**1.00**	**0.58**	**1.00**

LatFormer outperforms SotA neural program synthesis methods with access to both input-output pairs (IO) and natural language descriptions (NL)

EPFL
ETH
Roche

Conclusion

ICML
International Conference
On Machine Learning

LatFormer infuses **lattice symmetry** priors in attention and helps to **learn basic geometric transformations sample efficiently**.

- It achieves **SotA results** on **abstract geometric reasoning** tasks.
- It can be applied in conjunction with standard transformers as a module in bigger models for a broader range of applications.

Infusing Lattice Symmetry Priors in Attention Mechanisms for Sample-Efficient Abstract Geometric Reasoning

Mattia Atzeni [1] **Mrinmaya Sachan** [2] **Andreas Loukas** [3]