Comparative Analysis of Convolutional and Transformer Architectures in Go Policy Networks

Antoni Hanke

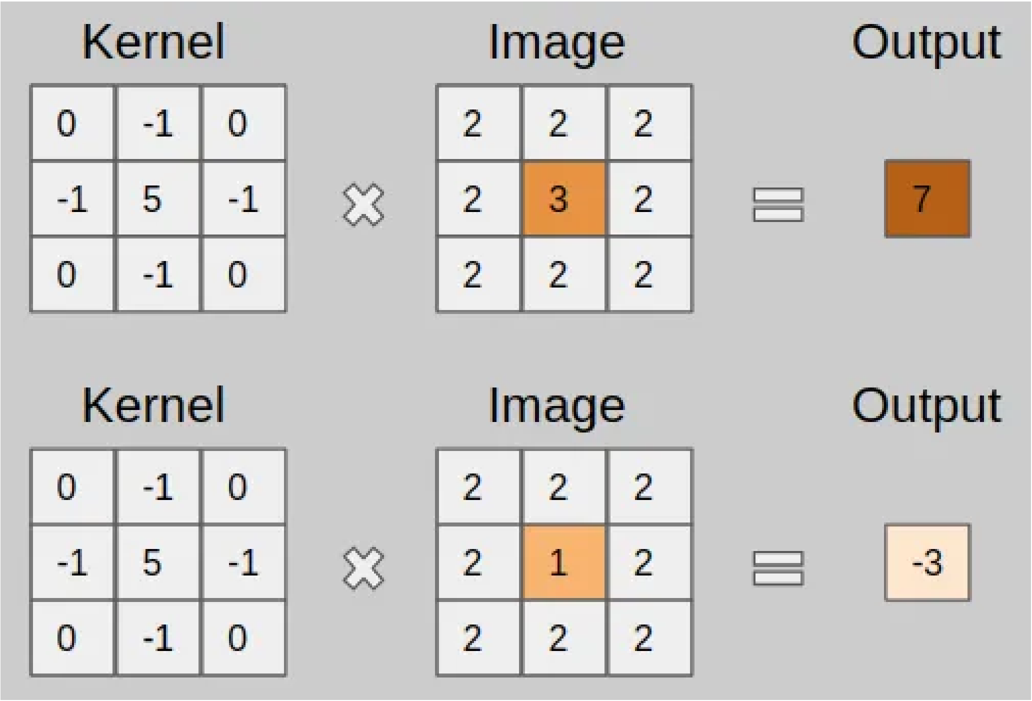
Wydział Matematyki, Informatyki i Mechaniki

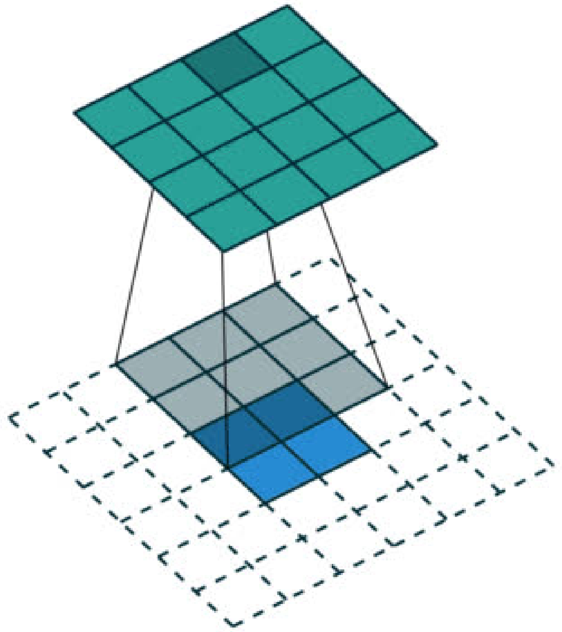
Uniwersytet Warszawski

Recap: Convolutional Neural Networks

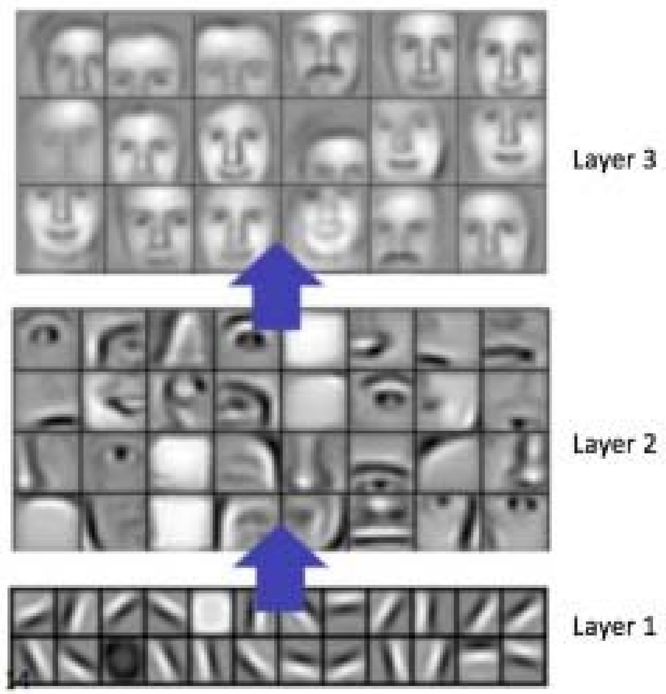
Computer way to extract features

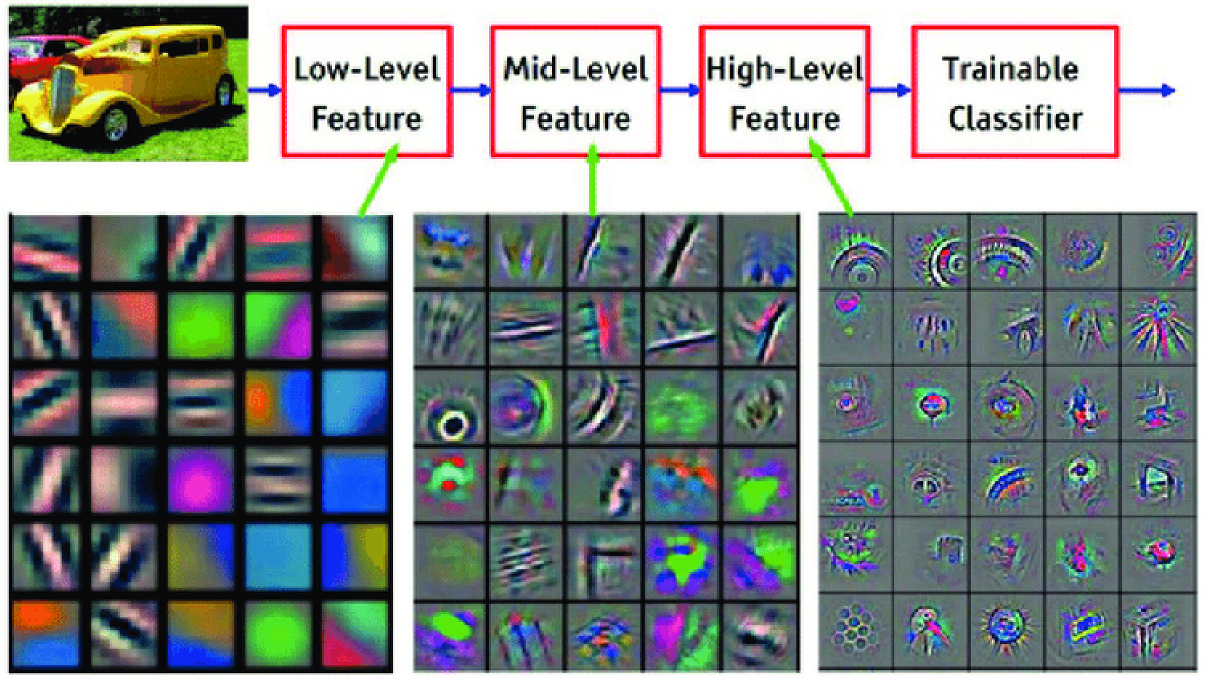
- Kernel multiplication allows to find basic features





- Applying next layers allows to combine features into meaningful properties





Sources: https://towardsdatascience.com/types-of-convolution-kernels-simplified-f040cb307c37

Recap: Transformer networks

Riddle:

The table is red. The chair is blue. What color is the table?

> The color of the table is red, as mentioned in the statement "The table is red."

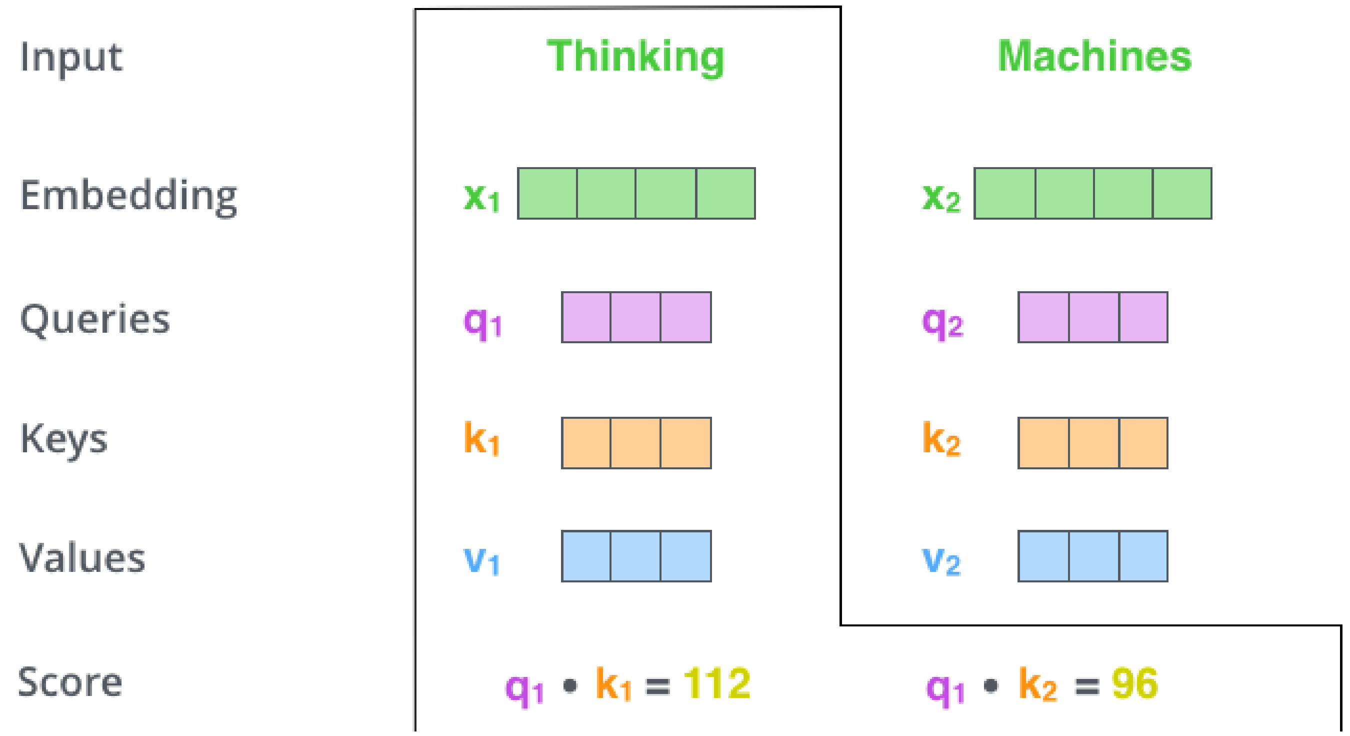
Yes, but how did he know?

Sources: chat.openai.com

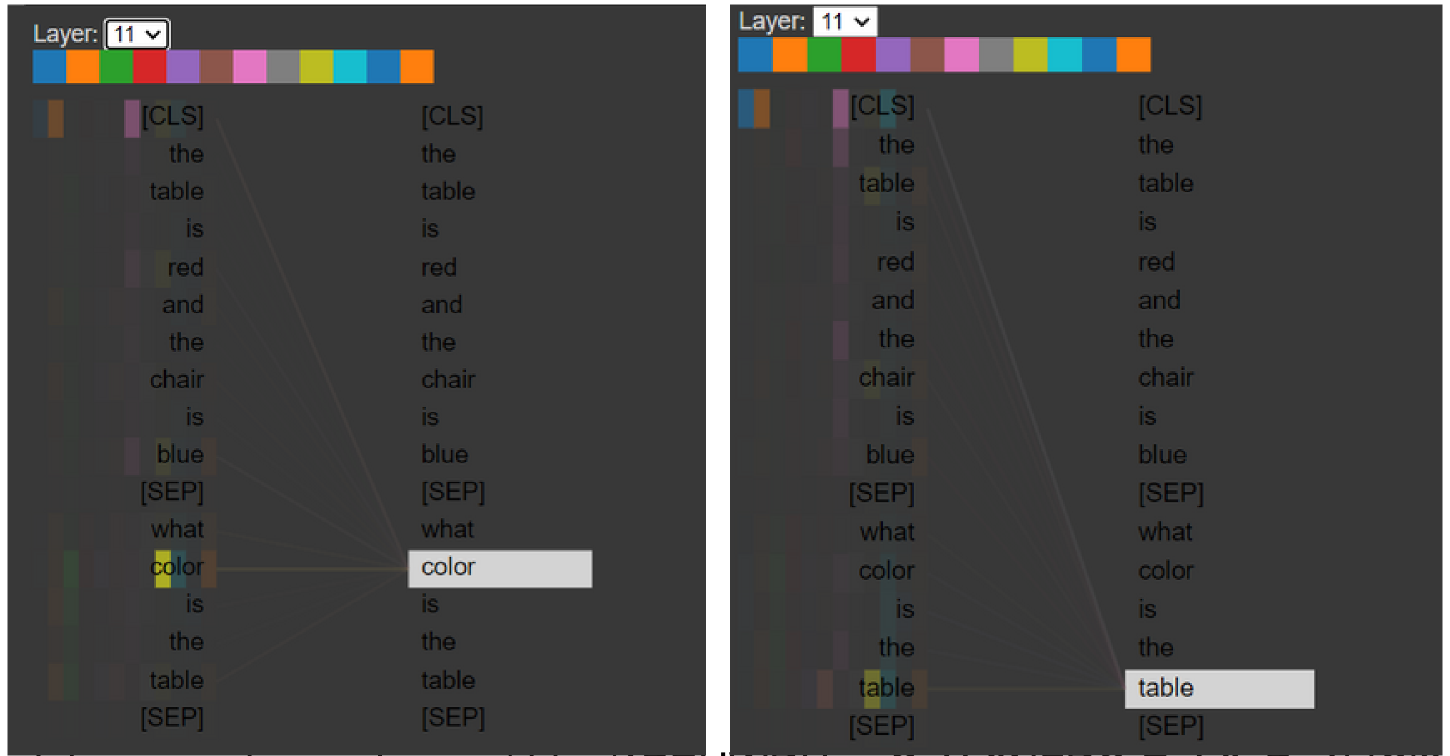
Recap: Transformers

Attention is all you need

- Key, Query, Value attention allow to learn which words refer to which



- Multiple attention layers allow for accurate reasoning even with complicated context



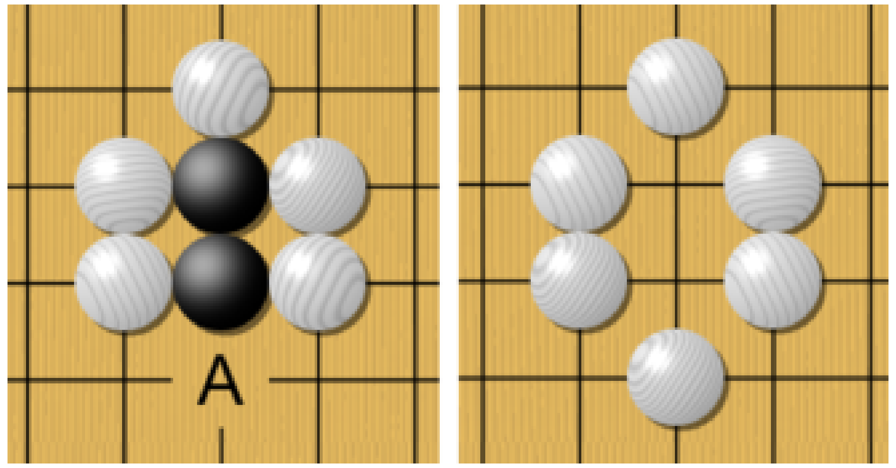
Sources: https://lih-verma.medium.com/query-key-and-value-in-attention-mechanism-3c3c6a2d4085, https://colab.research.google.com/drive/1PEHWRHrvxQvYr9NFRC-E_fr3xDq1htCj?ref=morioh.com&utm_source=morioh.com#scrollTo=fZAXH7hWyt58,

Few problems and choice of architectures

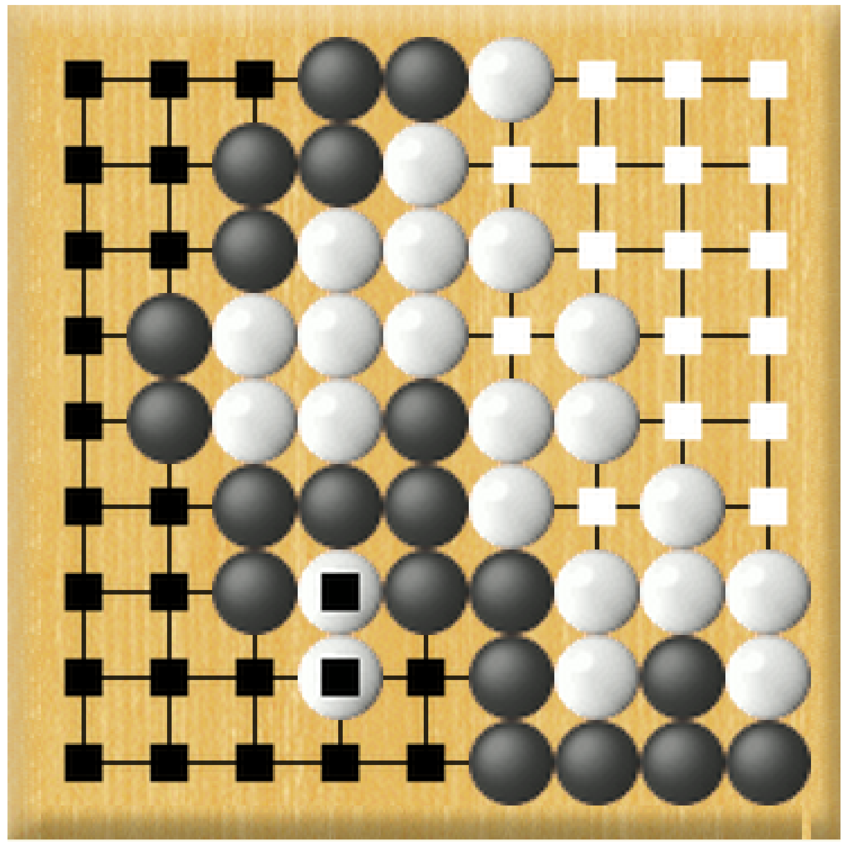
- Determine whether a dog is in the picture?
- Determine whether there are 3 red pixels on different corners of picture?
- Determine which human is pointing his finger?
- Determine which human is having a finger pointed at him?

Last recap: Go rules

- Players take turns to place a stone on the crossings board
- Once a group of stones is completely surrounded, it is captured and taken off the board



- After both players pass, the overall surrounded territory is evaluated. Winner is the player with more territory



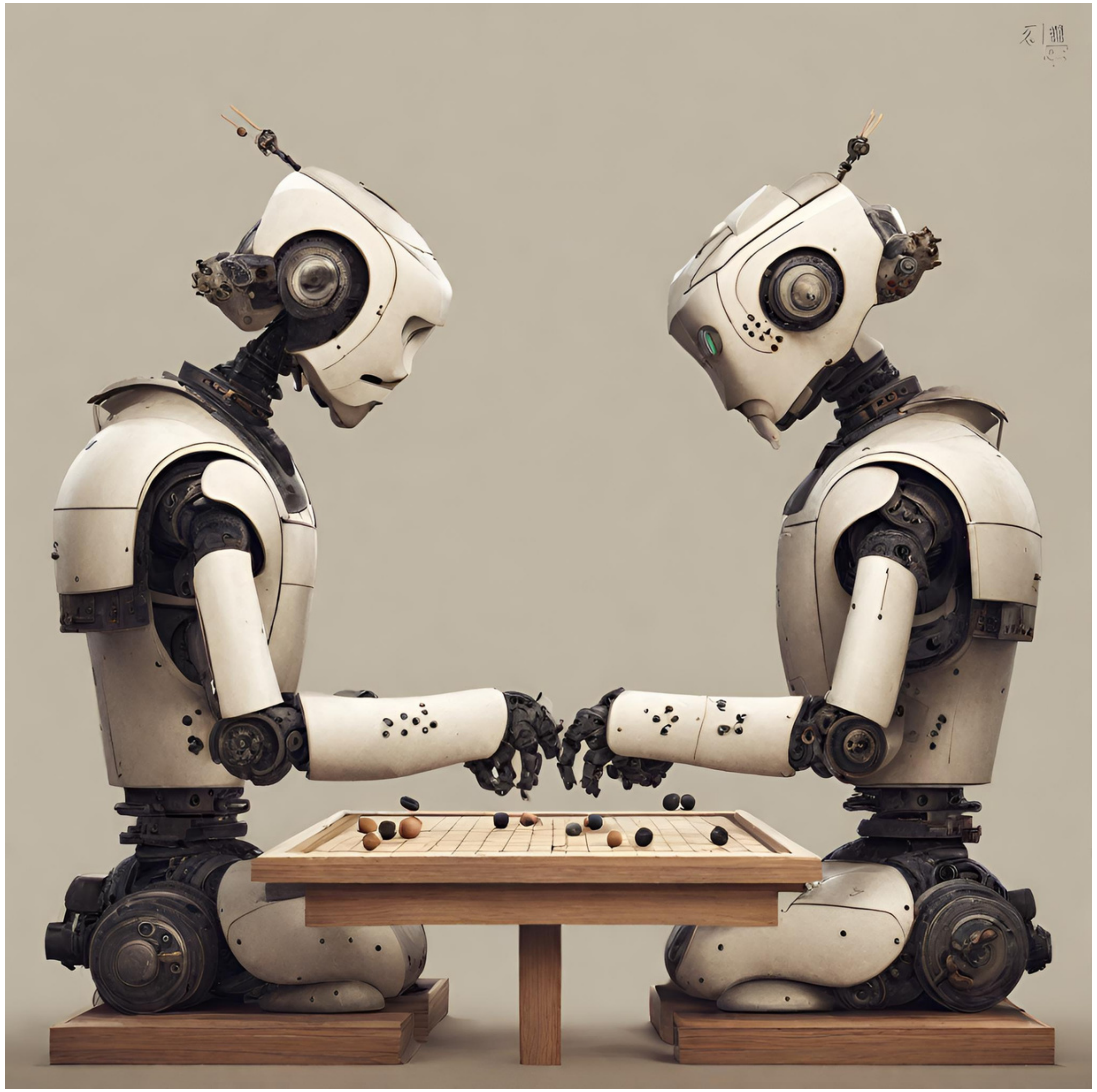
Sources: wikipedia.org, wolframalpha.com

Comparative Analysis of Convolutional and Transformer Architectures in Go Policy Networks

Antoni Hanke

Wydział Matematyki, Informatyki i Mechaniki

Uniwersytet Warszawski



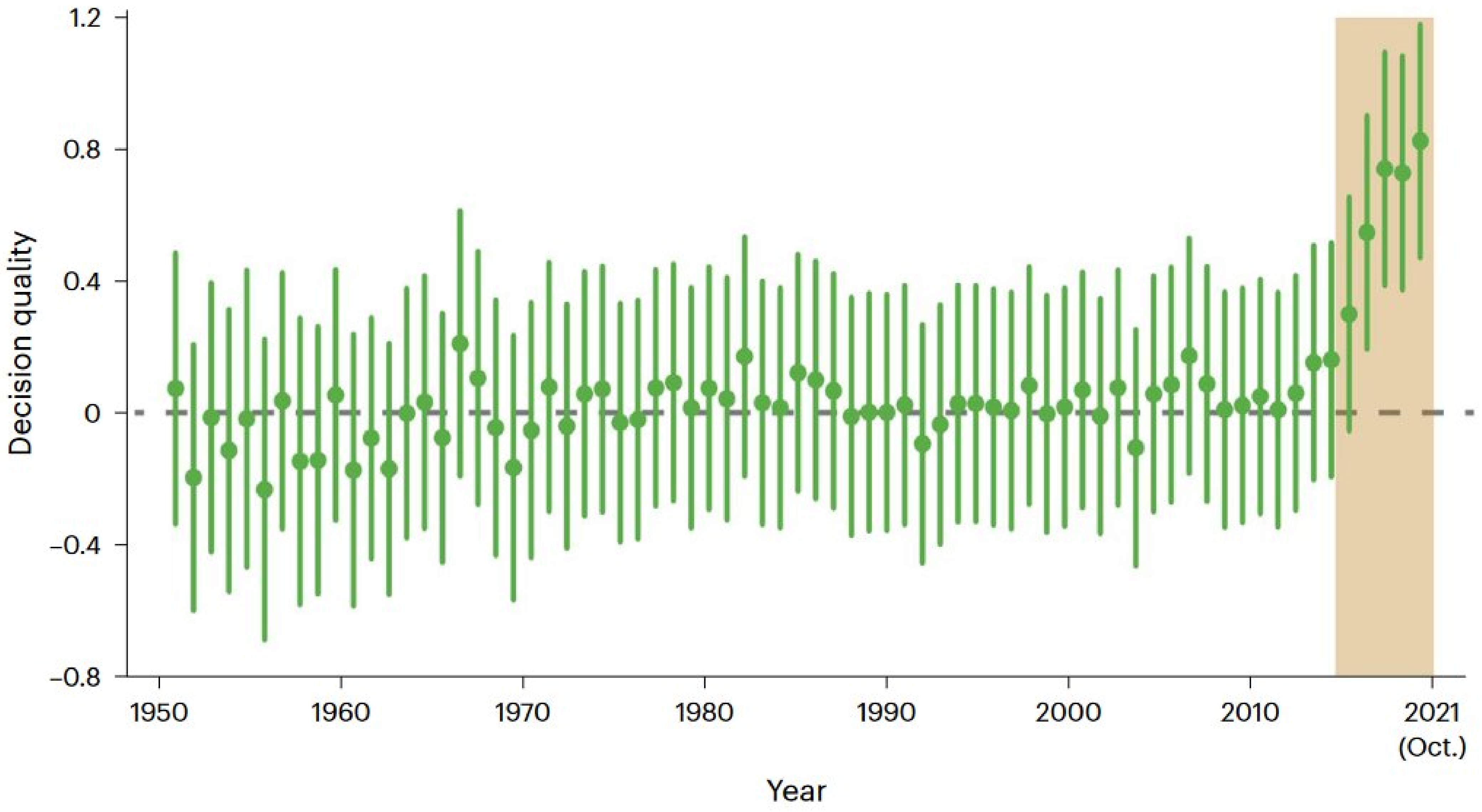
Why study how policies think?

"Know thy self, know thy enemy. A thousand battles, a thousand victories!"

Sun Tzu

Why study how policies think?

We can learn from them...

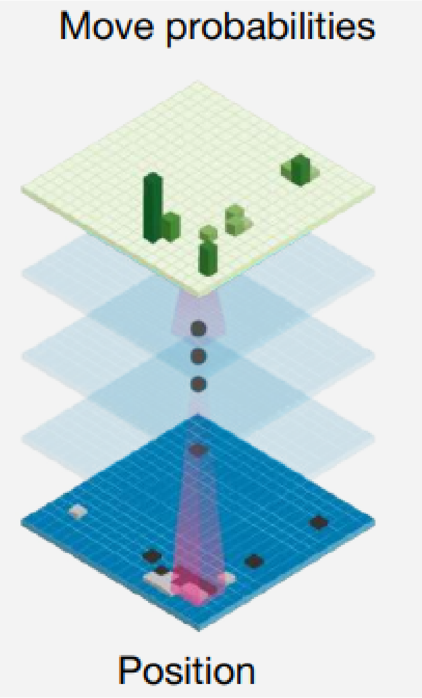


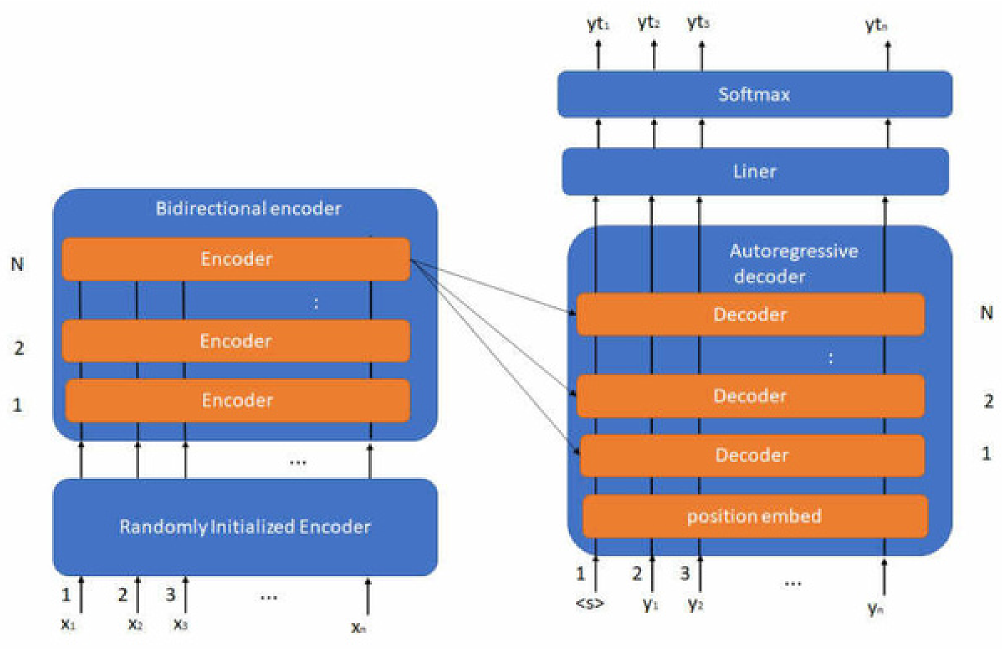
Decision quality of professional Go players before and after AlphaGo

Machine Culture, Brinkman et.al - 2023

https://arxiv.org/abs/2311.11388

Policy networks analyzed

<table>
<tr>
<td>


Convolution
40 residual connection blocks
256 channels
</td>
<td>


BART
8 encoder blocks
8 decoder blocks
8 heads
FFdim = 2048
d_model = 512
</td>
</tr>
<tr><td>~60M Parameters</td><td>~60M Parameters</td></tr>
<tr><td>Behavioral cloning on 500K human master games (4d+)</td><td>Behavioral cloning on 500K human master games (4d+)</td></tr>
<tr><td>48h training on 8xA100</td><td>48h training on 8xA100</td></tr>
</table>

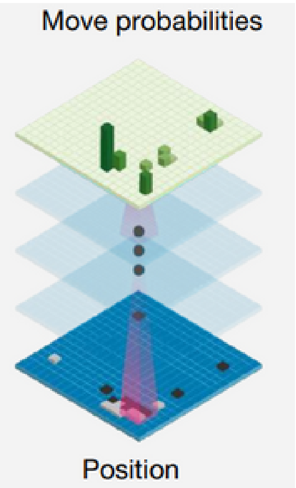
Both networks achieved human master strength (~3d)

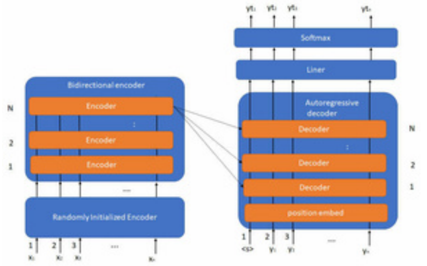
Oracle helps us evaluate both policy networks

Oracle - a superhuman strength policy (KataGo)
~9d+ strength, winning 100% games against our policies

Katago

Our networks



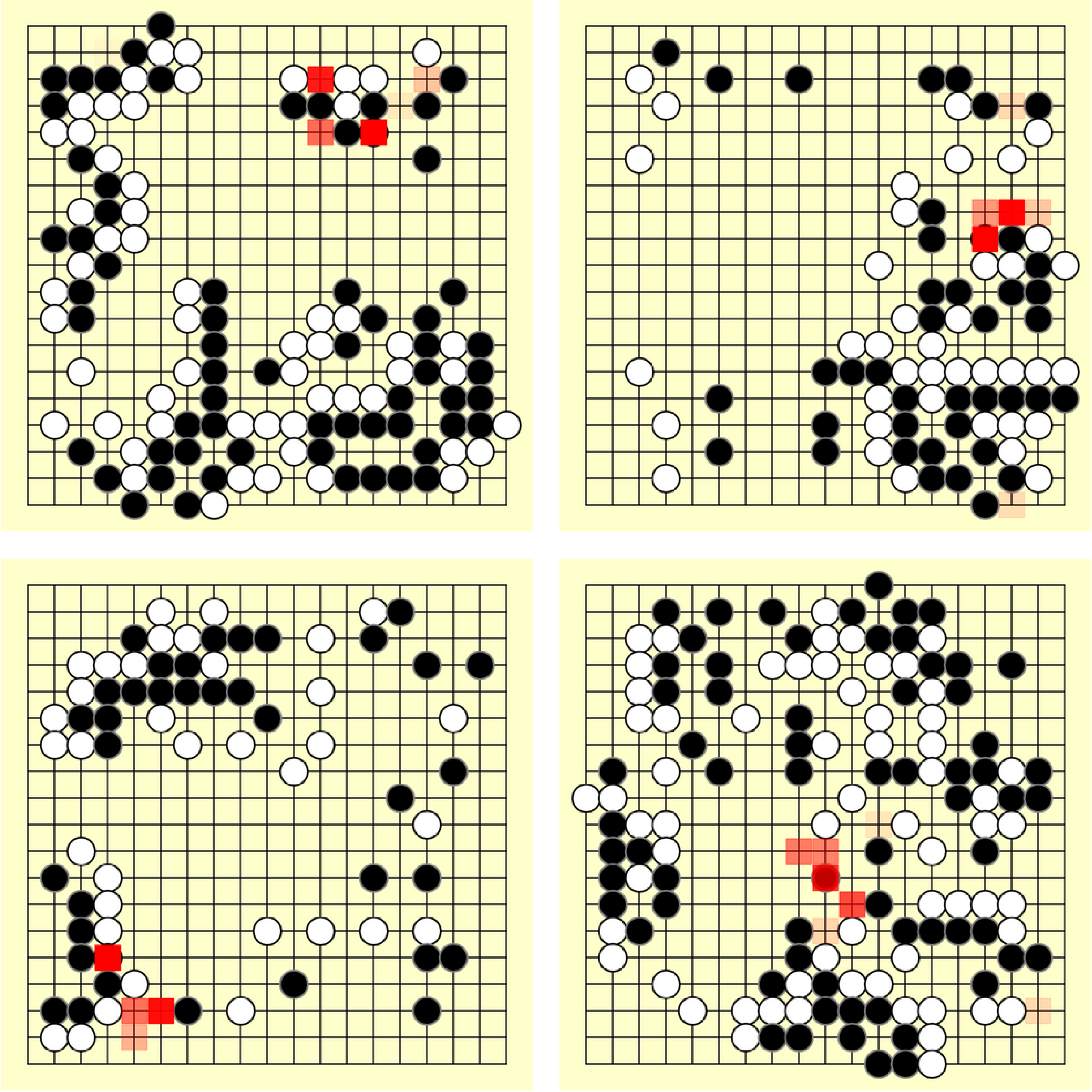


Board dispersity - a measure of position focus

$$dispersity(b) = \sum_{m \in moves} dist(m_{best}, m) \cdot p_{oracle}(m)$$

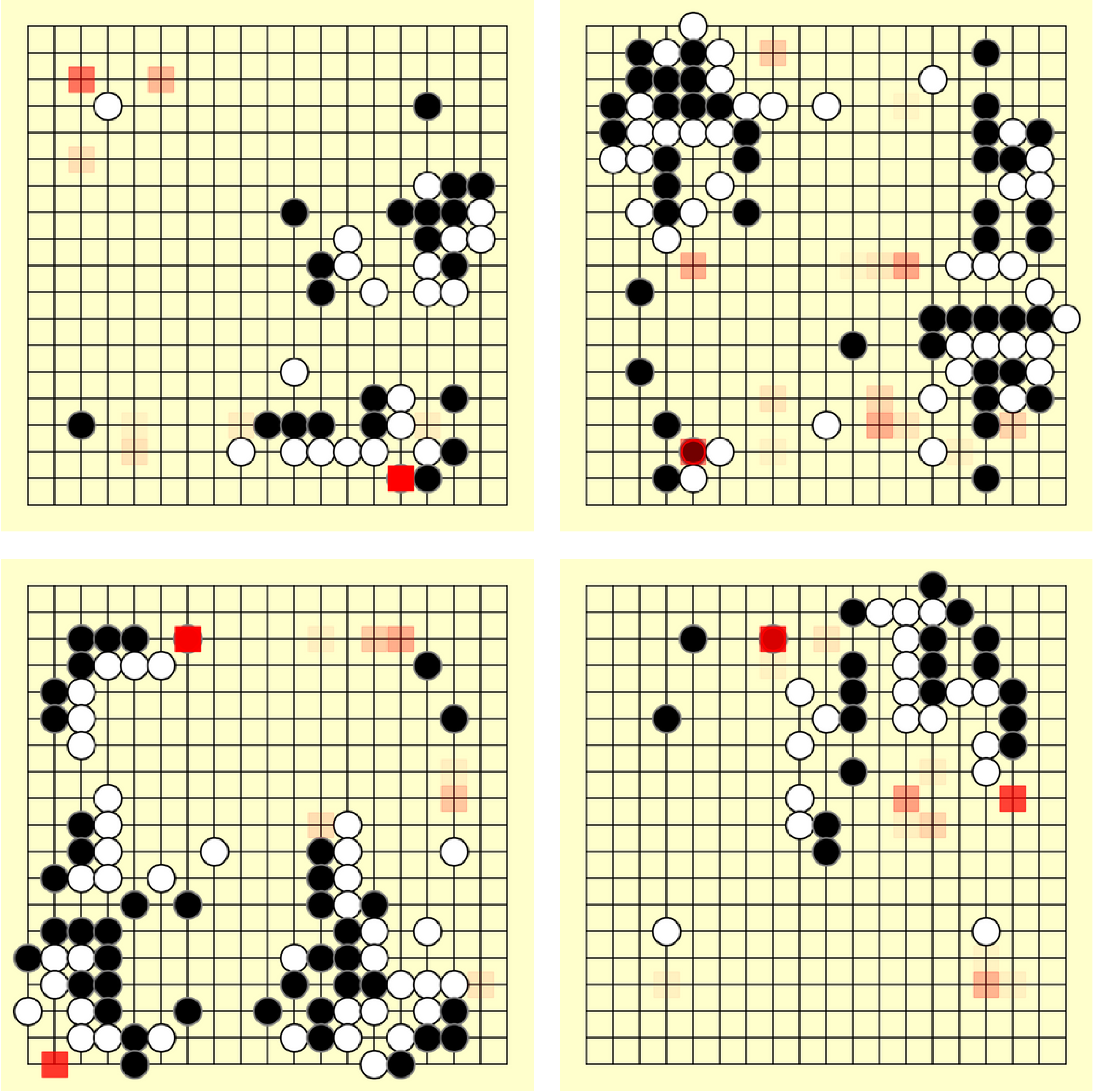
Low dispersity

Single globally important position



High dispersity

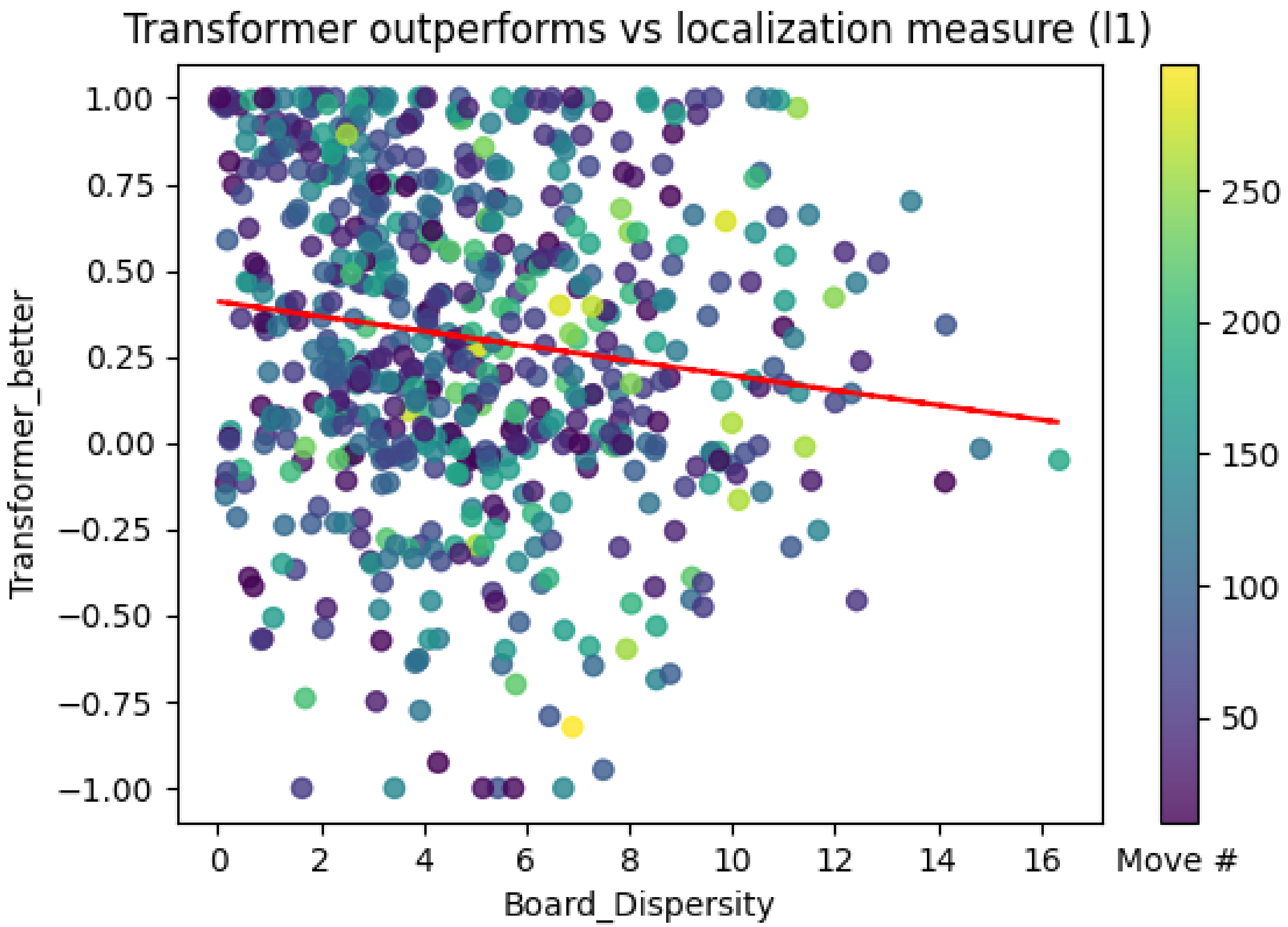
Multiple independent equivalent regions



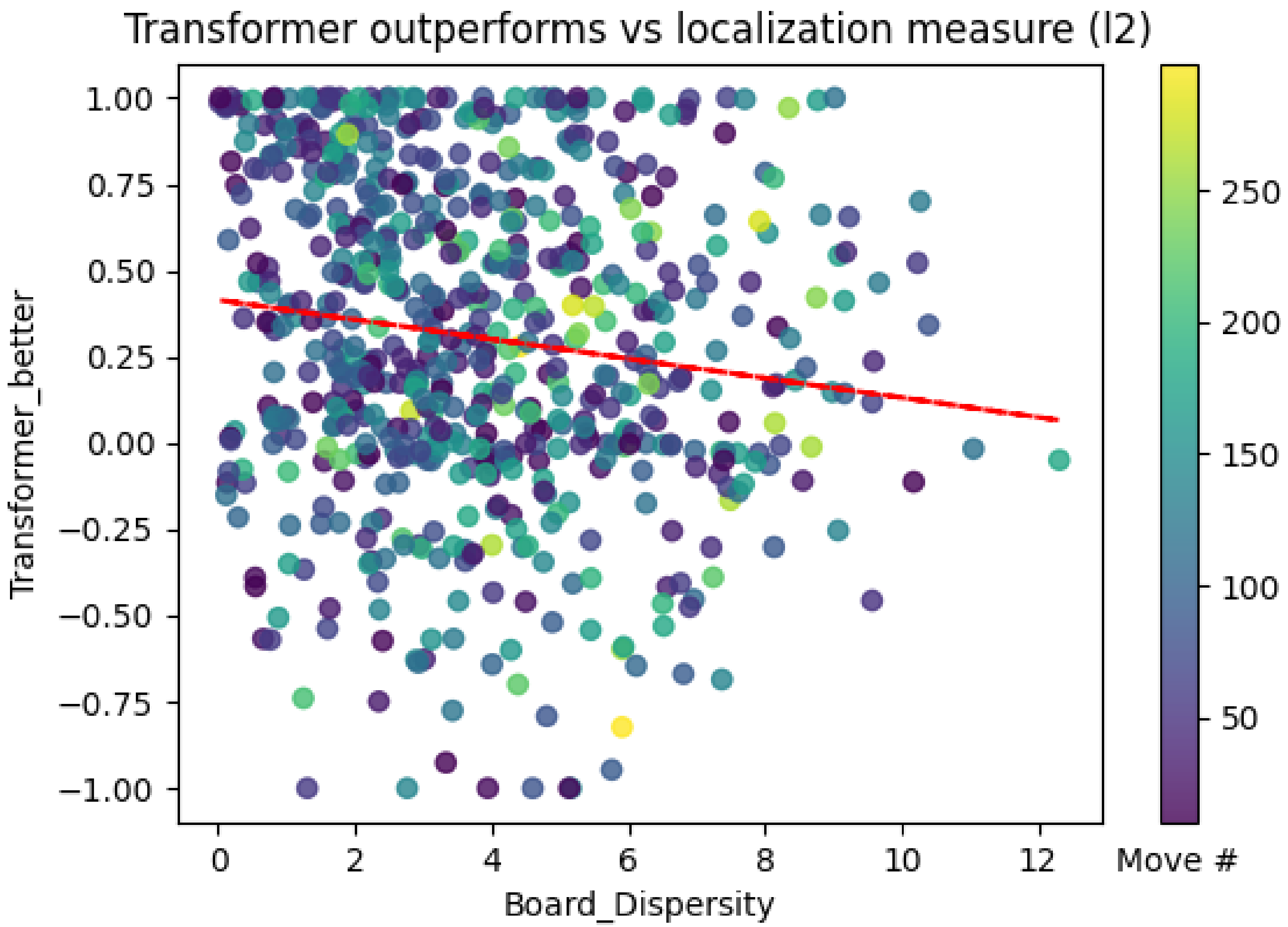
Policy performance

$$performacne_p = \frac{p(m_{best})}{\max\limits_{m \in moves} [p(m)]}$$

p – policy, m_{best} – best oracle move, $p(m)$ – probability of policy selecting move m



Spearman Correlation: -0.162, p-value: 4e-05

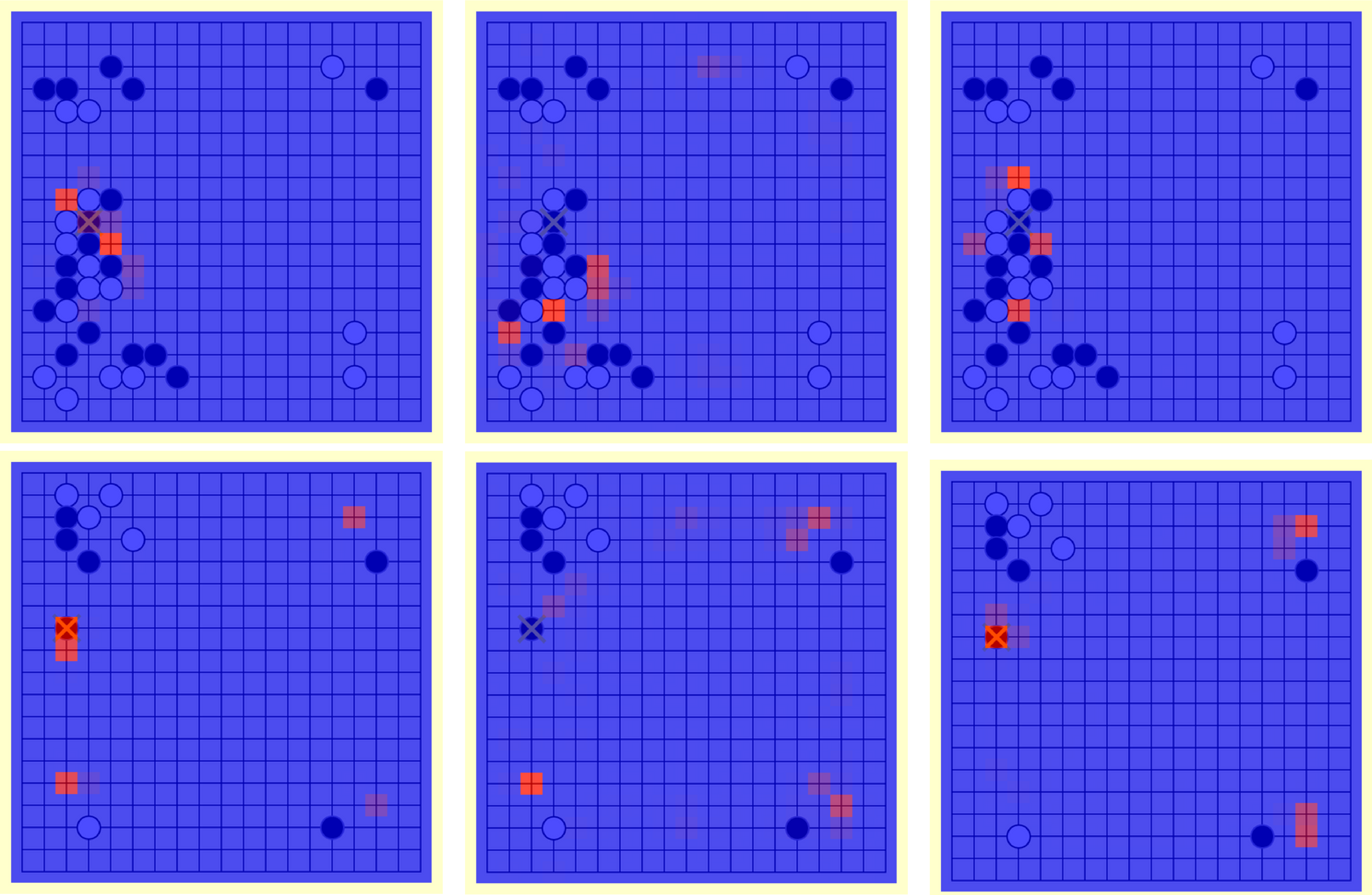


Spearman Correlation: -0.162, p-value: 4e-05

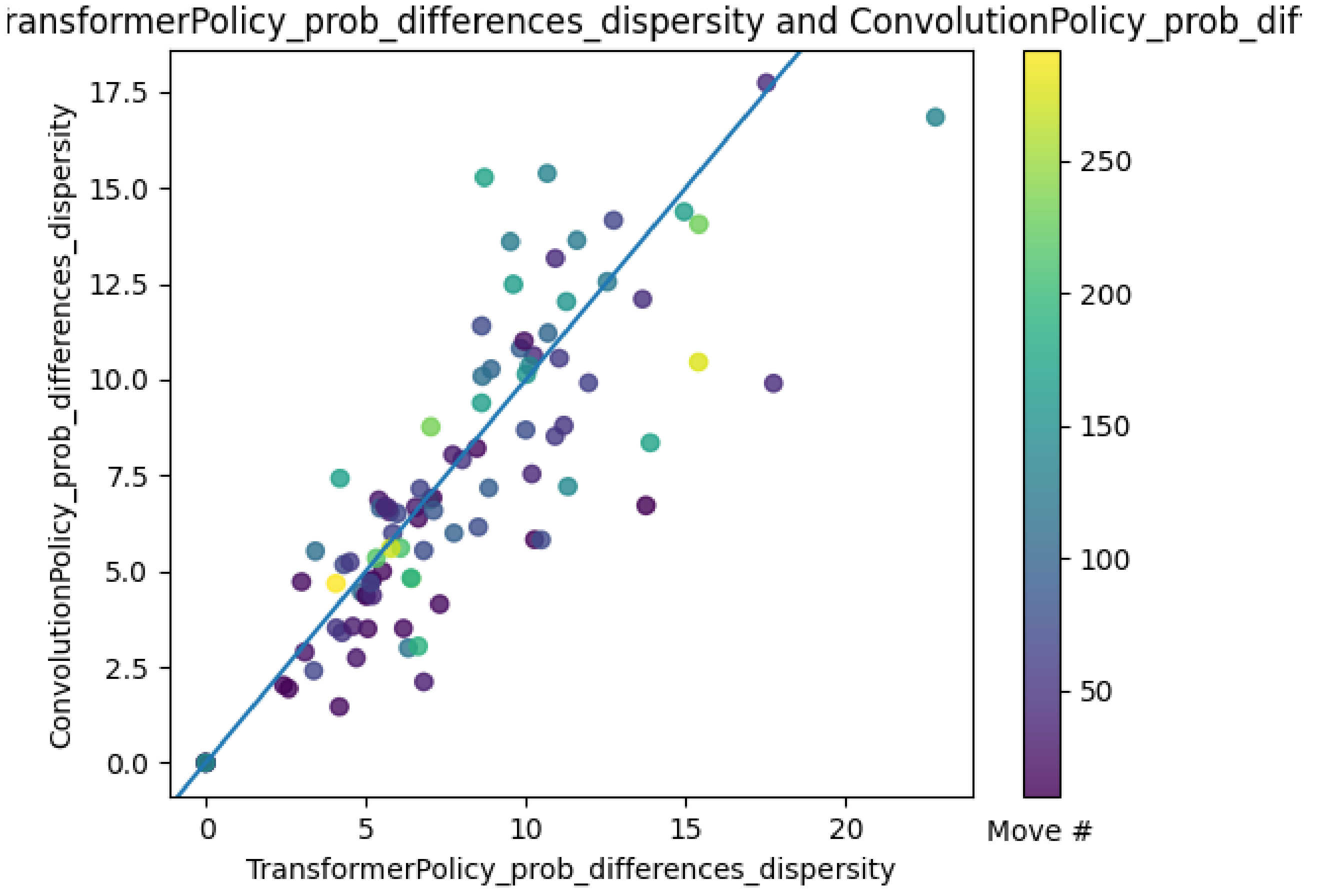
Transformer is better on low dispersity boards (single globally important position)

Ceteris Paribus probability difference disparity

$$probabilities\_dispersity_p(b, m_{removed}) = \frac{\sum_{s \in board\_squares} |p_b(s) - p_{b'}(s)| \cdot dist(s, m_{removed})}{\sum_{s \in board\_squares} |p_b(s) - p_{b'}(s)|}$$



Convolution Transformer Oracle



Convolution	Transformer	Oracle
7.23	7.77	8.72

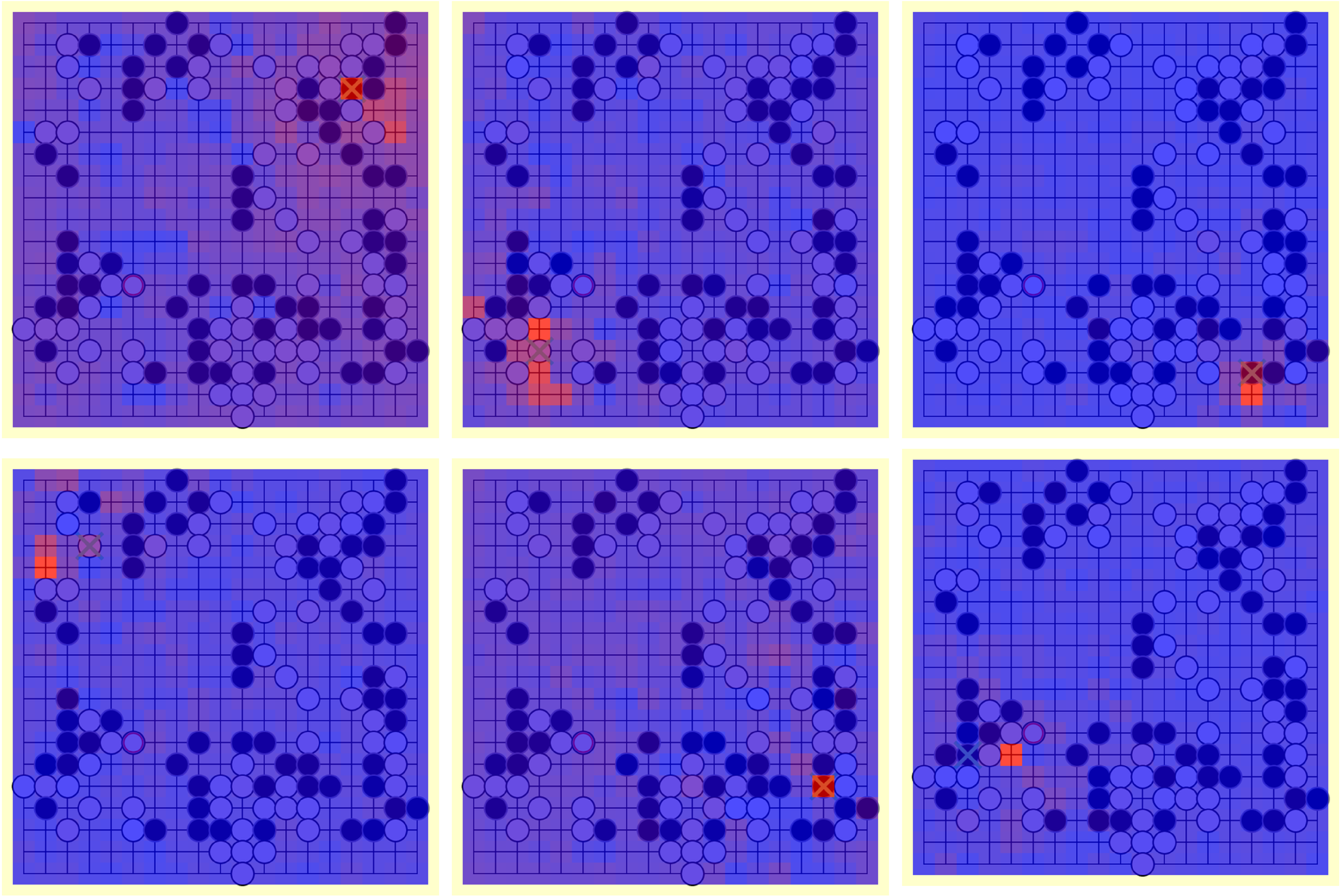
Wilcoxon singed pair test

P-value: 0.039

Changing a position influences Transformer's decisions further

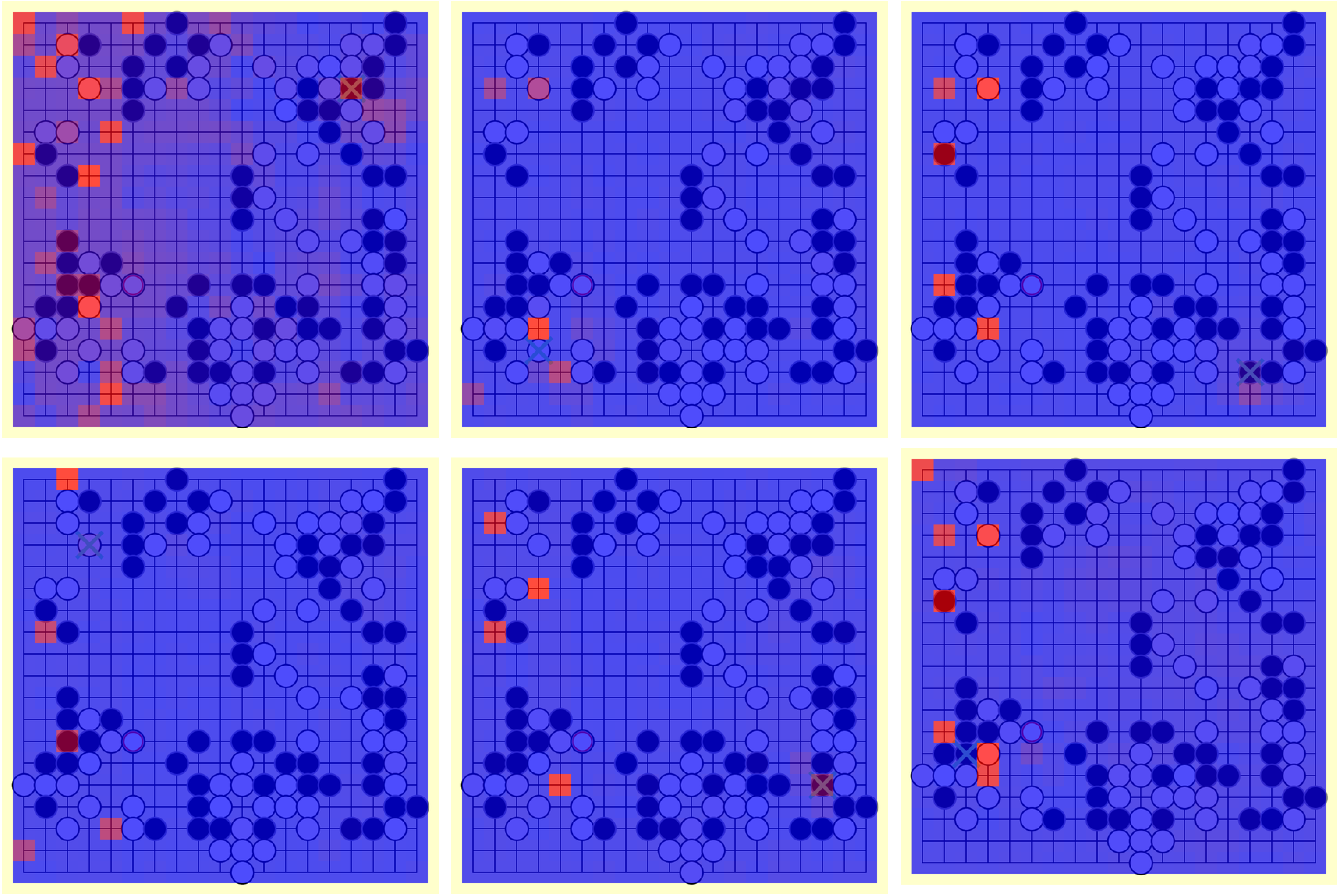
Last layer Ceteris Paribus logit differences

Convolution



Last layer Ceteris Paribus logit differences

Transformer



Conclusions

- Transformer’s global attention helps with understanding a single universal situation
- Transformer’s calculations are very pin-point and precise; Convolution’s are more gradient
- Convolution is slower at transmitting information over large distances upon localized change
- Applications where **precise understanding the global context is important might benefit from utilizing Transformers** instead of Convolutions

Further planned work

- Linear probings: We attach a simple linear classifier at various points of the network and train it to recognize whether certain features are on the board
- Concept-conditional explanations to find non-human concepts recognized by the 2 networks

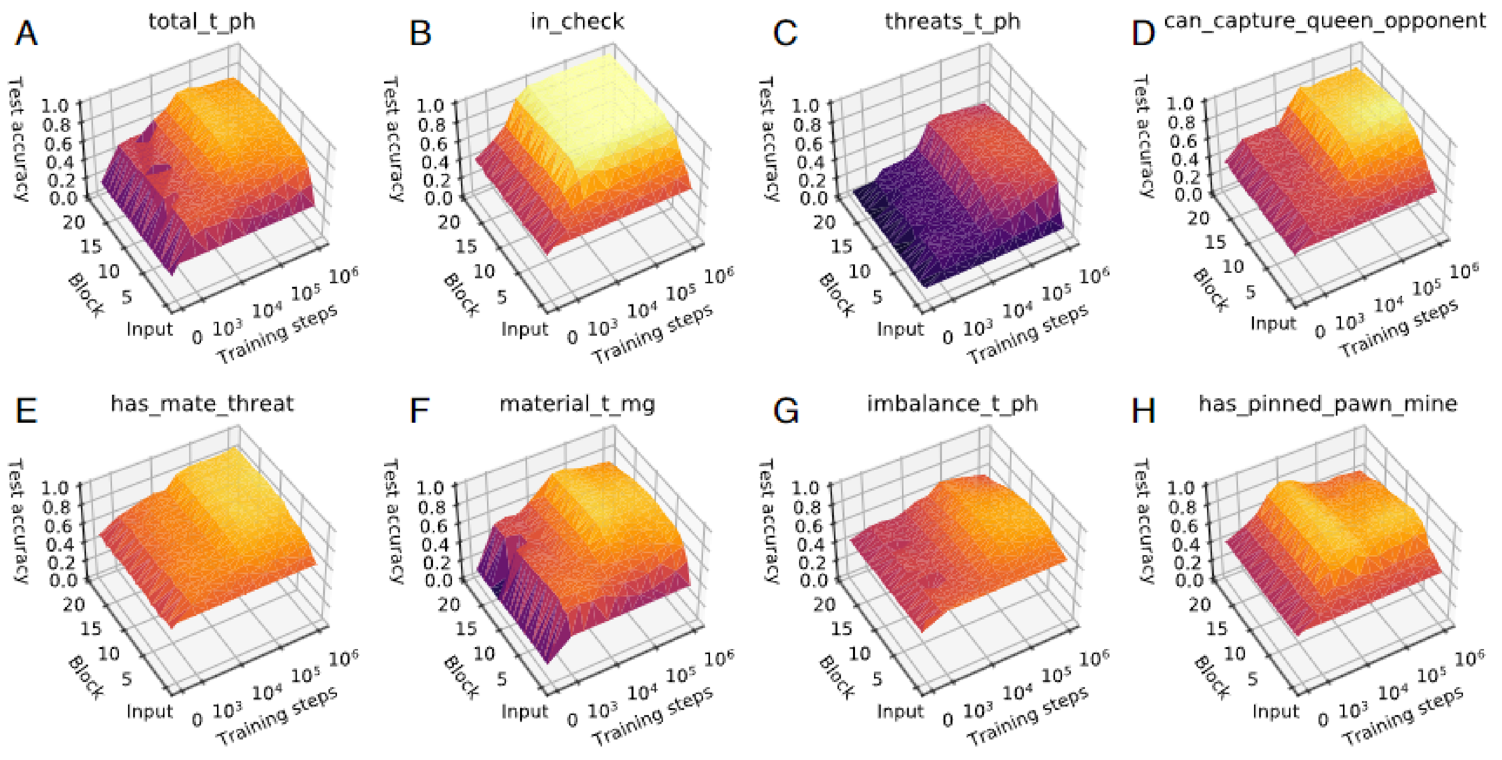
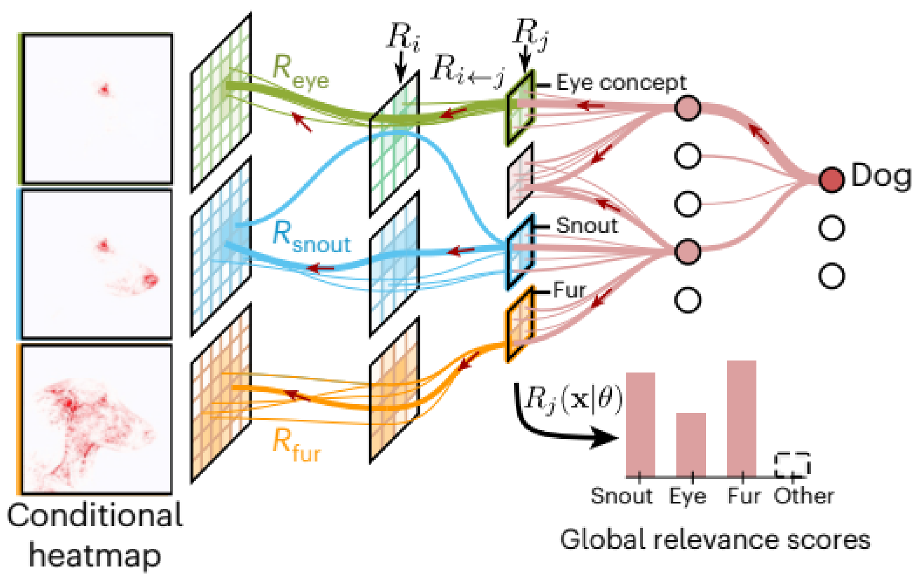


Fig. 2. What–when–where plots for a selection of Stockfish 8 and custom concepts. Following Fig. 1, we count a ResNet "block" as a layer. (*A*) Stockfish 8's evaluation of total score. (*B*) Is the playing side in check? (*C*) Stockfish 8's evaluation of threats. (*D*) Can the playing side capture the opponent's queen? (*E*) Could the opposing side checkmate the playing side in one move? (*F*) Stockfish 8's evaluation of "material score." (*G*) Stockfish 8's material score. Past 10^5 training steps this becomes less predictable from AlphaZero's later layers. (*H*) Does the playing side have a pawn that is pinned to the king?

b Concept-conditioned explanation (CRP)



$$\boxed{R_{i\leftarrow j}^{(l-1,l)}(\mathbf{x}|\theta\cup\theta_l)=\frac{z_{ij}}{z_j}\sum_{c_l\in\theta_l}\delta_{jc_l}R_j^l(\mathbf{x}|\theta)}$$