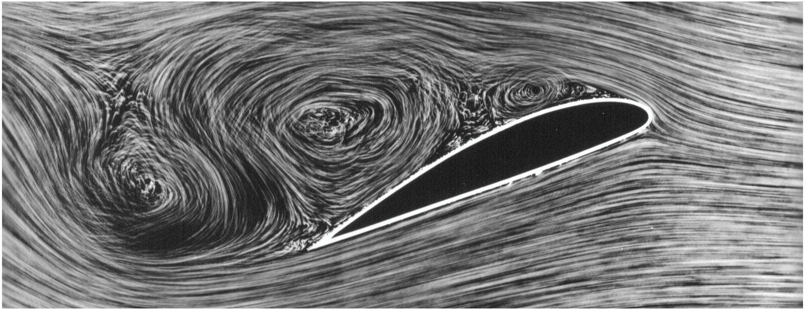


Subsampled Bi-Transformer

for RANS Surrogate modelling

Machine Learning for Physical Simulation Challenge – IRT SystemX

Anthony Kalaydjian – Master student @ ENSTA/EPFL – anthony.kalaydjian@epfl.ch
Anton Balykov – Master student @ EPFL – anton.balykov@epfl.ch
Adrien Chan-Hon Tong – Onera Université Paris Saclay – adrien.chan_hon_tong@onera.fr

DaGrapix/ml4physim

Introduction

Models used as surrogates to approximate solutions for physical problems are built upon physical data which is highly interconnected by nature.

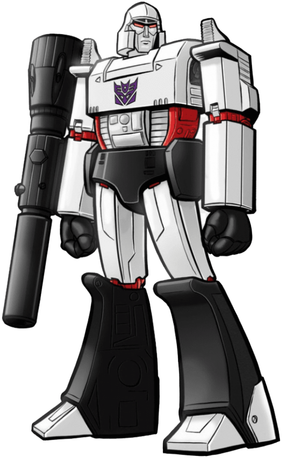
How can we use these relationships to our advantage?

Outline

1. Transformers
2. Our method
3. Results

Outline

Our Motivation



Attention Mechanism [Vaswani et al., 2017]

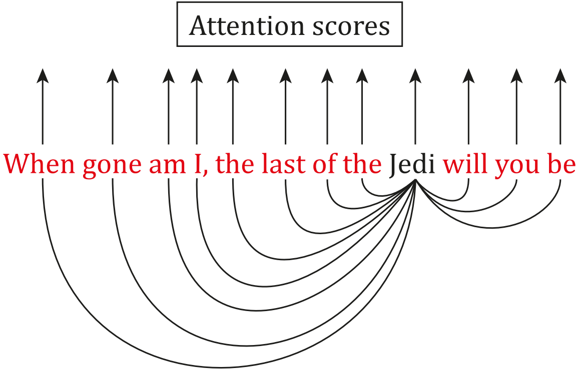


Figure 1: Attention in NLP

Attention Mechanism on Mesh

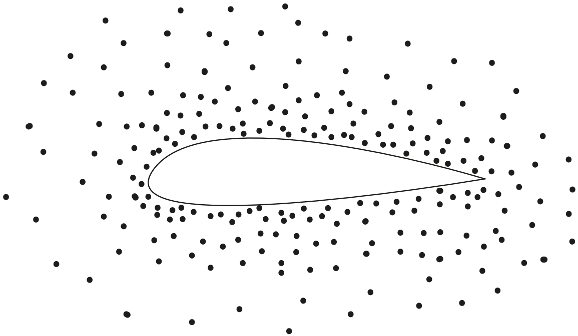


Figure 2: Original mesh

There's just one small problem...

Attention context windows:

- GPT 3.5 turbo context window: 16,385 tokens
- (NEW) GPT 4 turbo context window: 128,000 tokens[1]
- BUT 1 simulation = 179,761 points on average

Memory efficient attention:

- Linformer [Wang et al., 2020] : $O(N^2) \rightarrow O(N)$
- Reformer [Kitaev et al., 2020] : $O(N^2) \rightarrow O(N \log N)$

[1] https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

Outline

Our Method

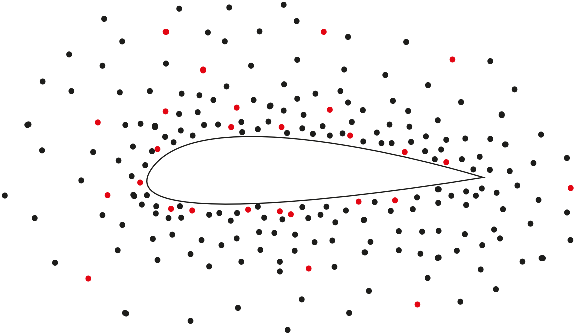


Figure 3: Sampled skeleton

Our Method

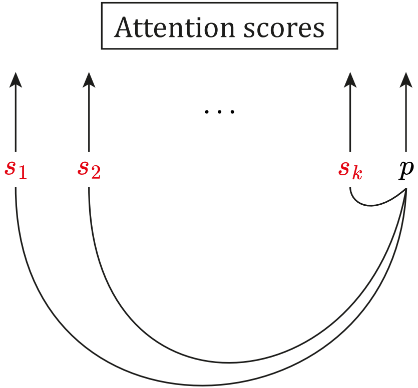


Figure 4: Pointcloud Attention

Bi-Transformer 1/4

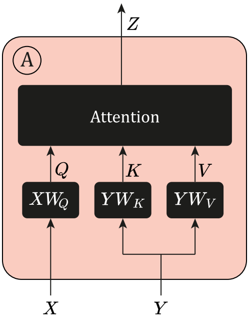


Figure 5: Attention block

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T/\sqrt{d_k})V$$

Bi-Transformer 2/4

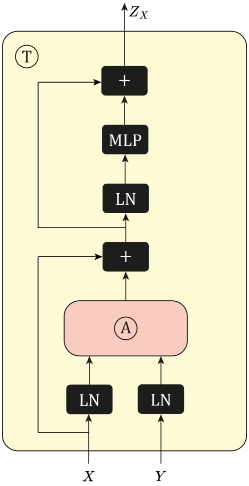


Figure 6: Transformer block

Bi-Transformer 3/4

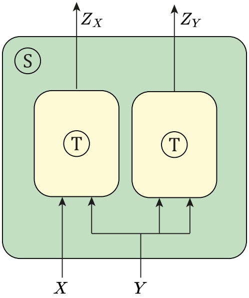


Figure 7: Bi-transformer block

Bi-Transformer 4/4

Architecture:

- 3 layers
- Size of skeleton: $k = 1000$
- Bi-transformer MLP: [32, 64, 64, 64, 32]
- Encoder MLP: [7, 64, 64, 32]
- Decoder MLP: [32, 64, 64, 32, 16, 4]
- Activation functions: ReLU
- Batch-size: 50,000

Benefits:

- Capture global context
- $O(kN)$ attention
- Any batch-size !

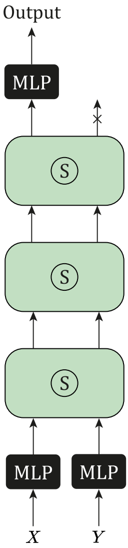


Figure 8: Full architecture

Outline

Results

Global	ML	Physics	OOD
51.25	0.62	0.5	0.38

Table 1: Results

Conclusion & perspectives

In summary:

- Attention captures similarities in the data
- But it is expensive

$\rightarrow$ Subsample the context for each point

Further conclusions:

- Ensembling method performed poorly
- Use different sampling methods
- Use Physics!
  - Physics informed loss
  - Mixture of models with physical head

Thank you for your **attention**!

References

Kitaev, N., Kaiser, Ł., and Levskaya, A. (2020).
Reformer: The efficient transformer.
arXiv preprint arXiv:2001.04451.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017).
Attention is all you need.
In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Wang, S., Li, B. Z., Khabsa, M., Fang, H., and Ma, H. (2020).
Linformer: Self-attention with linear complexity.
arXiv preprint arXiv:2006.04768.