Vision Transformer (ViT) и его применение в науке

Анвар Курмуков

Старший научный сотрудник, AIRI

Прежде чем начать: matrix multiplication

Как можно умножать матрицу на вектор?

- Слева:

- Справа:

В чем разница?

Линейная модель

Предположения

1. Значение целевой переменной наблюдения равно линейной комбинации признаков (описывающих это наблюдение):

2. *Наблюдения в выборке independent identically distributed.*

Пример 1. Кредитный скоринг

Имя	Пол	...	Выдать кредит
Маруся	1	...	0
Алиса	1	...	1
Олег	0	...	0

Пример 2. Предсказание части речи

«Однажды в студеную зимнюю пору я из лесу вышел был сильный мороз...»

token	Embedding, 512	Часть речи
Однажды	...	нар.
в студеную	...	прил.
зимнюю	...	прил.
пору	...	сущ.
я	...	мест.
из лесу	...	сущ.
вышел	...	глаг.
	...	

Как добавить связь между наблюдениями в линейную модель?

Олег и Алиса – муж и жена.

Есть *n* сущностей, как нам описать наличие или отсутствие связи между этими сущностями?

Имя	Пол	...	Выдать кредит
Маруся	1	...	0
Алиса	1	...	1
Олег	0	...	0

n x n → число

Последовательность частей речи в предложении

token	Embedding, 512	Часть речи
Однажды	…	нар.
в студеную	…	прил.
зимнюю	…	прил.
пору	…	сущ.
я	…	мест.
из лесу	…	сущ.
вышел	…	глаг.
	…	

Как создавать граф?

Here, score is referred as a *content-based* function for which we consider three different alternatives:

$$\text{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s) = \begin{cases} \boldsymbol{h}_t^\top \bar{\boldsymbol{h}}_s & \textit{dot} \\ \boldsymbol{h}_t^\top \boldsymbol{W}_a \bar{\boldsymbol{h}}_s & \textit{general} \\ \boldsymbol{v}_a^\top \tanh\left(\boldsymbol{W}_a [\boldsymbol{h}_t; \bar{\boldsymbol{h}}_s]\right) & \textit{concat} \end{cases}$$

https://arxiv.org/pdf/1508.04025.pdf

Можно ли идти в глубину?

Все что нам нужно это внимание!

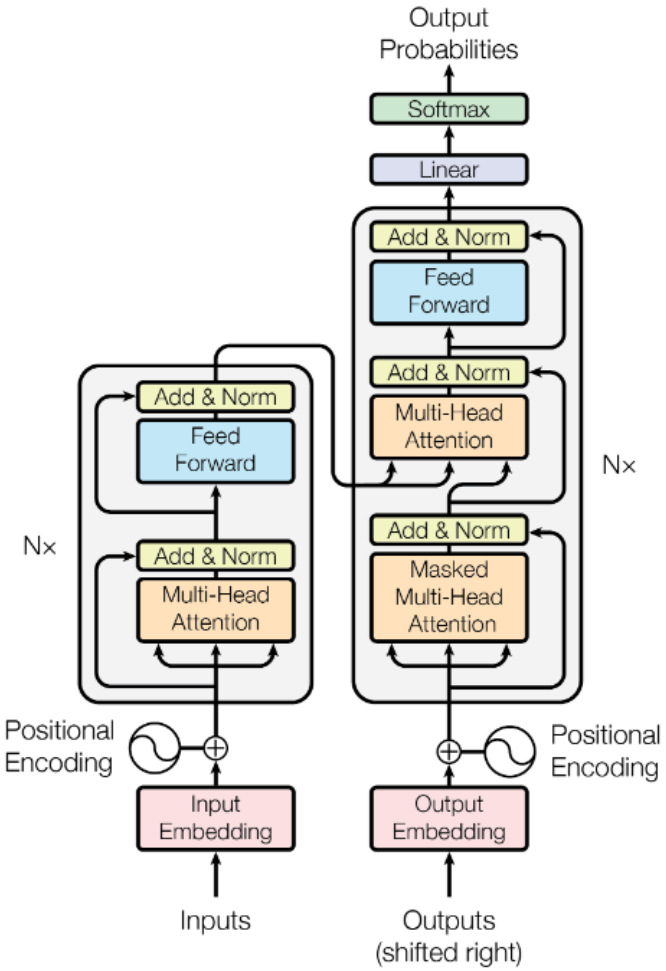
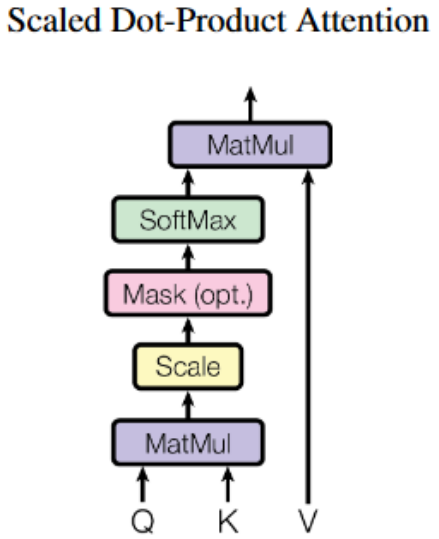
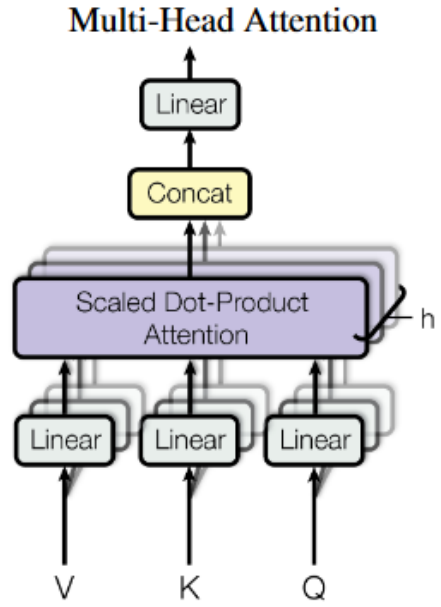


Figure 1: The Transformer - model architecture.





https://arxiv.org/pdf/1706.03762.pdf

Некоторые технические детали без которых ничего не заведется

1. Non-linearities
2. Attention vs Self-attention
3. Нормализация матрицы attention
4. Layer normalization vs batch normalization
5. Positional encoding
6. Masked attention
7. ...

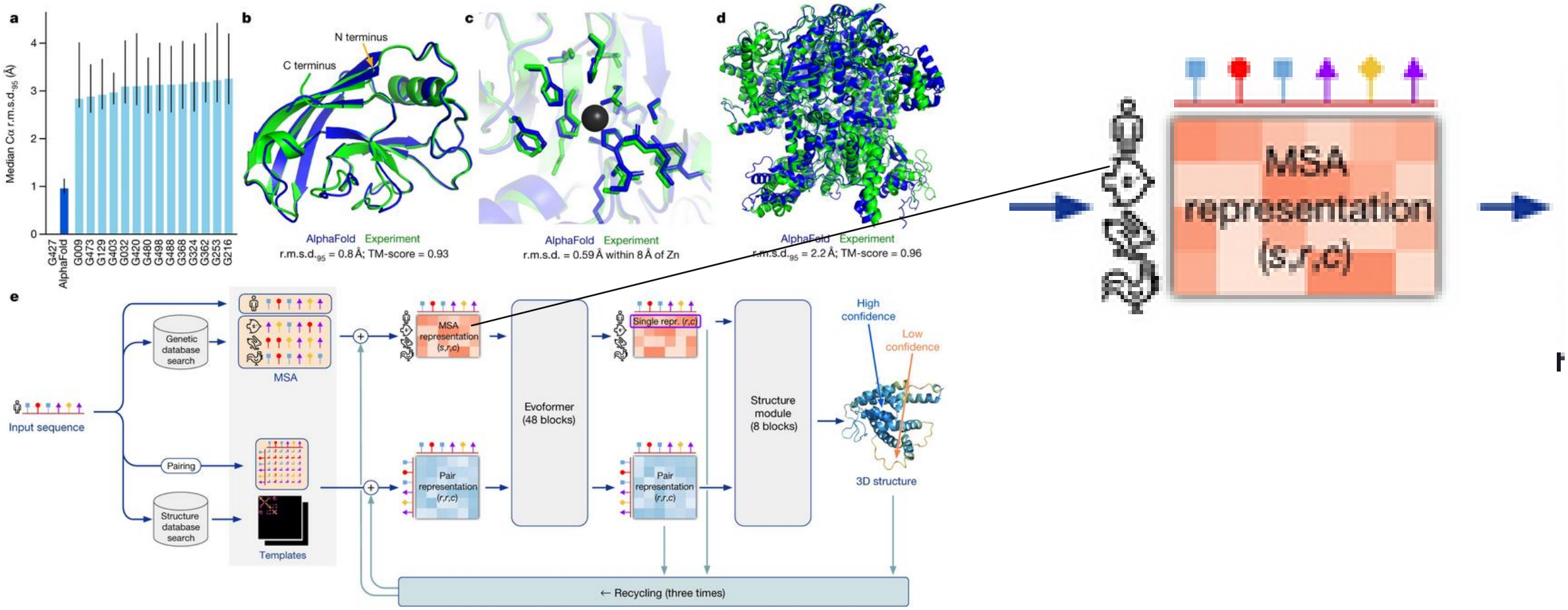
Что дальше?

- GPT/BERT
- Graph Attention Network
- Alpha/Rosetta Fold
- Vision Transformer (ViT)
- Swin Transformer
- MLP-Mixer
- Foundation models
- ...

Highly accurate protein structure prediction with AlphaFold

Fig. 1: AlphaFold produces highly accurate structures.

From: Highly accurate protein structure prediction with AlphaFold



https://www.nature.com/articles/s41586-021-03819-2

AIRI

Image is worth 16x16 words

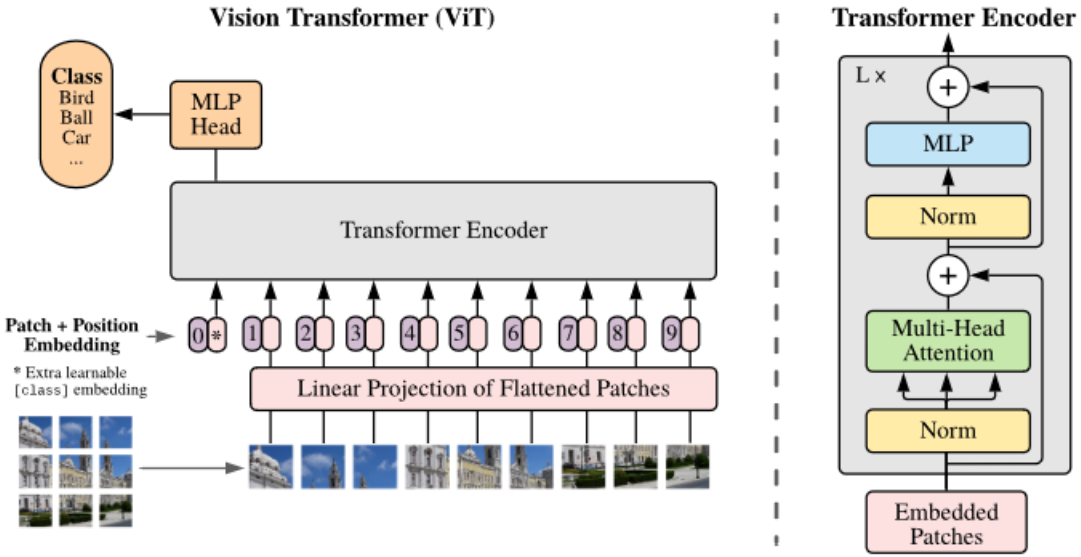


Figure 1: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

https://arxiv.org/pdf/2010.11929.pdf

AIRI

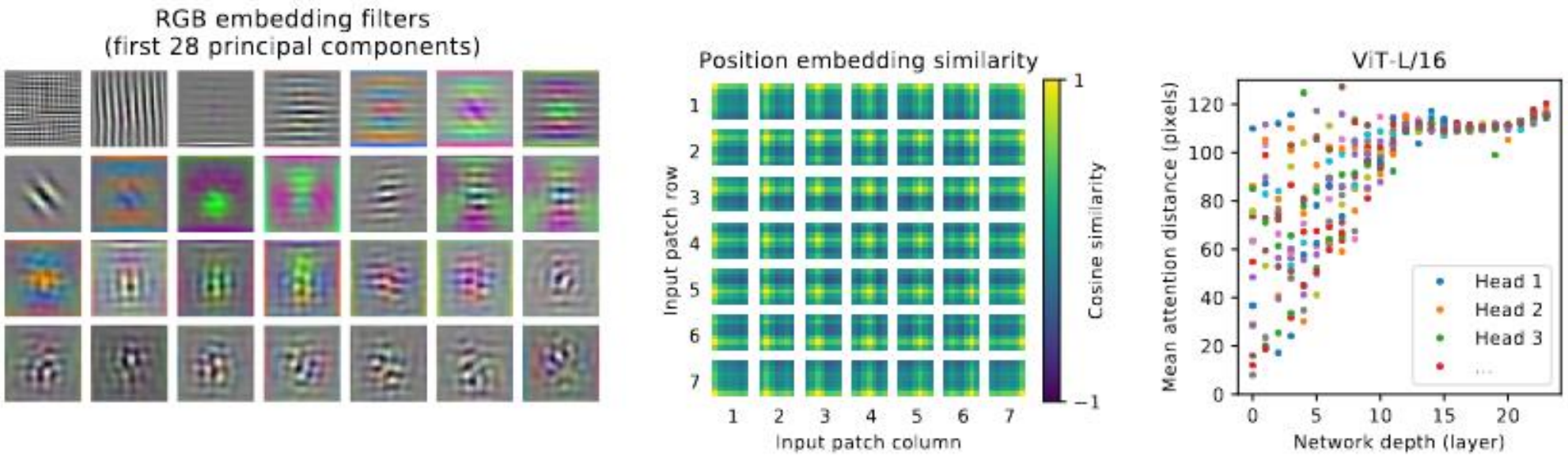


Figure 7: **Left:** Filters of the initial linear embedding of RGB values of ViT-L/32. **Center:** Similarity of position embeddings of ViT-L/32. Tiles show the cosine similarity between the position embedding of the patch with the indicated row and column and the position embeddings of all other patches. **Right:** Size of attended area by head and network depth. Each dot shows the mean attention distance across images for one of 16 heads at one layer. See Appendix D.7 for details.

UNETR: Transformers for 3D Medical Image Segmentation

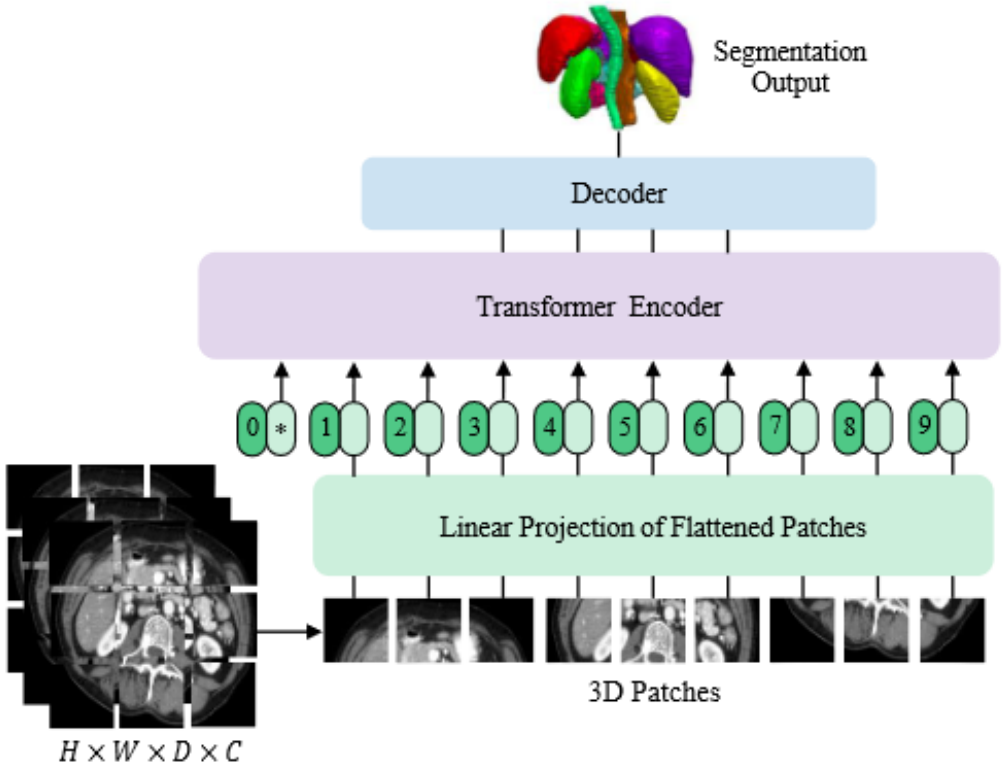
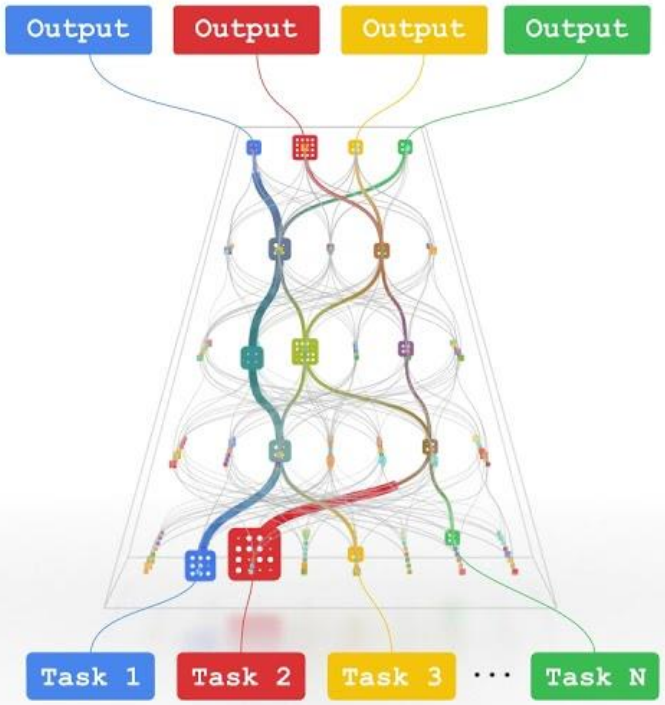


Figure 1. Overview of UNETR. Our proposed model consists of a transformer encoder that directly utilizes 3D patches and is connected to a CNN-based decoder via skip connection.

https://arxiv.org/pdf/2103.10504.pdf

AIRI

Foundation models



Google Pathways
Sberbank Ru-DALLE, FusionBrain

Pathways: A single model that can generalize across millions of tasks.

Questions?

Анвар Курмуков, PhD

kurmukovai@gmail.com