NLP &
Outfit Recommendation

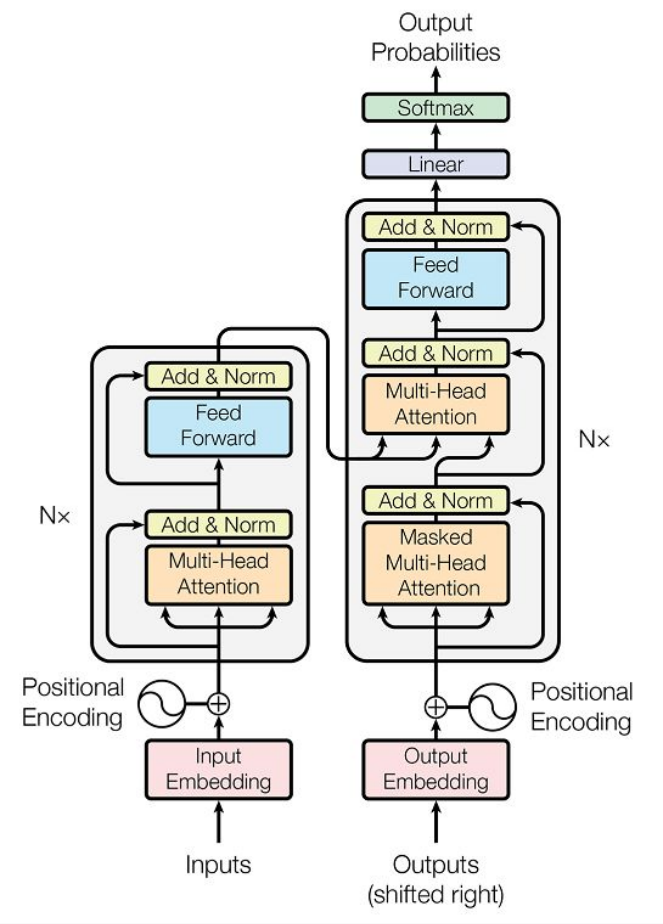
David Nepožitek

Natural Language Processing

- Syntax (grammar induction, stemming,...),
- Semantics (translation, language generation, chatbots, sentiment analysis, question answering,...),
- Speech (speech recognition, text-to-speech)
- and much more

Approaches

1. Rule-based
   - Grammars, patterns, heuristics etc.
2. “Traditional” Machine Learning
   - Mostly probabilistic modeling, decision trees etc.
3. Neural Networks
   - Vector representations of words are learned
   - Learning rules thanks to the large amount of data

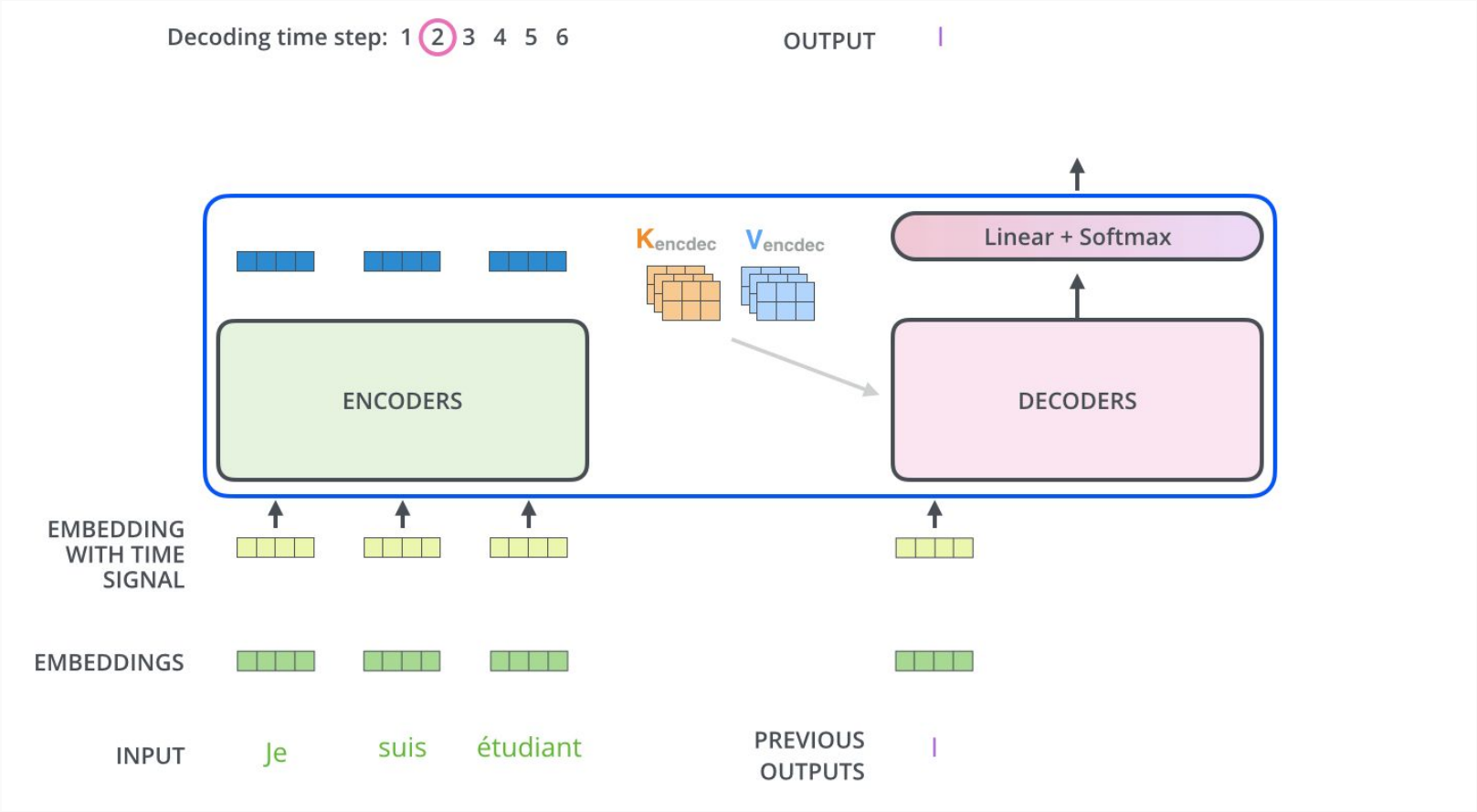


The Transformer

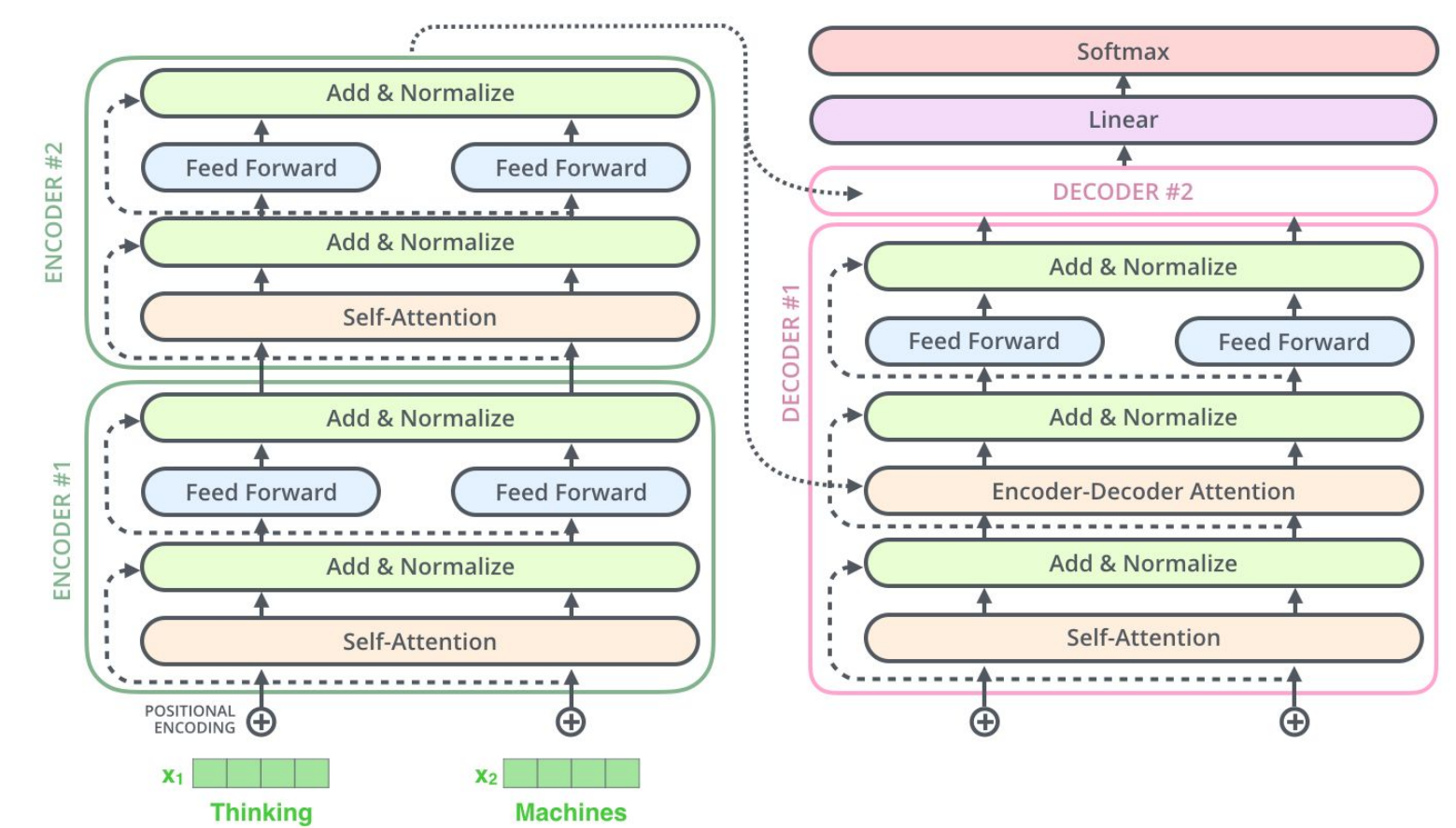
Vaswani, Ashish et al. “Attention is All you Need.” NIPS (2017).

Translation Task Example

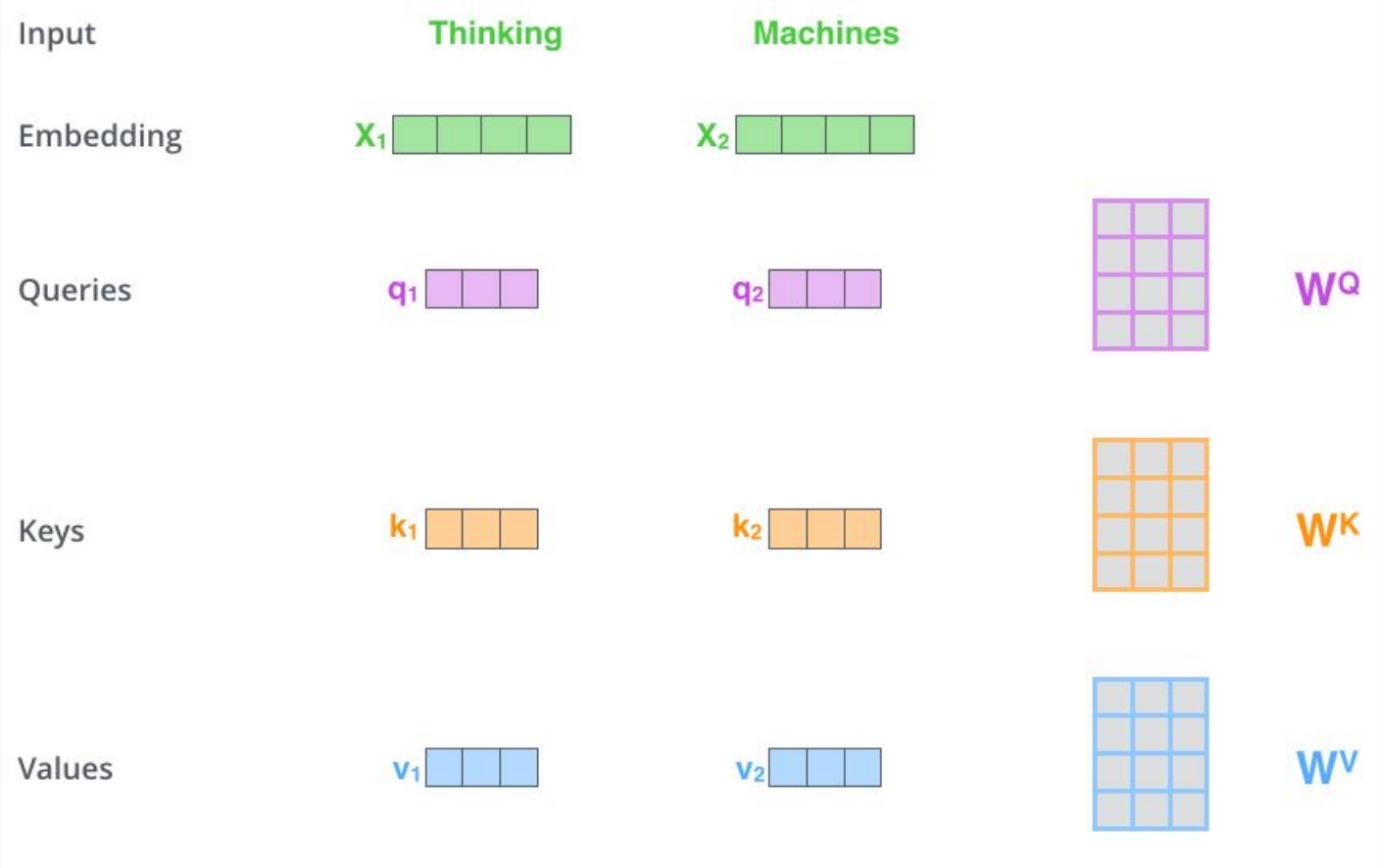
High Level Look



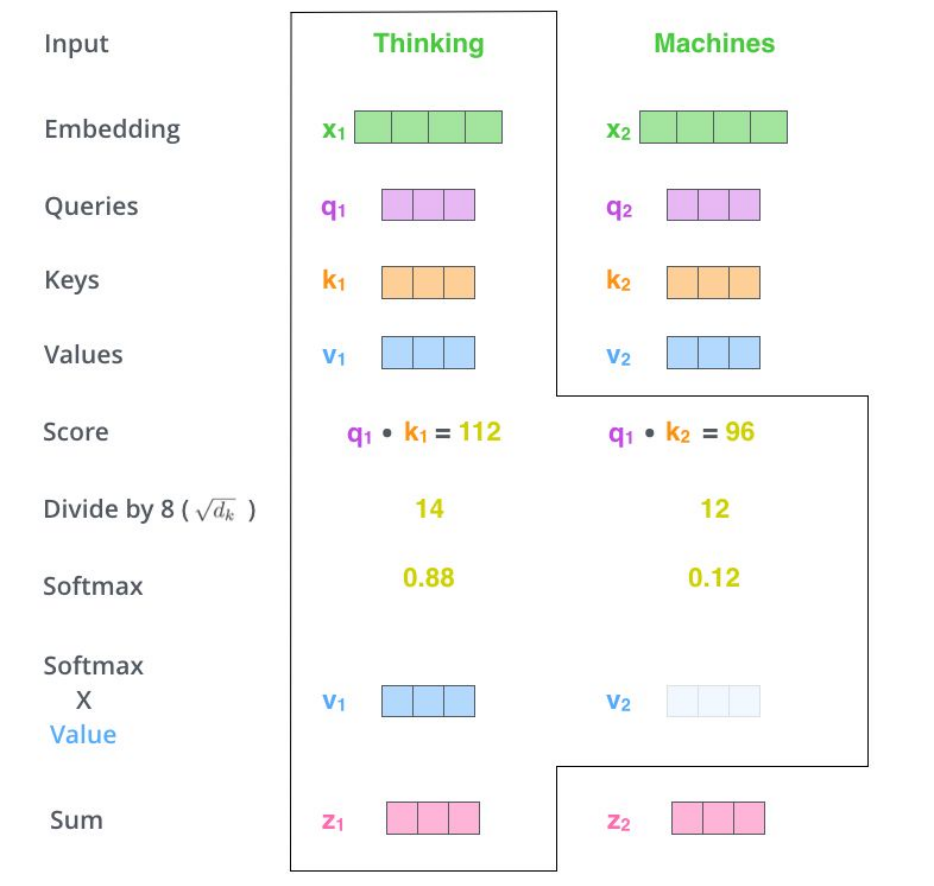
Closer Look into the Transformer



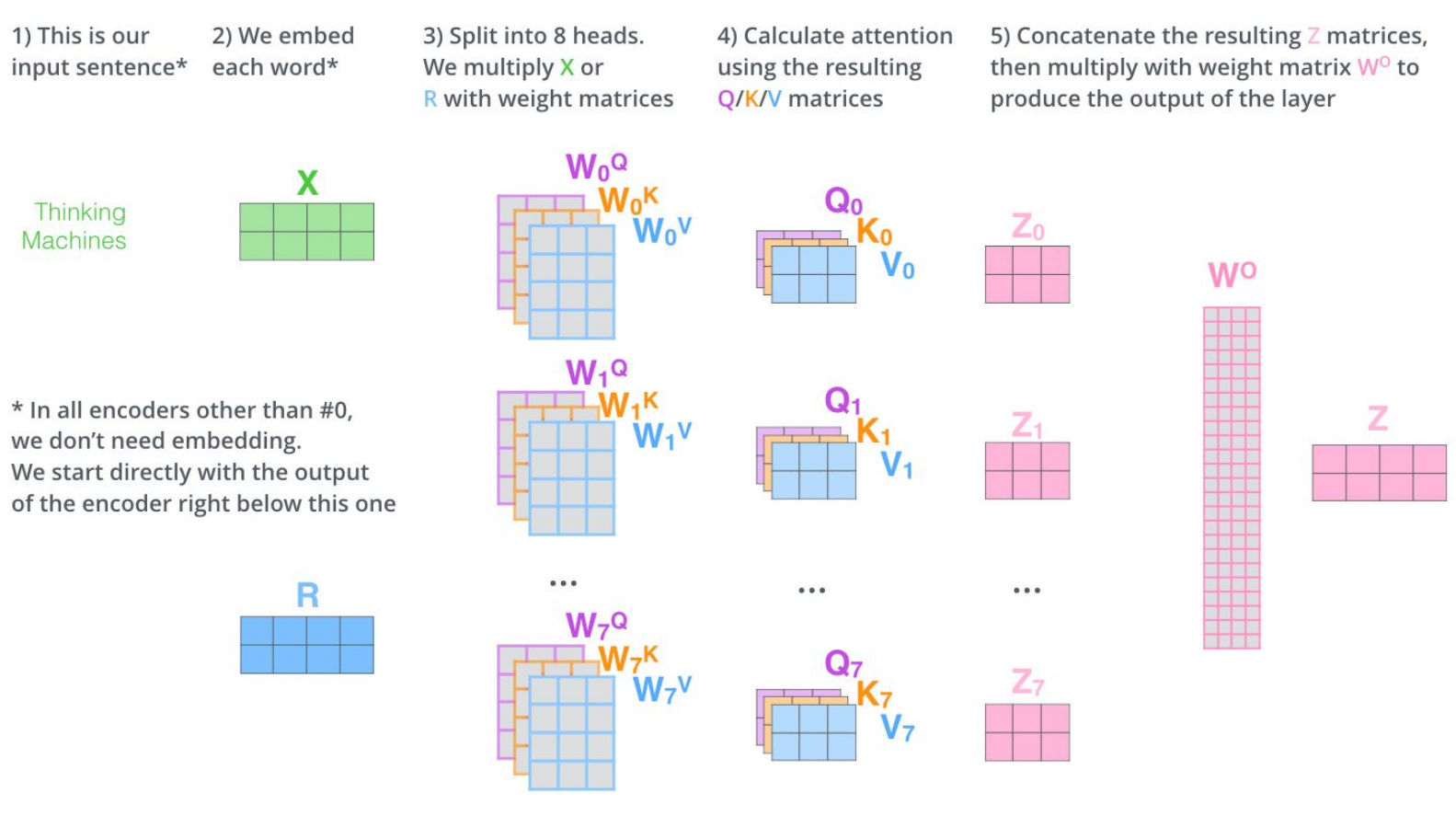
Self-Attention



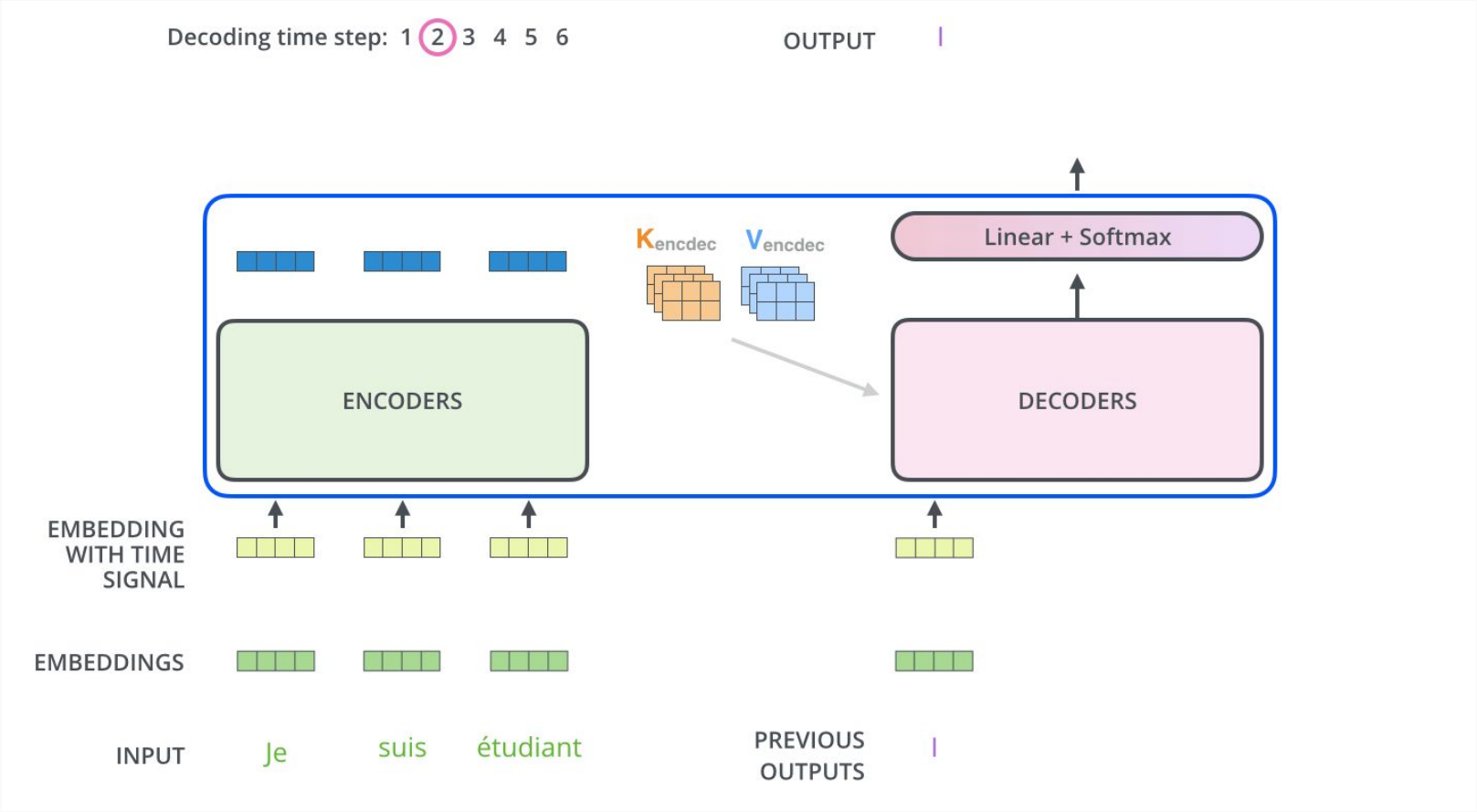
Self-Attention



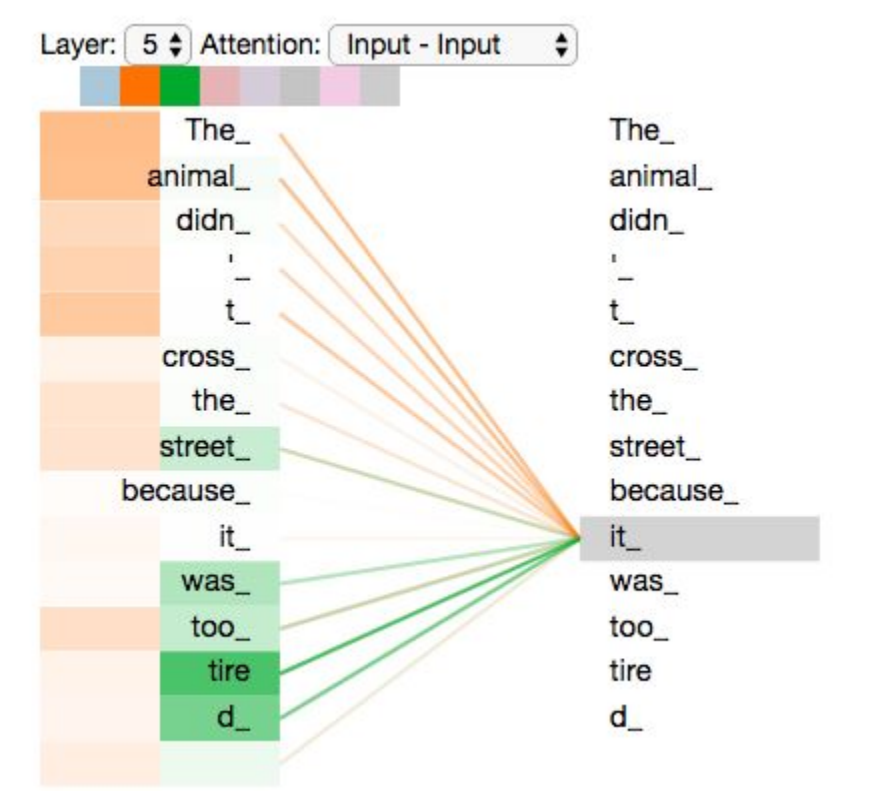
Multi-head Self-Attention



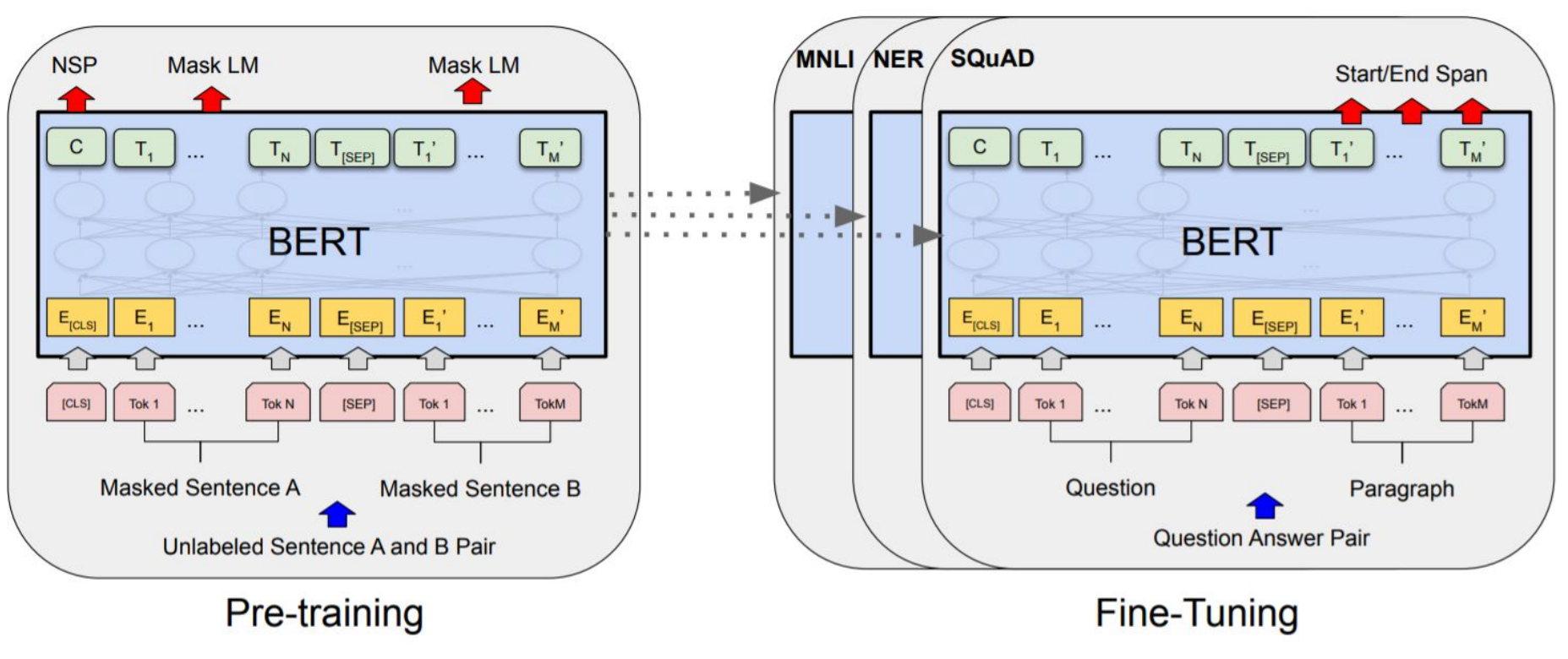
Closer Look into the Transformer



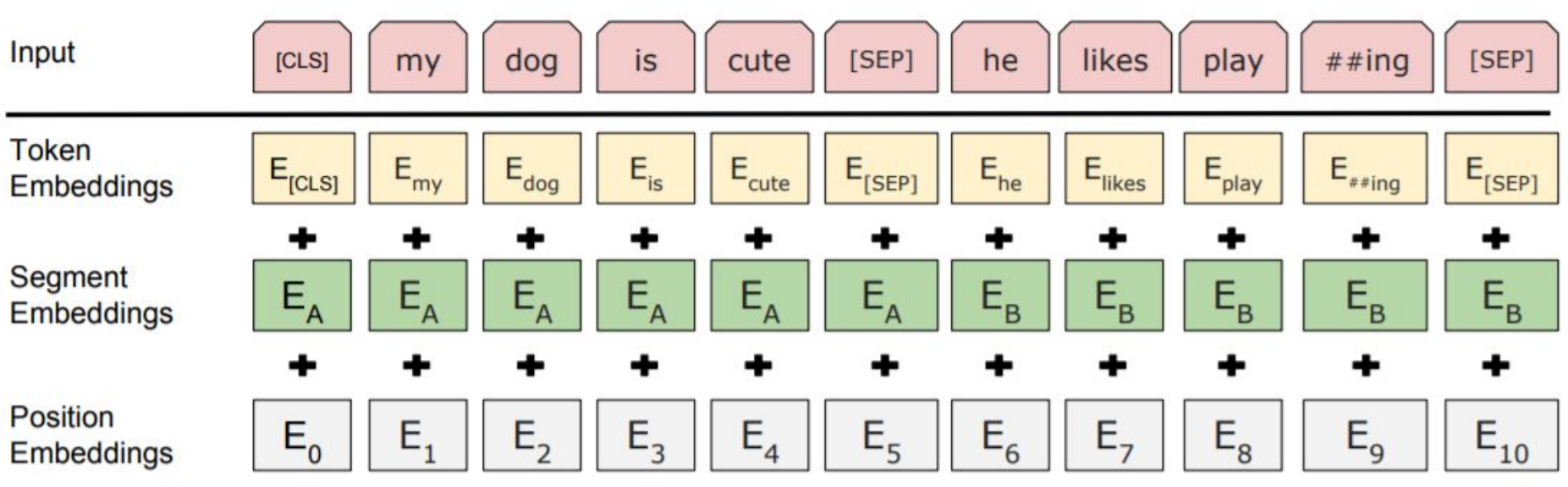
Attention Visualization



Pre-training



BERT Embedding



Conclusion, results, takeaways

- Context meaning can captured
- Superhuman results at various tasks
- Lot of parameters (largest models have ~ 11B params)
  - Large corpus is needed (C4 dataset has ~ 750 GB)
  - The model has great capacity
- The computation can be parallelised

Rank	Name	Model	URL	Score
1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8
2	T5 Team - Google	T5		88.9
3	Facebook AI	RoBERTa		84.6
4	IBM Research AI	BERT-mtl		73.5
5	SuperGLUE Baselines	BERT++		71.5
		BERT		69.0
		Most Frequent Class		47.1
		CBoW		44.5
		Outside Best		-

SuperGLUE Benchmark (super.gluebenchmark.com)

Reading Comprehension with Commonsense Reasoning Dataset

PASSAGE

(**CNN**) -- A day after her sister **Serena**'s comeback was ended in **Eastbourne**, **Venus Williams** suffered a similar fate losing in the quarterfinals to **Daniela Hantuchova**. The **Slovak** battled hard in blustery conditions on the south coast of **England** as she recorded a 6-2 5-7 6-2 win -- her first over **Venus** in 11 meetings. **Williams** had been out of action for five months with an abdominal injury before returning for the warm-up tournament ahead of **Wimbledon** and showed flashes of her old self in the second set. **Hantuchova** told the **WTA**'s official web site: "I was not thinking about our other matches at all. I was just focusing on my game today.

- **Daniela Hantuchova** knocks **Venus Williams** out of **Eastbourne** 6-2 5-7 6-2
- It is the first time **Hantuchova** has beaten **Williams** in 11 matches
- **Slovak** will now face fifth seed **Petra Kvitova** after she beat **Agnieszka Radwanska**
- **Mario Bartoli** will face **Australian Sam Stosur** in other semifinal

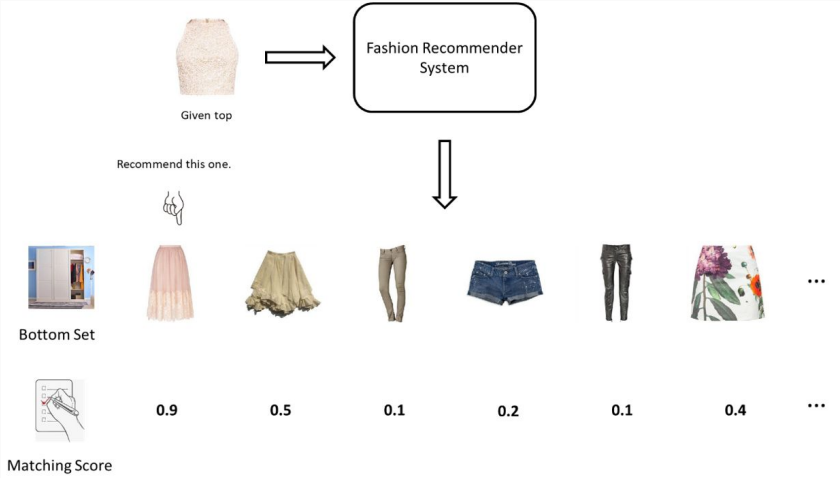
QUERY

Hantuchova breezed through the first set in just under 40 minutes after breaking Williams' serve twice to take it 6-2 and led the second 4-2 before **X** hit her stride.

Fashion Recommendation

Challenges:

- Just a few “standard” attributes
- Various customers’ preferences
- Changing trends
- Short lifetime of items



Approaches

1. Rule-based
2. **Machine Learning**
    - Usually uses CNN for images features extraction

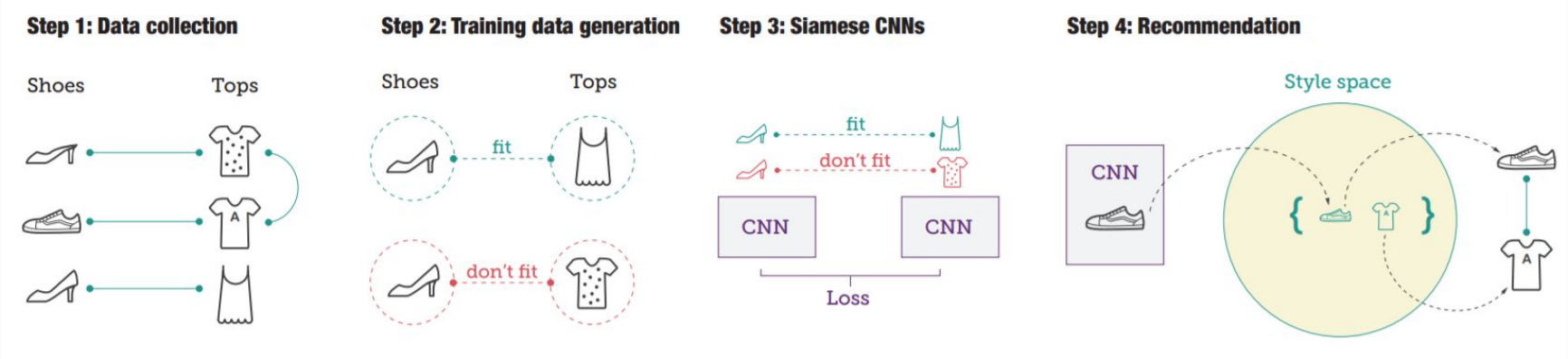
Examples:

- Sequence modelling task - Bi-LSTM network
- Fixed number of items in one outfit with fully-connected NN
- Dyadic Co-occurrences (Siamese Network)

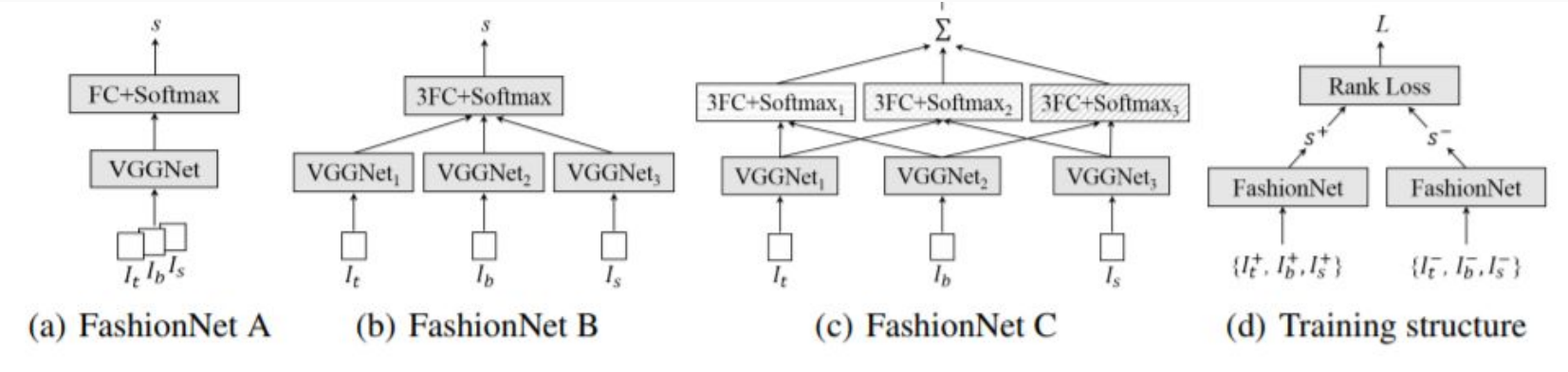
Polyvore Dataset

- 21 889 user-created outfits from polyvore.com
- Contains images and basic information about the products

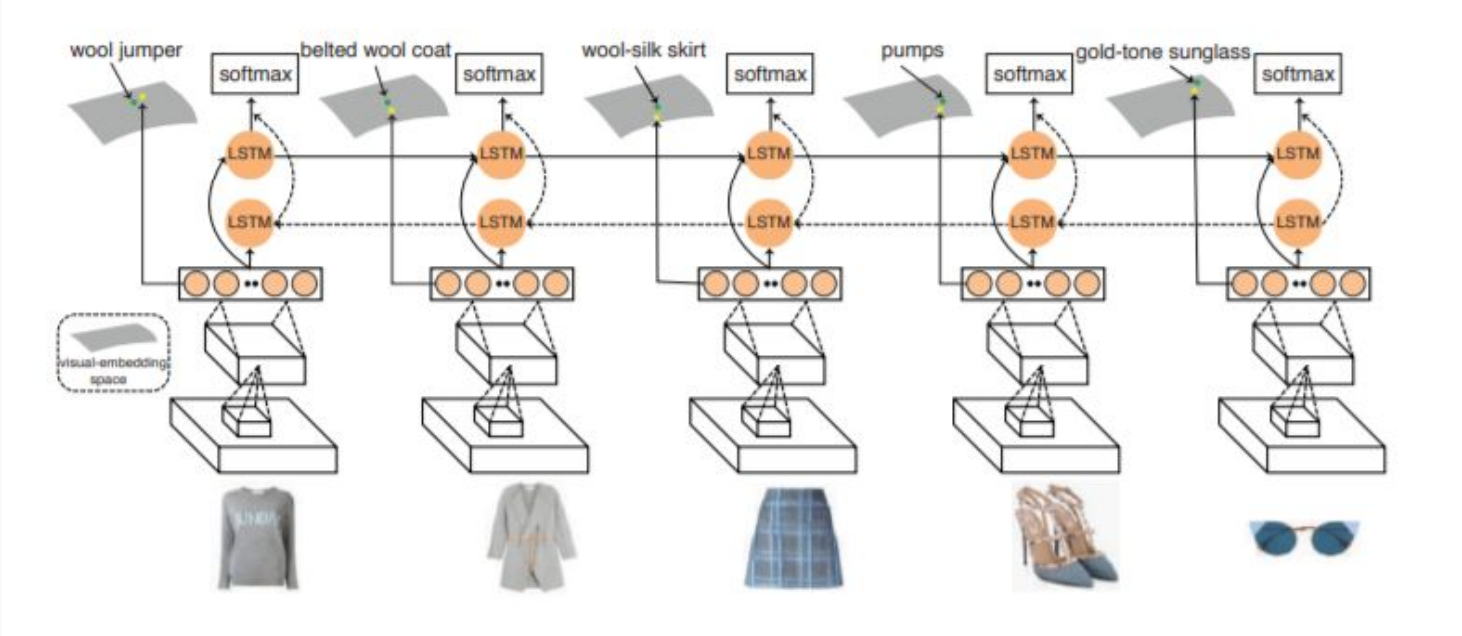
Dyadic Co-occurrences (Siamese Network)



Fixed number of items in one outfit with fully-connected NN



Sequence modelling task (Bi-LSTM network)



Self-attention for Outfit Generation

Intuition behind the self-attention:

- When choosing shoes, pay attention to the color of the belt

Approach:

- Extract features with CNN
- Treat fashion item as a word and an outfit as its context (sentence)
- Use special tokens/embeddings representing product category
- Train on masked outfits

Challenges

- How to embed fashion products?
  - NLP transformers use sub-word embedding and the dictionary has a fixed size
  - Are image representations learned on a classification task good enough?
  - What about positional embeddings?
- Which part of transformer to use?
  - GPT uses only decoder blocks
  - BERT uses only encoder blocks
  - T5 states that the full transformer is the most convenient choice
- How big?
  - Relatively small dataset
  - A lot of parameters

InceptionV3 ImageNet embeddings

Style2Vec

References

- Vaswani, Ashish et al. “Attention is All you Need.” NIPS (2017). Devlin, Jacob et al. “BERT: Pre-training of Deep
- Bidirectional Transformers for Language Understanding.” NAACL-HLT (2019). Raffel, Colin et al. “Exploring the
- Limits of Transfer Learning with a Unified Text-to-Text Transformer.” ArXiv abs/1910.10683 (2019): n. pag.
- He, Tong, and Yang Hu. "FashionNet: Personalized Outfit Recommendation with Deep Neural Network." arXiv preprint arXiv:1810.02443 (2018).
- Han, Xintong et al. “Learning Fashion Compatibility with Bidirectional LSTMs.” Proceedings of the 2017 ACM on Multimedia Conference - MM ’17 (2017): n. pag. Crossref. Web.
- Veit, Andreas, et al. "Learning visual clothing style with heterogeneous dyadic co-occurrences." Proceedings of the IEEE International Conference on Computer Vision. 2015.
- Klein, Guillaume, et al. "Opennmt: Open-source toolkit for neural machine translation." arXiv preprint arXiv:1701.02810 (2017). (http://nlp.seas.harvard.edu/2018/04/03/attention.html)
- Alammar, Jay (2018). The Illustrated Transformer [Blog post]. Retrieved from https://jalammar.github.io/illustrated-transformer/

Thanks