

Transformers

Recap: Convolution

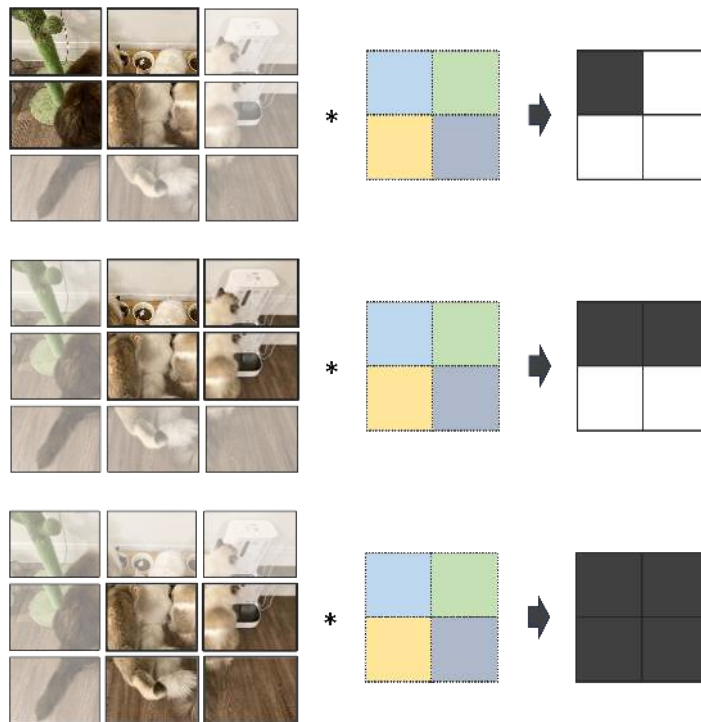
Input: $x \in \mathbb{R}^{C_1 \times H \times W}$

Output: $y \in \mathbb{R}^{C_2 \times (H-h+1) \times (W-w+1)}$

Parameters:

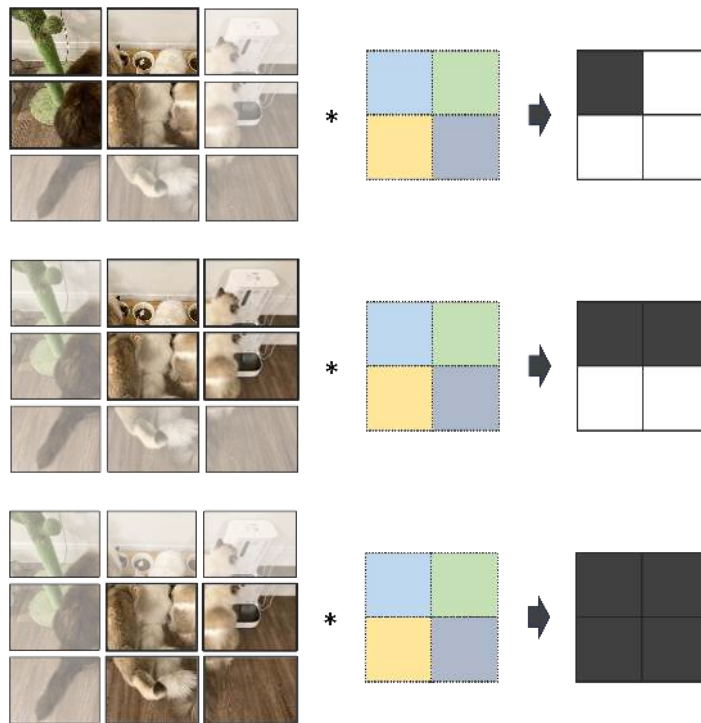
- Kernel: $\omega \in \mathbb{R}^{C_1 \times C_2 \times h \times w}$
- Bias: (optional) $b \in \mathbb{R}^{C_2}$

$$\underbrace{y_{i,j,k}}_{\text{output}} = \underbrace{b_i}_{\text{bias}} + \sum_{l=1}^{C_1} \sum_{m=0}^{h-1} \sum_{n=0}^{w-1} \underbrace{x_{l,j+m,k+n}}_{\text{input}} \cdot \underbrace{\omega_{i,l,m,n}}_{\text{kernel}}$$



Recap: Convolution

Good at parsing relatively **structured** patterns



A New Problem

Sentiment Analysis for Movie Reviews

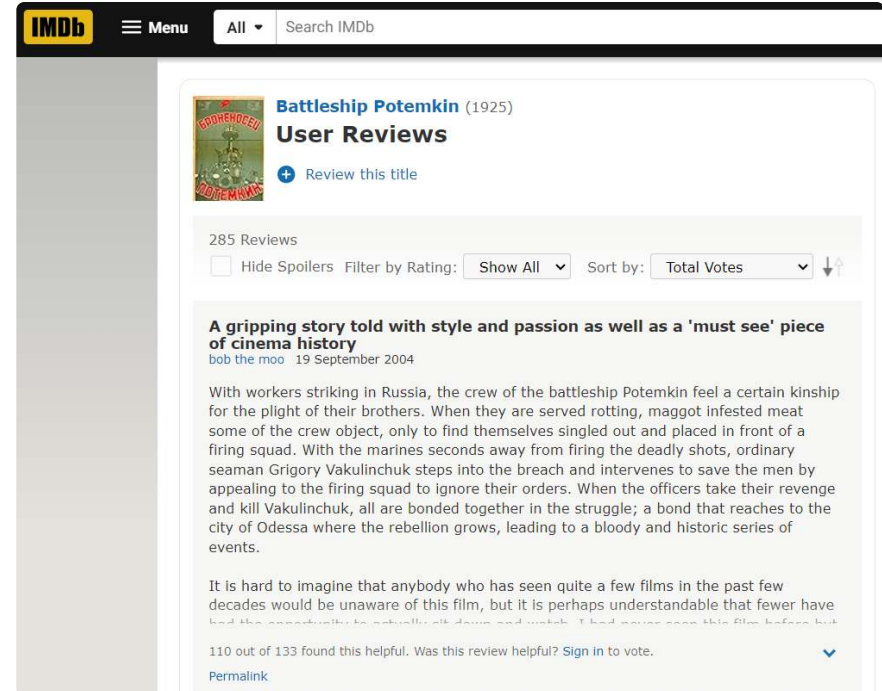
- Predict if review is positive 😊 or negative 😞

😊 My kid likes this movie

😊 My little kid likes this animated movie

😞 My kid does not like this movie

Real-world examples: reviews on IMDB¹:



IMDb Menu All Search IMDb

Battleship Potemkin (1925)
User Reviews
+ Review this title

285 Reviews
 Hide Spoilers Filter by Rating: Show All Sort by: Total Votes

A gripping story told with style and passion as well as a 'must see' piece of cinema history
bob the moo 19 September 2004

With workers striking in Russia, the crew of the battleship Potemkin feel a certain kinship for the plight of their brothers. When they are served rotting, maggot infested meat some of the crew object, only to find themselves singled out and placed in front of a firing squad. With the marines seconds away from firing the deadly shots, ordinary seaman Grigory Vakulinchuk steps into the breach and intervenes to save the men by appealing to the firing squad to ignore their orders. When the officers take their revenge and kill Vakulinchuk, all are bonded together in the struggle; a bond that reaches to the city of Odessa where the rebellion grows, leading to a bloody and historic series of events.

It is hard to imagine that anybody who has seen quite a few films in the past few decades would be unaware of this film, but it is perhaps understandable that fewer have had the opportunity to actually sit down and watch. I had never seen this film before but

110 out of 133 found this helpful. Was this review helpful? Sign in to vote.

Permalink

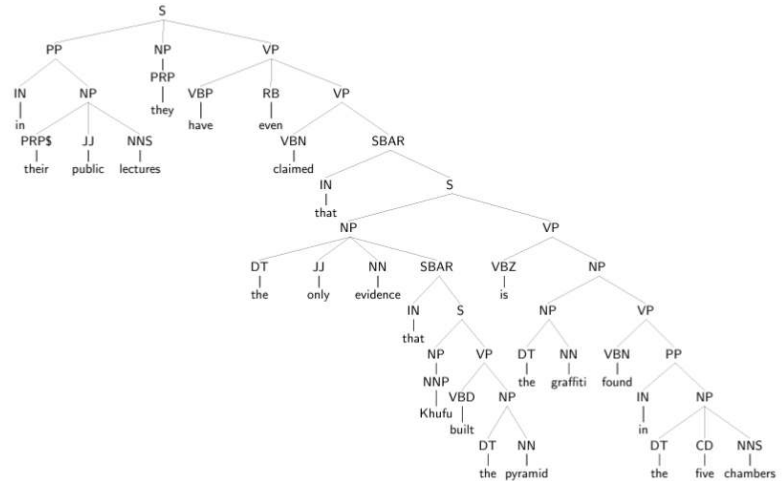
How Is the Task Different From Previous Examples

Images

- Fixed input (resolution)
- Fixed structure

Language

- Variable length
- Diverse structure (tree syntax¹!)



Solution: Transformer

A deep network based on the *attention* mechanism¹:

