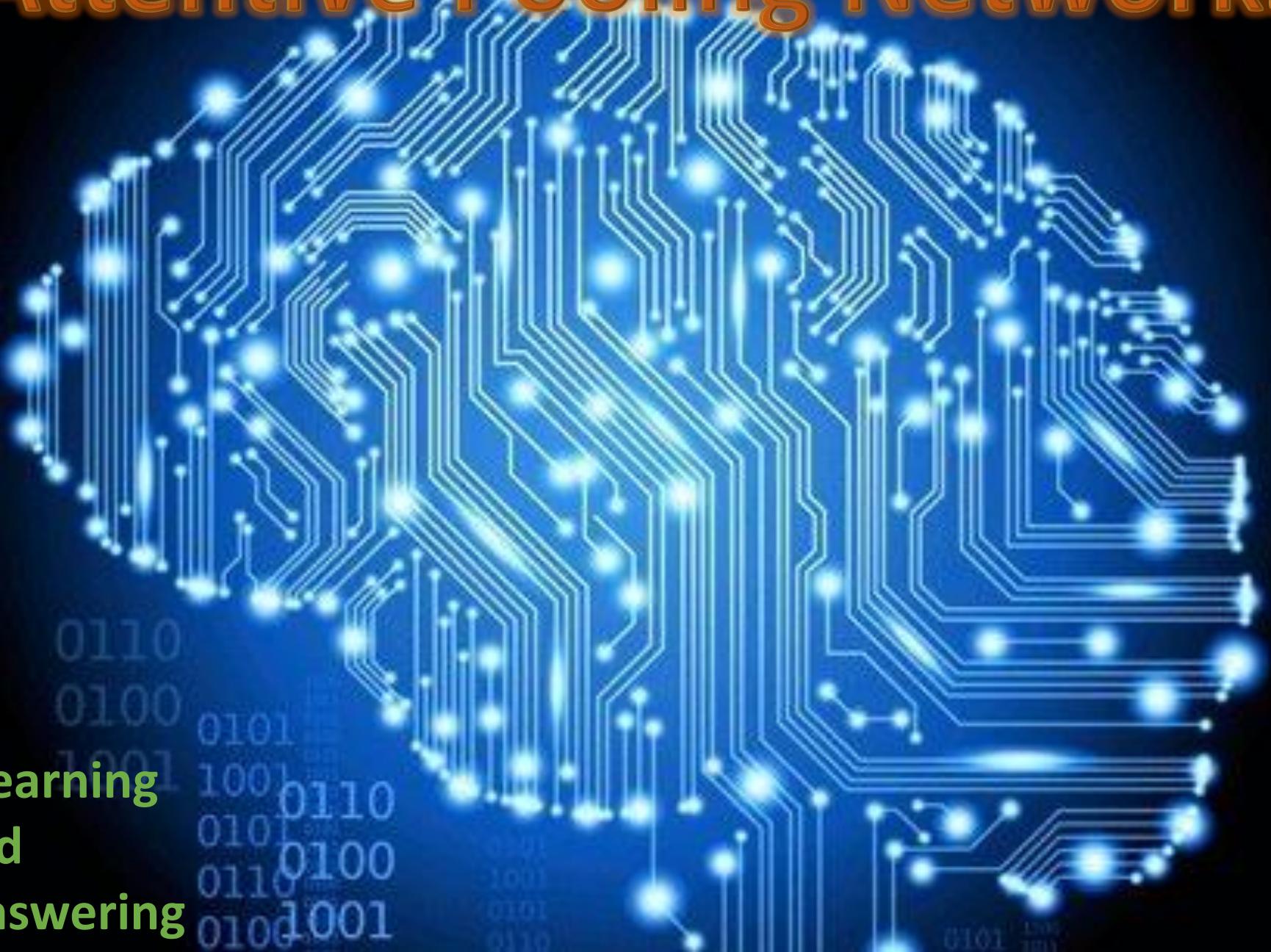


Attentive Pooling Networks



Machine Learning
based
Question Answering

Eran Raveh
15-Jun-2016

What we had so far...

Answering

End-to-end systems
(semi-)rule-based
transparent models (w/ some linguistic motivations)
Very little use of deep learning

Generating

Patterns + deep (?) learning
Generalizes based on knowledgebase



Overview for Today

Attentive Pooling Network

IBM Watson team (2016)

Santos, Cicero dos, et al. "Attentive pooling networks." arXiv preprint arXiv:1602.03609 (2016).

Introducing a two-way (question \leftrightarrow answer) attention mechanism, which makes the network aware of the current input



Motivation

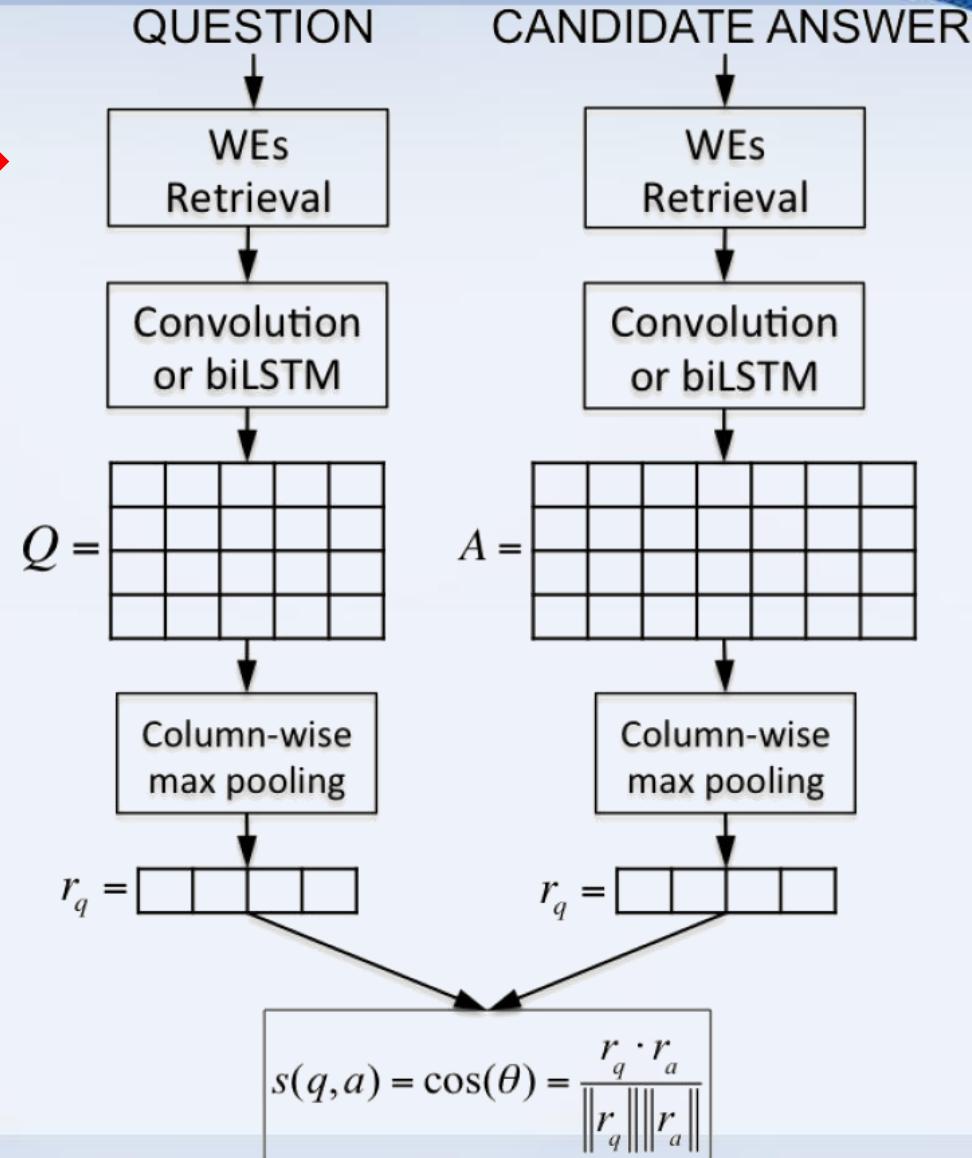
- ❖ Use attention mechanism for pair-wise ranking tasks
- ❖ Make the system aware of current candidate (bi-directional attention)
- ❖ Joint learning of inputs and their similarity
- ❖ **Remember:** at the end, it's a find-best-candidate task



QA-CNN

Common to most NLP tasks →

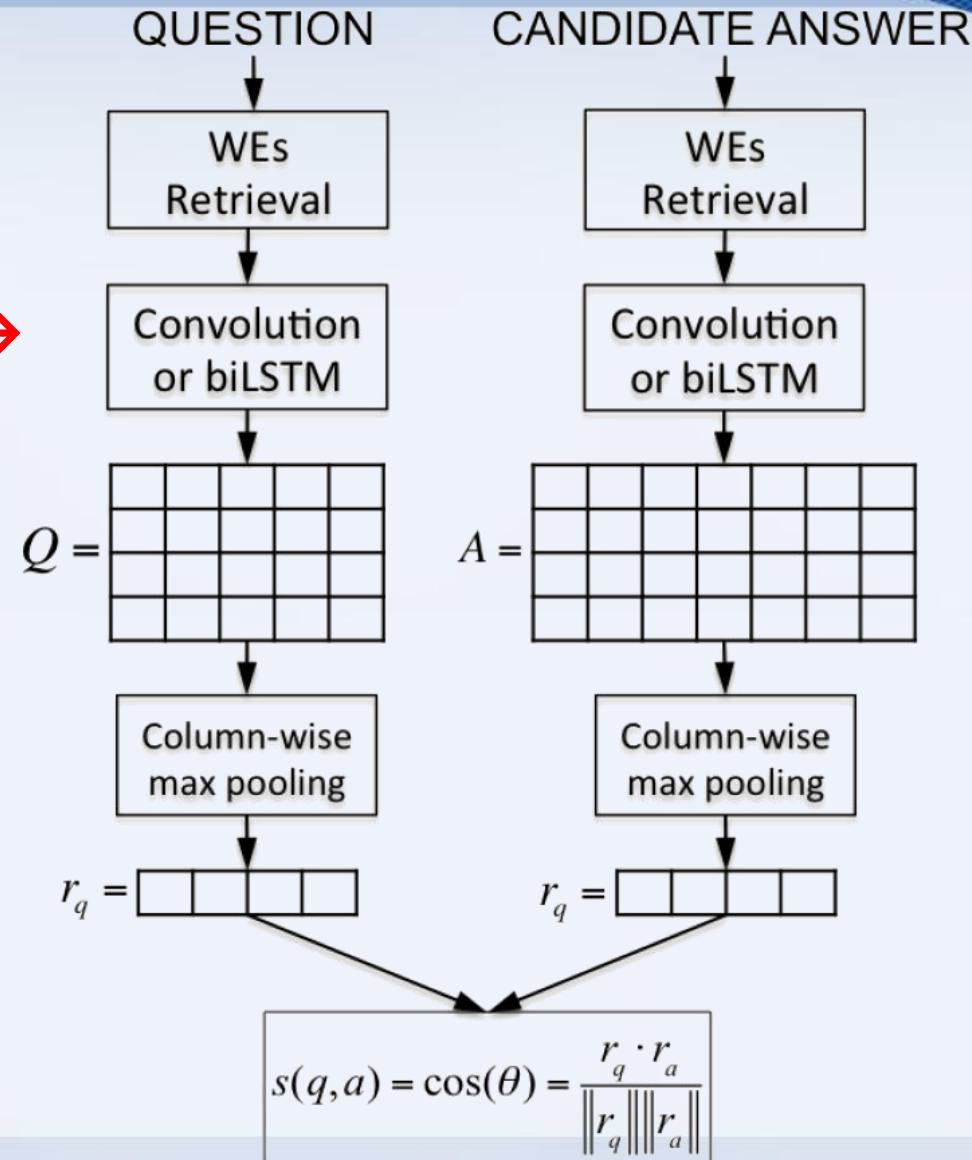
Goal – represent text as vectors. In this case, word-level moving window



QA-CNN

Pattern recognizing →

Goal – learn feature patterns using
pre-defined filters
(only convolution will be demonstrated here)



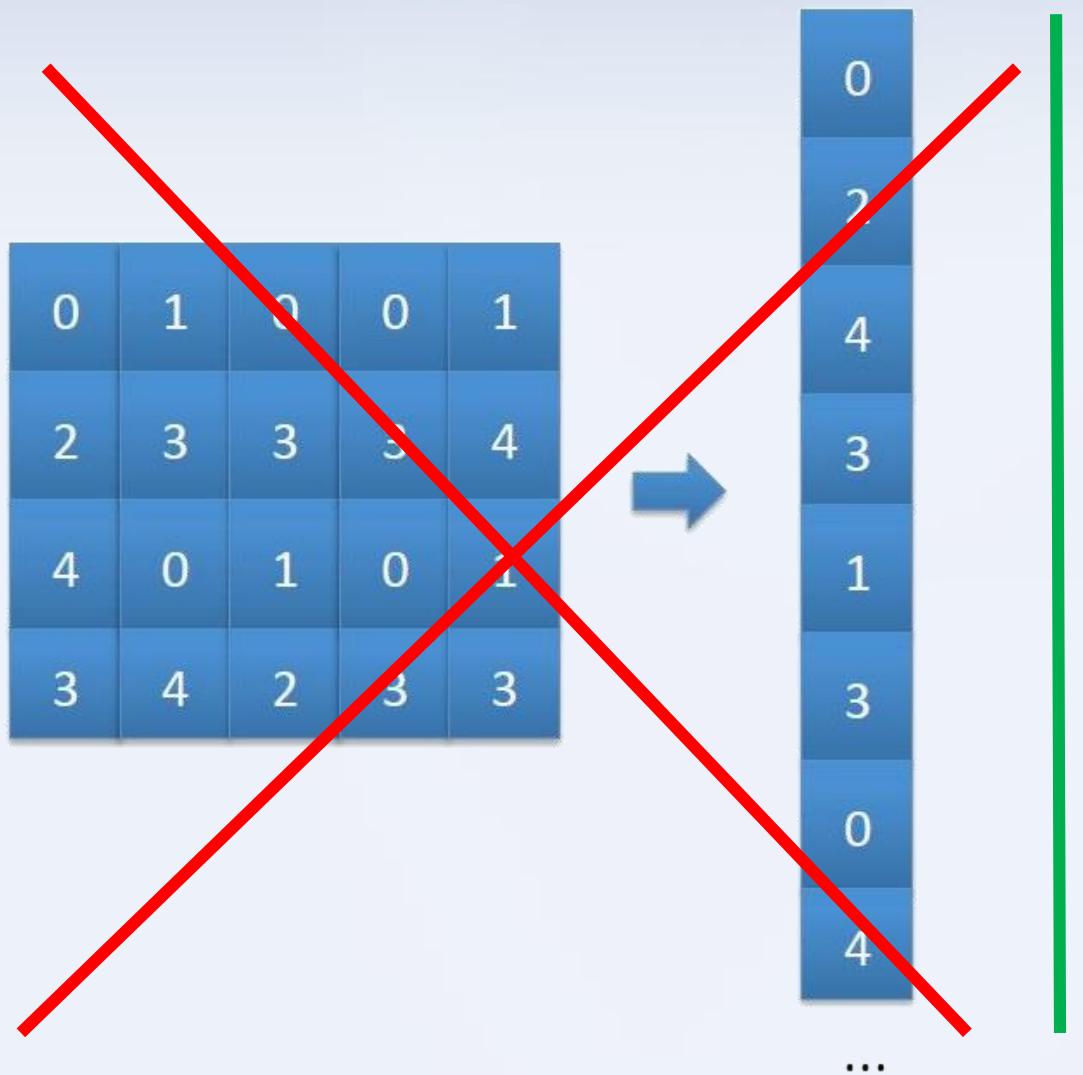
Convolutional Neural Networks (CNNs)

Convolutional Neural Nets

- ❖ A deep learning technique (popular since 2012)
- ❖ Matrix as input
- ❖ Used for learning patterns (a.o.t. sequences), e.g. for digit recognition, image classification etc.
- ❖ Relatively small number of parameters
- ❖ Typically the input of a fully connected DNN



Why a Matrix?



- ❖ Parameter explosion
- ❖ Order important for pattern recognition



Filters

35	40	41	45	50
40	40	42	46	52
42	46	50	55	55
48	52	56	58	60
56	60	65	70	75



0	1	0		
0	0	0		
0	0	0		

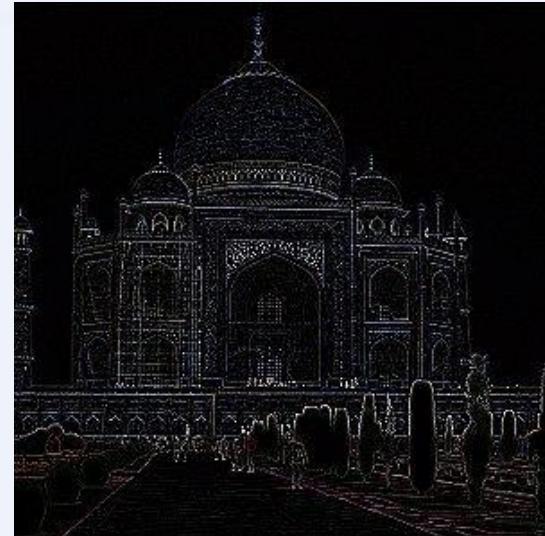
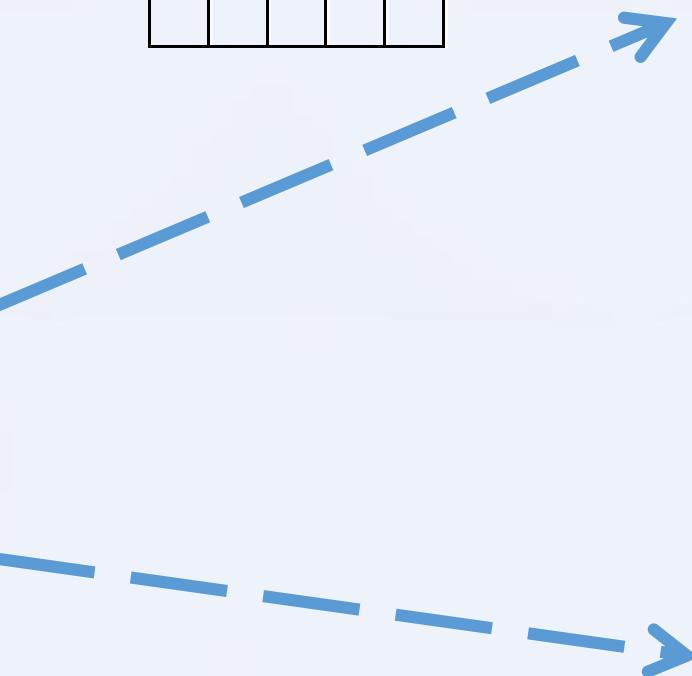
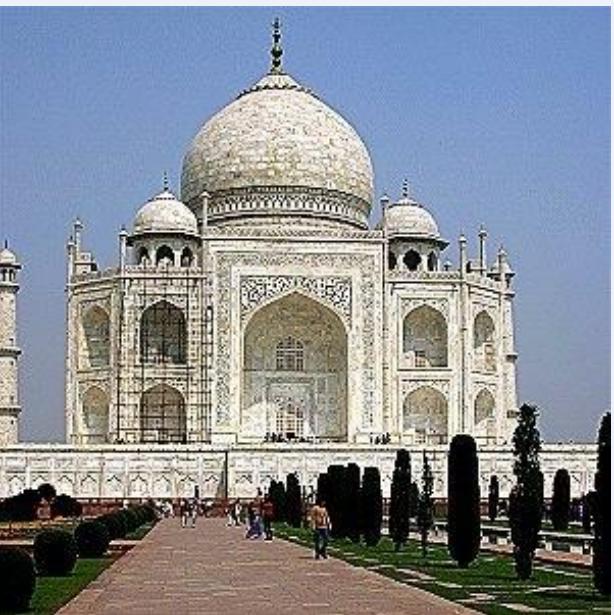


42				



Blurring

0	0	0	0	0
0	1	1	1	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0



Edge-detection

0	1	0
1	-4	1
0	1	0

<https://docs.gimp.org/en/images/filters/examples/generic-taj-convmatrix-sharpen.jpg; /generic-taj-convmatrix-edge-detect.jpg; /generic-taj-convmatrix-blur.jpg>



UNIVERSITÄT
DES
SAARLANDES

Convolutions

- ❖ Applying a filter to a matrix input result in a matrix of convolved feature
- ❖ Typically, multiple filters are used

1 x1	1 x0	1 x1	0	0
0 x0	1 x1	1 x0	1	0
0 x1	0 x0	1 x1	1	1
0	0	1	1	0
0	1	1	0	0

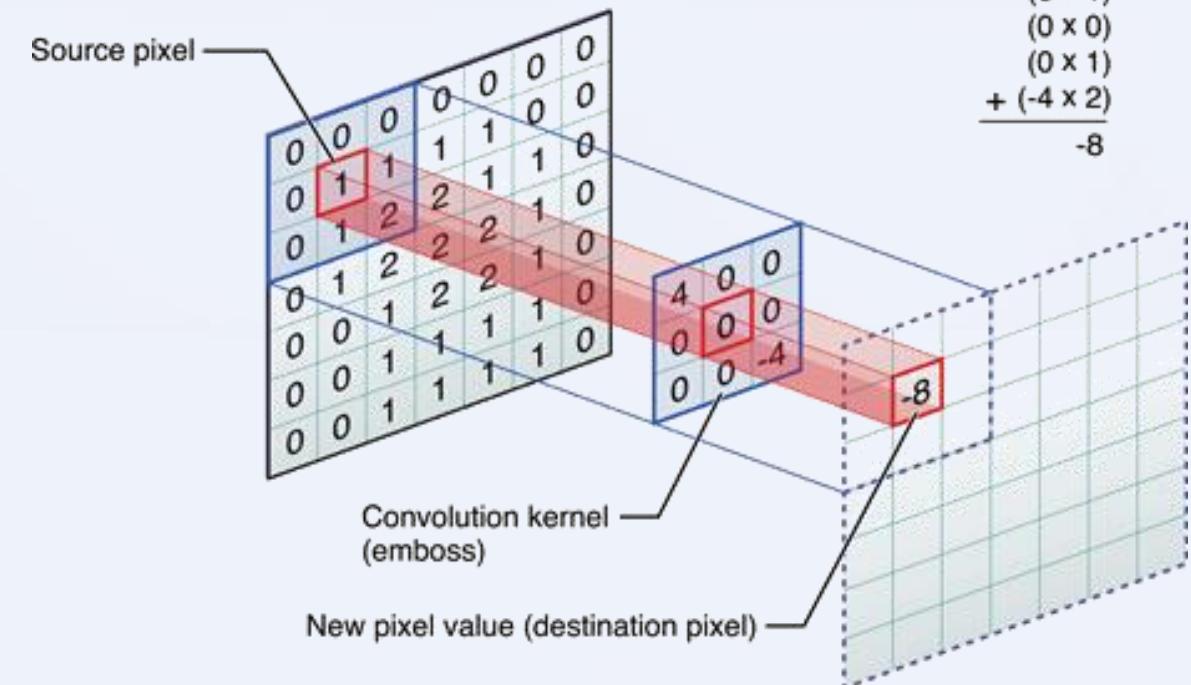
Image

4		

Convolved Feature

Center element of the kernel is placed over the source pixel. The source pixel is then replaced with a weighted sum of itself and nearby pixels.

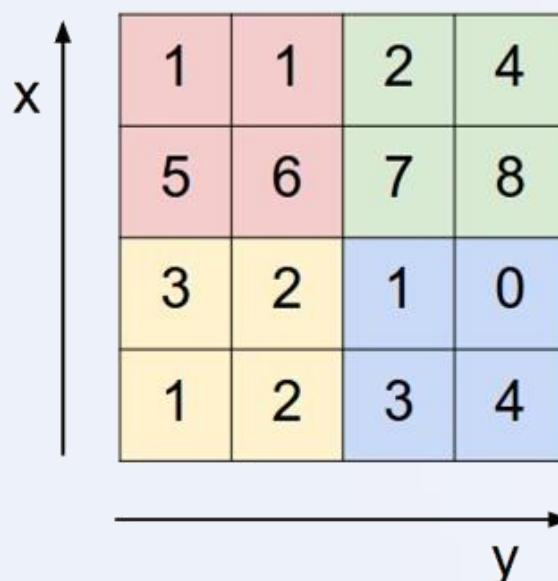
$$\begin{array}{r} (4 \times 0) \\ (0 \times 0) \\ (0 \times 0) \\ (0 \times 0) \\ (0 \times 1) \\ (0 \times 1) \\ (0 \times 0) \\ (0 \times 1) \\ + (-4 \times 2) \\ \hline -8 \end{array}$$



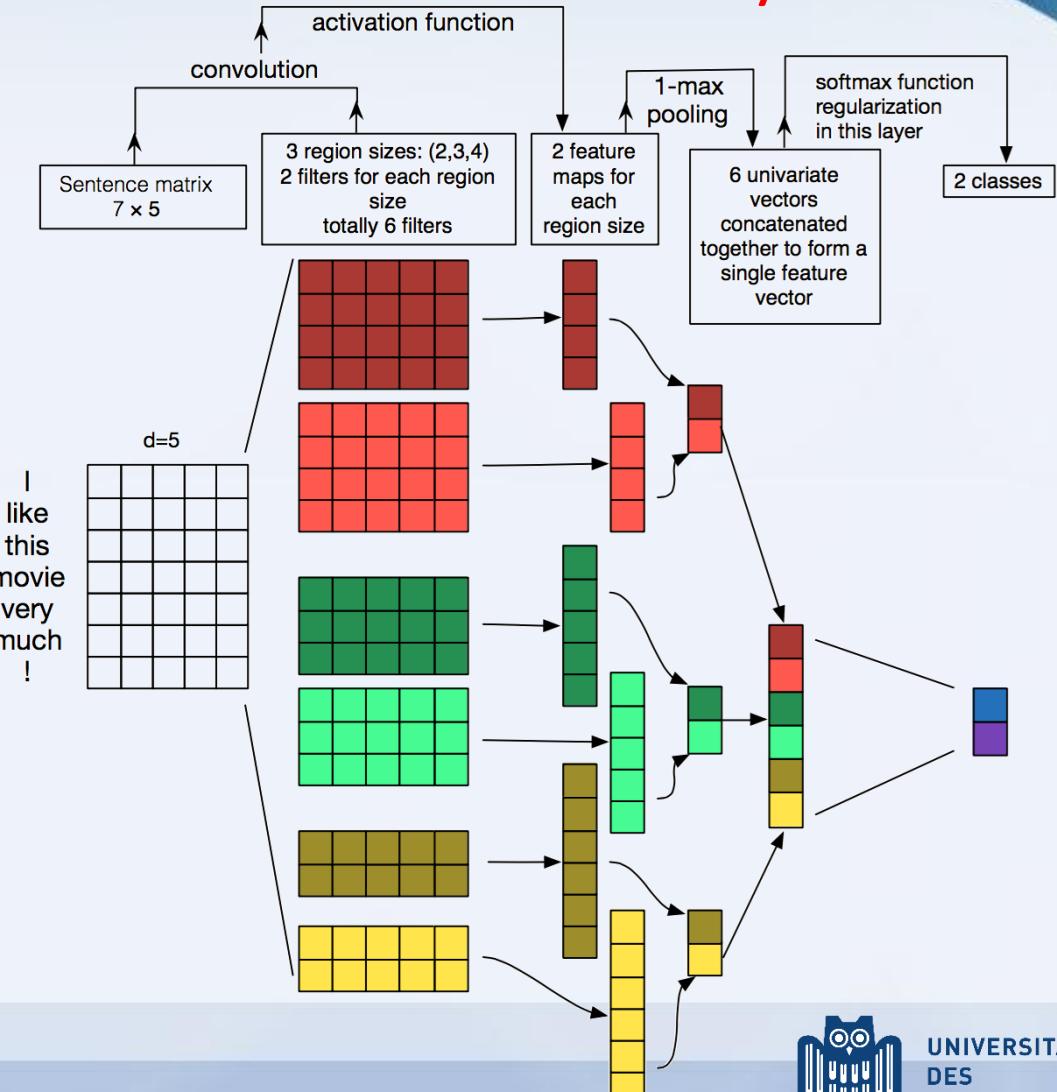
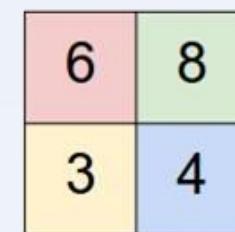
See also: <http://cs231n.github.io/assets/conv-demo/index.html>

Max Pooling

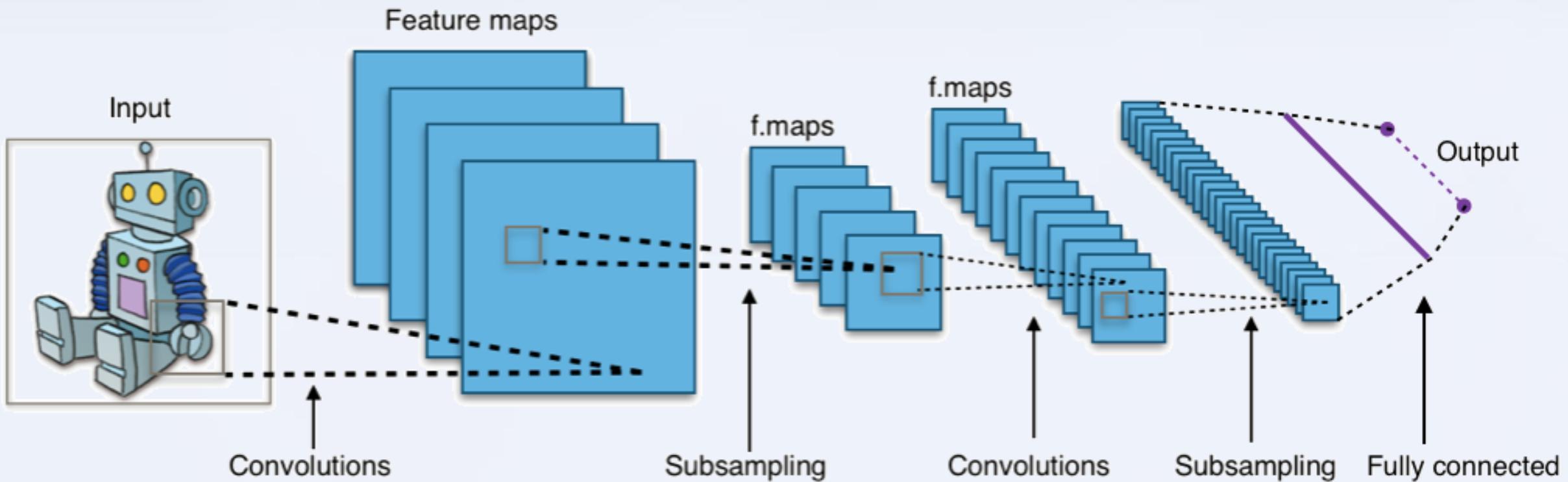
- ❖ From the convolved features, take those that are most dominant. Why?



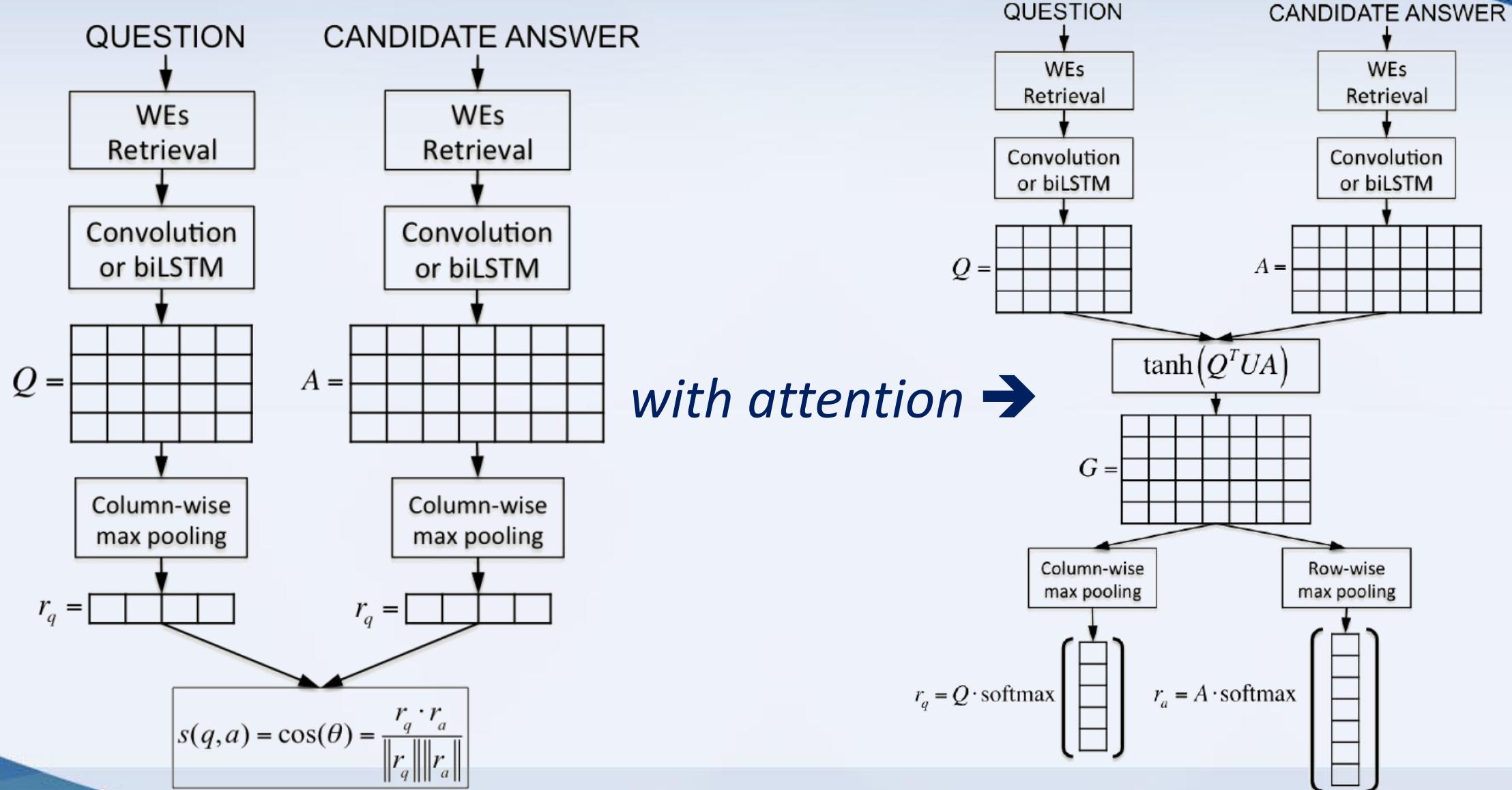
max pool with 2x2 filters
and stride 2



CNN – The Whole Picture



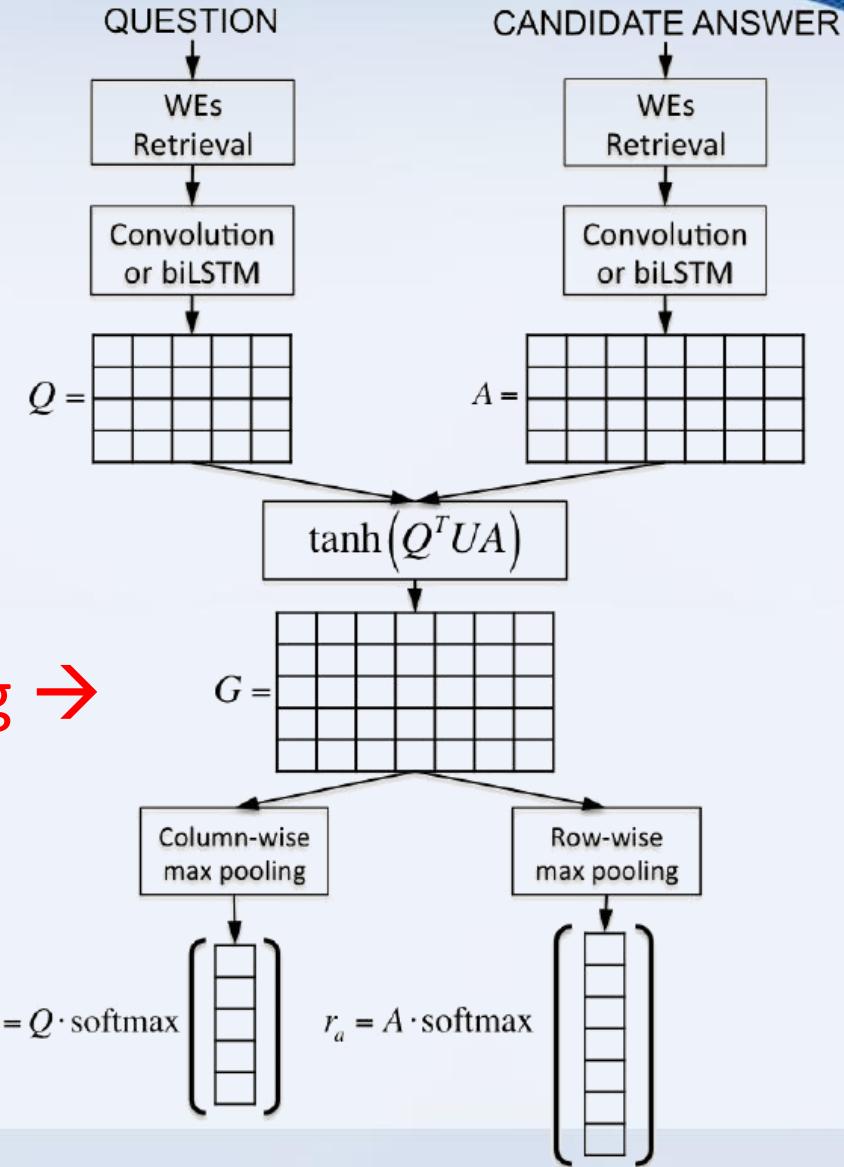
From QA-CNN to AP-CNN



AP-CNN

Goal – Let the question and the answer directly influence on the representation of one another

Attentive pooling →



Attentive Neural Networks

Paying Attention

Attention – From Latin: *attendere / attentio*

The act or state of applying the mind to something.
A condition of readiness for such attention
involving especially a selective narrowing or
focusing of consciousness and receptivity

(Merriam-Webster Dictionary)



Paying Attention

Use only the part of the Information that is relevant to the task



More Examples



**What is the color of the
coat?**



**What is the color of the
umbrella?**

More Examples [2]



A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

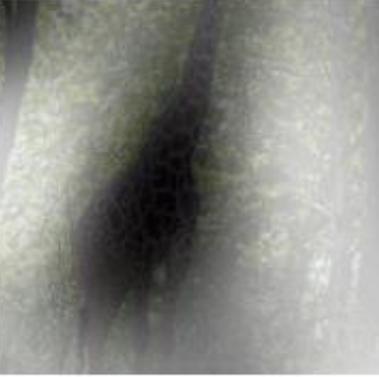
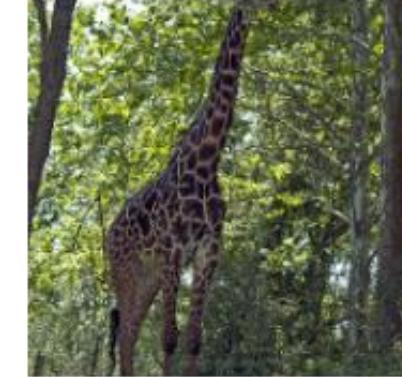
A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.

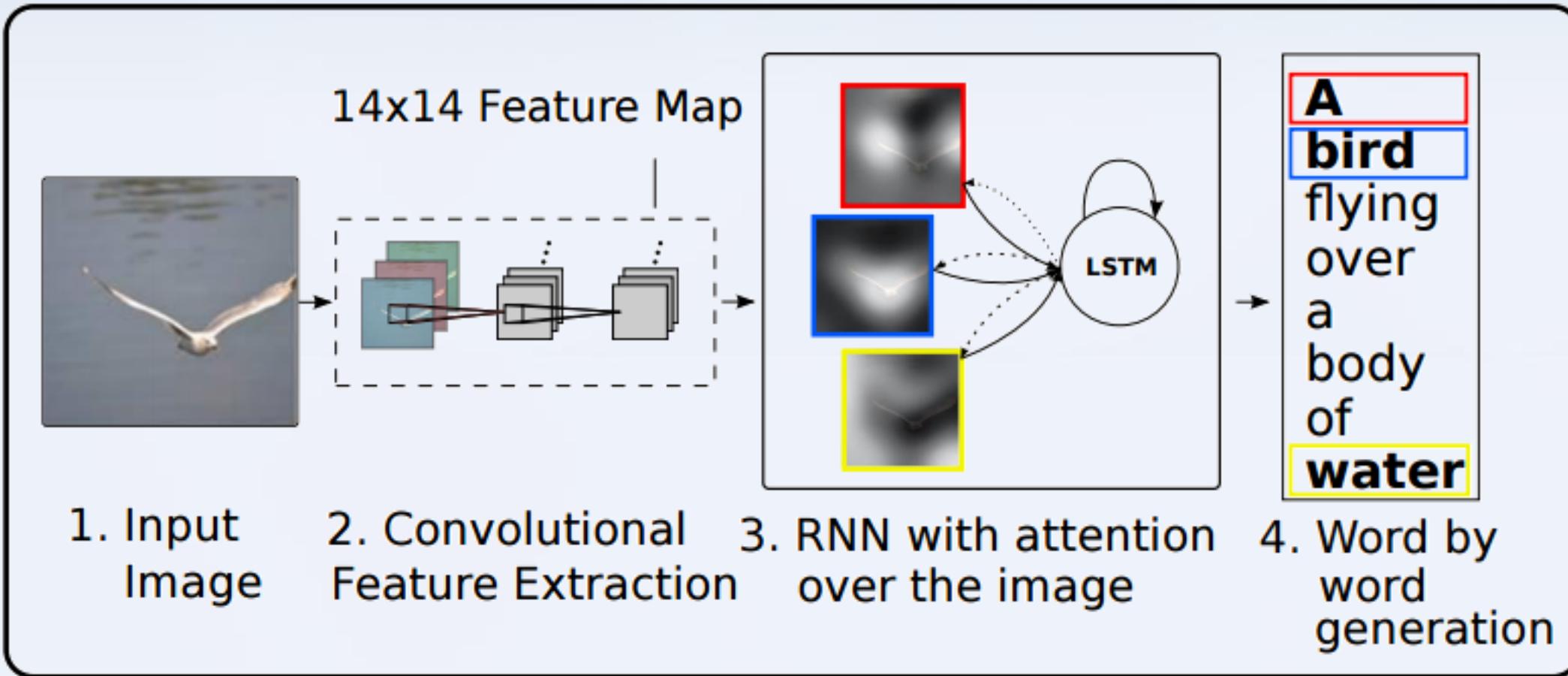


A group of people sitting on a boat in the water.

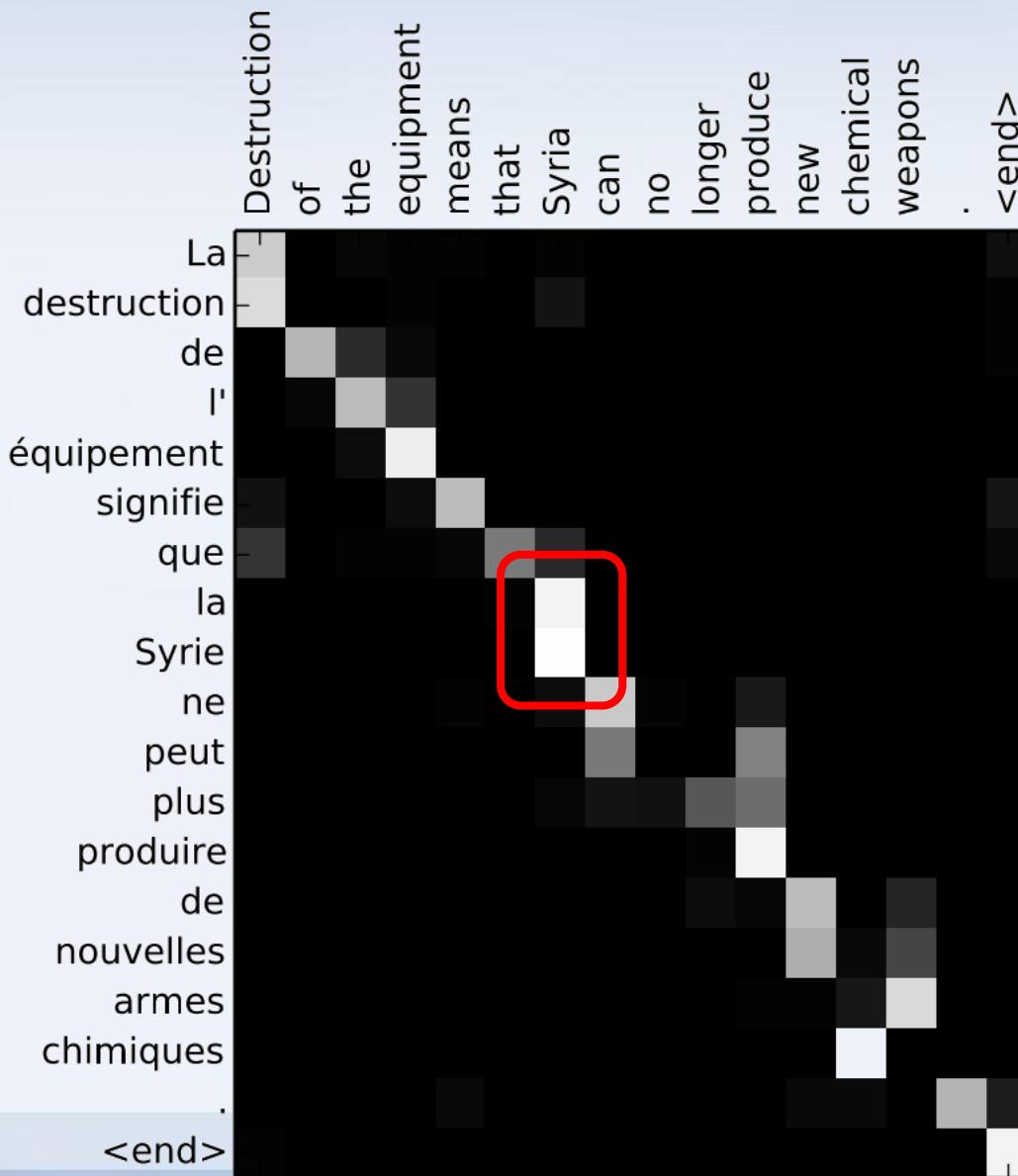


A giraffe standing in a forest with trees in the background.

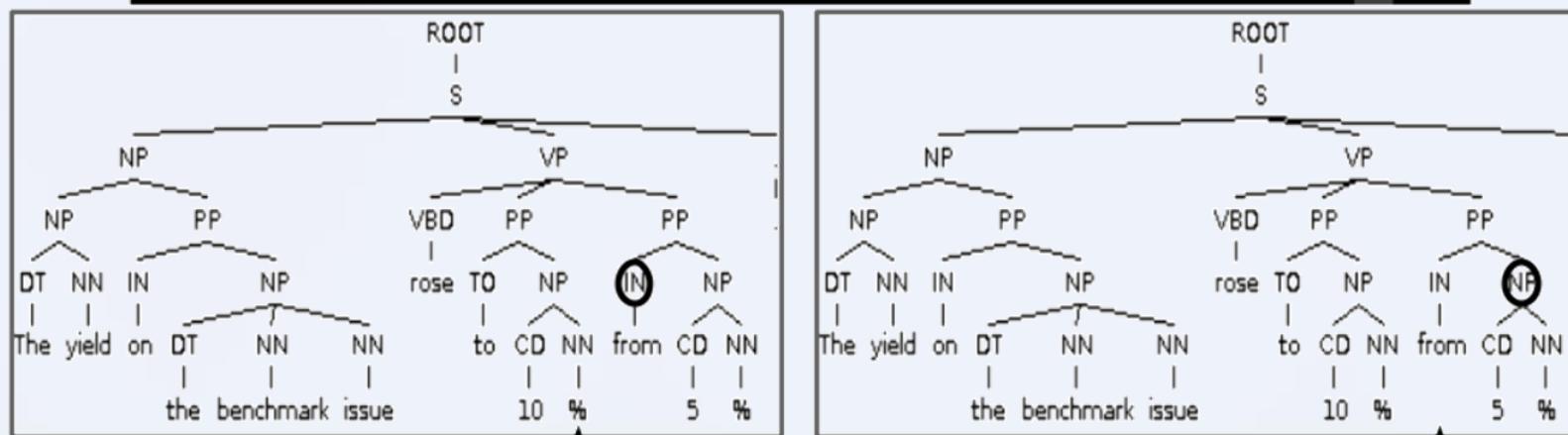
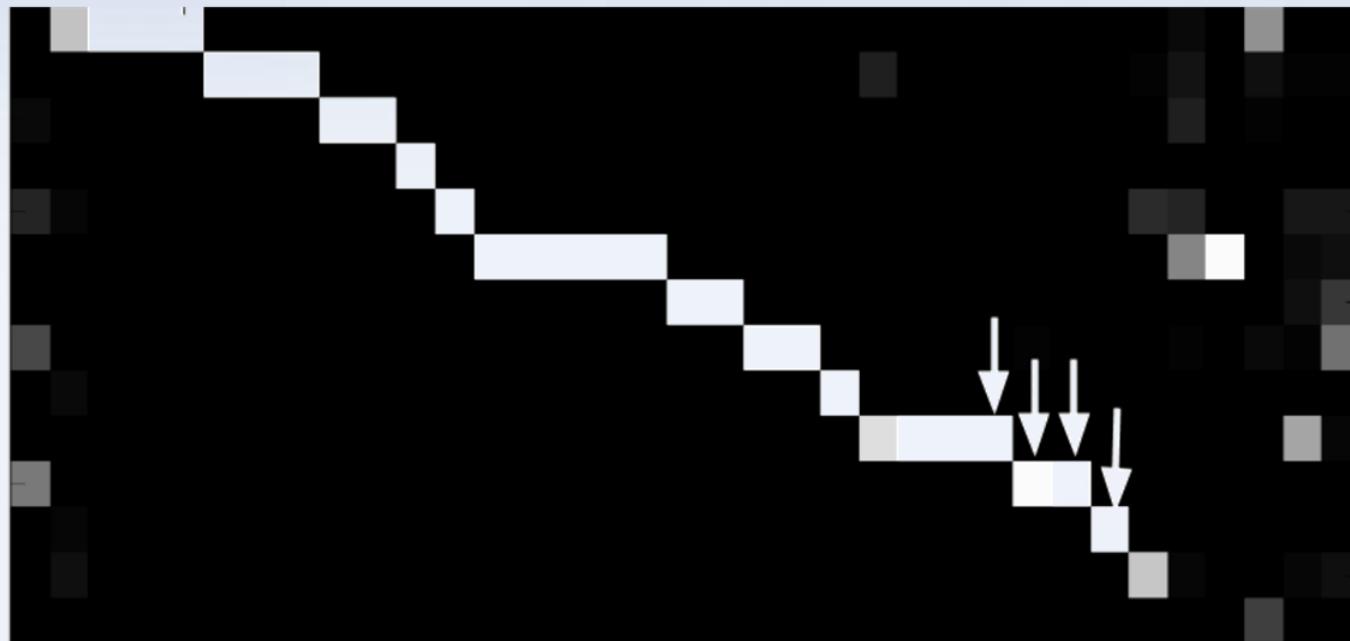
Behind the Scenes



Attention in NLP



Attention in NLP [2]



Attention in NLP [3]

by ent423 ,ent261 correspondent updated 9:49 pm et ,thu march 19, 2015 (ent261) a ent114 was killed in a parachute accident in ent45 ,ent85 ,near ent312 ,a ent119 official told ent261 on wednesday .he was identified thursday as special warfare operator 3rd class ent23 ,29 ,of ent187 , ent265 .`` ent23 distinguished himself consistently throughout his career .he was the epitome of the quiet professional in all facets of his life ,and he leaves an inspiring legacy of natural tenacity and focused

...

ent119 identifies deceased sailor as X ,who leaves behind a wife

by ent270 ,ent223 updated 9:35 am et ,mon march 2 ,2015 (ent223) ent63 went familial for fall at its fashion show in ent231 on sunday ,dedicating its collection to `` mamma '' with nary a pair of `` mom jeans " in sight .ent164 and ent21 , who are behind the ent196 brand ,sent models down the runway in decidedly feminine dresses and skirts adorned with roses ,lace and even embroidered doodles by the designers ' own nieces and nephews .many of the looks featured saccharine needlework phrases like `` i love you ,

...

X dedicated their fall fashion show to moms

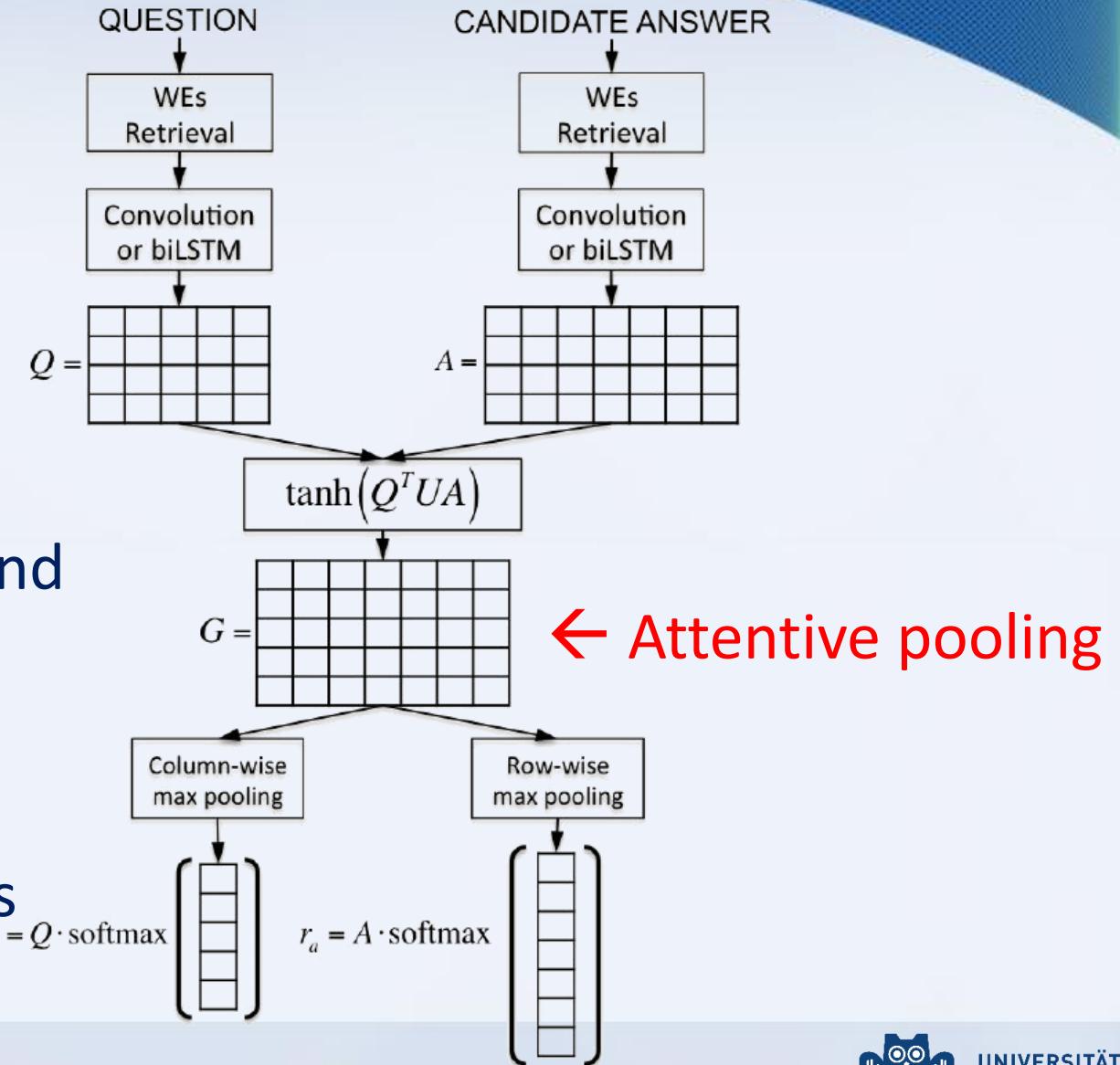
Putting it all together

Attentive Pooling Revisited

❖ Attention – pay attention to what's asked and you will find a better answer

❖ Pooling –

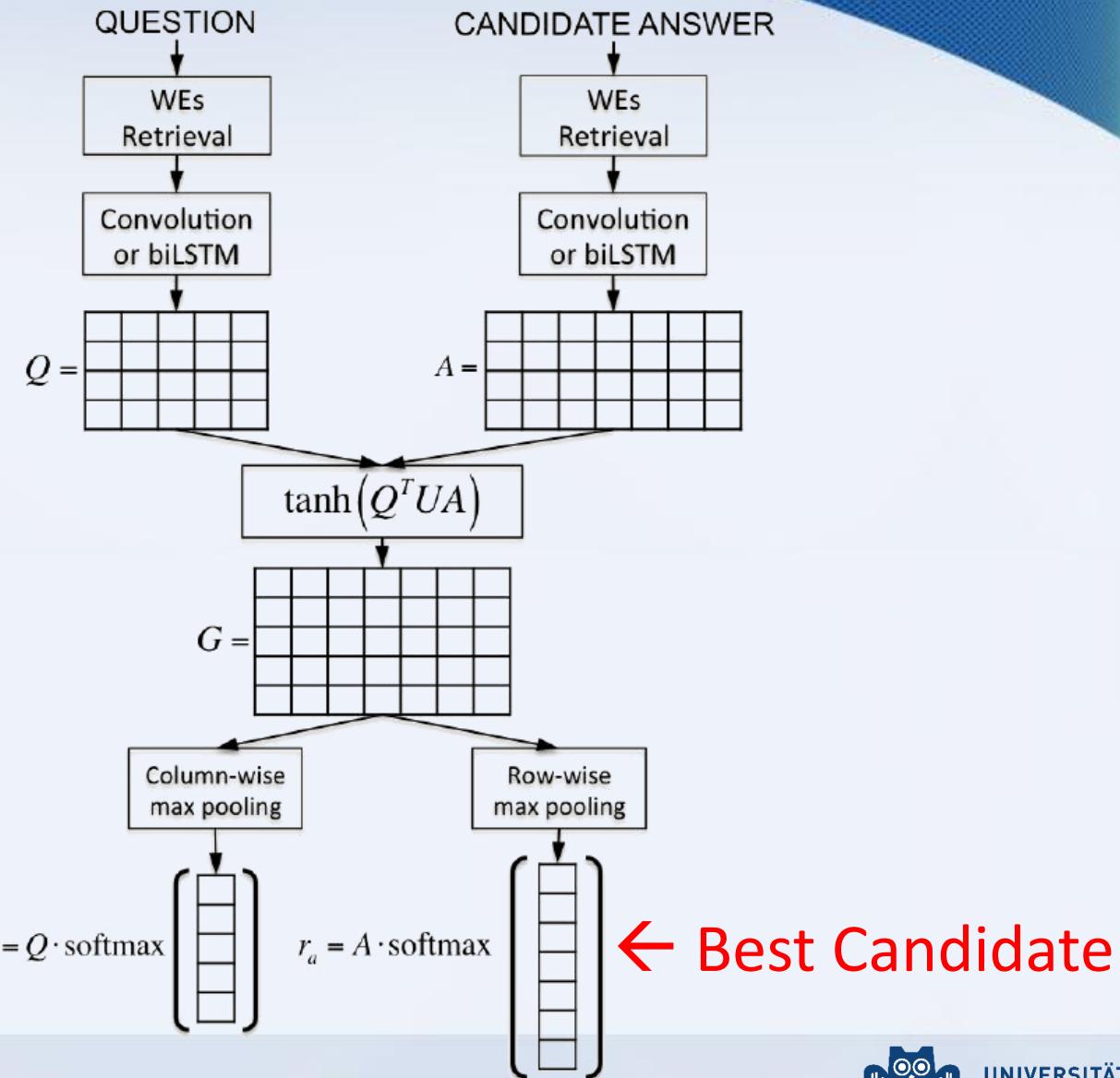
- Let the features from the question and from the answer interact with each other, and
- from the features you come across, always take the most influential ones



Scoring and Ranking

- ❖ Finally, take the part from the attentive pooling with the highest features

- ❖ No need for similarity measure between question and answer vectors



Experimental Setting

Datasets

DATASET	TRAIN	DEV	TEST	AVG. M	AVG. L	AVG. # CAND. ANS.	AVG. L/M
INSURANCEQA	12887	1000	1800x2	7	95	500	13.8
TREC-QA	1162	65	68	8	28	38	4.2
WIKIQA	873	126	243	6	25	9	5.0

NN architecture

Hyp.	Hyperpar. Name	AP-CNN	QA-CNN
d	WORD EMB. SIZE	100/300	100/300
c	CONV. FILTERS / HID. VEC. SIZE	400	4000
k	CONTEXT WINDOW SIZE	3/4	2
mbs	MINIBATCH SIZE	20	1
m	LOSS MARGIN	0.5	0.009
λ	INIT. LEARNING RATE	1.1	0.05

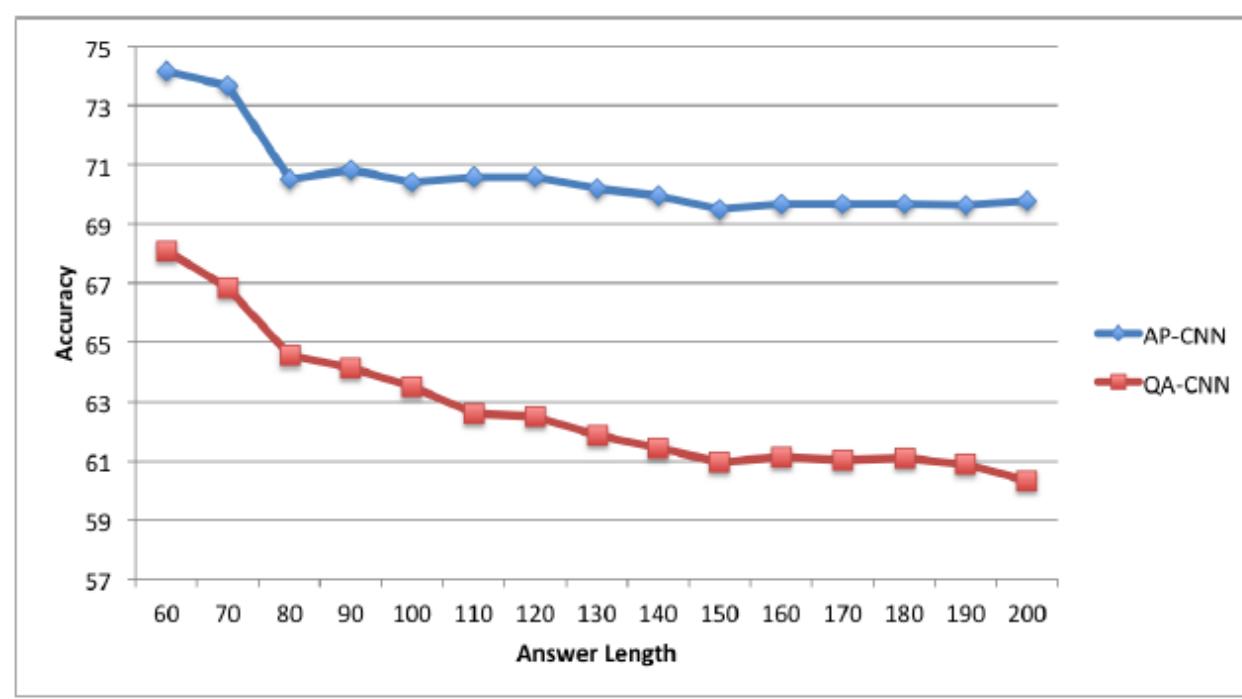
← word2vec

$$\leftarrow \lambda_t = \frac{\lambda}{t}$$



Results

- ❖ AP-CNN outperforms QA-CNN
- ❖ AP-CNN needs 10x less filters than QA-CNN



System	Dev	Test1	Test2
QA-CNN	61.6	60.2	56.1
QA-BiLSTM	66.6	66.6	63.7
AP-CNN	68.8	69.8	66.3
AP-BiLSTM	68.4	71.7	66.4

- ❖ AP-CNN stabilizes, i.e. more robust to long answers

Visualizations

QUESTION:

what be auto insurance lapse

ANSWER:

an auto insurance lapse mean under the term and condition of the policy coverage provide that policy be discontinue typically a policy lapse for non payment of premium coverage can be discontinue for other reason for example if an insured decide cancel the policy or conversley if the company decide to non-renew because the risk no longer meet the company guidlines as file with the state insurance department the state of CT and many other require liability coverage be in force as a condition of the registration of a car thus if a car insurance policy lapse and the state DMV therefore be notify the result can be a fine or suspension of the car registration or both likewise many car insurance company will not offer coverage for those who have not maintain continuous coverage in a case such as that , the car ower will end up be offer coverage usually in a non-standard company at a much high premium

QUESTION:

how much do a pool add to home insurance

ANSWER:

the primary concern of add a pool be the liability exposure if someone not in your household be hurt use the pool you may be hold responsible and/or sue if a judgement be bring against you it can mean 100's thousand in settlement if you live in a typical neighborhood and your yard / pool be fence and secure most insurance company will charge little or no additional dollar for the exposure of the pool if the yard / pool be not fence most company will either require a sign exclusion of coverage for injury arise out of the use of the pool or deny you coverage altogether there be exception to the fencing requirement if the home be in a rural area with no close neighbor



Summary

- ❖ Introduced Bi-directional attention mechanism for pairwise ranking
- ❖ Pair's components directly influence the representation of each other
- ❖ AP systems perform better than non-attentive systems in QA tasks
- ❖ AP-QA systems are also more robust to long answer-candidates



Thank you for your (max-pooled) Attention

Discussion

- ❖ Would a greater window size yield better/worse results? Why?
- ❖ What's the difference between attentive embedding and any other string matching technique, e.g. n-gram matching?
- ❖ What other NLP tasks can benefit from attention mechanisms?

