

# Convolutional Neural Networks

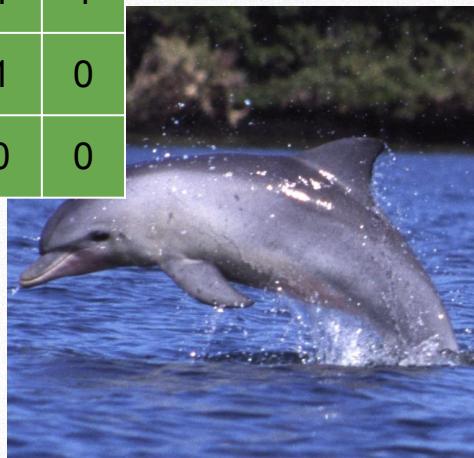
Felix Hill



# Convolutions

1	0	1
0	1	0
1	0	1

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



# Convolutions

<sup>1</sup> 1	<sup>0</sup> 1	<sup>1</sup> 1	0	0
<sup>0</sup> 0	<sup>1</sup> 1	<sup>0</sup> 1	1	0
<sup>1</sup> 0	<sup>0</sup> 0	<sup>1</sup> 1	1	1
0	0	1	1	0
0	1	1	0	0

4		

# Convolutions

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

4	3	

# Convolutions

1	1	1 <sup>1</sup>	0 <sup>0</sup>	0	1
0	1	1 <sup>0</sup>	1 <sup>1</sup>	0	0
0	0	1 <sup>1</sup>	1 <sup>0</sup>	1	1
0	0	1	1	0	
0	1	1	0	0	

4	3	4

# Convolutions

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

4	3	4
2		

# Convolutions

1	1	1	0	0
0	<sup>1</sup> 01	<sup>0</sup> 11	1 <sup>1</sup>	0
0	<sup>0</sup> 01	<sup>1</sup> 11	1 <sup>0</sup>	1
0	<sup>1</sup> 01	<sup>0</sup> 11	1 <sup>1</sup>	0
0	1	1	0	0

4	3	4
2	4	

# Convolutions

1	1	1	0	0
0	1	<sup>1</sup> 1	<sup>0</sup> 1	<sup>1</sup> 0
0	0	<sup>0</sup> 1	<sup>1</sup> 1	<sup>0</sup> 1
0	0	<sup>1</sup> 1	<sup>0</sup> 1	<sup>1</sup> 0
0	1	1	0	0

4	3	4
2	4	3

# Convolutions

1	1	1	0	0
0	1	1	1	0
<sup>1</sup> 0	<sup>0</sup> 0	1 <sup>1</sup>	1	1
<sup>0</sup> 0	<sup>1</sup> 0	1 <sup>0</sup>	1	0
<sup>1</sup> 0	<sup>0</sup> 1	1 <sup>1</sup>	0	0

4	3	4
2	4	3
2		

# Convolutions

1	1	1	0	0
0	1	1	1	0
0	0 <sup>1</sup>	1 <sup>0</sup>	1 <sup>1</sup>	1
0	0 <sup>0</sup>	1 <sup>1</sup>	1 <sup>0</sup>	0
0	1 <sup>1</sup>	1 <sup>0</sup>	0 <sup>1</sup>	0

4	3	4
2	4	3
2	3	

# Convolutions

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

4	3	4
2	4	3
2	3	5

# Convolutions - filter # 2

0	1	0
1	1	1
0	1	0

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

4	3	4
2	4	3
2	3	5

# Convolutions - filter # 2

<sup>0</sup> 1	<sup>1</sup> 1	<sup>0</sup> 1	0	0
<sup>1</sup> 0	<sup>1</sup> 1	<sup>1</sup> 1	1	0
<sup>0</sup> 0	<sup>1</sup> 0	<sup>0</sup> 1	1	1
0	0	1	1	0
0	1	1	0	0

4	3	4
2	4	3
2	3	5

3		

# Convolutions - filter # 2

1	1	1	0	0
0	1	1	1	0
0	0	<sup>0</sup> 1	<sup>1</sup> 1	<sup>0</sup> 1
0	0	<sup>1</sup> 1	<sup>1</sup> 1	<sup>1</sup> 0
0	1	<sup>0</sup> 1	<sup>1</sup> 0	<sup>0</sup> 0

4	3	4
2	4	3
2	3	5

3	5	3
2	4	5
2	4	3

# What do filters 'do'?

0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1

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

4	3	4
2	4	3
2	3	5

3	5	3
2	4	5
2	4	3

# What do filters 'do'?

0	1	0
1	1	1
0	1	0

1	0	1
0	1	0
1	0	1

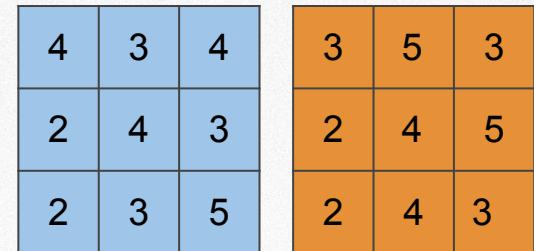
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

4	3	4
2	4	3
2	3	5

3	5	3
2	4	5
2	4	3

# Convolutions - filter # 3

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



10	1	1
1	4	2
3	2	1

# Convolutions - filter # 4

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

4	3	4
2	4	3
2	3	5

3	5	3
2	4	5
2	4	3

10	1	1
1	4	2
3	2	1

1	5	0
2	4	2
2	11	3

# Subsampling / Max Pooling

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

4	

4	3	4
2	4	3
2	3	5

3	5	3
2	4	5
2	4	3

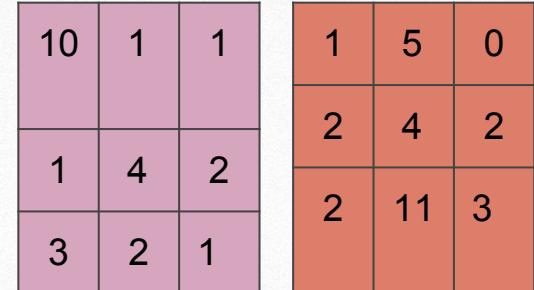
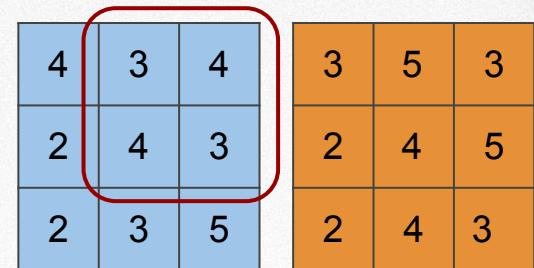
10	1	1
1	4	2
3	2	1

1	5	0
2	4	2
2	11	3

# Subsampling / Max Pooling

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

4	4



# Subsampling / Max Pooling

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

4	4
4	

4	3	4
2	4	3
2	3	5

3	5	3
2	4	5
2	4	3

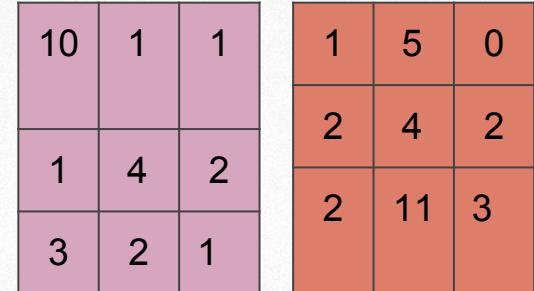
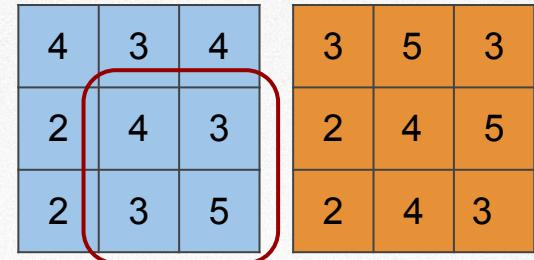
10	1	1
1	4	2
3	2	1

1	5	0
2	4	2
2	11	3

# Subsampling / Max Pooling

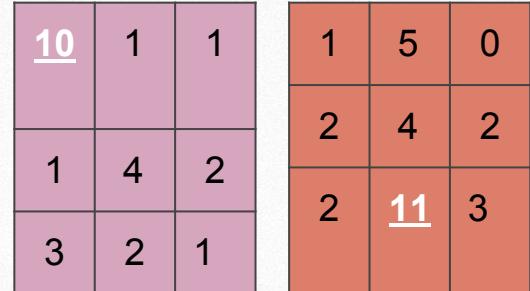
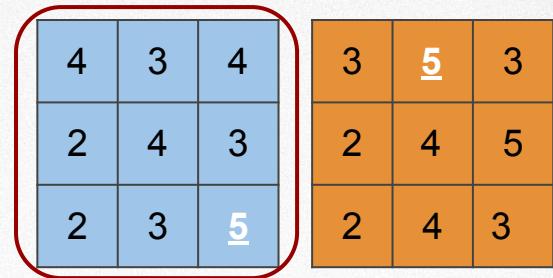
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

4	4
4	5



# (Total) max pooling

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



# Max pooling

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

5	5	10	11
---	---	----	----

4	3	4
2	4	3
2	3	5
2	4	3

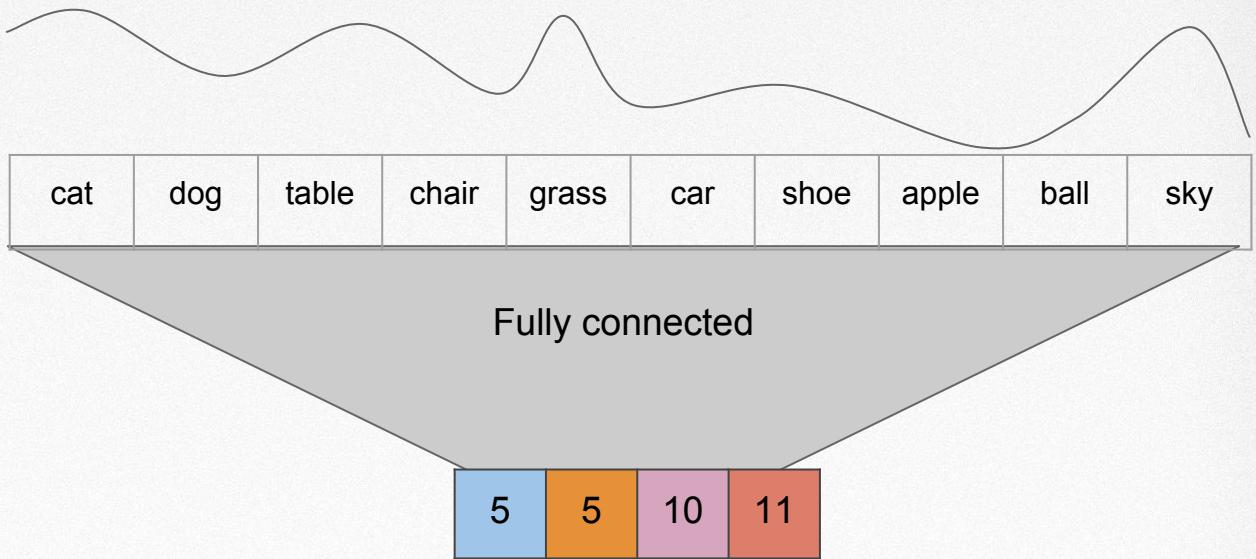
3	5	3
2	4	5
2	4	3
2	4	3

10	1	1
1	4	2
3	2	1

1	5	0
2	4	2
2	11	3

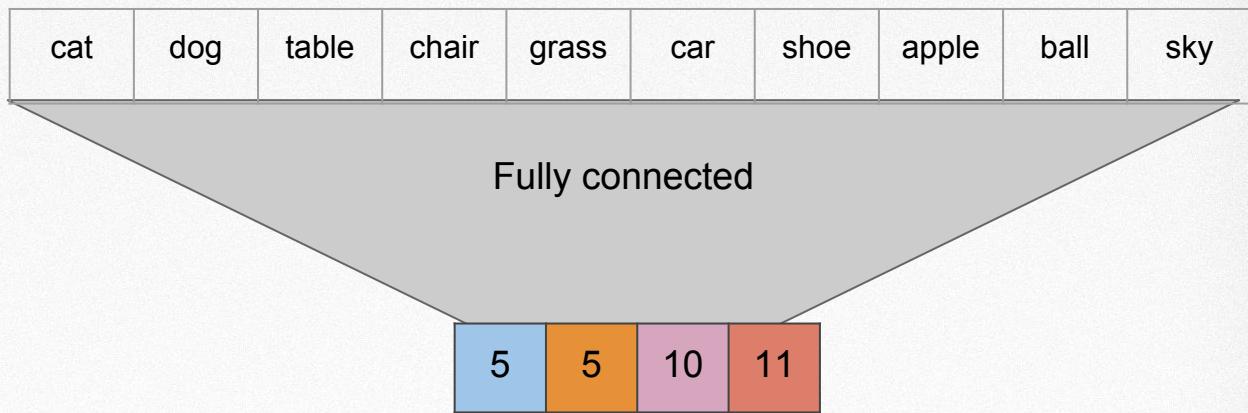
# Fully-connected final layer

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



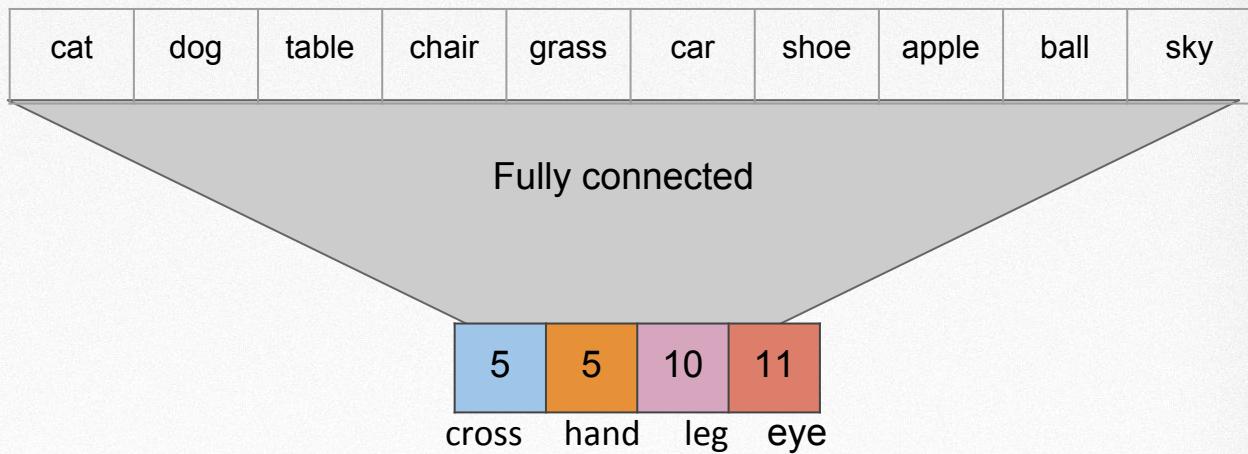
# Fully-connected final layer

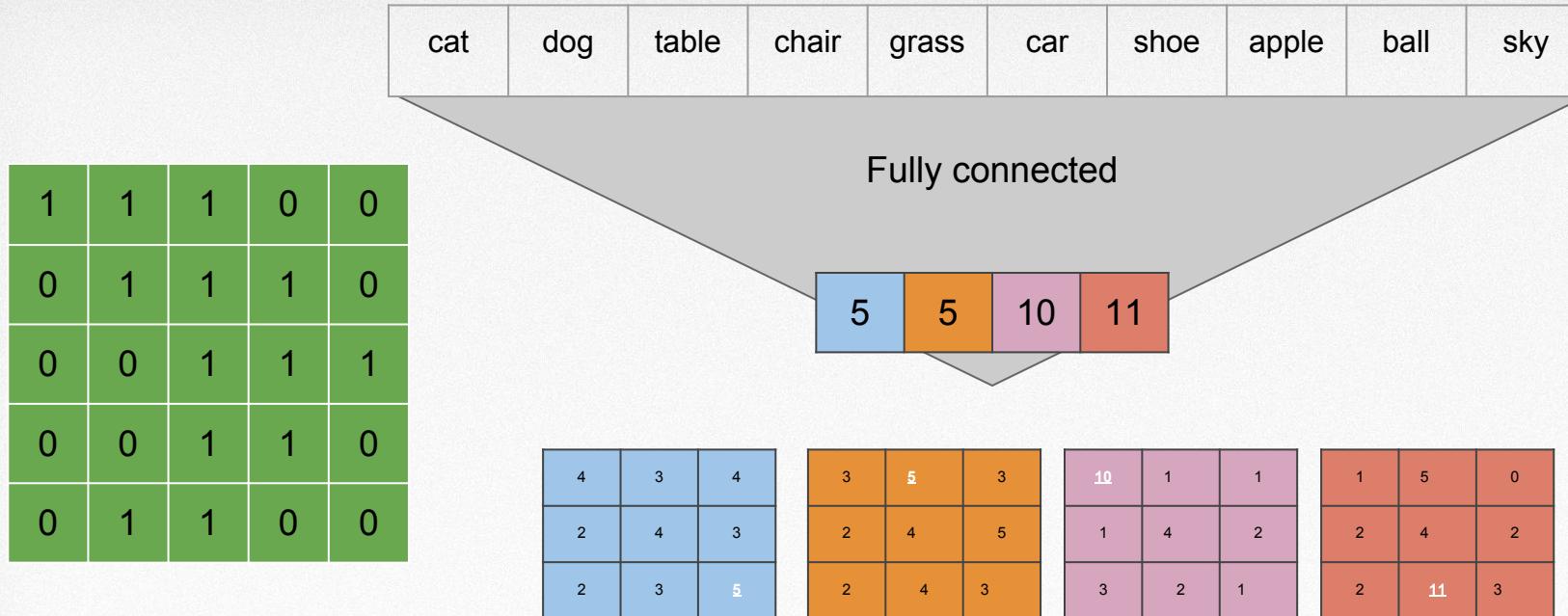
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0



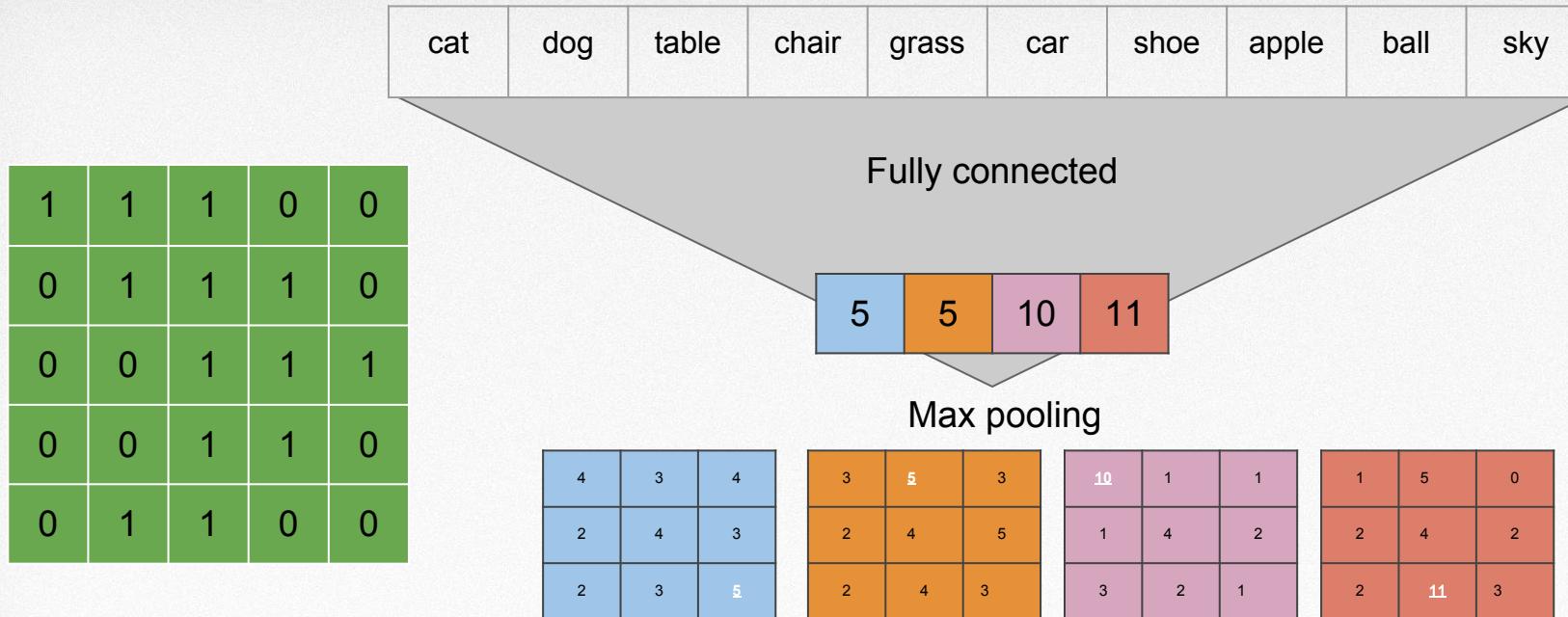
# Fully-connected final layer

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



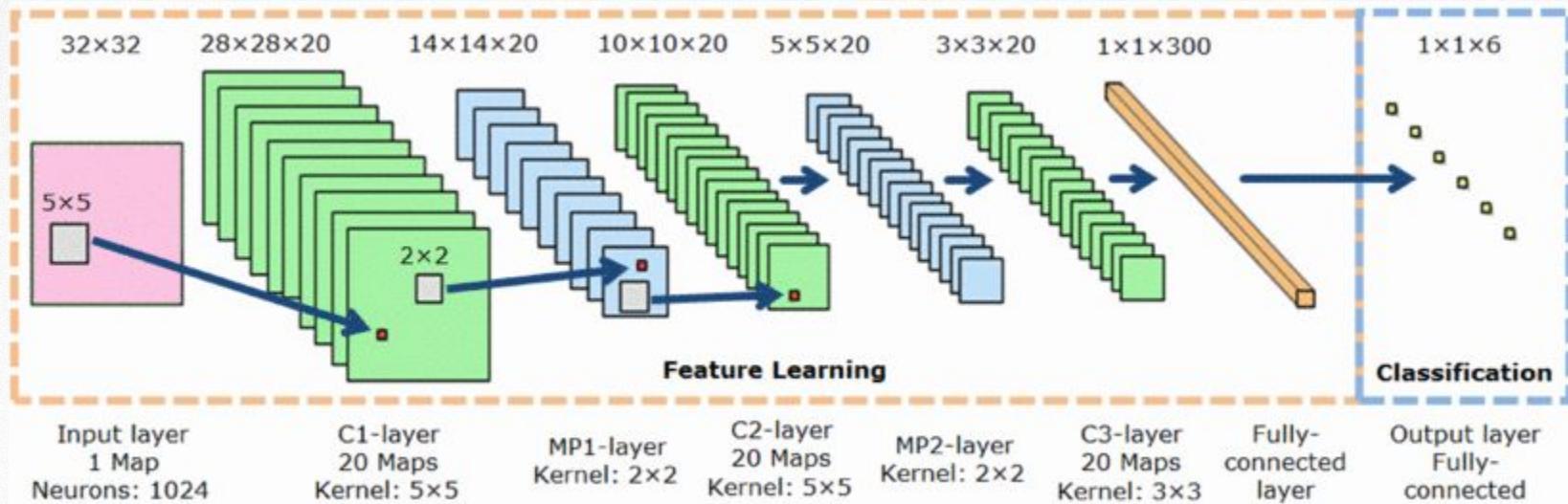


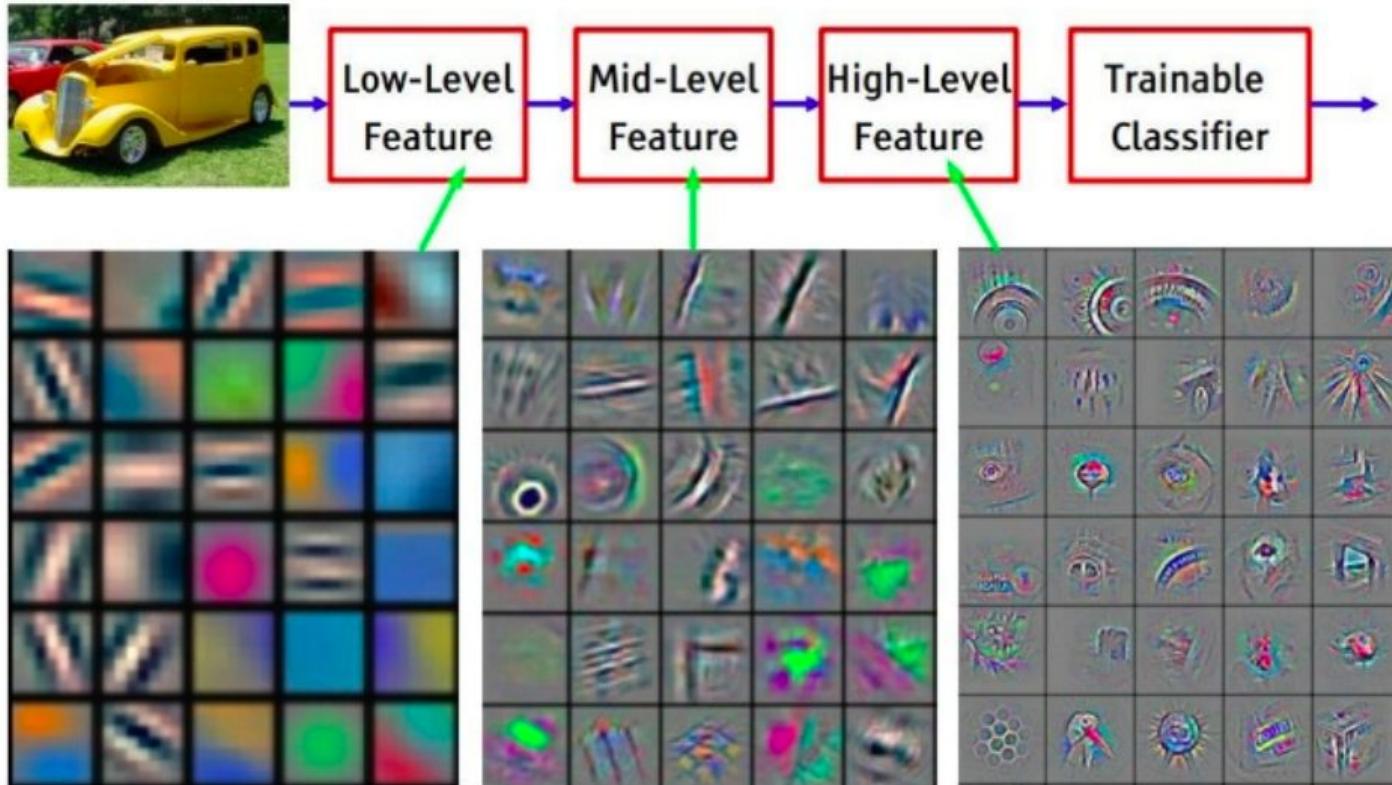
How many weights does this network have??



# How many weights does this network have??

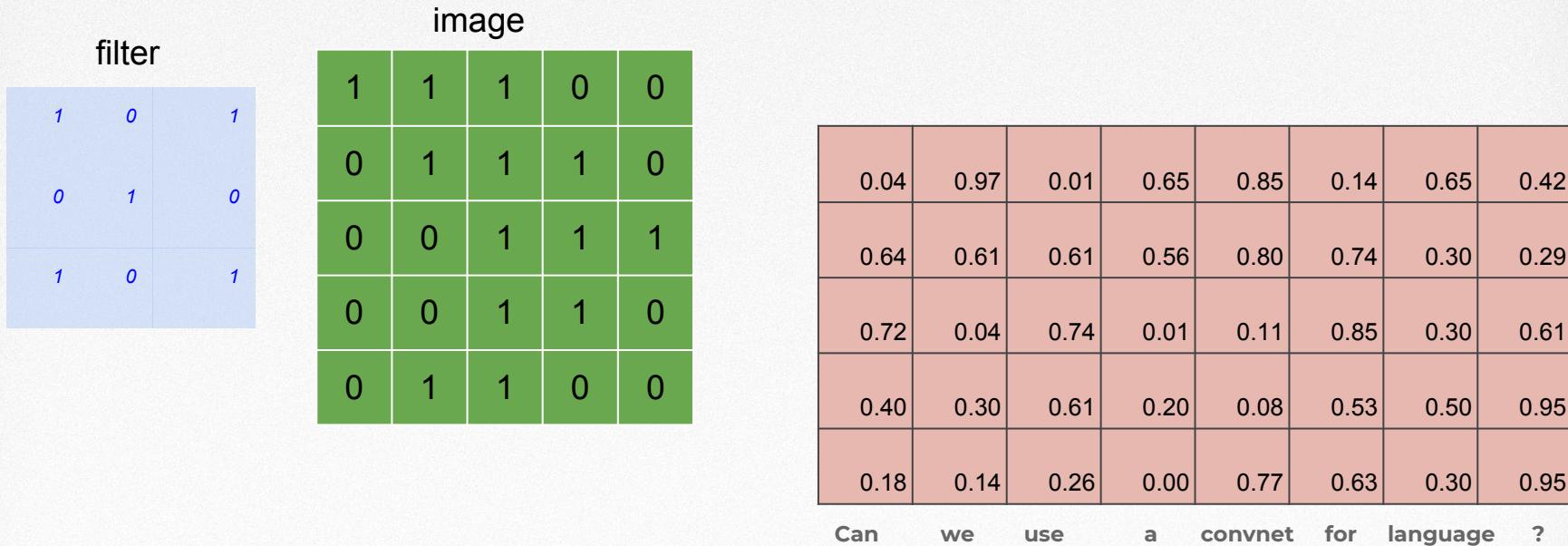
# For natural image, Conv Nets are 8+ layers deep





From Zeiler & Fergus, 2013

# Can we use a conv net for language?



# Can we use a conv net for language?

filter

0.97	0.86	0.70
0.35	0.28	0.26
0.15	0.28	0.18
0.84	0.94	0.09
0.61	0.93	0.21

See e.g Collobert & Weston 2007, Kalchbrenner & Blunsom 2014

0.04	0.97	0.01	0.65	0.85	0.14	0.65	0.42
0.64	0.61	0.61	0.56	0.80	0.74	0.30	0.29
0.72	0.04	0.74	0.01	0.11	0.85	0.30	0.61
0.40	0.30	0.61	0.20	0.08	0.53	0.50	0.95
0.18	0.14	0.26	0.00	0.77	0.63	0.30	0.95

Can      we      use      a      convnet      for      language      ?

# Can we use a conv net for language?

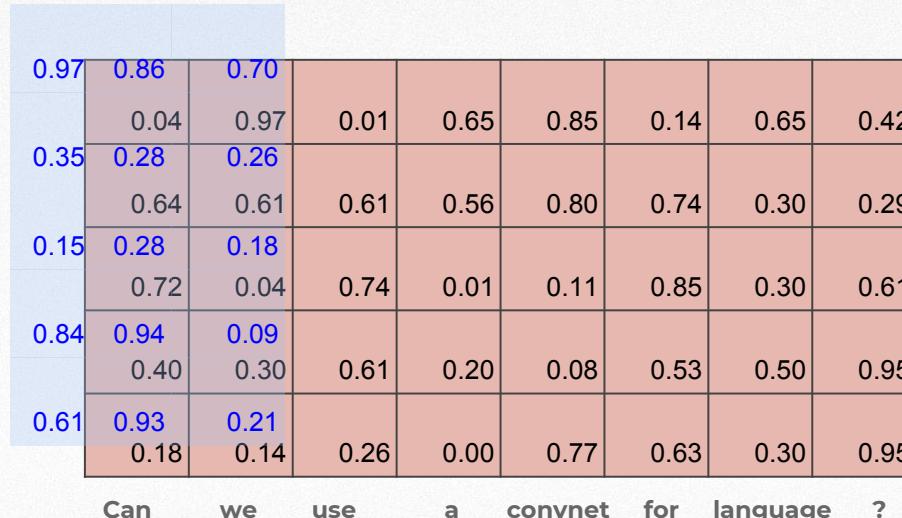
3.41

		0.97 0.86	0.70	0.04	0.97	0.01	0.65	0.85	0.14	0.65	0.42
0.35 0.28		0.26	0.64	0.61	0.61	0.56	0.80	0.74	0.30	0.29	
0.15 0.28		0.18	0.72	0.04	0.74	0.01	0.11	0.85	0.30	0.61	
0.84 0.94		0.09	0.40	0.30	0.61	0.20	0.08	0.53	0.50	0.95	
0.61 0.93		0.21	0.18	0.14	0.26	0.00	0.77	0.63	0.30	0.95	

Can    we    use    a    convnet    for    language    ?

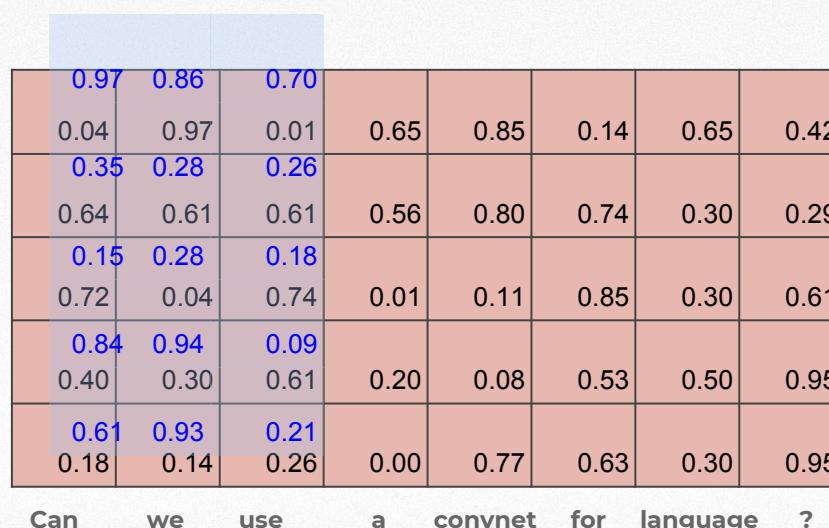
# Can we use a conv net for language?

3.41	2.71								
------	------	--	--	--	--	--	--	--	--



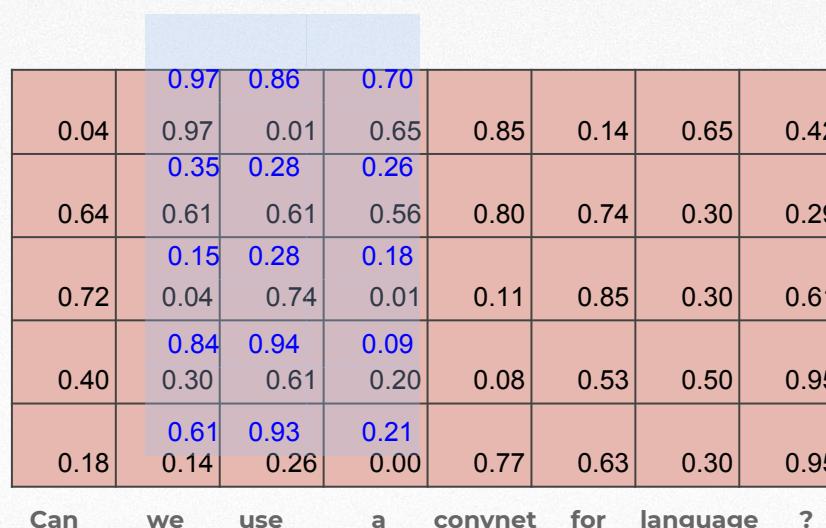
# Can we use a conv net for language?

3.41	2.71	4.32							
------	------	------	--	--	--	--	--	--	--



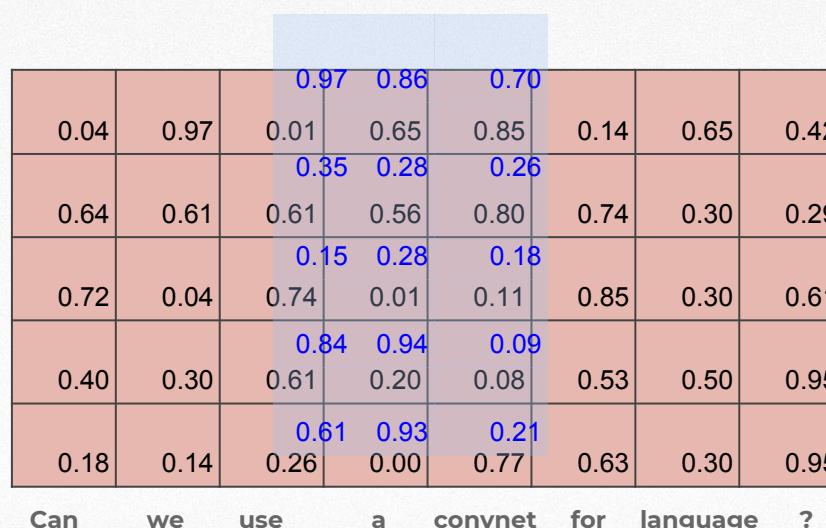
# Can we use a conv net for language?

3.41	2.71	4.32	3.21						
------	------	------	------	--	--	--	--	--	--



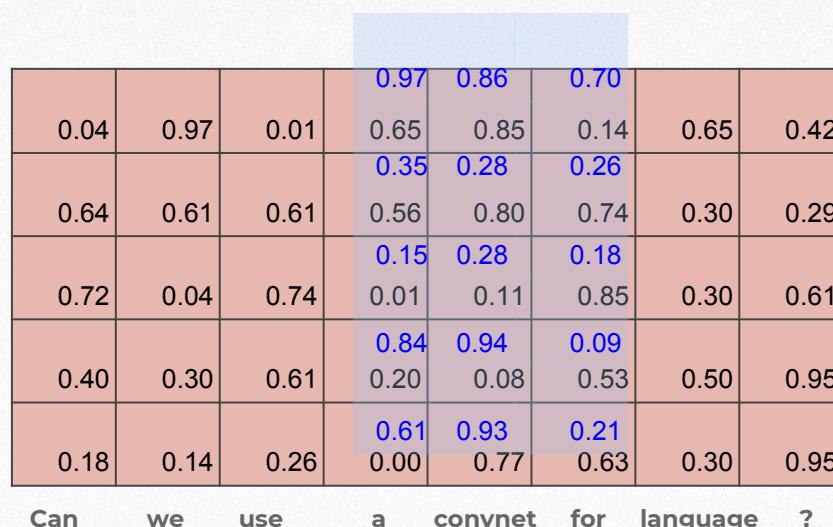
# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81					
------	------	------	------	------	--	--	--	--	--



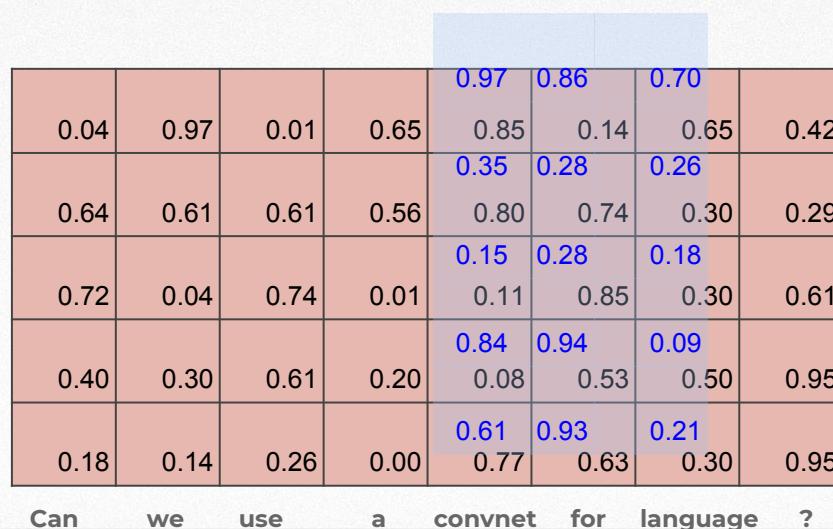
# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81	2.95				
------	------	------	------	------	------	--	--	--	--



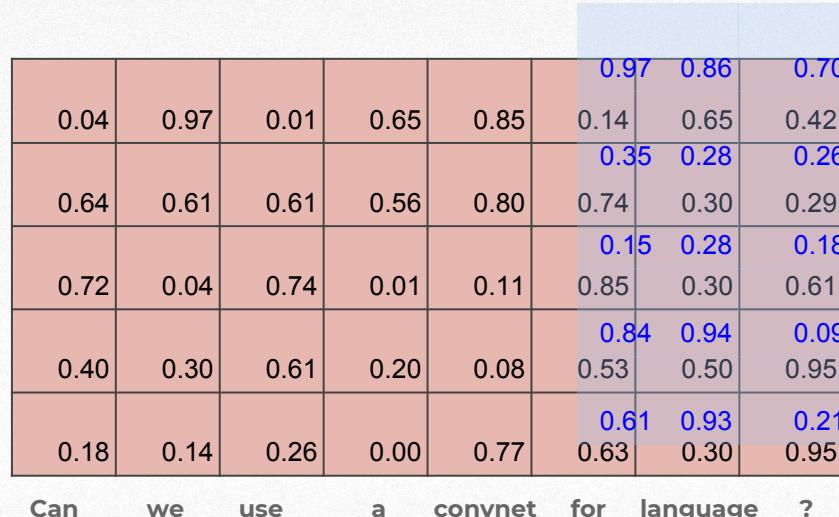
# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81	2.95	5.43			
------	------	------	------	------	------	------	--	--	--



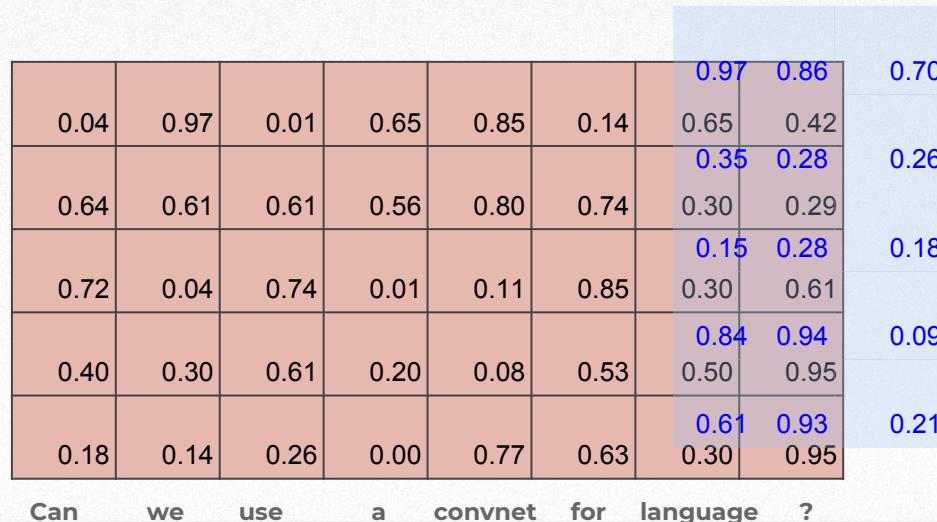
# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34		
------	------	------	------	------	------	------	------	--	--



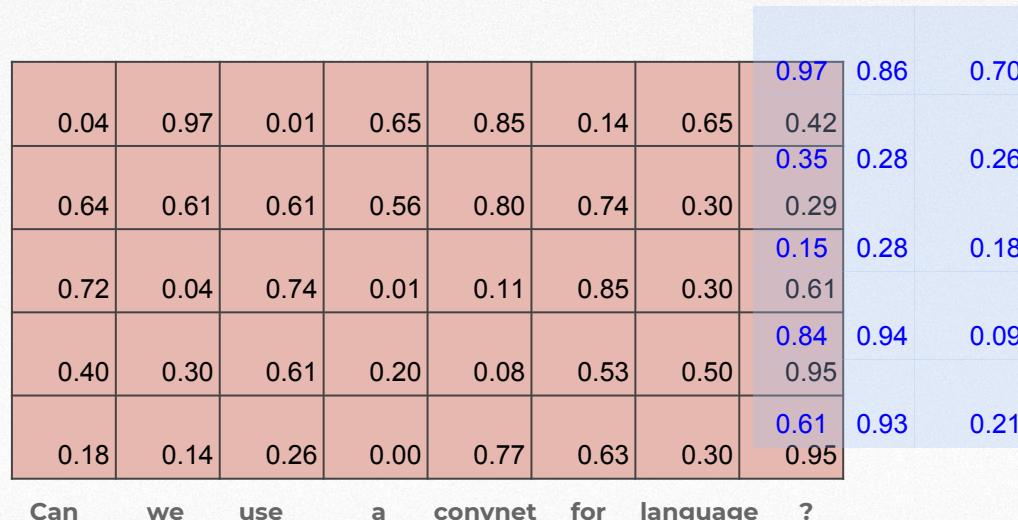
# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34	2.22	
------	------	------	------	------	------	------	------	------	--



# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34	2.22	1.96
------	------	------	------	------	------	------	------	------	------



# Can we use a conv net for language?

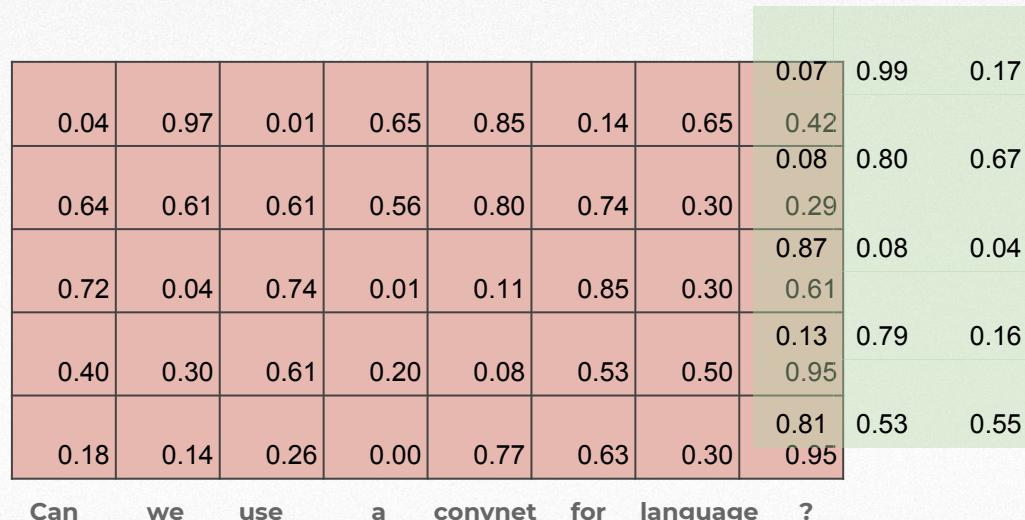
3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34	2.22	1.96
0.87									

0.07	0.99	0.17							
0.08	0.80	0.04	0.97	0.01	0.65	0.85	0.14	0.65	0.42
0.67									
0.64		0.61	0.61	0.56	0.80	0.74	0.30	0.29	
0.87	0.08	0.04							
0.72		0.04	0.74	0.01	0.11	0.85	0.30	0.61	
0.13	0.79	0.16							
0.40		0.30	0.61	0.20	0.08	0.53	0.50	0.95	
0.81	0.53	0.55							
0.18		0.14	0.26	0.00	0.77	0.63	0.30	0.95	

Can    we    use    a    convnet    for    language    ?

# Can we use a conv net for language?

3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34	2.22	1.96
0.87	0.28	0.64	4.30	3.66	2.71	4.90	2.55	0.30	0.80



# Five 'independent' analyses of the sentence

3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34	2.22	1.96
0.87	0.28	0.64	4.30	3.66	2.71	4.90	2.55	0.30	0.80
4.51	3.84	1.63	1.71	1.67	3.51	4.69	4.01	3.55	4.68
0.68	2.43	4.51	4.30	1.69	0.26	3.52	1.67	3.27	2.96
2.68	2.43	4.51	0.30	3.69	0.26	3.52	2.67	4.27	2.96

# Five 'independent' analyses of the sentence.... start over again...?

3.41	2.71	4.32	3.21	2.81	2.95	5.43	3.34	2.22	1.96
0.87	0.28	0.64	4.30	3.66	2.71	4.90	2.55	0.30	0.80
4.51	3.84	1.63	1.71	1.67	3.51	4.69	4.01	3.55	4.68
0.68	2.43	4.51	4.30	1.69	0.26	3.52	1.67	3.27	2.96
2.68	2.43	4.51	0.30	3.69	0.26	3.52	2.67	4.27	2.96

# Or, max-pooling over time



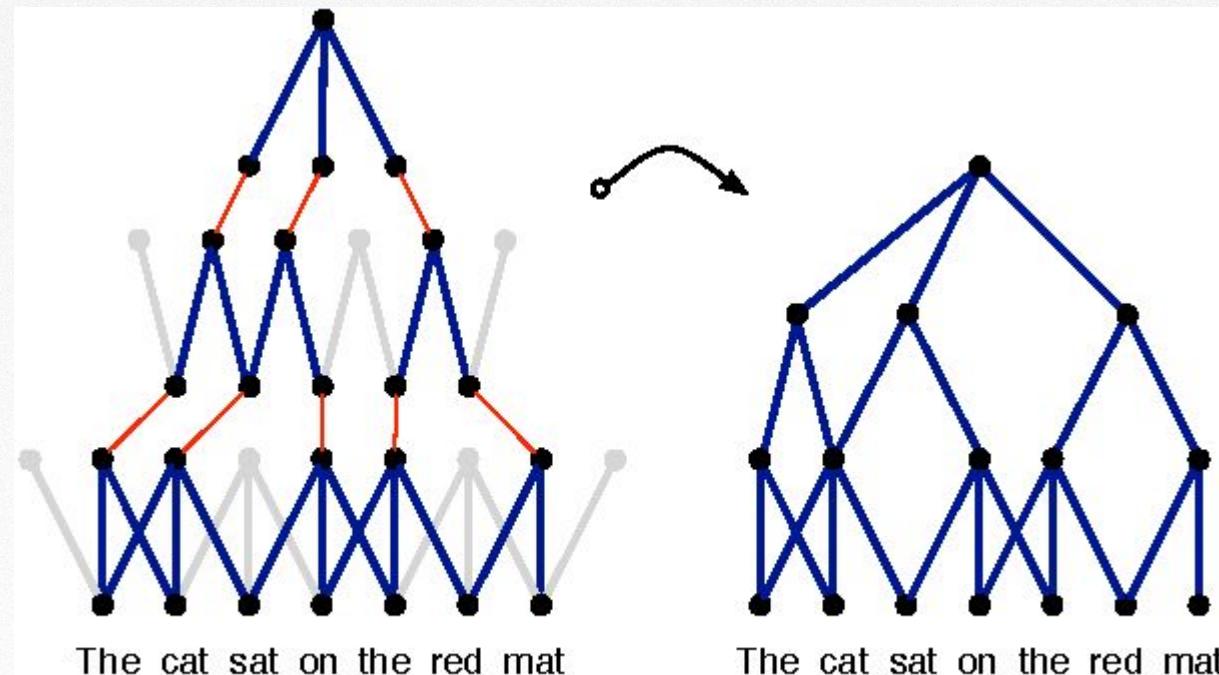
3.41	2.71	<b>4.32</b>	3.21	2.81	2.95	5.43	3.34	2.22	1.96
0.87	0.28	0.64	<b>4.30</b>	3.66	2.71	4.90	2.55	0.30	0.80
<b>4.51</b>	3.84	1.63	1.71	1.67	3.51	4.69	4.01	3.55	4.68
0.68	2.43	<b>4.51</b>	4.30	1.69	0.26	3.52	1.67	3.27	2.96
2.68	2.43	3.51	0.30	<b>3.69</b>	0.26	3.52	2.67	4.27	2.96

Fill in the gaps...



3.41	2.71	<b>4.32</b>	3.21	2.81	2.95	5.43	3.34	2.22	1.96	
0.87	0.28	0.64	<b>4.30</b>	3.66	2.71	4.90	2.55	0.30	0.80	
<b>4.51</b>	3.84	1.63	1.71	1.67	3.51	4.69	4.01	3.55	4.68	
0.68	2.43	<b>4.51</b>	4.30	1.69	0.26	3.52	1.67	3.27	2.96	
2.68	2.43	3.51	0.30	<b>3.69</b>	0.26	3.52	2.67	4.27	2.96	

# Max Pooling gives us a 'syntactic' analysis



From Kalchbrenner & Blunsom, 2014

# Analyse the first-layer feature-maps

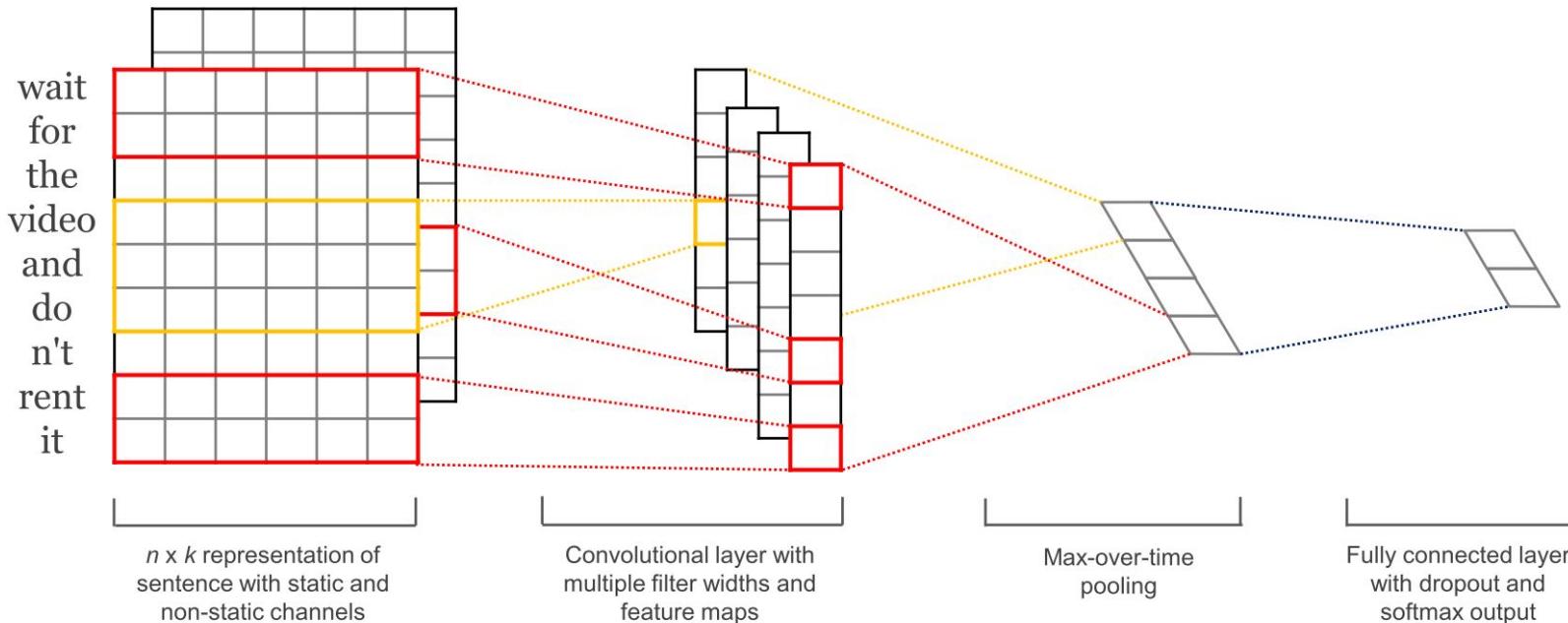
	1.00	0.59	0.85	0.10	0.70	1.00	0.56		
		0.04	0.97	0.01	0.65	0.85	0.14	0.65	0.42
	0.07	0.18	0.11	0.18	0.05	0.30	0.73		
		0.64	0.61	0.61	0.56	0.80	0.74	0.30	0.29
	0.93	0.90	0.53	0.25	0.66	0.58	0.03		
		0.72	0.04	0.74	0.01	0.11	0.85	0.30	0.61
	0.27	0.64	0.51	0.71	0.98	0.75	0.74		
		0.40	0.30	0.61	0.20	0.08	0.53	0.50	0.95
	0.58	0.28	0.26	0.58	0.78	0.34	0.82		
		0.18	0.14	0.26	0.00	0.77	0.63	0.30	0.95

Can we use a convnet for language ?

POSITIVE						"NOT"					
lovely	comedic	moments and several	fine	performances	n't	have	any	huge	laughs	in	its
good	script	, good dialogue	, funny		no	movement,	,	no	,	not	much
sustains	throughout	is daring ,	inventive and		n't	stop	me	from	enjoying	much	of
well	written	, nicely acted	and beautifully		not	that	kung	pow	is	n't	funny
remarkably	solid	and subtly satirical	tour de		not	a	moment	that is	not		false
NEGATIVE						'TOO'					
,	nonexistent	plot	and pretentious	visual style	,	too	dull	and	pretentious	to	be
it	fails	the	most basic	test as	either	too	serious	or	too	lighthearted,	
so	stupid	,	so ill	conceived,	too	slow	,	too	long	and	too
,	too	dull	and pretentious	to be	feels	too	formulaic	and	too	familiar	to
hood	rats	butt	their ugly	heads in	is	too	predictable	and	too	self	conscious

From Kalchbrenner & Blunsom, 2014

# Kim (2014) - pre-trained word embs + convnet



# Kim (2014) - pre-trained word embs + convnet

Model	MR	SST-1	SST-2	Subj	TREC	CR	MPQA
1: Learned embeddings	76.1	45.0	82.7	89.6	91.2	79.8	83.4
2: Google w2v embeddings	81.0	45.5	86.8	93.0	92.8	84.7	<b>89.6</b>
3: Google w2v initialise	<b>81.5</b>	<b>48.0</b>	87.2	93.4	93.6	84.3	89.5
1 and 2	81.1	47.4	<b>88.1</b>	93.2	92.2	<b>85.0</b>	89.4

# Kim (2014) - pre-trained word embs + convnet

Model	MR	SST-1	SST-2	Subj	TREC	CR	MPQA
1: Learned embeddings	76.1	45.0	82.7	89.6	91.2	79.8	83.4
2: Google w2v embeddings	81.0	45.5	86.8	93.0	92.8	84.7	<b>89.6</b>
3: Google w2v initialise	<b>81.5</b>	<b>48.0</b>	87.2	93.4	93.6	84.3	89.5
1 and 2	81.1	47.4	<b>88.1</b>	93.2	92.2	<b>85.0</b>	89.4

# Summary

- Convolutions and convnets have applications in language processing
- They can model hierarchical processing ('syntax'?)

## Food for thought

- They are *faster* to compute than RNNs (why?)
- They are most commonly used in *character-level models* (why?, how?)

# References

- Visualising and understanding convolutional networks (Zeiler & Fergus, 2013)
- A convolutional neural network for modelling sentences (Kalchbrenner & Blunsom, 2014)
- Convolutional Neural Networks for Sentence Classification (Kim, 2014)