Lecture 5
Common neural networks for different clinical applications

Serena Yeung

BIODS 388

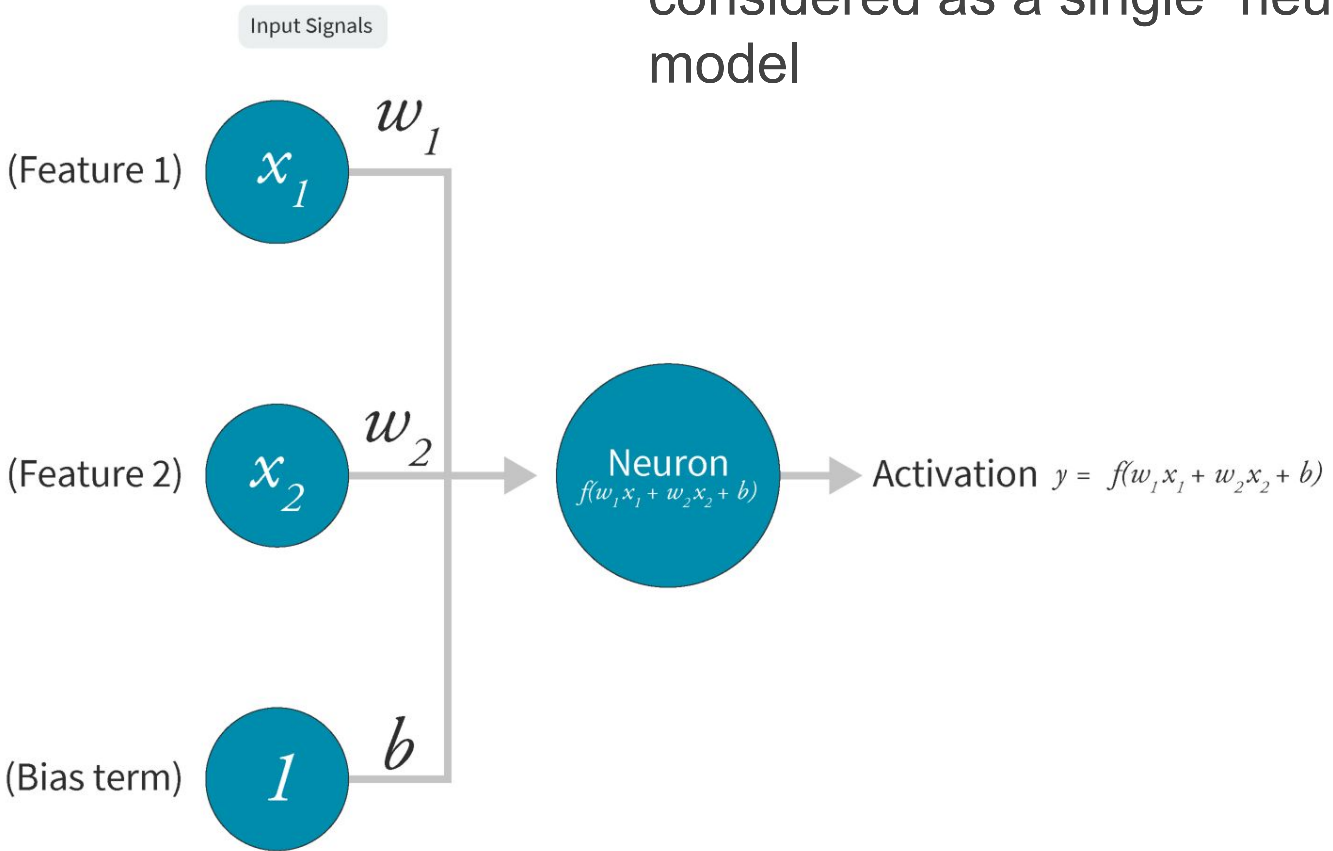
Last time...

Deep learning: Machine learning models based on "deep" neural networks comprising millions (sometimes billions) of parameters organized into hierarchical layers.

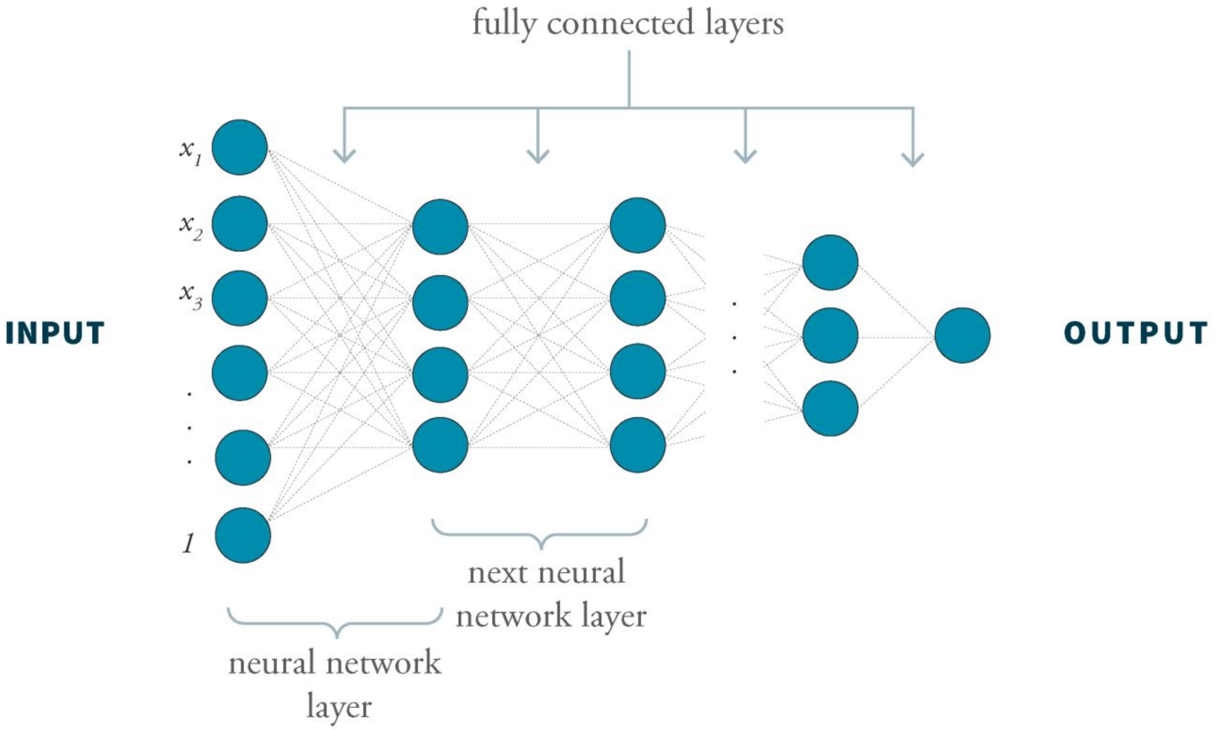
Features are multiplied and added together repeatedly, with the outputs from one layer of parameters being fed into the next layer -- before a prediction is made.

Last time...

The logistic regression with sigmoid that we just saw can be considered as a single "neuron" model

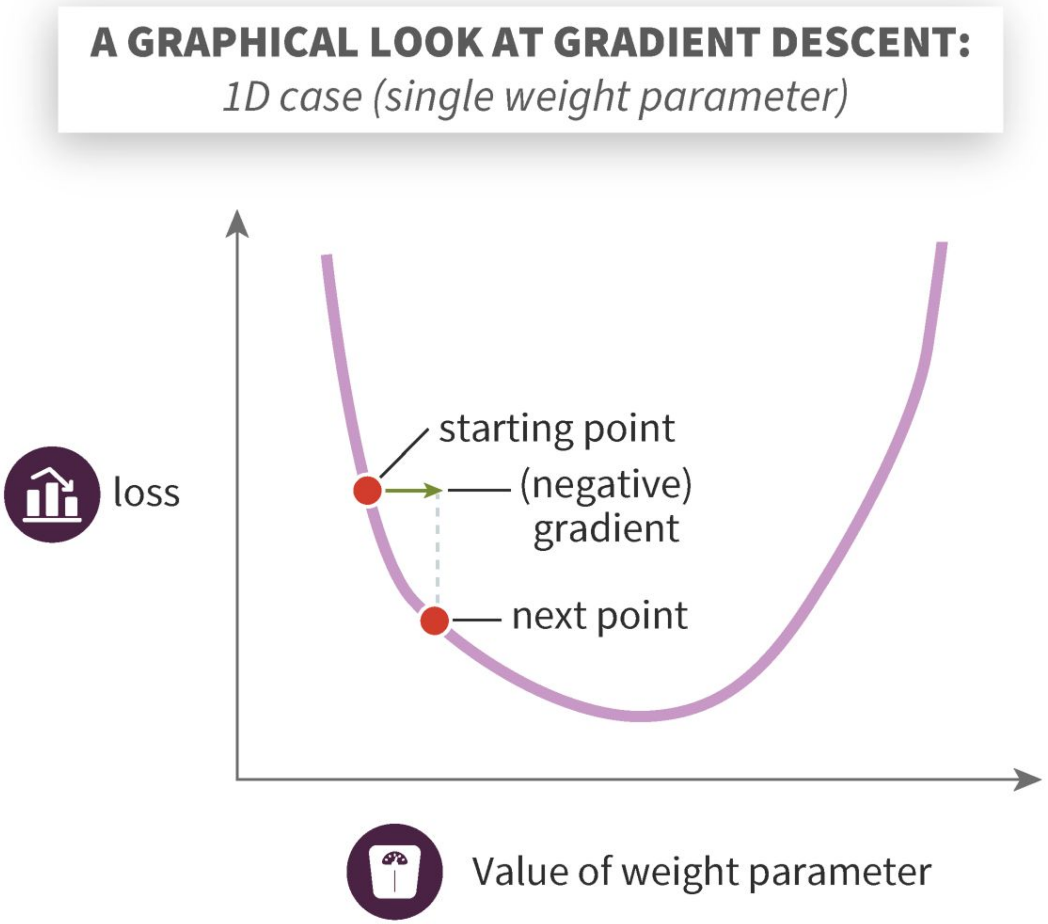


Last time...



Fully connected layer: all neurons in the layer takes as input the full input to the layer (also called dense layer or linear layer)

Last time...


A GRAPHICAL LOOK AT GRADIENT DESCENT:
1D case (single weight parameter)
starting point
(negative) gradient
next point
loss
Value of weight parameter

Last time...

UNSTRUCTURED DATA

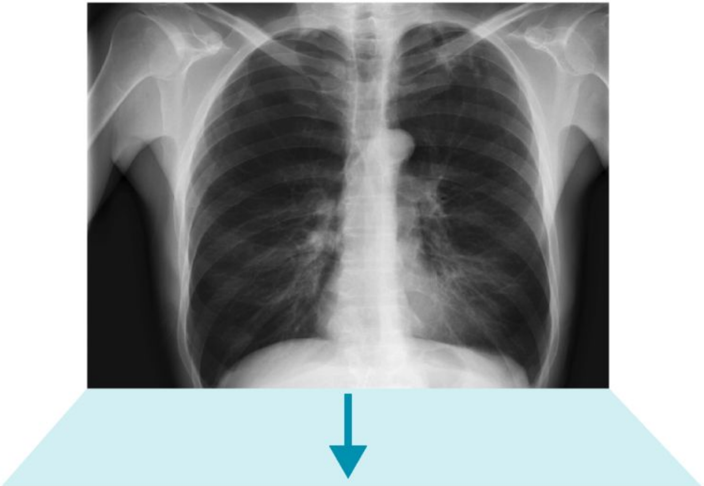
"...PMH of **metastatic breast cancer, R lung malignant** effusion, and **R lung empyema** who presents with increased drainage from **R lung pleurx** tract..."

STRUCTURED DATA

Last time...

GRAYSCALE IMAGES

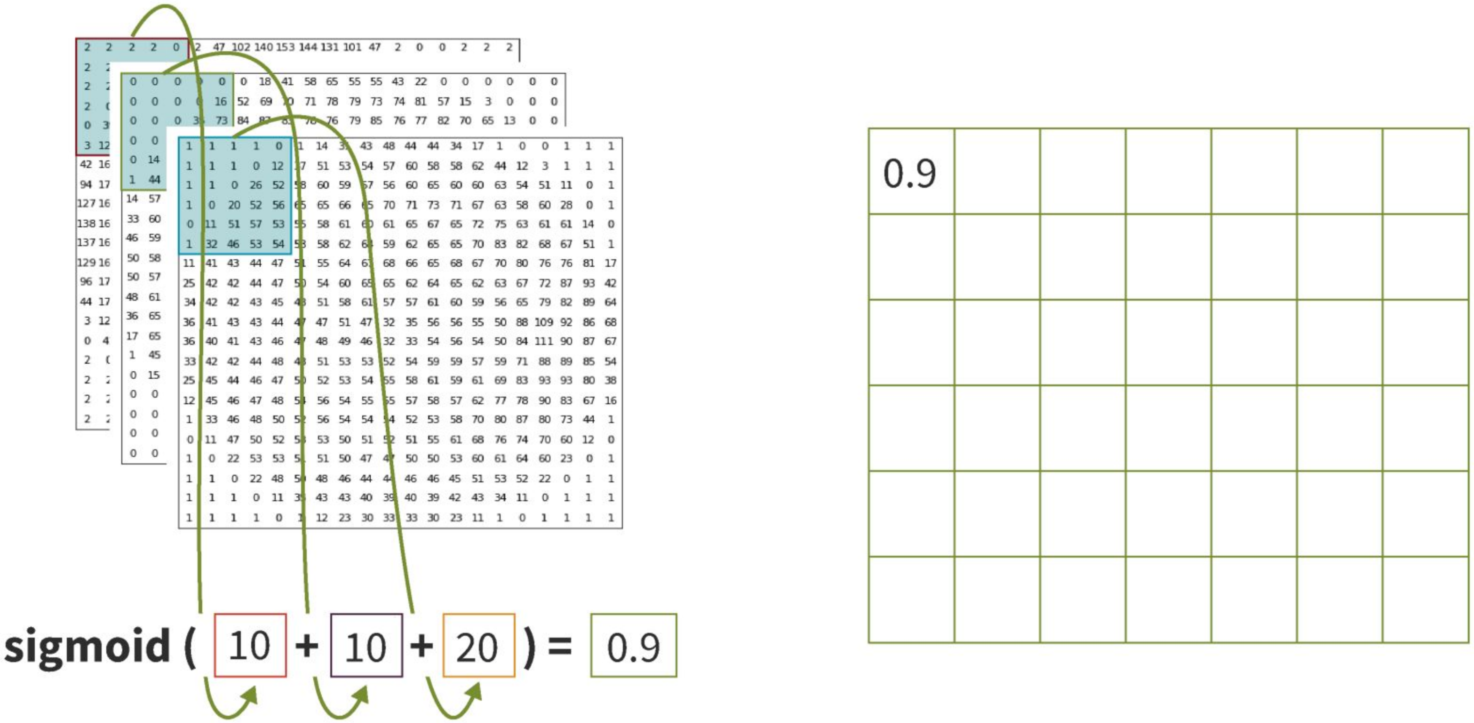
- Pixels → number grid
- Pixel brightness → grid value



```
[[  2  2  1  37   1  10  66  60  77  94  78  69  64  23  12  45  28  45 ]
 [ 58  1  9  13  17  29  56  72  65  64  59  58  39  18  15  12   7   1 ]
 [ 71 49 53  38  30  41  73  73  80  71  69  69  72  45  45  49  36  59 ]
 [ 88 60 73  50  59  59  54  51  71  81  69  50  54  75  56  61  80  67 ]
 [ 94 91 86  59  65  57  57  52  64  88  66  56  55  54  70  64 109 114 ]
 [ 94 95 84  74  70  41  48  55  74  85  84  60  50  46  70  82  92 122 ]
 [ 85 85 95  83  54  37  59  60  84  97  82  50  38  44  56  92 111 112 ]
 [ 81 87 94  92  54  54  56  54  79  96  79  48  36  44  62 103 107 145 ]
 [ 67 83 91  87  60  59  61  71  91 108  86  65  53  40  63 101 110 121 ]
 [ 49 73 88  72  66  73  78  84 107 120 102  71  57  39  56  89 114 103 ]
 [ 31 61 84  65  73  80  92 103 117 128 114  76  66  57  52  89 111  91 ]
 [  6 51 82  84  92  90  92 114 128 135 122 109  73  69  69  84 109  66 ]
 [  2 44 72  87  95 104 113 124 138 141 130 122  96  77  68  76 104  10 ]
 [  0 37 74  84 102 113 115 131 146 146 133 124 113  94  83  96  90   1 ]
 [  0 33 67  90 113 126 130 140 148 147 136 130 117  95  91  81  71   1 ]
 [  0 33 68  98 122 139 141 144 153 149 135 127 122 108  96  76  65   1 ]
 [  0 36 81 105 127 144 151 151 155 149 125 114 113 121 105  76  49   1 ]
 [  0 39 90 114 131 151 155 157 161 153 122  96 102 107 110  66  50   1 ]]
```

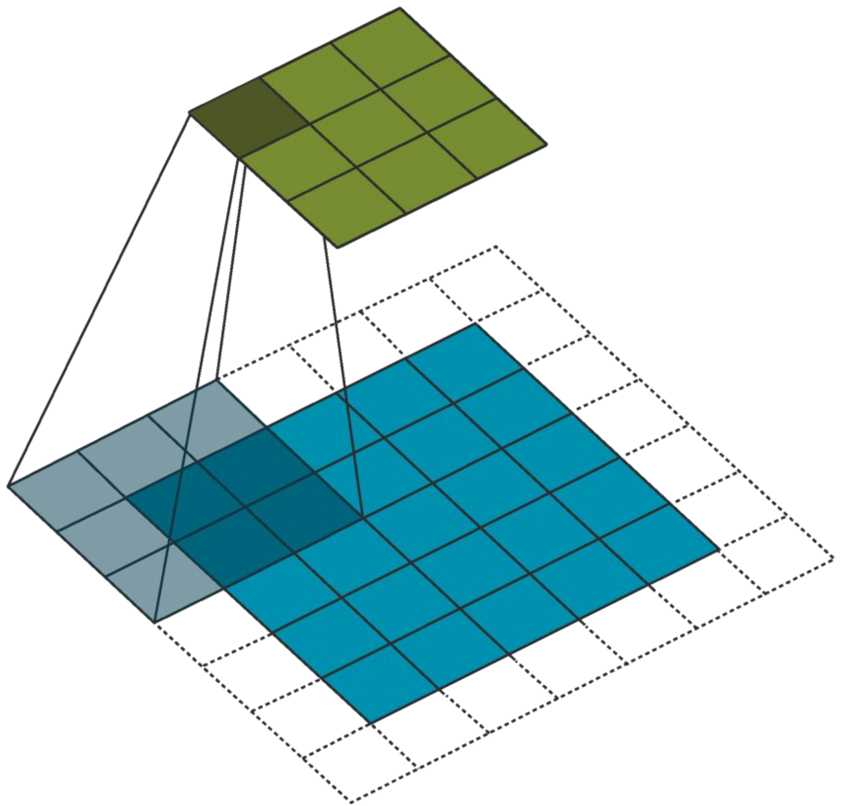
Last time...

IDEA: ANALYZE THE IMAGE BY REGION



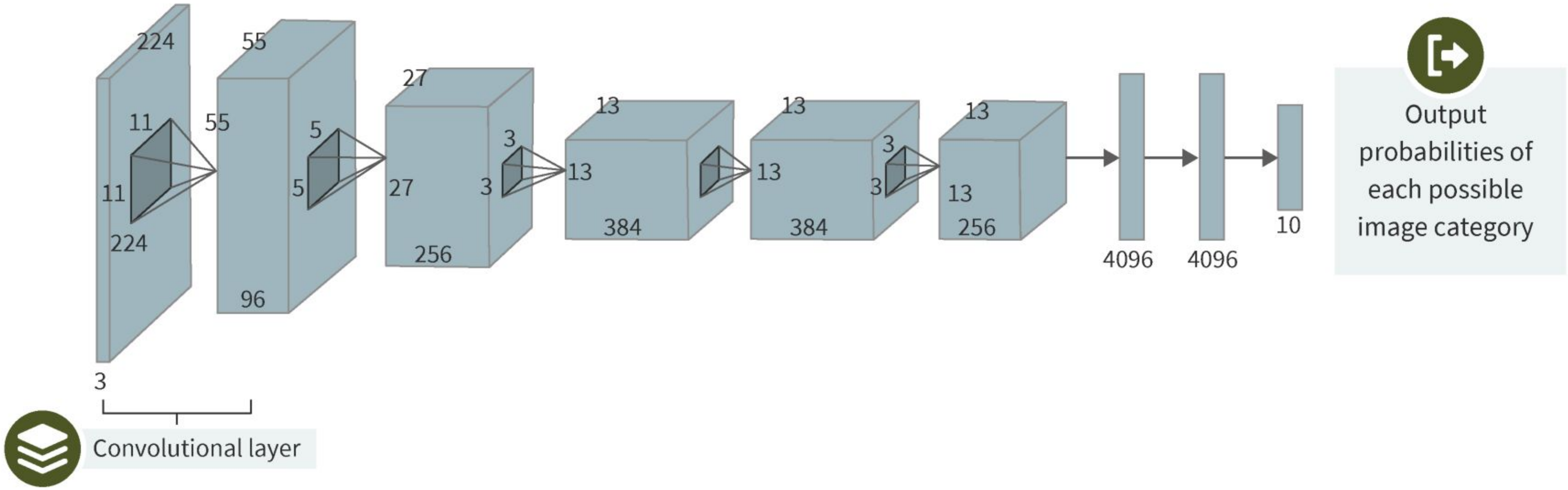
Last time...

ANOTHER VISUALIZATION OF A CONVOLUTIONAL FILTER



Last time...

CONVOLUTIONAL NEURAL NETWORKS (CNNs)



Major advantages:

- Parameter efficient (filters "scan" the entire image for objects)
- Preserves spatial information (activations are the result of regional information)

Today’s Agenda

- More on a different class of unstructured data: text
- More on neural network variants for different types of AI tasks

UNSTRUCTURED DATA

Physician Note

"...PMH of **metastatic breast cancer, R lung malignant** effusion, and **R lung empyema** who presents with increased drainage from **R lung pleurx** tract..."

STRUCTURED DATA

Word Embeddings

Numerical representation of vocabulary of words, where “similar” words have similar representations

[0 0 1 0 0 0 0 …. 0] X (N x D embedding matrix) = [0.5 0.8 0.2]

1xN word input (one-hot selection of word from a length-N vocabulary)

0.5	0.2	0.1
0.6	0.1	0.6
0.5	0.8	0.2
0.7	0.9	0.3
0.3	0.5	0.1
…	…	…
0.7	0.8	0.1

N x D embedding matrix

D-dim word embedding

Aside: how do we learn good word embeddings?

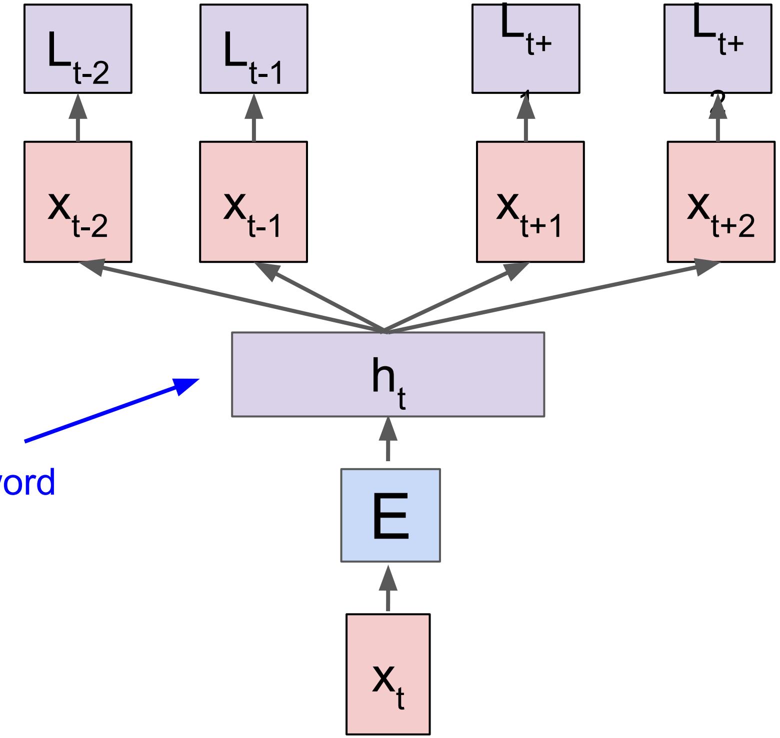
With a deep learning approach! Can design a loss function for a neural network, that captures the desiderata that similar words should occur in similar surrounding context. So, the numerical word embedding for a particular word should be able to be used to give a reasonable prediction of the identity of neighboring words.

Skip-gram model

Can train using a classification loss (e.g. cross-entropy loss we've already seen, extended to more than 2 classes) based only on the text structure, without any external labels!

Use word embedding vector to predict the word identity of a set of neighboring positions

(Each is an N-way classification if the dictionary has N words)



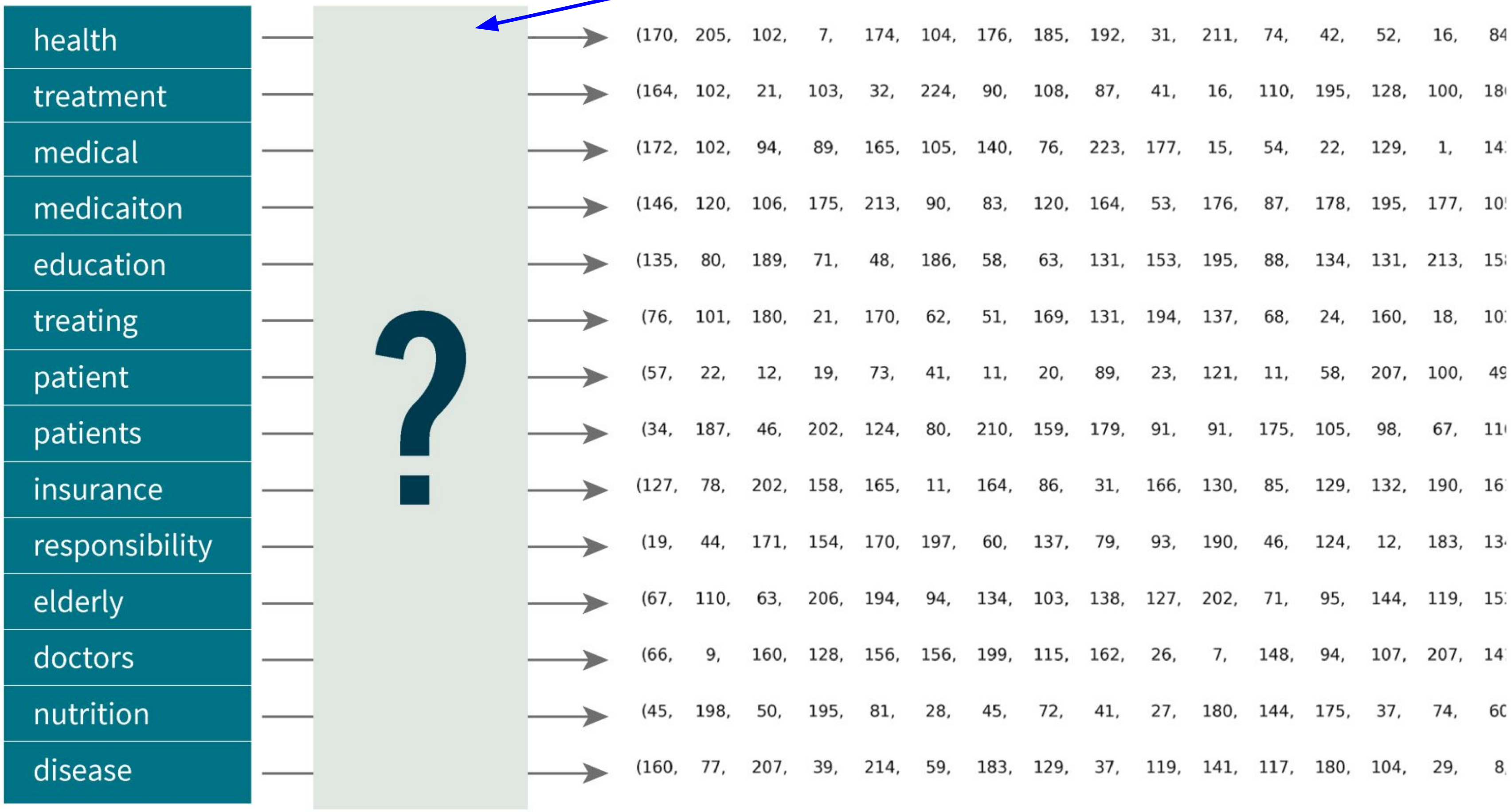
Word embedding (feature vector), of word at the t-th position

Word embeddings

health	→	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,	80,	120,	101,
treatment	→	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,	111,	211,	215,
medical	→	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,	176,	47,	191,
medicaiton	→	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,	56,	74,	203,
education	→	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,	101,	171,	163,
treating	→	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,	5,	20,	112,
patient	→	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,	48,	43,	11,
patients	→	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,	28,	195,	220,
insurance	→	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,	67,	82,	28,
responsibility	→	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,	48,	119,	179,
elderly	→	(67,	110,	63,	206,	194,	94,	134,	103,	138,	127,	202,	71,	95,	144,	119,	152,	109,	95,	47,
doctors	→	(66,	9,	160,	128,	156,	156,	199,	115,	162,	26,	7,	148,	94,	107,	207,	141,	37,	174,	81,
nutrition	→	(45,	198,	50,	195,	81,	28,	45,	72,	41,	27,	180,	144,	175,	37,	74,	60,	208,	197,	109,
disease	→	(160,	77,	207,	39,	214,	59,	183,	129,	37,	119,	141,	117,	180,	104,	29,	8,	144,	183,	112,

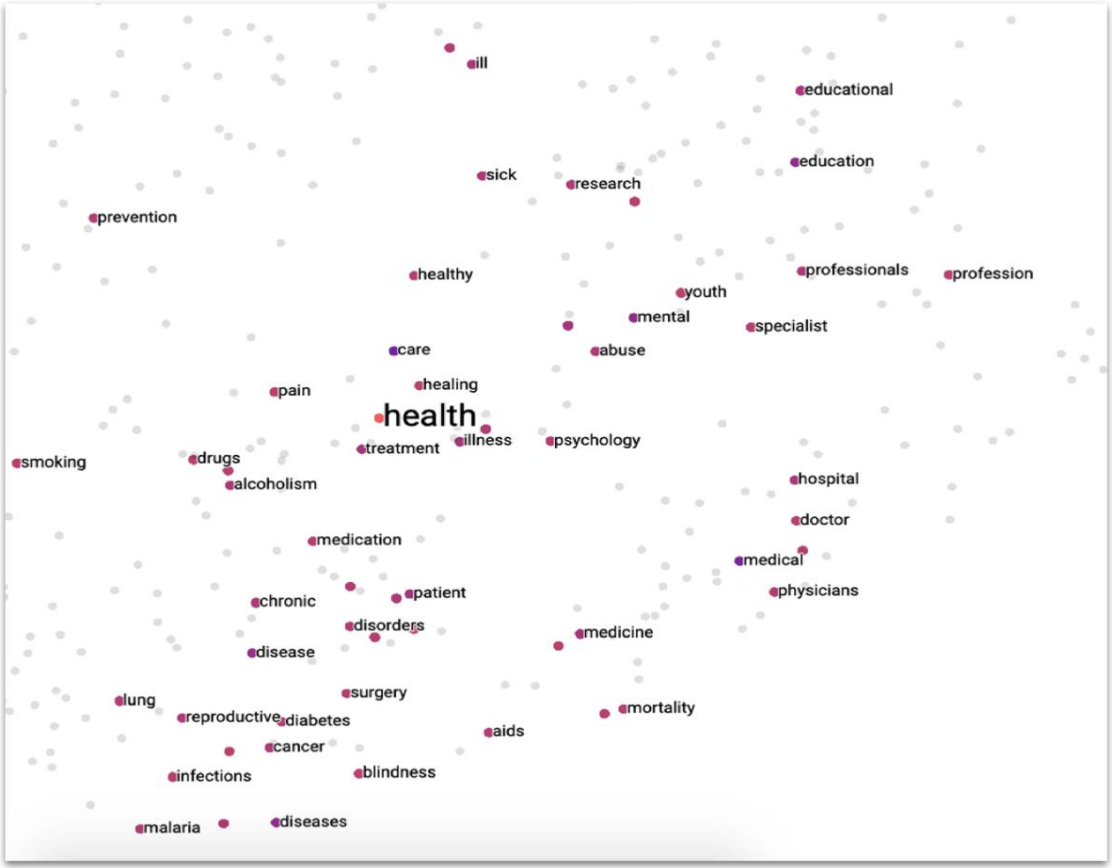
Word embeddings

Learning the word embedding function



Similar words should have similar word embeddings (2D version)

health	(178,168)
treatment	(165, 50)
medical	(209,159)
medicaiton	(208,171)
education	(73, 213)
treating	(207, 68)
patient	(185, 1)
patients	(36, 187)
insurance	(176, 52)
responsibility	(165, 87)
elderly	(94, 17)
doctors	(133,199)
nutrition	(53, 142)
disease	(217,144)



A SENTENCE CAN BE REPRESENTED AS A SEQUENCE OF WORD EMBEDDINGS

the	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
quick	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
brown	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
fox	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
jumped	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
over	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
the	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
lazy	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
dog	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
twice	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,

10-word sentence with
1024D embeddings → **10,240 features**

EXAMPLE OF A SEQUENCE-BASED TASK

Task: Identify if a given sentence describes a hypermetabolic region.

There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,

EXAMPLE OF A SEQUENCE-BASED TASK

Task: Identify if a given sentence describes a hypermetabolic region.

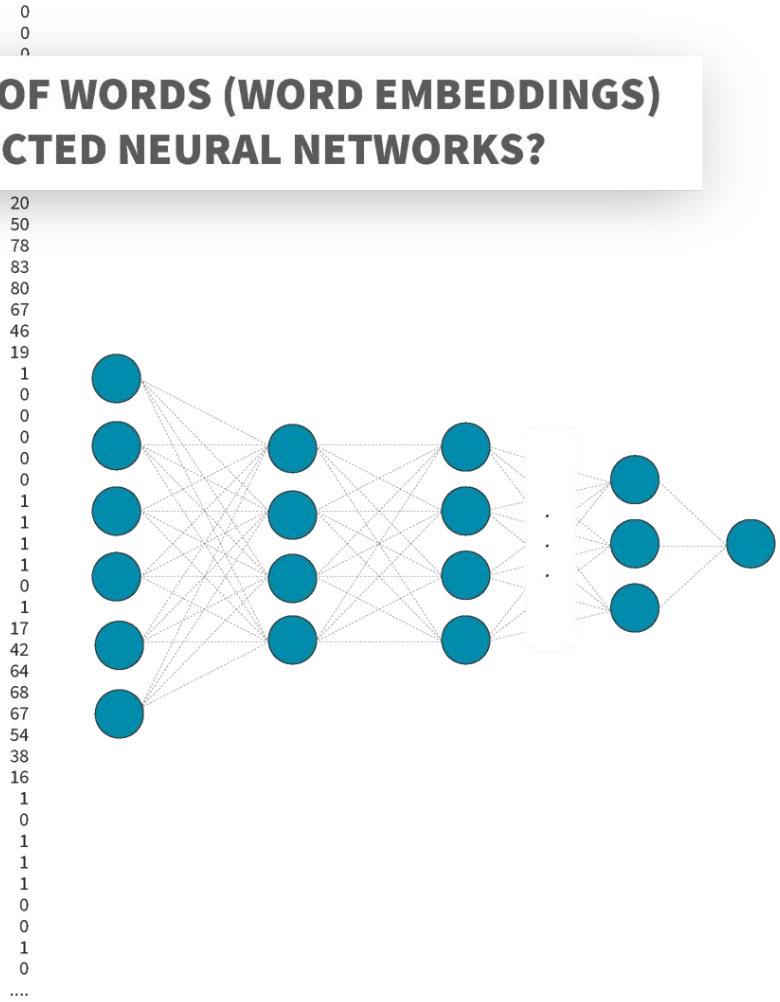
There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,

Discriminating factors:

1. Mentions of FDG uptake
2. Qualitative evaluation of the region and uptake.

CAN WE REPRESENT SEQUENCES OF WORDS (WORD EMBEDDINGS) AS INPUT TO FULLY-CONNECTED NEURAL NETWORKS?

First attempt: Connect the embeddings!

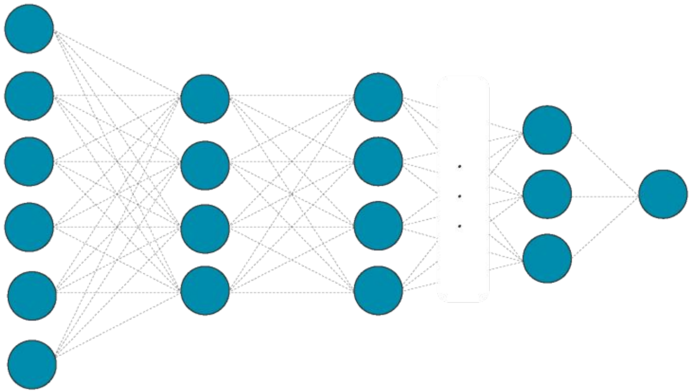


CAN WE REPRESENT SEQUENCES OF WORDS (WORD EMBEDDINGS) AS INPUT TO FULLY-CONNECTED NEURAL NETWORKS?

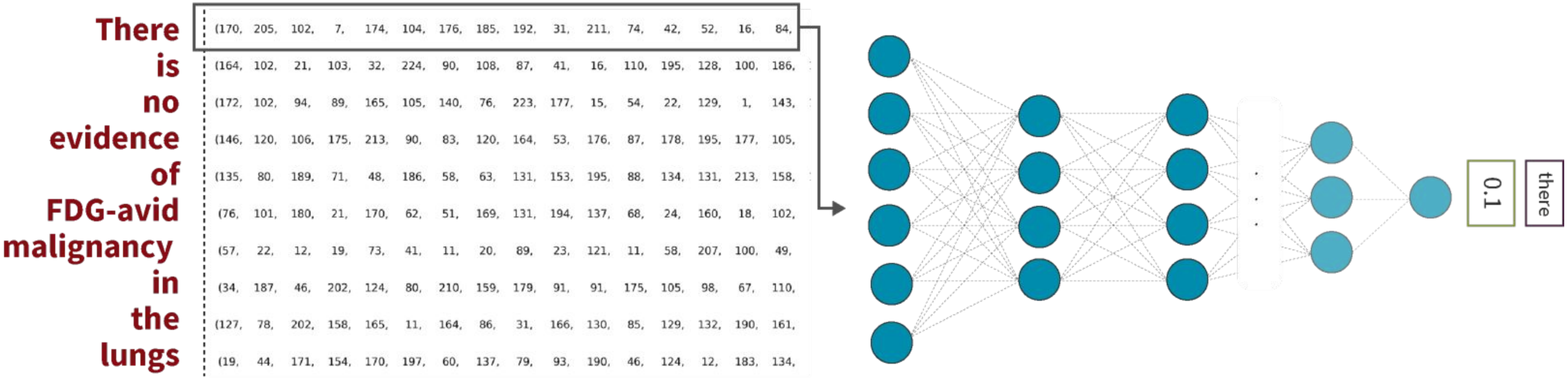
First attempt: Connect the embeddings!

Problems:

- Requires a HUGE number of parameters
- Highly sensitive to word displacement
- Destroys temporal information
- Cannot handle variable-sized input

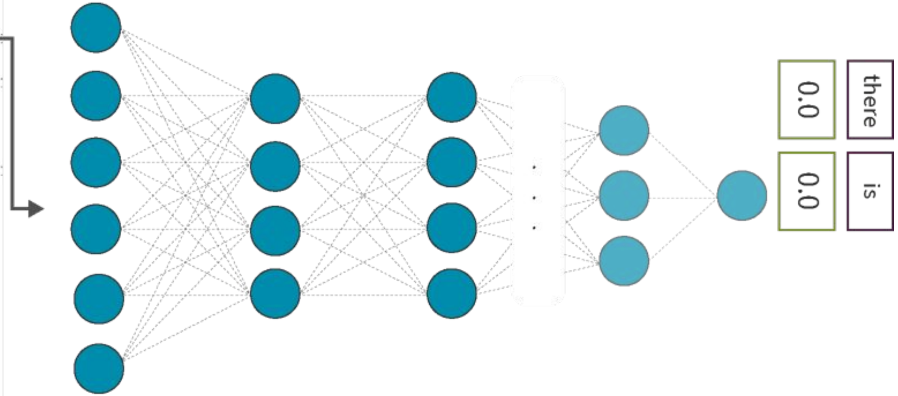


IDEA #1: PROCESS ONE WORD AT A TIME



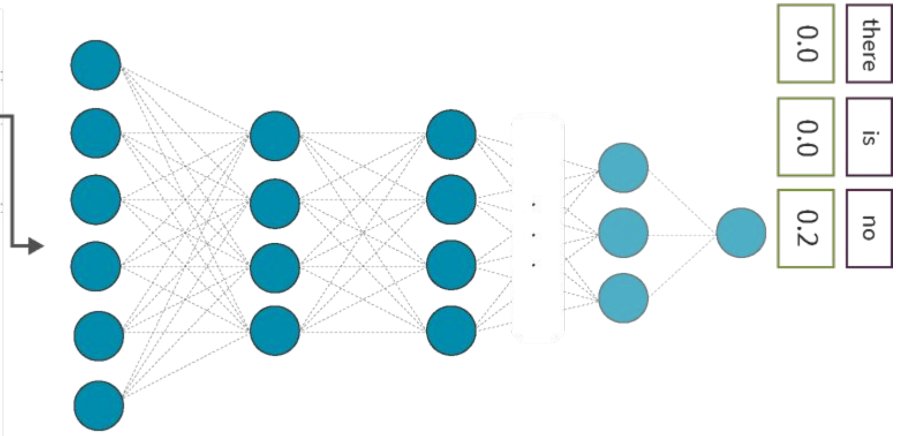
IDEA #1: PROCESS ONE WORD AT A TIME

There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,

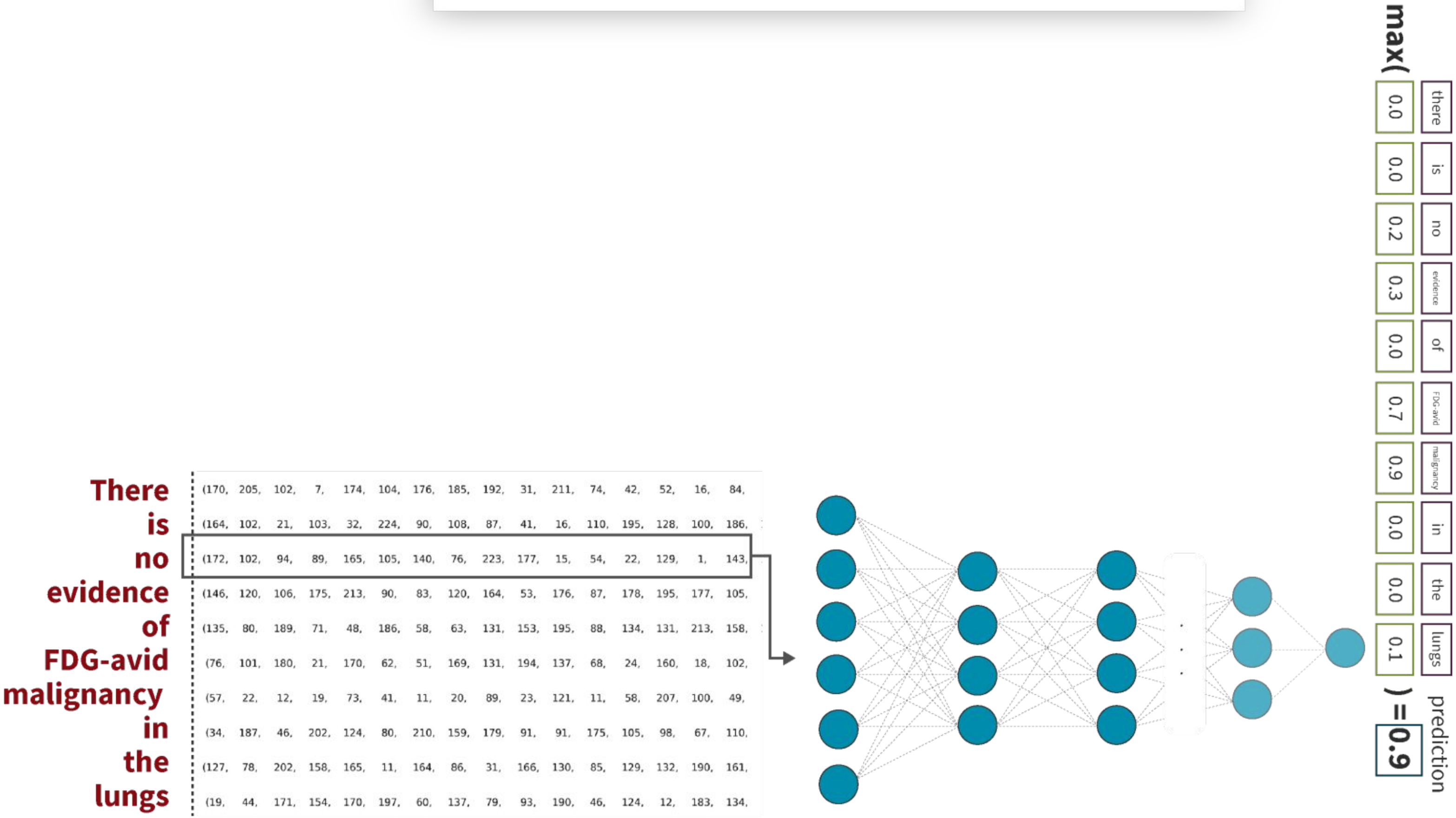


IDEA #1: PROCESS ONE WORD AT A TIME

There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



IDEA #1: PROCESS ONE WORD AT A TIME

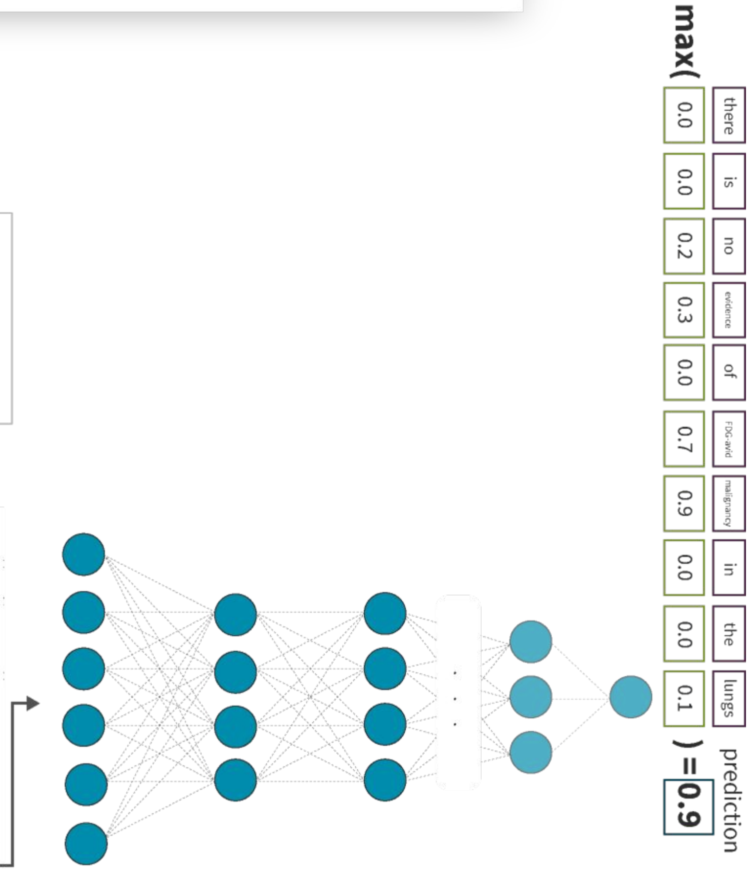


IDEA #1: PROCESS ONE WORD AT A TIME

Problems:

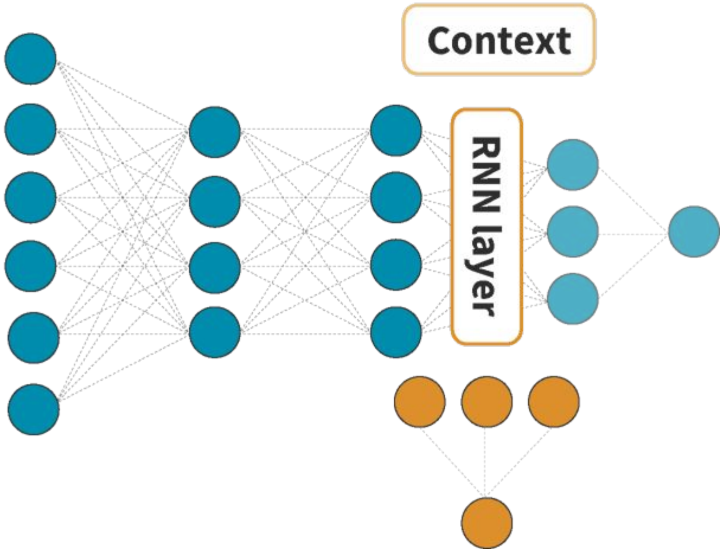
probabilities do not take into account relationships that words have with each other

There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



IDEA #2: PASS ALONG INFORMATION

There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,

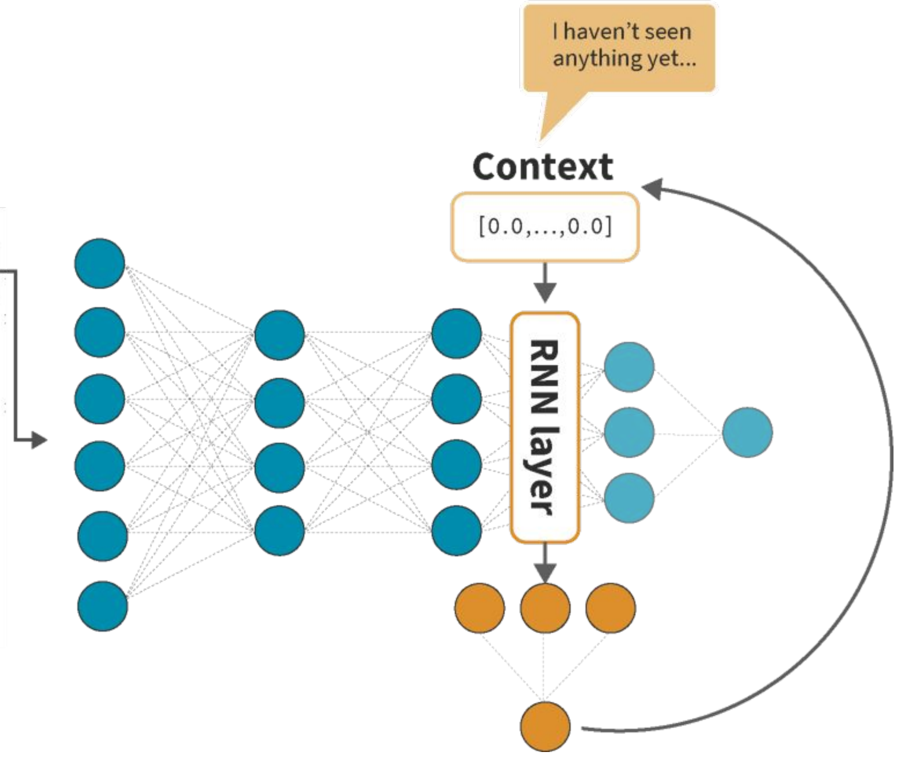


IDEA #2: PASS ALONG INFORMATION

IDEA #2: PASS ALONG INFORMATION

The “context” provides information from words already seen that could be helpful in the future.

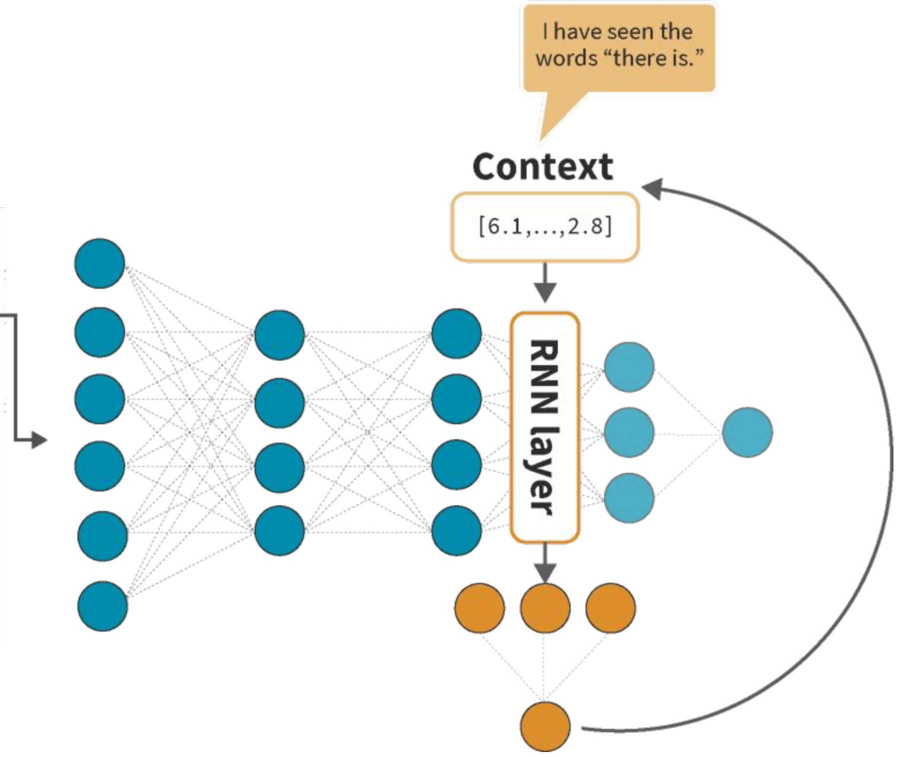
There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



IDEA #2: PASS ALONG INFORMATION

The "context" provides information from words already seen that could be helpful in the future.

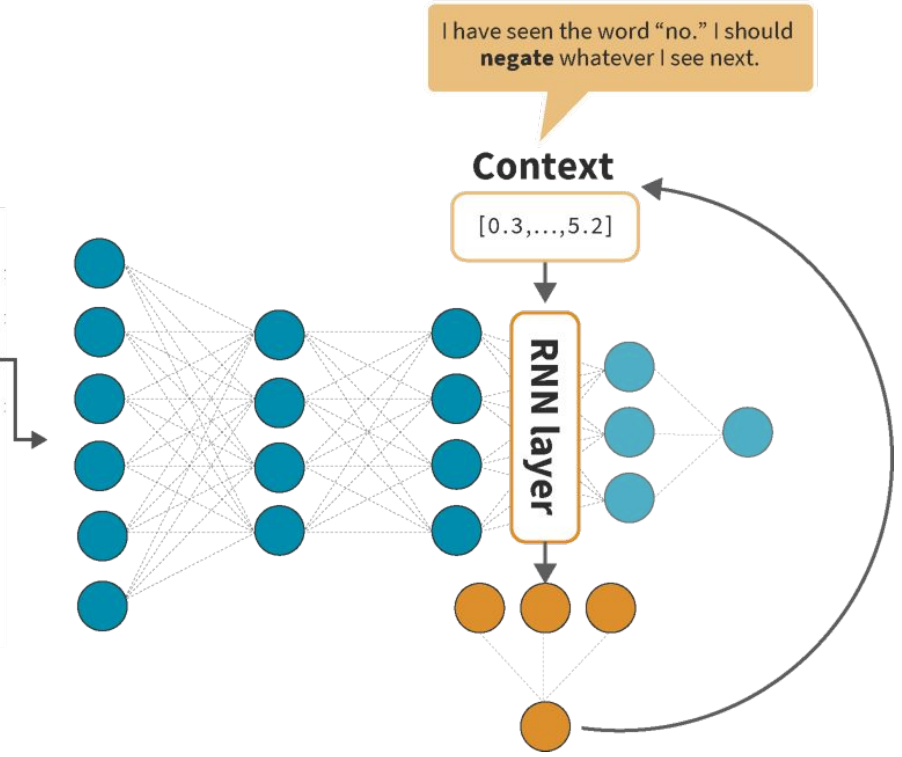
There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



IDEA #2: PASS ALONG INFORMATION

The "context" provides information from words already seen that could be helpful in the future

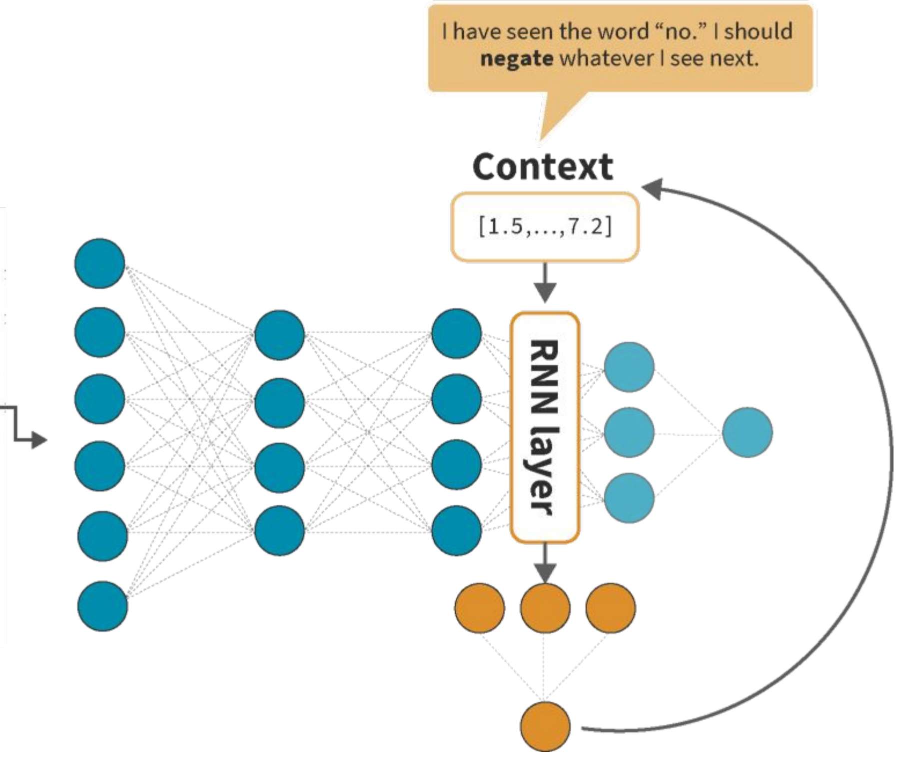
There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



IDEA #2: PASS ALONG INFORMATION

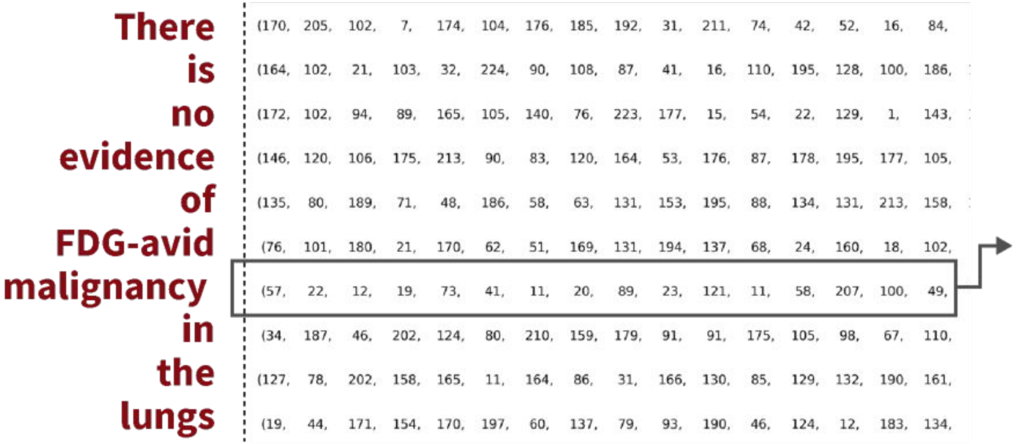
The "context" provides information from words already seen that could be helpful in the future

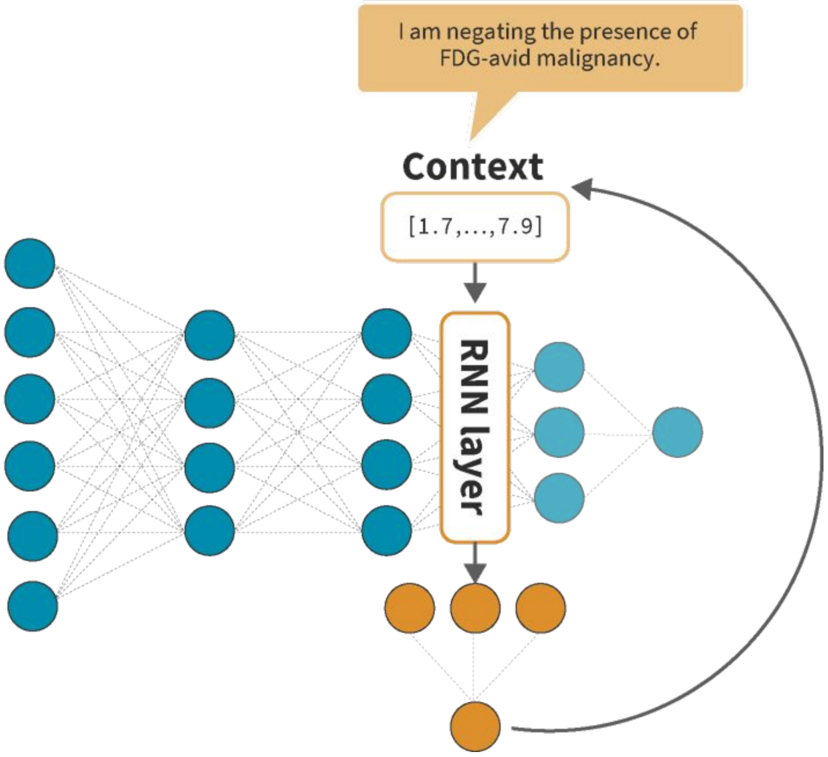
There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



IDEA #2: PASS ALONG INFORMATION

The “context” provides information from words already seen that could be helpful in the future

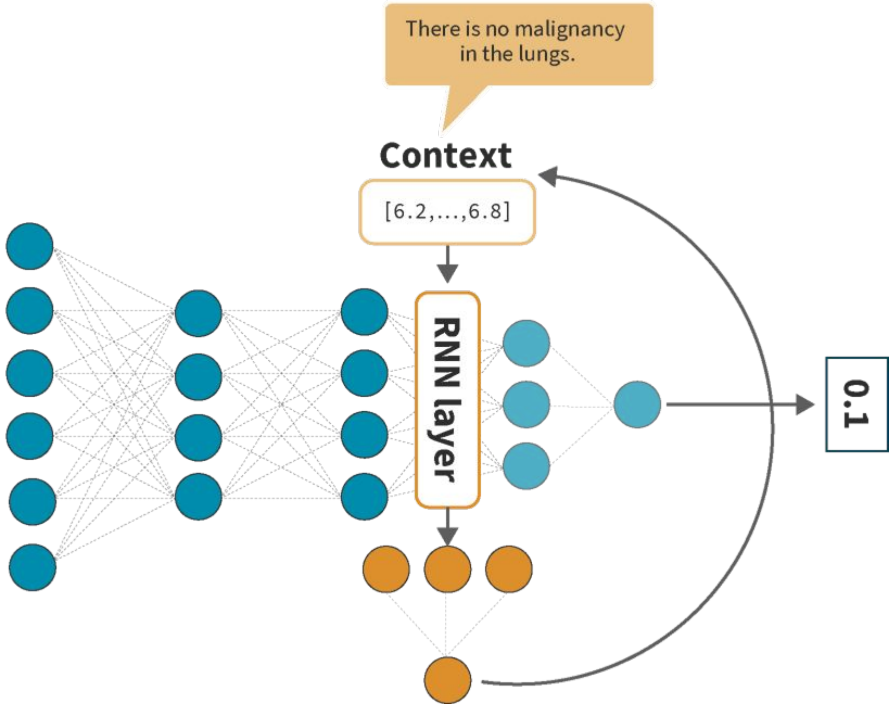




IDEA #2: PASS ALONG INFORMATION

The "context" provides information from words already seen that could be helpful in the future

There	(170,	205,	102,	7,	174,	104,	176,	185,	192,	31,	211,	74,	42,	52,	16,	84,
is	(164,	102,	21,	103,	32,	224,	90,	108,	87,	41,	16,	110,	195,	128,	100,	186,
no	(172,	102,	94,	89,	165,	105,	140,	76,	223,	177,	15,	54,	22,	129,	1,	143,
evidence	(146,	120,	106,	175,	213,	90,	83,	120,	164,	53,	176,	87,	178,	195,	177,	105,
of	(135,	80,	189,	71,	48,	186,	58,	63,	131,	153,	195,	88,	134,	131,	213,	158,
FDG-avid	(76,	101,	180,	21,	170,	62,	51,	169,	131,	194,	137,	68,	24,	160,	18,	102,
malignancy	(57,	22,	12,	19,	73,	41,	11,	20,	89,	23,	121,	11,	58,	207,	100,	49,
in	(34,	187,	46,	202,	124,	80,	210,	159,	179,	91,	91,	175,	105,	98,	67,	110,
the	(127,	78,	202,	158,	165,	11,	164,	86,	31,	166,	130,	85,	129,	132,	190,	161,
lungs	(19,	44,	171,	154,	170,	197,	60,	137,	79,	93,	190,	46,	124,	12,	183,	134,



Another more recent approach for reasoning over sequences, which ingests the entire sequence at once rather than sequentially: “Transformer” architectures

FUNDAMENTAL CONCEPT: *the meaning of a word changes based on context*

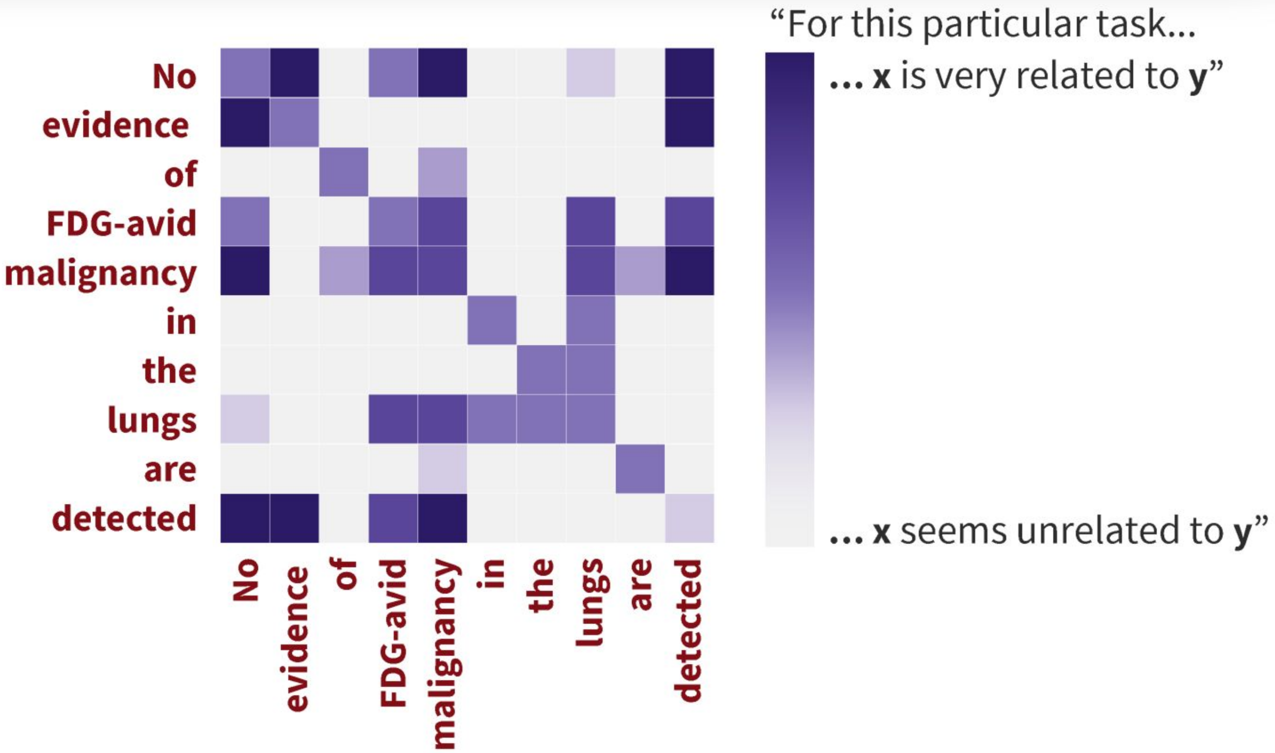
FUNDAMENTAL CONCEPT: *the meaning of a word changes based on context*

"I broke my **leg** while running."

vs.

"I was trailing in the first **leg** of the marathon."

SELF-ATTENTION LAYERS CALCULATE WEIGHTING SCORES FOR THE STRENGTH OF CONTEXTUAL SIGNAL FROM ANY WORD TO ANOTHER



ELEMENTS AT EACH POSITION ARE TRANSFORMED TO INCORPORATE WEIGHTED CONTEXT FROM OTHER ELEMENTS

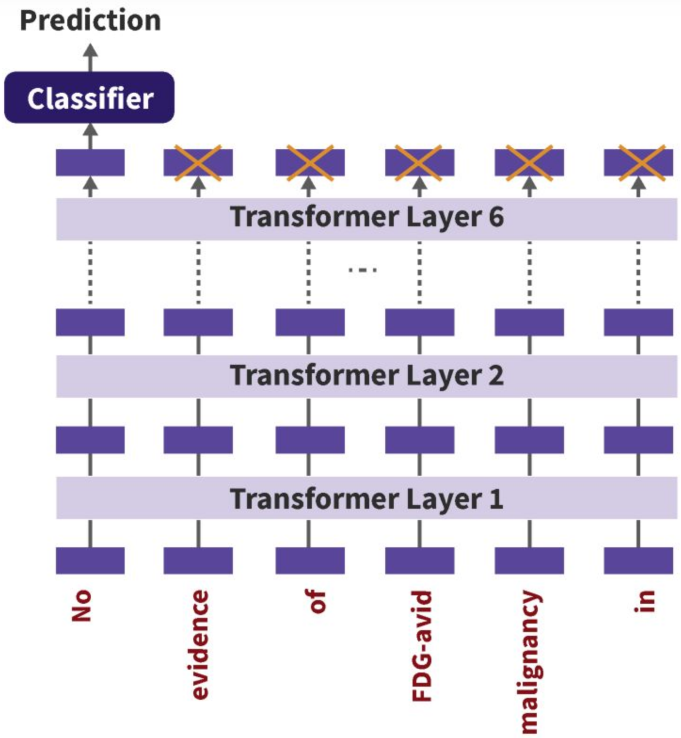
Feature values at each element position

Contextual weighting scores from each input element to "detection"

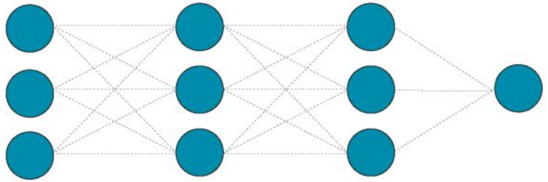
0.3	X	(100,	73,	180	21,	170,	62,	51,	169,	131,	194,	137,	21,	71,	222,	201,	151,
0.2	X	(225,	125,	12,	19,	73,	41,	11,	20,	89,	23,	121,	19,	29,	10,	108,	109,
0.0	X	(0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,
0.1	X	(370,	130,	202,	158,	124,	11,	210,	159,	31,	166,	158,	138,	358,	188,	198,	18,
0.3	X	(160,	56,	171,	154,	165,	197,	164,	86,	78,	137,	154,	194,	44,	64,	174,	14,
0.0	X	(0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,
0.0	X	(0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,
0.0	X	(0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,
0.0	X	(0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,
+ **0.0**	X	(0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,

Output = weighted sum (160, 77, 207, 39, 214, 59, 183, 129, 37, 119, 141, 117, 180, 104, 29, 8,

TRANSFORMER ARCHITECTURES: NEURAL NETWORKS WITH STACKS OF SELF-ATTENTION LAYERS

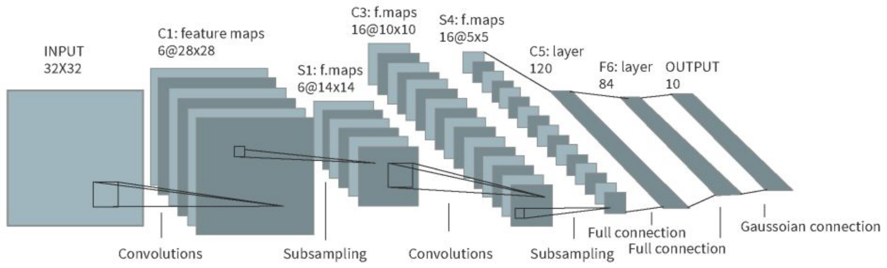


MAJOR CLASSES OF NEURAL NETWORK ARCHITECTURES



Fully connected neural networks

(fully connected layers, good for structured data inputs)

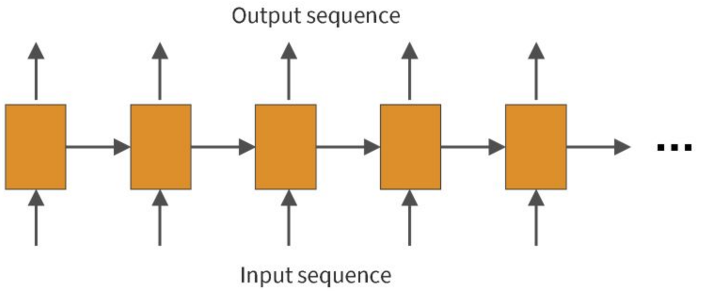


Convolutional neural networks

(convolutional layers, good for image inputs)

Recurrent neural networks

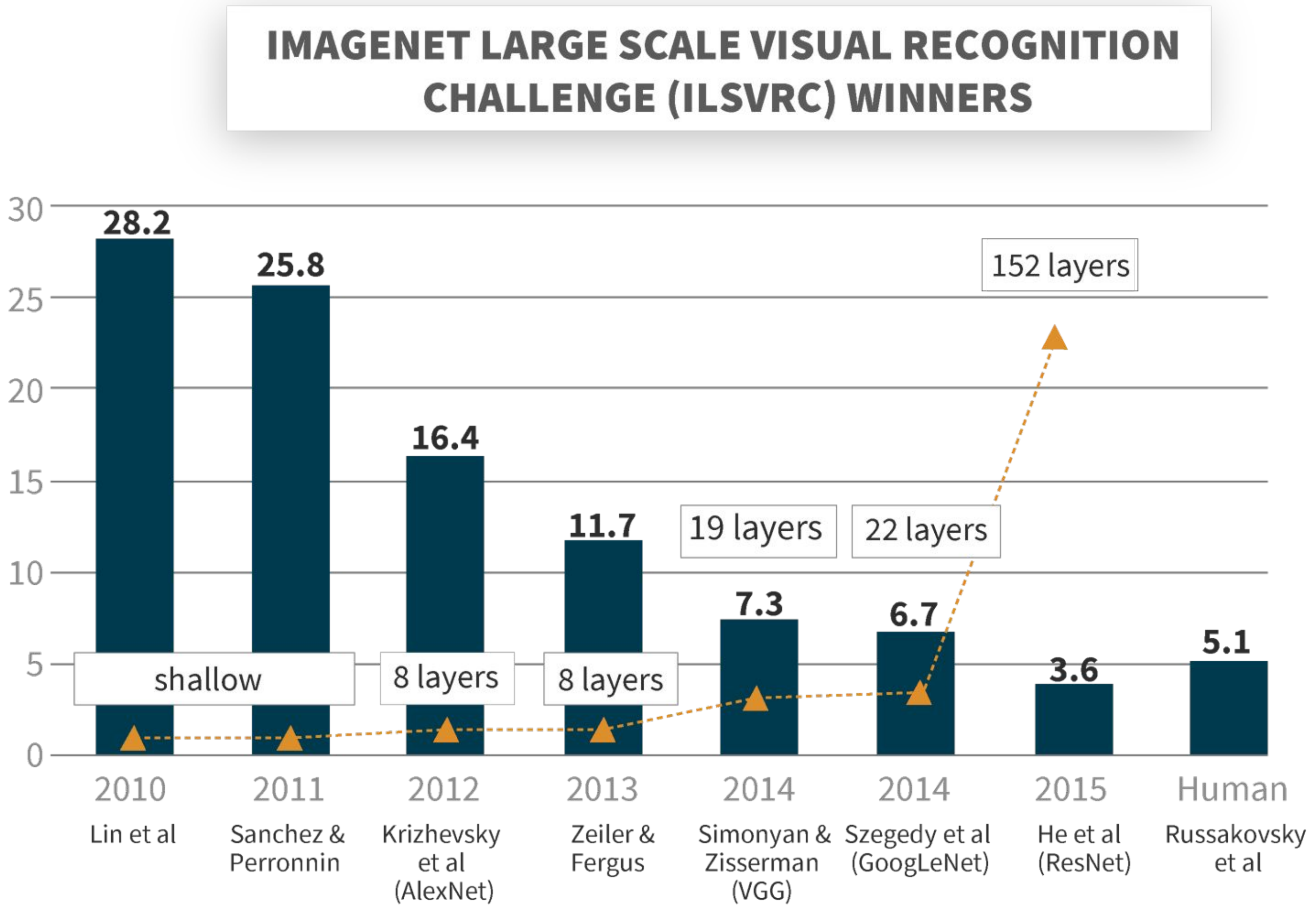
(fully connected layers modeling recurrence relation across sequence, good for sequence inputs)

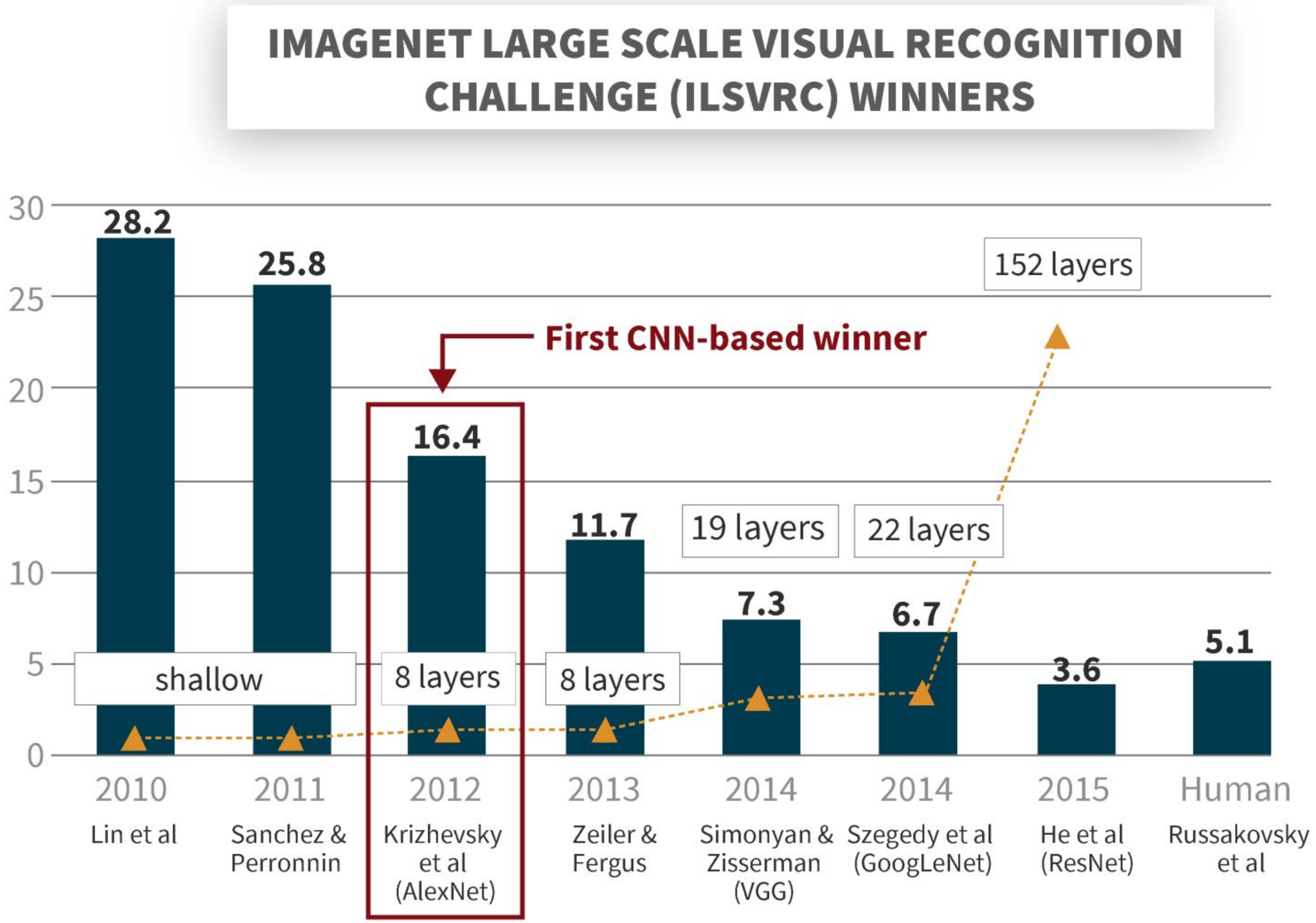


Let’s revisit some more neural network architectures for different AI outputs

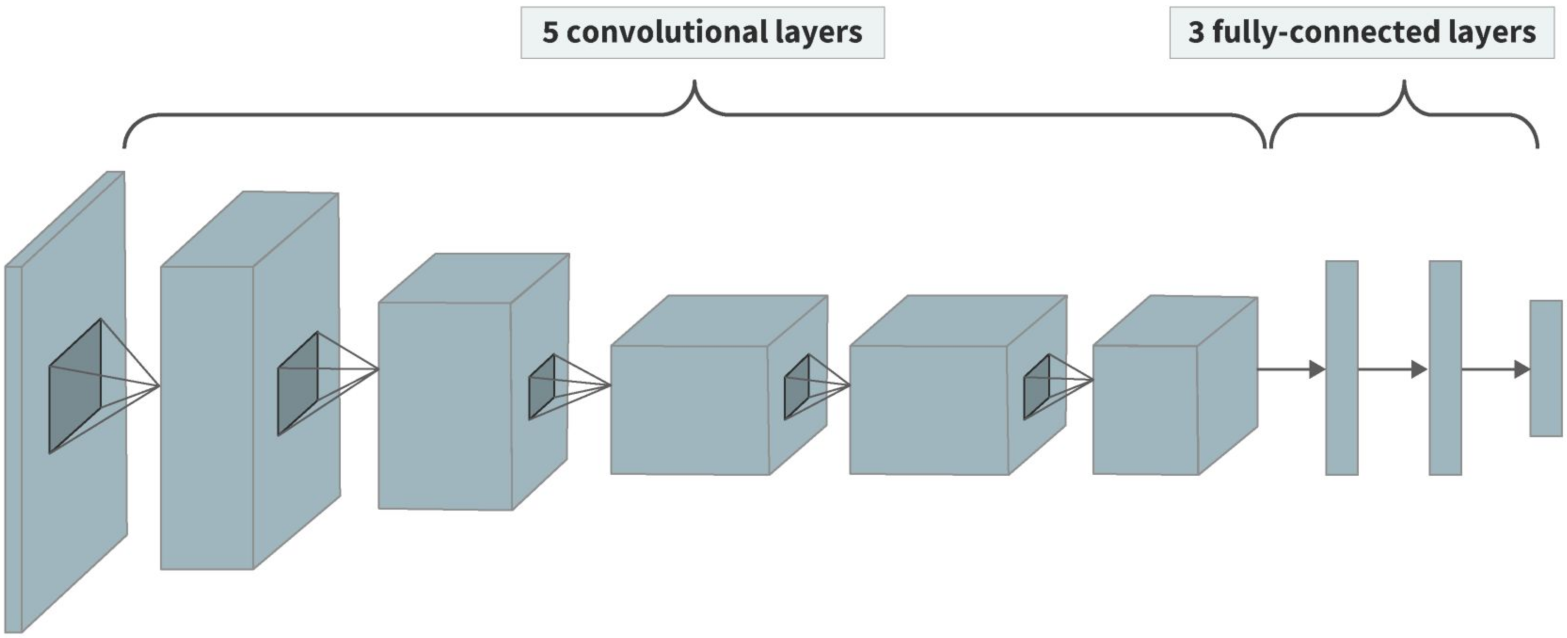
Let’s revisit some more neural network architectures for different AI outputs

First, some commonly used architectures for **classification**

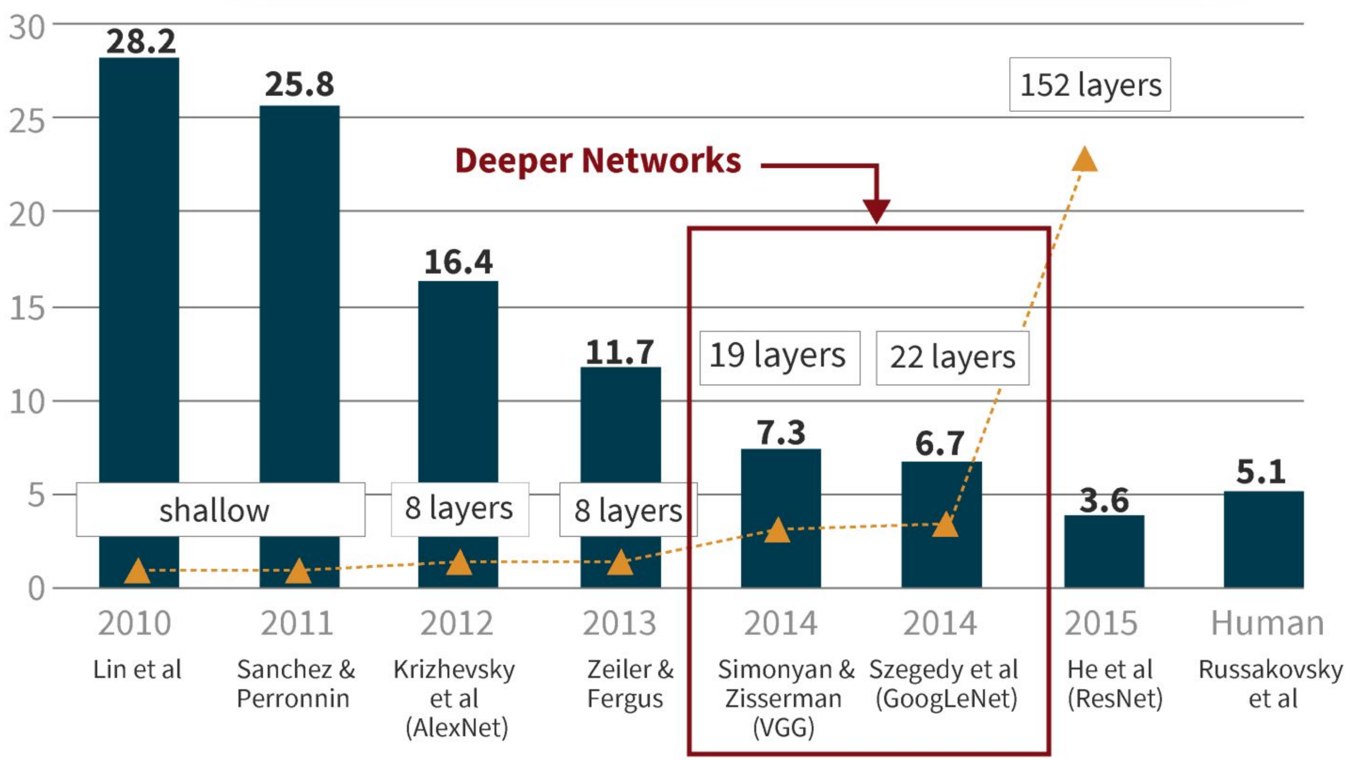

IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE (ILSVRC) WINNERS
30
25
20
15
10
5
0
28.2
25.8
16.4
11.7
7.3
6.7
3.6
5.1
shallow
8 layers
8 layers
19 layers
22 layers
152 layers
2010
Lin et al
2011
Sanchez & Perronnin
2012
Krizhevsky et al (AlexNet)
2013
Zeiler & Fergus
2014
Simonyan & Zisserman (VGG)
2014
Szegedy et al (GoogLeNet)
2015
He et al (ResNet)
Human
Russakovsky et al


IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE (ILSVRC) WINNERS
30
25
20
15
10
5
0
28.2
25.8
16.4
11.7
7.3
6.7
3.6
5.1
First CNN-based winner
shallow
8 layers
8 layers
19 layers
22 layers
152 layers
2010
Lin et al
2011
Sanchez & Perronnin
2012
Krizhevsky et al (AlexNet)
2013
Zeiler & Fergus
2014
Simonyan & Zisserman (VGG)
2014
Szegedy et al (GoogLeNet)
2015
He et al (ResNet)
Human
Russakovsky et al

ALEXNET (2012)



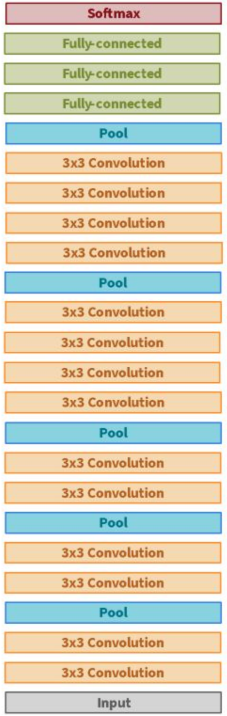
IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE (ILSVRC) WINNERS



VGGNET (2014)

Smaller filter sizes (3x3 convolutions)

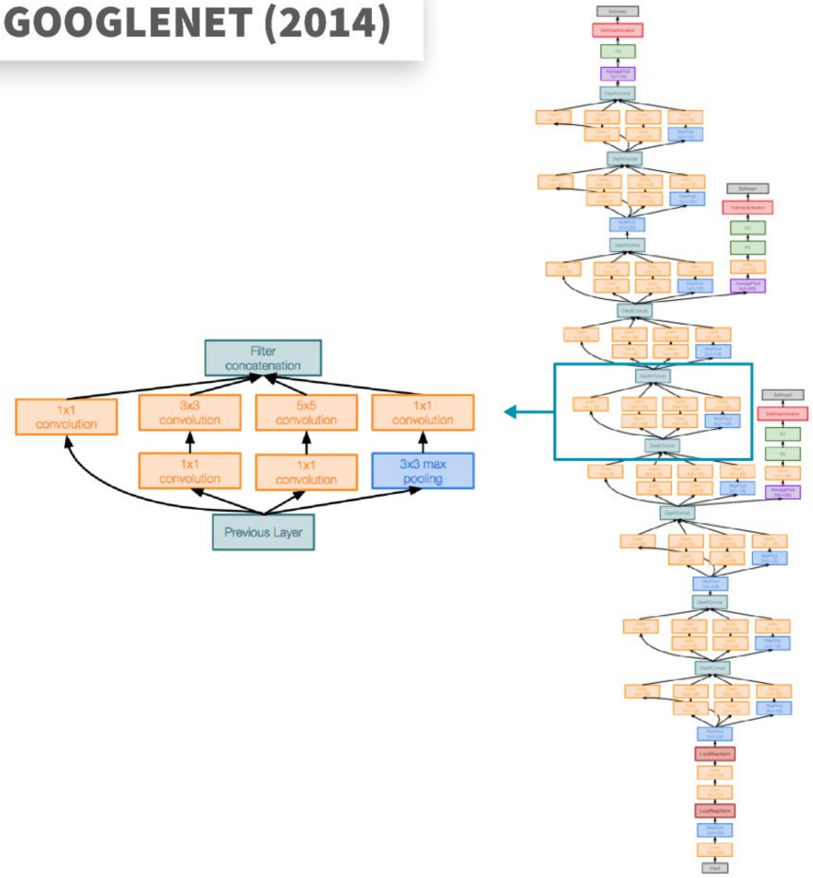
Deeper network (19 layers)



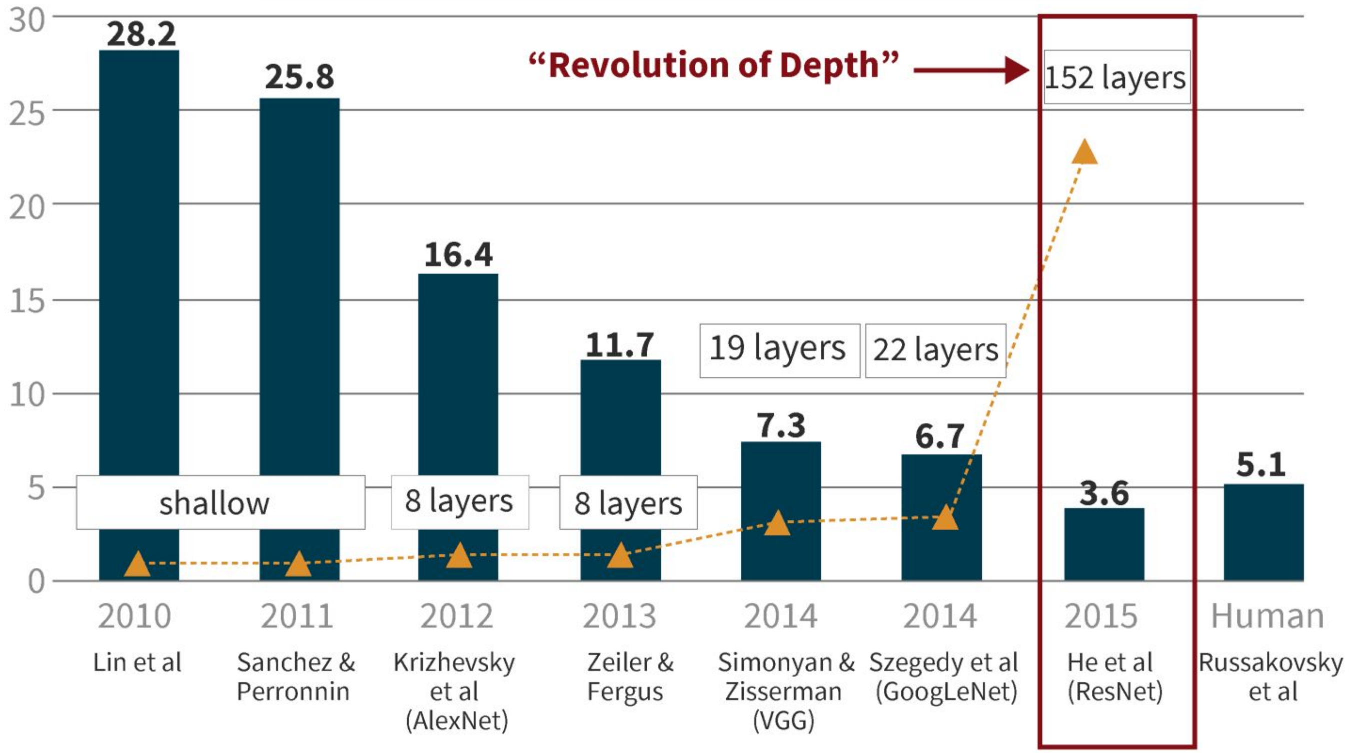
VGG19

GOOGLENET (2014)

Stack of efficient “Inception modules”
Total depth of 22 layers

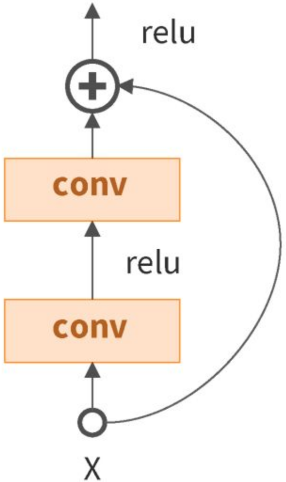


IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE (ILSVRC) WINNERS

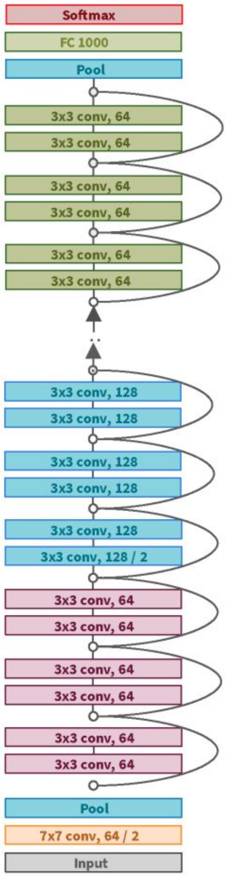


RESNET (2015)

Stack of residual blocks with skip connections
Allows easier optimization of loss function to enable very deep networks

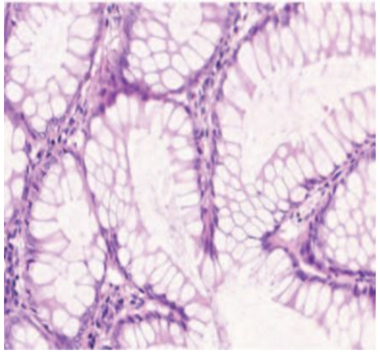


Residual block



Richer visual recognition tasks: segmentation and detection

Classification

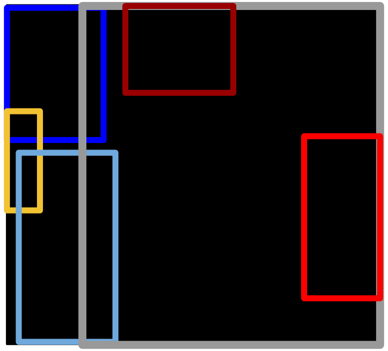


Output:
one category label for image (e.g., colorectal glands)

Semantic Segmentation

Output:
category label for each pixel in the image

Detection



Output:
Spatial bounding box for each **instance** of a category object in the image

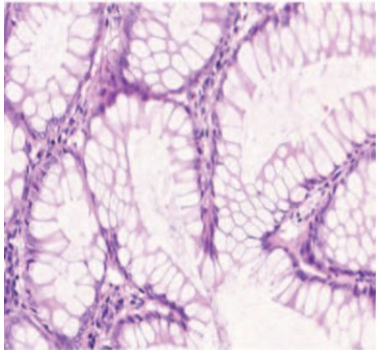
Instance Segmentation

Output:
Category label and instance label for each pixel in the image

Figures: Chen et al. 2016. https://arxiv.org/pdf/1604.02677.pdf

Richer visual recognition tasks: segmentation and detection

Classification

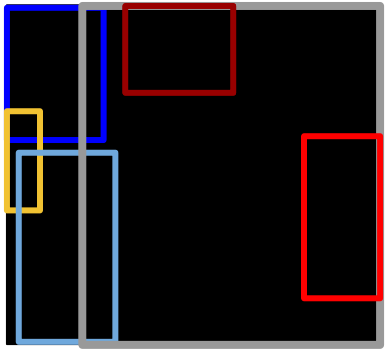


Output:
one category label for image (e.g., colorectal glands)

Semantic Segmentation

Output:
category label for each pixel in the image

Detection



Output:
Spatial bounding box for each **instance** of a category object in the image

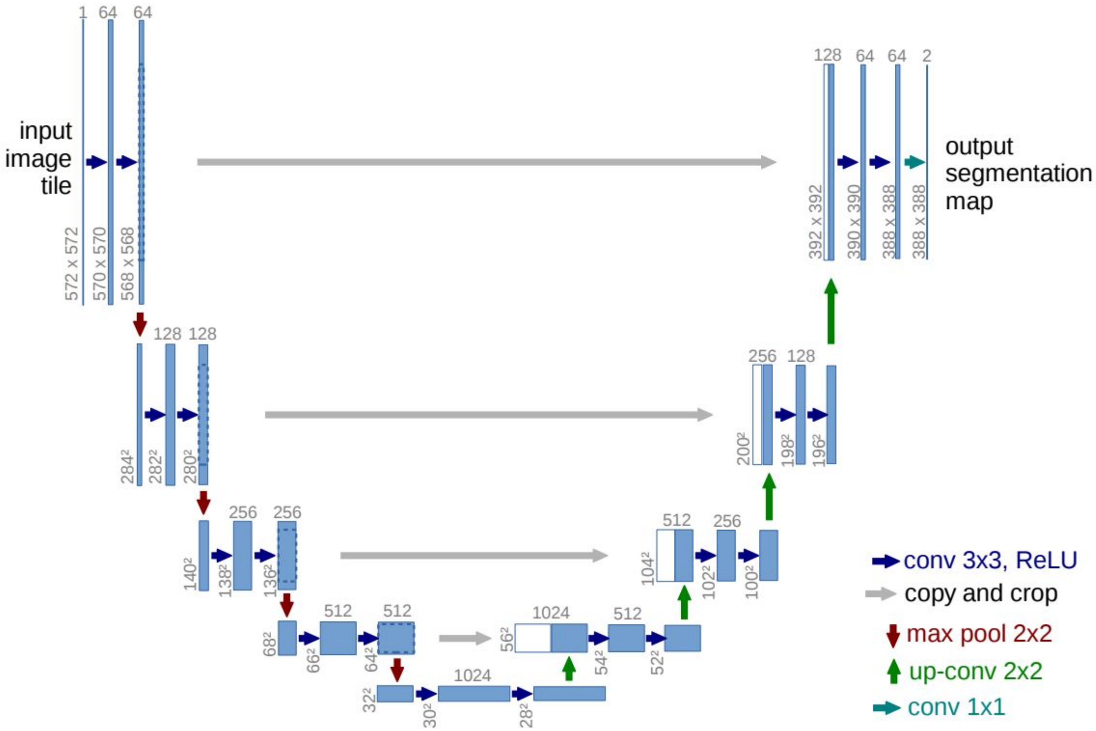
Instance Segmentation

Output:
Category label and instance label for each pixel in the image

Distinguishes between different instances of an object

Figures: Chen et al. 2016. https://arxiv.org/pdf/1604.02677.pdf

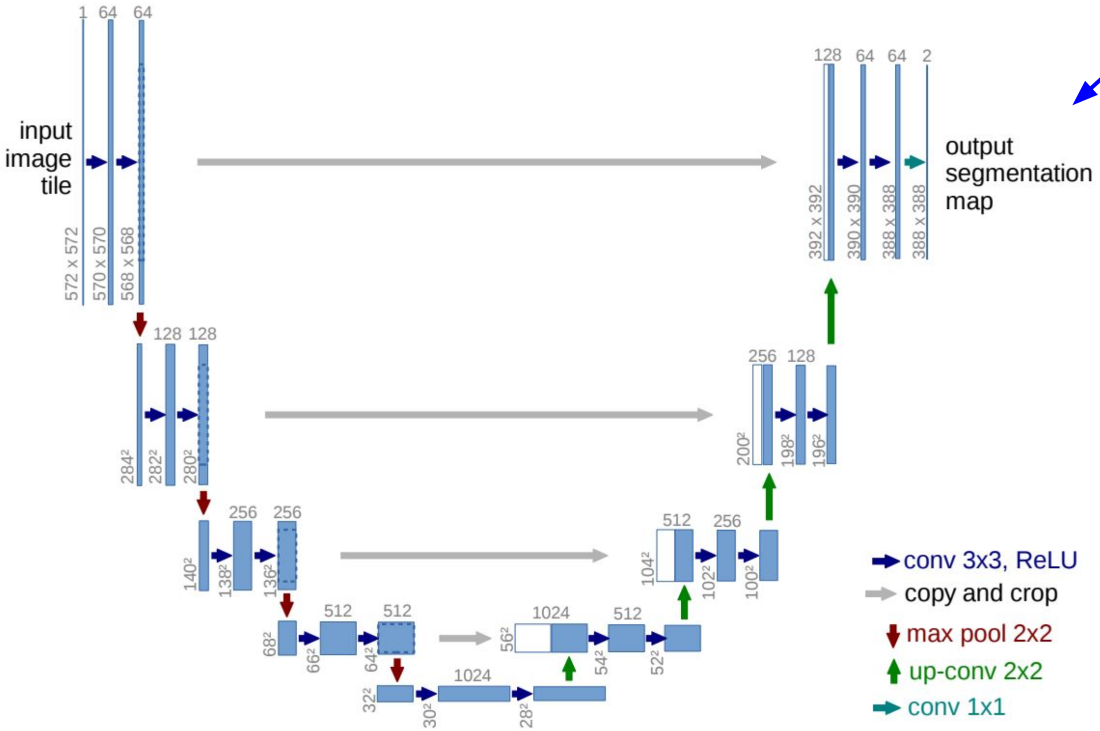
Semantic segmentation: U-Net



Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

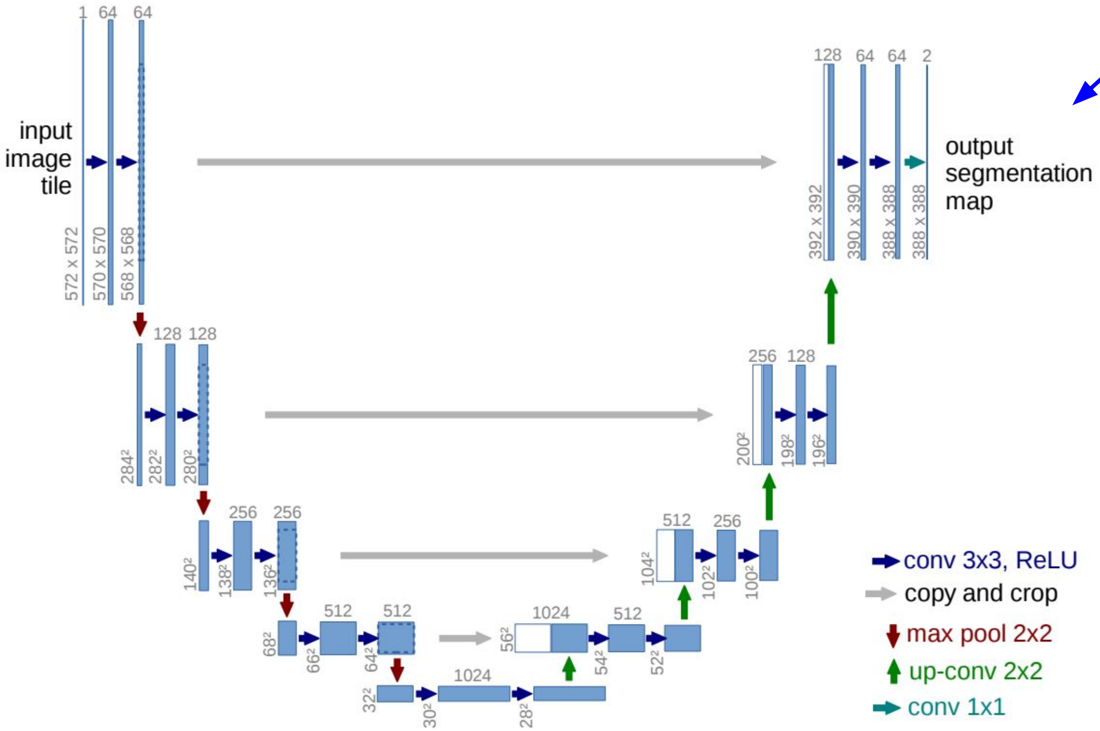
Semantic segmentation: U-Net

Output is an image mask: width x height x # classes



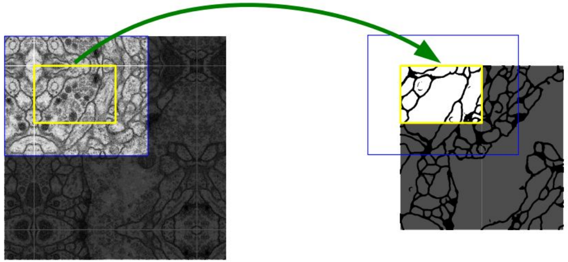
Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Semantic segmentation: U-Net



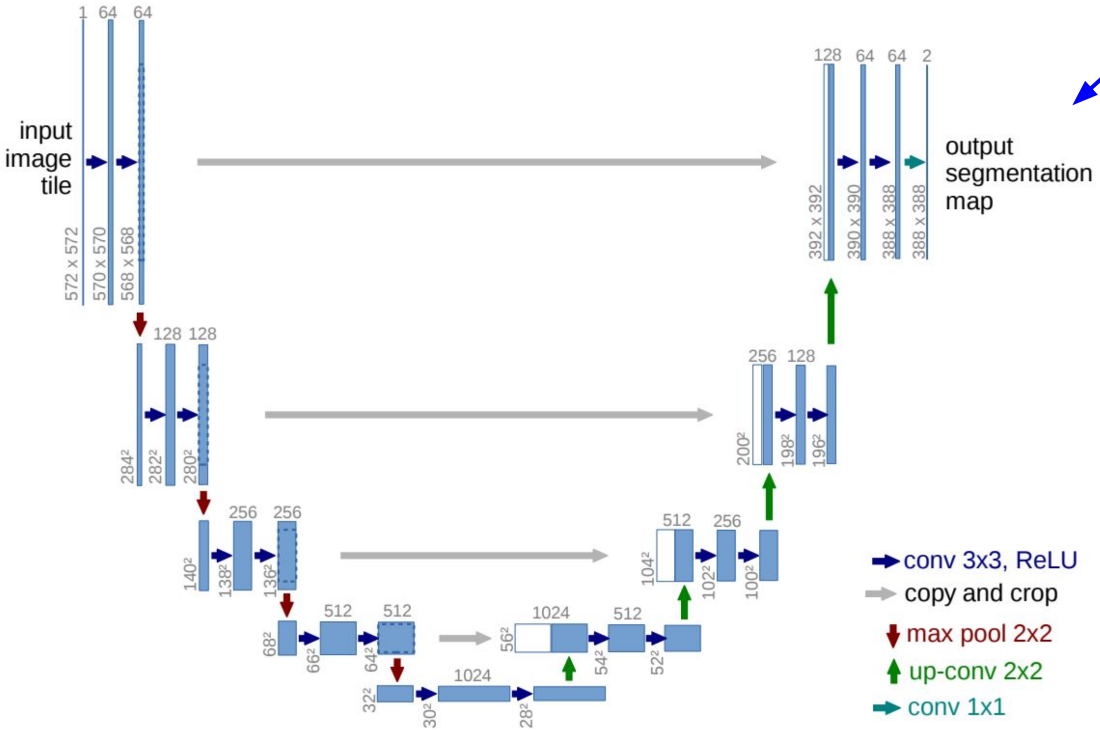
Output is an image mask: width x height x # classes

Output image size a little smaller than original, due to convolutional operations w/o padding



Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

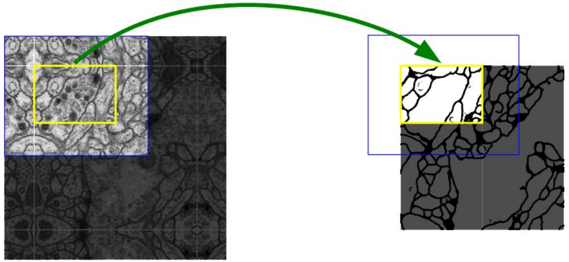
Semantic segmentation: U-Net



Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Output is an image mask: width x height x # classes

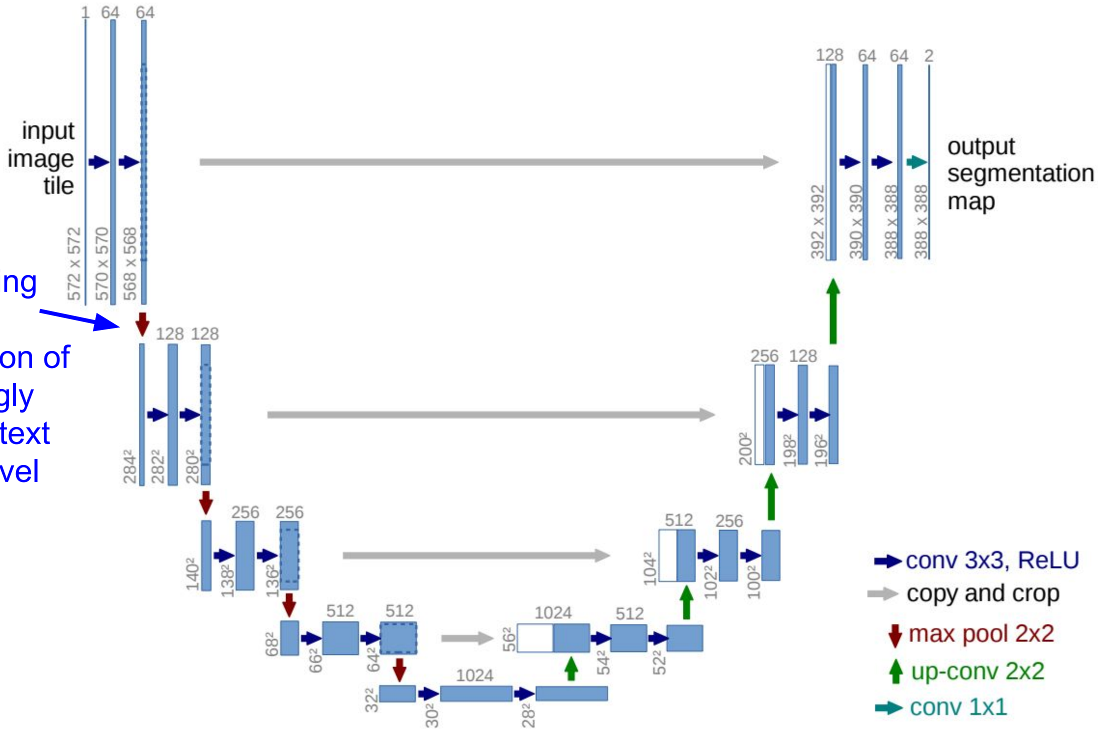
Output image size a little smaller than original, due to convolutional operations w/o padding



Gives more “true” context for reasoning over each image area. Can tile to make predictions for arbitrarily large images

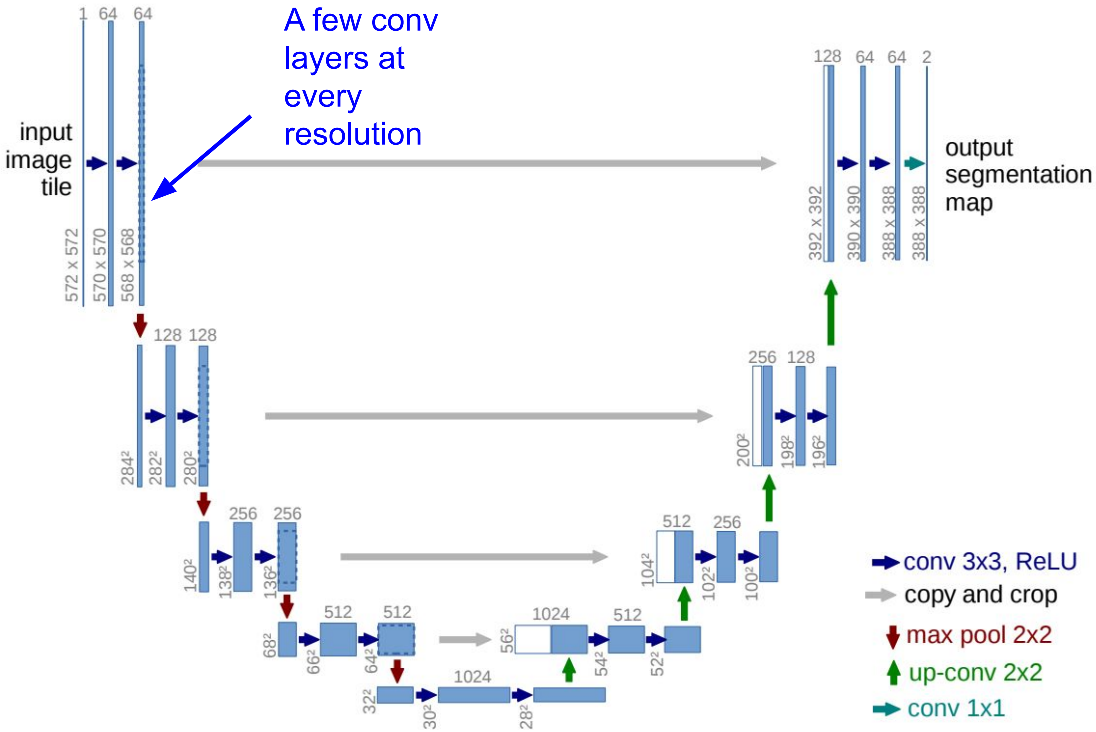
Semantic segmentation: U-Net

Max pooling enables aggregation of increasingly more context (higher level features)



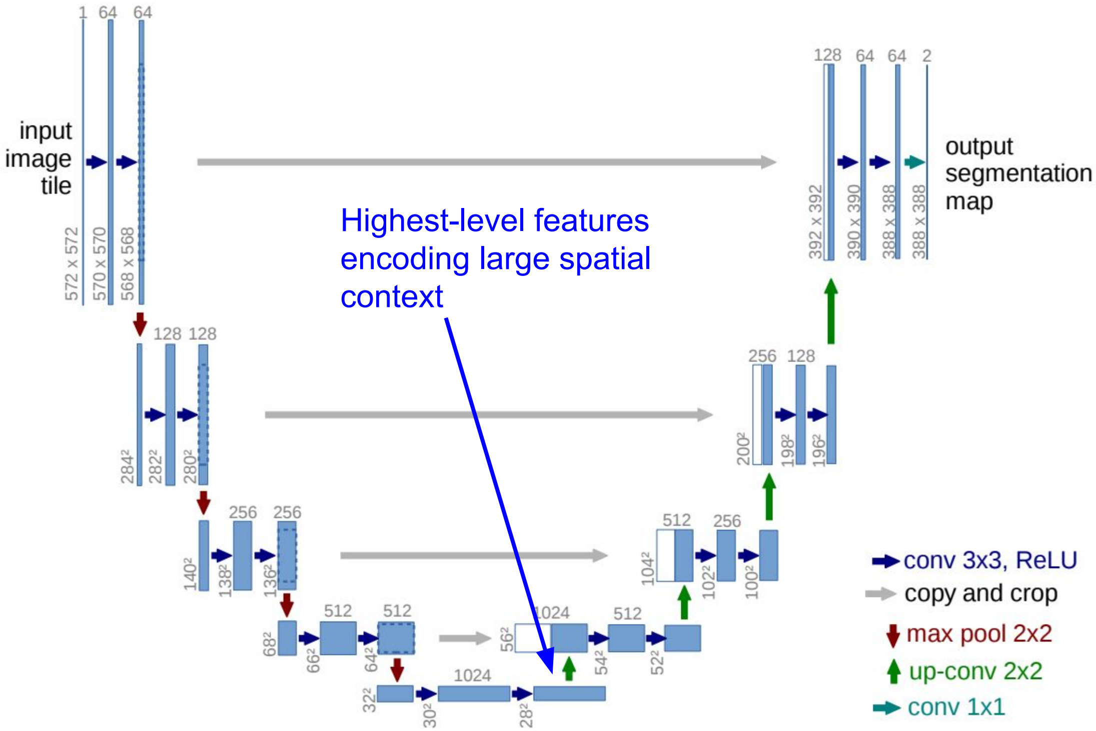
Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Semantic segmentation: U-Net



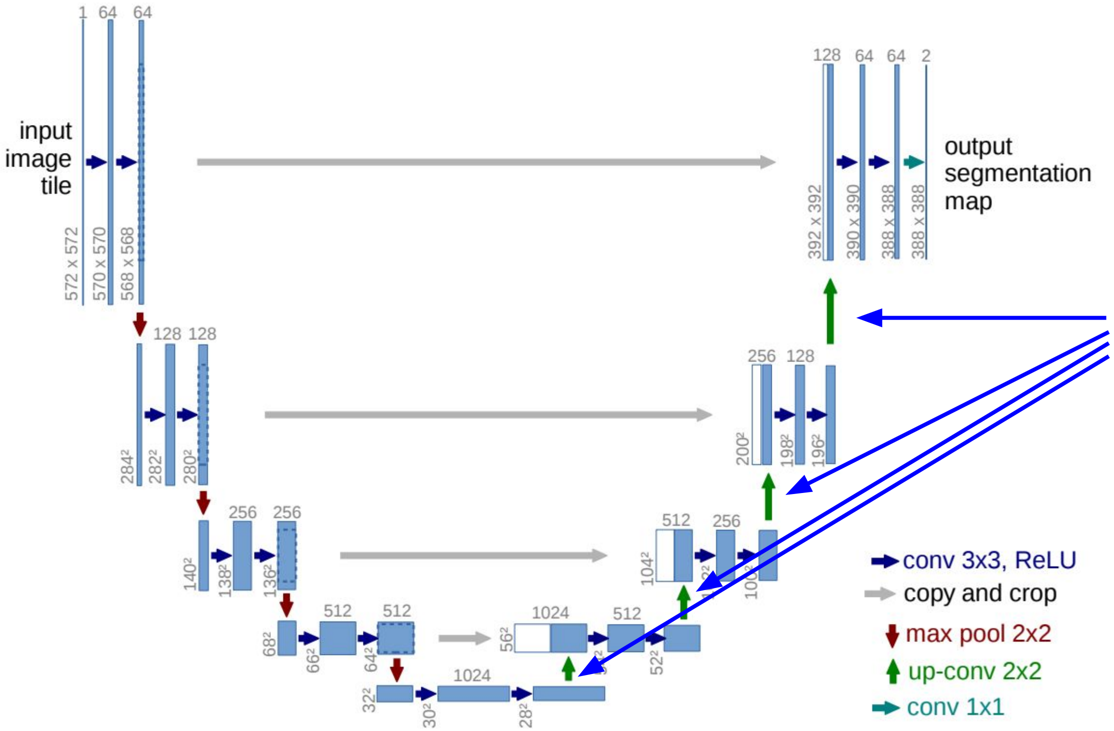
Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Semantic segmentation: U-Net



Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Semantic segmentation: U-Net

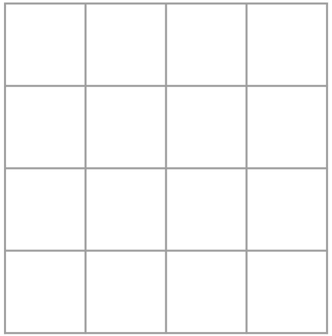


Up-convolutions to go from the global information encoded in highest-level features, back to individual pixel predictions

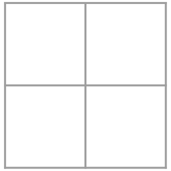
Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Up-convolutions

Recall: Normal 3 x 3 convolution, stride 2 pad 1



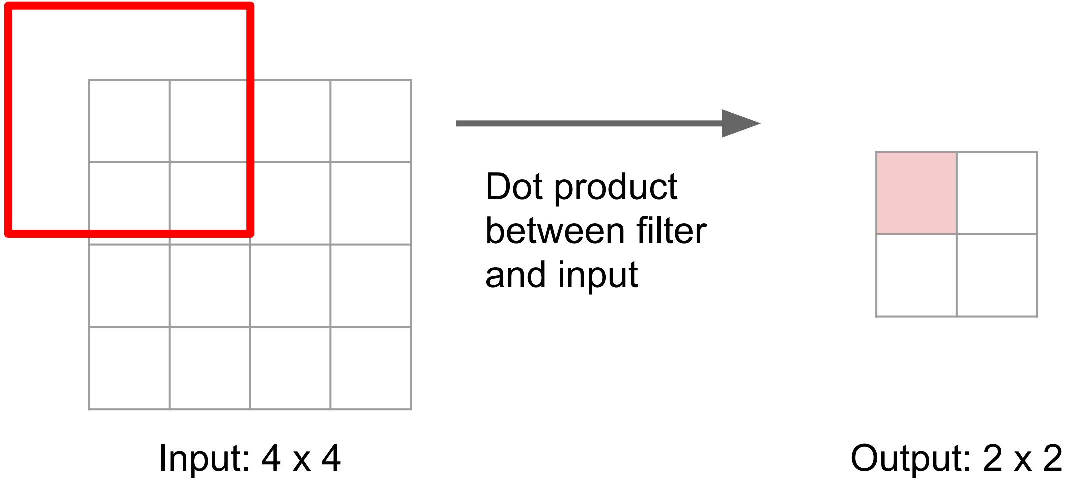
Input: 4 x 4



Output: 2 x 2

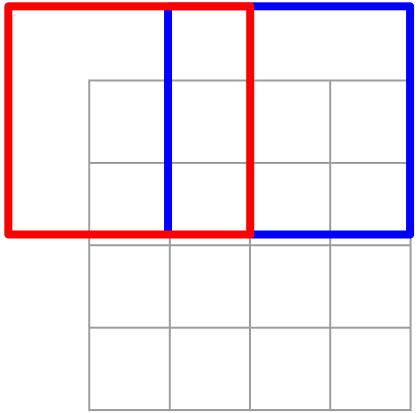
Up-convolutions

Recall: Normal 3 x 3 convolution, <u>stride 2</u> pad 1

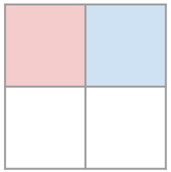


Up-convolutions

Recall: Normal 3 x 3 convolution, stride 2 pad 1



Element-wise multiplication between filter and input



Input: 4 x 4

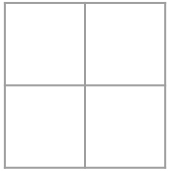
Output: 2 x 2

Filter moves 2 pixels in the input for every one pixel in the output

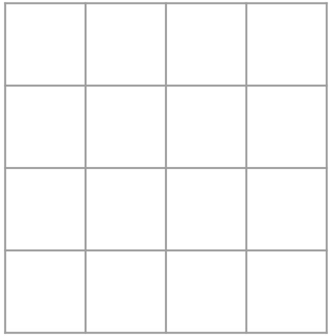
Stride gives ratio between movement in input and output

Up-convolutions

3 x 3 **transpose** convolution, stride 2 pad 1



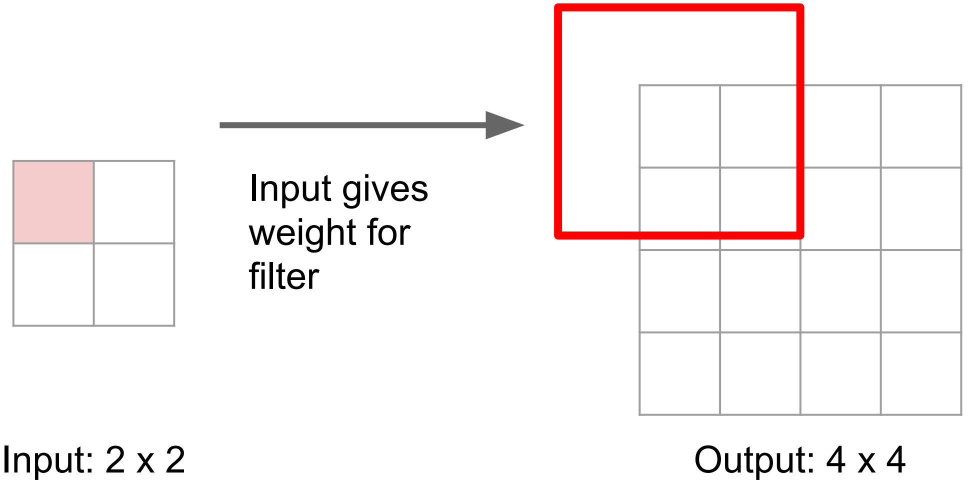
Input: 2 x 2



Output: 4 x 4

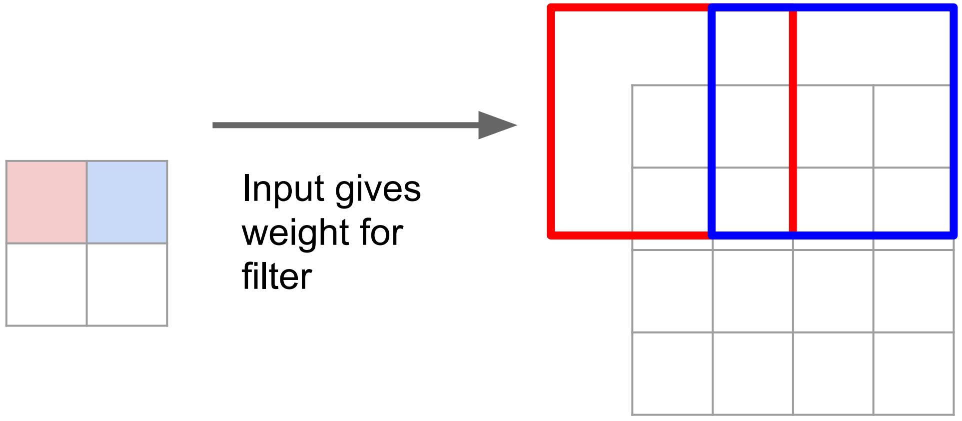
Up-convolutions

3 x 3 **up-convolution**, stride 2 pad 1



Up-convolutions

3 x 3 **up-convolution**, stride 2 pad 1



Input: 2 x 2

Output: 4 x 4

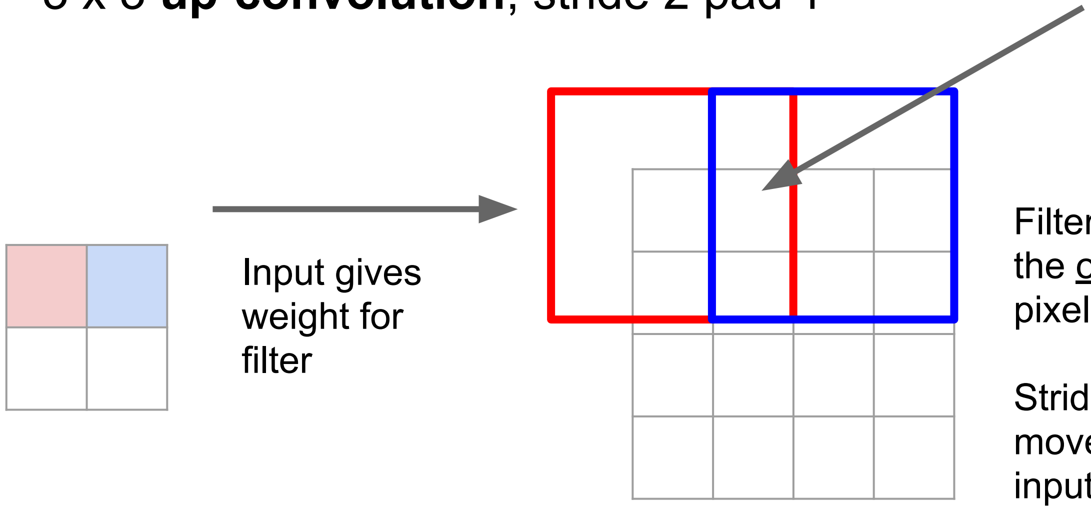
Filter moves 2 pixels in the output for every one pixel in the input

Stride gives ratio between movement in output and input

Up-convolutions

3 x 3 **up-convolution**, stride 2 pad 1

Sum where output overlaps



Filter moves 2 pixels in the output for every one pixel in the input

Stride gives ratio between movement in output and input

Input: 2 x 2

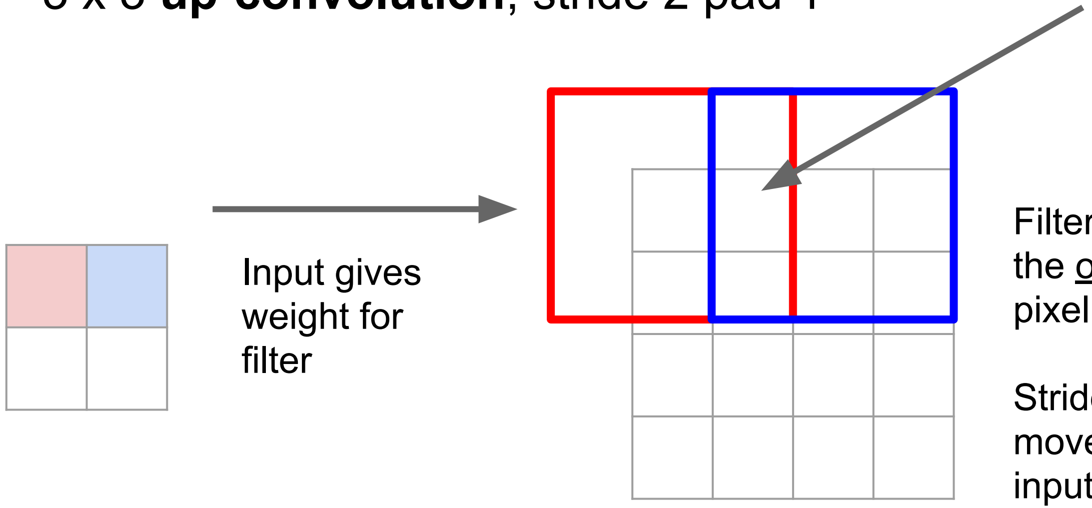
Output: 4 x 4

Up-convolutions

Other names:
-Transpose convolution
-Fractionally strided convolution
-Backward strided convolution

3 x 3 **up-convolution**, stride 2 pad 1

Sum where output overlaps



Input gives weight for filter

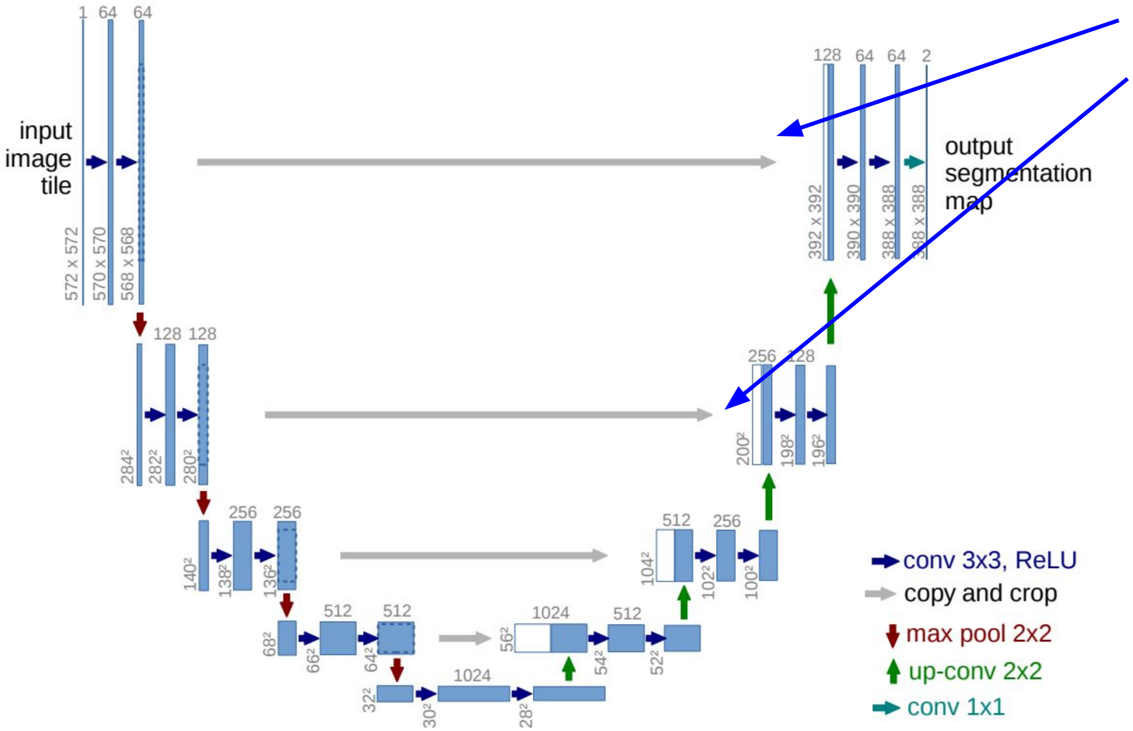
Input: 2 x 2

Output: 4 x 4

Filter moves 2 pixels in the output for every one pixel in the input

Stride gives ratio between movement in output and input

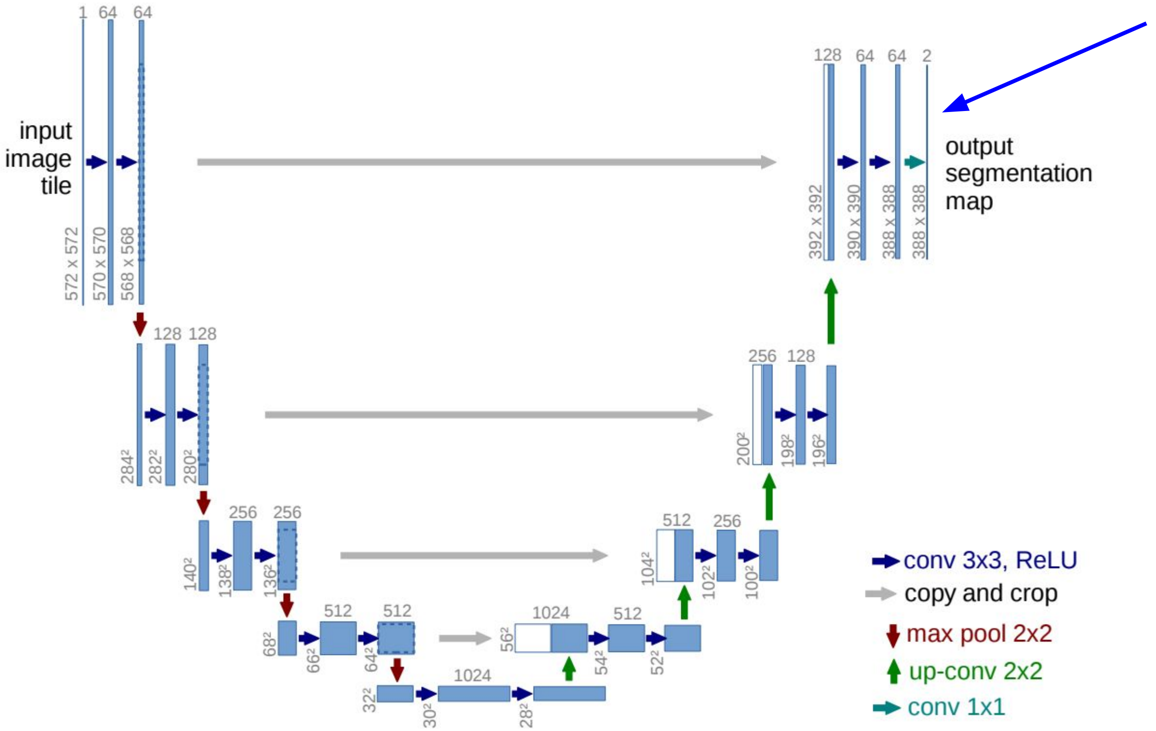
Semantic segmentation: U-Net



Concatenate with same-resolution feature map during downsampling process to combine high-level information with low-level (local) information

Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

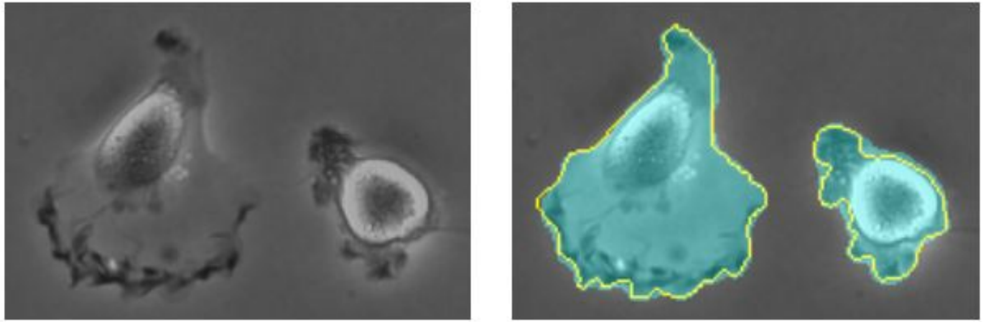
Semantic segmentation: U-Net



Train with classification loss (e.g. binary cross entropy) on every pixel, sum over all pixels to get total loss

Ronneberger et al. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Semantic segmentation: U-Net cell segmentation

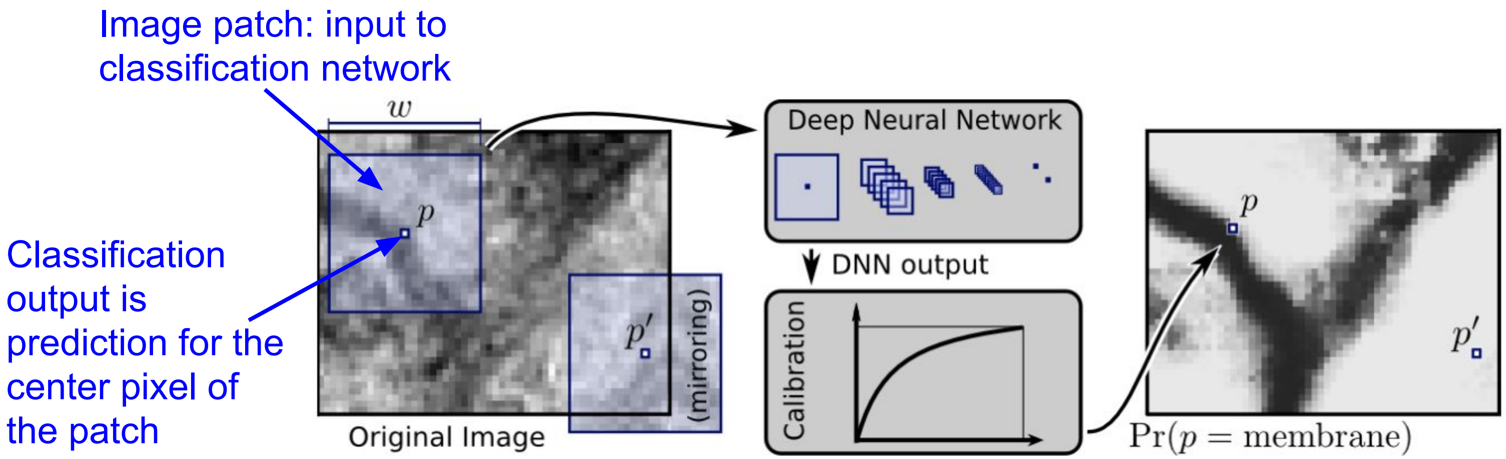


Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

Very small dataset: 30 training images of size 512x512, in the ISBI 2012 Electron Microscopy (EM) segmentation challenge. Used excessive data augmentation to compensate.

Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

Aside: segmentation through sliding-window pixel classification

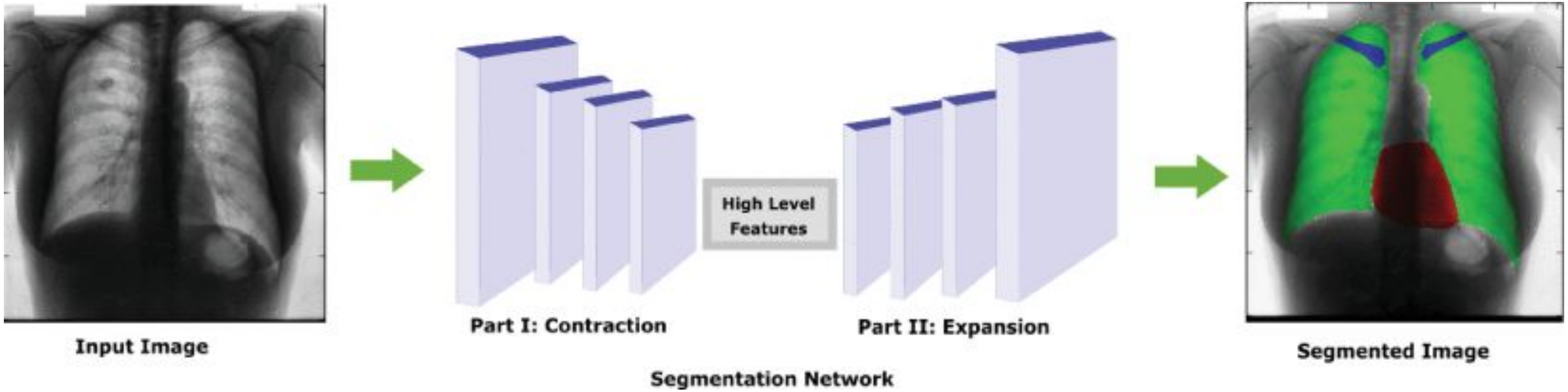


Note: a simple approach to segmentation can also be applying a classification CNN on image patches in a dense, sliding-window fashion (e.g. Ciresan et al.). But fully convolutional approaches such as U-Net generally achieve better performance.

Ciresan et al. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. NeurIPS, 2012.

Novikov et al. 2018

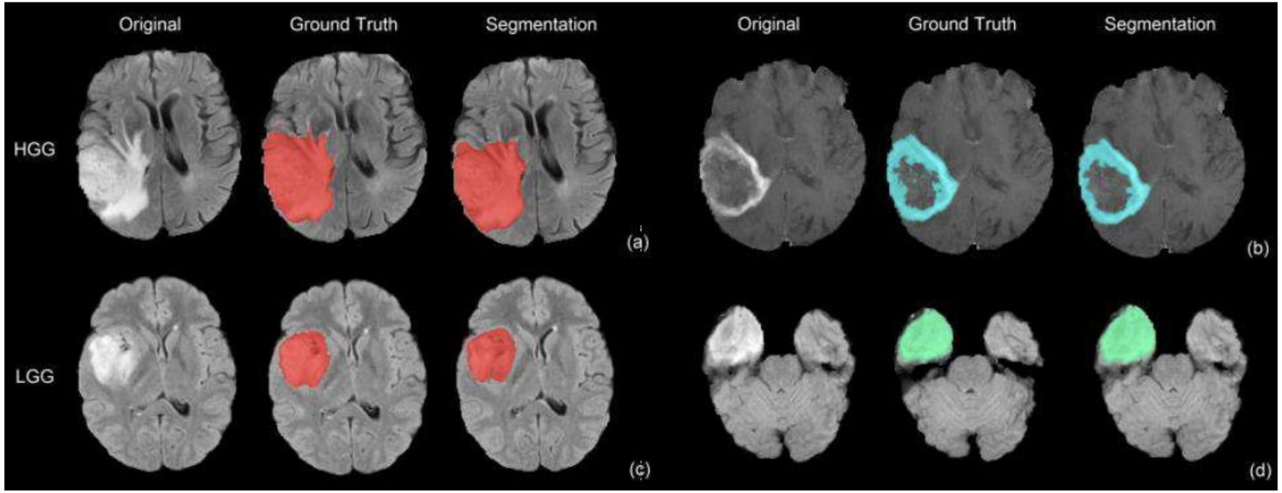
- Chest x-ray segmentation of lungs, clavicles, and heart
- JSRT dataset of 247 chest-xrays at 2048x2048 resolution. (But downsampled to 128x128 and 256x256!)
- Used a U-Net based segmentation network with a few modifications



Novikov et al. Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs. IEEE Trans. on Medical Imaging, 2018.

Dong et al. 2017

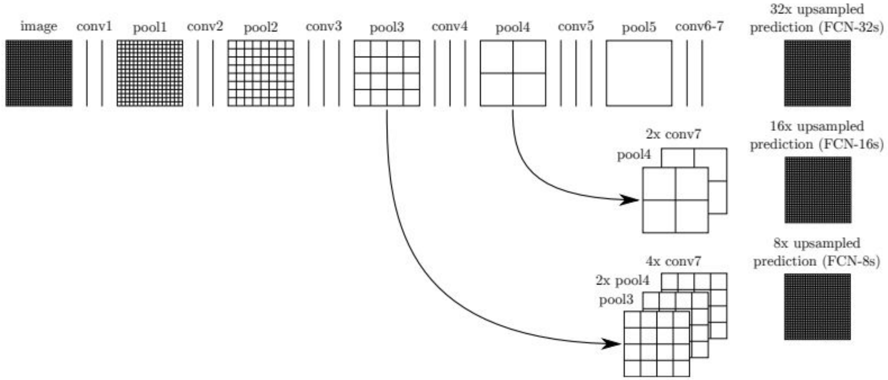
- Segmentation of tumors in brain MR image slices
- BRATS 2015 dataset: 220 high-grade brain tumor and 54 low-grade brain tumor MRIs
- U-Net architecture, Dice loss function



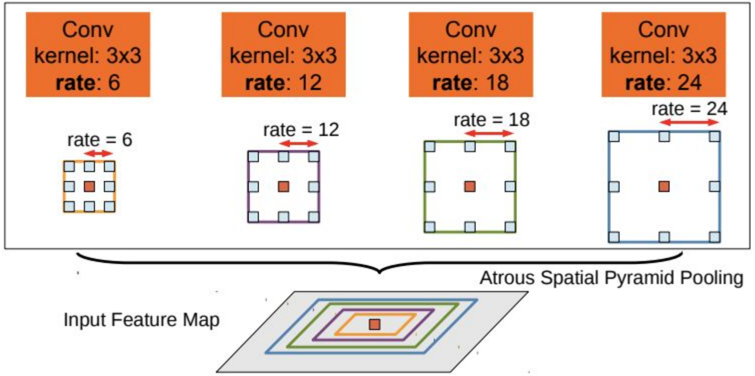
Dong et al. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. MIUA, 2017.

Other segmentation architectures

- **Fully convolutional networks (FCN)**
- Pre-cursor to U-Net, similar in structure but simpler upsampling pathway



- **DeepLab (v1-v3)**
- Uses “atrous convolutions” to control a filter’s field of view
- Parallel atrous convolutions with different rates for multi-scale features



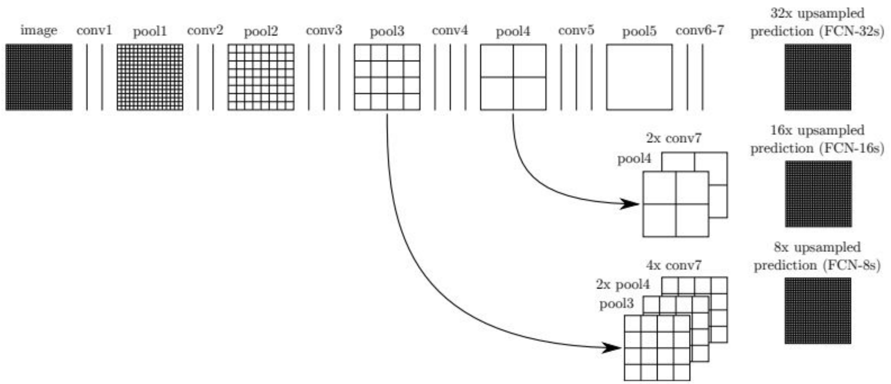
Shelhamer*, Long*, et al. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015.

Chen et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE TPAMI, 2017.
Chen et al. Rethinking Atrous Convolution for Semantic Image Segmentation. 2917.

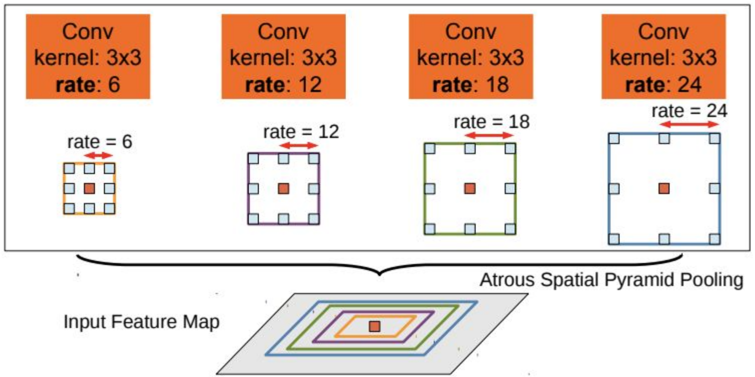
Other segmentation architectures

Can try DeepLab v3+ for segmentation projects!

- **Fully convolutional networks (FCN)**
- Pre-cursor to U-Net, similar in structure but simpler upsampling pathway



- **DeepLab (v1-v3+)**
- Uses "atrous convolutions" to control a filter's field of view
- Parallel atrous convolutions with different rates for multi-scale features

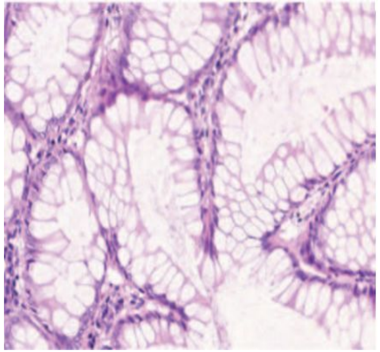


Shelhamer*, Long*, et al. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015.

Chen et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE TPAMI, 2017.
Chen et al. Rethinking Atrous Convolution for Semantic Image Segmentation. 2917.

Richer visual recognition tasks: segmentation and detection

Classification

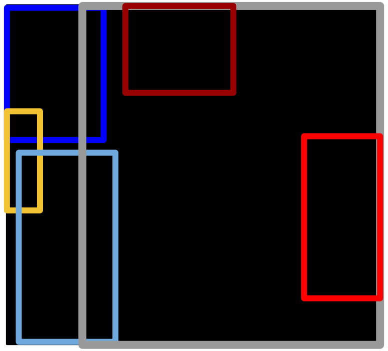


Output:
one category label for image (e.g., colorectal glands)

Semantic Segmentation

Output:
category label for each pixel in the image

Detection



Output:
Spatial bounding box for each **instance** of a category object in the image

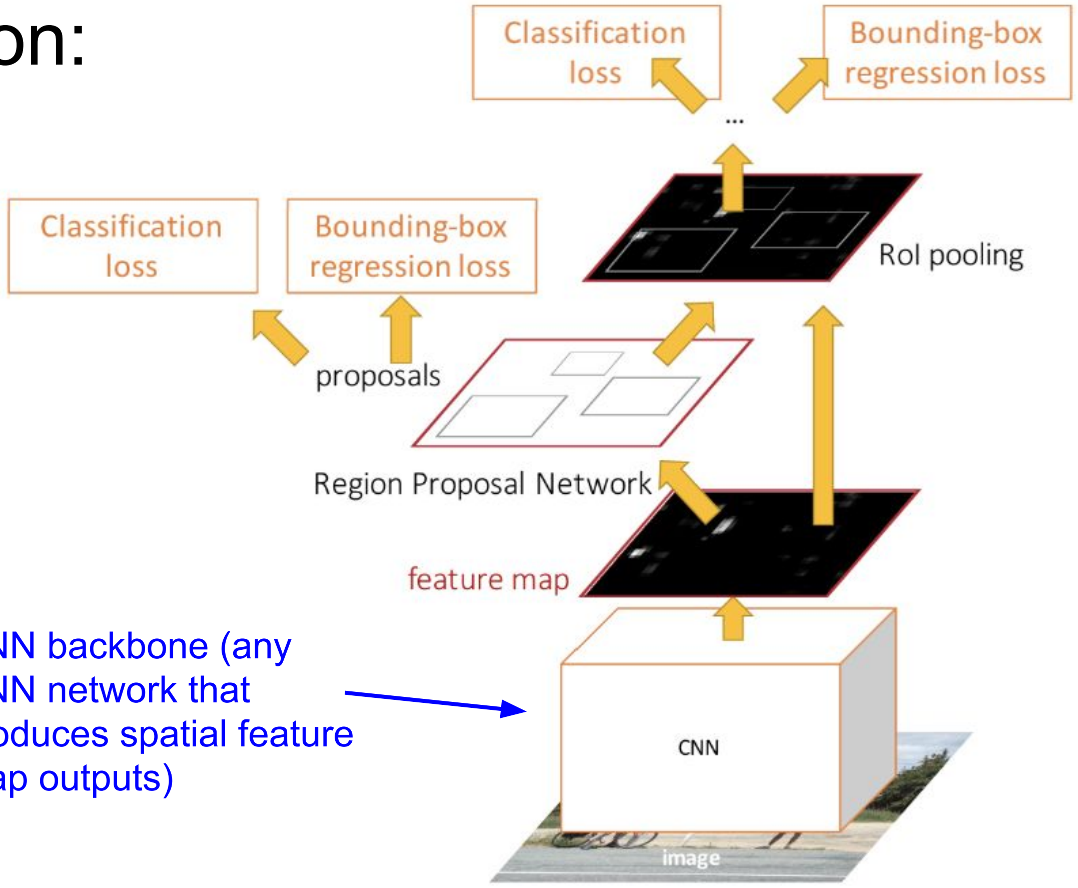
Instance Segmentation

Output:
Category label and instance label for each pixel in the image

Distinguishes between different instances of an object

Figures: Chen et al. 2016. https://arxiv.org/pdf/1604.02677.pdf

Object detection: Faster R-CNN

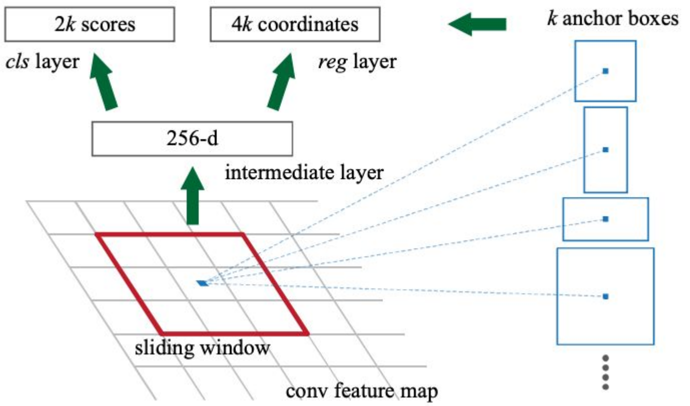


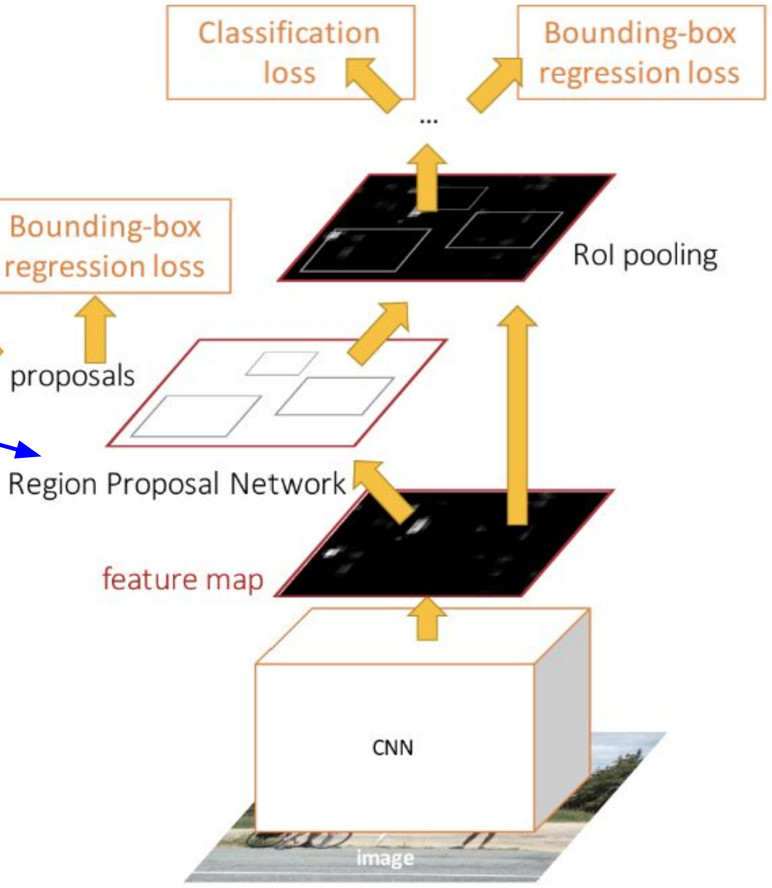
CNN backbone (any CNN network that produces spatial feature map outputs)

Ren et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2015.

Object detection: Faster R-CNN

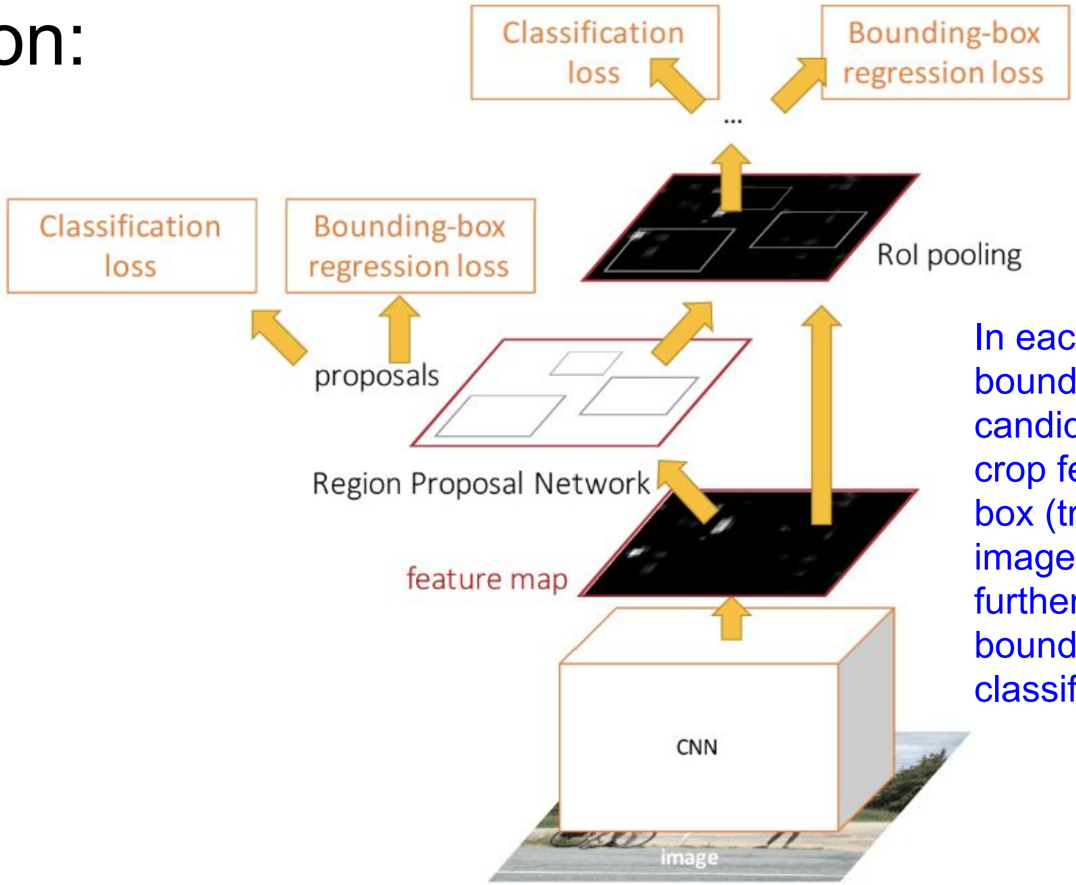
Regress to bounding box “candidates” from “anchor boxes” at each location





Ren et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2015.

Object detection: Faster R-CNN

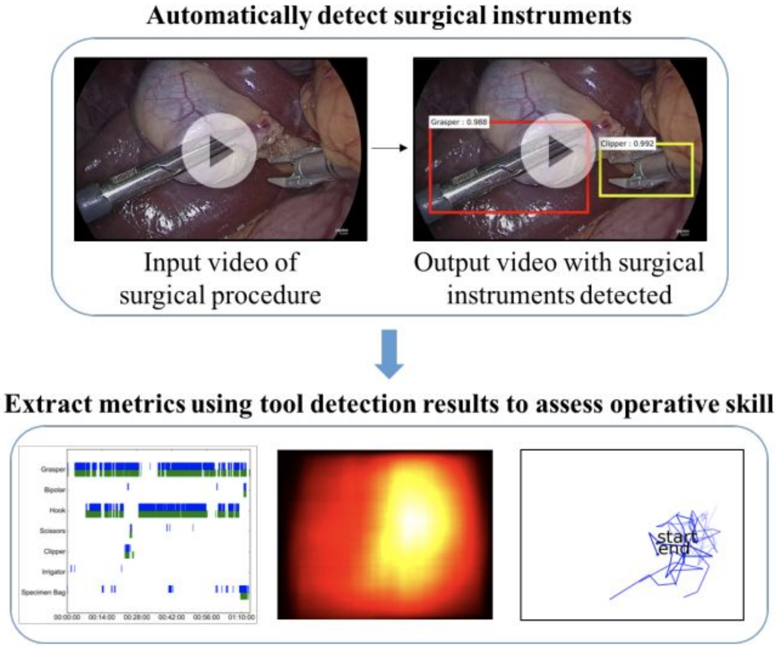


In each of top bounding box candidate locations, crop features within box (treat as own image) and perform further refinement of bounding box + classification

Ren et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2015.

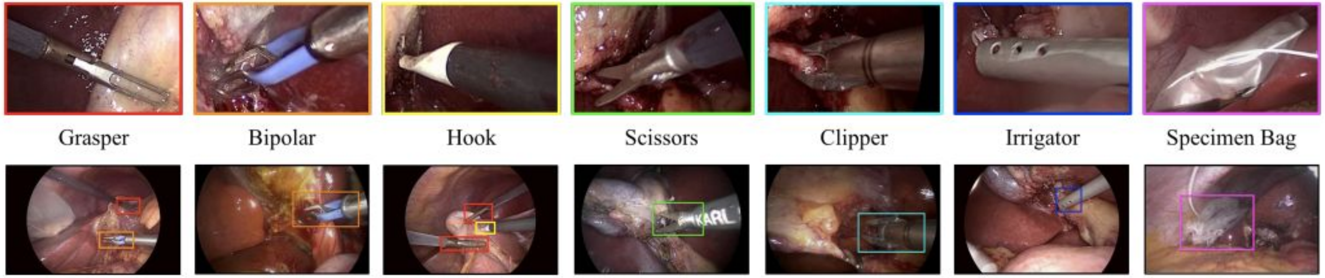
Jin et al. 2018

- Detection of surgical instruments in surgery videos (in each video frame)
- Surgical instrument movement over the course of a video can be used to extract metrics such as tool switching, and spatial trajectories, that can be used to assess and provide feedback on operative skill.
- Used M2cai16-tool dataset of 15 surgical videos. Annotated 2532 frames with bounding boxes of 7 tools.



Jin et al. Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks. WACV, 2018.

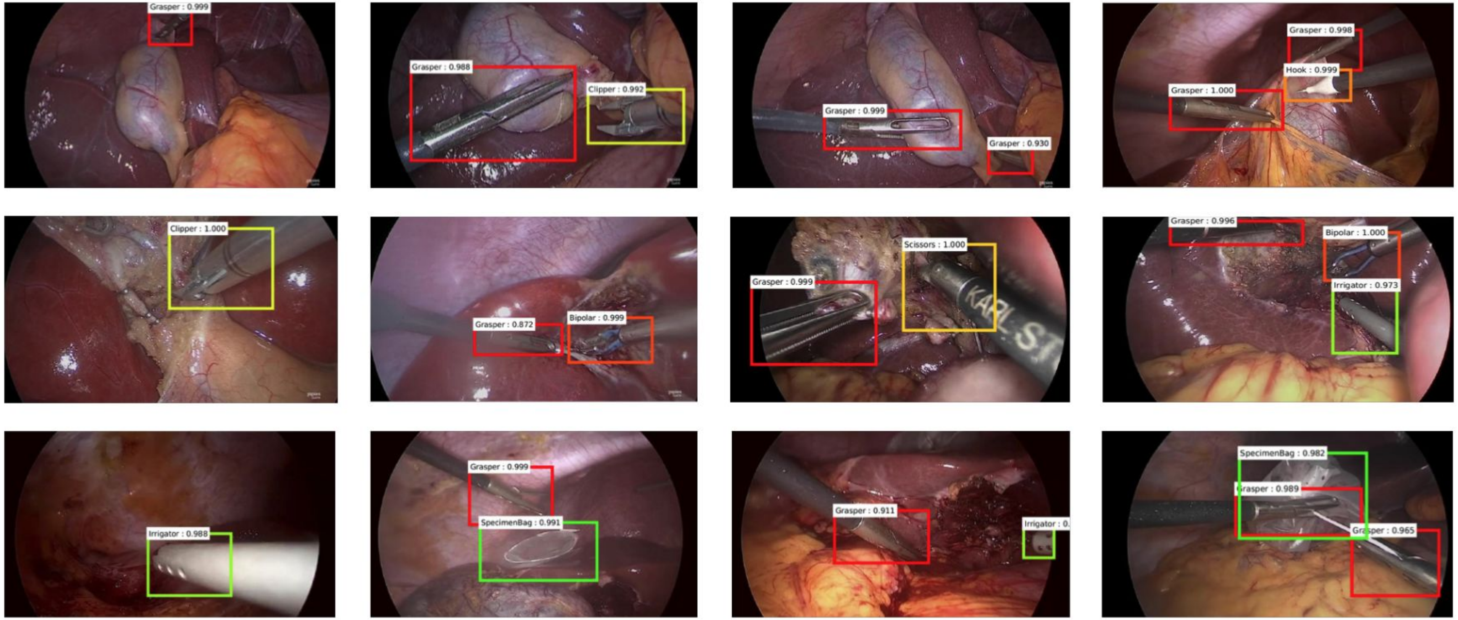
Jin et al. 2018



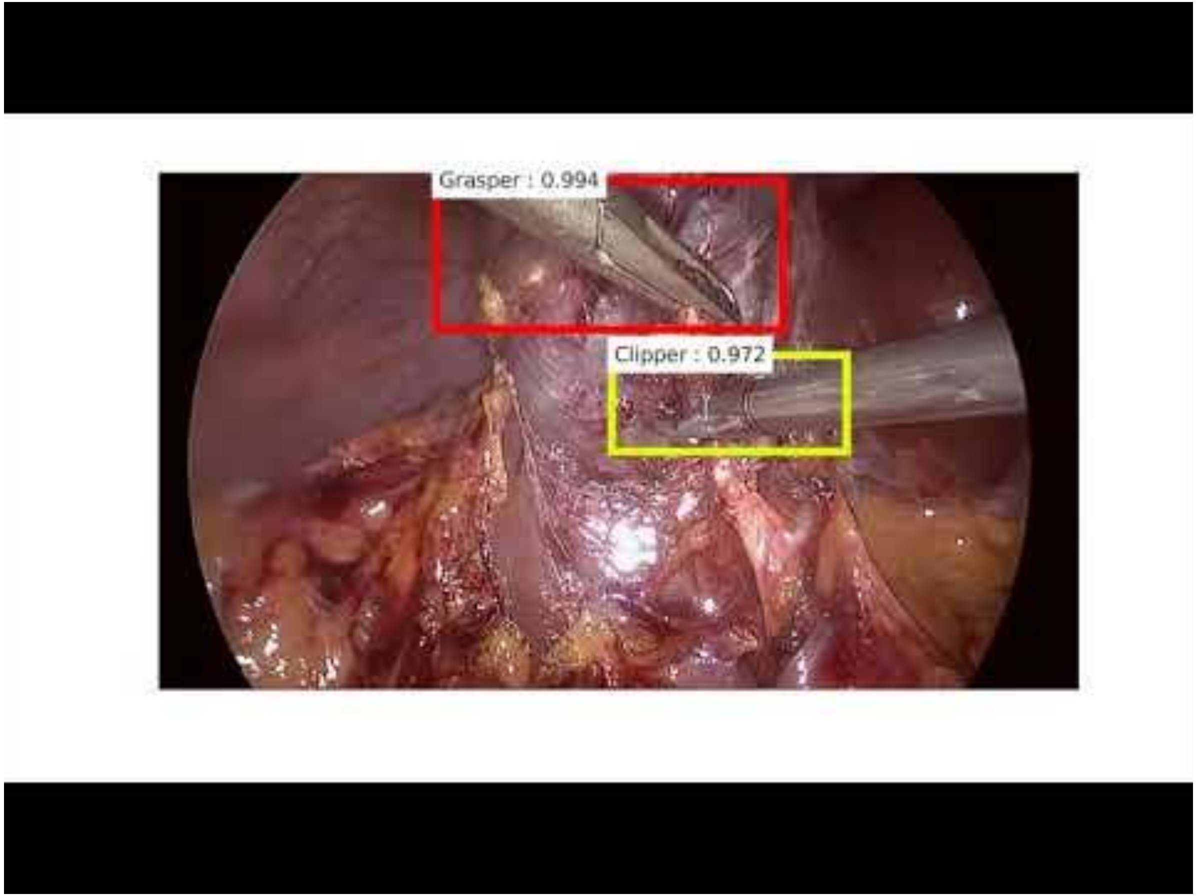
Tool	AP
Grasper	48.3
Bipolar	67.0
Hook	78.4
Scissors	67.7
Clipper	86.3
Irrigator	17.5
Specimen Bag	76.3
mAP	63.1

Jin et al. Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks. WACV, 2018.

Jin et al. 2018



Jin et al. Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks. WACV, 2018.



Jin et al. Tool Detection and Operative Skill Assessment in Surgical Videos Using Region-Based Convolutional Neural Networks. WACV, 2018.

Other object detection architectures

- RCNN, Fast RCNN: older and slower predecessors to Faster-RCNN
- YOLO, SSD: single-stage detectors that change region proposal generation -> region classification two-stage pipeline into a single stage.
    - Faster, but lower performance. Struggles more with class imbalance relative to two-stage networks that filter only top object candidate boxes for the second stage.
- RetinaNet: single-stage detector that uses a “focal loss” to adaptively weight harder examples over easy background examples. Able to outperform Faster R-CNN on some benchmark tasks, while being more efficient.

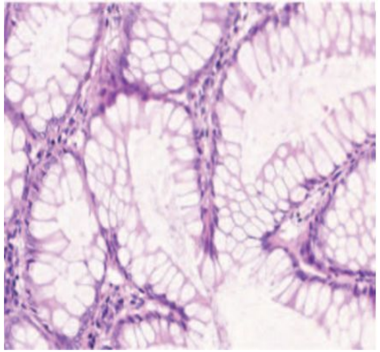
Other object detection architectures

- RCNN, Fast RCNN: older and slower predecessors to Faster-RCNN
- YOLO, SSD: single-stage detectors that change region proposal generation -> region classification two-stage pipeline into a single stage.
    - Faster, but lower performance. Struggles more with class imbalance relative to two-stage networks that filter only top object candidate boxes for the second stage.
- RetinaNet: single-stage detector that uses a “focal loss” to adaptively weight harder examples over easy background examples. Able to outperform Faster R-CNN on some benchmark tasks, while being more efficient.

RetinaNet also worth trying
for object detection projects!

Richer visual recognition tasks: segmentation and detection

Classification

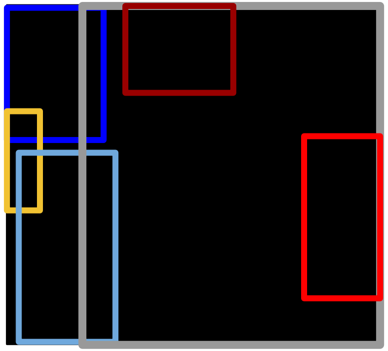


Output:
one category label for image (e.g., colorectal glands)

Semantic Segmentation

Output:
category label for each pixel in the image

Detection



Output:
Spatial bounding box for each **instance** of a category object in the image

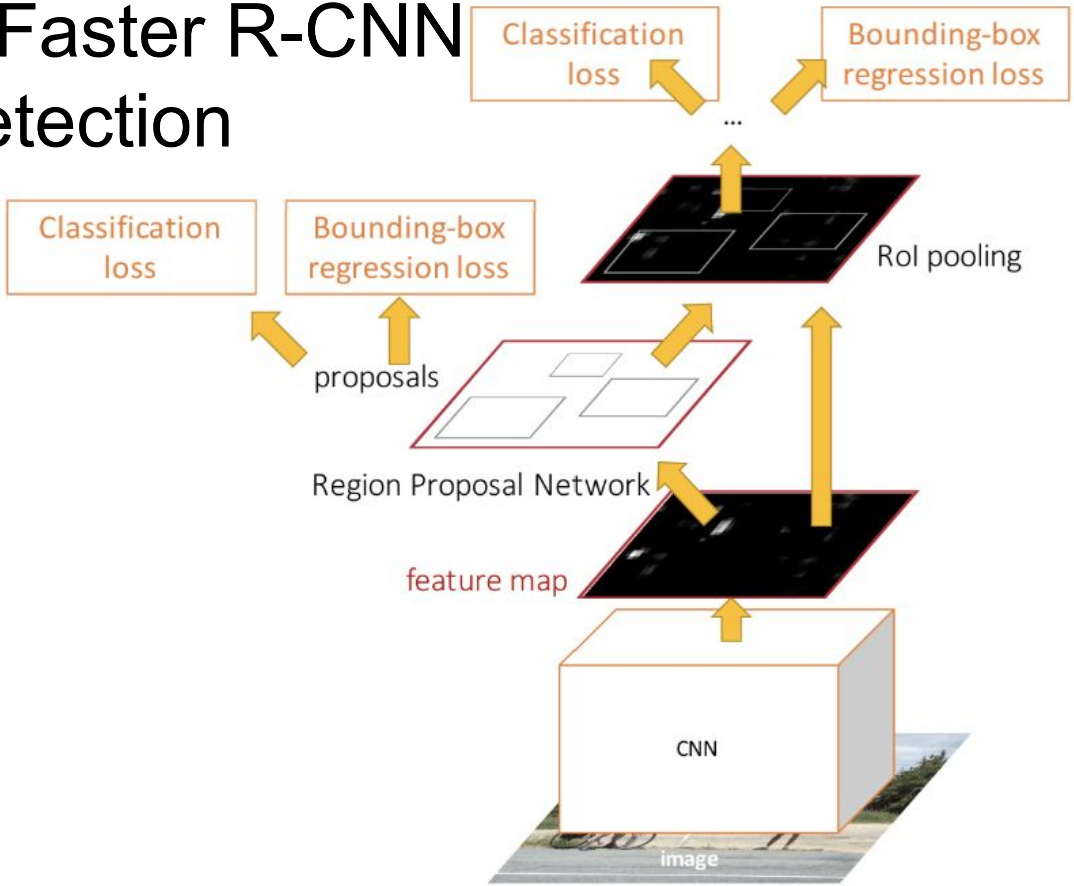
Instance Segmentation

Output:
Category label and instance label for each pixel in the image

Distinguishes between different instances of an object

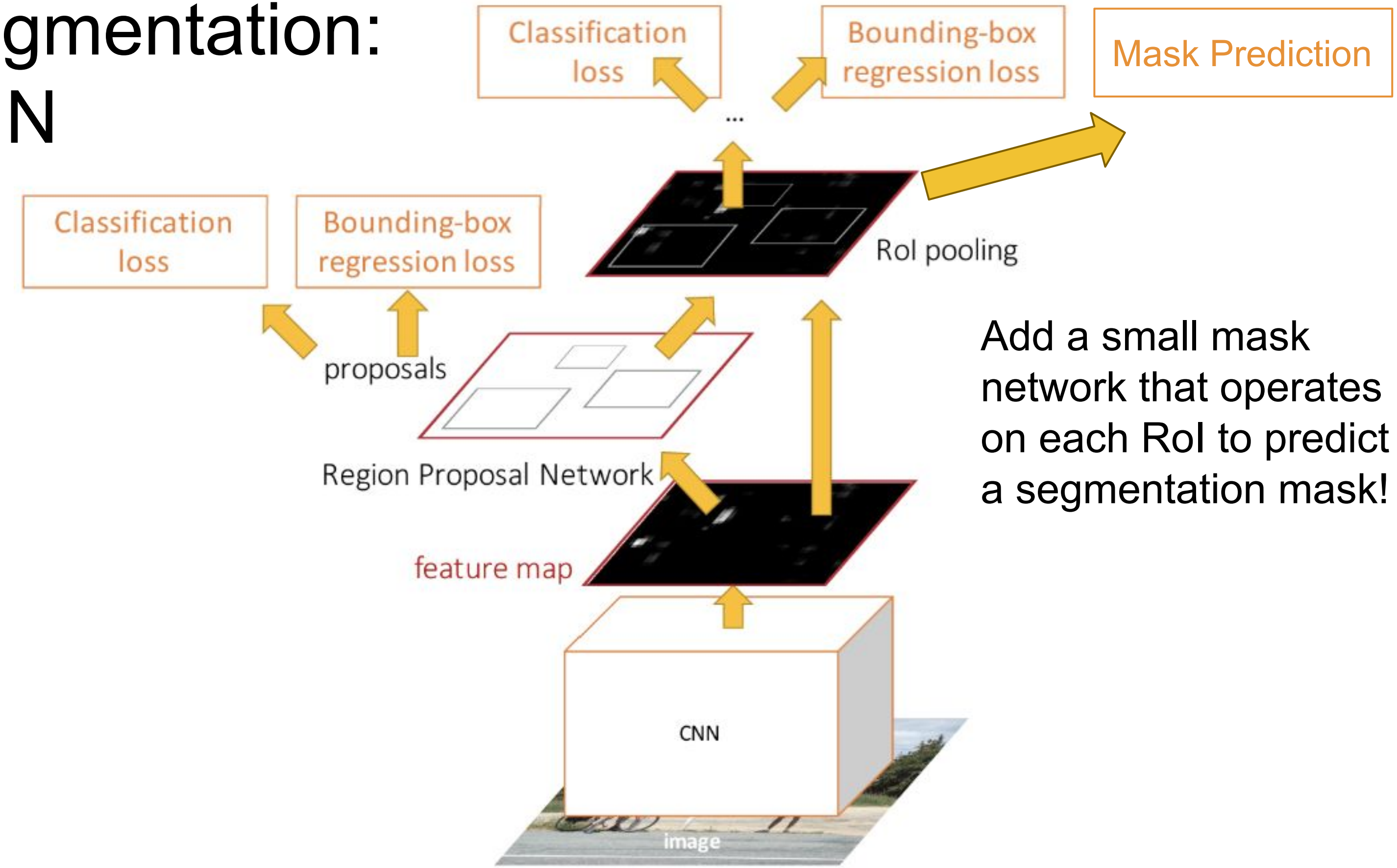
Figures: Chen et al. 2016. https://arxiv.org/pdf/1604.02677.pdf

Remember Faster R-CNN for object detection



Ren et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2015.

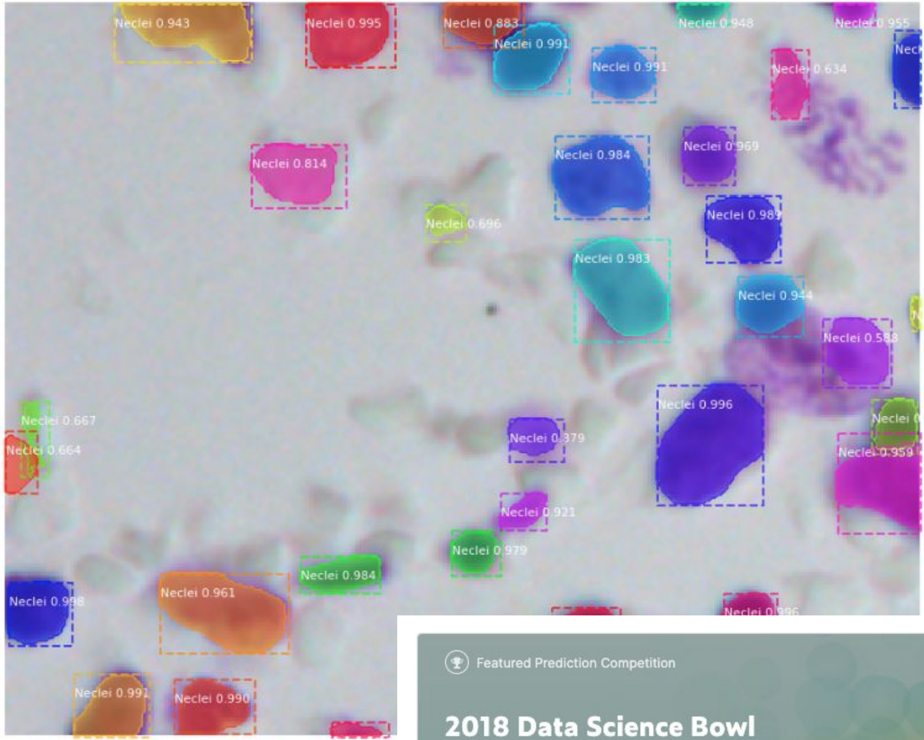
Instance segmentation: Mask R-CNN



Add a small mask network that operates on each RoI to predict a segmentation mask!

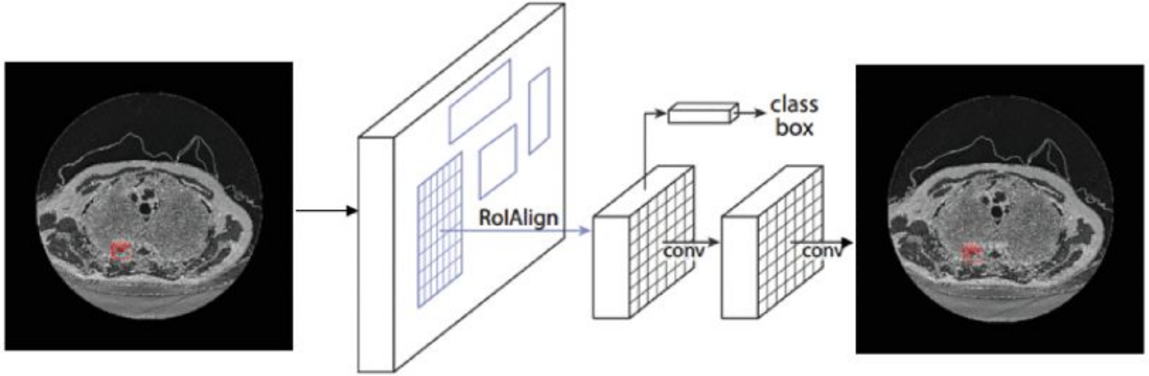
He et al. Mask R-CNN, 2017.

Example: instance segmentation of cell nuclei



Lung nodule segmentation

- E.g. Liu et al. 2018
  - Dataset: Lung Nodule Analysis (LUNA) challenge, 888 512x512 CT scans from the Lung Image Data Consortium database (LIDC-IDRI).
  - Performed 2D instance segmentation in 2D CT slices



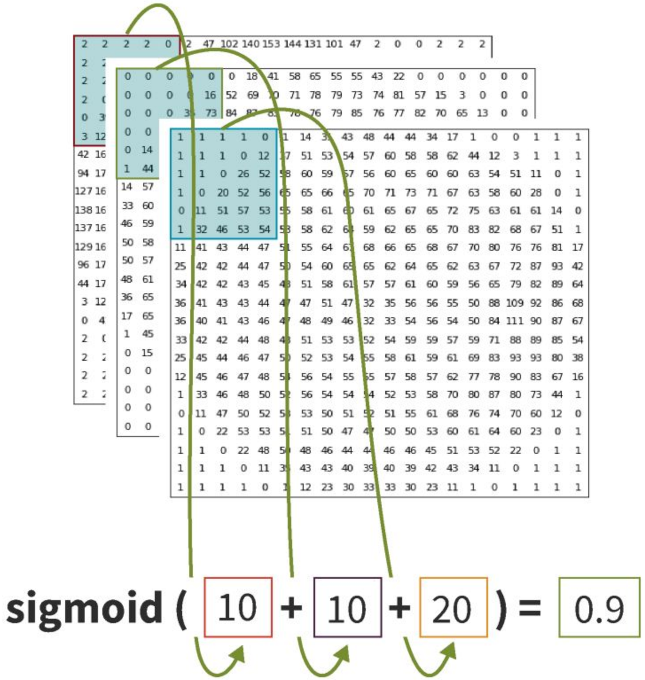
We will see other ways to handle 3D medical data types in the next lecture

Liu et al. Segmentation of Lung Nodule in CT Images Based on Mask R-CNN. 2018.

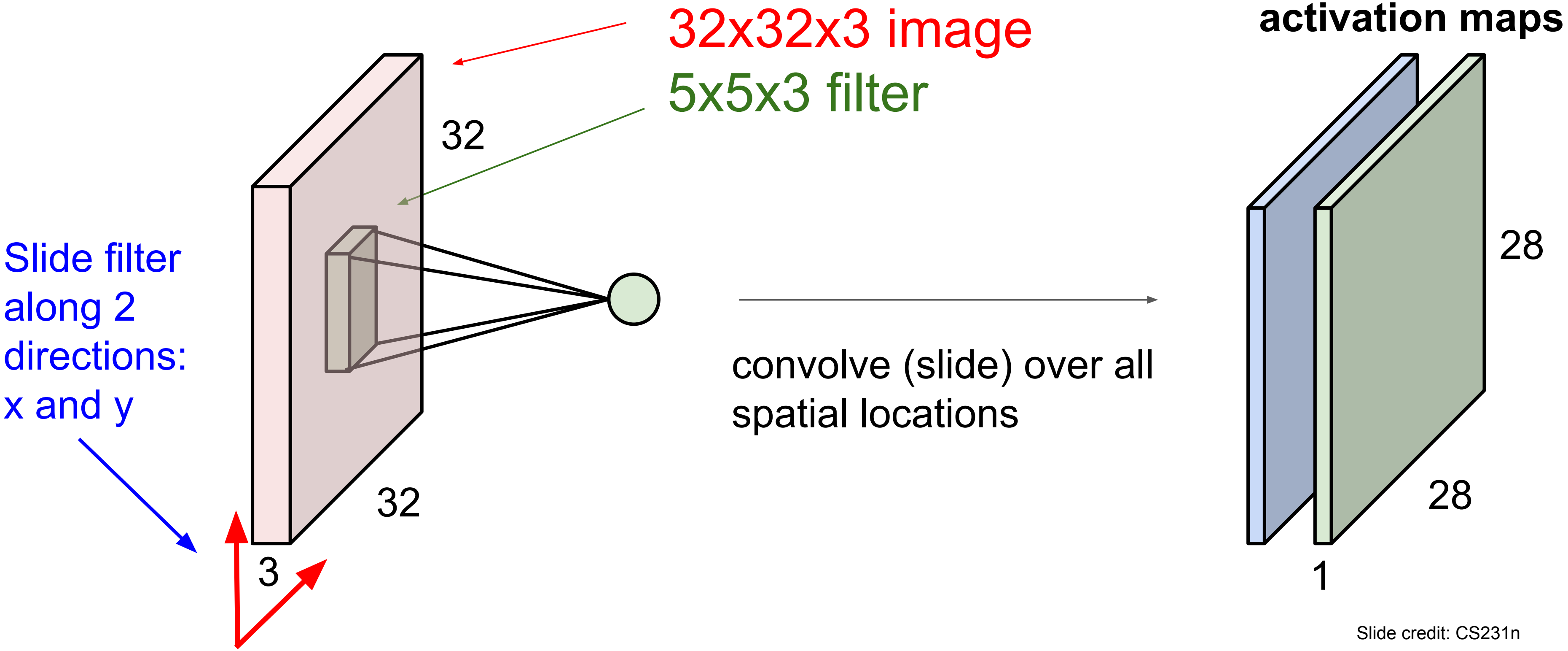
How do we handle 3D data?

Remember 2D convolutions

IDEA: ANALYZE THE IMAGE BY REGION



Remember 2D convolutions



Slide credit: CS231n

3D convolutions

Slide filter along **3** directions: x, y, and z!

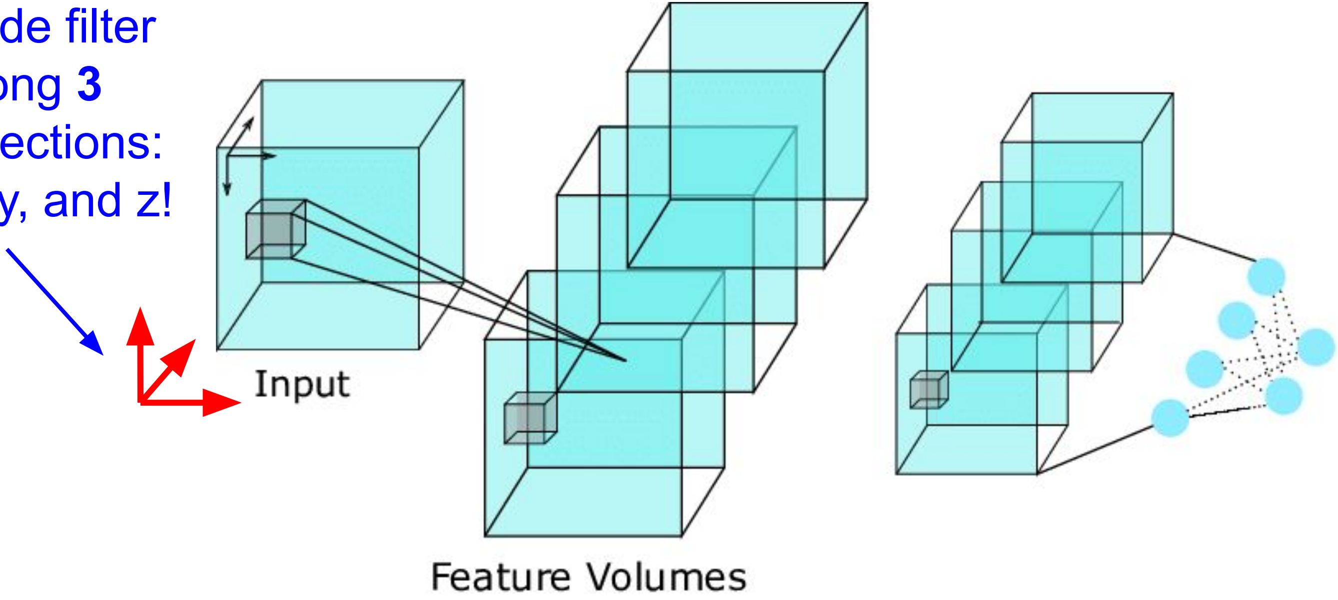
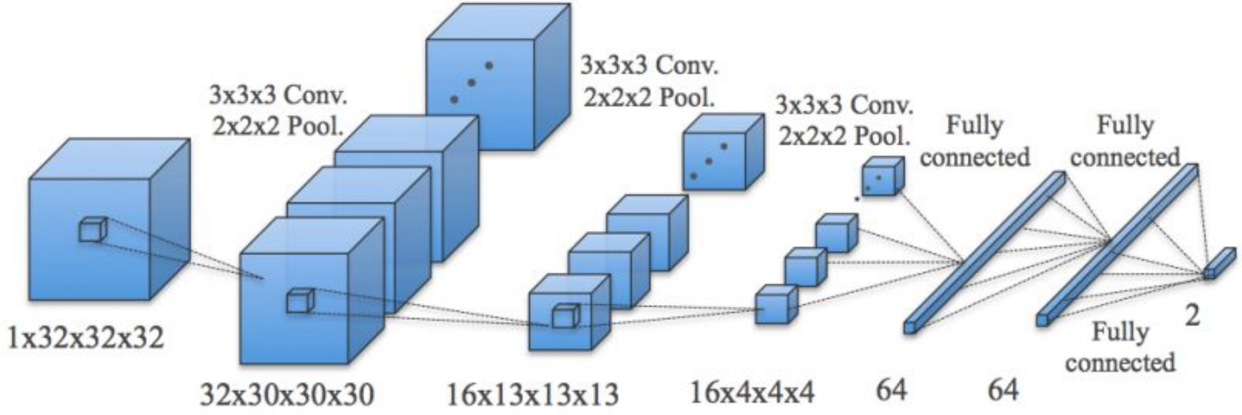


Figure credit:
https://www.researchgate.net/profile/Deepak_Mishra19/publication/330912338/figure/fig1/AS:723363244810254@1549474645742/Basic-3D-CNN-architecture-the-3D-filter-is-convolved-with-the-video-in-three-dimensions.png

Huang et al. 2017

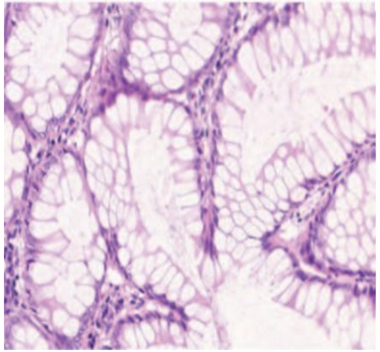
- Simple 3D CNN for lung nodule classification
- Used image processing approaches to extract candidate nodules, then 3D CNN to classify the surrounding volume
- Used the Lung Image Database Consortium (LIDC) Dataset, with 99 3D CT scans



Huang et al. Lung Nodule Detection in CT Using 3D Convolutional Neural Networks. ISBI 2017.

For richer visual recognition tasks, can also extend respective CNN architectures to use 3D convolutions

Classification

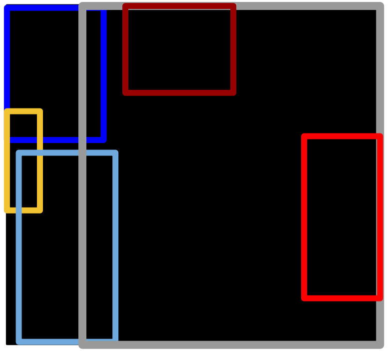


Output:
one category label for image (e.g., colorectal glands)

Semantic Segmentation

Output:
category label for each pixel in the image

Detection



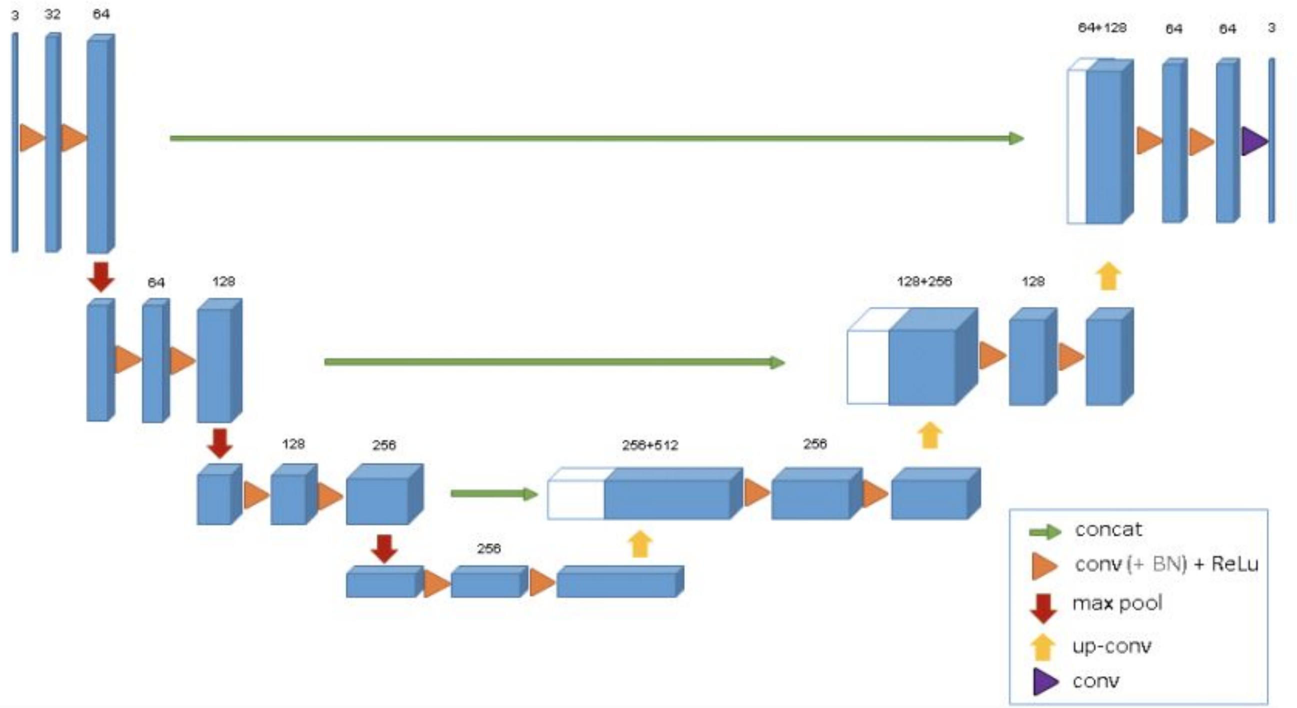
Output:
Spatial bounding box for each **instance** of a category object in the image

Instance Segmentation

Output:
Category label and instance label for each pixel in the image

Figures: Chen et al. 2016. https://arxiv.org/pdf/1604.02677.pdf

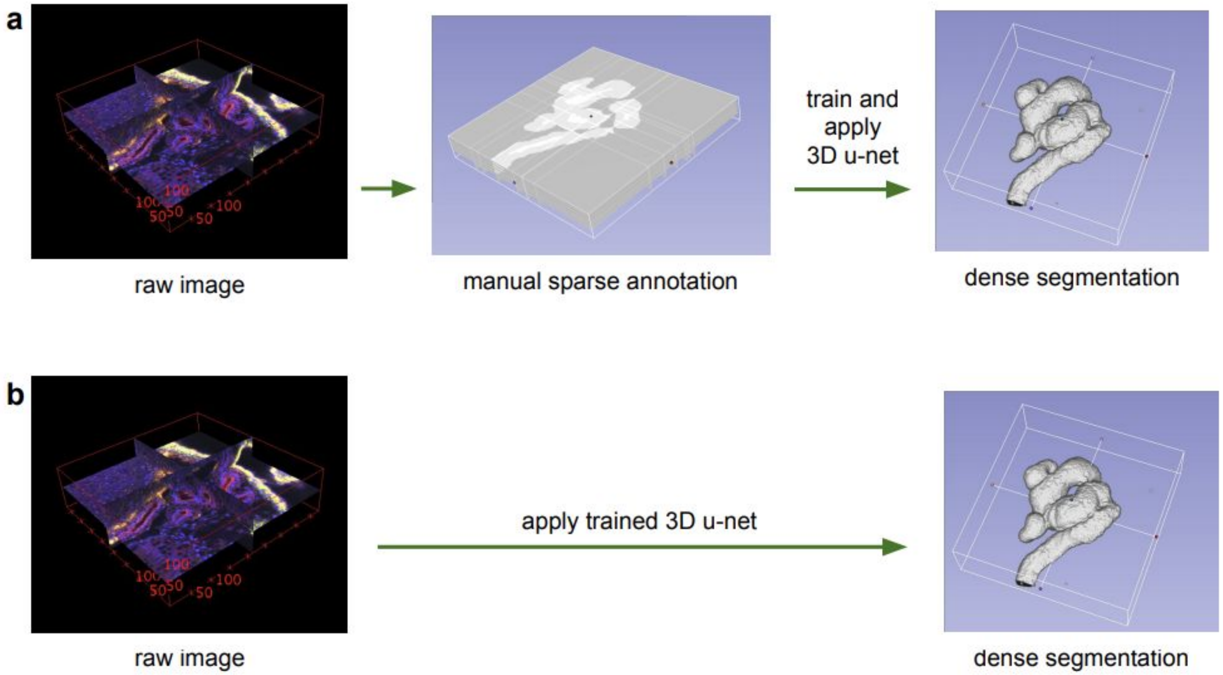
E.g. 3D U-Net



Cicek et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. MICCAI 2016.

E.g. 3D U-Net

Ex: 3D segmentation of Xenopus kidney in confocal microscopic data

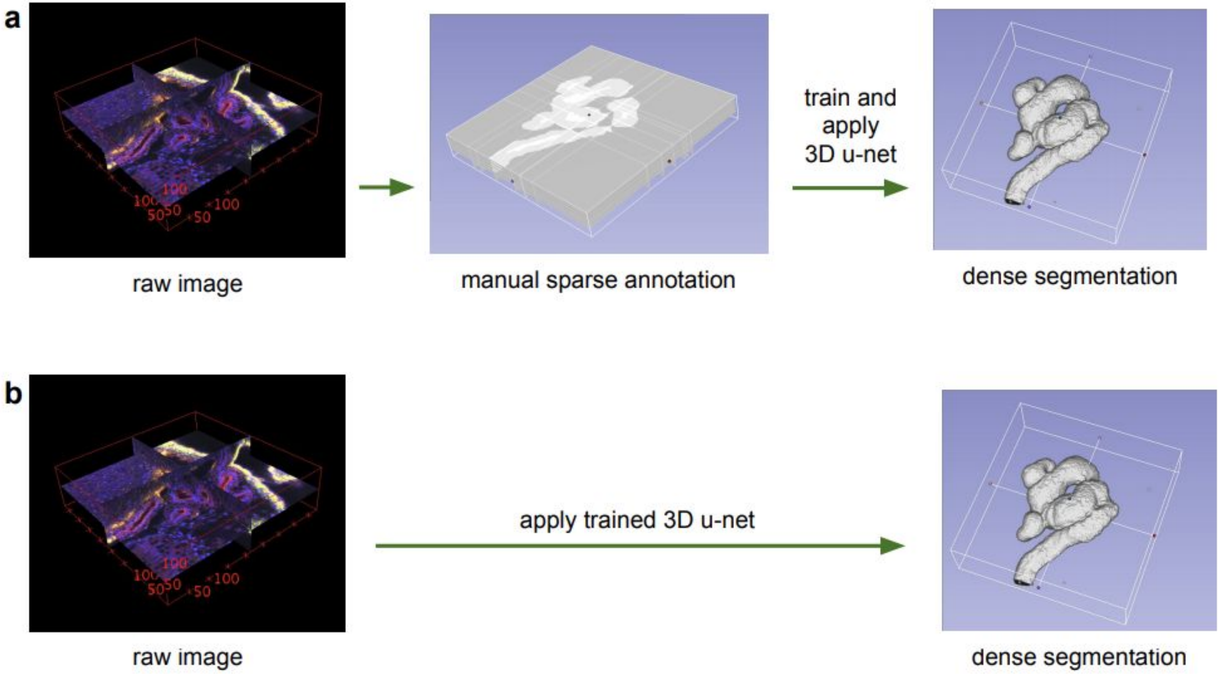


Cicek et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. MICCAI 2016.

E.g. 3D U-Net

Ex: 3D segmentation of Xenopus kidney in confocal microscopic data

Spatial dims: ~ 250 x 250 x 60.
3 channels: each channel corresponds to a different type of data capture



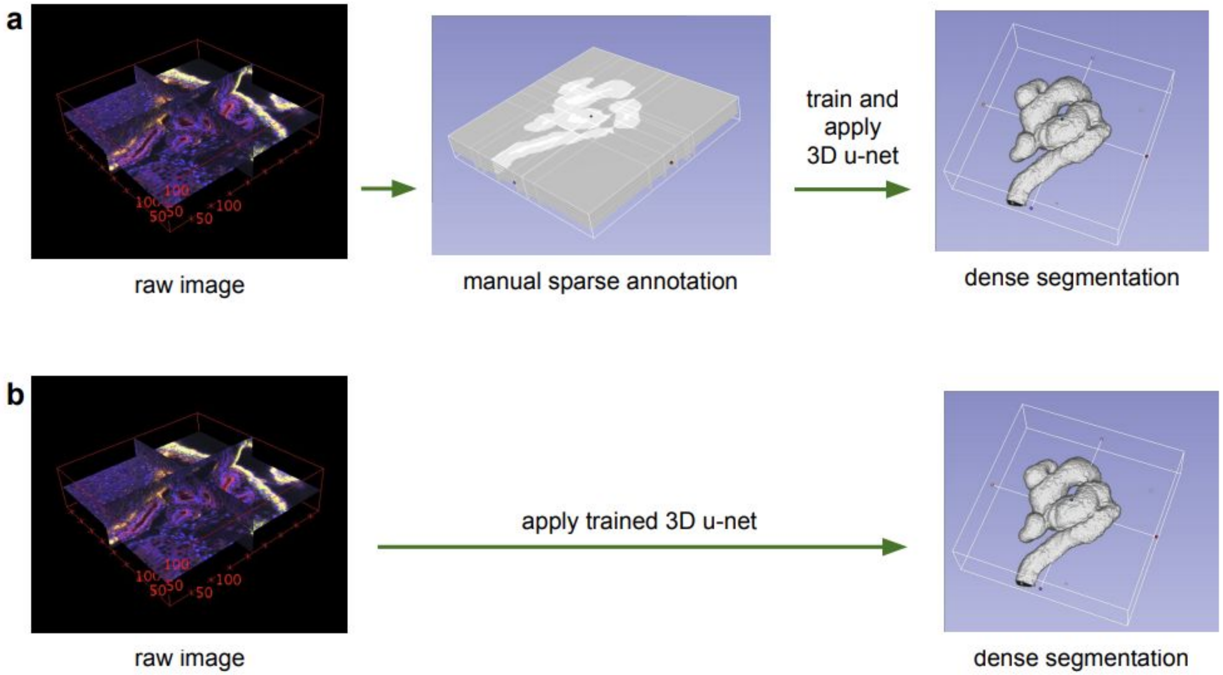
Cicek et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. MICCAI 2016.

E.g. 3D U-Net

Ex: 3D segmentation of Xenopus kidney in confocal microscopic data

Spatial dims: ~ 250 x 250 x 60.
3 channels: each channel corresponds to a different type of data capture

Used only 3 samples total! (with total of 77 annotated 2D slices). Leverages fact that each sample contains many instances of same repetitive structures w/ variation.



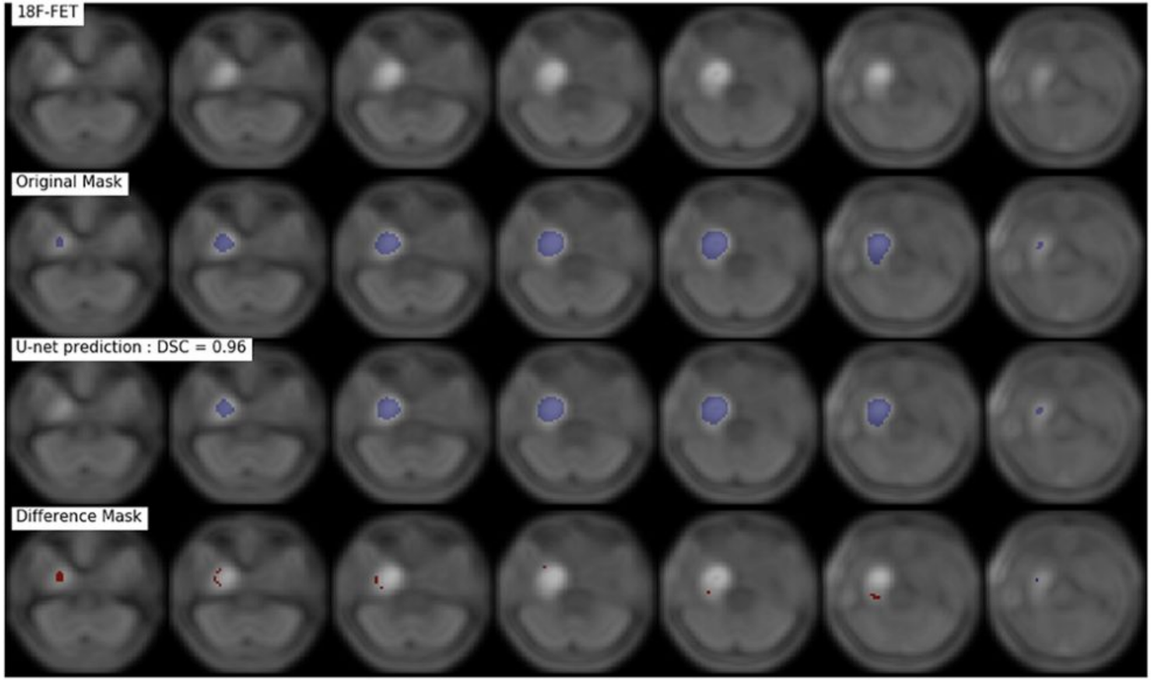
Cicek et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. MICCAI 2016.

Ex: Brain lesion segmentation

Training set: 37 PET scans (3D volumes)

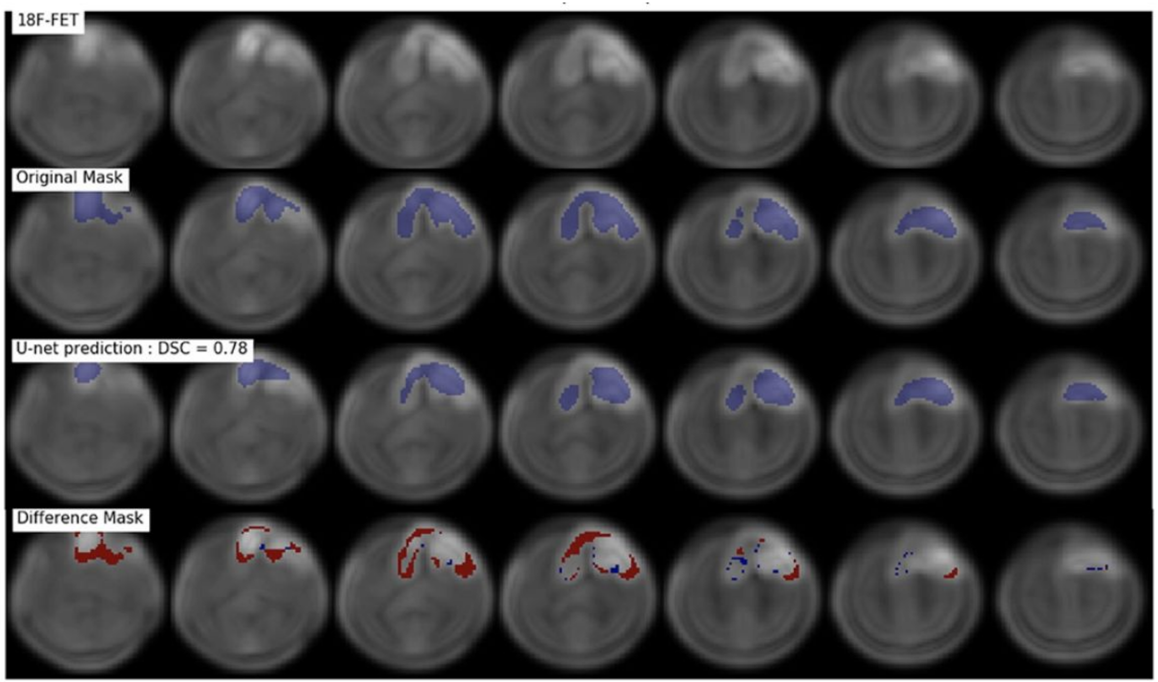
Evaluation set: 11 PET scans

Volumes resized to 64x64x40 for computational efficiency



Blanc-Durand et al. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. PLoS One, 2018.

Ex: Brain lesion segmentation

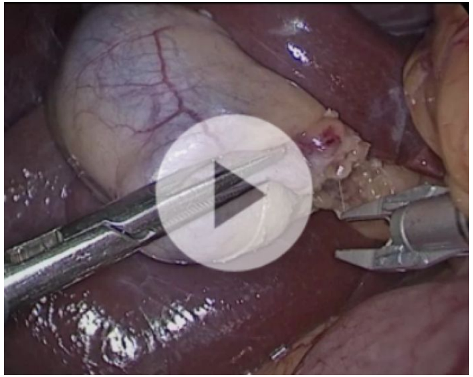


Blanc-Durand et al. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. PLoS One, 2018.

Video data (high dimensional in time)

E.g. in:

Surgery



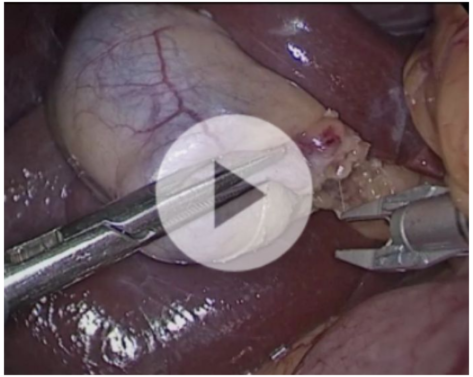
Hospital patient monitoring

Psychology

Video data (high dimensional in time)

E.g. in:

Surgery



Hospital patient monitoring

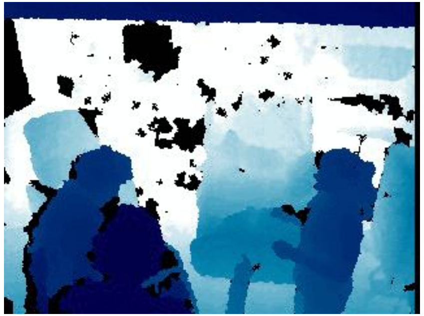
Psychology

Either 3D convolutional approaches or recurrent neural network approaches (we have time sequences!) can work

Detecting patient mobilization activities in the ICU

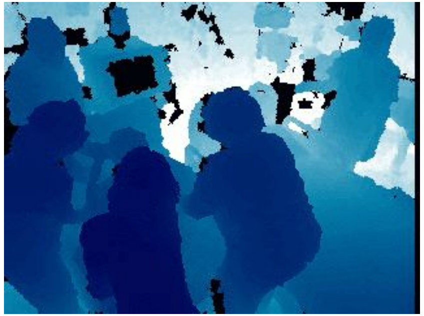
Get patient out of bed

Sit patient in chair

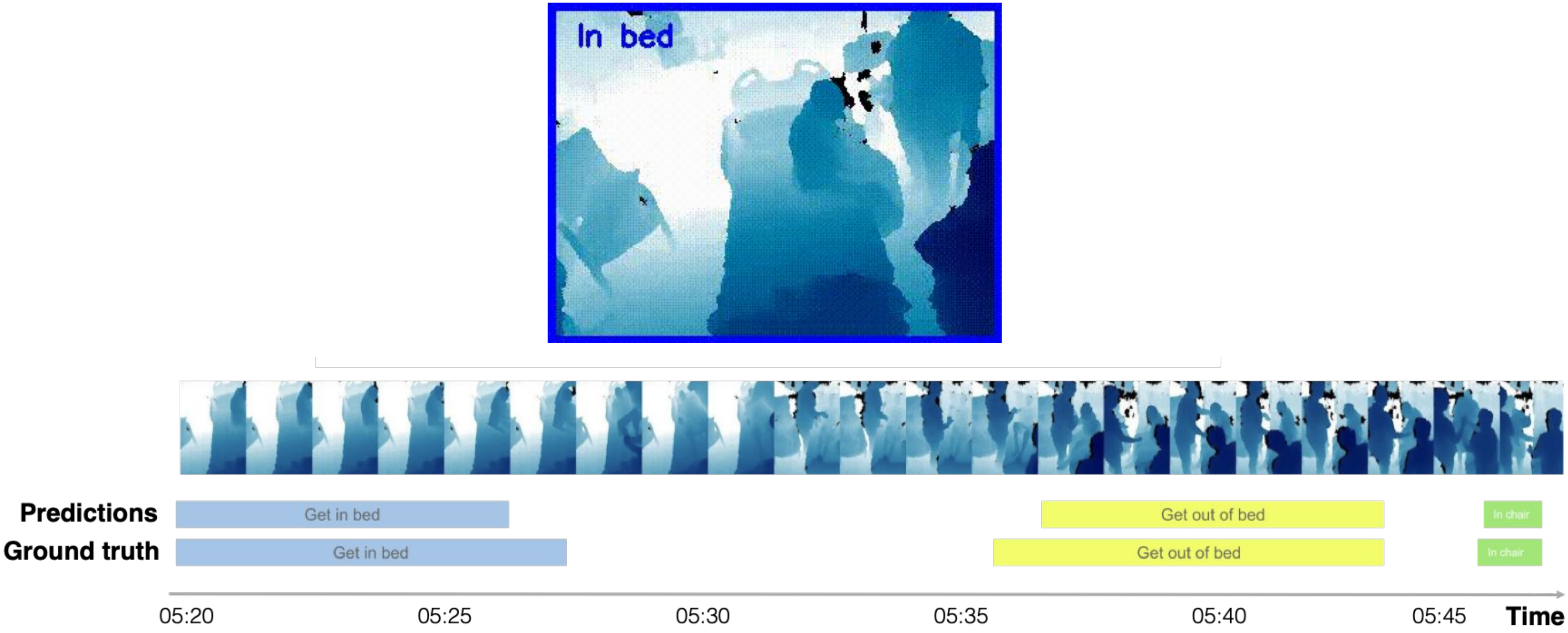


Get patient in bed

Get patient out of chair

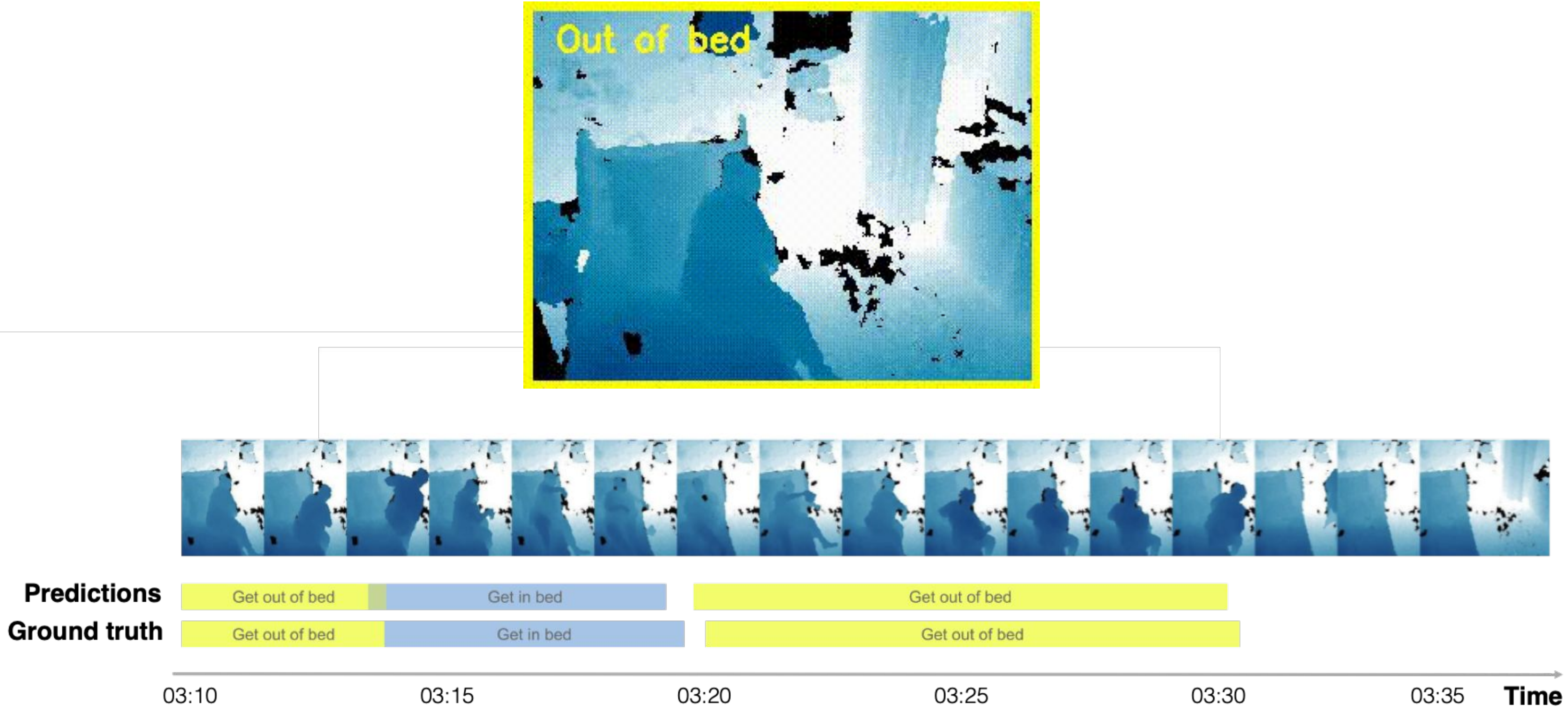


Detecting patient mobilization activities in the ICU



Yeung*,Salipur*, et al. A Computer Vision System for Deep Learning-Based Detection of Patient Mobilization Activities in the ICU. npj Digital Medicine, 2019.

Detecting patient mobilization activities in the ICU



Yeung*,Salipur*, et al. A Computer Vision System for Deep Learning-Based Detection of Patient Mobilization Activities in the ICU. npj Digital Medicine, 2019.

Summary

Today we covered:

- Recurrent and transformer neural networks for unstructured *text sequence* data
- Neural network architectures for more complex AI tasks

Next time: evaluation metrics for AI in healthcare