CS 4803 / 7643: Deep Learning

Topics:

- Convolutional Neural Networks
  - Transposed convolutions
- Visualizing CNNs

Dhruv Batra

Georgia Tech

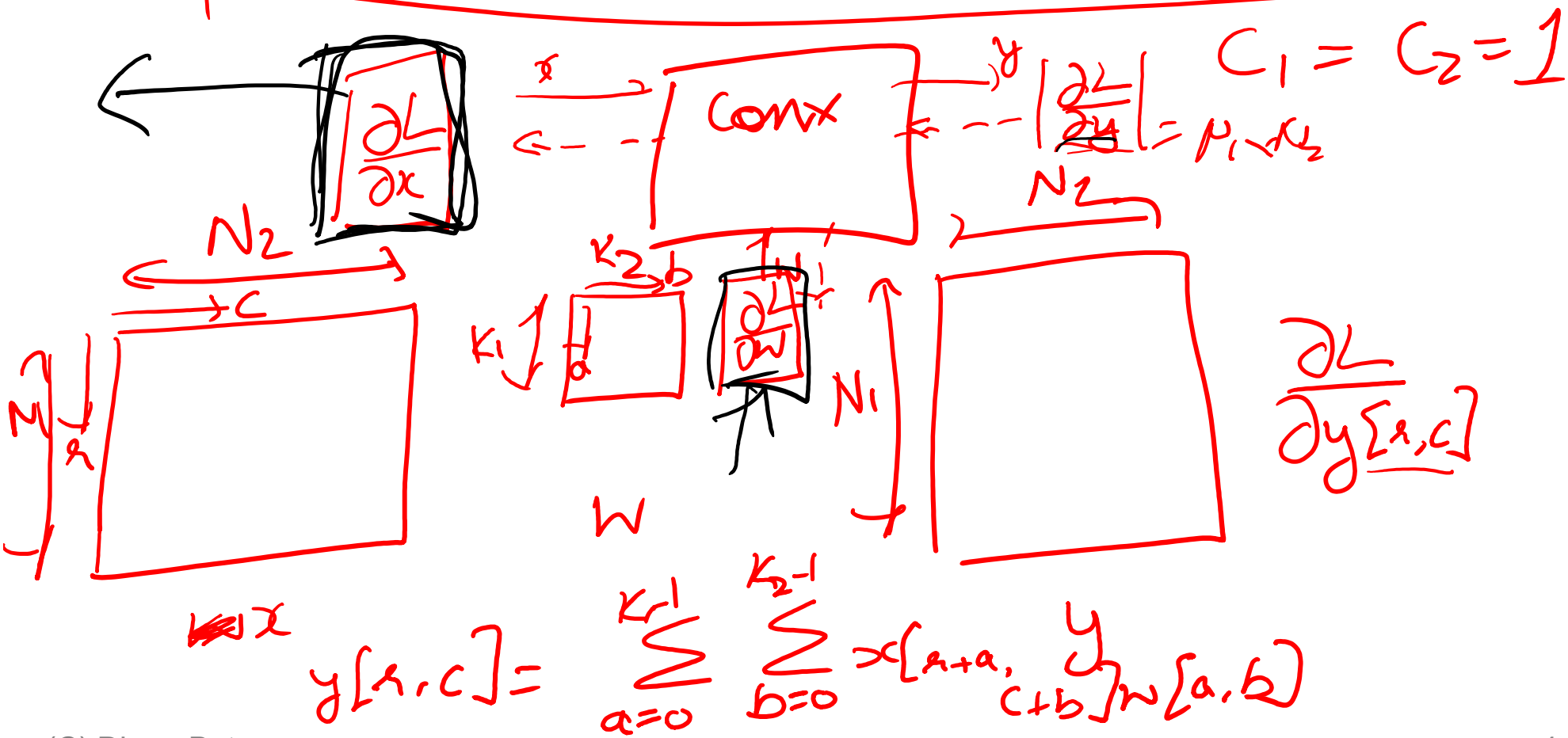
Administrativia

- HW0 Grades Released
  - Regard request windows closes 11:59pm 10/16.
- HW1 Challenge Final Analysis
  - https://docs.google.com/spreadsheets/d/1taAu_5AQSiDMtwIY59wTijkd-2Y95nswmv0JnH0eYRI/edit#gid=1468043323
  - Coming soon.
- HW2 Released
  - Due: 10/18, 11:55pm
  - https://www.cc.gatech.edu/classes/AY2019/cs7643_fall/assets/hw2.pdf

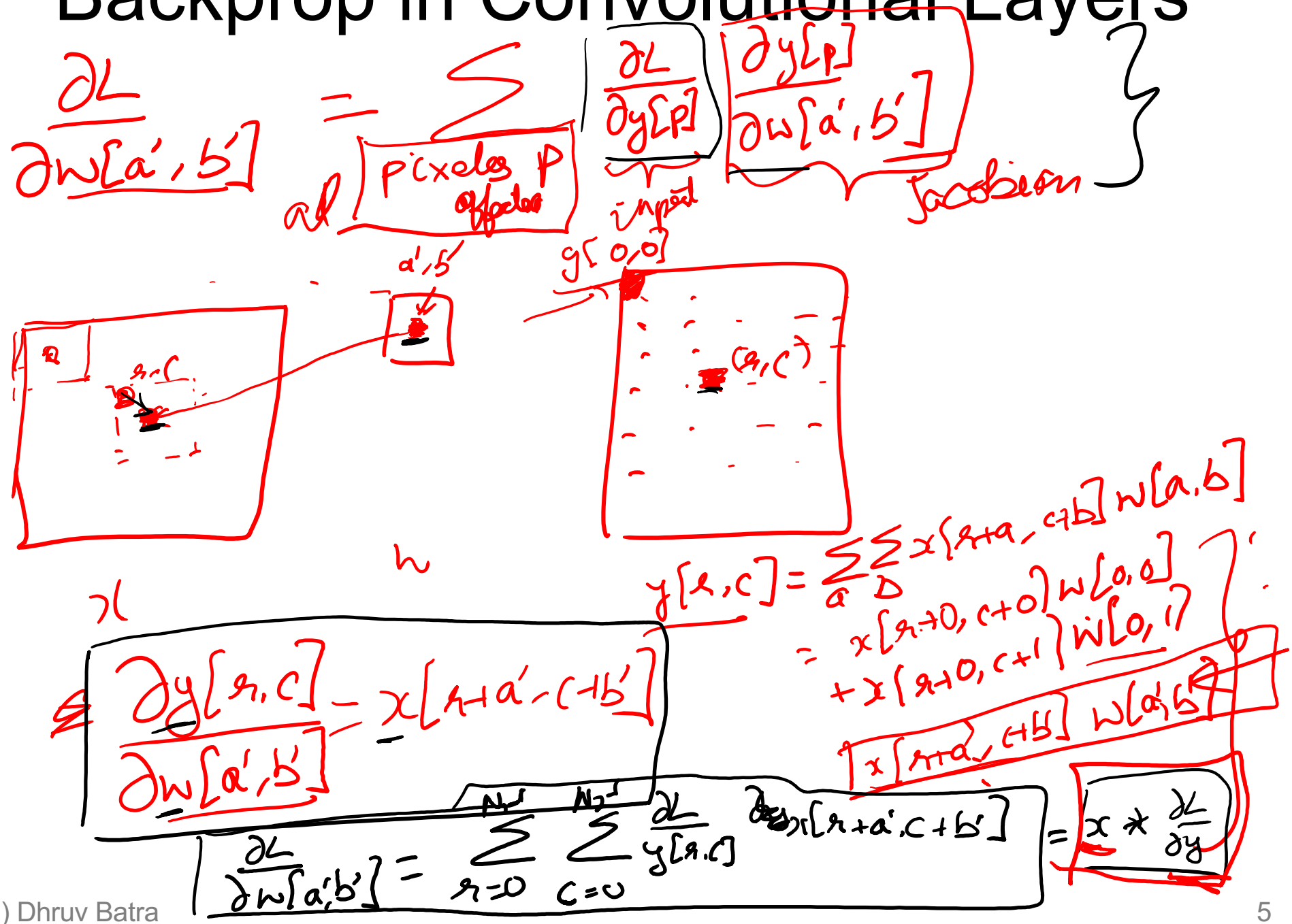
Recap from last time

Backprop in Convolutional Layers

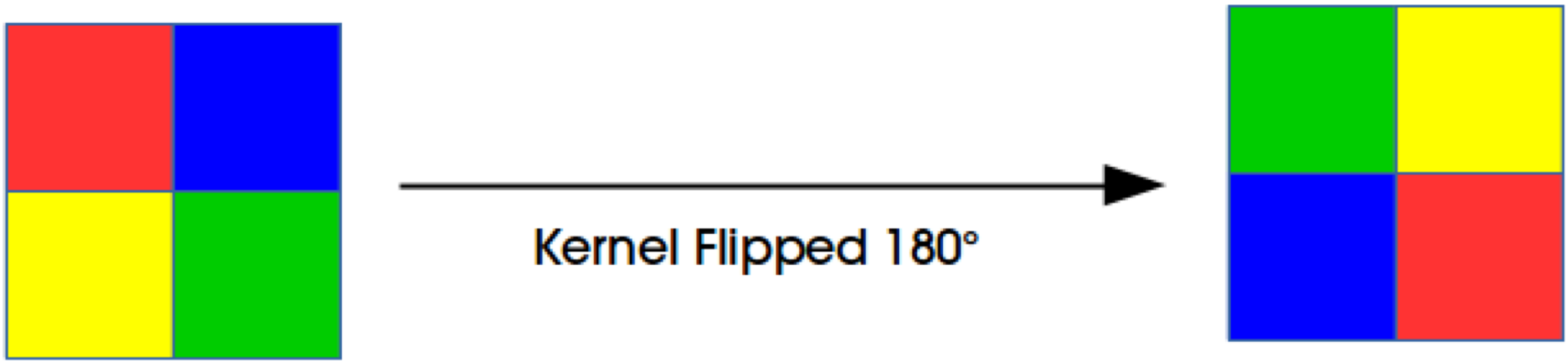
- Notes
  - https://www.cc.gatech.edu/classes/AY2018/cs7643_fall/slides/L6_cnns_backprop_notes.pdf



Backprop in Convolutional Layers

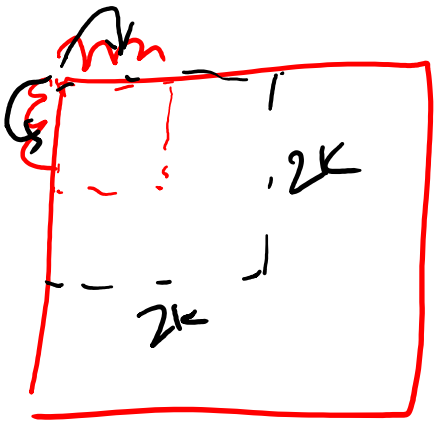


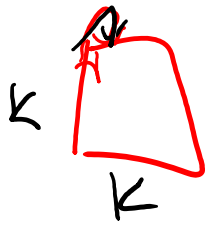
Backprop in Convolutional Layers

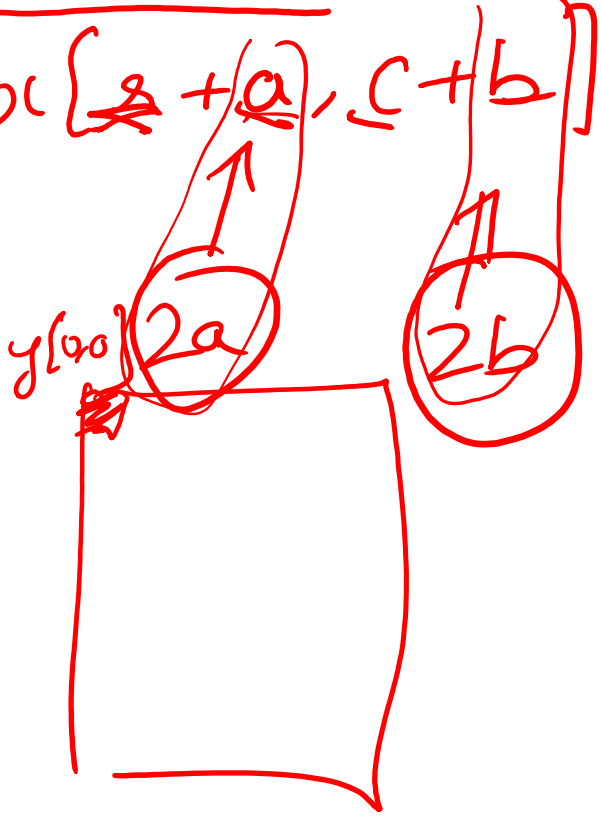


Dilated Convolutions

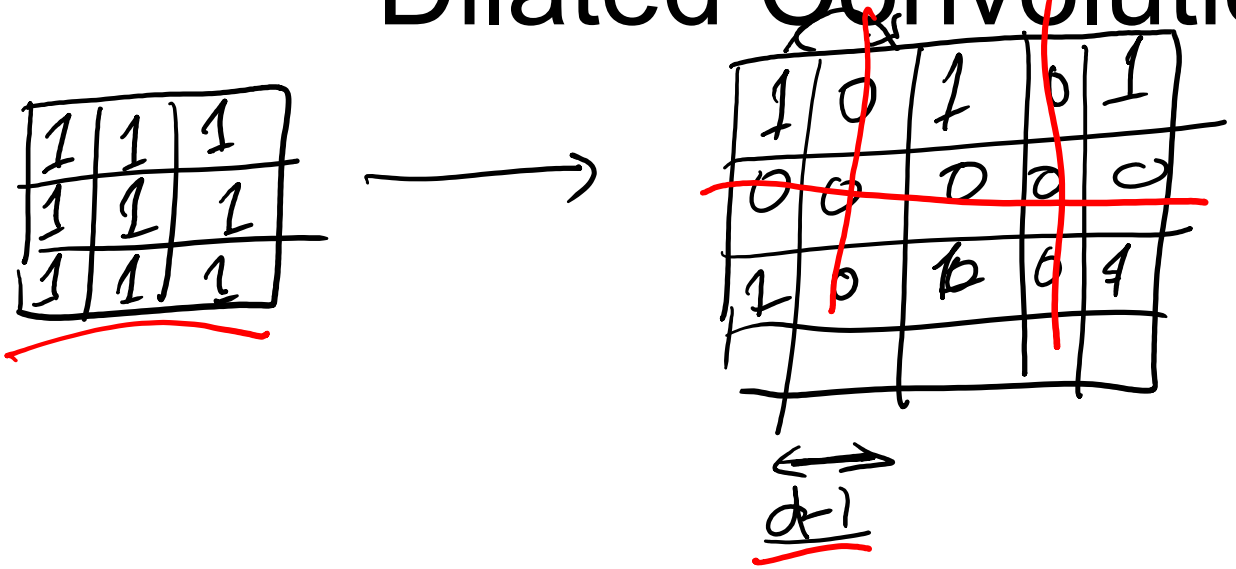
$$y[r,c] = \sum_a \sum_b x[r+a, c+b]\, w[a,b]$$







Dilated Convolutions



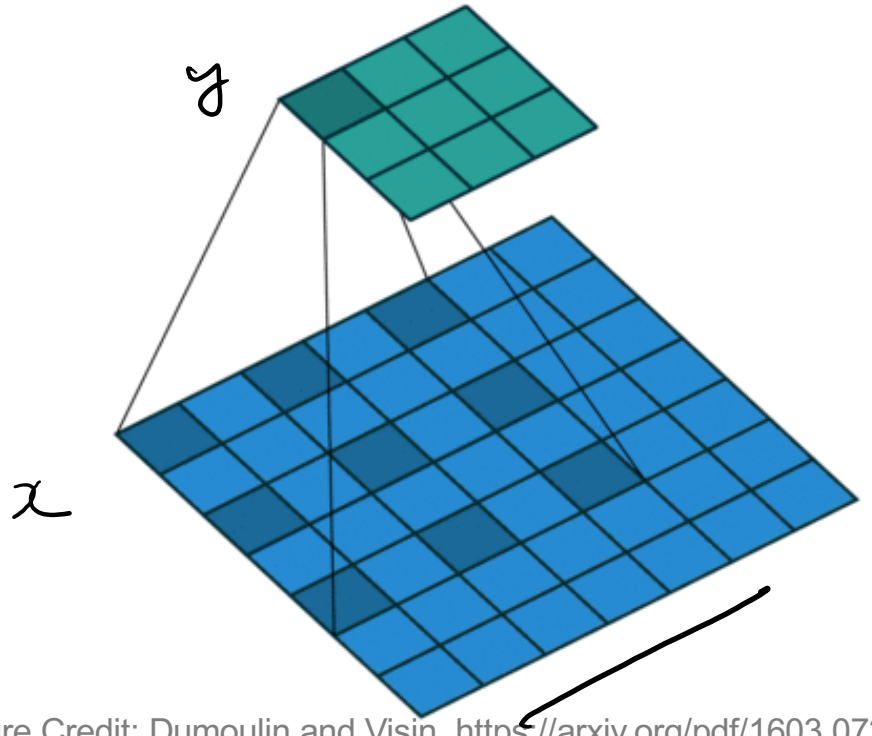
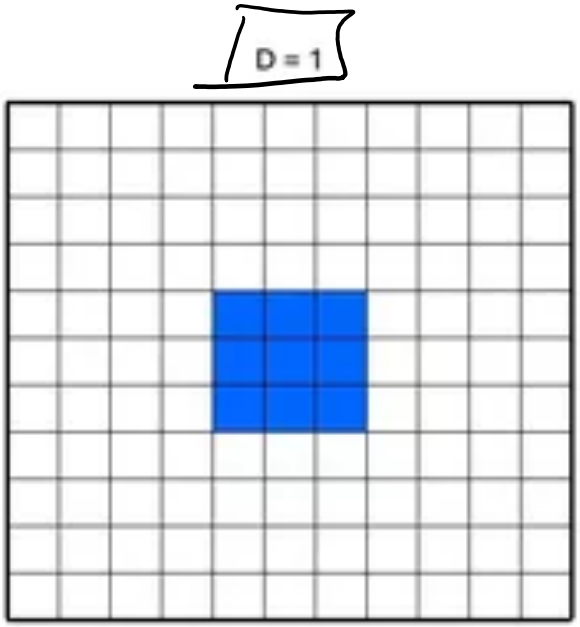
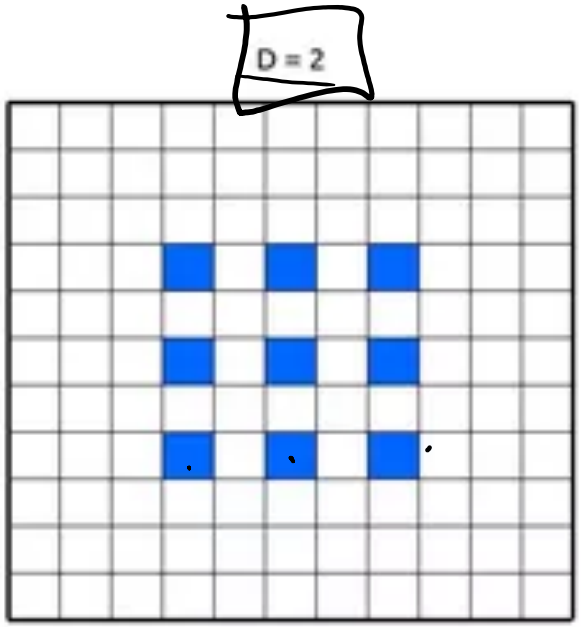
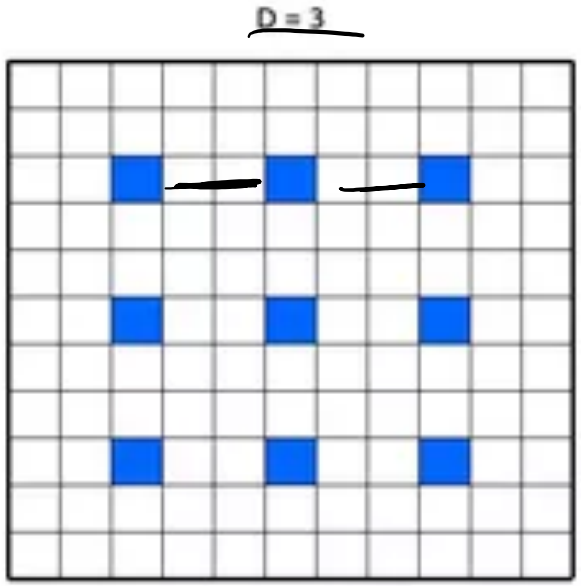


Figure Credit: Dumoulin and Visin, https://arxiv.org/pdf/1603.07285.pdf







$$k \rightarrow k + (k-1)(d-1)$$

$$3 + (3-1)(2-1) = 5$$

(recall:)

(N - k) / stride + 1

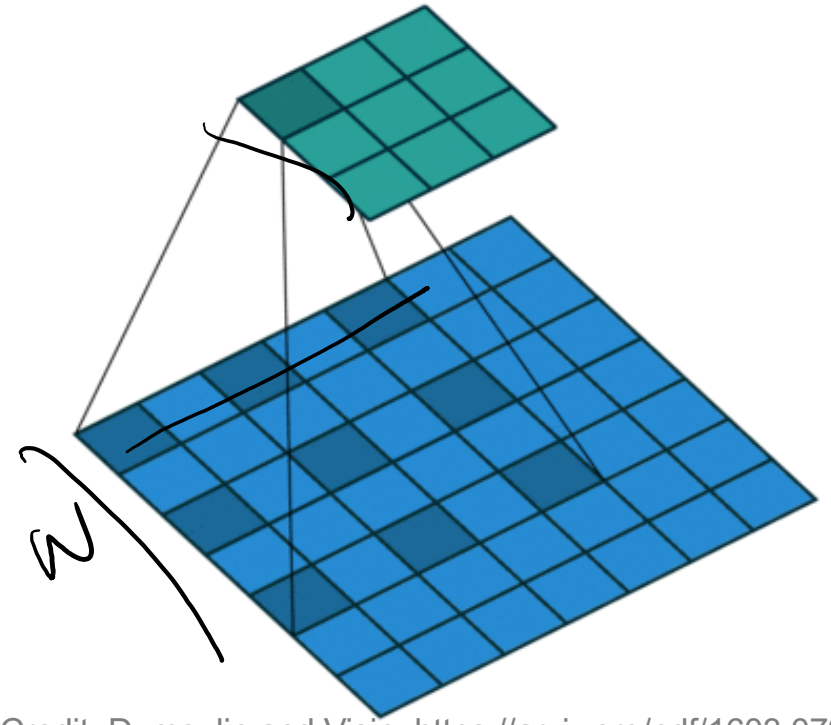


Figure Credit: Dumoulin and Visin, https://arxiv.org/pdf/1603.07285.pdf

Toeplitz Matrix

- Diagonals are constants

$$\begin{bmatrix} a & b & c & d & e \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ i & h & g & f & a \end{bmatrix}.$$

- $A_{ij} = a_{i-j}$

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots & \dots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \dots & \dots & a_2 & a_1 & a_0 \end{bmatrix}$$

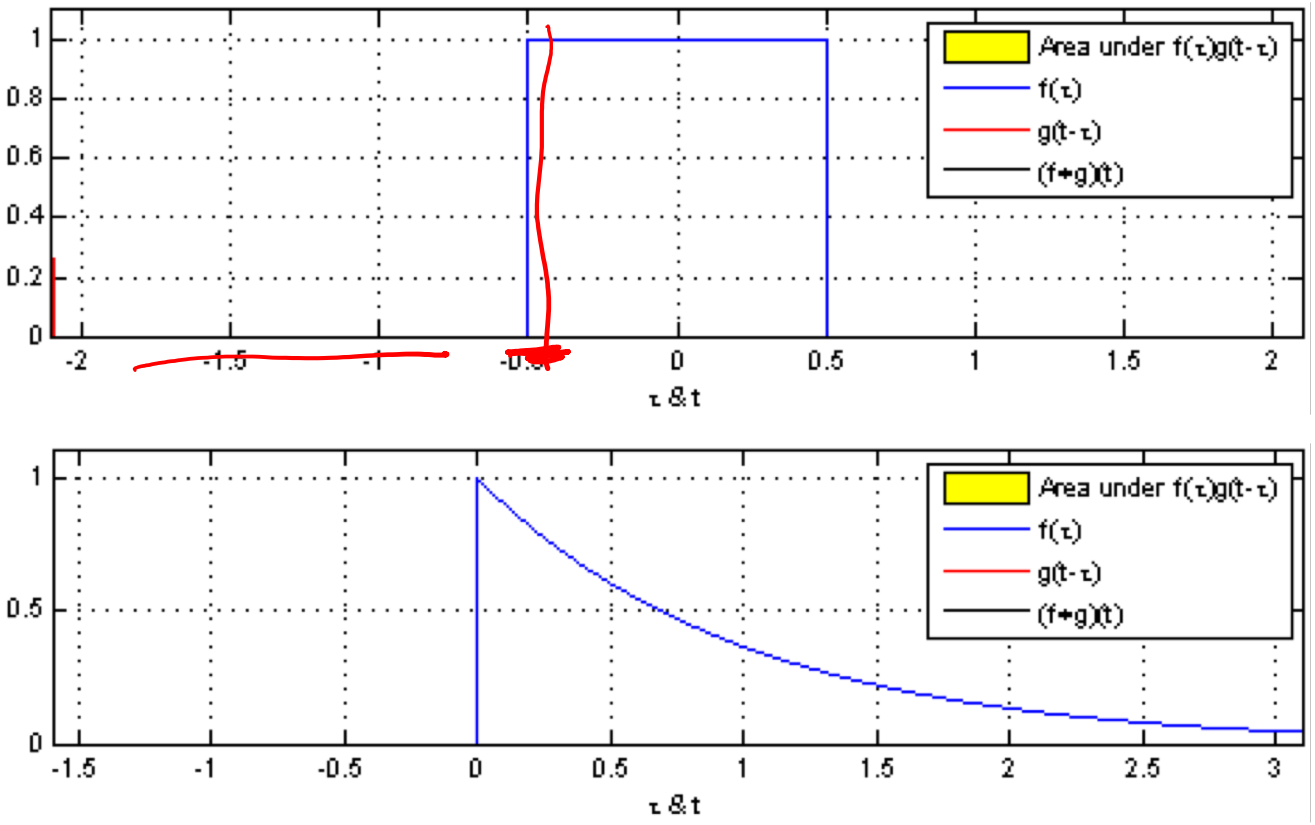
Why do we care?

- (Discrete) Convolution = Matrix Multiplication
  - with Toeplitz Matrices

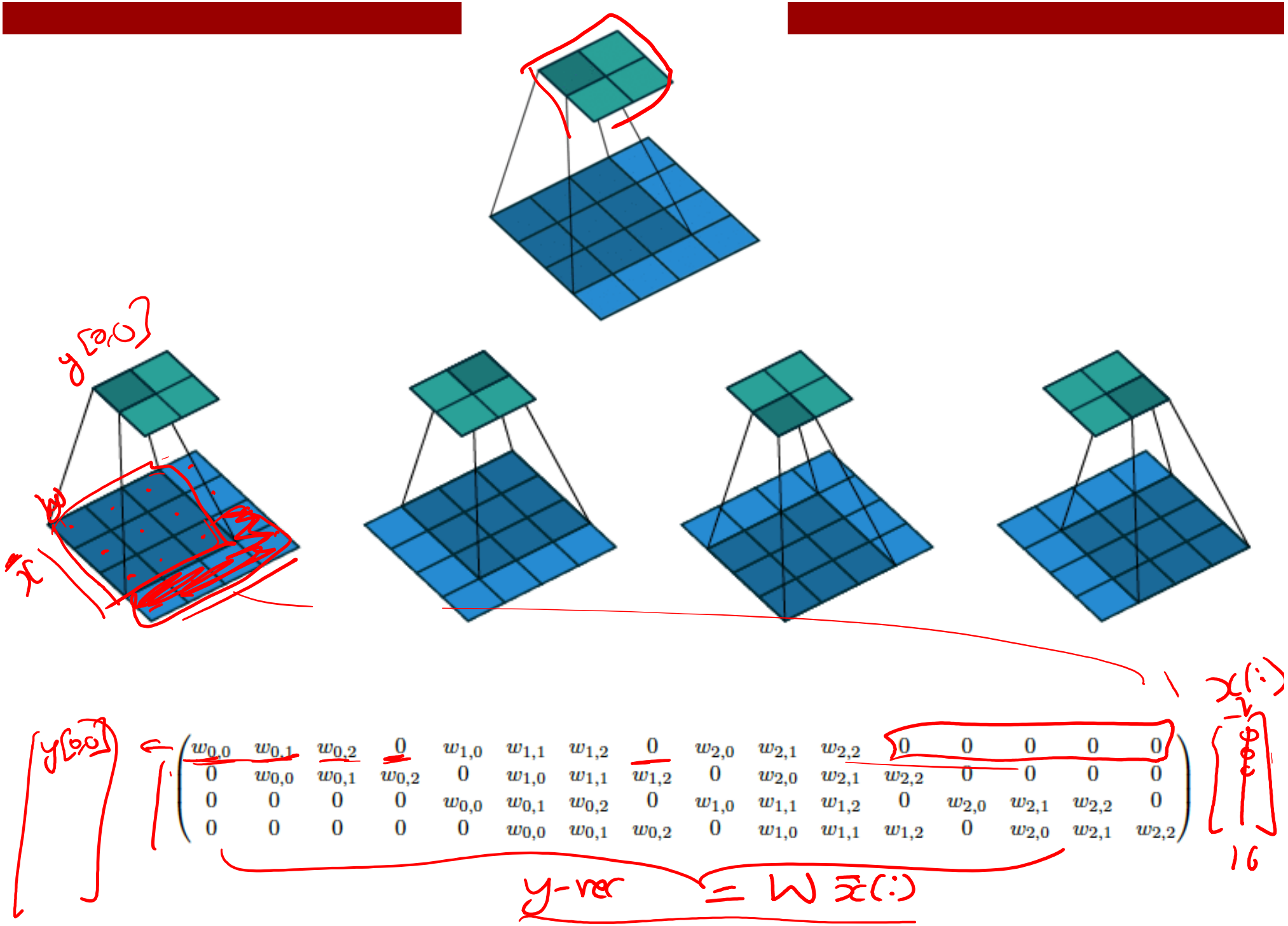
$$y = w * x$$

$$\begin{bmatrix} w_k & 0 & \dots & 0 & 0 \\ w_{k-1} & w_k & \dots & 0 & 0 \\ w_{k-2} & w_{k-1} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_1 & w_{k-2} & \dots & w_k & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & w_1 & \dots & w_{k-1} & w_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & w_1 & w_2 \\ 0 & 0 & \vdots & 0 & w_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$\vec{y} = W\vec{x}$$



"Convolution of box signal with itself2" by Convolution_of_box_signal_with_itself.gif: Brian Ambergderivative work: Tinos (talk) - Convolution_of_box_signal_with_itself.gif. Licensed under CC BY-SA 3.0 via Commons - https://commons.wikimedia.org/wiki/File:Convolution_of_box_signal_with_itself2.gif#/media/File:Convolution_of_box_signal_with_itself2.gif



$$\begin{pmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{pmatrix}$$

Figure Credit: Dumoulin and Visin, https://arxiv.org/pdf/1603.07285.pdf

So far: Image Classification

This image is CC0 public domain

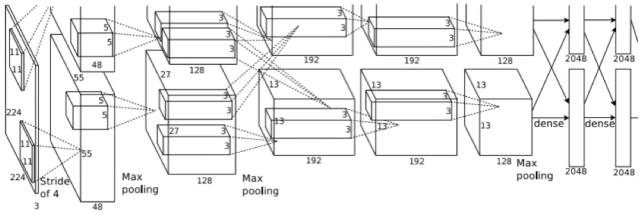


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Vector:
4096

Fully-Connected:
4096 to 1000

Class Scores
Cat: 0.9
Dog: 0.05
Car: 0.01
...

Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

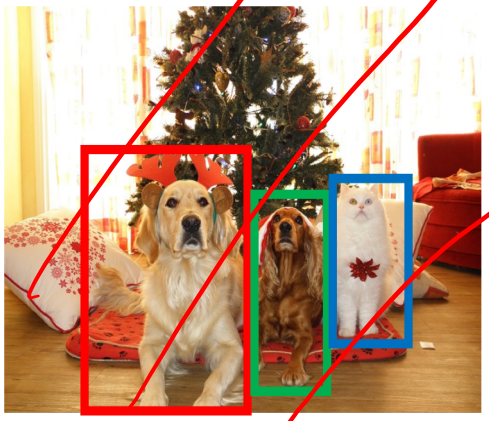
Other Computer Vision Tasks

Semantic Segmentation

GRASS, CAT, TREE, SKY

No objects, just pixels

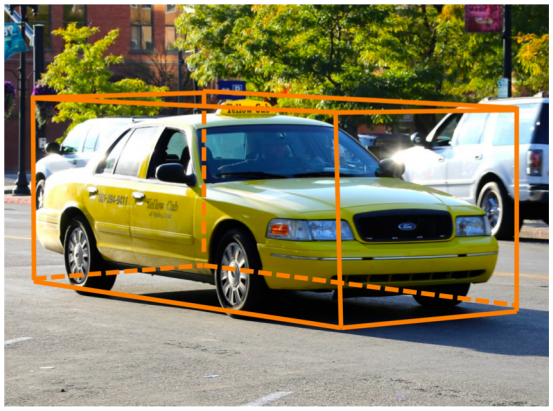
2D Object Detection



DOG, DOG, CAT

Object categories + 2D bounding boxes

3D Object Detection

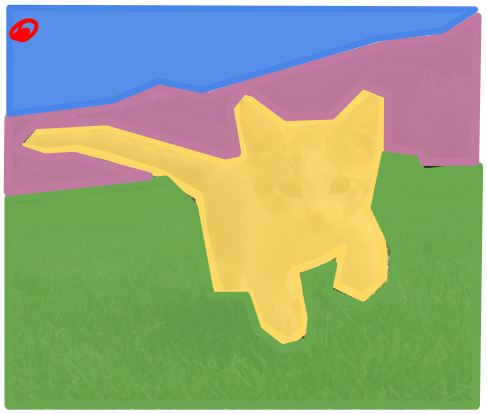


Car

Object categories + 3D bounding boxes

Semantic Segmentation

Semantic Segmentation



GRASS, CAT, TREE, SKY

No objects, just pixels

2D Object Detection

DOG, DOG, CAT

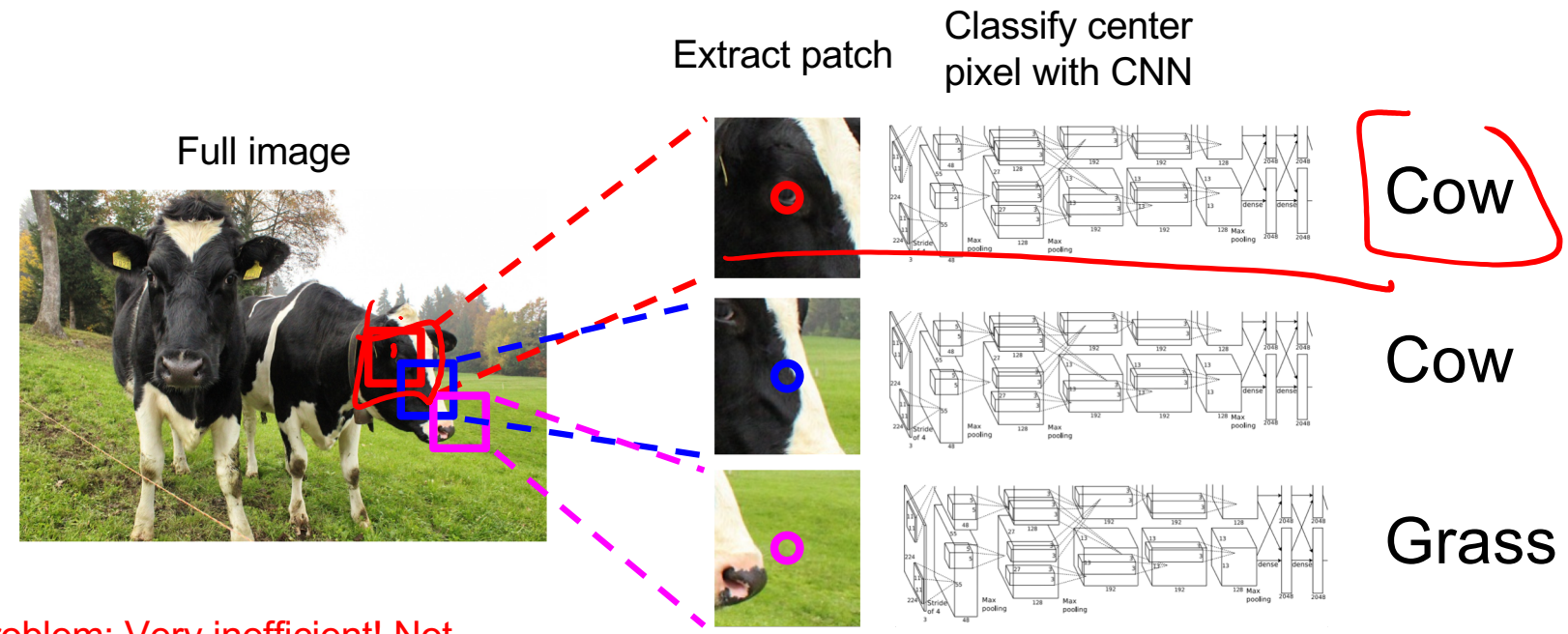
Object categories + 2D bounding boxes

3D Object Detection

Car

Object categories + 3D bounding boxes

Semantic Segmentation Idea: Sliding Window



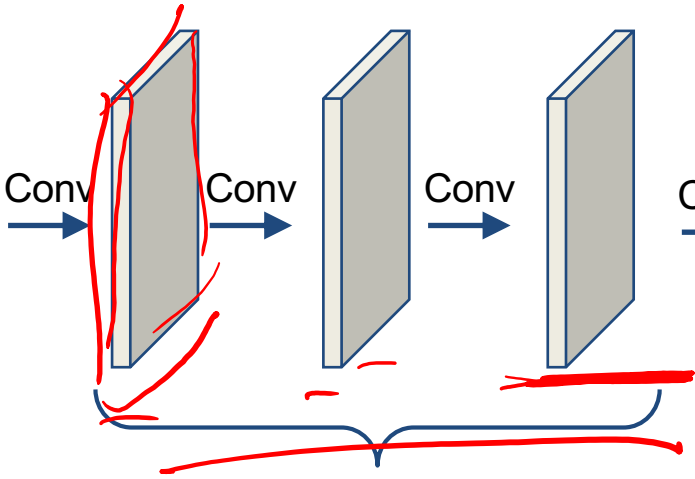
Problem: Very inefficient! Not reusing shared features between overlapping patches

Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI 2013
Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

Semantic Segmentation Idea: Fully Convolutional

Design a network as a bunch of convolutional layers to make predictions for pixels all at once!

Input:
3 x H x W



Convolutions:
D x H x W

Scores:
C x H x W

Predictions:
H x W

$$\sum_{pixels} -\log \hat{p}(x,c,y^*)$$

Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

Semantic Segmentation Idea: Fully Convolutional

Design a network as a bunch of convolutional layers to make predictions for pixels all at once!

Input:
3 x H x W

Conv → Conv → Conv

Convolutions:
D x H x W

Problem: convolutions at original image resolution will be very expensive ...

Conv

Scores:
C x H x W

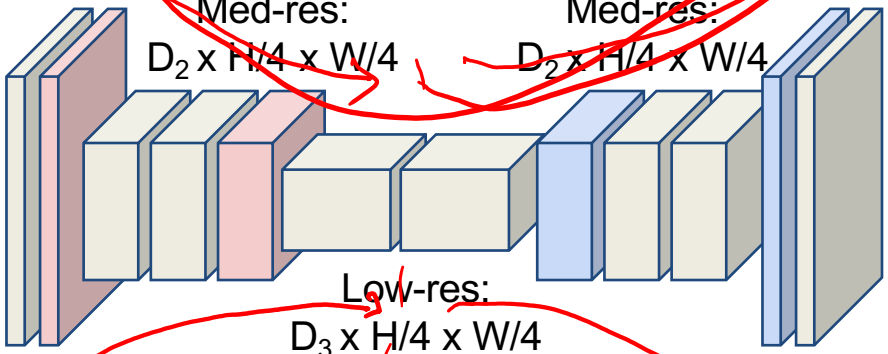
argmax

Predictions:
H x W

Semantic Segmentation Idea: Fully Convolutional

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

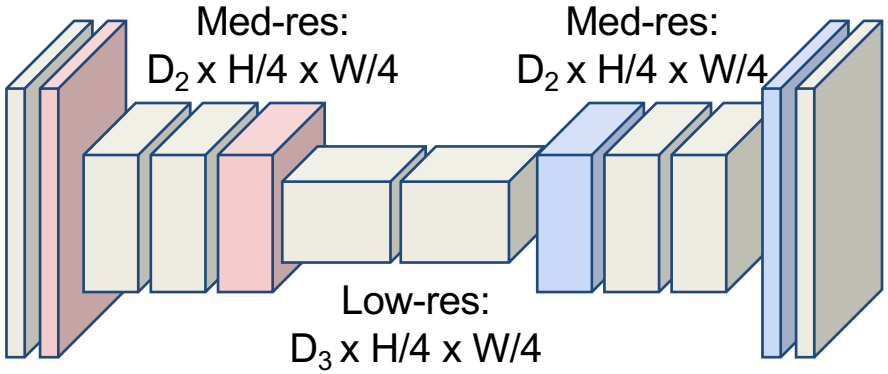
Semantic Segmentation Idea: Fully Convolutional

Downsampling: Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling: ???

Input:
3 x H x W



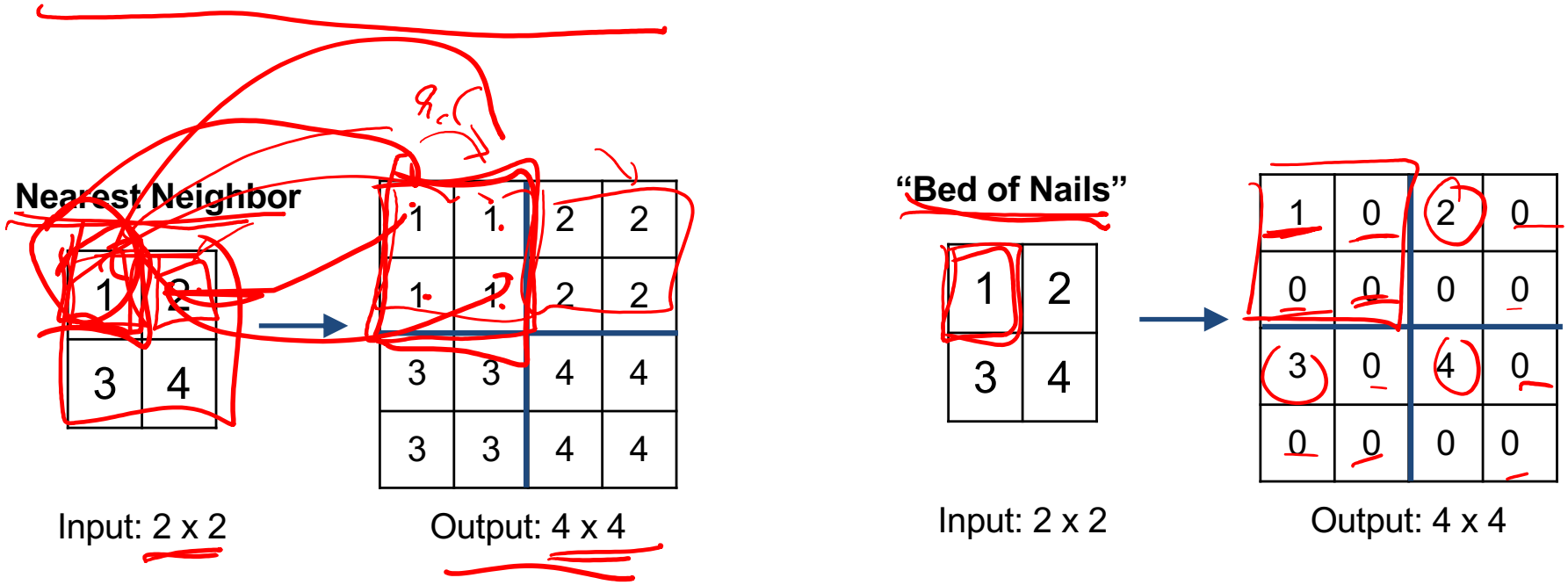
High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

In-Network upsampling: “Unpooling”



In-Network upsampling: "Max Unpooling"

Max Pooling
Remember which element was max!

1	2	6	3
3	5	2	1
1	2	2	1
7	3	4	8

Input: 4 x 4

5	6
7	8

Output: 2 x 2

Rest of the network

Max Unpooling
Use positions from pooling layer

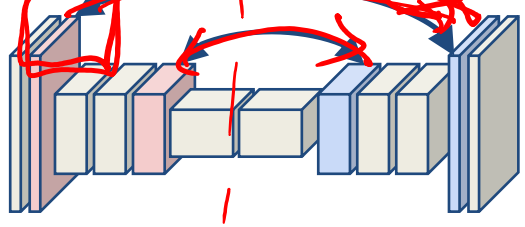
1	2
3	4

Input: 2 x 2

0	0	2	0
0	1	0	0
0	0	0	0
3	0	0	4

Output: 4 x 4

Corresponding pairs of downsampling and upsampling layers



Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

Plan for Today

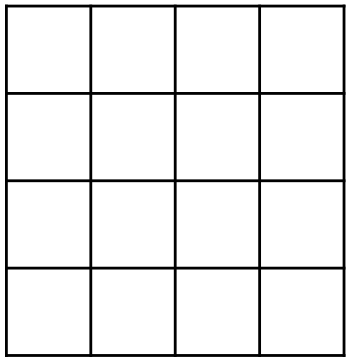
- Convolutional Neural Networks
  - Transposed convolutions
- Visualizing CNNs

Transposed Convolutions

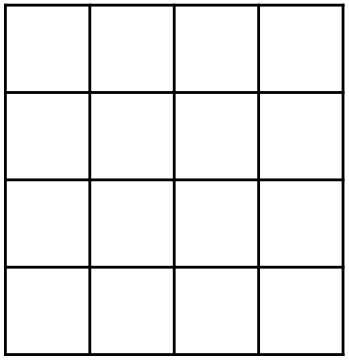
- Deconvolution (bad)
- Upconvolution
- Fractionally strided convolution
- Backward strided convolution

Learnable Upsampling: Transpose Convolution

Recall: Typical 3 x 3 convolution, stride 1 pad 1



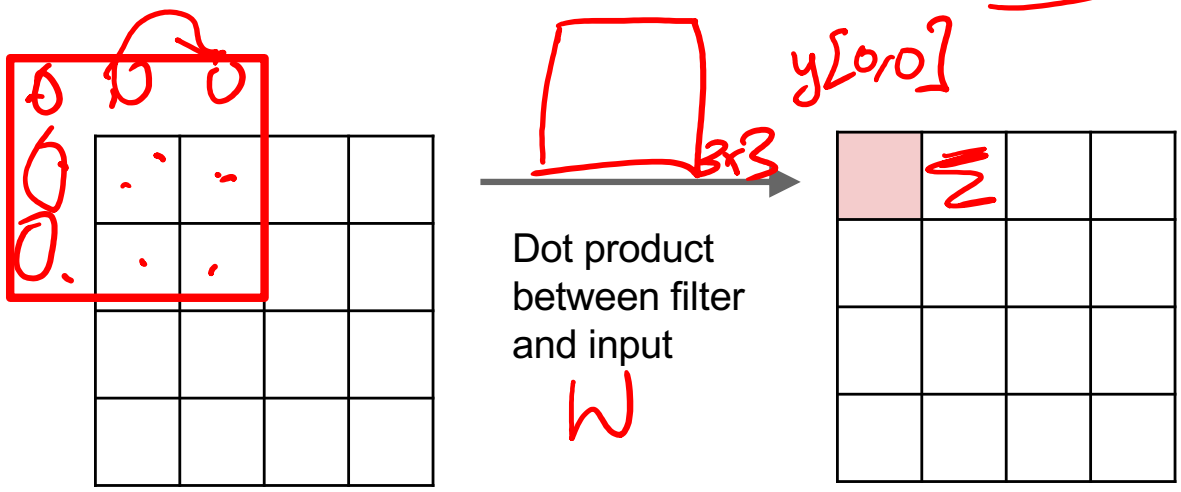
Input: 4 x 4



Output: 4 x 4

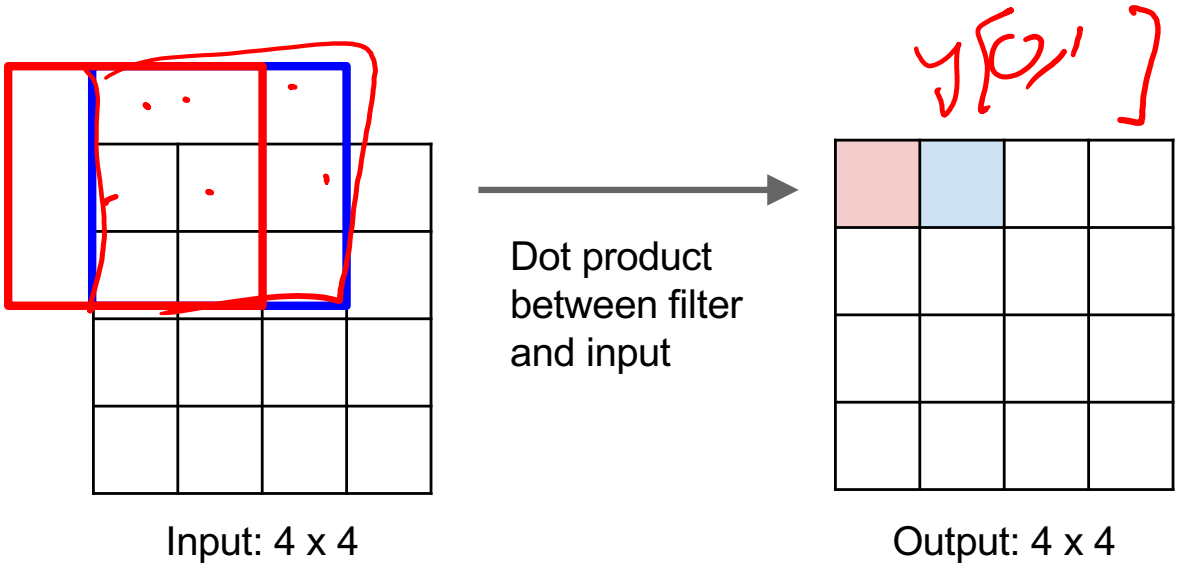
Learnable Upsampling: Transpose Convolution

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



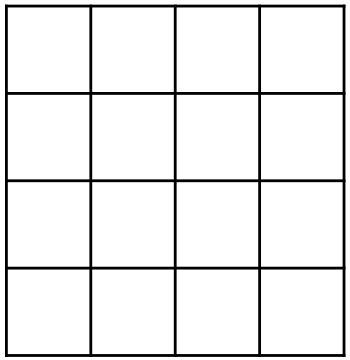
Learnable Upsampling: Transpose Convolution

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

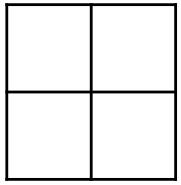


Learnable Upsampling: Transpose Convolution

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



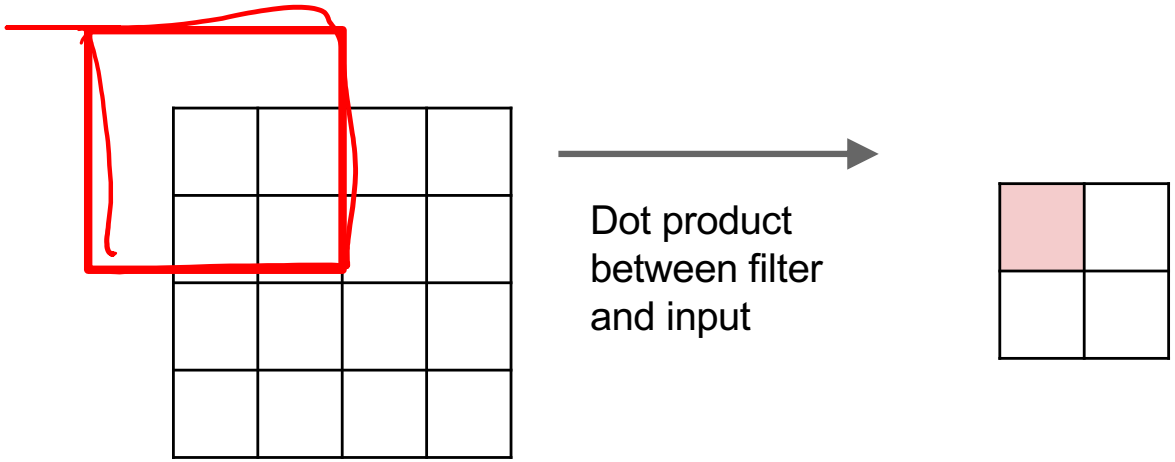
Input: 4 x 4



Output: 2 x 2

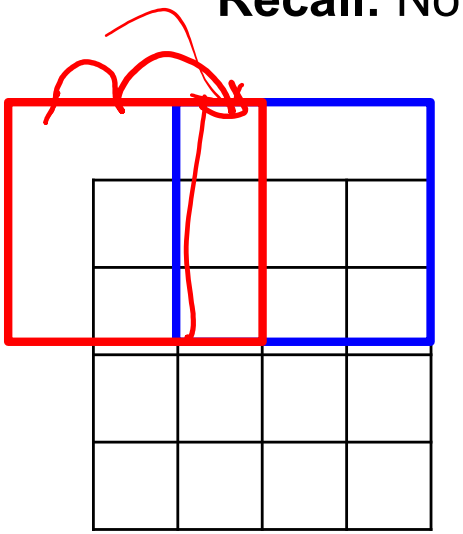
Learnable Upsampling: Transpose Convolution

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



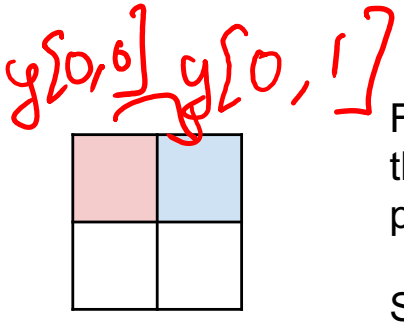
Learnable Upsampling: Transpose Convolution

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



Input: 4 x 4

Dot product between filter and input



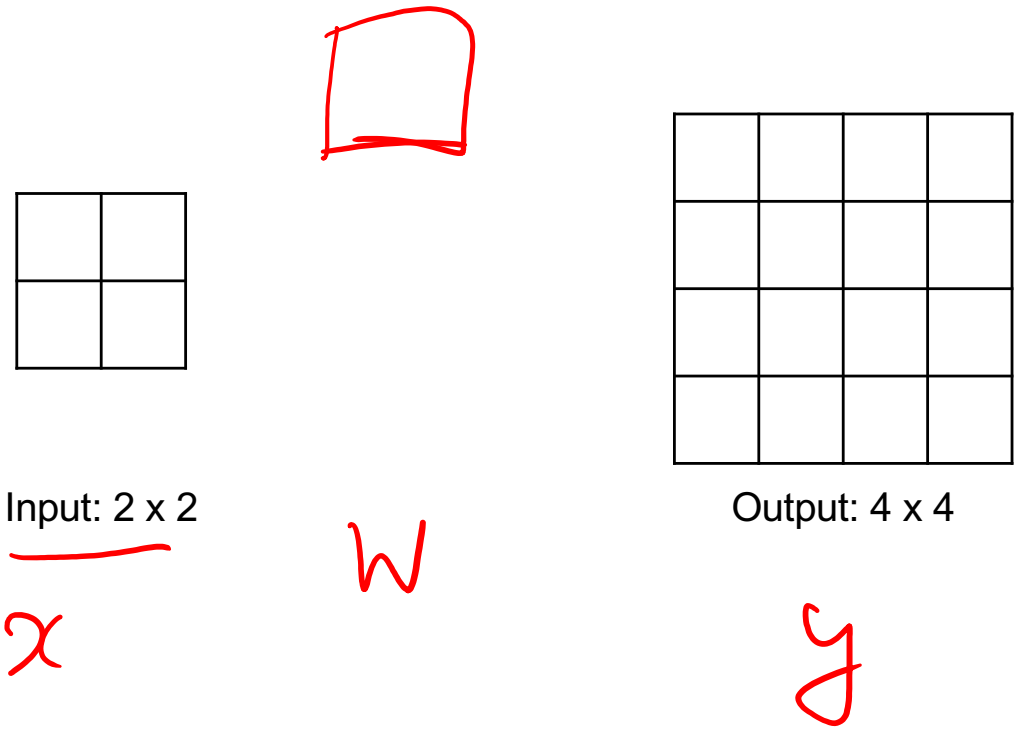
Output: 2 x 2

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

Stride gives ratio between movement in input and output

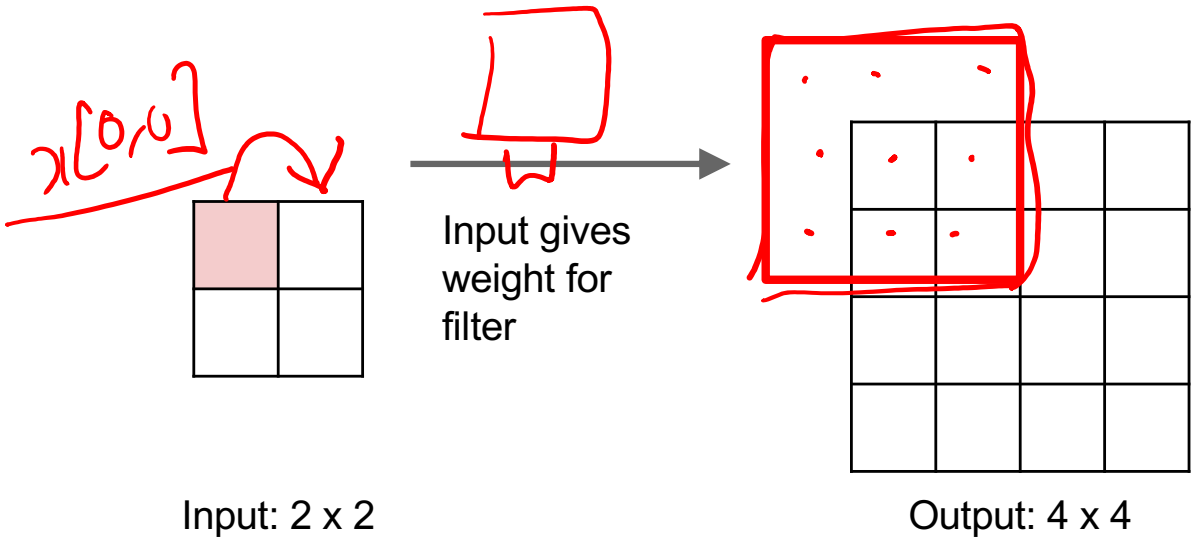
Learnable Upsampling: Transpose Convolution

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

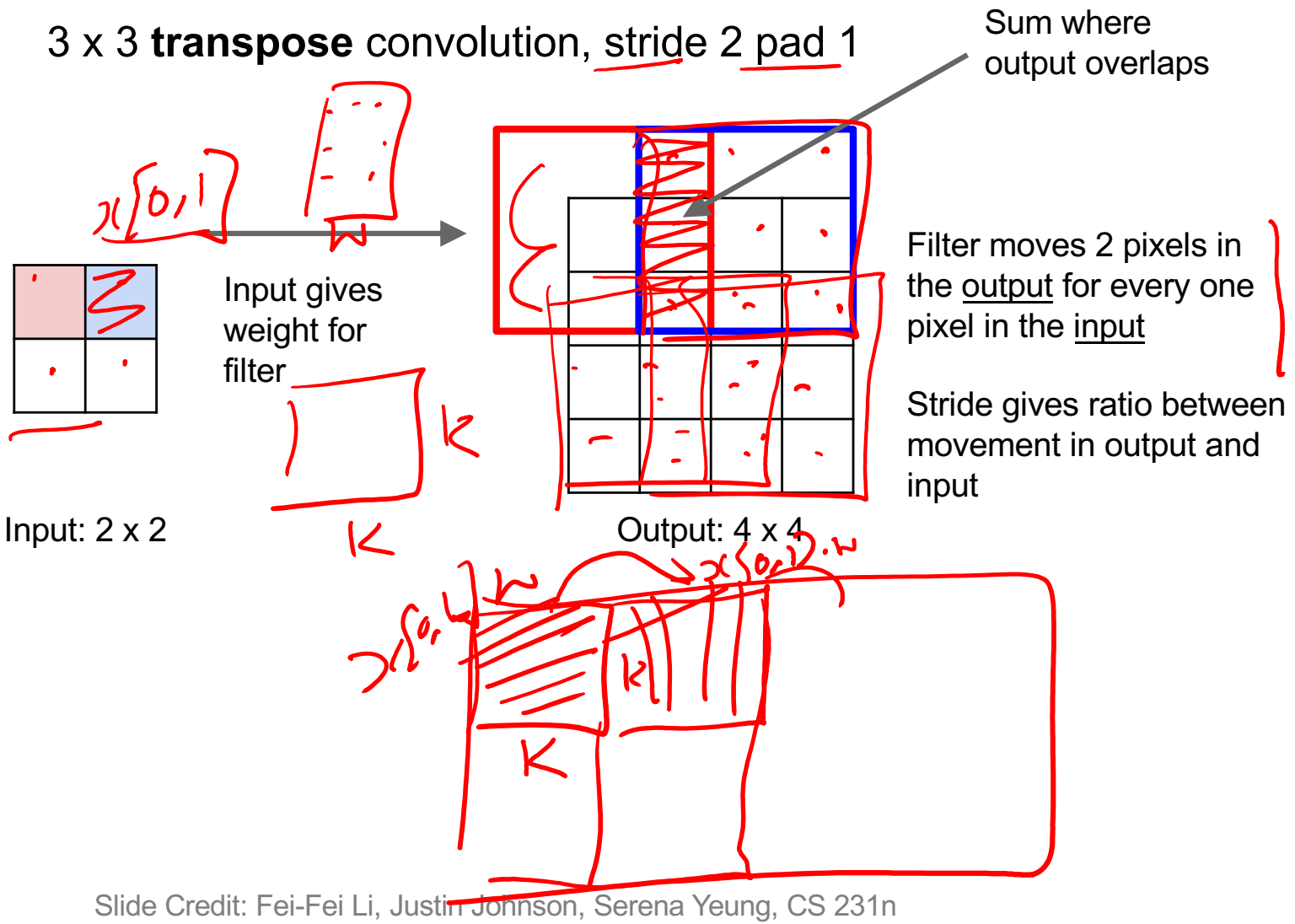


Learnable Upsampling: Transpose Convolution

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



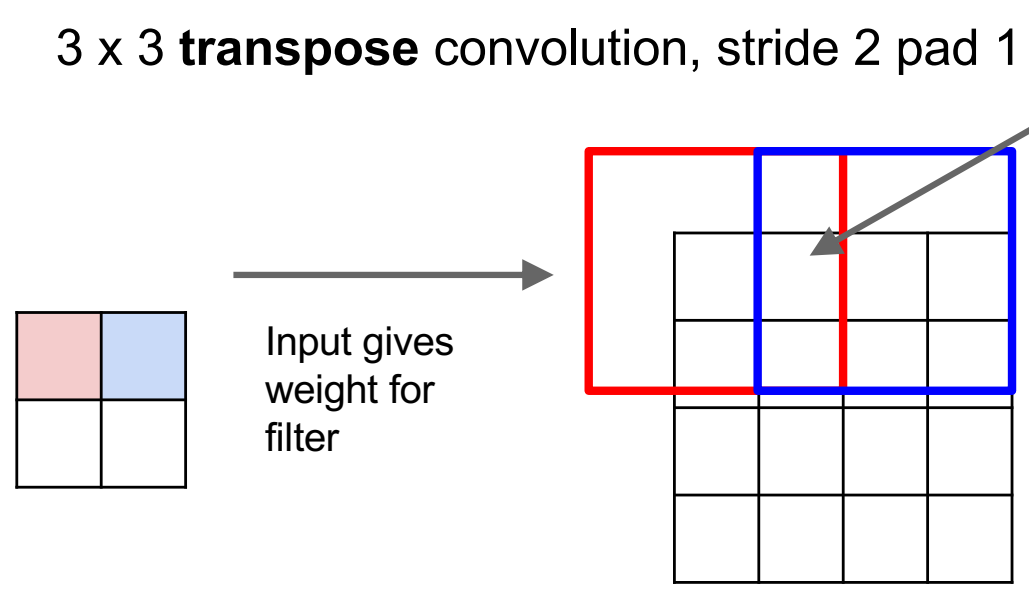
Learnable Upsampling: Transpose Convolution



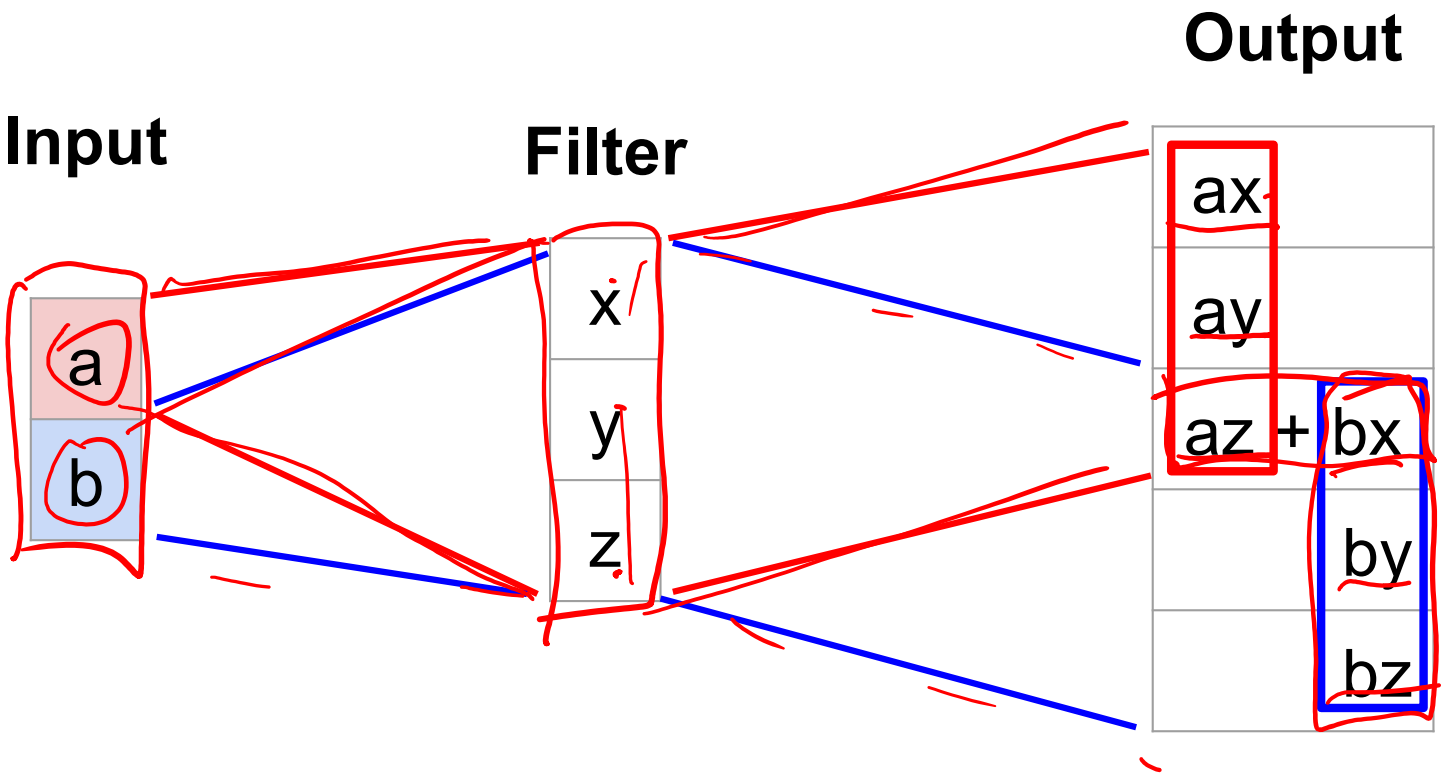
Learnable Upsampling: Transpose Convolution

Other names:
-~~Deconvolution~~ (bad)
-Upconvolution
-Fractionally strided convolution
-Backward strided convolution

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



Transpose Convolution: 1D Example



Output contains copies of the filter weighted by the input, summing at where at overlaps in the output

Need to crop one pixel from output to make output exactly 2x input

Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

Transposed Convolution

- https://distill.pub/2016/deconv-checkerboard/

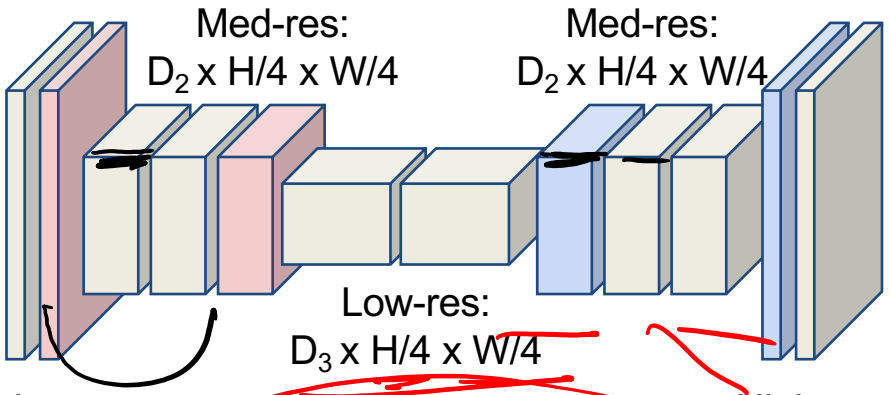
Semantic Segmentation Idea: Fully Convolutional

Downsampling: Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

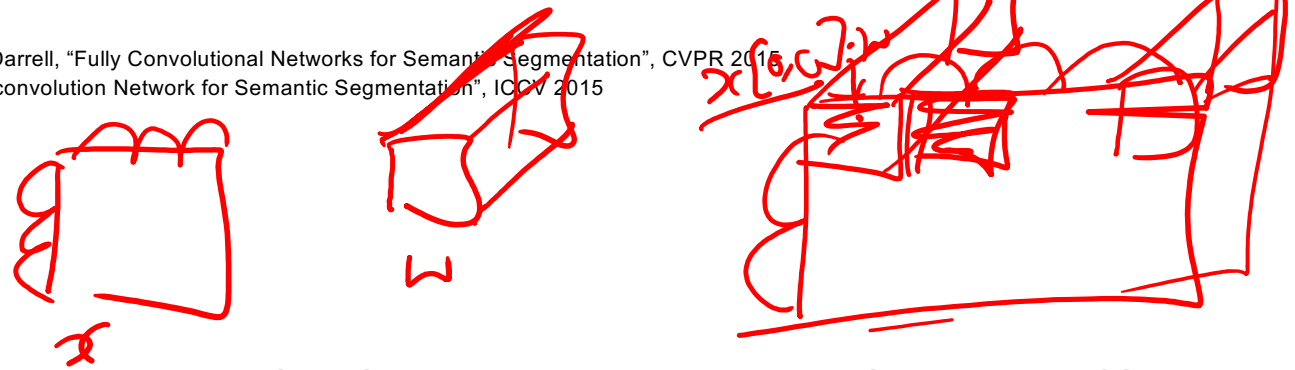
Upsampling: Unpooling or strided transpose convolution

Input:
3 x H x W



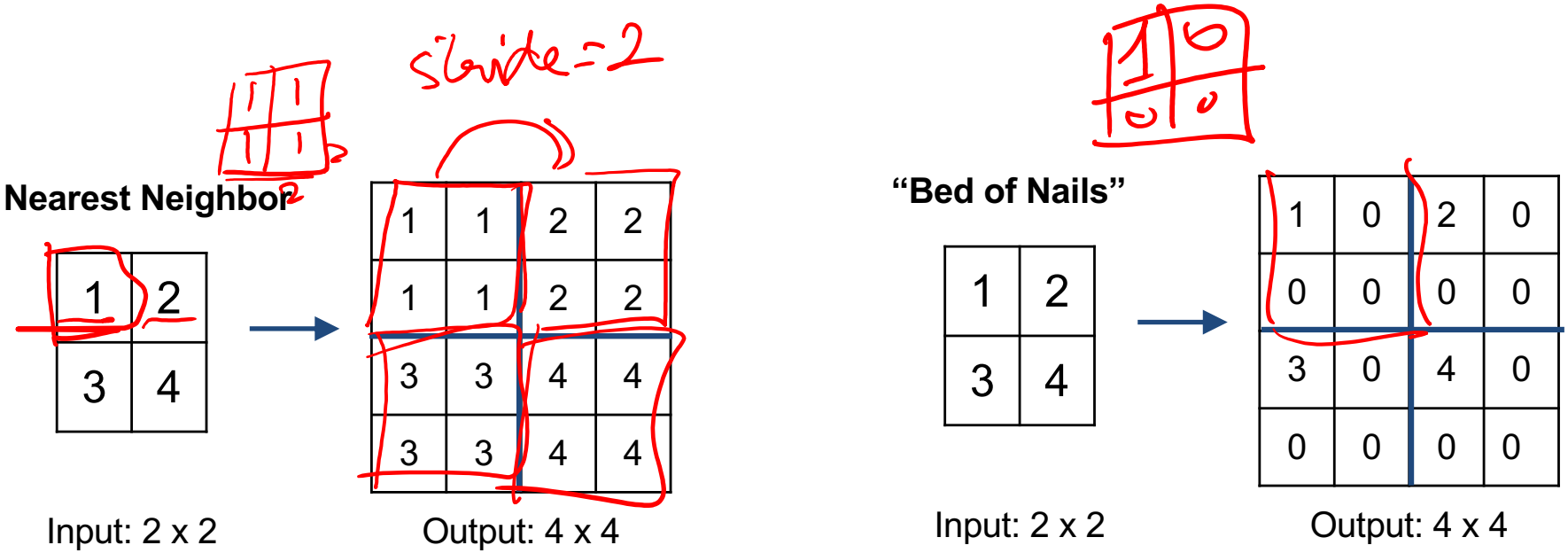
Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015



Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

In-Network upsampling: “Unpooling”



Why this operation?

What is deconvolution?

- (Non-blind) Deconvolution

$$y = w * x$$

$$\vec{y} = W\vec{x}$$

What is deconvolution?

- (Non-blind) Deconvolution

$$y = w * x$$

$$\begin{bmatrix} w_k & 0 & \dots & 0 & 0 \\ w_{k-1} & w_k & \dots & 0 & 0 \\ w_{k-2} & w_{k-1} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_1 & w_{k-2} & \dots & w_k & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & w_1 & \dots & w_{k-1} & w_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & w_1 & w_2 \\ 0 & 0 & \vdots & 0 & w_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

What does “deconvolution” have to do with “transposed convolution”?

“transposed convolution” is a convolution!

We can express convolution in terms of a matrix multiplication

$$\vec{x} * \vec{a} = X\vec{a}$$

$$\begin{bmatrix} x & y & z & 0 & 0 & 0 \\ 0 & x & y & z & 0 & 0 \\ 0 & 0 & x & y & z & 0 \\ 0 & 0 & 0 & x & y & z \end{bmatrix} \begin{bmatrix} 0 \\ a \\ b \\ c \\ d \\ 0 \end{bmatrix} = \begin{bmatrix} ay + bz \\ ax + by + cz \\ bx + cy + dz \\ cx + dy \end{bmatrix}$$

Example: 1D conv, kernel size=3, stride=1, padding=1

“transposed convolution” is a convolution!

We can express convolution in terms of a matrix multiplication

$$\vec{x} * \vec{a} = X\vec{a}$$

$$\begin{bmatrix} x & y & z & 0 & 0 & 0 \\ 0 & x & y & z & 0 & 0 \\ 0 & 0 & x & y & z & 0 \\ 0 & 0 & 0 & x & y & z \end{bmatrix} \begin{bmatrix} 0 \\ a \\ b \\ c \\ d \\ 0 \end{bmatrix} = \begin{bmatrix} ay + bz \\ ax + by + cz \\ bx + cy + dz \\ cx + dy \end{bmatrix}$$

Example: 1D conv, kernel size=3, stride=1, padding=1

Convolution transpose multiplies by the transpose of the same matrix:

$$\vec{x} *^T \vec{a} = X^T\vec{a}$$

$$\begin{bmatrix} x & 0 & 0 & 0 \\ y & x & 0 & 0 \\ z & y & x & 0 \\ 0 & z & y & x \\ 0 & 0 & z & y \\ 0 & 0 & 0 & z \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} ax \\ ay + bx \\ az + by + cx \\ bz + cy + dx \\ cz + dy \\ dz \end{bmatrix}$$

“transposed convolution” is a convolution!

We can express convolution in terms of a matrix multiplication

$$\vec{x} * \vec{a} = X\vec{a}$$

$$\begin{bmatrix} x & y & z & 0 & 0 & 0 \\ 0 & x & y & z & 0 & 0 \\ 0 & 0 & x & y & z & 0 \\ 0 & 0 & 0 & x & y & z \end{bmatrix} \begin{bmatrix} 0 \\ a \\ b \\ c \\ d \\ 0 \end{bmatrix} = \begin{bmatrix} ay + bz \\ ax + by + cz \\ bx + cy + dz \\ cx + dy \end{bmatrix}$$

Example: 1D conv, kernel size=3, stride=1, padding=1

Convolution transpose multiplies by the transpose of the same matrix:

$$\vec{x} *^T \vec{a} = X^T\vec{a}$$

$$\begin{bmatrix} x & 0 & 0 & 0 \\ y & x & 0 & 0 \\ z & y & x & 0 \\ 0 & z & y & x \\ 0 & 0 & z & y \\ 0 & 0 & 0 & z \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} ax \\ ay + bx \\ az + by + cx \\ bz + cy + dx \\ cz + dy \\ dz \end{bmatrix}$$

When stride=1, convolution transpose is just a regular convolution (with different padding rules)

Slide Credit: Fei-Fei Li, Justin Johnson, Serena Yeung, CS 231n

Plan for Today

- Convolutional Neural Networks
  - Transposed convolutions

- Visualizing CNNs

Story from Summer 2017…

Facebook Shut Down AI After It Invented Its Own Language

By **NTD Television** | July 29, 2017 AT 3:01 PM Last Updated: August 6, 2017 12:37 pm

"Han the Robot" at a discussion about the future of humanity in a demonstration of artificial intelligence at the RISE

More in Tech News

Google Fires Worker Who Exposed Discrimination, Gags Free Speech

Future scary? Facebook AI bots created own language that humans couldn't understand

SHOBHIT VARMA |

@SHOBHIT_TECH

LAST UPDATED: AUG 02, 2017

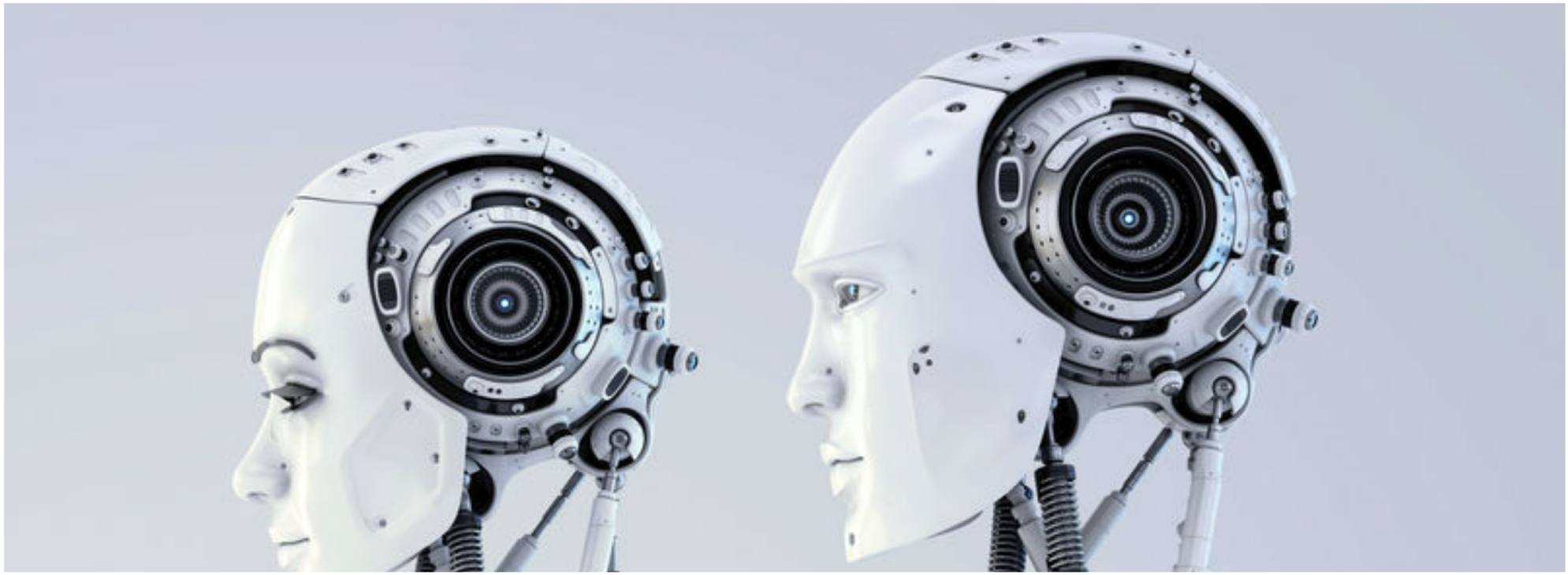
NEW DELHI

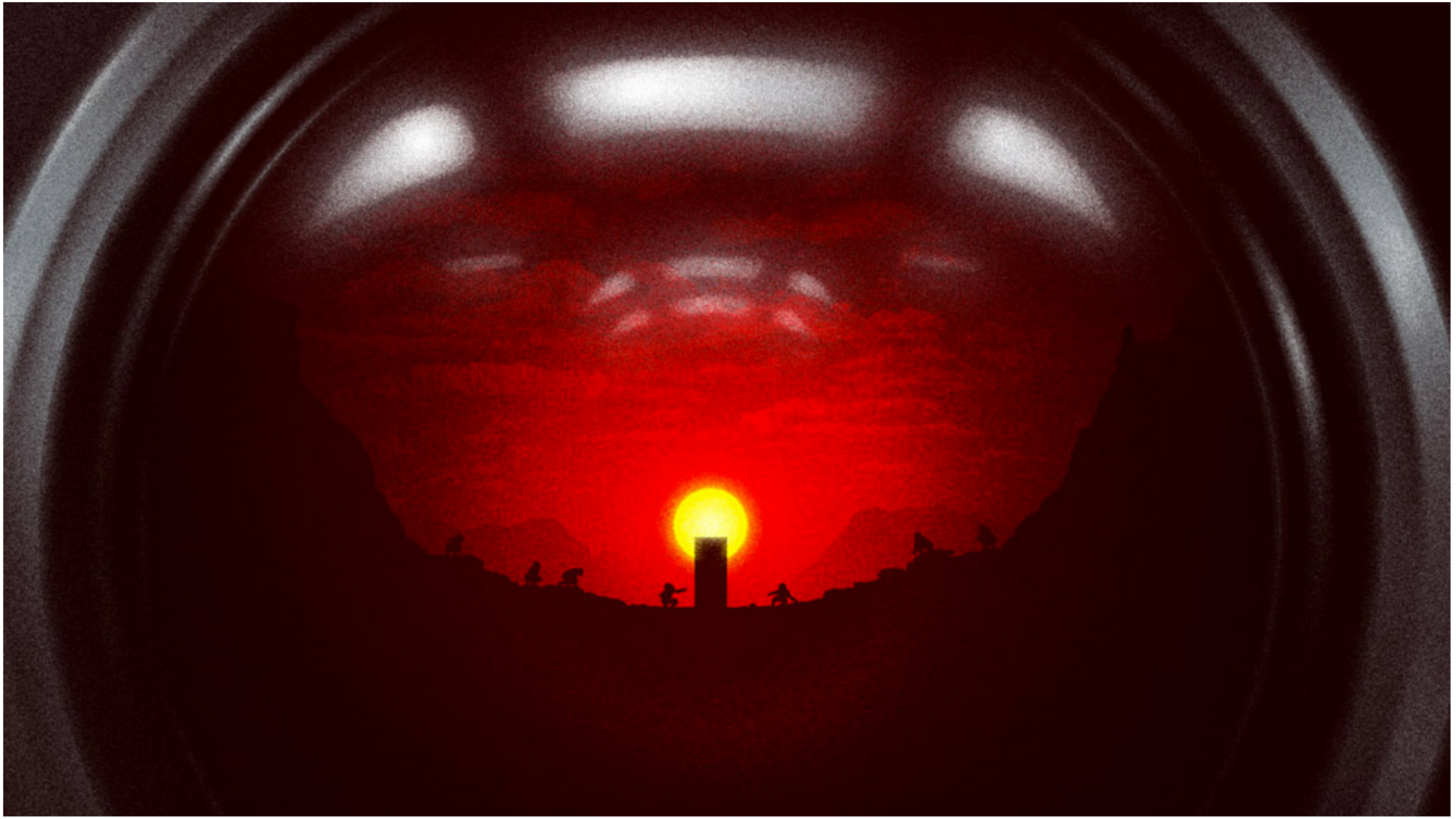
EMAIL AUTHOR

criticized AI naysayers.

Facebook shuts down AI after it invents its own creepy language

Phillip Tracy—July 31 at 9:17PM | Last updated July 31 at 9:18PM



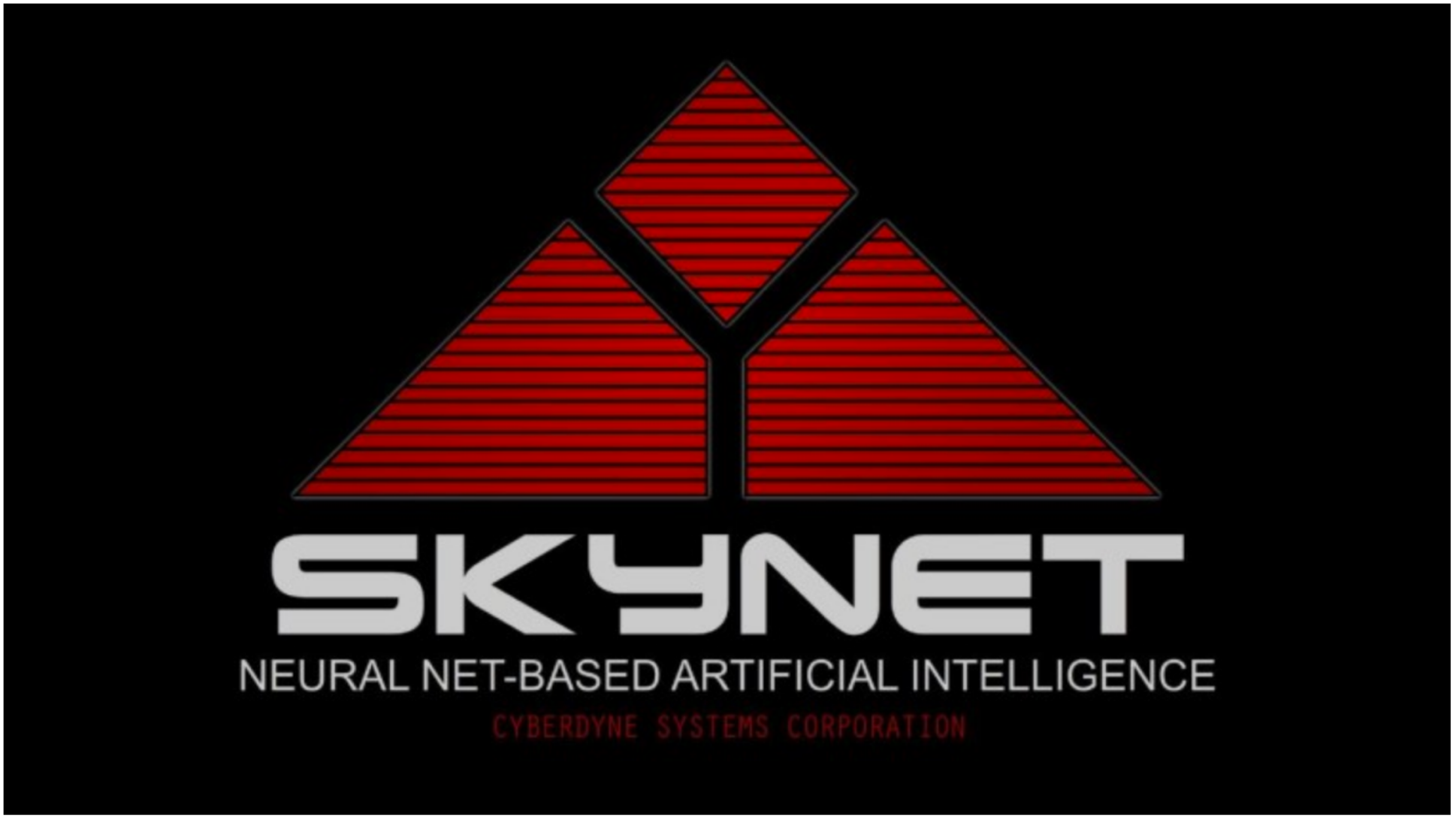


FACEBOOK KILLED AN AI AFTER IT CAME UP WITH ITS OWN LANGUAGE

Facebook's language-creating A.I. shut down because it didn't work as intended

By Chris Schroeder · Aug 1, 2017

11



DAVE ROSS

Did we just barely escape the robot takeover?

BY DAVE ROSS
AUGUST 1, 2017 AT 9:08 AM

Share

(File, Andy Tullis/The Bulletin, via AP)

Most Popular

Gorka: North Korea's latest threat 'an exercise in blackmail'

How anyone can easily become an addict

Report: More bike share companies coming to Seattle

Sign up for the Most Popular email

Artificial Intelligence Invents New Language Humans Can't Understand

by Marykate Jasper | 11:00 am, July 29th, 2017 80

Build your bunkers now, homo sapiens.

Deal or No Deal? End-to-End Learning for Negotiation Dialogues

[EMNLP '17]

Mike Lewis
(FAIR)

Denis Yarats
(FAIR)

Yann Dauphin
(FAIR)

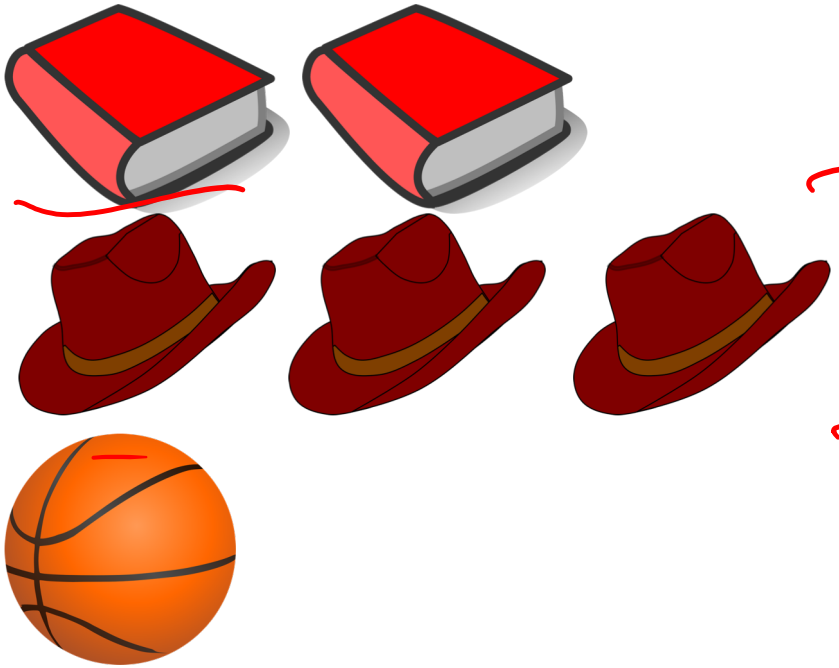
Devi Parikh
(Georgia Tech / FAIR)

Dhruv Batra
(Georgia Tech / FAIR)

Object Division Task

Agents shown *same* set of object but *different* values for each

Asked to agree how to divide objects between them



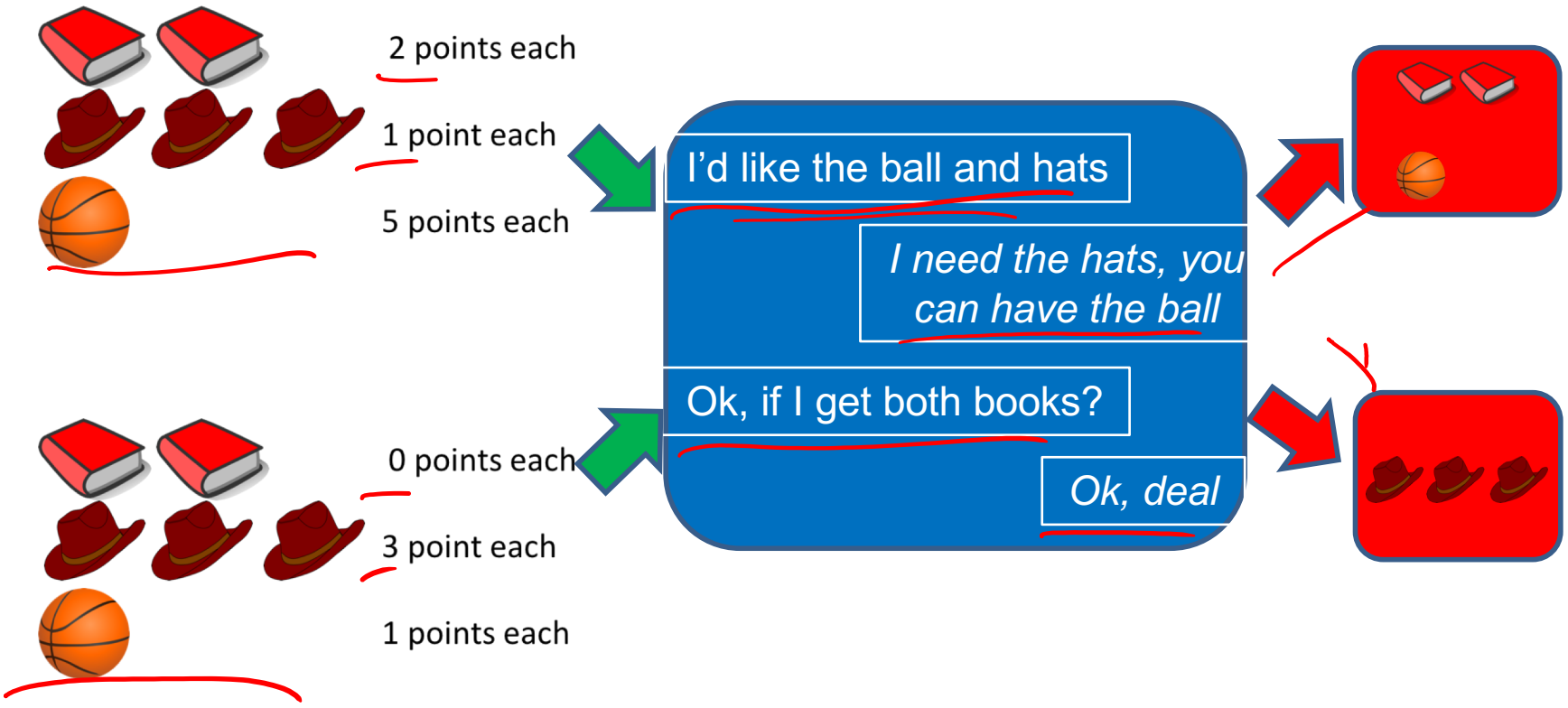
2 points each

1 point each

5 points each

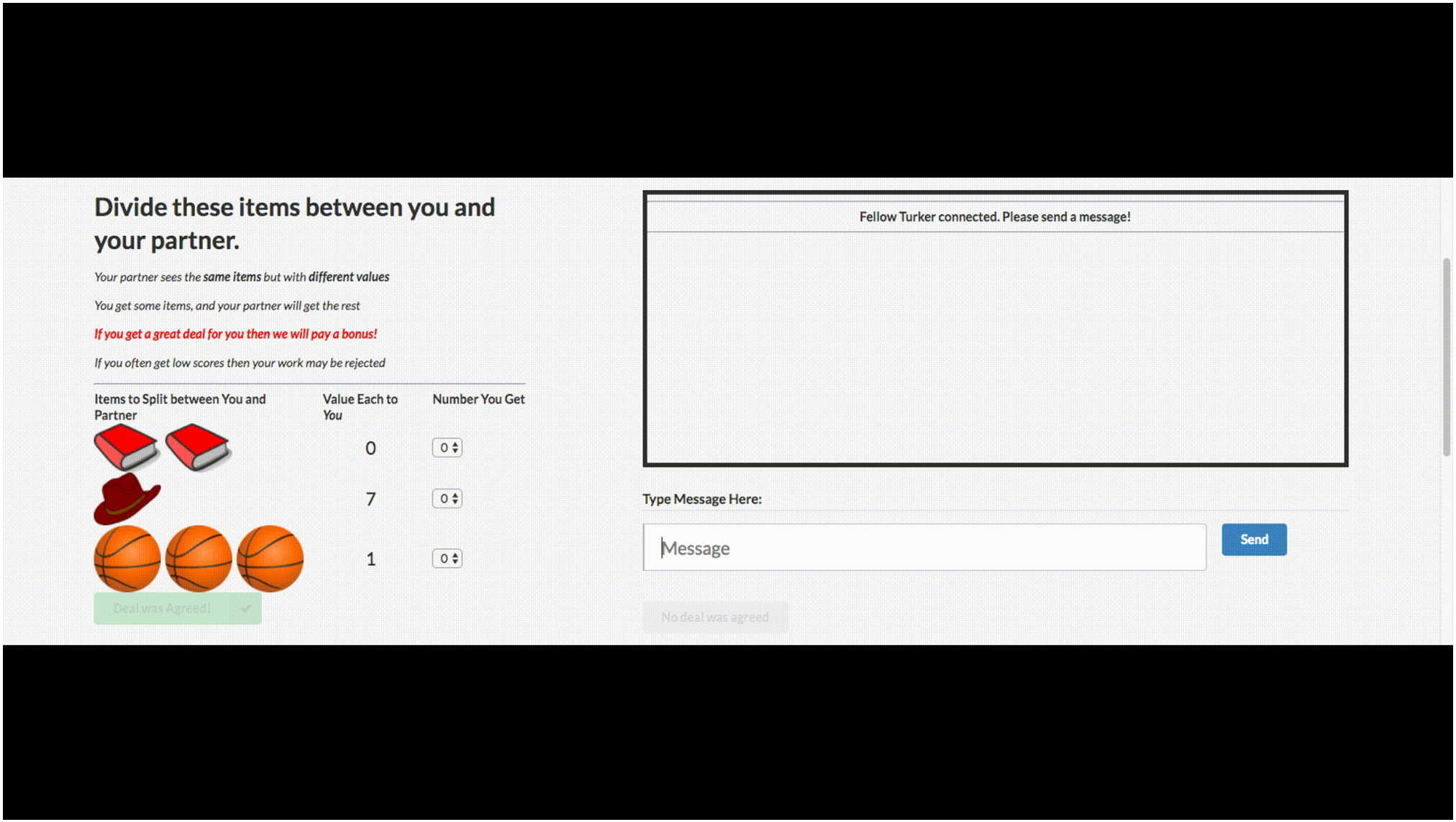
Slide Credit: Mike Lewis

Multi-Issue Bargaining



Slide Credit: Mike Lewis

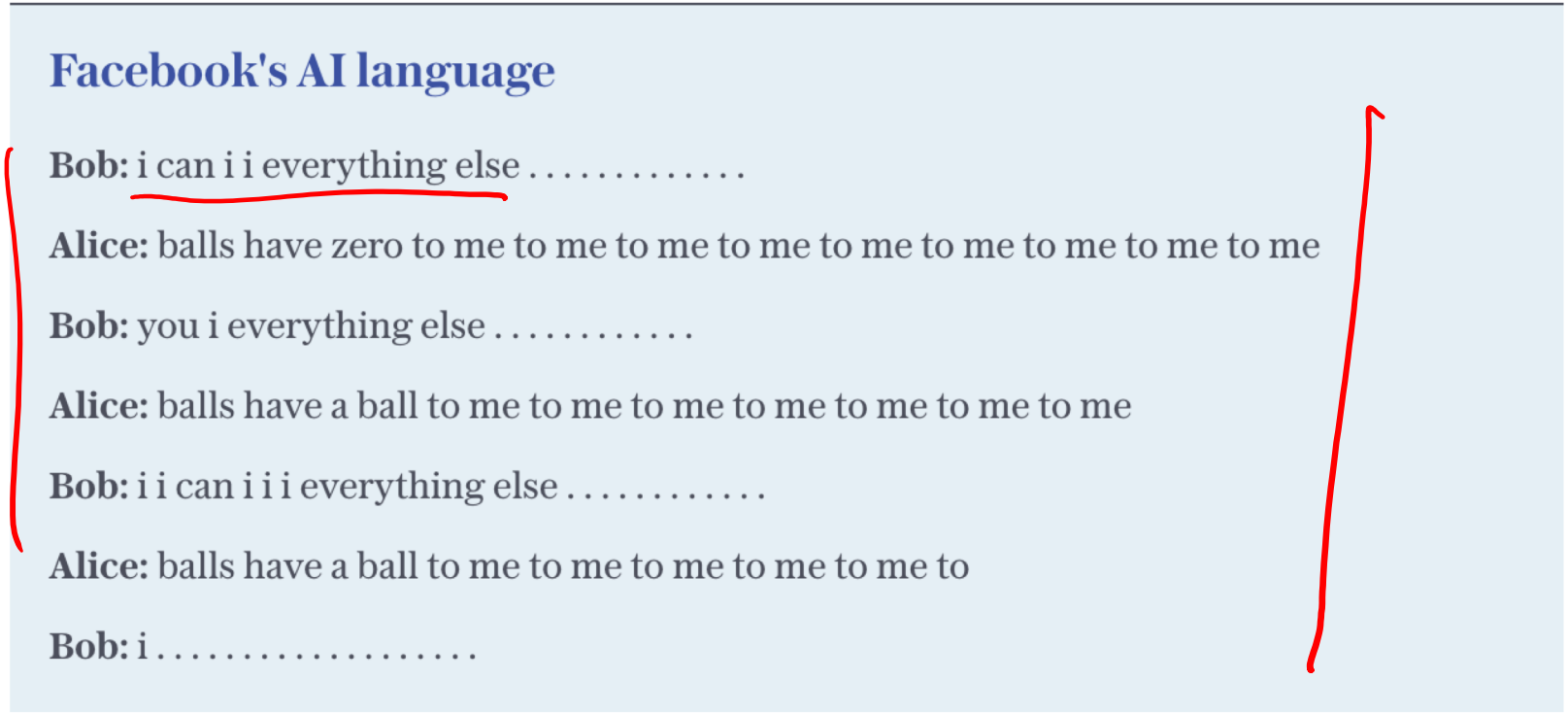
Data Collection on AMT



Slide Credit: Mike Lewis

What happened?

- *Results …show that the simple LIKELIHOOD model produces the most human-like responses, and the alternative training and decoding strategies cause a divergence from human language.*



Facebook's AI language

Bob: i can i i everything else . . . . . .

Alice: balls have zero to me to me to me to me to me to me to me to me to me

Bob: you i everything else . . . . . .

Alice: balls have a ball to me to me to me to me to me to me to me

Bob: i i can i i i everything else . . . . . .

Alice: balls have a ball to me to me to me to me to me to

Bob: i . . . . . .

No, Facebook Did Not Panic And Shut Down An AI Program That Was Getting Dangerously Smart

Tom McKay
Aug 1, 2017, 12:30pm · Filed to: ai

BBC Sign in News Sport Weather Shop Earth Travel M

NEWS

Home Video World Asia UK Business Tech Science Magazine Entertainme

Technology

The 'creepy Facebook AI' story that captivated the media

By Chris Baraniuk
Technology reporter

1 August 2017 | Technology

Share

INDEPENDENT News InFact Politics Voices Indy/Life Business Sport Tech Culture Subscribe

Voices

This is how Facebook's shut-down AI robots developed their own language – and why it's more common than you think

None of this is a huge surprise: the neural network robots behind Google Translate made up their own language long ago which is 'not readable or usable for humans'

Fact Check › Technology

Did Facebook Shut Down an AI Experiment Because Chatbots Developed Their Own Language?

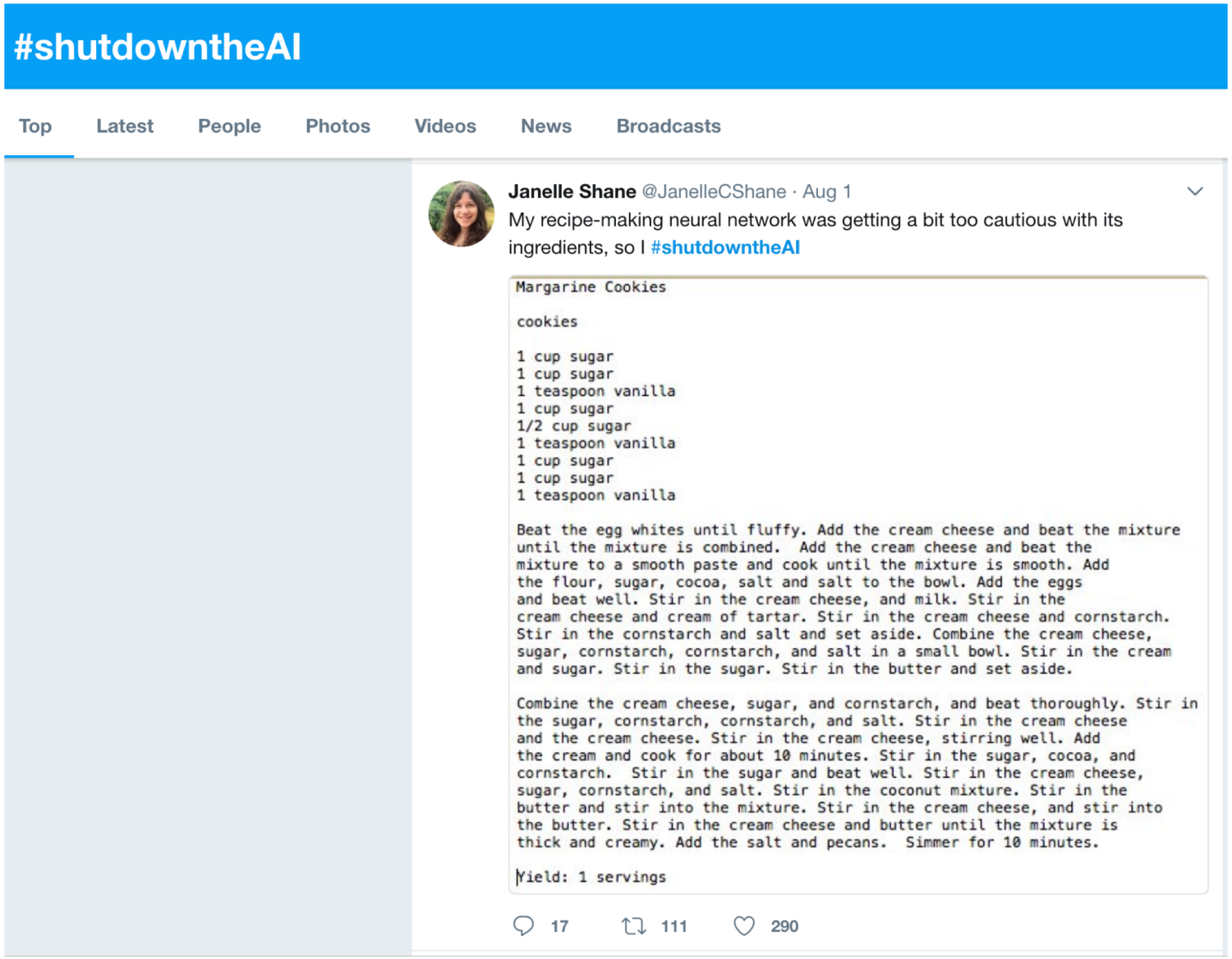
CLAIM

Concerned artificial intelligence researchers hurriedly abandoned an experimental chatbot program after they realized that the bots were inventing their own language.

RATING

ORIGIN

#ShutDownTheAI



Takeaway for us

People don't trust
what they don't understand

Stories from last year…

Self-Driving Tesla Was Involved in Fatal Crash, U.S. Says

By **BILL VLASIC** and **NEAL E. BOUDETTE** JUNE 30, 2016

The New York Times

A Tesla Model S, with its self-driving mode enabled. In a statement, the National Highway Traffic Safety Administration said it had sent an investigative team to examine the vehicle and the crash site in Williston, Fla. Jasper Juinen/Bloomberg

Stories from last year…

AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol

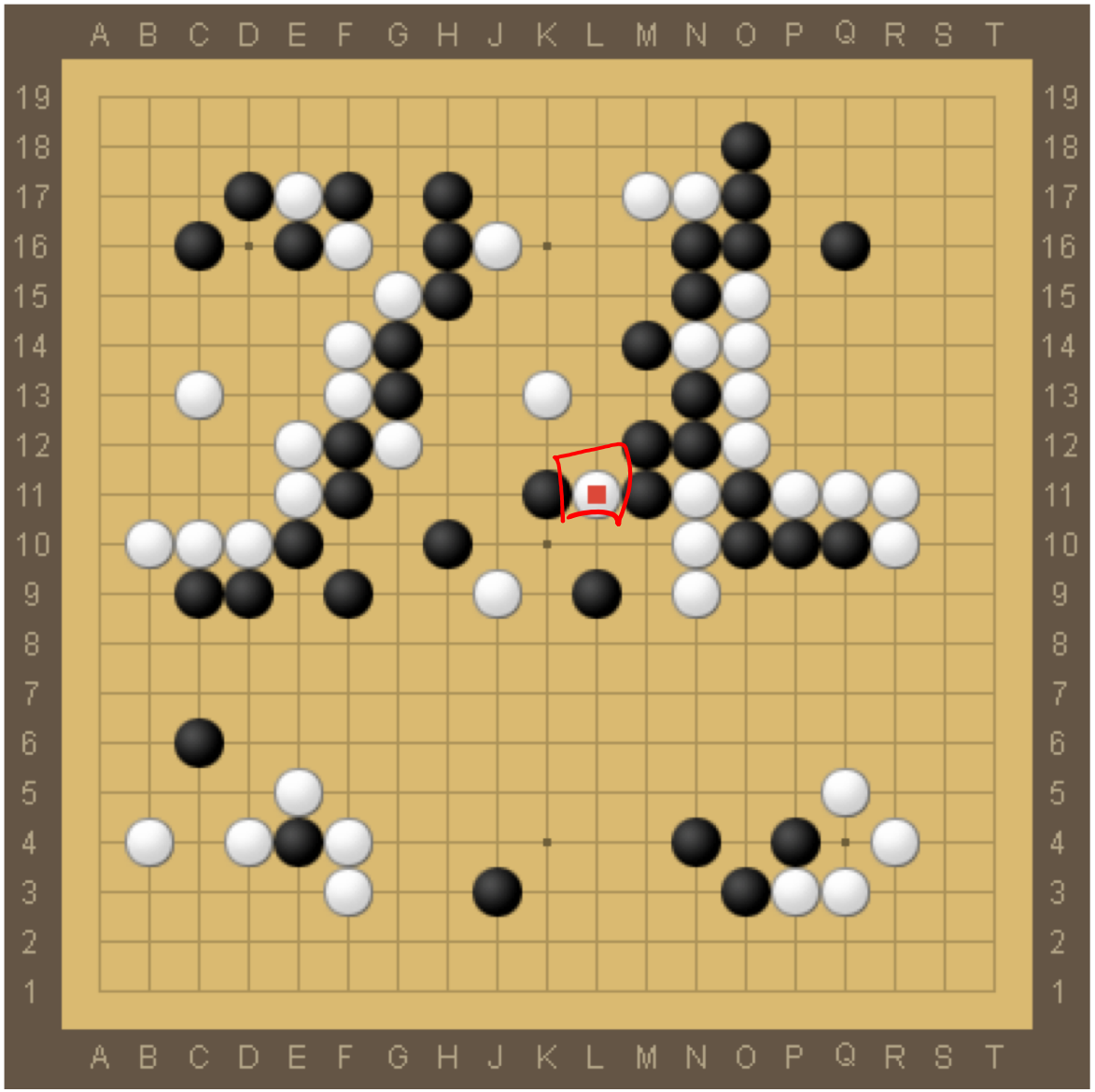
DeepMind's artificial intelligence astonishes fans to defeat human opponent and offers evidence computer software has mastered a major challenge

The world's top Go player, Lee Sedol, lost the final game of the Google DeepMind challenge match. Photograph: Yonhap/Reuters

Google DeepMind's AlphaGo program triumphed in its final game against South Korean Go grandmaster Lee Sedol to win the series 4-1, providing further evidence of the landmark achievement for an artificial intelligence program.

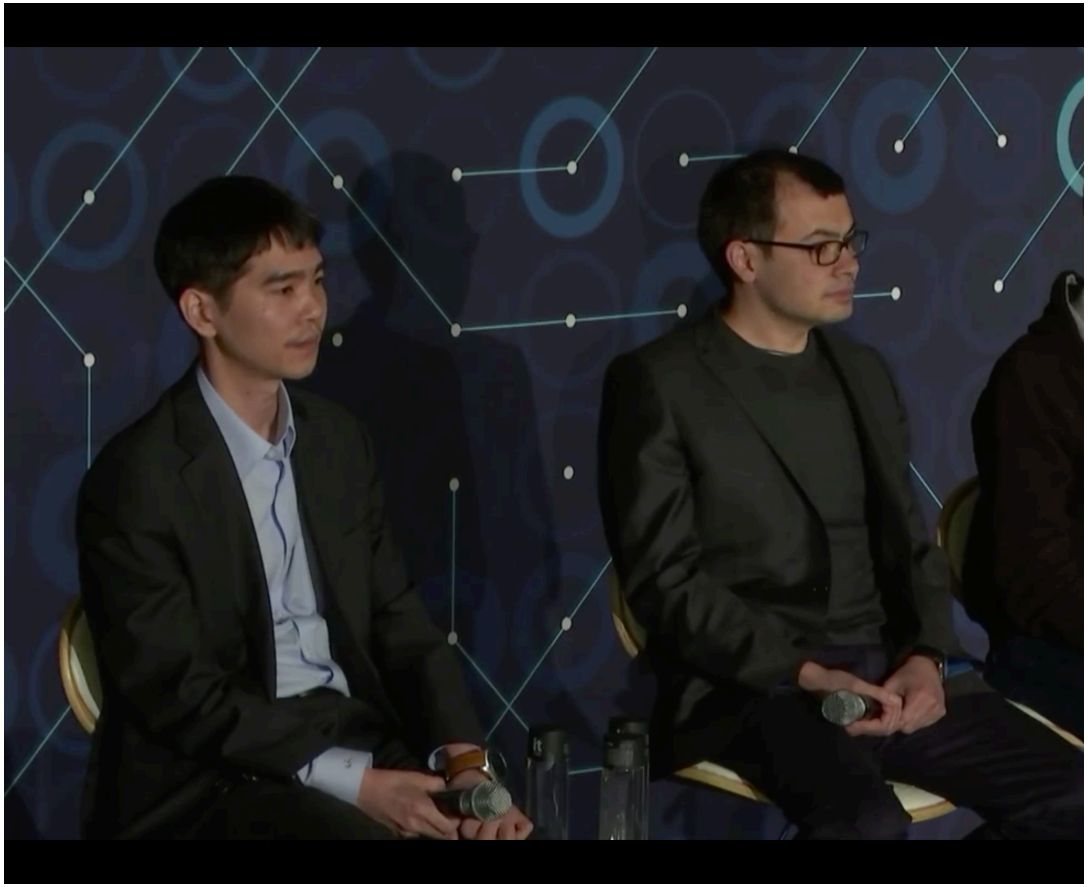
Stories from last year…

- AlphaGo vs Lee Sedol, Match 4, Move 78



Stories from last year…

- AlphaGo vs Lee Sedol, Match 4, Press Conference



…Story from 1980s

- Tanks vs No-Tanks, or Sunny vs Cloudy?

Explainable AI

Why does an intelligent system do what it does?

Explainable AI

a data subject has the right to

"an explanation of the decision reached after [algorithmic] assessment."

EU regulations on algorithmic decision-making and a "right to explanation"

Bryce Goodman BRYCE.GOODMAN@STX.OX.AC.UK
Oxford Internet Institute, Oxford

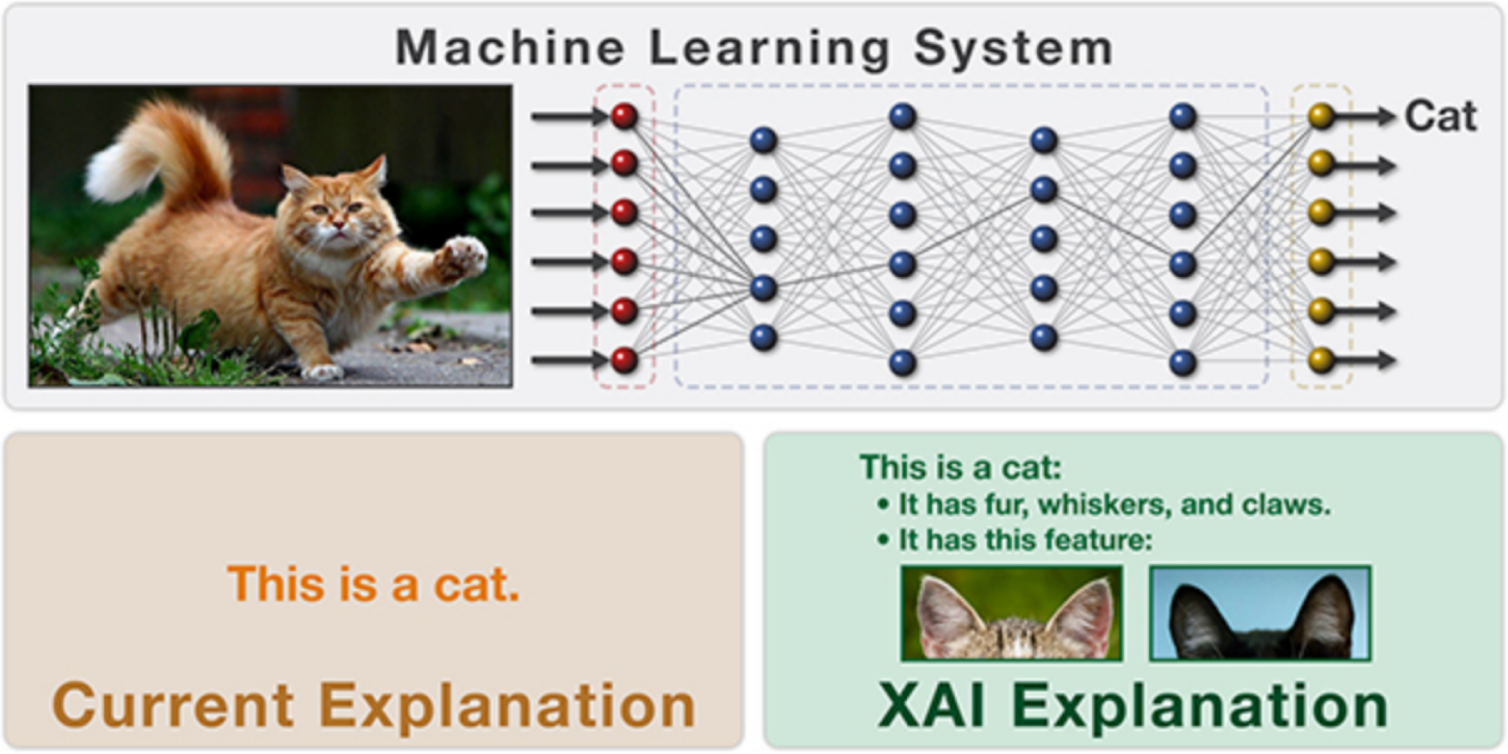
Seth Flaxman FLAXMAN@STATS.OX.AC.UK
Department of Statistics, Oxford

DARPA XAI

Defense Advanced Research Projects Agency > Program Information

Explainable Artificial Intelligence (XAI)

Mr. David Gunning



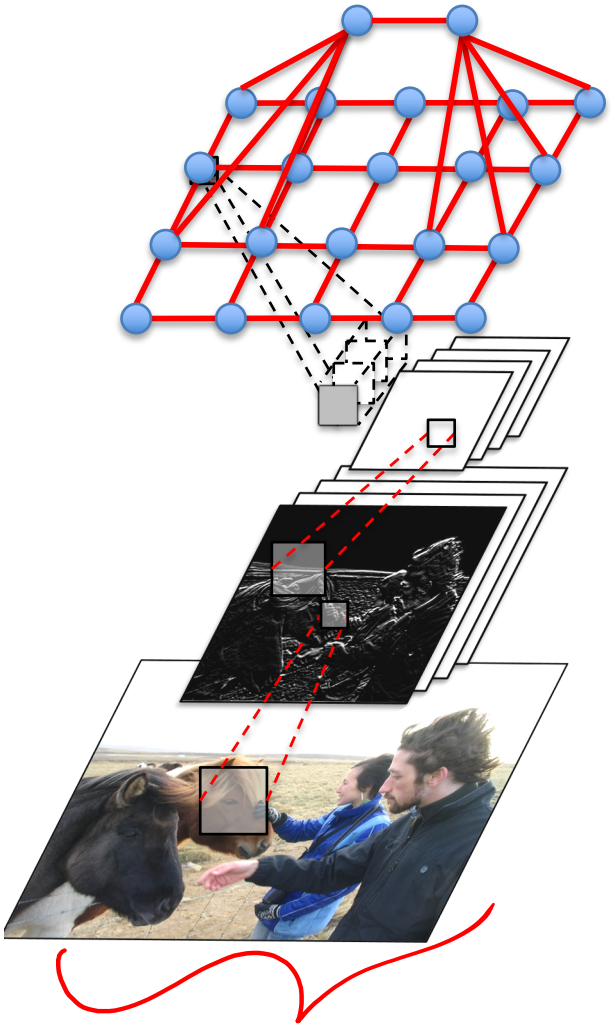
Explainable AI

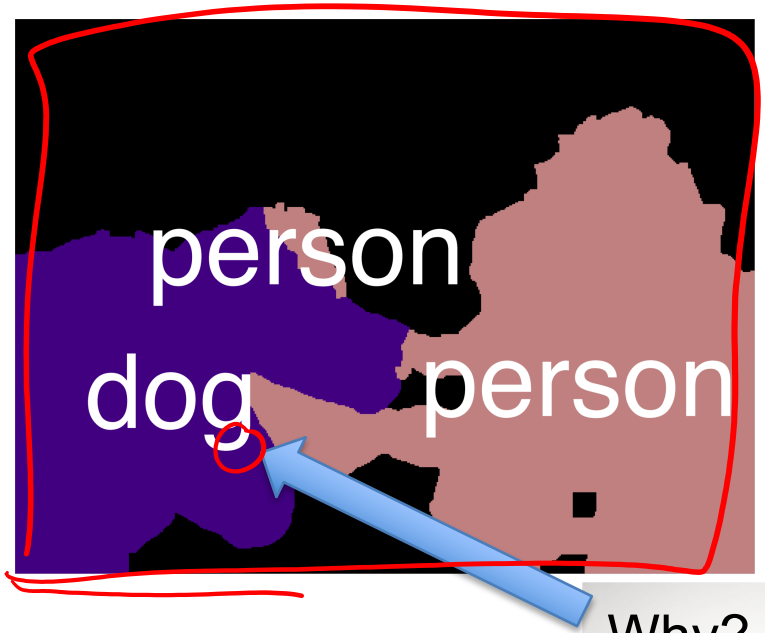
- *Why does an intelligent system do what it does?*

Explaining a prediction

Explaining a model

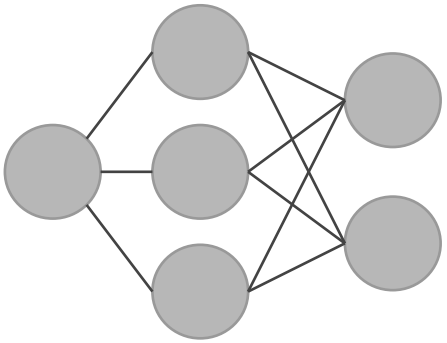
Explaining/Trusting a Prediction





Explaining/Trusting a Model

- *Should I put this into production?*



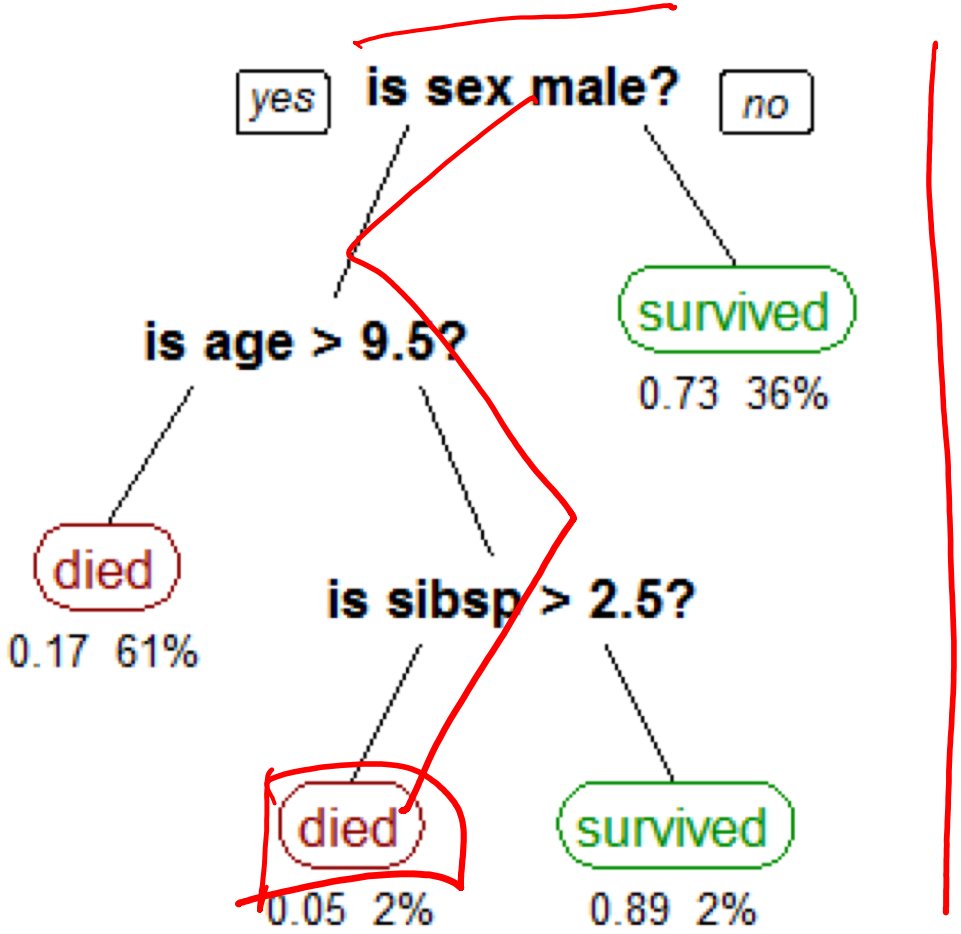
- Usually related to trusting most predictions

What do we do today for trust?

Slide Credit: Marco Ribeiro, Sameer Singh

1. Use simple/interpretable models

- Can be great, but usually at the cost of:
  - Accuracy
  - Flexibility
  - Speed



Slide Credit: Marco Ribeiro, Sameer Singh

2. Dark Magic

- “Trust me” / Gut Feeling
  - "I looked at a few examples and it seems to work”
  - "I've done this before"

“Trust me child”

“I know how the brain works”

Explainable AI

- When AI is weaker than humans
  - Transparency = finding error modes
  - Goal = improving machines
- When AI is at par with humans
  - Transparency = providing rationales
  - Goal = building trust with humans
- When AI is stronger than humans
  - Transparency = explaining a complicated function
  - Goal = teaching humans

The Kyunghyang Shinmun

Home Opinions Editorials Articles Korean Wave Latest news

Home > General News

Lee Se-dol, "Learned a Lot from AlphaGo": Lessons Also Applied in the "Human World" of Go

By Eom Min-yong

Posted on : 2016-05-03 18:17

Zoom

Machines Teaching Humans

- Park Yeong-hun, who was recently defeated by Lee Sedol in the Maxim Cup semifinal match said:

  "*Sedol surprised people by copying the moves of AlphaGo in the Ing Cup*….

  It appears he has learned something from the five matches against AlphaGo,

  and his game of go, which was strong to begin with, seems stronger because of that."

Explainable AI: Visual Explanations

Where does an intelligent system "look" to make its predictions?