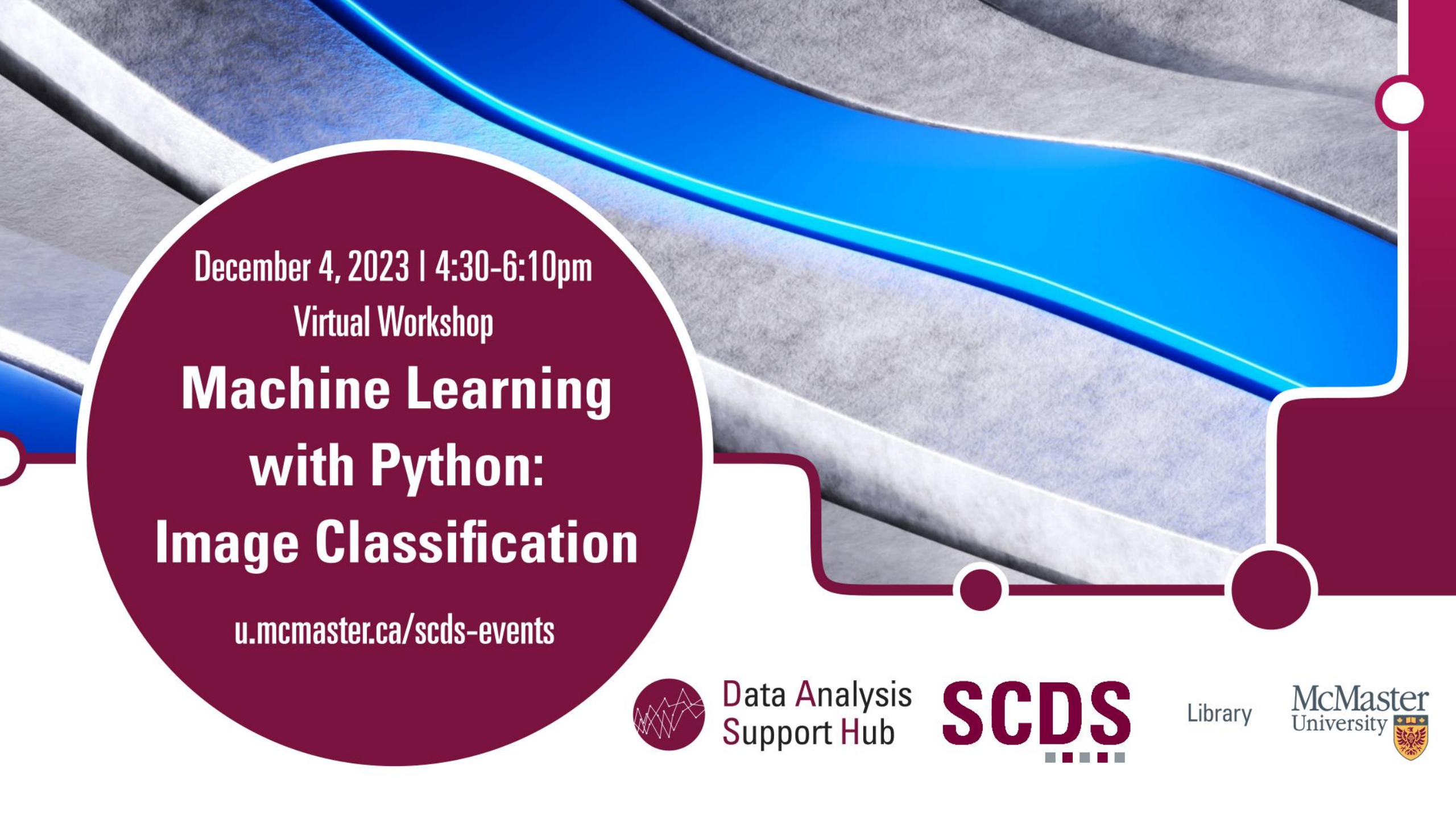


December 4, 2023 | 4:30-6:10pm

Virtual Workshop

Machine Learning with Python: Image Classification

u.mcmaster.ca/scds-events

Data Analysis Support Hub

SCDS

Library

McMaster University

Machine Learning with Python: Image classification

Seyed Amirreza Mousavi
Master's student at McMaster University

DASH: Data Analysis Support Hub Workshop Series
4/12/2023

Lewis & Ruth Sherman Centre for Digital Scholarship
scds.ca

McMaster University Library

McMaster University is located on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

Laslovarga, “Webster Falls in Winter, Waterdown, Hamilton, Ontario, Canada - Spencer Gorge / Webster's Falls Conservation Area,” 23 January 2011, Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Waterdawn_Webster_Falls_in_Winter8.jpg

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information: scds.ca/events/code-of-conduct/

Certificate Program

The Sherman Centre offers a Certificate of Attendance that rewards synchronous participation at 7 workshops. We also offer concentrations in Data Analysis and Visualization, Digital Scholarship, and Research Data Management.

Learn more about the Certificate Program: https://scds.ca/certificate-program

Verify your participation at a session: https://u.mcmaster.ca/verification

At an unspecified point during the workshop, a code will be read aloud. This is the answer to the third question of the form.

Book an Appointment with the DASH Team

Receive help from a member of the DASH team! DASH can assist with the following topics:

- Creating data visualizations, including charts, graphs, and scatter plots
- Figuring out which statistical tests to run (e.g., t-test, chi-square, etc.).
- Analyzing data with software including SPSS, Python, R, SAS, ArcGIS, MATLAB, and Excel
- Choosing which software package to use, including free and open-source software
- Troubleshooting problems related to file formats, data retrieval, and download
- Selecting methodology and type of data analysis to use in a thesis project

Book an appointment: https://library.mcmaster.ca/services/dash

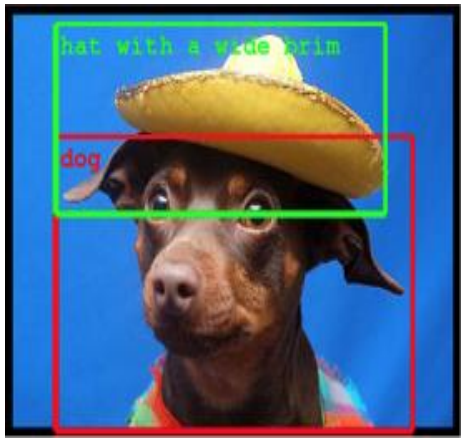
Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Why Deep Learning?

End-to-End Learning for Many Tasks



vision

speech

text

control

Deep Learning for Vision:

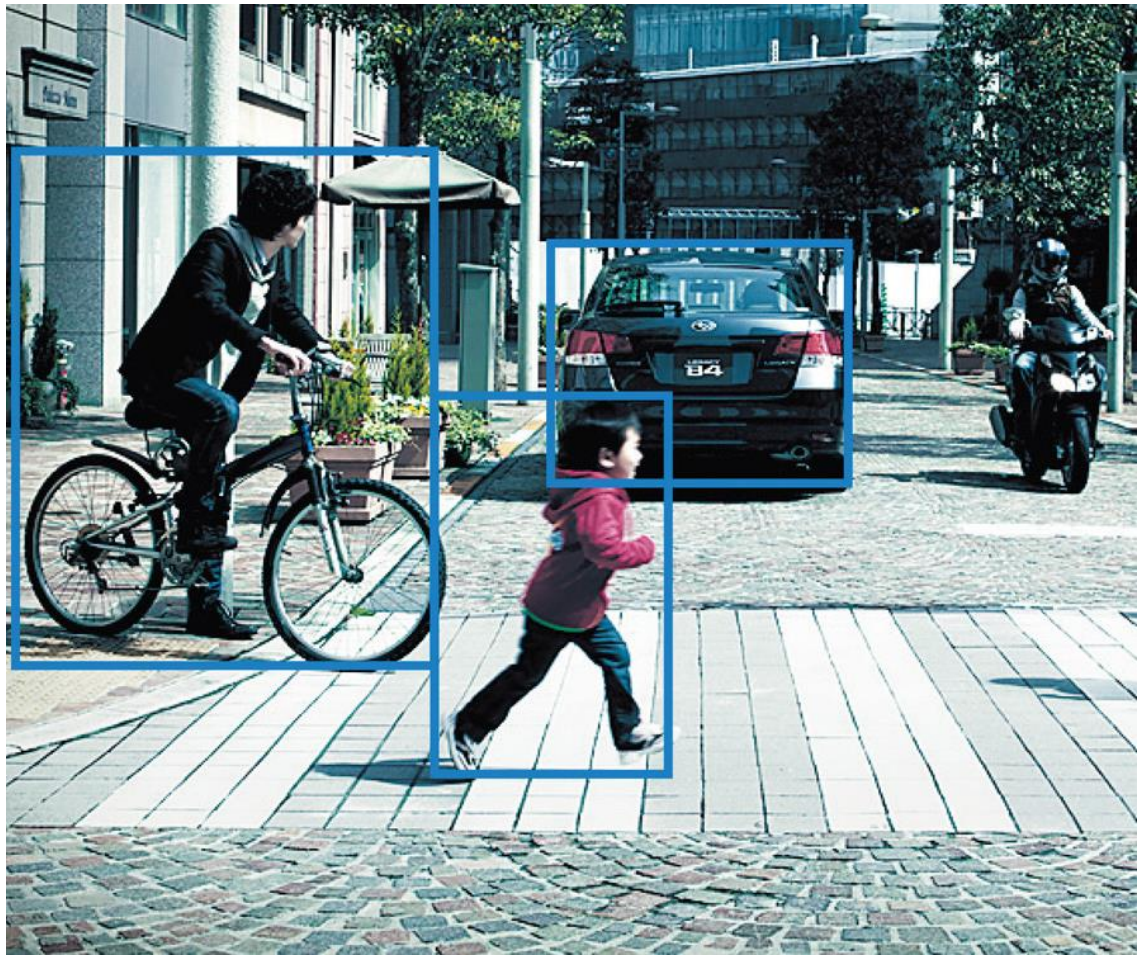
Machines are useful mainly to the extent that they **interact with the physical world**

Visual information is the richest source of information about the real world

Vision is the highest-bandwidth mode for machines to obtain real-world info

Embedded vision enables our things to be:

- **More responsive**
- **More personal and secure**
- **Safer, more autonomous**
- **Easier to use**



Visual Recognition Tasks

Classification

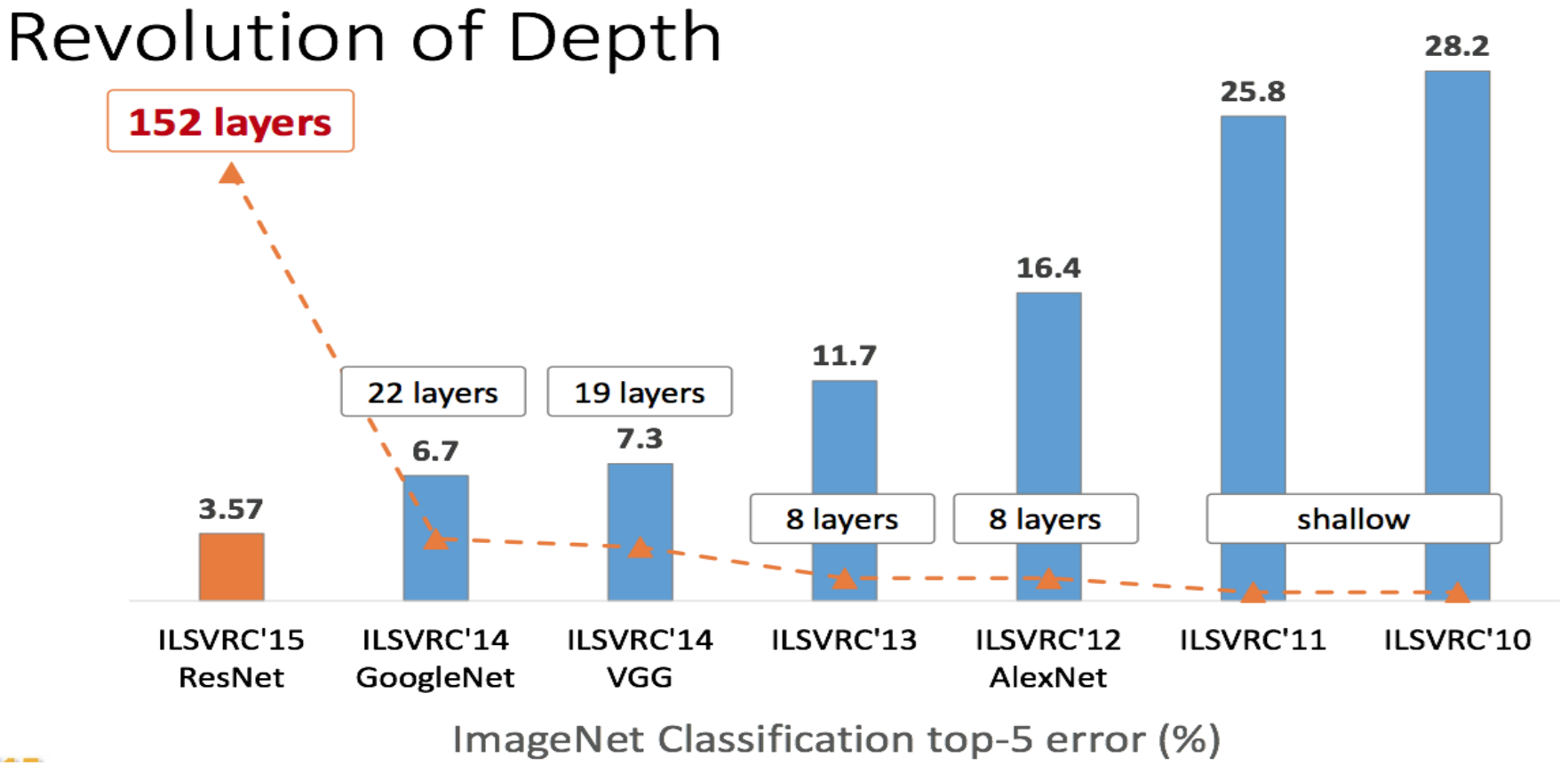
- what kind of image?
- which kind(s) of objects?

Challenges

- appearance varies by lighting, pose, context, ...
- clutter
- fine-grained categorization (horse or exact species)

- ❑ dog
- ☑ car
- ☑ horse
- ❑ bike
- ❑ cat
- ❑ bottle
- ☑ person

Image Classification: ILSVRC 2010-2015

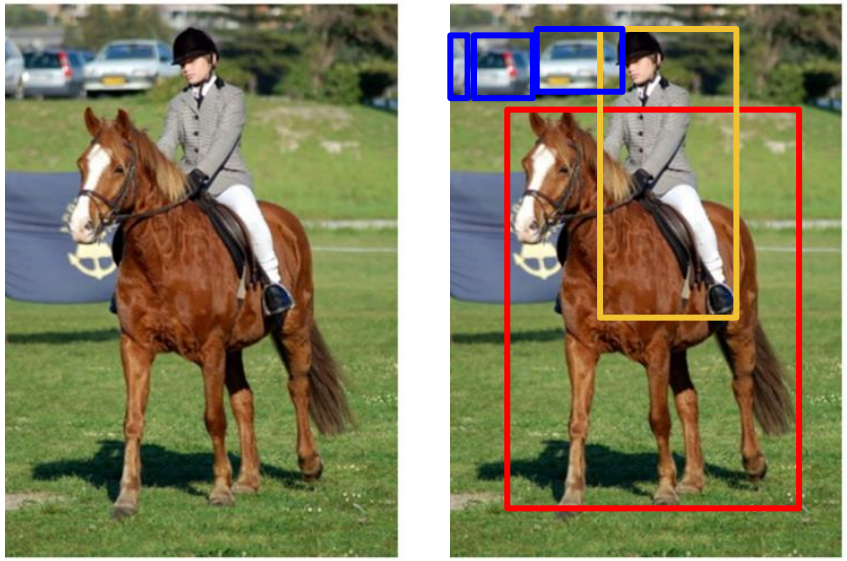


- ☐ dog
- ☑ car
- ☑ horse
- ☐ bike
- ☐ cat
- ☐ bottle
- ☑ person

Visual Recognition Tasks

Detection

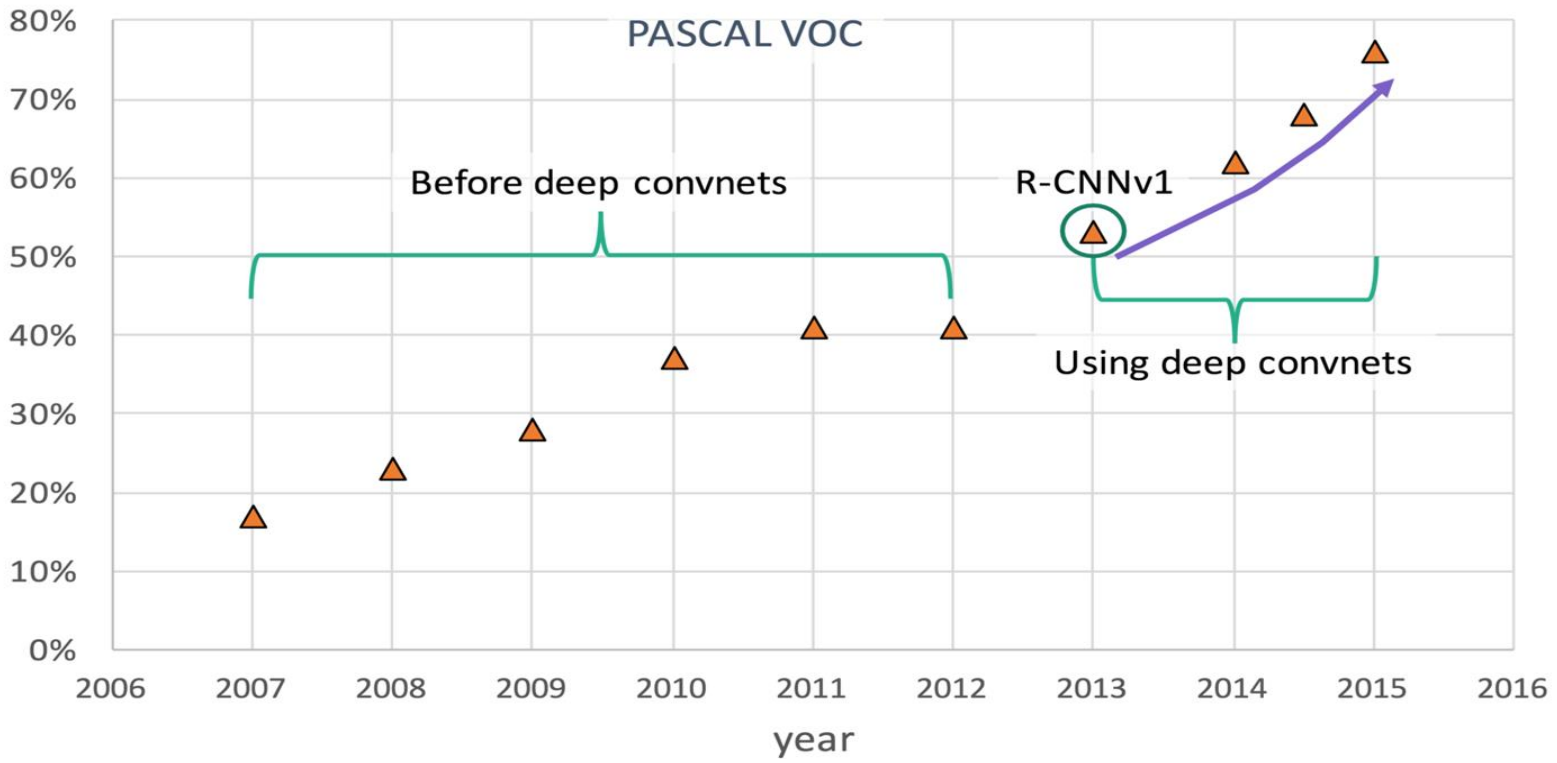
- what objects are there?
- where are the objects?

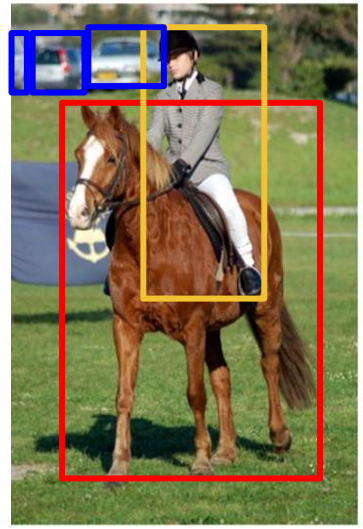


Challenges

- localization
- multiple instances
- small objects

Detection


PASCAL VOC
80%
70%
60%
50%
40%
30%
20%
10%
0%
Before deep convnets
R-CNNv1
Using deep convnets
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
year



Visual Recognition Tasks

Semantic Segmentation

- what kind of thing
  is each pixel part of?
- what kind of stuff
  is each pixel?

Challenges

- tension between
  recognition and localization
- amount of computation

Deep Learning is

Stacking **Layers**

and

Learning **End-to-End**

Stacking Layers

A **layer** is a transformation

$x' = \mathbf{layer}(x)$

Deep networks are layered models made by stacking different types of transformation

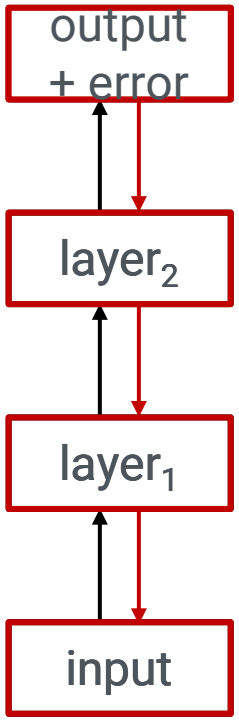
$x_2 = \mathbf{layer_1}(x_1)$

$x_3 = \mathbf{layer_2}(x_2)$

...

How do layers stack?

Layered Networks



Networks run layer-by-layer, composing the input-output transformation of each layer

x_1 = **layer$_1$**(input)

out = **layer$_2$**(x_1)

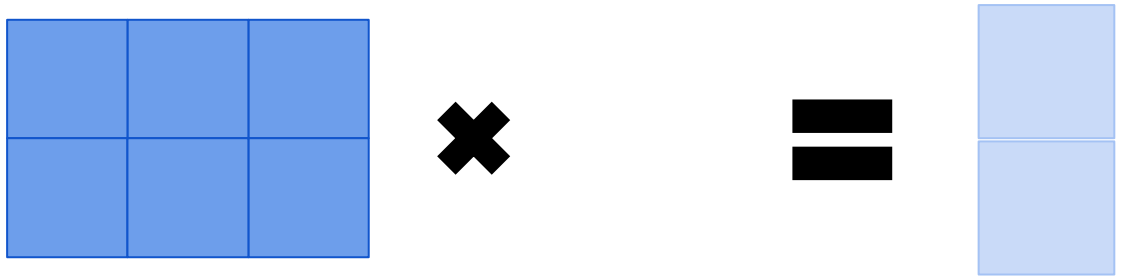
During learning, the error is passed back layer-by-layer to tune the transformations

What kind of layers should we stack?

The Simplest Layers

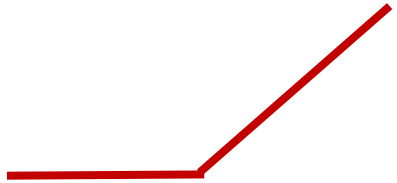
Matrix Multiplication

$$x' = Wx + b$$



Non-linearity

$$x' = \max(0, x)$$



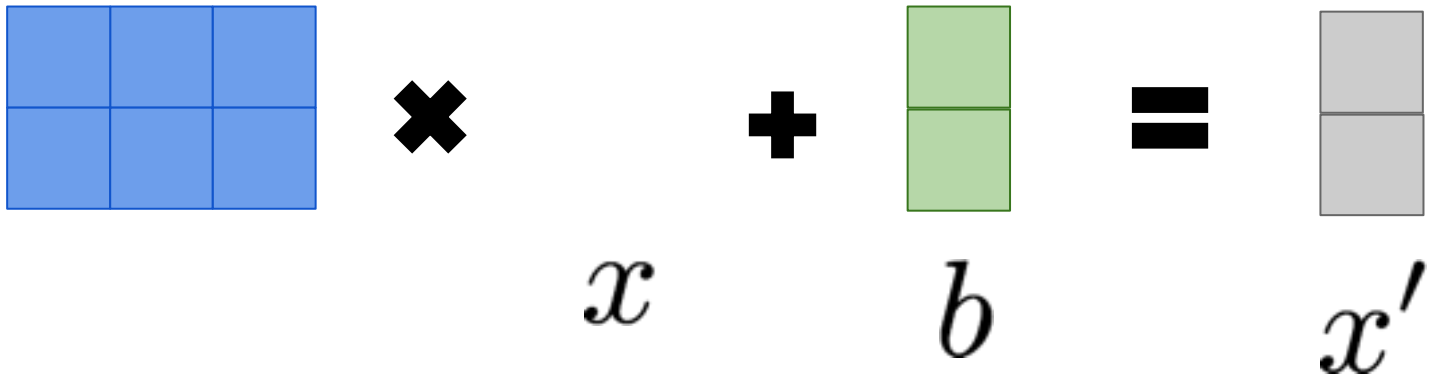
(for example)

Matrix Multiplication

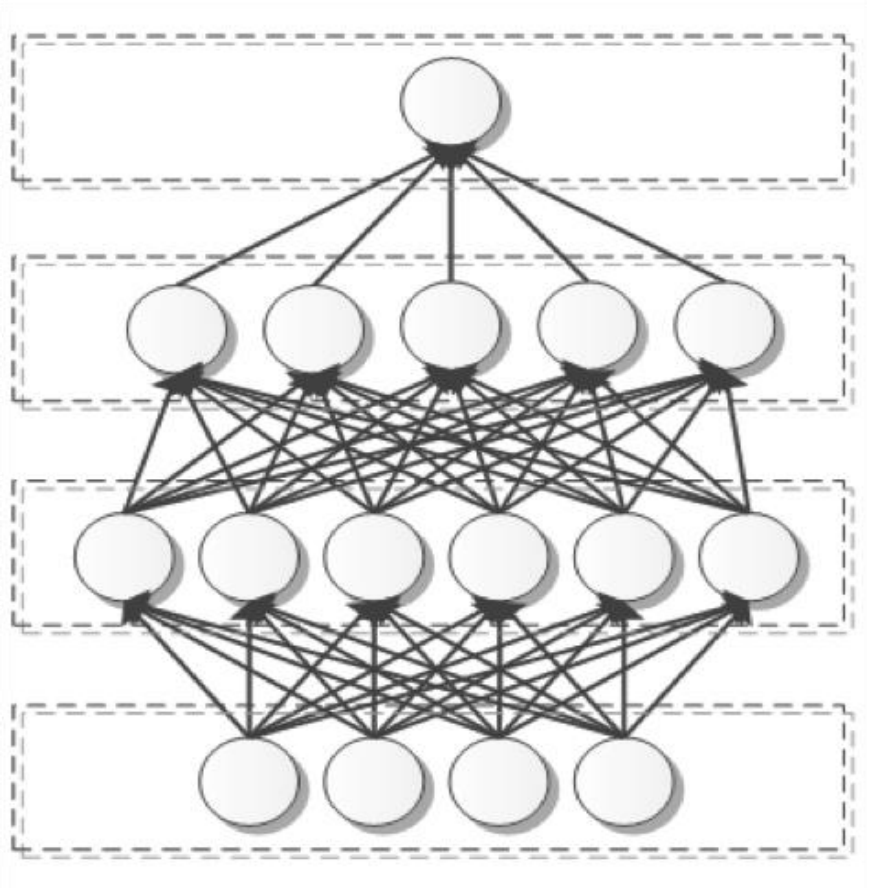
Multiply input *x* by weights *W* and add bias *b*

Learns linear transformations

$$x' = Wx + b$$



Matrix Multiplication == Fully Connected Layer



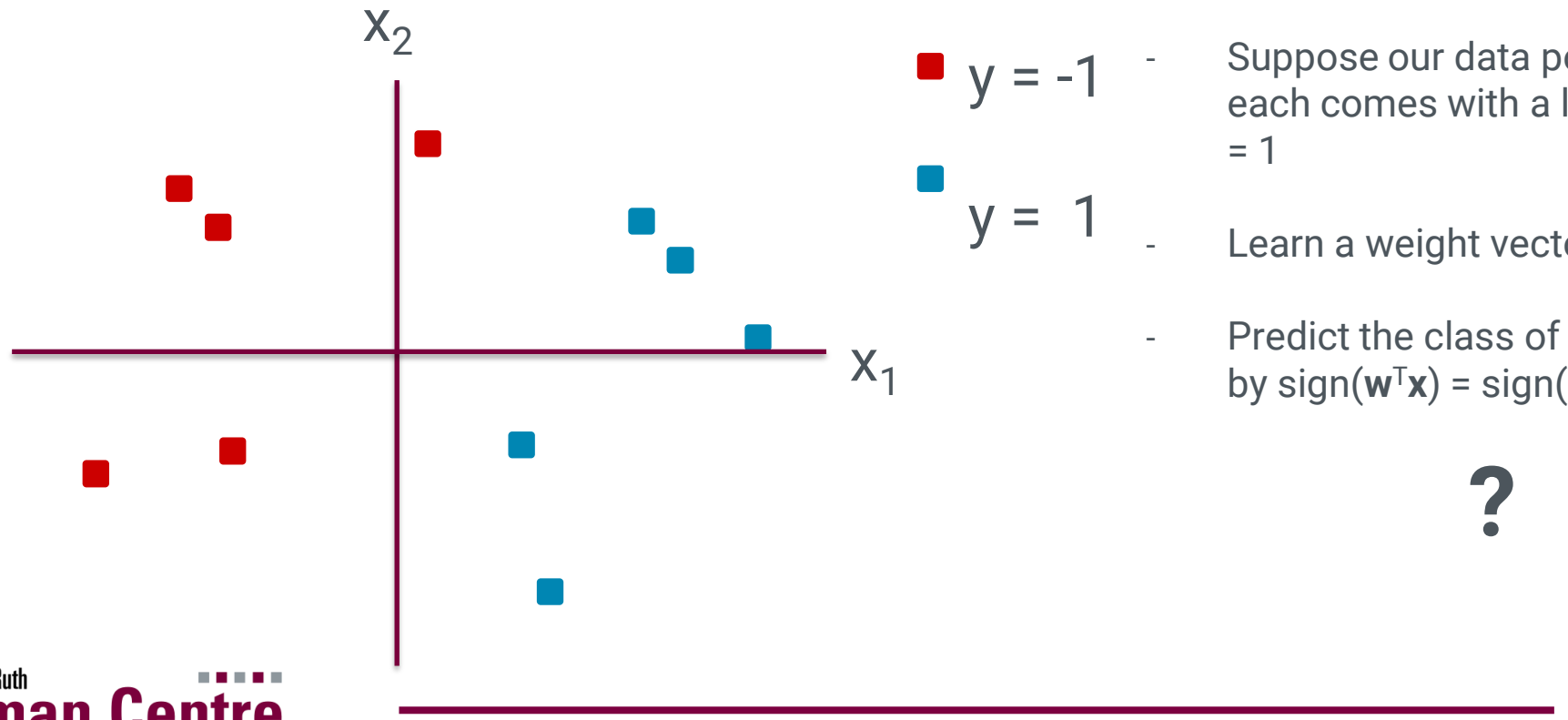
Output is a function of every input, or the input and output are "fully connected"

Abbreviated as **FC**

[figure credit BDTI]

Linear Classification

To classify we need to separate the data into red vs. blue

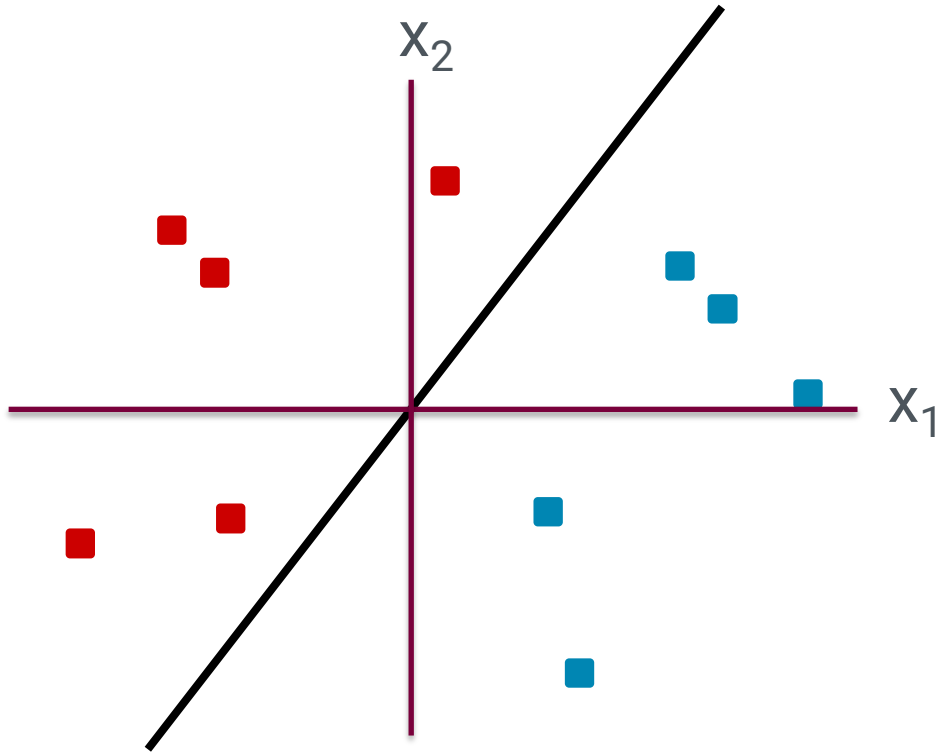


- Suppose our data points (**x**) are 2D and each comes with a label y, where y = -1 or y = 1
- Learn a weight vector $\mathbf{w} = [w_1; w_2]$
- Predict the class of a given **x** by $\text{sign}(\mathbf{w}^T\mathbf{x}) = \text{sign}(w_1x_1 + w_2x_2)$

?

Linear Classification

To classify we need to separate the data into red vs. blue



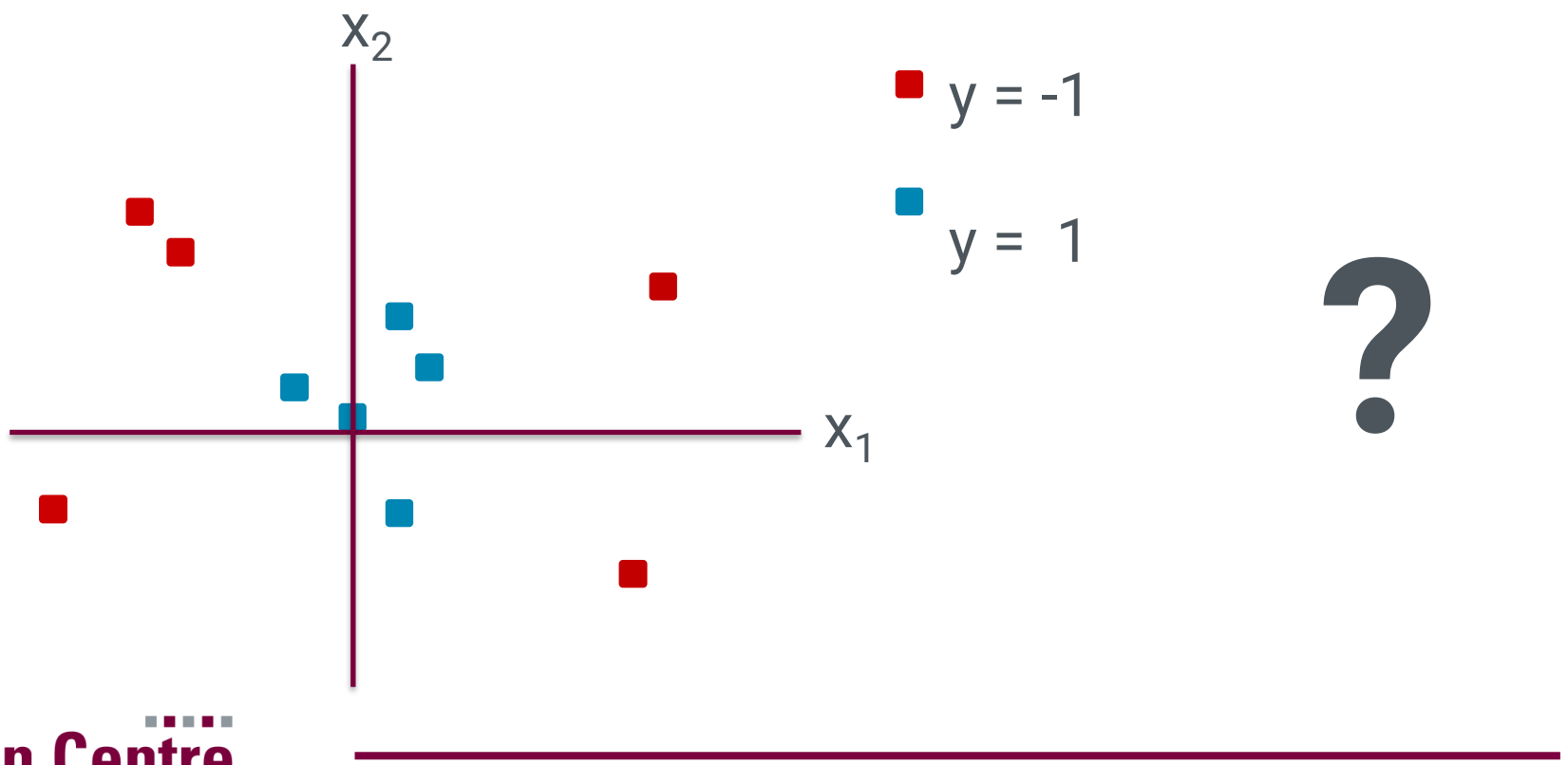
y = -1

y = 1

- Suppose our data points (**x**) are 2D and each comes with a label y, where y = -1 or y = 1
- Learn a weight vector $\mathbf{w} = [w_1; w_2]$
- Predict the class of a given **x** by $\text{sign}(\mathbf{w}^T\mathbf{x}) = \text{sign}(w_1x_1 + w_2x_2)$

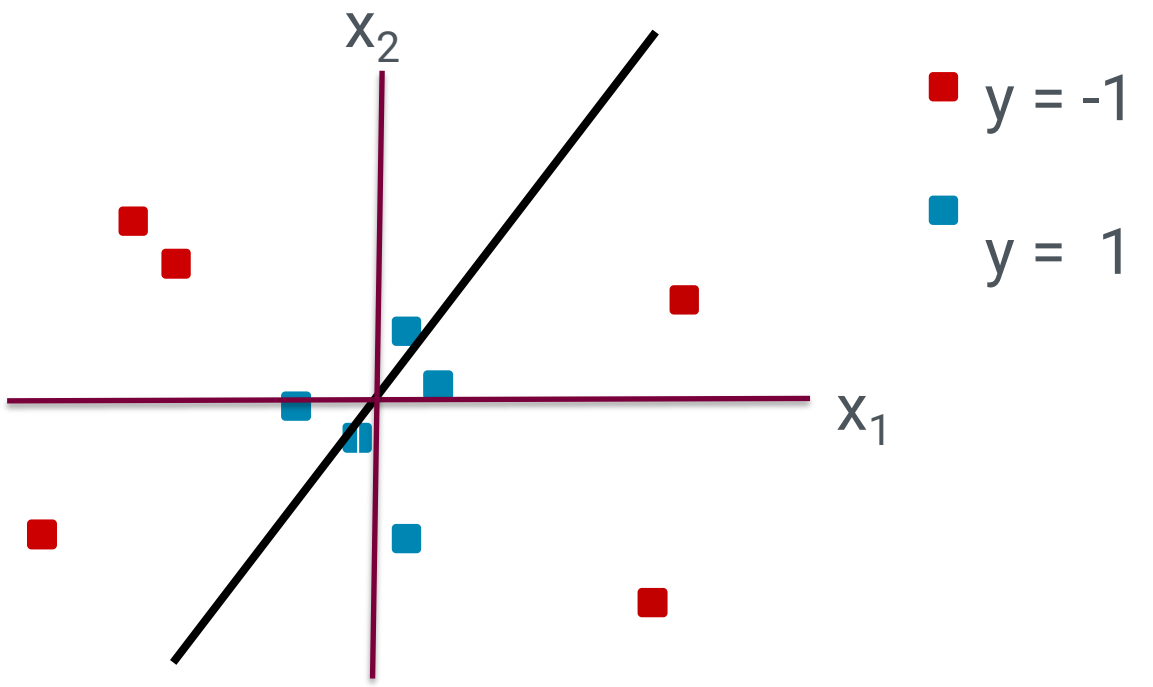
Linearity is Not Enough

To classify we need to separate the data into red vs. blue



Linearity is Not Enough

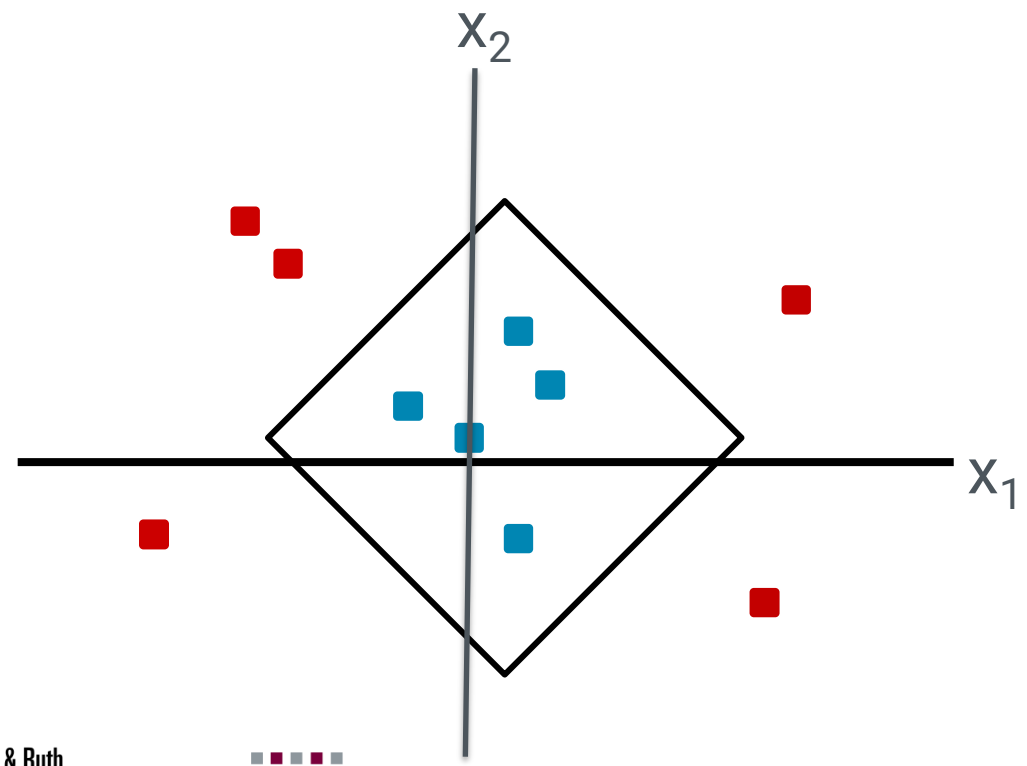
To classify we need to separate the data into red vs. blue



NO

Linearity is Not Enough

To classify we need to separate the data into red vs. blue



y = -1

y = 1

YES

Non-linearity!

The Limits of Linearity

Linear steps **collapse** and stay linear

$$x_2 = W_1 x_1$$

$$x_3 = W_2 x_2$$

$$\longrightarrow \quad x_3 = W_{\text{both}} x_1$$

$$W_{\text{both}} = W_2 \times W_1$$

Linear layers alone do not meaningfully stack

The Shallowest Deep Net

Deep nets are made by stacking learned linear layers and simple pointwise **non-linear** layers

Linear

$$x_2 = W_1 x_1$$

$$x_3 = W_2 x_2$$

add ReLU

Non-linear, Deep

$$x_2 = W_1 x_1$$

$$x_2^\circ = \max(0, x_2)$$

$$x_3 = W_2 x_2^\circ$$

Due to the Rectified Linear Unit (ReLU) non-linearity **max(0, x)**, x_3 cannot be computed as a linear function of x_1

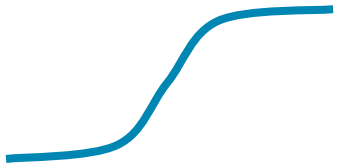
Non-linearity

ReLU

$$x' = \max(0, x)$$

Sigmoid

$$x' = 1/(1 + e^{-x})$$



Non-linearity is needed to deepen the representation

Many non-linearities or *activations* to choose from

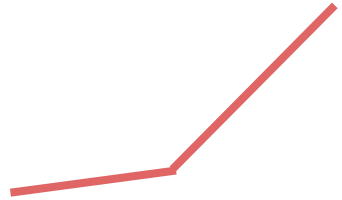
Yet More Non-linearities

ReLU

$$x' = \max(0, x)$$

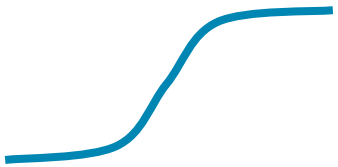
Leaky ReLU

$$x' = \max(0.1x, x)$$



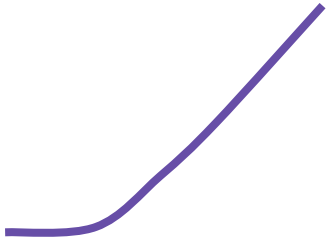
Sigmoid

$$x' = 1/(1 + e^{-x})$$



ELU

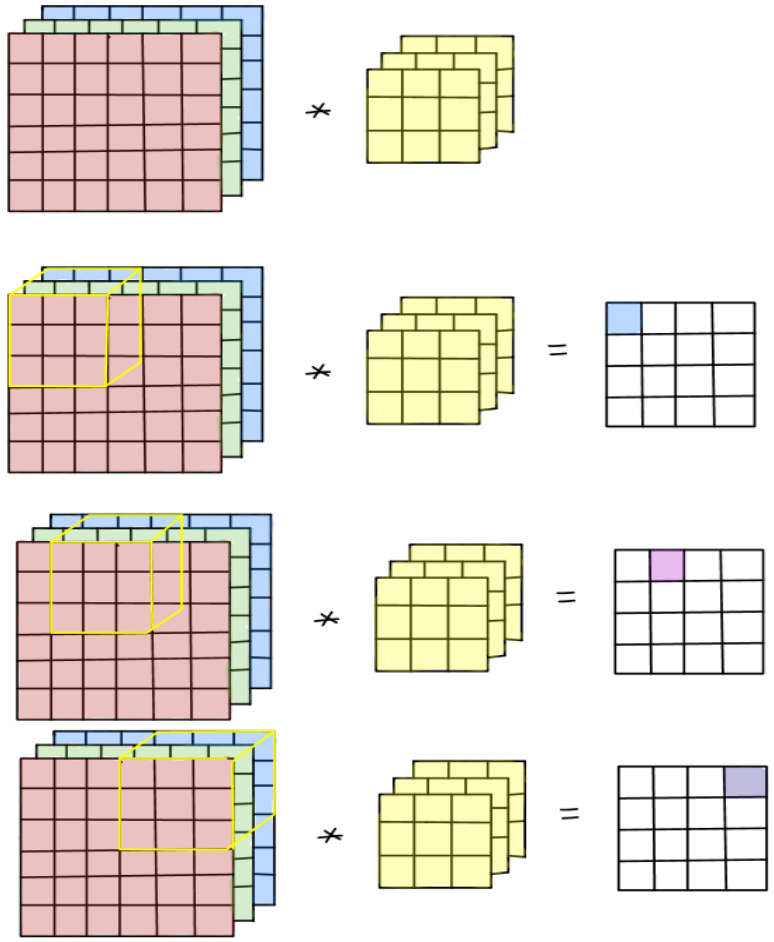
$$x' = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \leq 0 \end{cases}$$

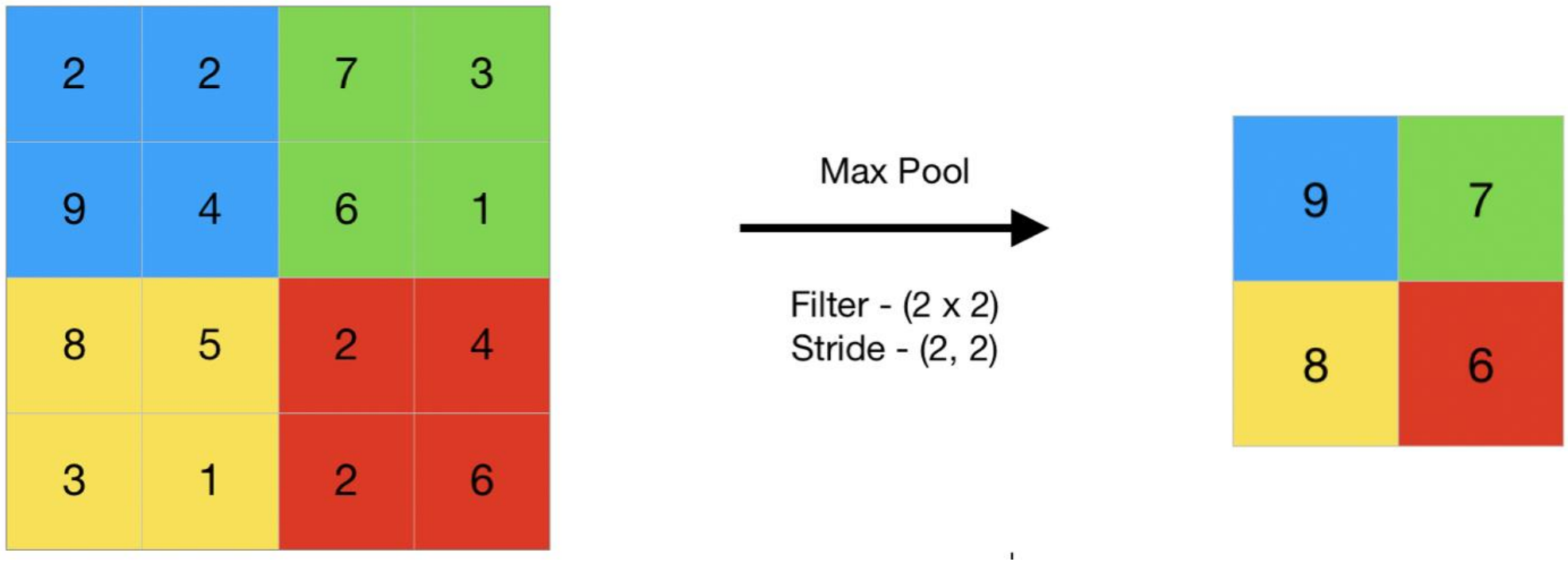


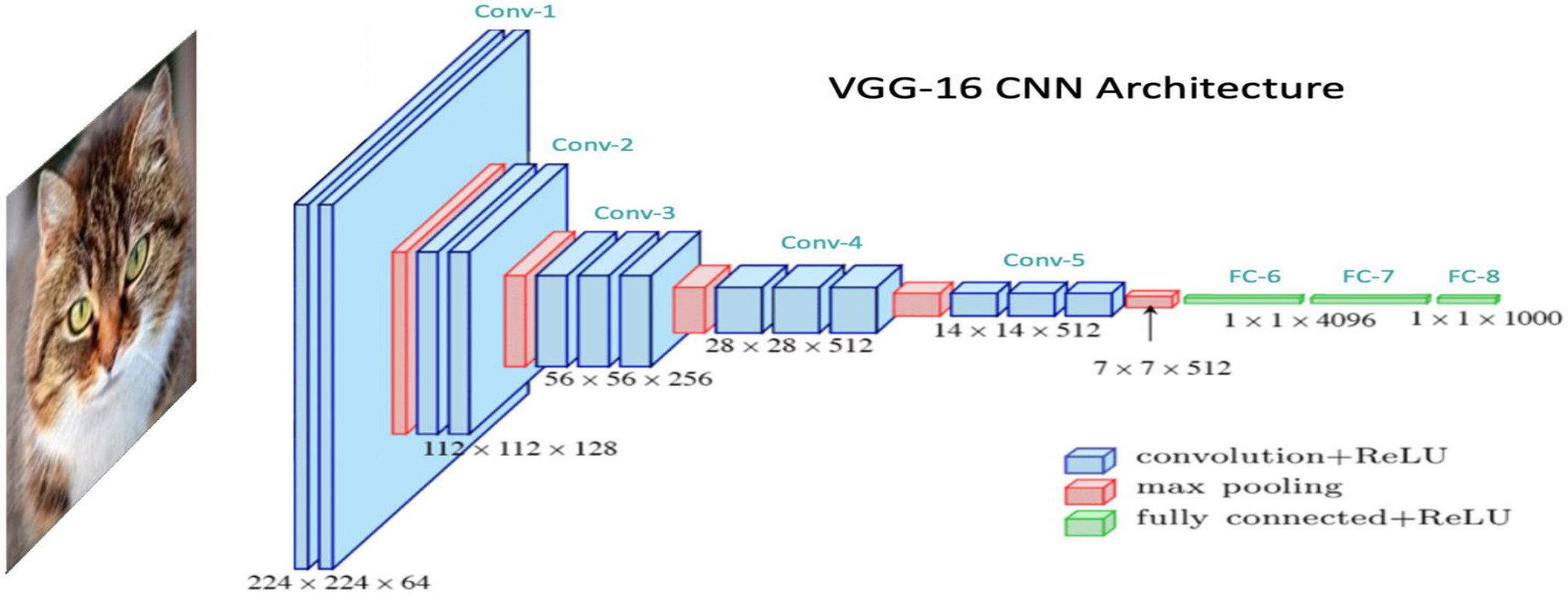
TanH

$$x' = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Visualizing how convolution and maxpooling work:




2 2 7 3
9 4 6 1
8 5 2 4
3 1 2 6
Max Pool
Filter - (2 x 2)
Stride - (2, 2)
9 7
8 6


VGG-16 CNN Architecture
Conv-1
Conv-2
Conv-3
Conv-4
Conv-5
FC-6
FC-7
FC-8
224 × 224 × 64
112 × 112 × 128
56 × 56 × 256
28 × 28 × 512
14 × 14 × 512
7 × 7 × 512
1 × 1 × 4096
1 × 1 × 1000
convolution+ReLU
max pooling
fully connected+ReLU

Contact

Speaker's Email: mousas27@mcmaster.ca

Book an appointment with DASH: https://library.mcmaster.ca/services/dash

Contact DASH: Data Analysis Support Hub: libdash@mcmaster.ca