Introduction to Deep Learning

Embedded AI Lab

CUTe, IIUM

Instructor:

Ts. Dr. Hasan Firdaus Bin Mohd Zaki

Agenda

Deep Learning

Image Classification

Image Classification (Project)

Image Regression

Object Detection Inference(Demo)

Outcomes

In this course, you'll build AI projects that can answer simple visual questions:

- Is my hand showing thumbs-up or thumbs-down?
- Does my face appear happy or sad?
- How many fingers am I holding up?
- Where's my nose?

WHAT IS DEEP LEARNING?


horse
truck
deer
airplane
bird

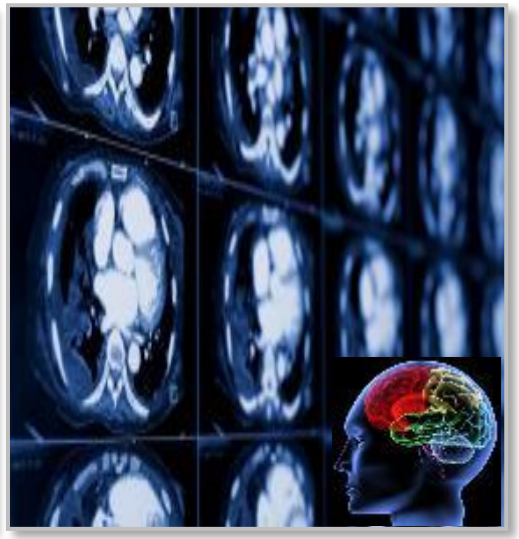
ACCOMPLISHING COMPLEX GOALS

SWEEPING ACROSS INDUSTRIES

Internet Services

- Image/Video classification
- Speech recognition
- Natural language processing

Medicine



- Cancer cell detection
- Diabetic grading
- Drug discovery

Media & Entertainment

- Video captioning
- Content based search
- Real time translation

Security & Defense

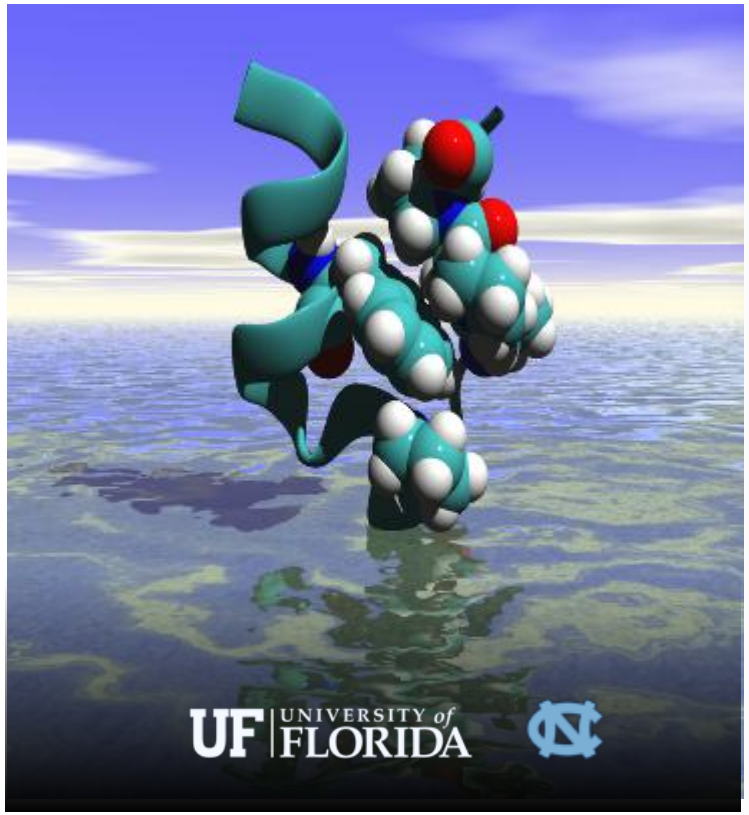
- Face recognition
- Video surveillance
- Cyber security

Autonomous Machines

- Pedestrian detection
- Lane Detection
- Lane tracking
- Recognize traffic signs

TRANSFORMING RESEARCH

“Seeing” Gravity In Real Time



Accelerating Drug Discovery

92%

believe AI will impact their work

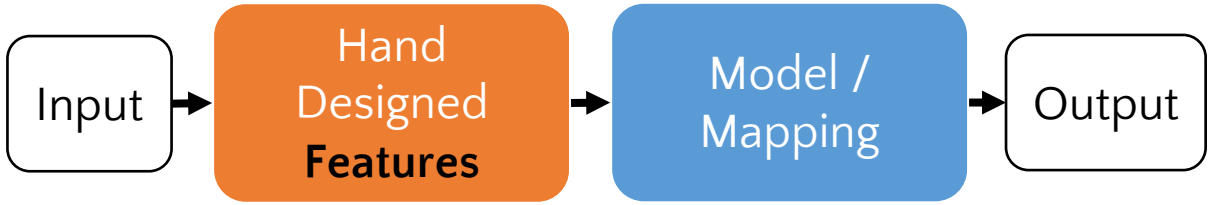
93%

using deep learning seeing positive results

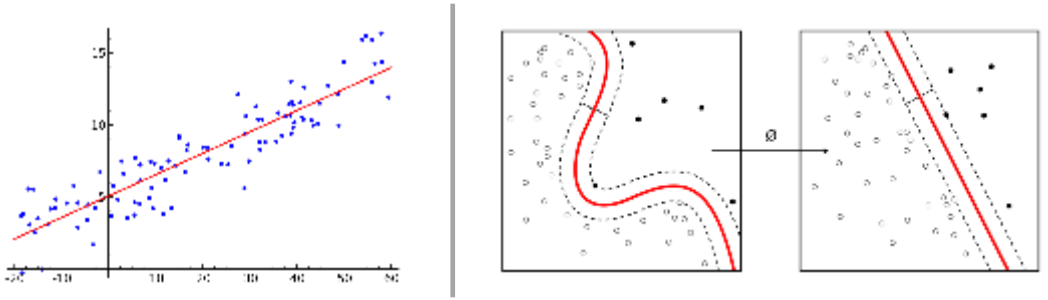
insideHPC.com Survey
November 2016

Difference in Workflow

Classic Machine Learning [1990 : now]

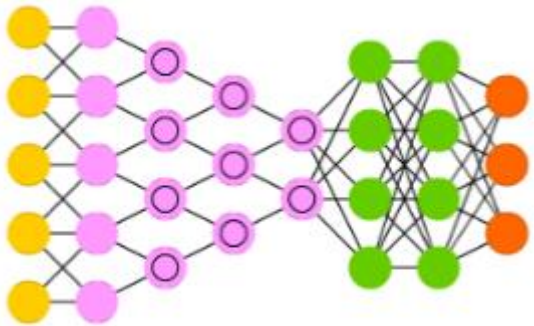


Examples [Regression and SVMs]



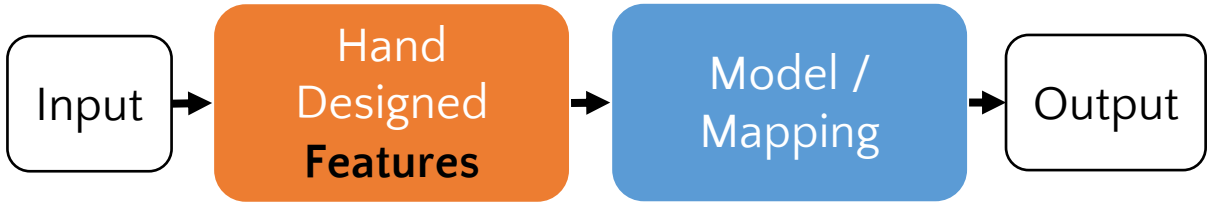
Deep/End-to-End Learning [2012 : now]

Example [Conv Net]

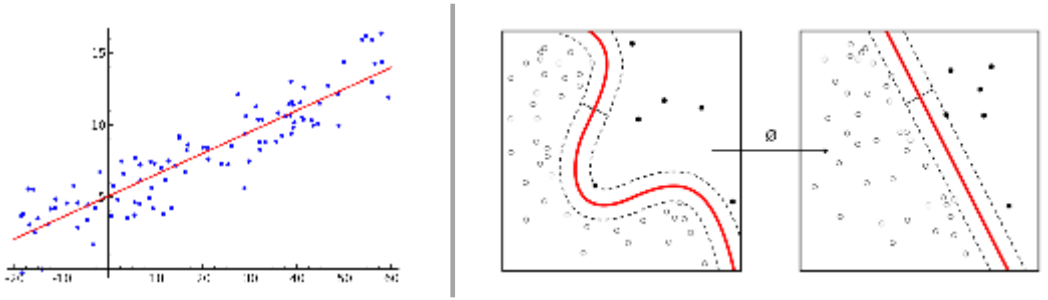


Traditional Workflow

Classic Machine Learning [1990 : now]



Examples [Regression and SVMs]



Challenge : How would you describe this image to someone (or something) blind?

Difficult: From the raw pixels.
Medium: From geometric primitives (lines, curves, colors)
Easy: Using any words that you may know

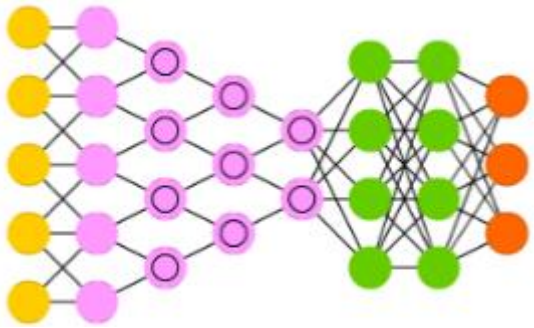
Deep Learning Workflow

Experience: Trust Neural Network to learn features and model by providing inputs and outputs.

Key Skill: Experience (data) creation

Deep/End-to-End Learning [2012 : now]

Example [Conv Net]



INPUT TO OUTPUT

Louie or Not Louie?

1 = Louie
0= Not Louie
Confidence/Predication?

INPUT TO OUTPUT

Louie or Not Louie?

1 = Louie
0= Not Louie
.85 = 85% confident Louie

Yes, this beagle is Louie!

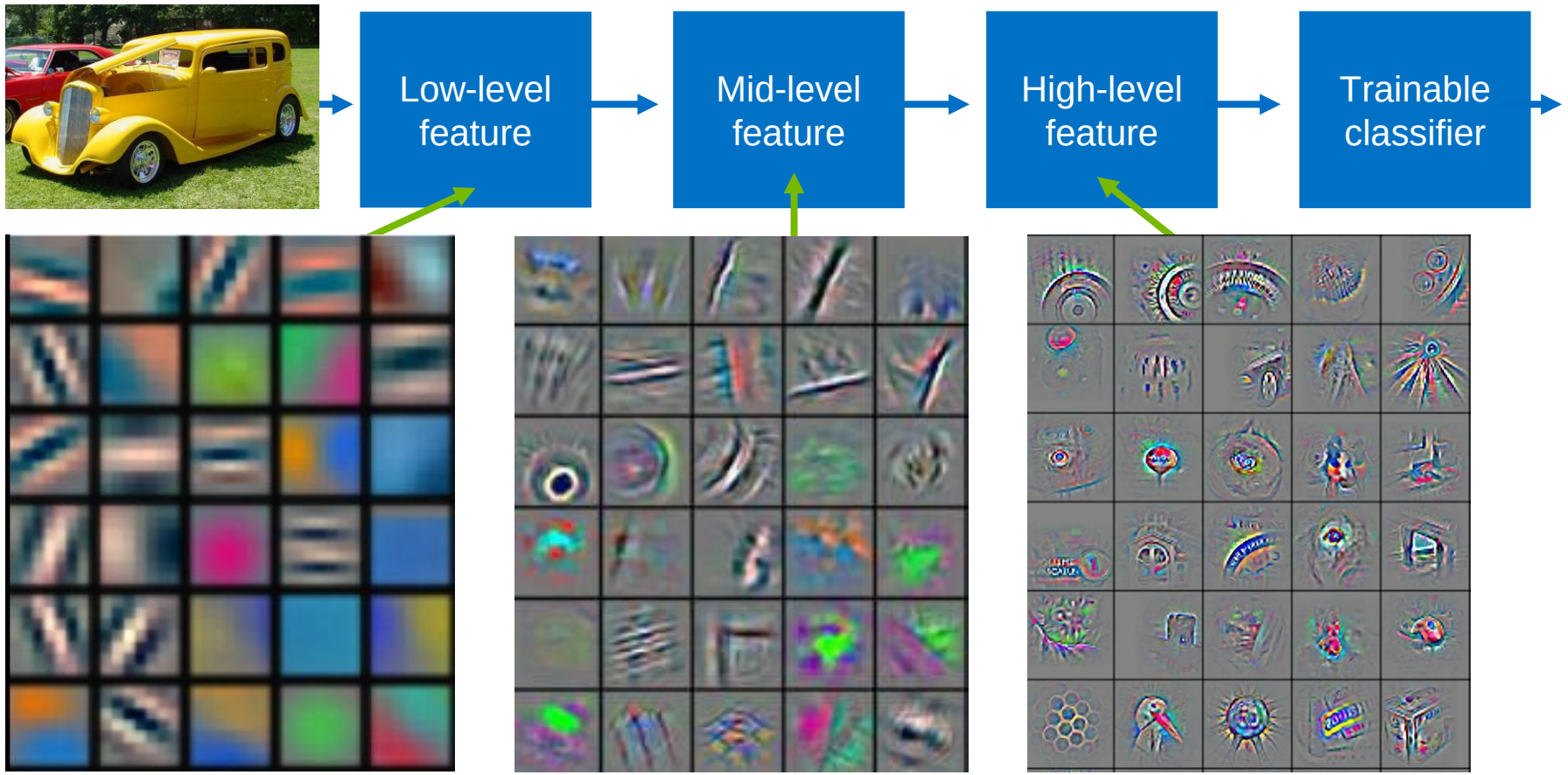
Deep learning =
Learning representations/features

– The traditional model of pattern recognition (since the late 50's)
  – Fixed/engineered features (or fixed kernel) + trainable classifier

– End-to-end learning / Feature learning / Deep learning
  – Trainable features (or kernel) + trainable classifier

Deep learning = learning hierarchical representations

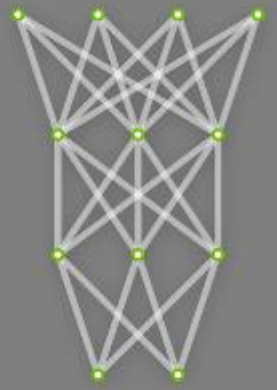
It's deep if it has more than one stage of non-linear feature transformation



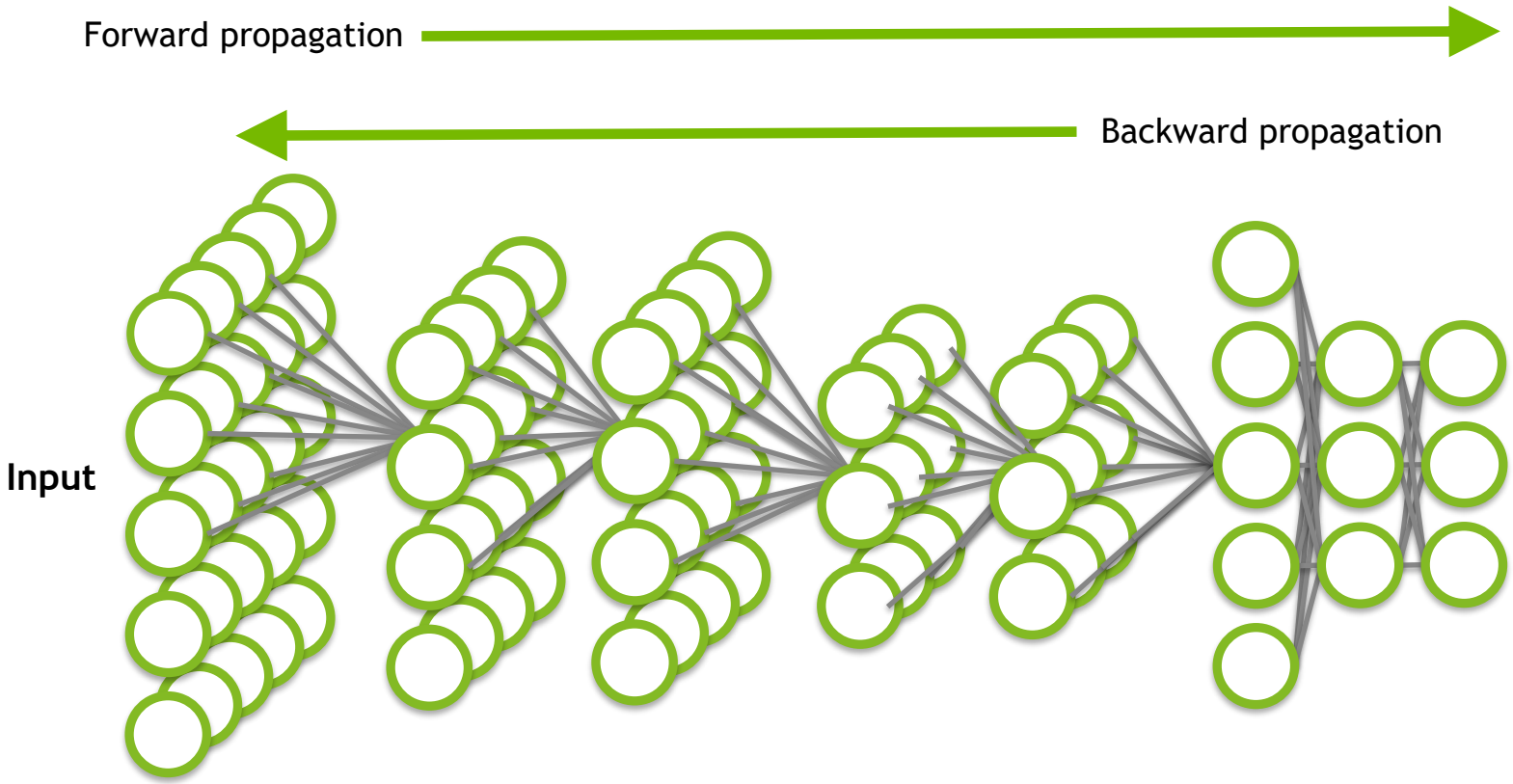
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

DEEP LEARNING

Untrained
Neural Network
Model



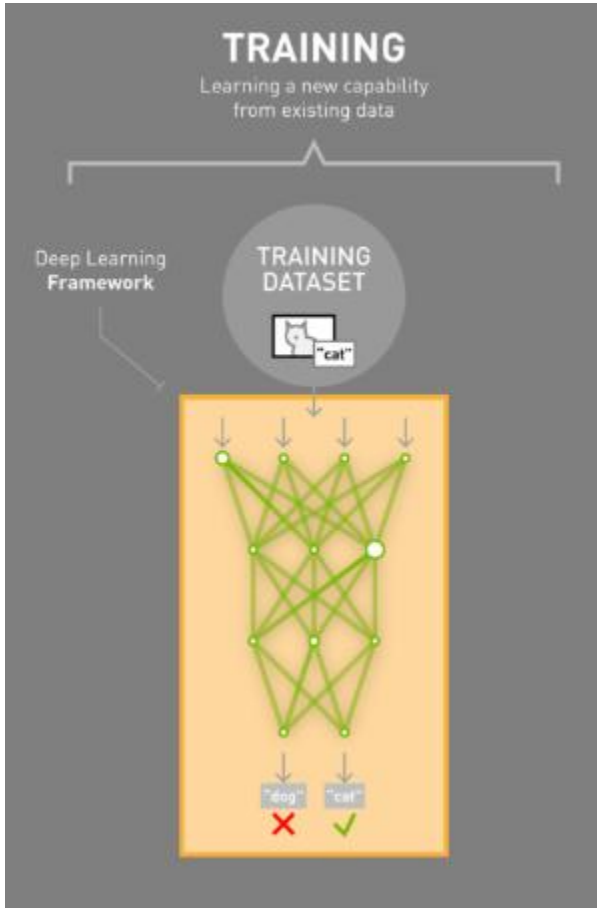
DEEP LEARNING APPROACH - TRAINING



Process

- Forward propagation yields an inferred label for each training image
- Loss function used to calculate difference between known label and predicted label for each image
- Weights are adjusted during backward propagation
- Repeat the process

Training a network with data



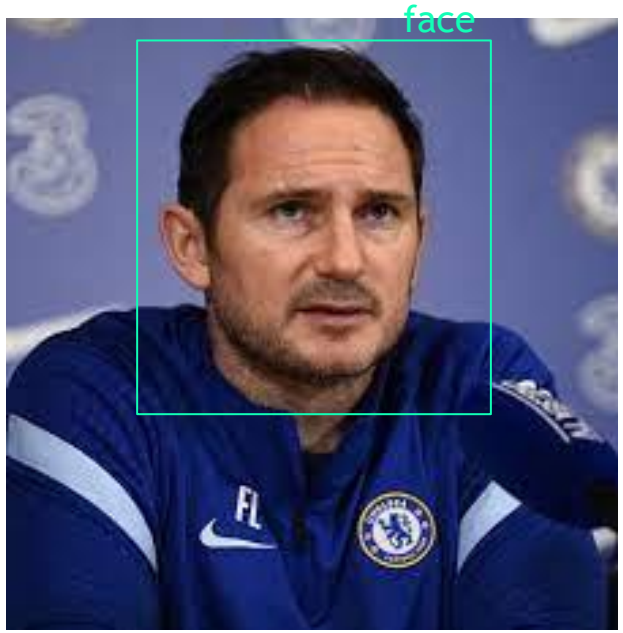
Three types of training protocols

- Purely Supervised
  - Initialize parameters randomly Train in supervised mode
    - Typically with SGD, using backprop to compute gradients
  - Used in most practical systems for speech and image recognition
- Unsupervised, layerwise + supervised classifier on top
  - Train each layer unsupervised, one after the other
  - Train a supervised classifier on top, keeping the other layers fixed
  - Good when very few labeled samples are available
- Unsupervised, layerwise + global supervised fine-tuning
  - Train each layer unsupervised, one after the other
  - Add a classifier layer, and retrain the whole thing supervised
  - Good when label set is poor (e.g. pedestrian detection)
- Unsupervised pre-training often uses regularized auto-encoders

detection +classification

person

classification

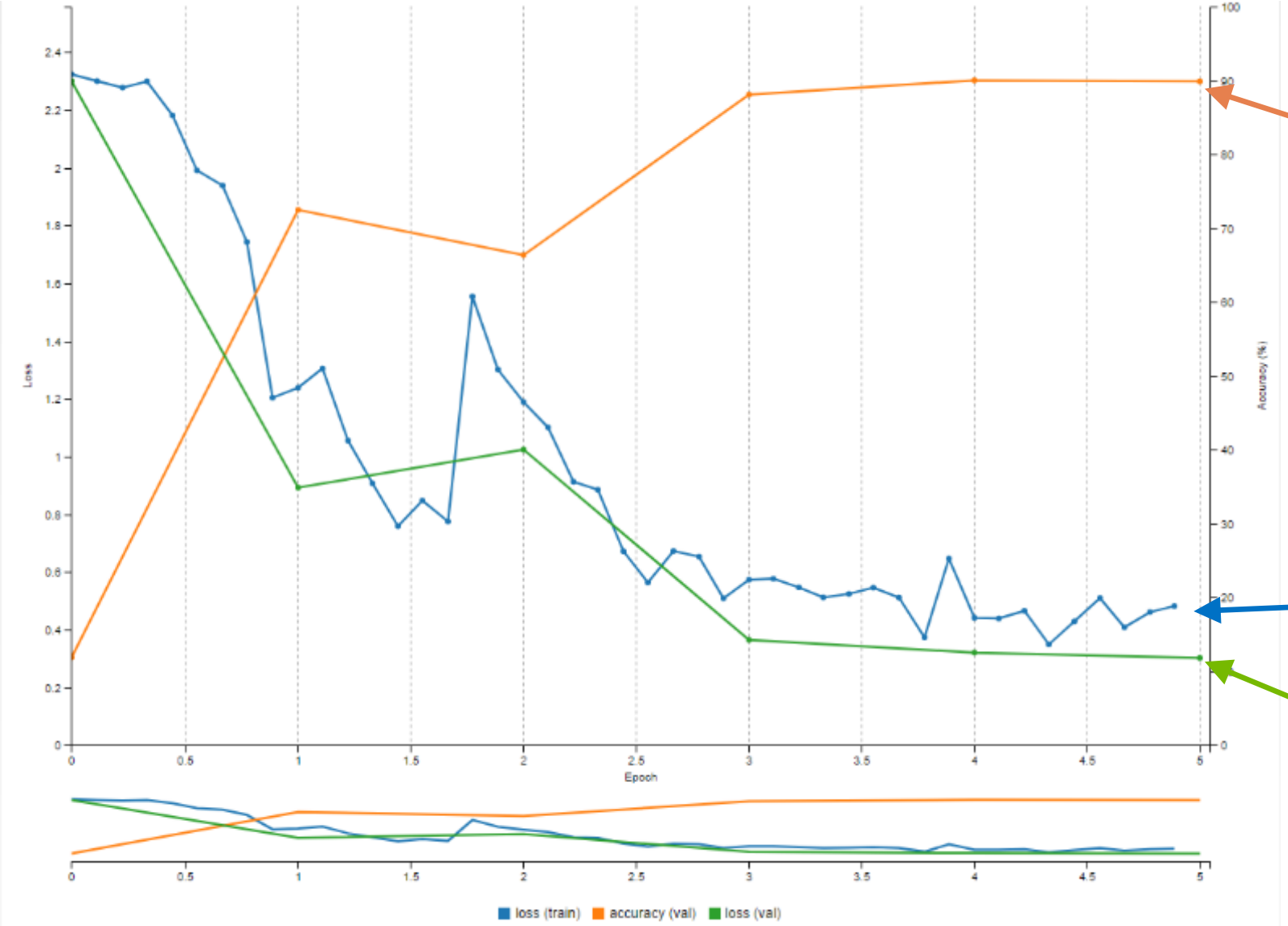


detection
+classification

regression

Evaluating Performance

EVALUATE THE MODEL



Accuracy obtained from validation dataset

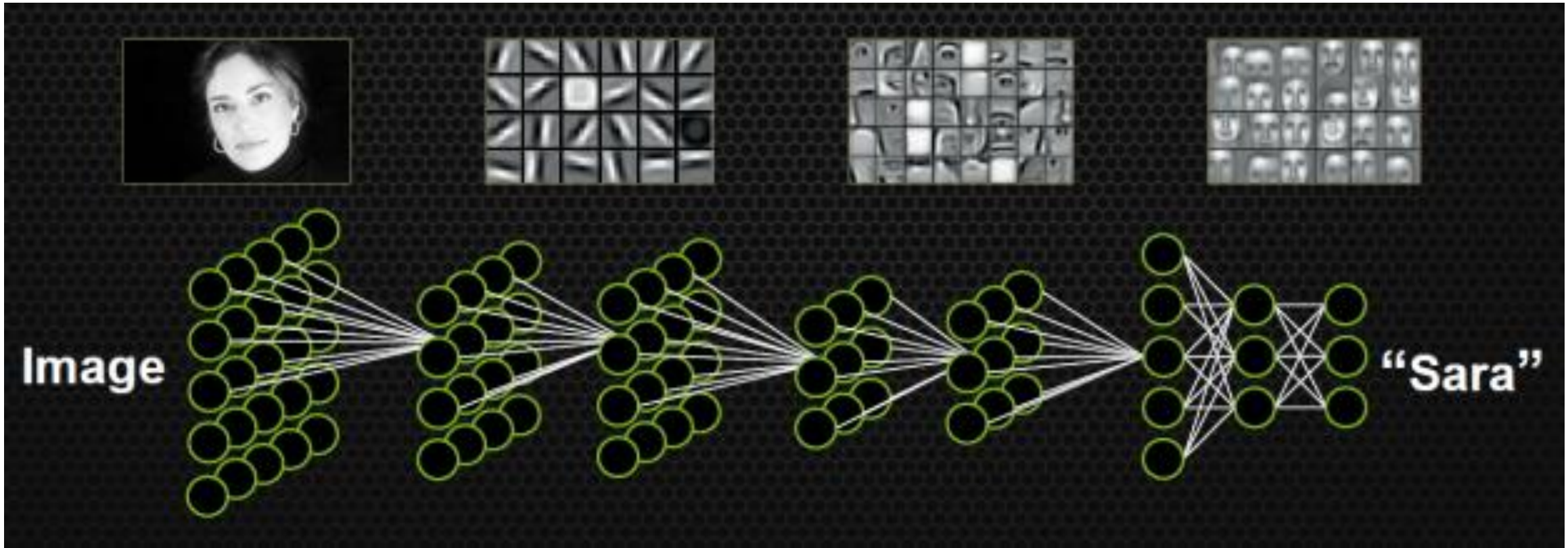
Loss function (Training)

Loss function (Validation)

Convolutional Neural Networks (CNNs)

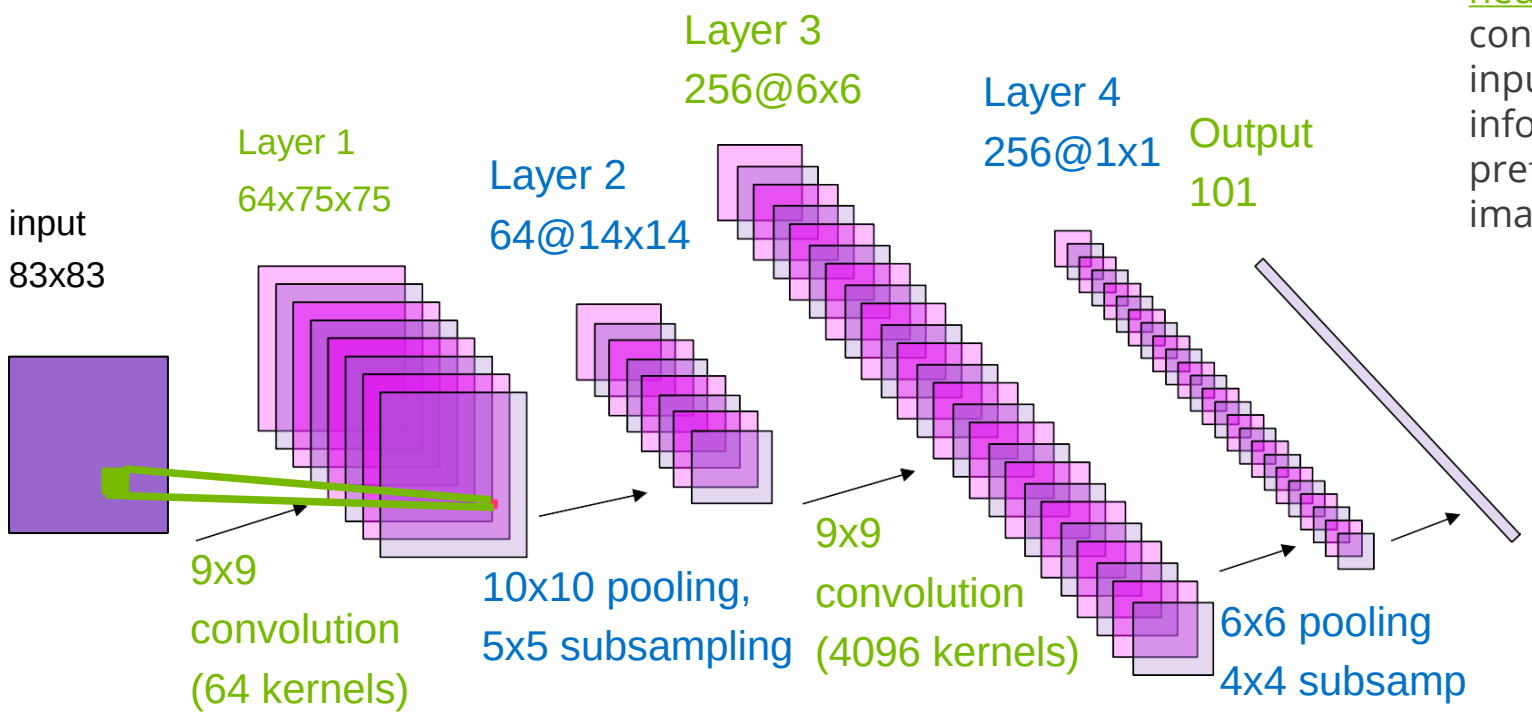
Artificial Neural Network

- An artificial neural network is a biologically inspired computational model
- At each layer, the network transforms input data by applying a nonlinear function to a weighted sum of the inputs.
- The intermediate outputs of one layer, called features, are used as the *input into the next layer*.
- The final layer is used to make a prediction



Convolutional Network (ConvNet)

A CNN is a class of artificial neural network that uses convolutional layers to filter inputs for useful information, and is the preferred network for image applications



- Non-Linearity: half-wave rectification (ReLU), shrinkage function, sigmoid
- Pooling: max, average, L1, L2, log-sum-exp
- Training: Supervised (1988-2006), Unsupervised+Supervised (2006-now)

Convolutional Neural Network (CNN)

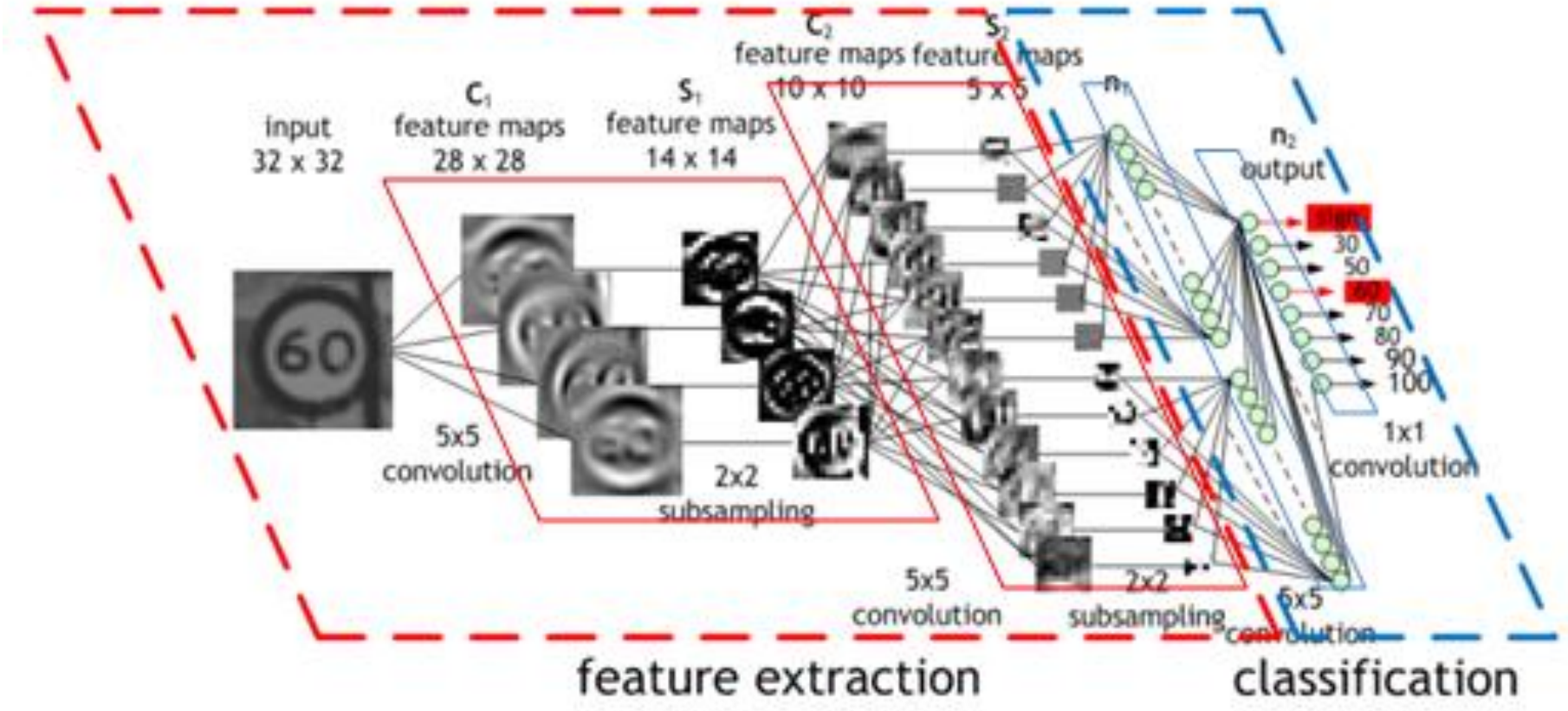


Figure 1: An input image of a traffic sign is filtered by 4 5x5 convolutional kernels which create 4 feature maps, these feature maps are subsampled by max pooling. The next layer applies 10 5x5 convolutional kernels to these subsampled images and again we pool the feature maps. The final layer is a fully connected layer where all generated features are combined and used in the classifier (essentially logistic regression). Image by Maurice Peemen.

Training and Hyperparameters

- The hyper-parameters for the model training are set.
  - **number of epochs,**
  - **batch size,**
  - **Learning rate,**
  - **Momentum**
- Loads the images for training and/or evaluation.
- The model determines a predicted output from the loaded input image.
- The difference between the predicted output and the actual label is used to calculate the "**loss".**
- If the model is in training mode, the loss is backpropagated into the network to improve the model.
- One **epoch** is a complete cycle of all images through the trainer network. This may take several minutes to execute.
- **Batch size** refers to the number of training examples utilized in one iteration.
- **Learning rate** controls how much to change the model in response to the estimated error each time the model weights are updated.

Classification Vs. Regression

#1. Type of Mapping Function

Regression

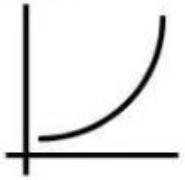
In these algorithms, the mapping function will be chosen of type which can align the values to the continuous output.

Classification

In these algorithms, the mapping function will be chosen of type which can align the values to the predefined classes.

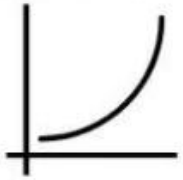
#2. Involves Prediction

Regression



For this type of algorithms, predicted data belongs to the category of continuous values.
(Like 23,34,45,67,28)

Classification



For this type of algorithm's predicted data, belongs to the category of discrete values.
(Like Either Yes or No, Belongs to A or B or C).

Classification Vs. Regression

#3. Method of Calculation

Regression

Root Mean Square Error will be calculated to identify the best fit of the dataset.

Classification

Accuracy will be calculated to identify the best fit of the dataset.

#4. Nature of the Predicted Data

Regression

The nature of the predicted data is ordered. (That is values predicted will be in some sequence).

Classification

The nature of the predicted data is unordered. (That is values predicted will not be in any sequence).

DEEP LEARNING INSTITUTE

Instructor:

Ts. Dr. Hasan Firdaus Bin Mohd Zaki

www.nvidia.com/dli