Data Science Portugal
October 08, 2019
Coimbra, Portugal

Human Action Recognition using Deep Learning methods

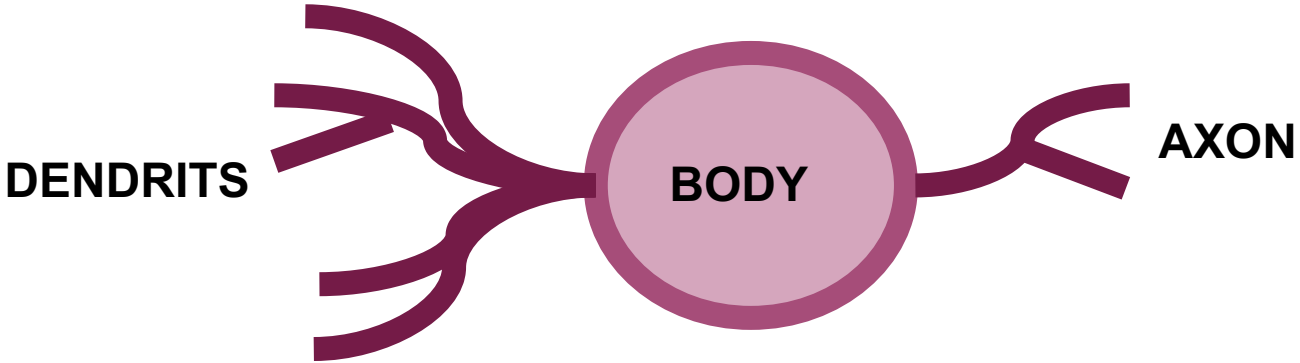
Hemerson Aparecido da Costa Tacon
Data Scientist @ Talkdesk

Summary

1. Artificial Neural Networks
2. Computer Vision Problems
3. Convolutional Neural Networks
4. What does a CNN learn?
5. Human Action Recognition
6. Deep Learning and HAR

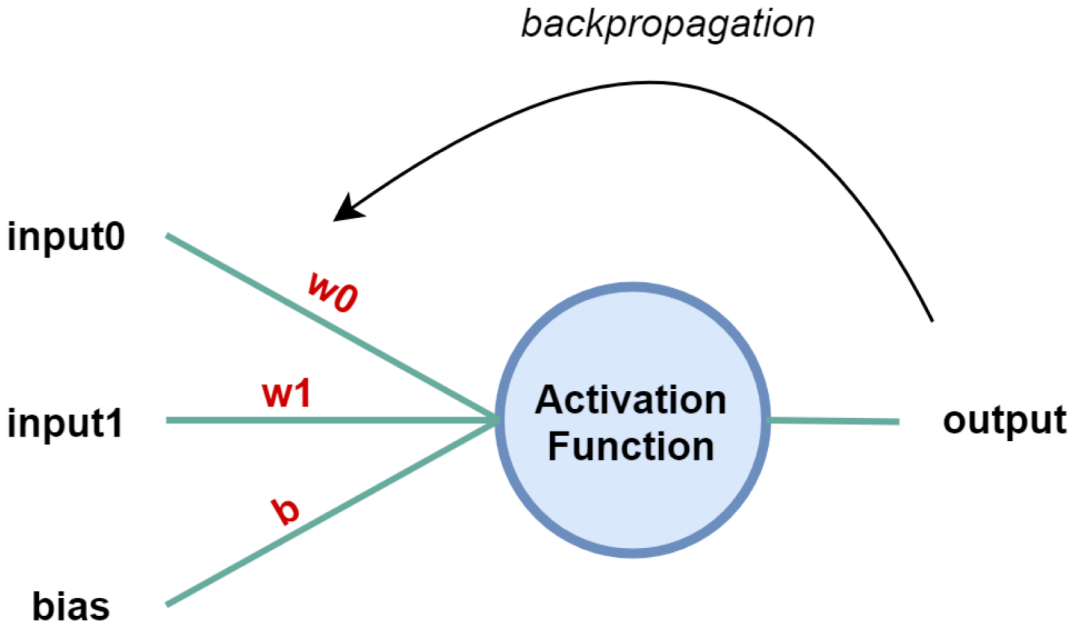
What are Neural Networks (NNs)?

- Biological inspiration



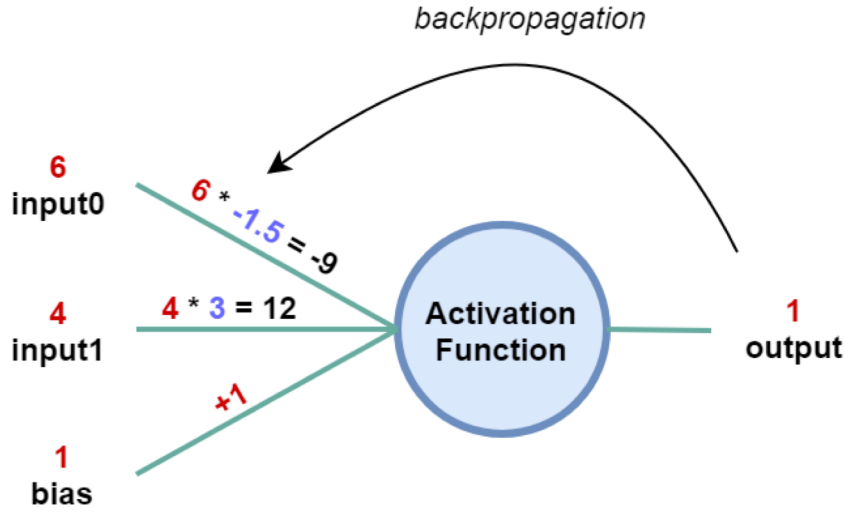
Basic elements of a NN

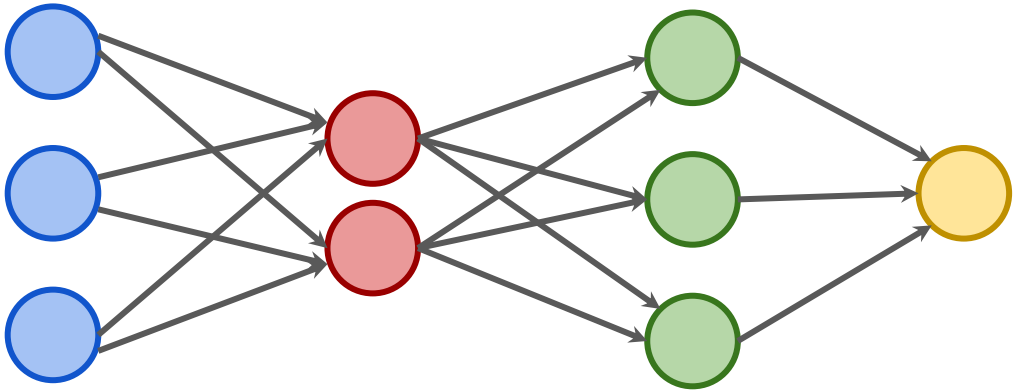
- Main components:
  - neuron;
  - activation function;
  - connections (weights);
  - cost function;
  - gradient descent;
  - backpropagation;



Simple Neural Network

- Perceptron
- Layer types:
  - input;
  - hidden;
  - output.
- Multilayer Perceptron

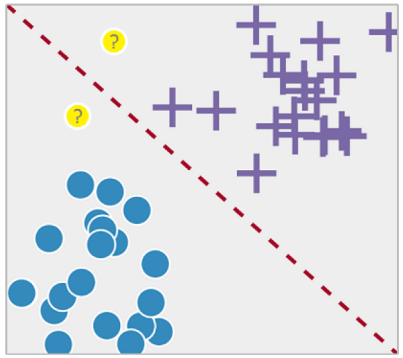


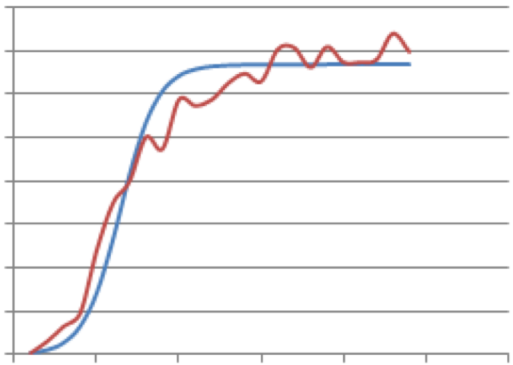


Use cases of Neural Networks

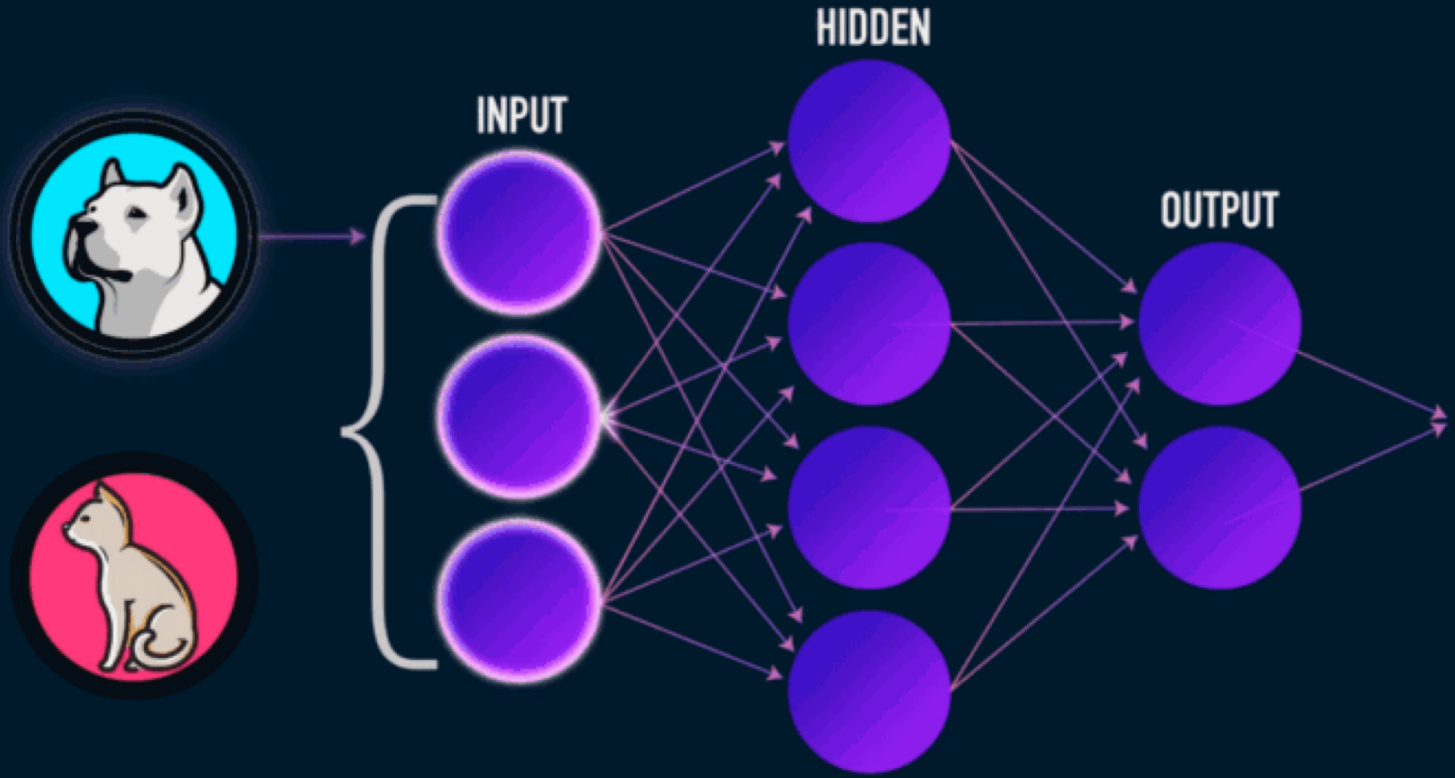
Some use cases:

- function approximation;
- time series forecasting;
- classification;
- pattern recognition.



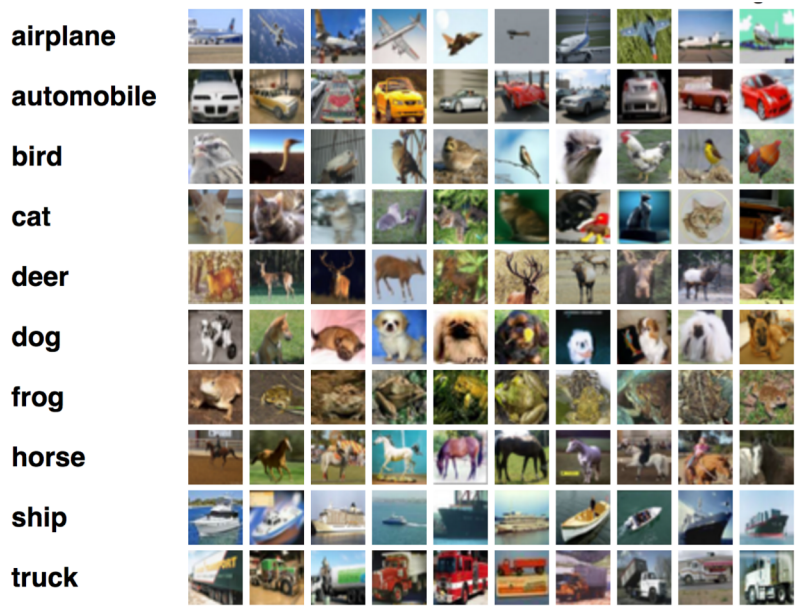


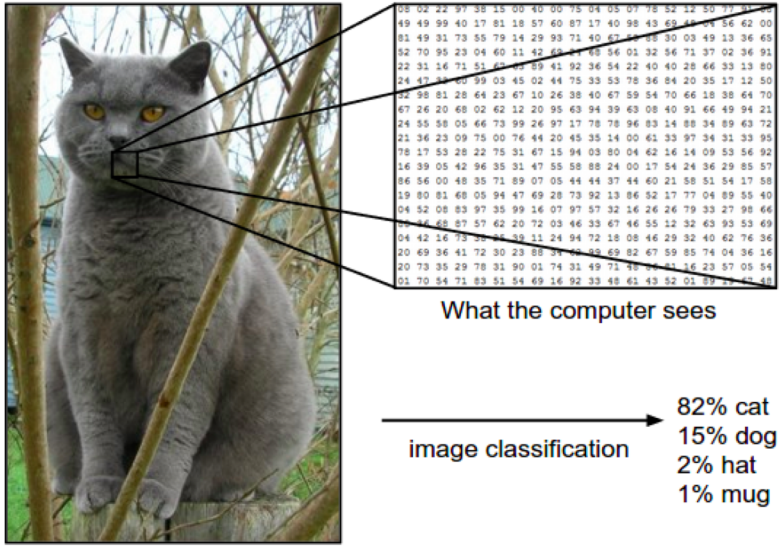
NNs for Computer Vision problems



NNs for Computer Vision problems

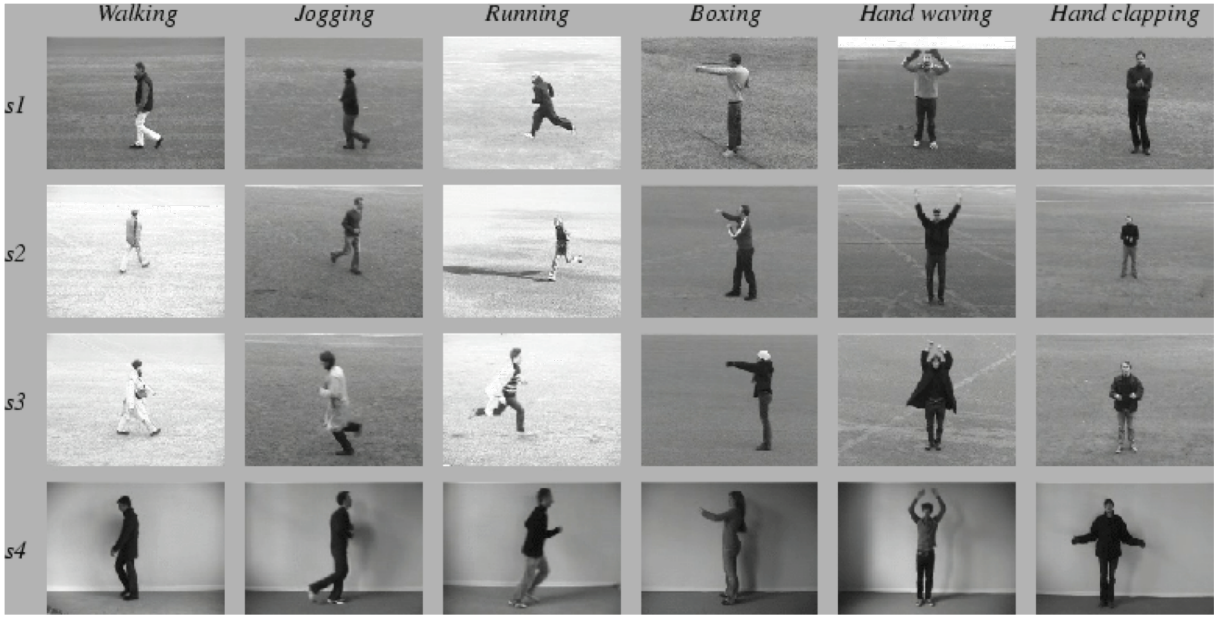
- Image classification





NNs for Computer Vision problems

- Human Action Recognition in video

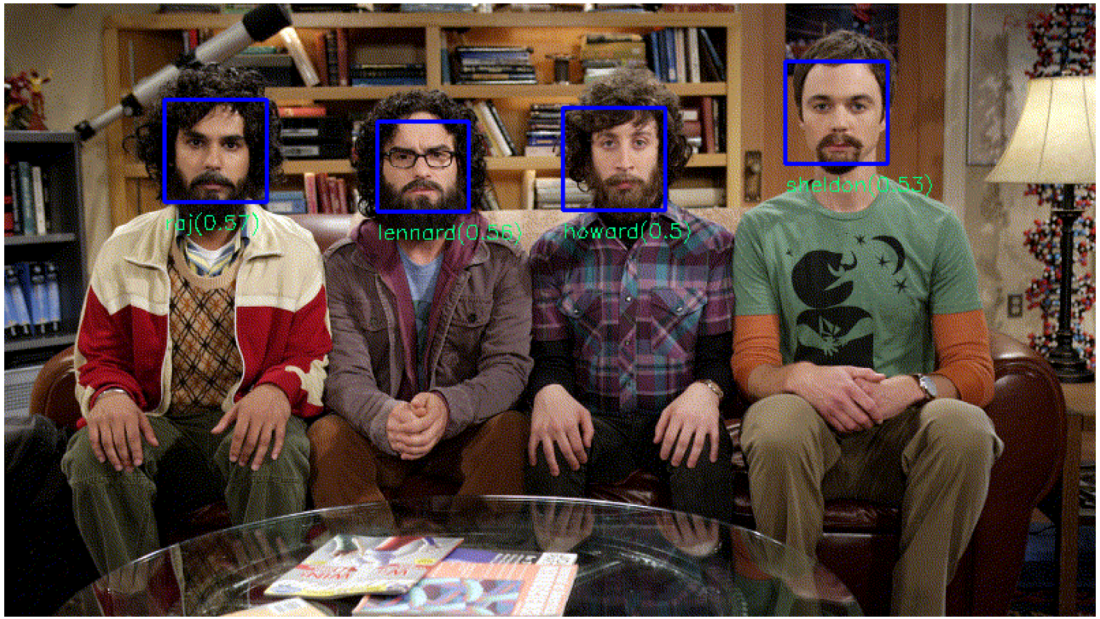


NNs for Computer Vision problems

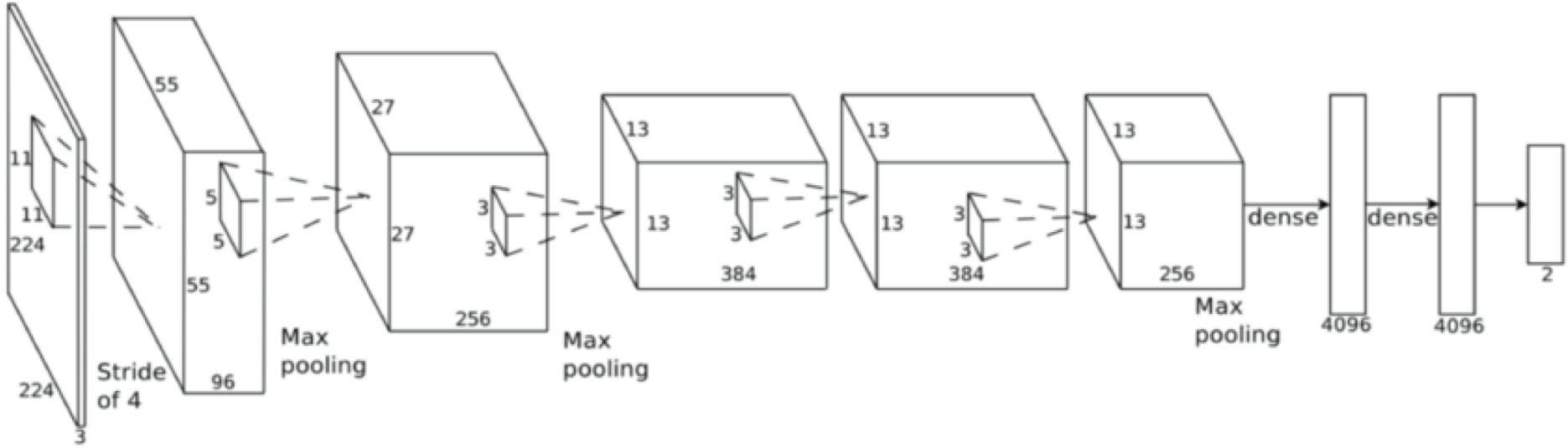
- Object detection

NNs for Computer Vision problems

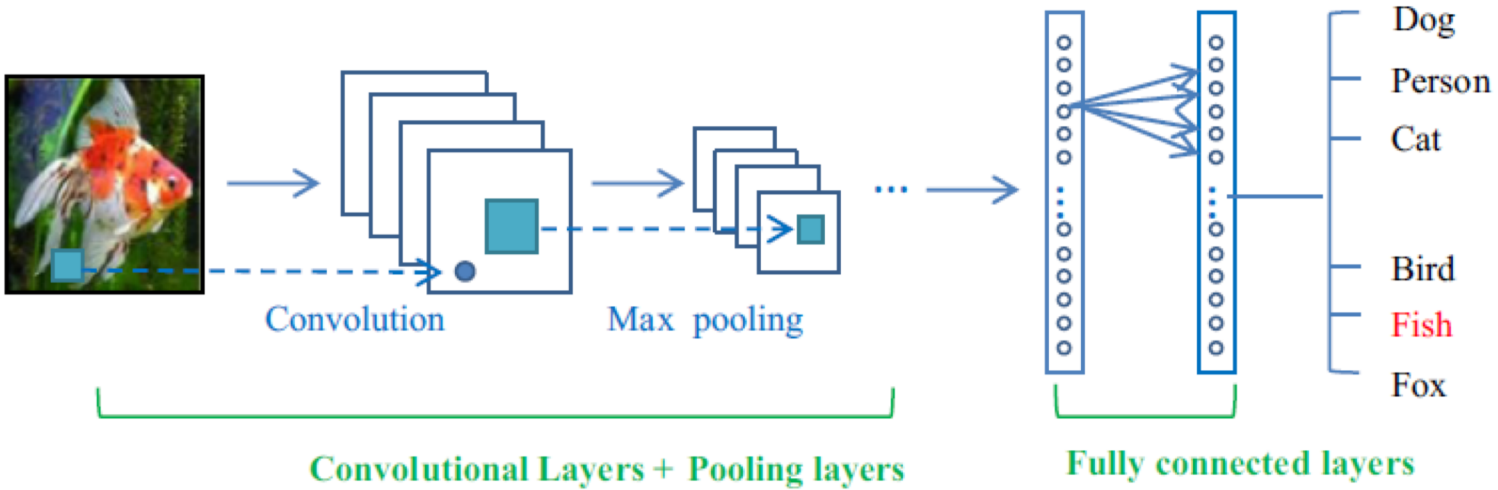
- Face recognition



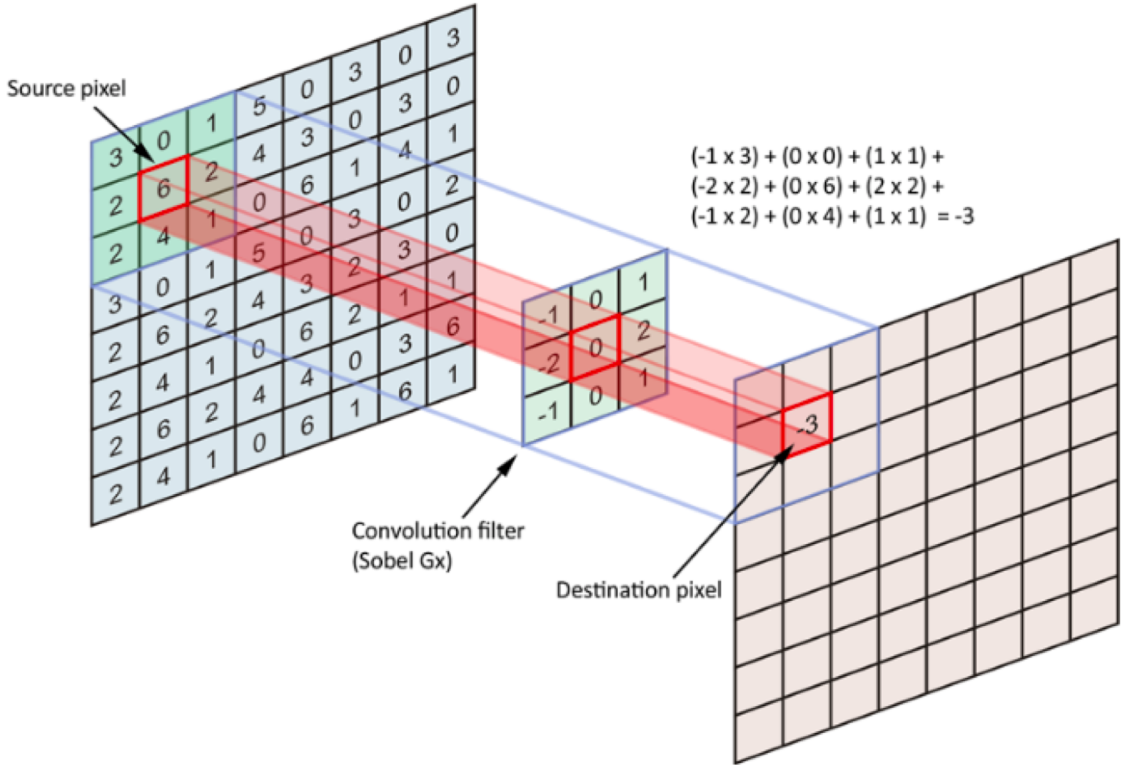
Convolutional Neural Networks (CNNs)



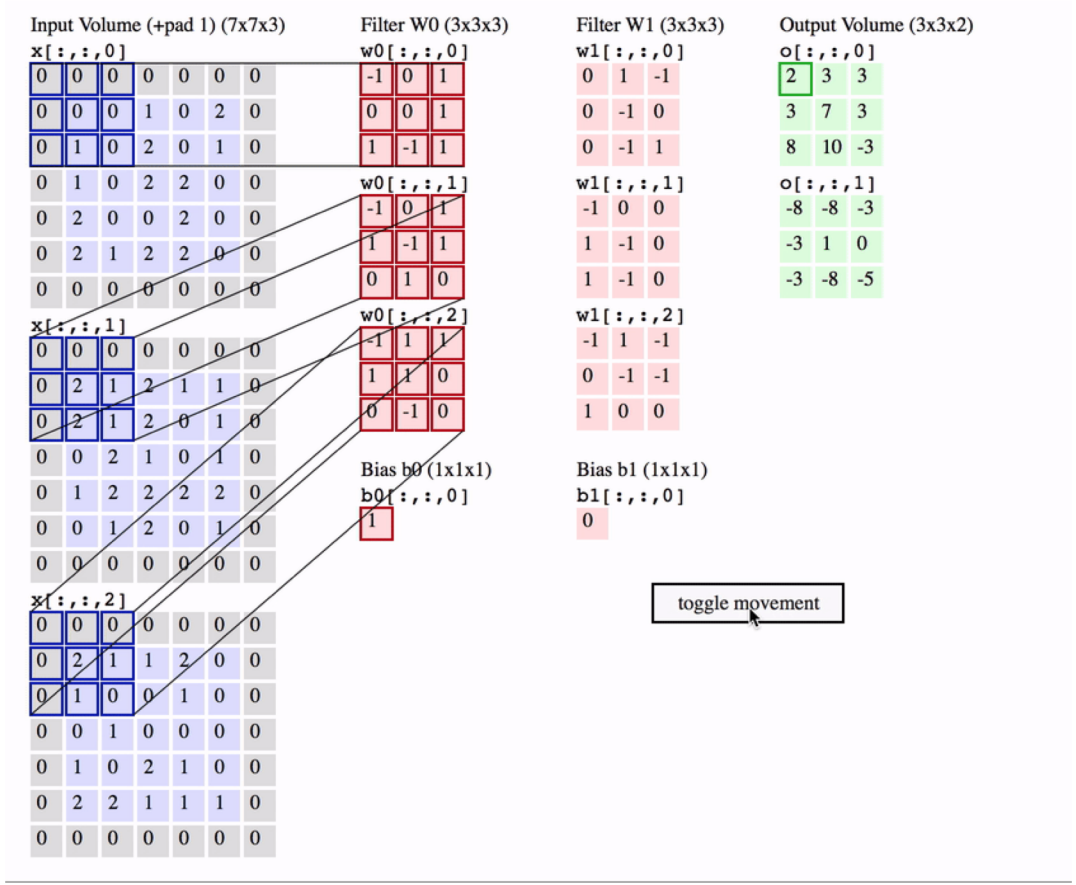
Convolutional Neural Networks (CNNs)



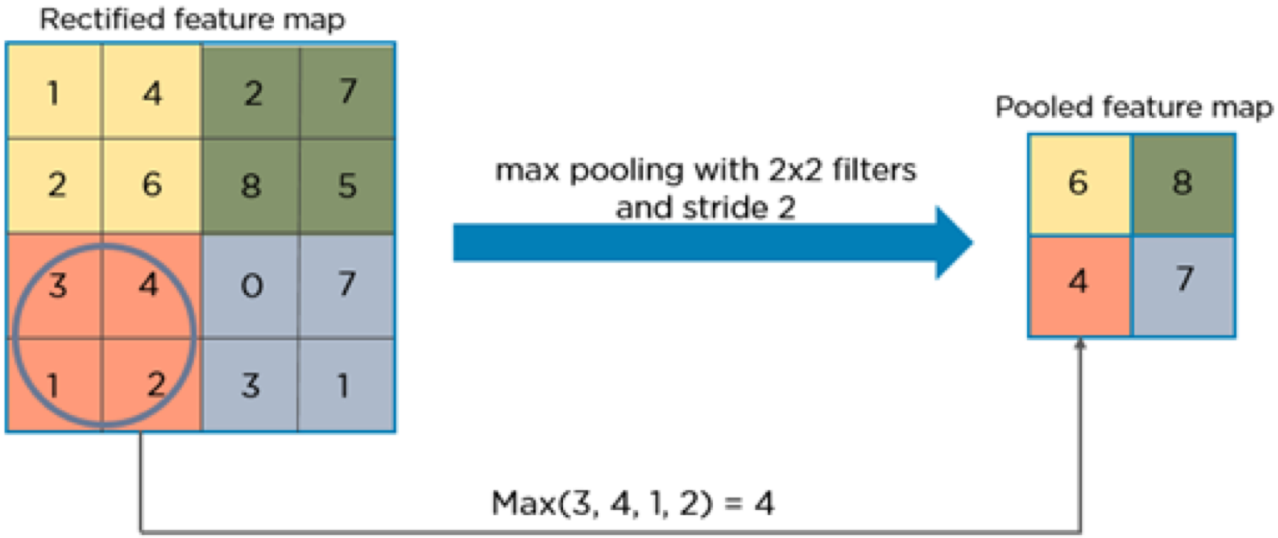
Convolution



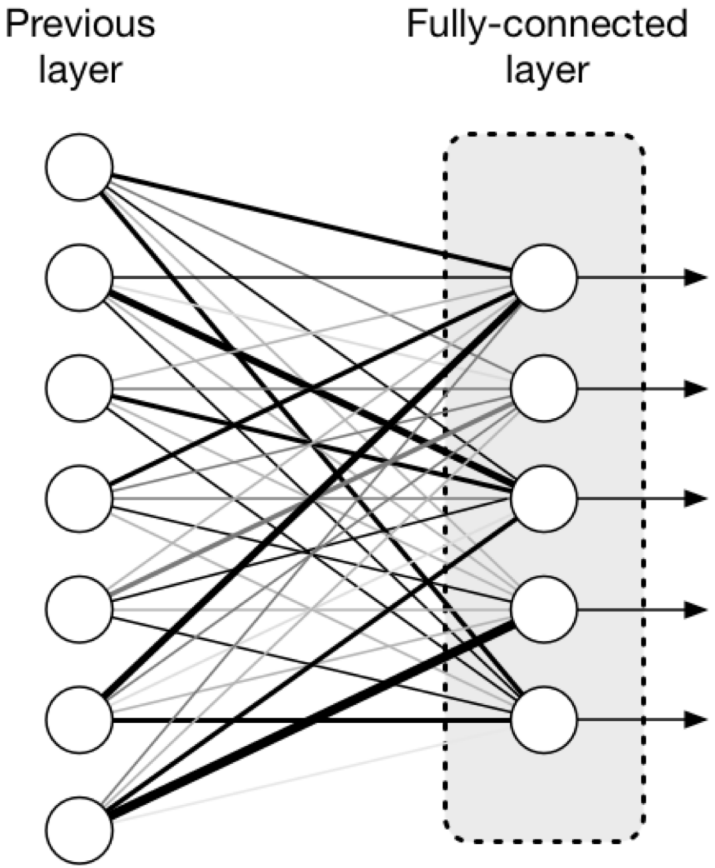
Convolution



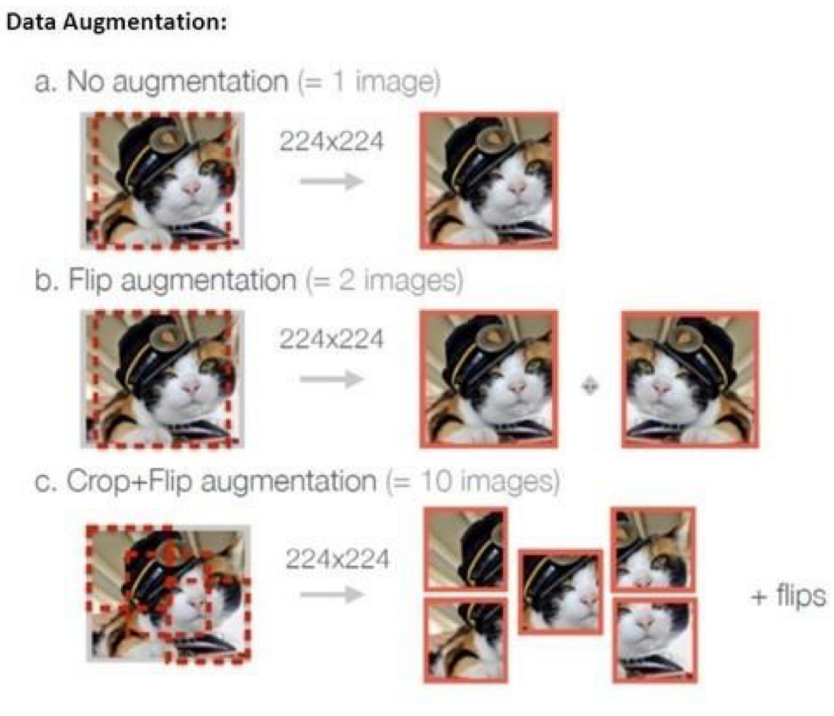
Pooling


Rectified feature map
1 4 2 7
2 6 8 5
3 4 0 7
1 2 3 1
max pooling with 2x2 filters
and stride 2
Pooled feature map
6 8
4 7
Max(3, 4, 1, 2) = 4

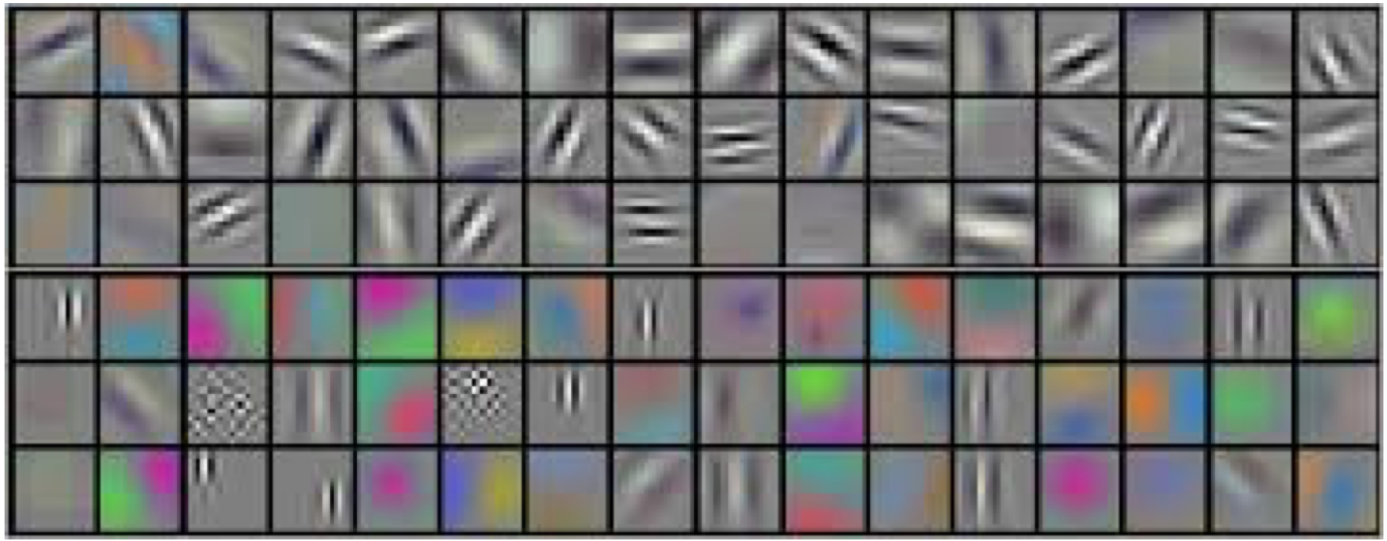
Fully connected layer



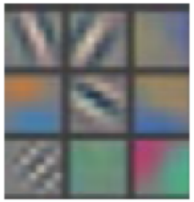
Data Augmentation



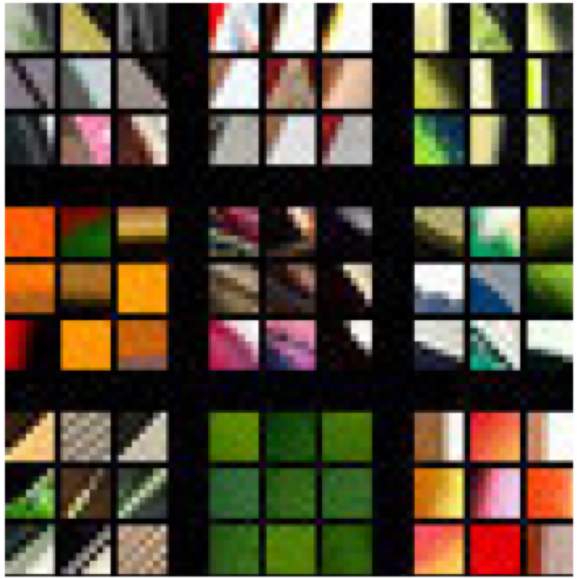
What does a CNN learn?



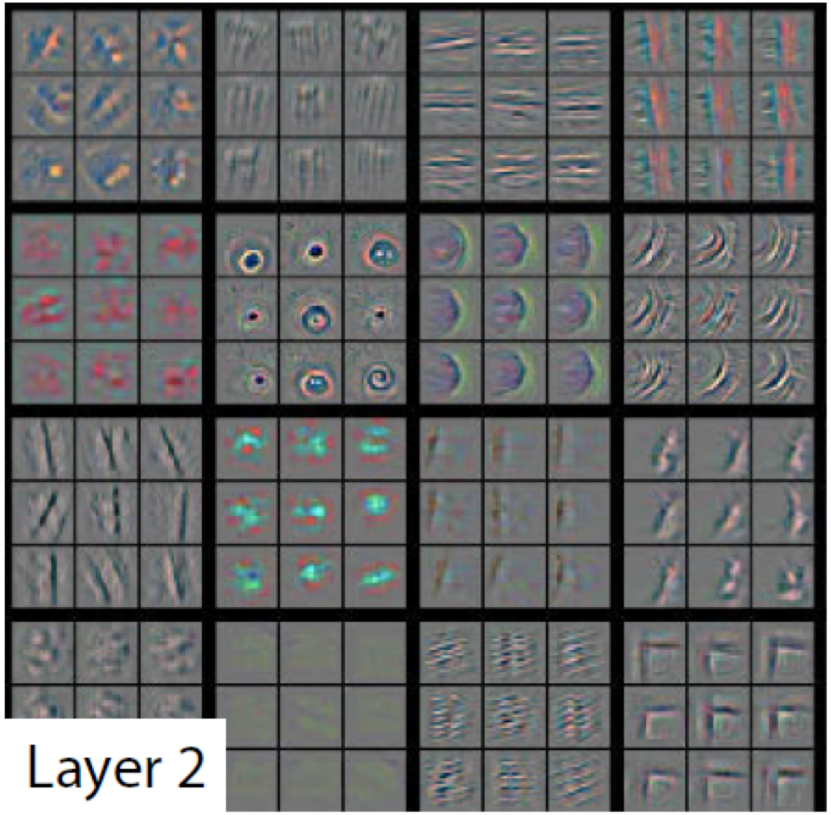
What does a CNN learn?

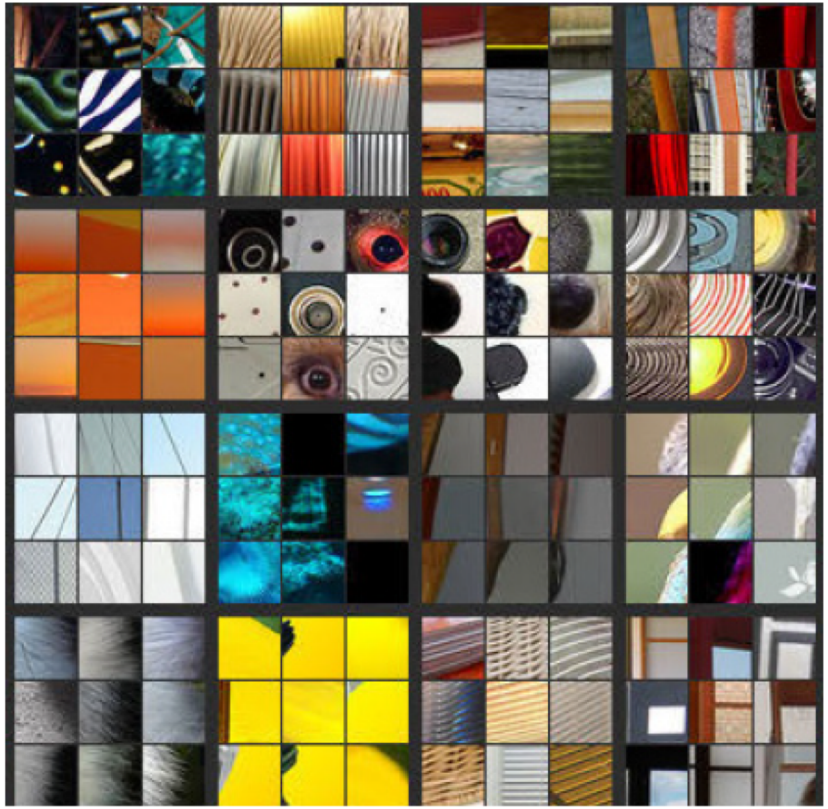


Layer 1

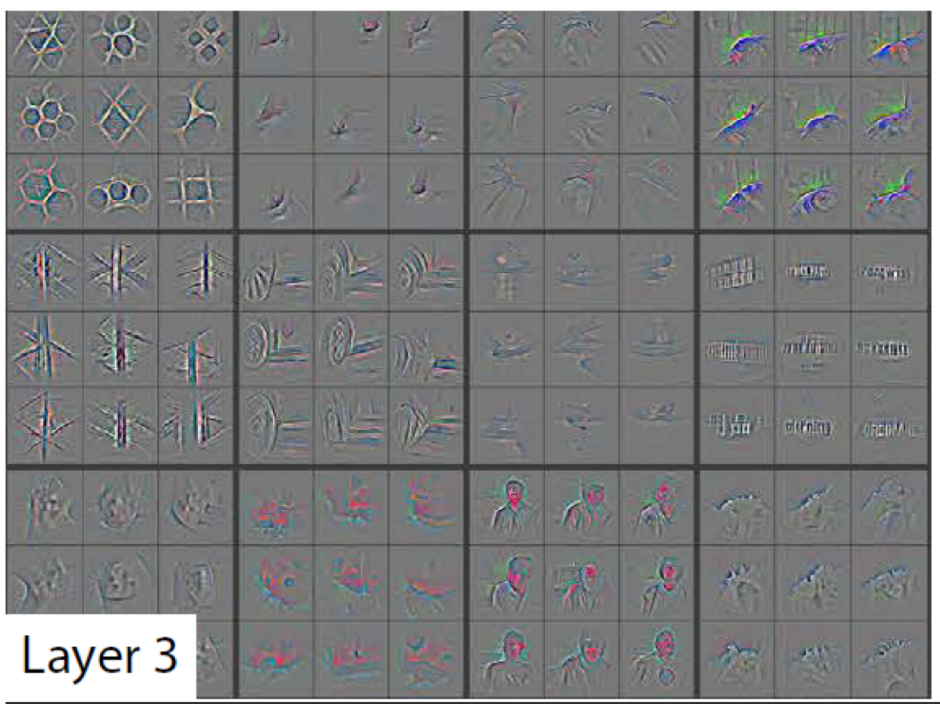


What does a CNN learn?

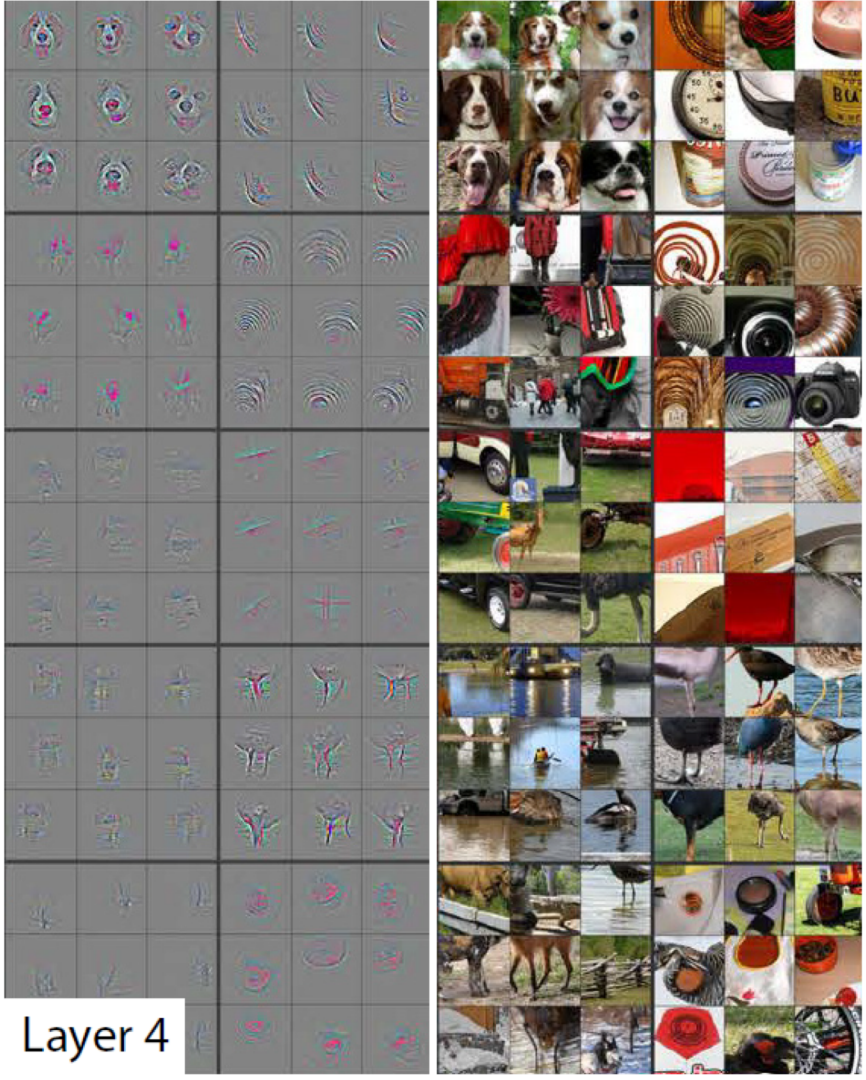




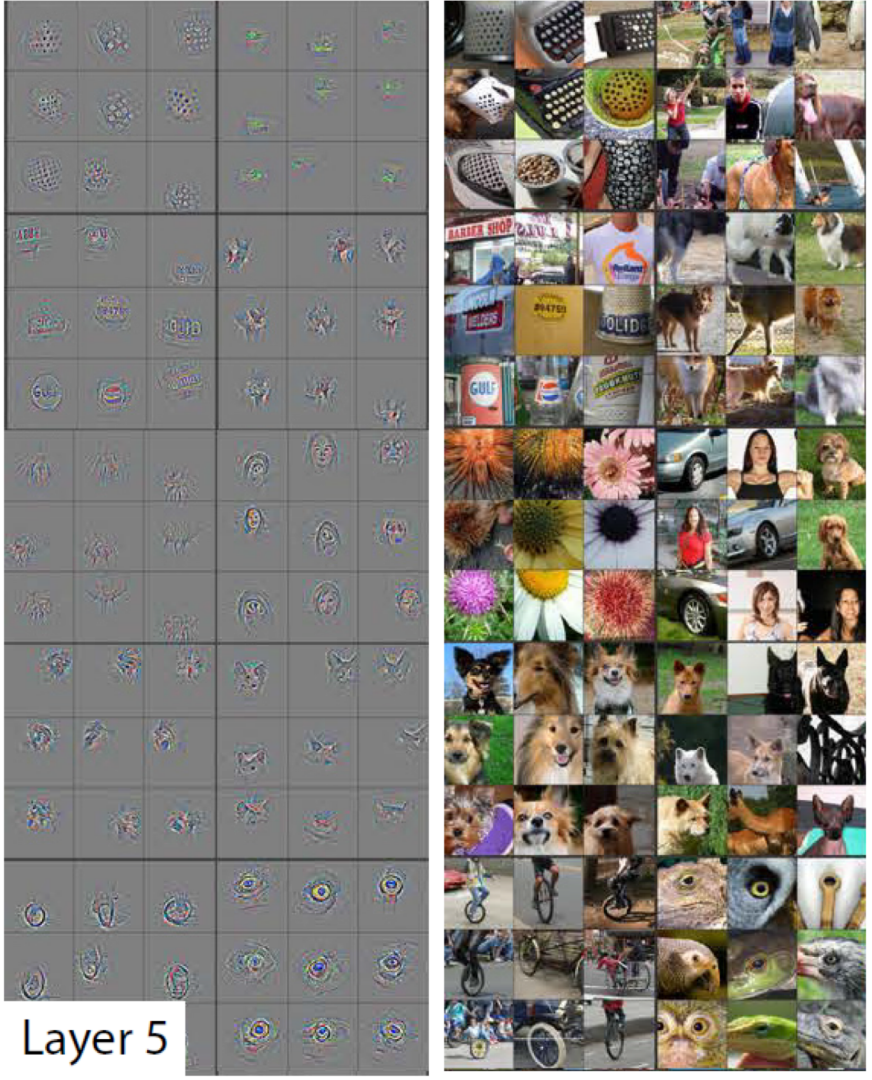
What does a CNN learn?



What does a CNN learn?



What does a CNN learn?



Human Action Recognition (HAR)

- **RECOGNIZE the ACTION that a HUMAN is performing in a scene**

Why it’s important?

- **Surveillance systems:**
  - Security, patients monitoring

- **Video indexing;**

- **Motion capture.**

Challenges

- Abrupt appearance changes:
  occlusions, viewpoint, light, background, cameras…

- Interclass similarity and intraclass variation

Hugging class

Karate

Taekwondo

Challenges

- The action concept is not well defined:
  - Ambiguous classes
  - “Action atomicity”



Action: open

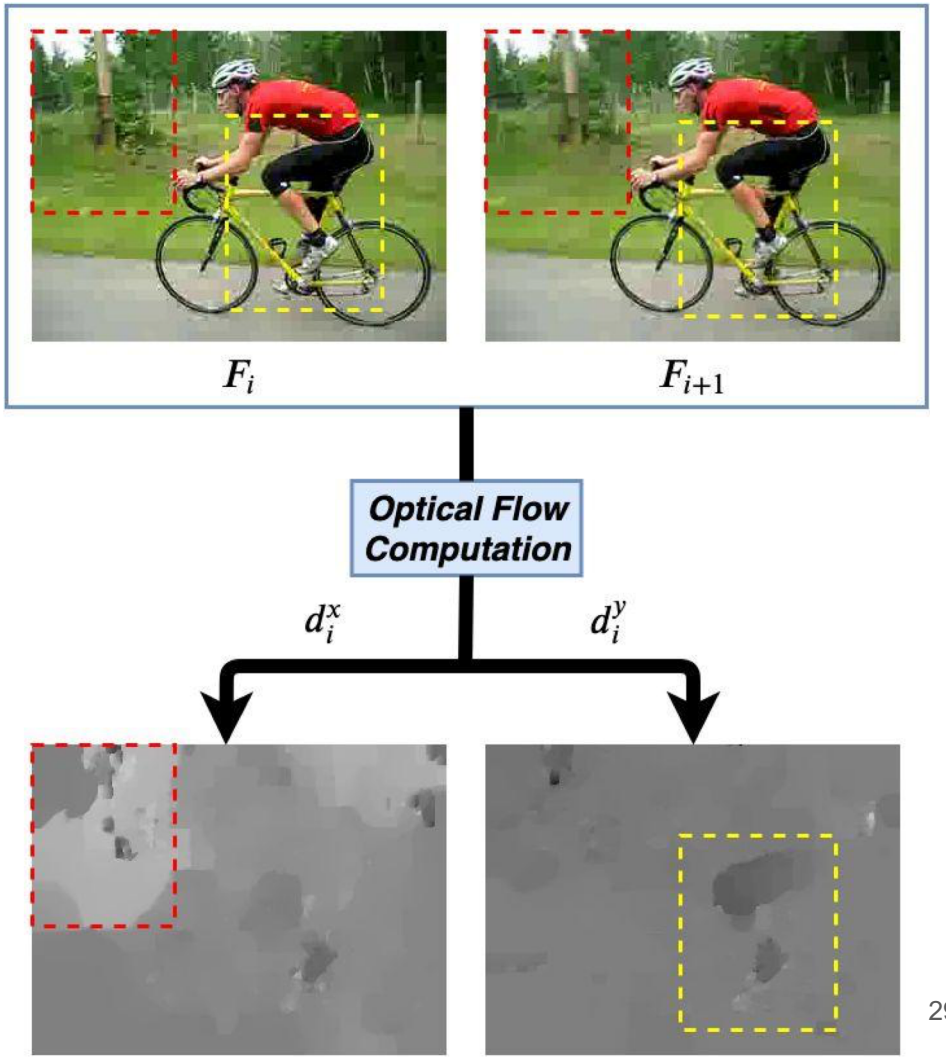
Action: rise the arms

Action: play soccer

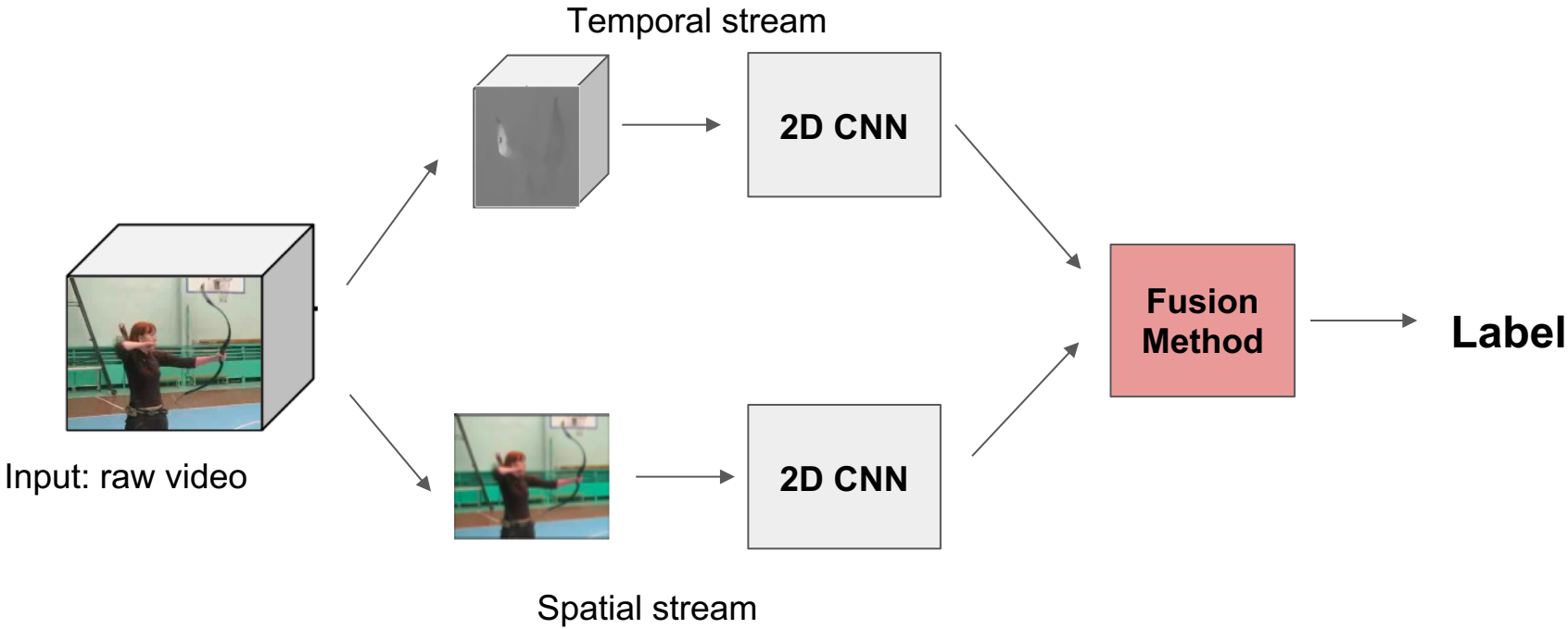
Action: horse riding

Video representation

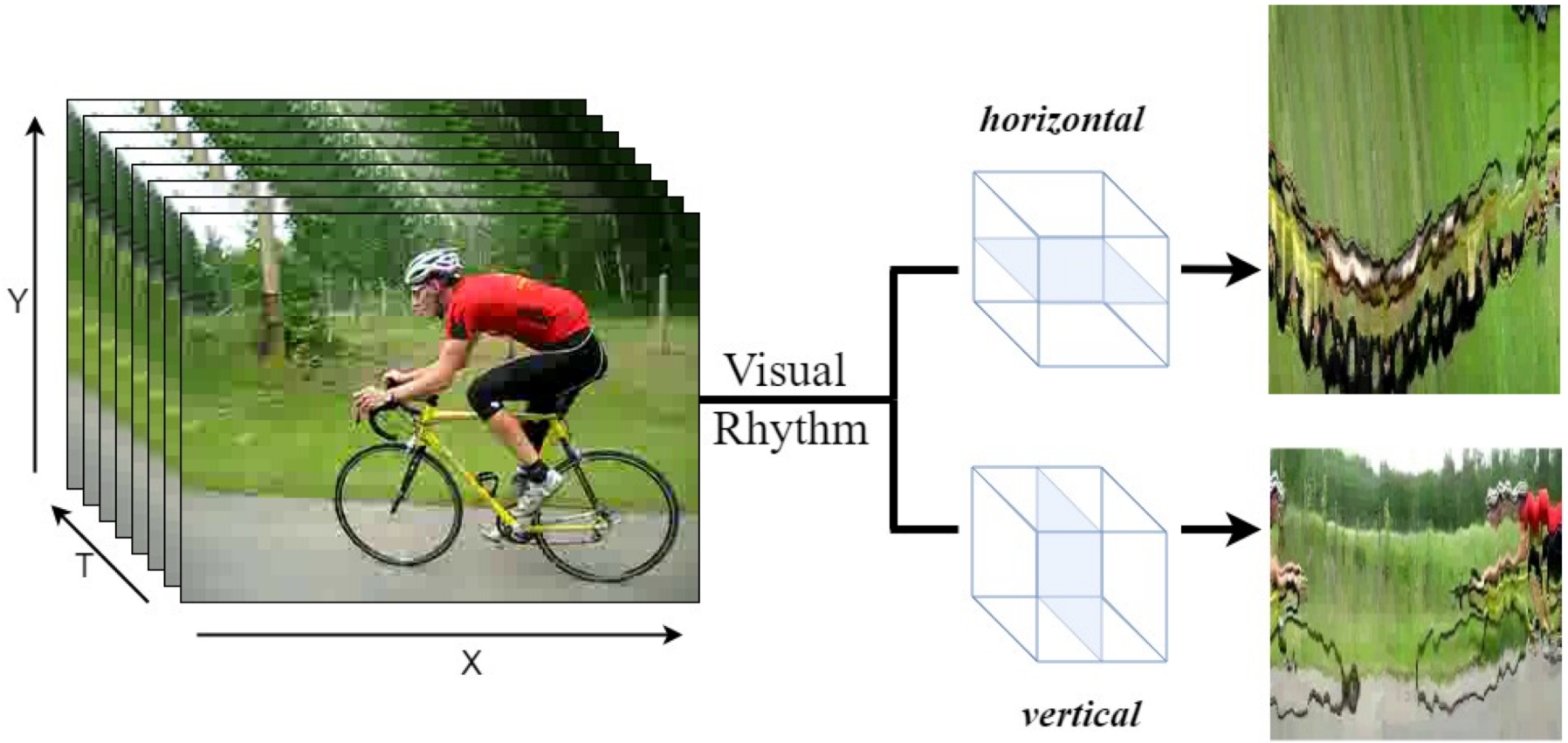
- RGB frame
- Optical Flow



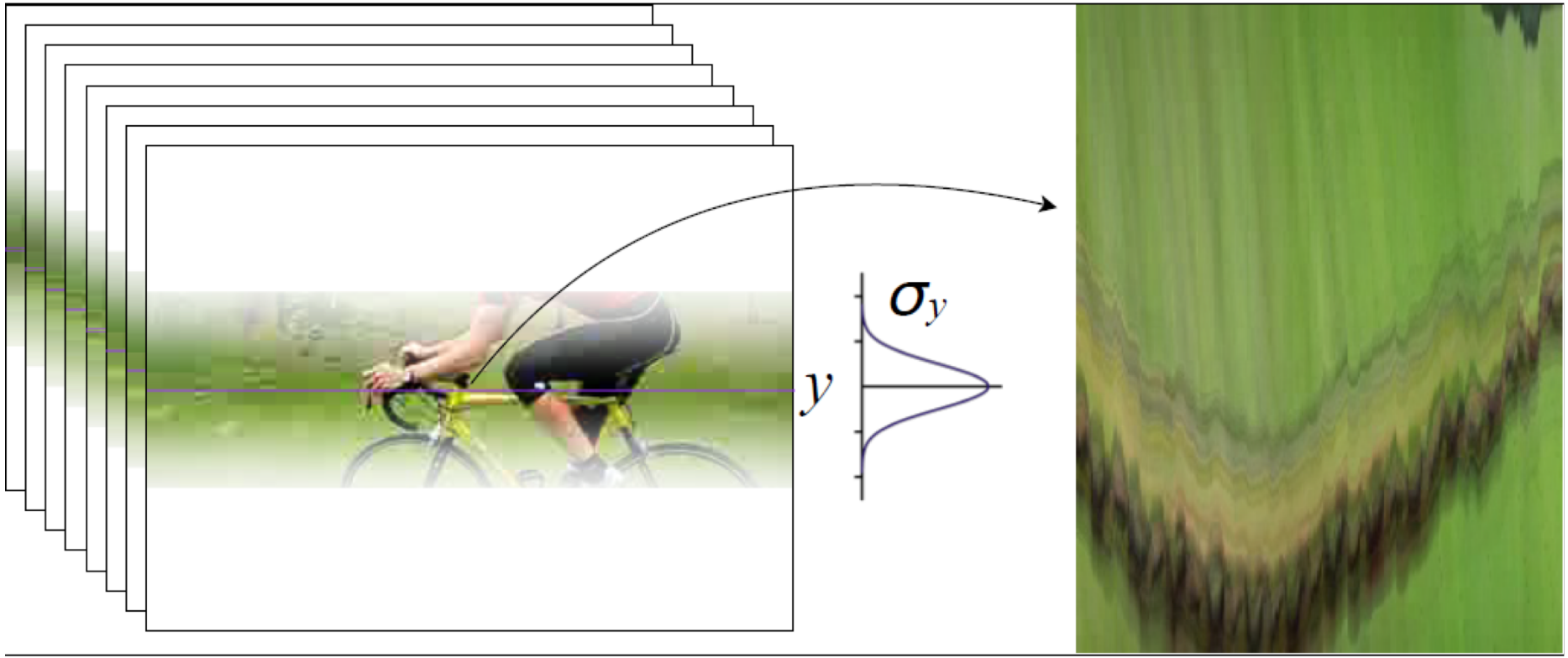
Two-Stream Architecture



Visual Rhythm

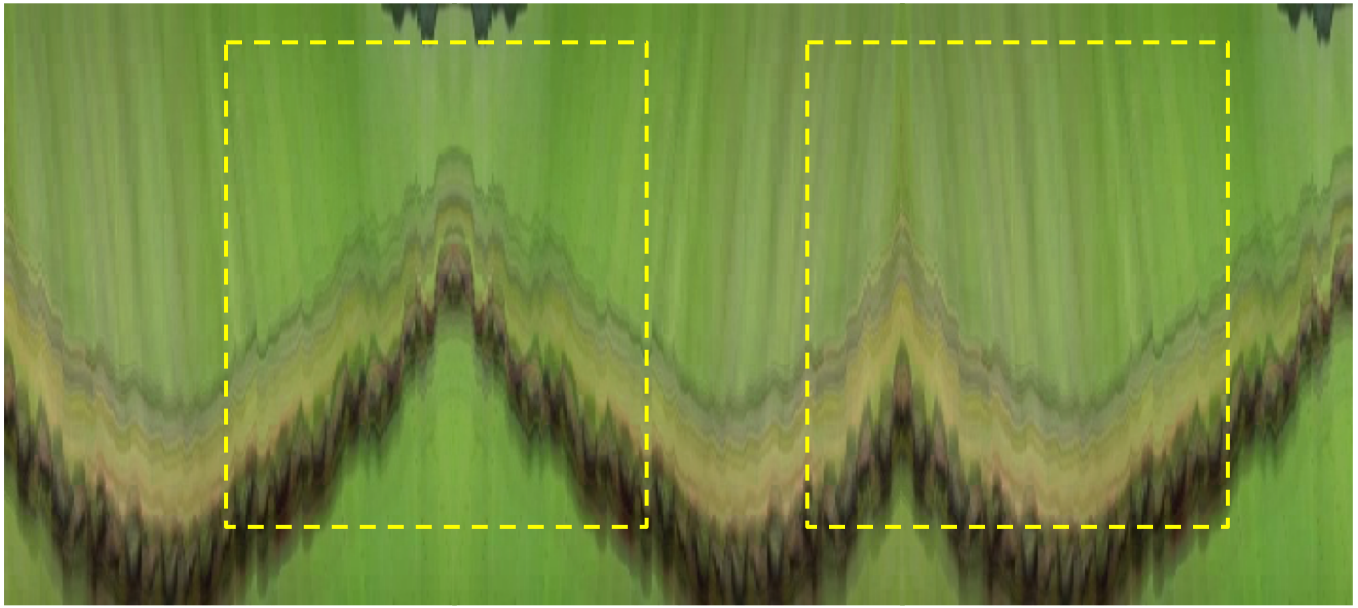


Weighted Visual Rhythm

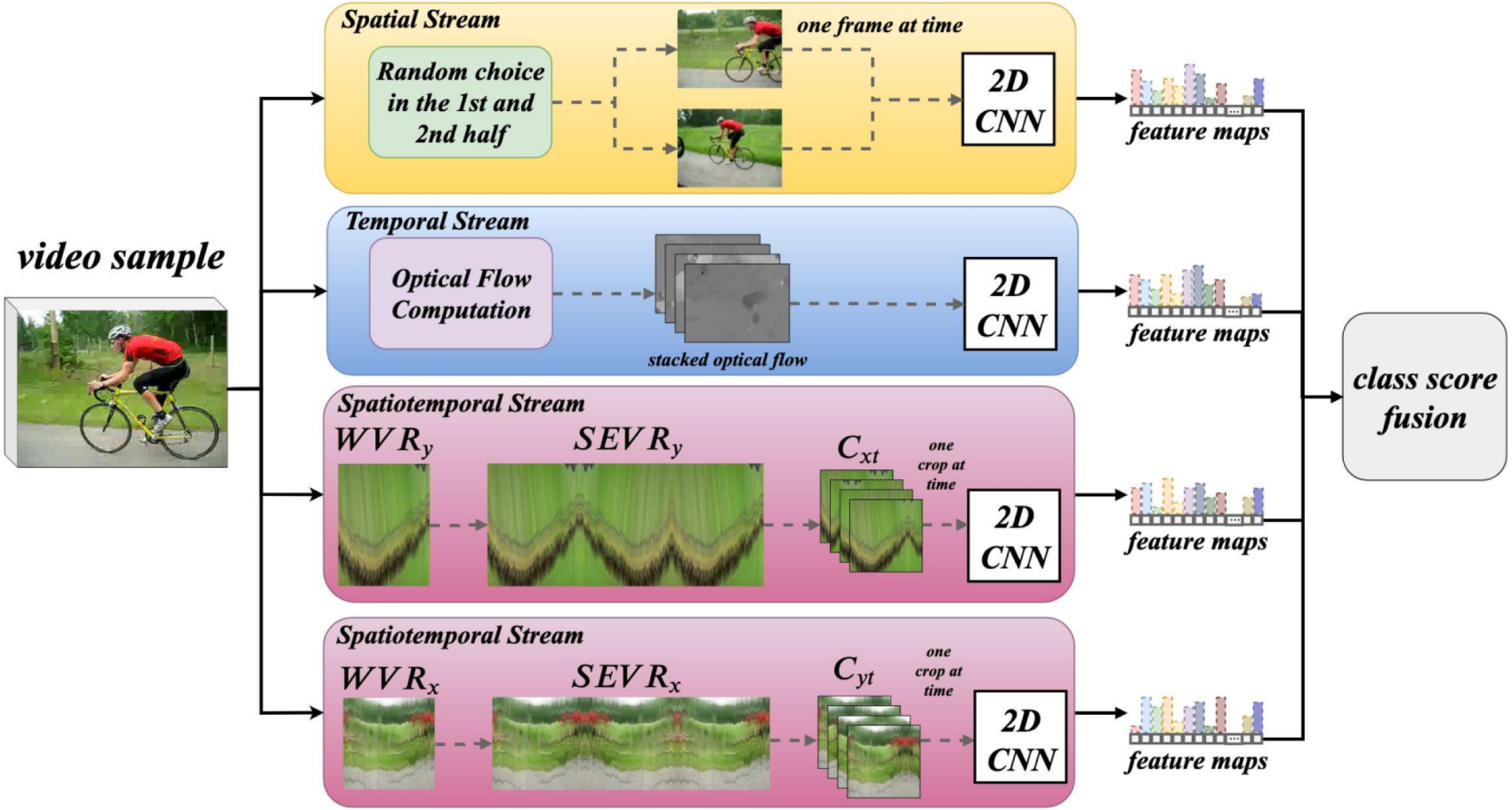


Weighted Visual Rhythm

Symmetric Extended Visual Rhythm



Multi-stream Architecture



How to assess the models?

- Datasets: UCF101 (13320 videos, 101 actions) and HMDB51 (6766 videos, 51 actions)

- Mean accuracy obtained from 3 splits of each dataset

Results

Streams	UCF101(%)	HMDB51(%)
OF + RGB*	93.21	66.43
OF + RGB* + $SEVR_y$	93.70	67.15
OF + RGB* + $SEVR_y$ + $SEVR_x$	94.06	67.73

References

- Two-stream Convolutional Networks for Action Recognition in Videos
- Towards Good Practices for very Deep Two-Stream Convnets
- Multi-stream Convolutional Neural Networks for Action Recognition in Video Sequences Based on Adaptive Visual Rhythms
- Human Action Recognition Using Convolutional Neural Networks with Symmetric Time Extension of Visual Rhythms
- Data augmentation of visual rhythms using symmetric extension for deep learning video based human action recognition

Thank you!

Questions?

- Get in touch with me:
  - e-mail: hemerson.tacon@talkdesk.com
  - LinkedIn: https://www.linkedin.com/in/hemerson-tacon/
  - GitHub: https://github.com/HemersonTacon