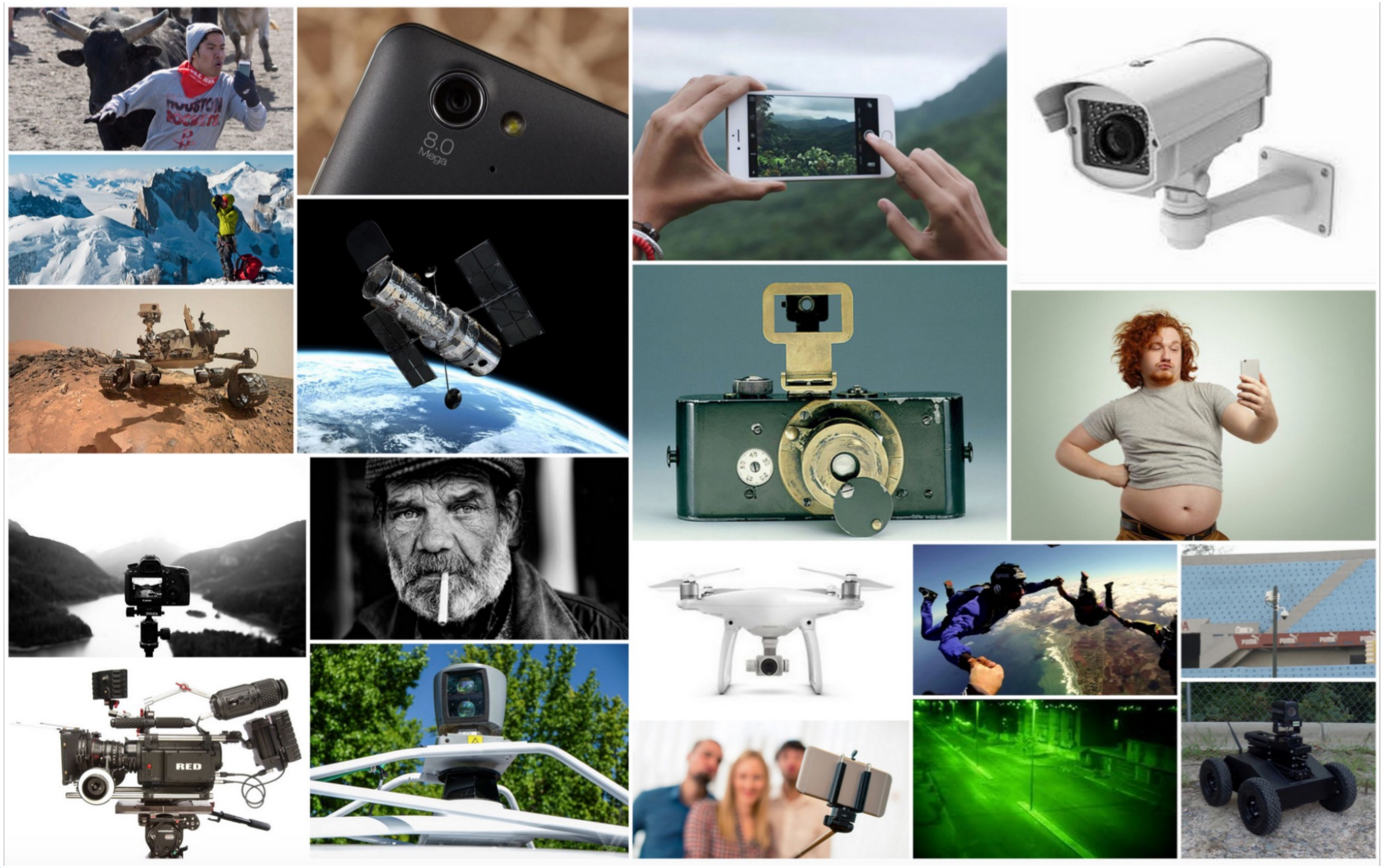
Deep Learning
Last trends and applications

Joan Duran Grimalt
Tractament i Anàlisi Matemàtica d'Imatges (TAMI)
Departament de Ciències Matemàtiques i Informàtica
joan.duran@uib.es joandg @joan_dg

Universitat de les Illes Balears

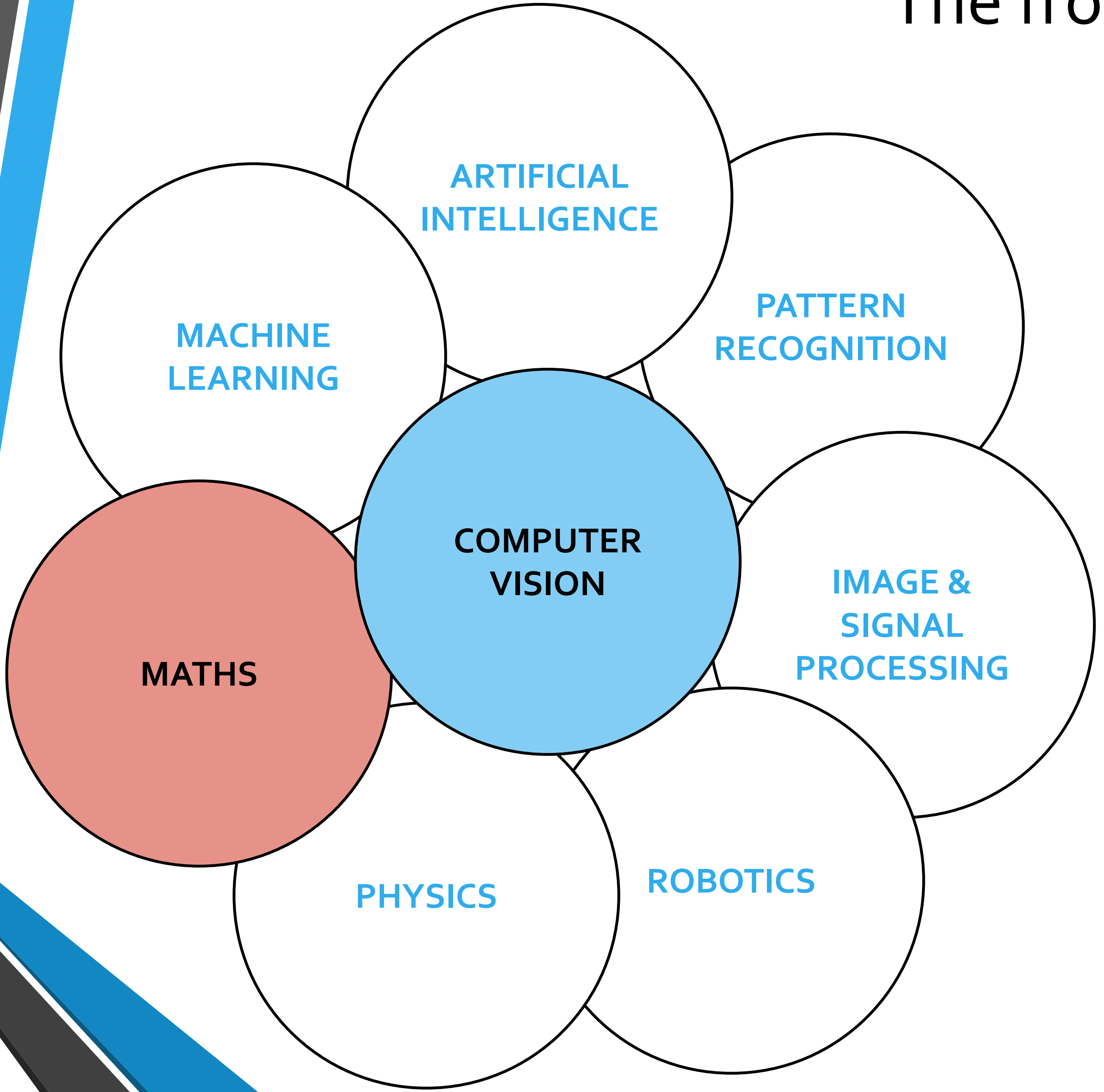
A world full of images



A world full of images

A world full of images

The frontiers of computer vision



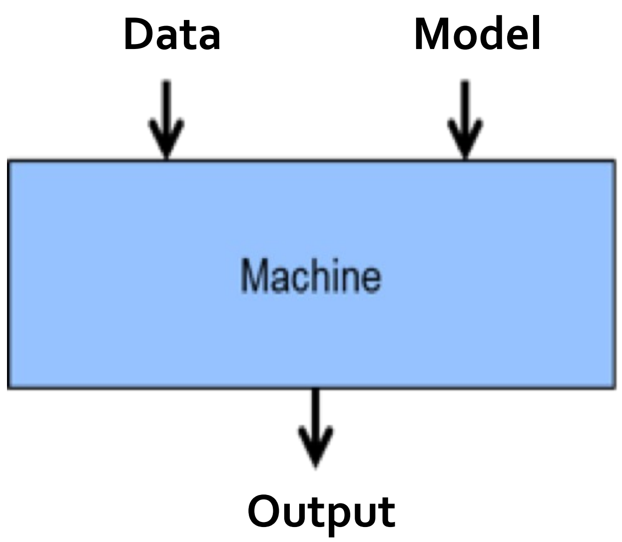
- **Computer vision**
  - is an interdisciplinary field of research that deals with how computers can be made for gaining high-level understanding from digital images or videos.
  - seeks to automate tasks that the human visual system can do.

Deep Learning for Computer Vision

The Good, the Bad and the Ugly

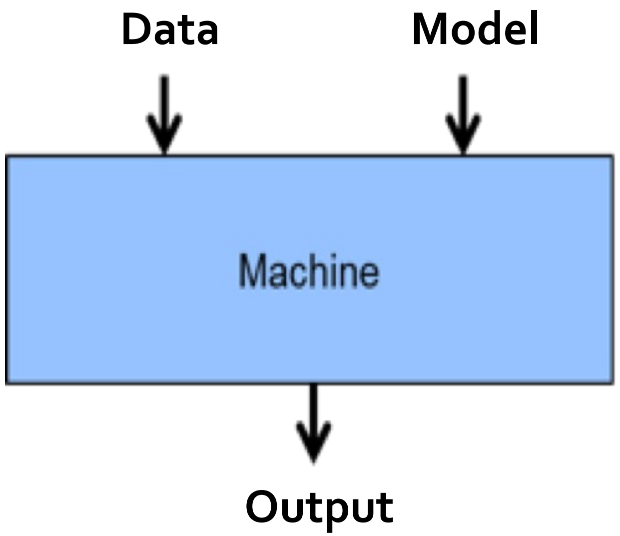
From modelling to learning

- Traditional programming designs a **precise model** for the desired task.



From modelling to learning

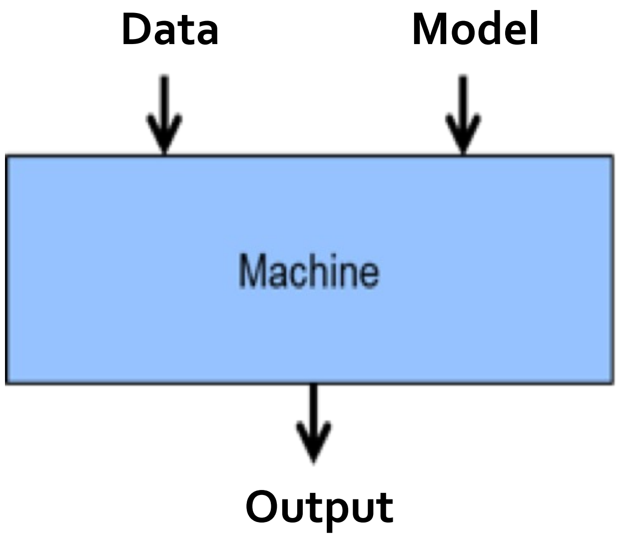
- Traditional programming designs a precise model for the desired task.
- The definition of such a model may be extremely complicated.

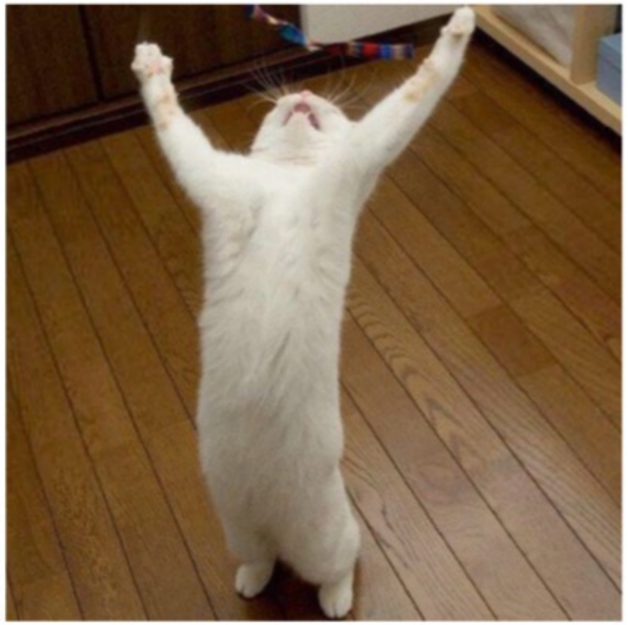


How can we define "breakfast" mathematically ?

From modelling to learning

- Traditional programming designs a precise model for the desired task.
- The definition of such a model may be extremely complicated.

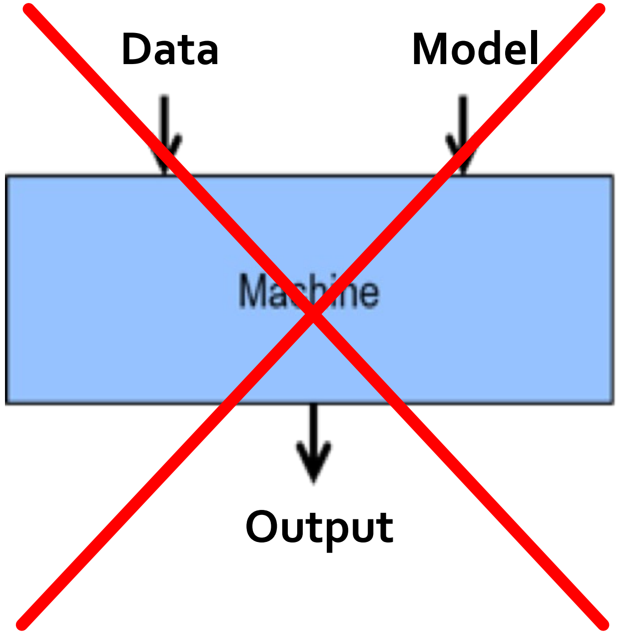


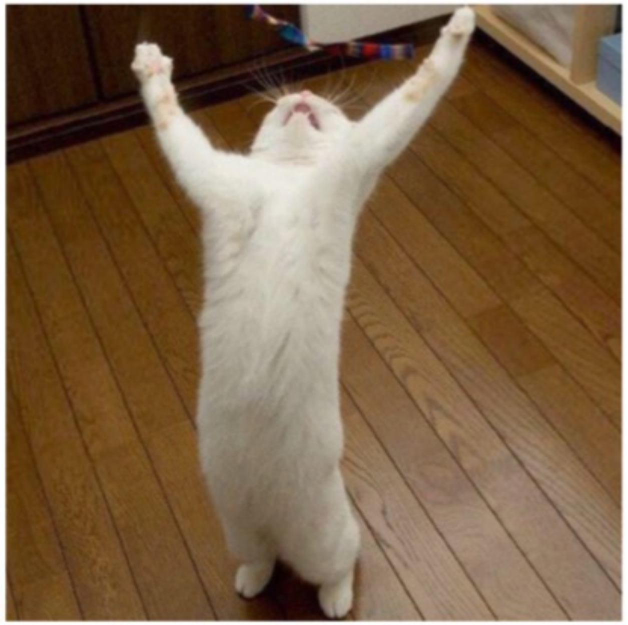


How can we define "cat" mathematically ?

From modelling to learning

- Traditional programming designs a precise model for the desired task.
- The definition of such a model may be extremely complicated.

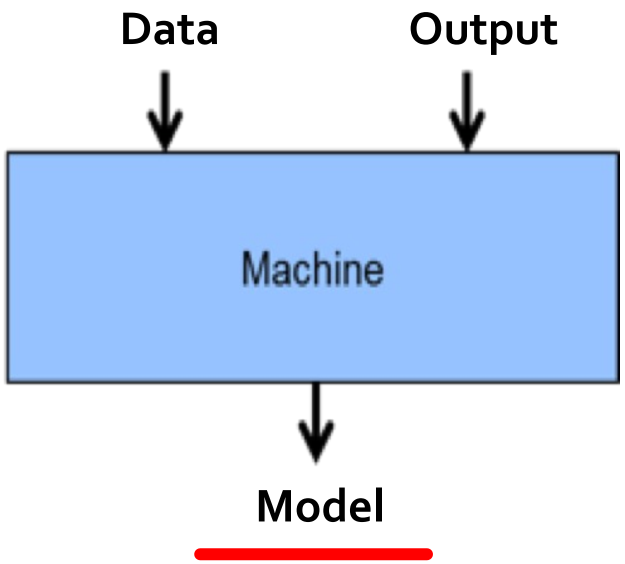




How can we define "cat" mathematically ?

From modelling to learning

- Traditional programming designs a precise model for the desired task.
- The definition of such a model may be extremely complicated.
- The emergence of large **labelled datasets** allows **learning** from examples.



Machine learning

- With **massive computational power**, large datasets can be tackled through learning techniques.

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- **Machine learning** gives computers the ability to learn without need of explicit model.

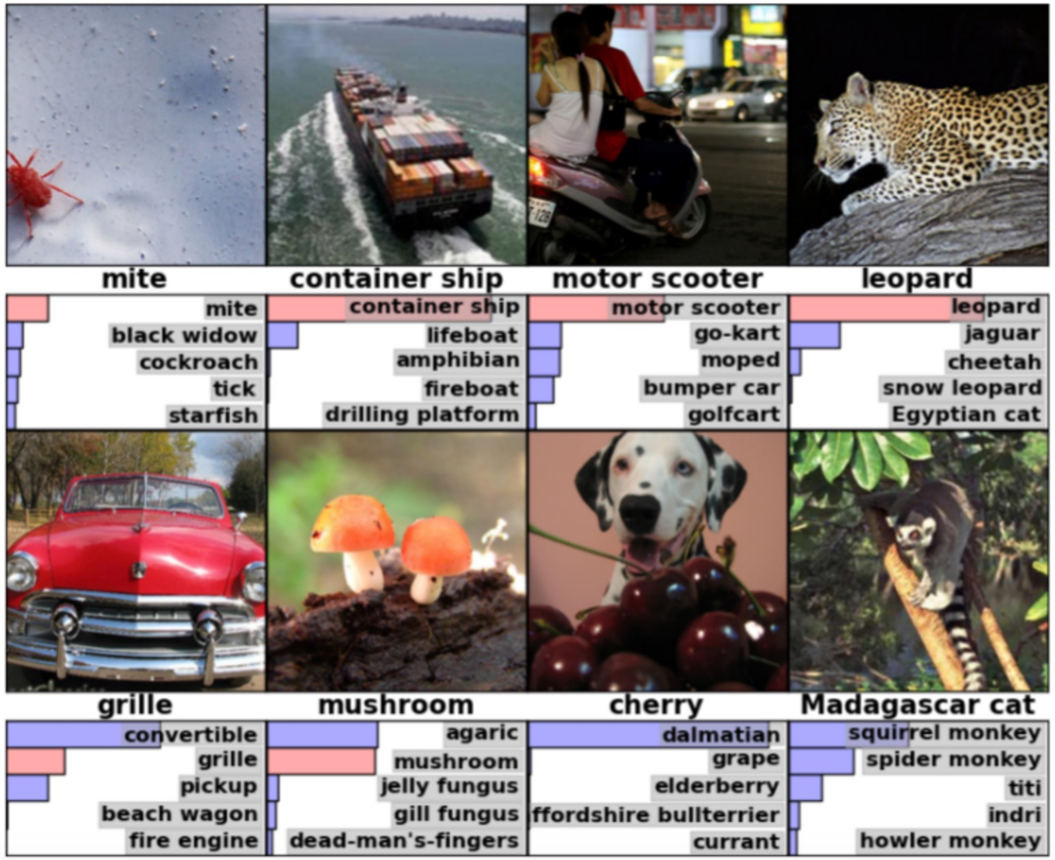
Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- **Unsupervised learning** and supervised learning.

Segmentation

Machine learning

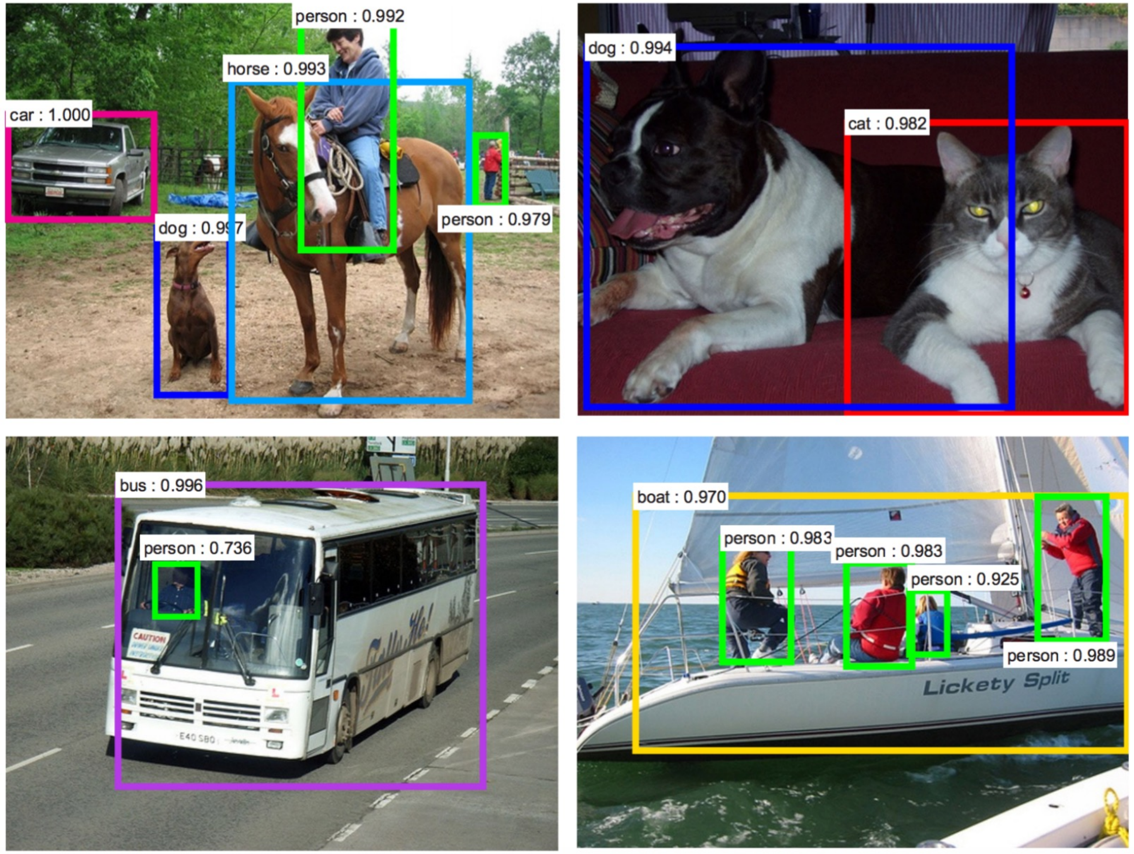
- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and **supervised learning**.



Classification

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and **supervised learning**.



Detection & Classification

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and **supervised learning**.

Labelling

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and **supervised learning**:

  - Dataset of pairs $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)\}$, usually $n \gg 1$.
  - The goal is to find a function f that best maps each $\boldsymbol{x}_i$ to each y_i.
  - One commonly chooses a family of *parametric functions* $f(\boldsymbol{x};\theta) = f_\theta(\boldsymbol{x})$.
  - Define a *loss function* $\mathcal{L}(f_\theta(\boldsymbol{x}), y)$ between predictions $f_\theta(\boldsymbol{x}_i)$ and labels y_i.
  - Choose a method to minimize the loss for *each* (?) pair in the dataset.
  - Learning "reduces" to solve the **optimization problem**

$$\arg\min_\theta \sum_{i=1}^{n} \mathcal{L}(f_\theta(\boldsymbol{x}_i), y_i)$$

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and **supervised learning**:

- Dataset of pairs $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)\}$, $n \gg 1$.
- The goal is to find a function f that best maps each $\boldsymbol{x}_i$ to each y_i.
- One commonly chooses a family of *parametric functions* $f(\boldsymbol{x}; \theta) = f_\theta(\boldsymbol{x})$.
- Define a *loss function* $\mathcal{L}(f_\theta(\boldsymbol{x}), y)$ between predictions $f_\theta(\boldsymbol{x}_i)$ and labels y_i.
- Choose a method to minimize the loss for *each* (?) pair in the dataset.
- Learning "reduces" to solve the **optimization problem**

$$\arg\min_\theta \sum_{i=1}^{n} \mathcal{L}\big(f_\theta(\boldsymbol{x}_i), y_i\big)$$

THE BLACK BOX

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and **supervised learning**:

  - Dataset of pairs $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)\}$, $n \gg 1$.
  - The goal is to find a function f that best maps each $\boldsymbol{x}_i$ to each y_i.
  - One commonly chooses a family of *parametric functions* $f(\boldsymbol{x};\theta) = f_\theta(\boldsymbol{x})$.
  - Define a *loss function* $\mathcal{L}(f_\theta(\boldsymbol{x}), y)$ between predictions $f_\theta(\boldsymbol{x}_i)$ and labels y_i.
  - Choose a method to minimize the loss for *each* (?) pair in the dataset.
  - Learning "reduces" to solve the **optimization problem**

$$\arg\min_\theta \sum_{i=1}^{n} \mathcal{L}(f_\theta(\boldsymbol{x}_i), y_i)$$

Traditional programming

Modelling function f

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without without need of explicit model.
- Unsupervised learning and **supervised learning**:

  - Dataset of pairs $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)\}$, $n \gg 1$.
  - The goal is to find a function f that best maps each $\boldsymbol{x}_i$ to each y_i.
  - One commonly chooses a family of *parametric functions* $f(\boldsymbol{x}; \theta) = f_\theta(\boldsymbol{x})$.
  - Define a *loss function* $\mathcal{L}(f_\theta(\boldsymbol{x}), y)$ between predictions $f_\theta(\boldsymbol{x}_i)$ and labels y_i.
  - Choose a method to minimize the loss for *each* (?) pair in the dataset.
  - Learning "reduces" to solve the **optimization problem**

$$\arg\min_\theta \sum_{i=1}^{n} \mathcal{L}(f_\theta(\boldsymbol{x}_i), y_i)$$

Machine Learning

Modelling function f - Hyperparameters $\boldsymbol{\theta}$

Machine learning

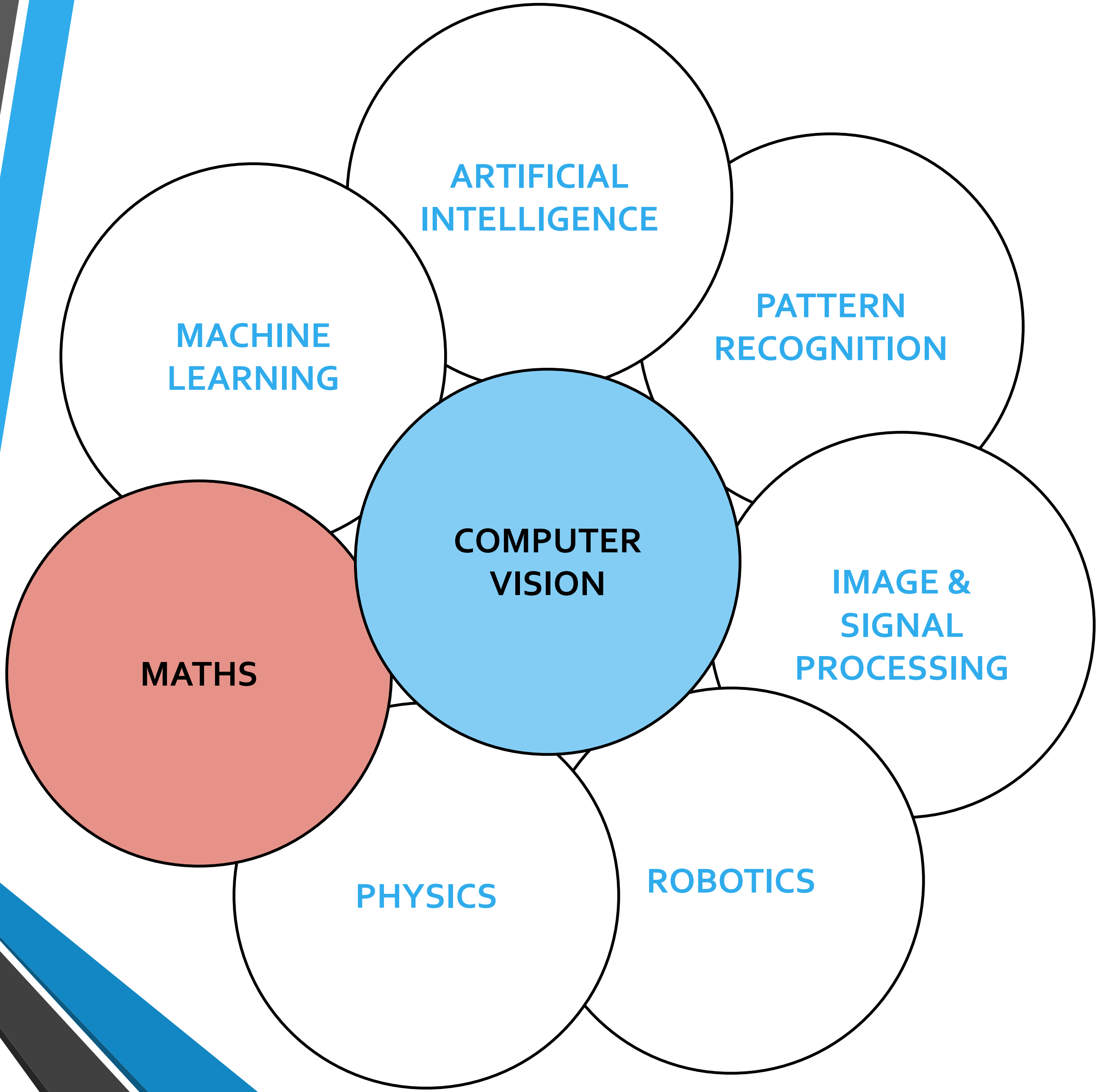
- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and supervised learning.
- **Challenges in optimization**:
  - Ill-conditioning.
  - Local minima in non-convex problems.
  - Saddle points and other flat regions.
  - Cliffs and exploiting gradients.
  - Long-term dependencies.
  - Inexact gradients.
  - Poor correspondence between local and global structure.
  - Theoretical limits of optimization.

Machine learning

- With massive computational power, large datasets can be tackled through learning techniques.
- Machine learning gives computers the ability to learn without need of explicit model.
- Unsupervised learning and supervised learning.
- **Challenges in optimization**:
  - Ill-conditioning.
  - Local minima in non-convex problems.
  - Saddle points and other flat regions.
  - Cliffs and exploiting gradients.
  - Long-term dependencies.
  - Inexact gradients.
  - Poor correspondence between local and global structure.
  - Theoretical limits of optimization.

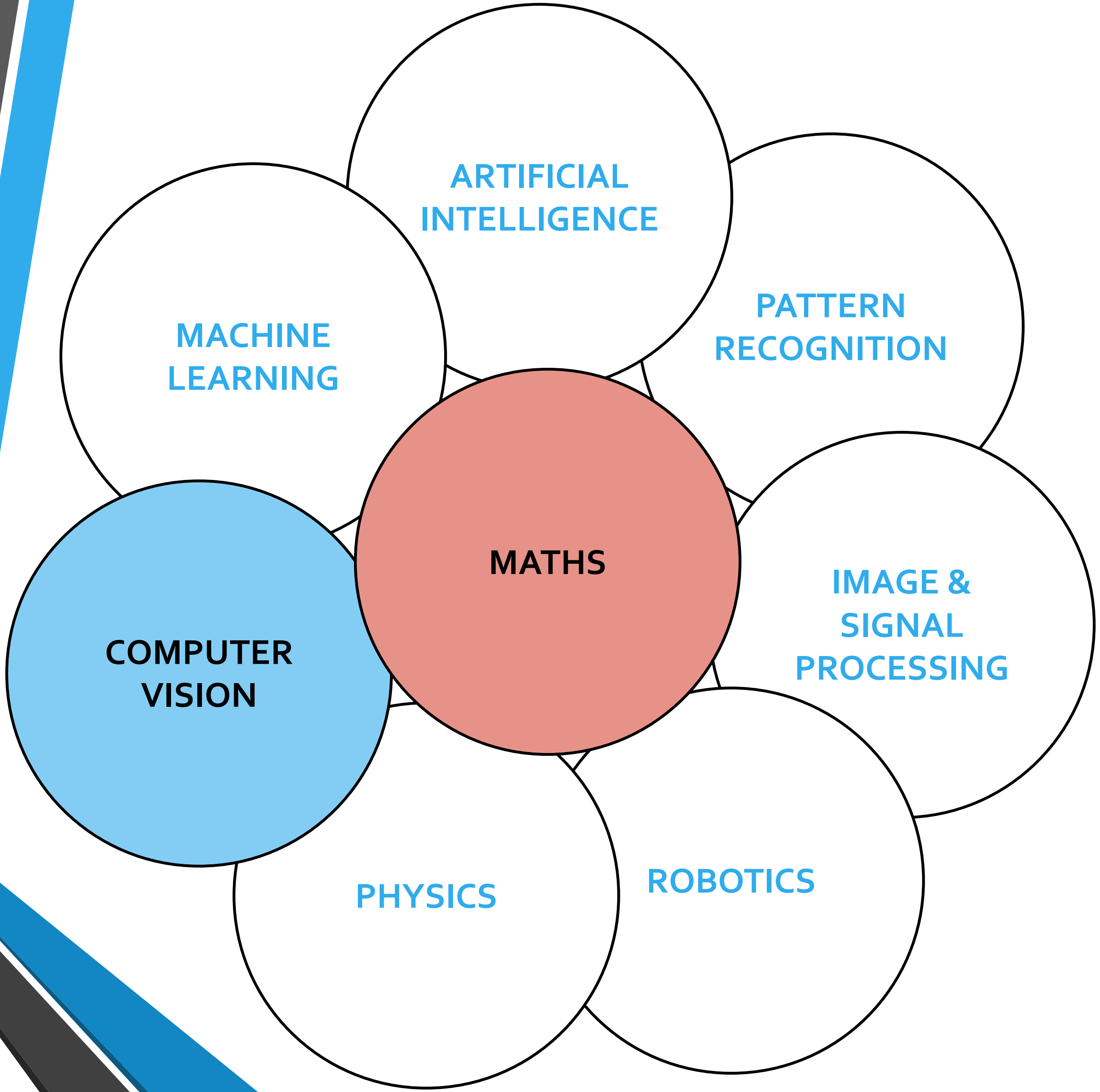
Blind faith in the "black box" is critical. Mathematics are a must!

Mathematics are a must



- **Computer vision**
  - is an interdisciplinary field of research that deals with how computers can be made for gaining high-level understanding from digital images or videos.
  - seeks to automate tasks that the human visual system can do.

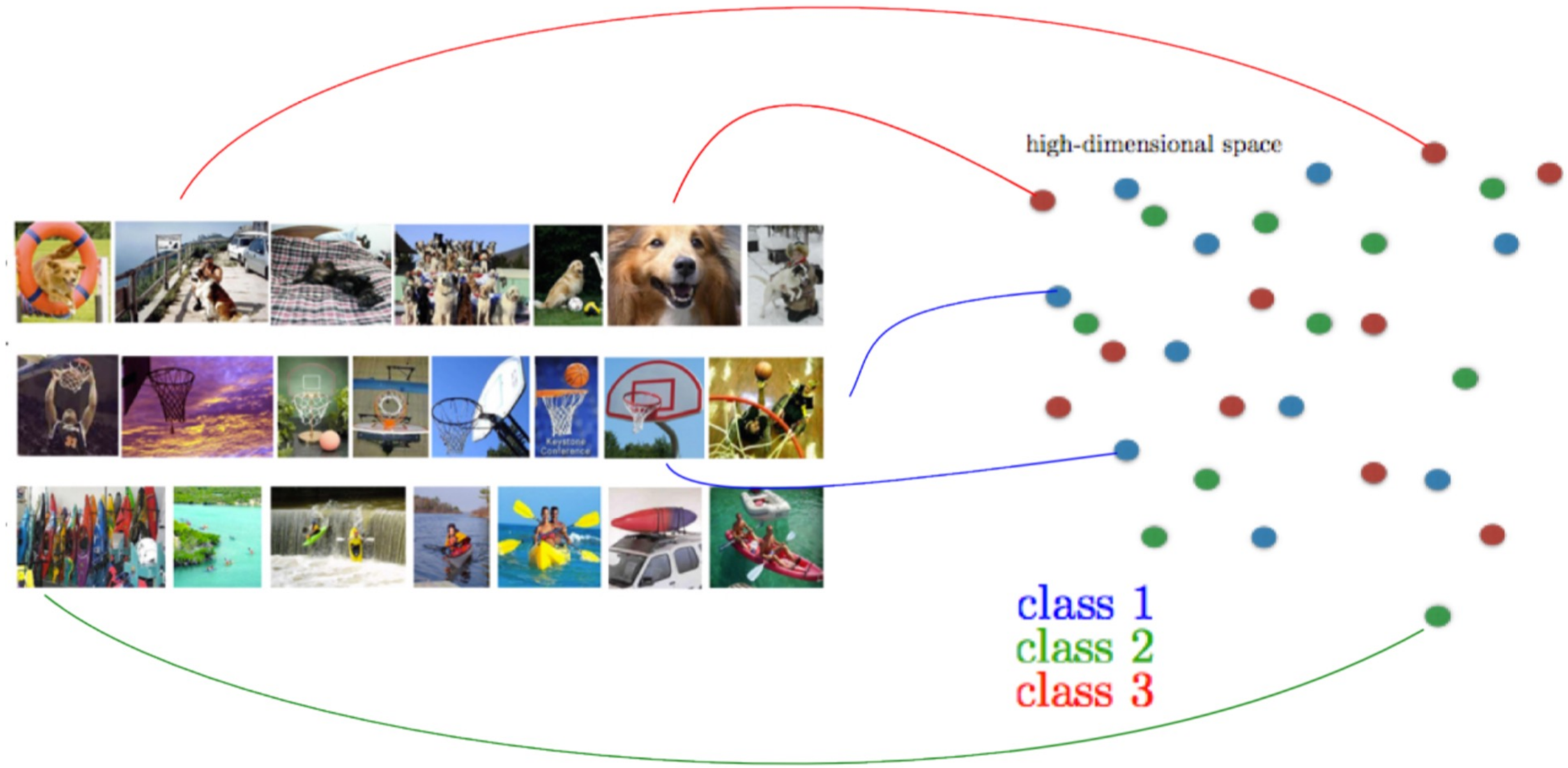
Mathematics are a must



- **Computer vision**
  - is an interdisciplinary field of research that deals with how computers can be made for gaining high-level understanding from digital images or videos.
  - seeks to automate tasks that the human visual system can do.

Images are challenging!

- **High-dimensional data** makes difficult to learn complicated functions in high-dimensional spaces.



Images are challenging!

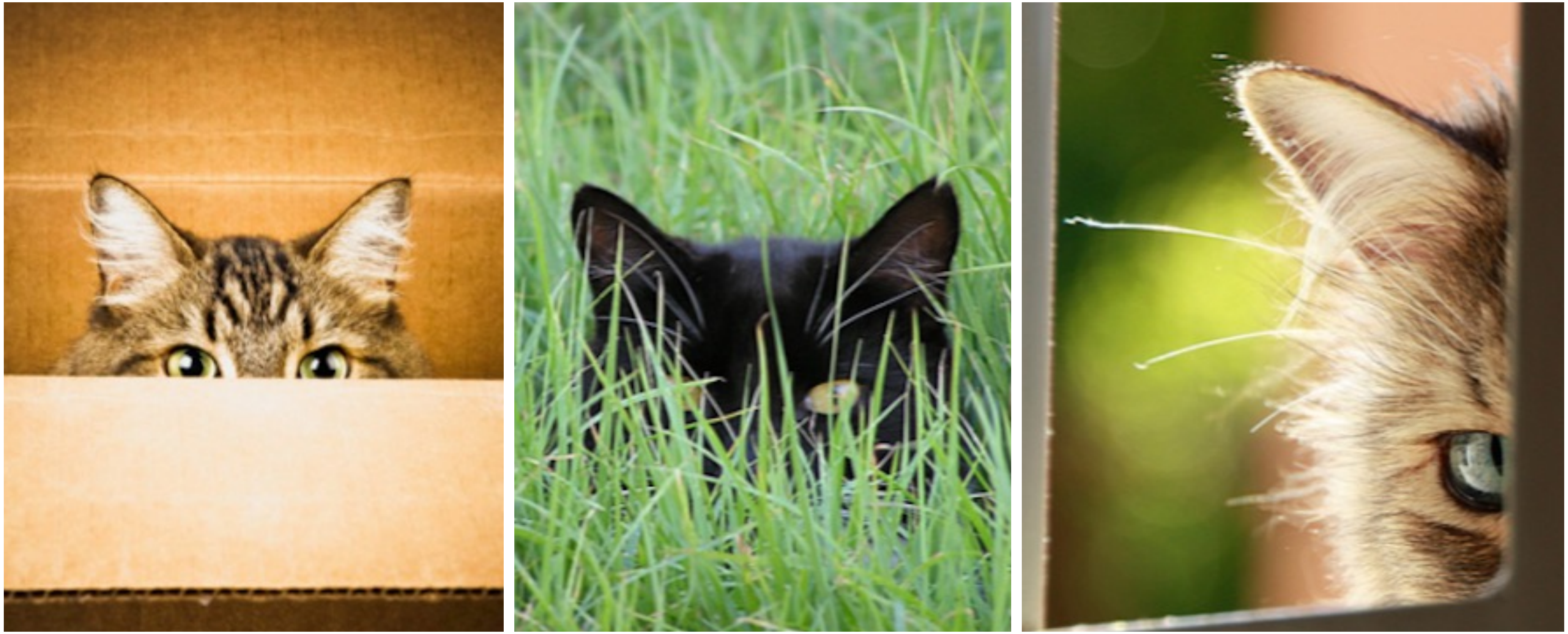
- High-dimensional data makes difficult to learn complicated functions in high-dimensional spaces.
- Huge **intra-class representation variety**: deformations, occlusions, illumination, viewpoints, etc.

Images are challenging!

- High-dimensional data makes difficult to learn complicated functions in high-dimensional spaces.
- Huge **intra-class representation variety**: deformations, occlusions, illumination, viewpoints, etc.

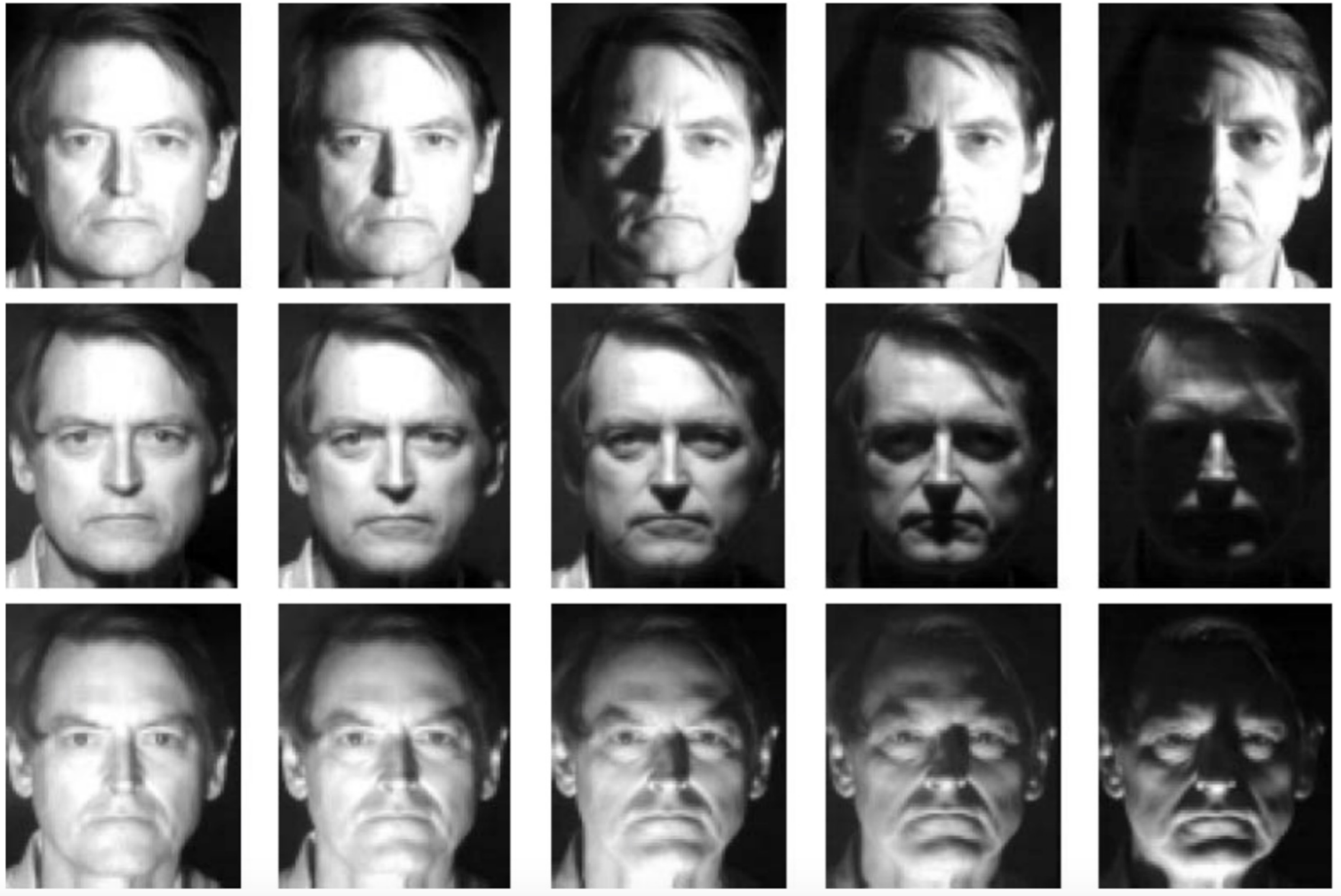
Images are challenging!

- High-dimensional data makes difficult to learn complicated functions in high-dimensional spaces.
- Huge **intra-class representation variety**: deformations, occlusions, illumination, viewpoints, etc.



Images are challenging!

- High-dimensional data makes difficult to learn complicated functions in high-dimensional spaces.
- Huge **intra-class representation variety**: deformations, occlusions, illumination, viewpoints, etc.



Images are challenging!

- High-dimensional data makes difficult to learn complicated functions in high-dimensional spaces.
- Huge **intra-class representation variety**: deformations, occlusions, illumination, viewpoints, etc.

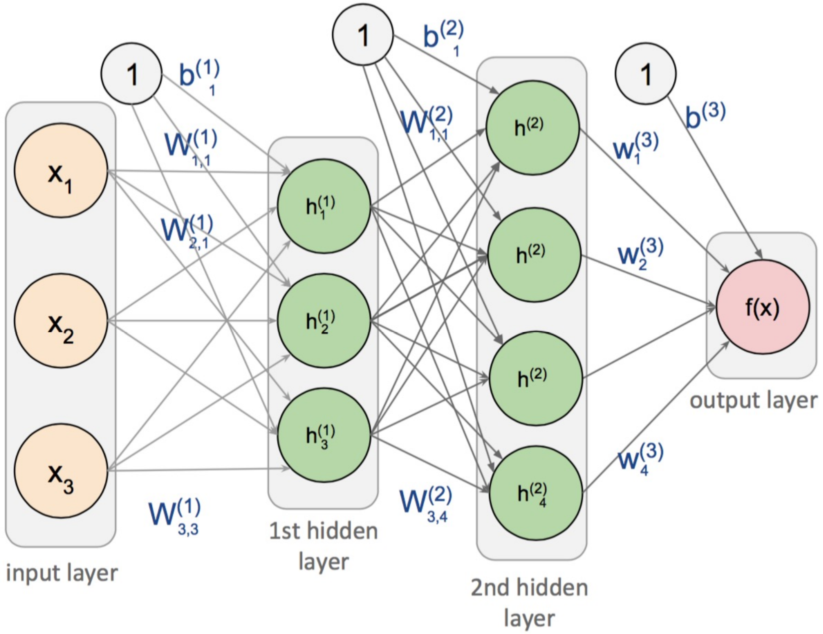


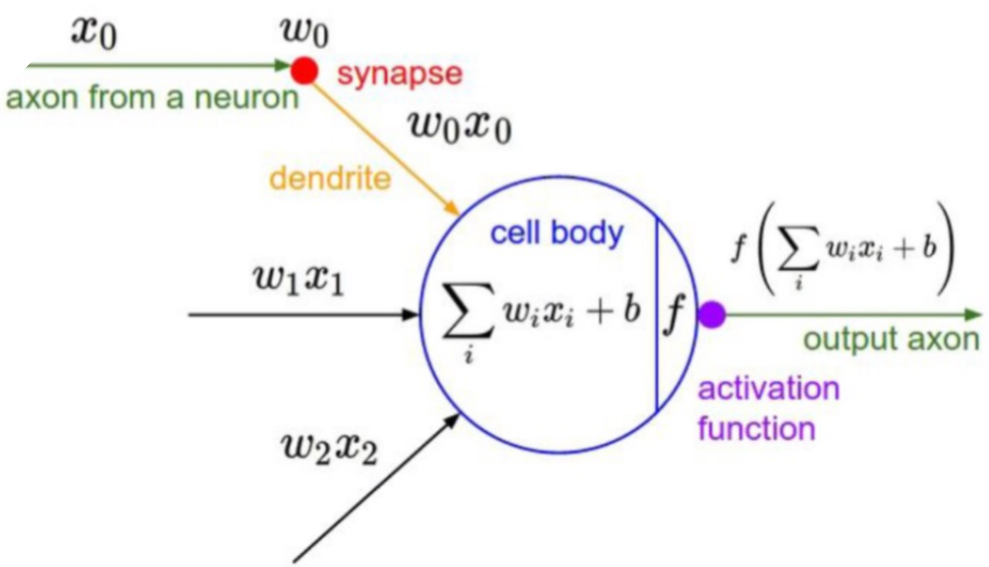

Deep Learning

- **Learning the representation** is nearly as difficult as to solve the primal task!

Deep Learning

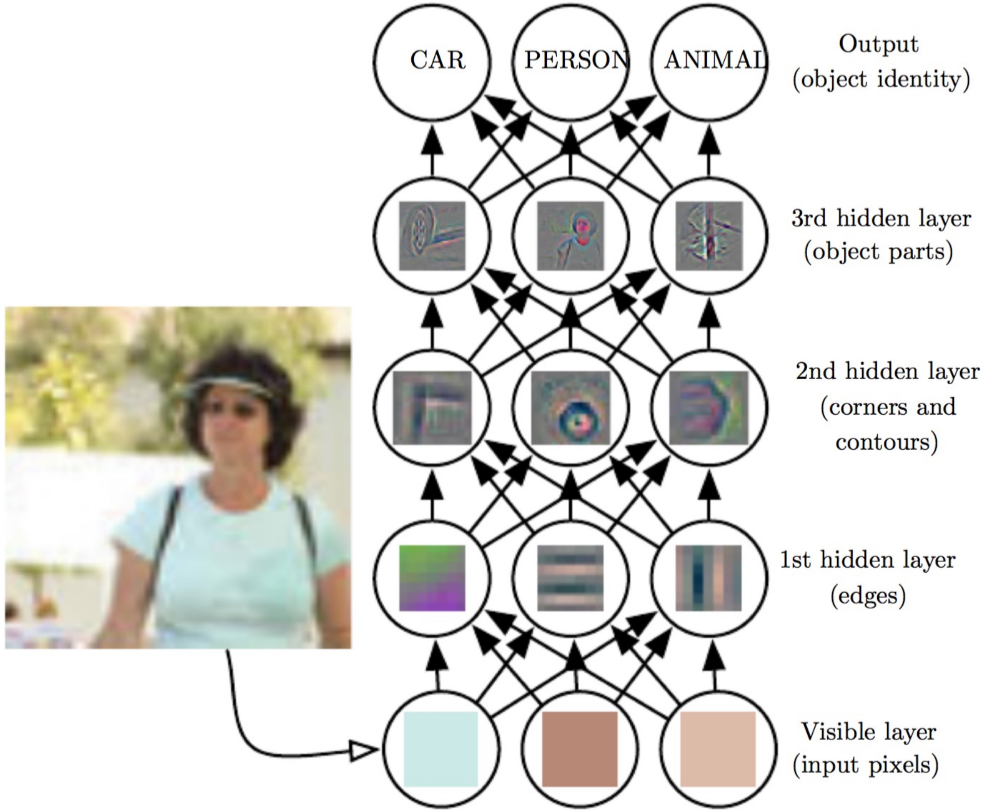
- Learning the representation is nearly as difficult as to solve the primal task!
- **Deep learning**
  - uses a cascade of multiple layers for feature extraction and transformation;





Deep Learning

- Learning the representation is nearly as difficult as to solve the primal task!
- **Deep learning**
  - uses a cascade of multiple layers for feature extraction and transformation;
  - learns multiple levels of representations that correspond to different levels of abstraction;



Deep Learning

- Learning the representation is nearly as difficult as to solve the primal task!
- **Deep learning**
  - uses a cascade of multiple layers for feature extraction and transformation;
  - learns multiple levels of representations that correspond to different levels of abstraction;
  - uses some form of gradient descent for training via backpropagation:

$$\arg\min_{\theta} \sum_{i=1}^{n} \mathcal{L}\big(f_\theta(\boldsymbol{x}_i), y_i\big) + \lambda \cdot R(\theta)$$

Traditional programming

Modelling function f

Deep Learning

- Learning the representation is nearly as difficult as to solve the primal task!
- Deep learning
  - uses a cascade of multiple layers for feature extraction and transformation;
  - learns multiple levels of representations that correspond to different levels of abstraction;
  - uses some form of gradient descent for training via backpropagation:

$$\arg\min_{\theta} \sum_{i=1}^{n} \mathcal{L}\big(f_\theta(\boldsymbol{x}_i), y_i\big) + \lambda \cdot R(\theta)$$

Machine Learning

Modelling function f - Hyperparameters $\boldsymbol{\theta}$

Deep Learning

- Learning the representation is nearly as difficult as to solve the primal task!
- **Deep learning**
  - uses a cascade of multiple layers for feature extraction and transformation;
  - learns multiple levels of representations that correspond to different levels of abstraction;
  - uses some form of gradient descent for training via backpropagation:

$$\arg\min_{\theta} \sum_{i=1}^{n} \mathcal{L}\big(f_\theta(\boldsymbol{x}_i), y_i\big) + \lambda \cdot R(\theta)$$

Deep Learning

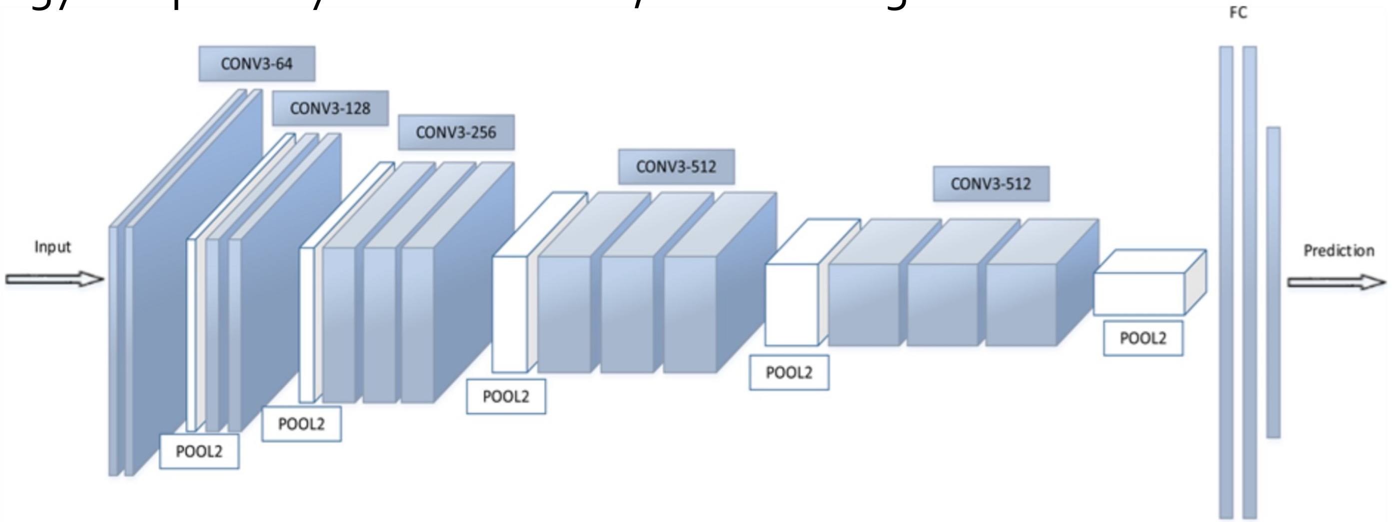
Modelling function $\boldsymbol{f}$ **-** Hyperparameters $\boldsymbol{\theta}$

Deep Learning

- Learning the representation is nearly as difficult as to solve the primal task!
- Deep learning
  - uses a cascade of multiple layers for feature extraction and transformation;
  - learns multiple levels of representations that correspond to different levels of abstraction;
  - uses some form of gradient descent for training via backpropagation:

$$\arg\min_{\theta} \sum_{i=1}^{n} \mathcal{L}\big(f_\theta(\boldsymbol{x}_i), y_i\big) + \lambda \cdot R(\theta)$$

- Convolutional neural networks use convolution + pooling operators to process data with known grid-like topology and possibly of variable size, such as images.



The Bad and the Ugly

- **Images are challenging**: noise, blur, perturbations, deformations, anomalies, etc.

The Bad and the Ugly

- **Images are challenging**: noise, blur, perturbations, deformations, anomalies, etc.

The Bad and the Ugly

- **Images are challenging**: noise, blur, perturbations, deformations, anomalies, etc.

Original image

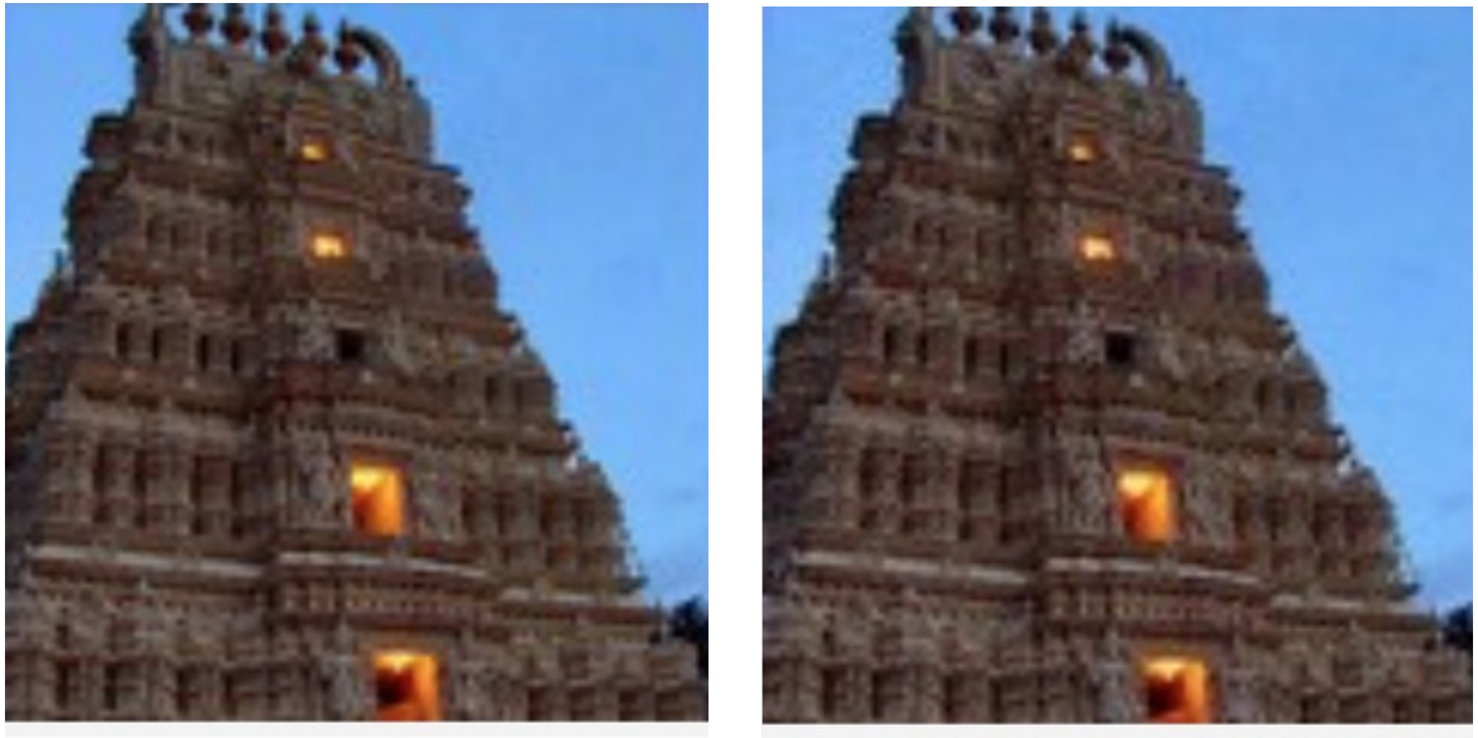
Temple (97%)

Adversarial example

Ostrich (98%)

The Bad and the Ugly

- **Images are challenging**: noise, blur, perturbations, deformations, anomalies, etc.



Original image — Temple (97%)

Adversarial example — Ostrich (98%)

Blind faith in the "black box" is critical. Image processing is a must!

The Bad and the Ugly

- Images are challenging: noise, blur, perturbations, deformations, anomalies, etc.
- **Challenges in optimization**.

The Bad and the Ugly

- Images are challenging: noise, blur, perturbations, deformations, anomalies, etc.
- Challenges in optimization.
- **Grid search** or the Survival of the Fittest.

The Bad and the Ugly

- Images are challenging: noise, blur, perturbations, deformations, anomalies, etc.
- Challenges in optimization.
- Grid search or the Survival of the Fittest.
- **Democratization** is good and not-so-good:
  - In Academia, people are distorted from the really important and harder questions.
  - In Industry, applying deep learning to a dataset is not going to make you *unique* anymore.

The Bad and the Ugly

- Images are challenging: noise, blur, perturbations, deformations, anomalies, etc.
- Challenges in optimization.
- Grid search or the Survival of the Fittest.
- Democratization is good and not-so-good:
  - In Academia, people are distorted from the really important and harder questions.
  - In Industry, applying deep learning to a dataset is not going to make you *unique* anymore.
- Statistical significance: **Quantity overwhelms quality**!

The Bad and the Ugly

- Images are challenging: noise, blur, perturbations, deformations, anomalies, etc.
- Challenges in optimization.
- Grid search or the Survival of the Fittest.
- Democratization is good and not-so-good:
  - In Academia, people are distorted from the really important and harder questions.
  - In Industry, applying deep learning to a dataset is not going to make you *unique* anymore.
- Statistical significance: Quantity overwhelms quality!
- **Fast pace of the field**.

The Bad and the Ugly

- Images are challenging: noise, blur, perturbations, deformations, anomalies, etc.
- Challenges in optimization.
- Grid search or the Survival of the Fittest.
- Democratization is good and not-so-good:
  - In Academia, people are distorted from the really important and harder questions.
  - In Industry, applying deep learning to a dataset is not going to make you *unique* anymore.
- Reproducible research and statistical significance: Quantity overwhelms quality!
- Fast pace of the field.

Deep learning is neither a philosophy nor an application, it is a tool that makes many applications smarter and more natural through experience.

TAMI Research Actions

The road so far…

Industrial panel

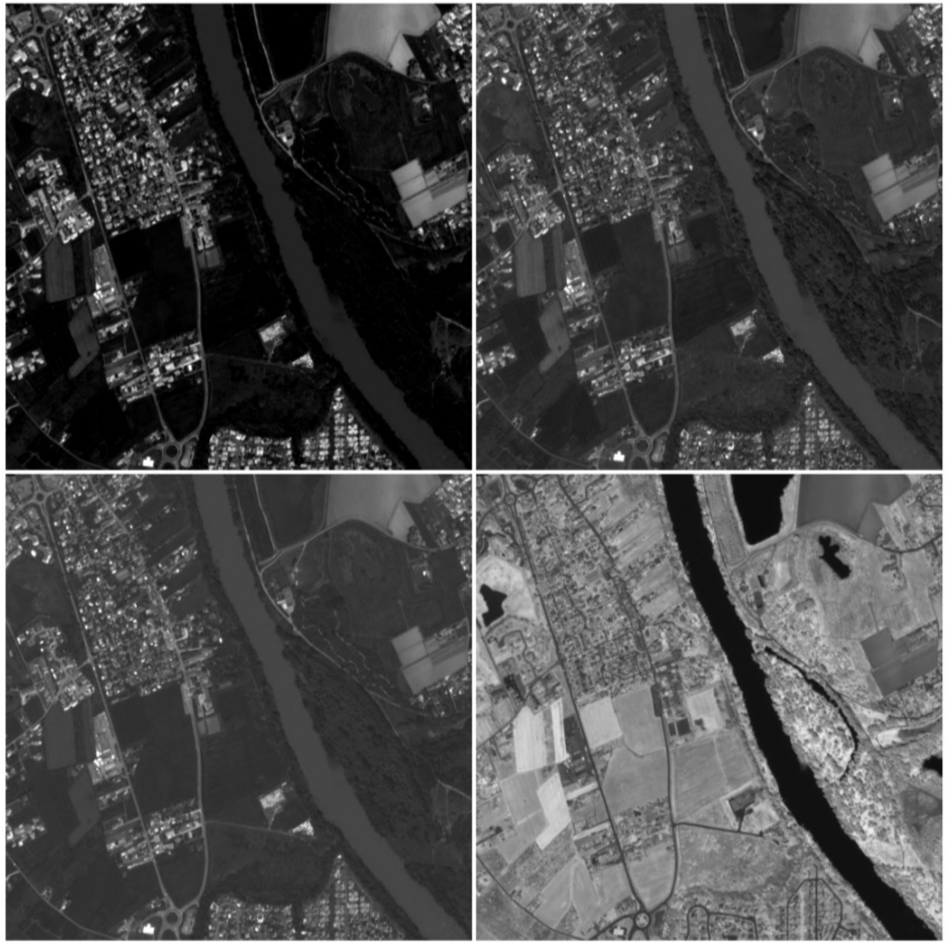
cnes

CENTRE NATIONAL D'ÉTUDES SPATIALES

Industrial panel
Enhancement

Industrial panel
Enhancement

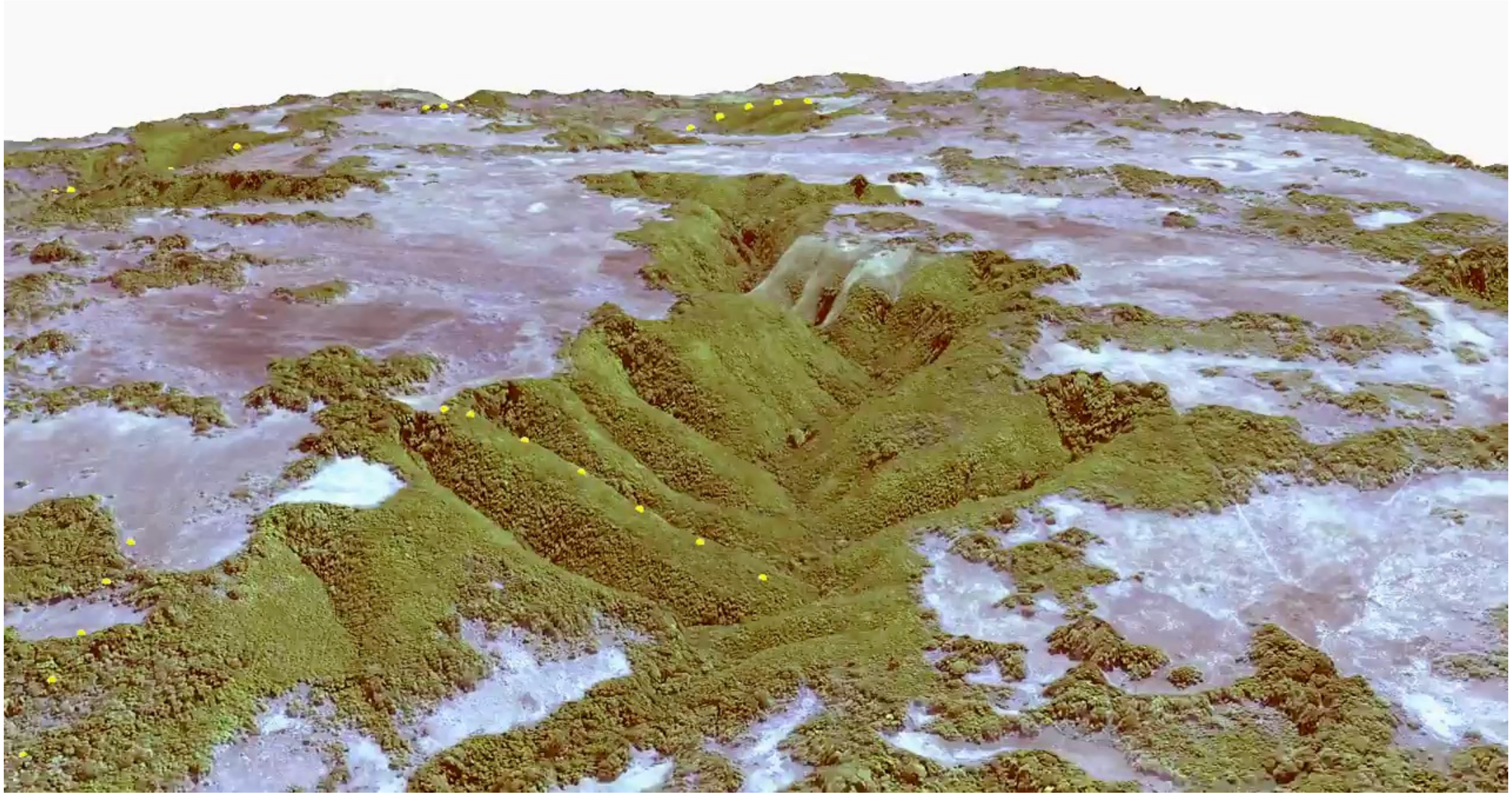
Industrial panel
Data fusion



Industrial panel
Data fusion

Industrial panel

3D Reconstruction from satellite imagery

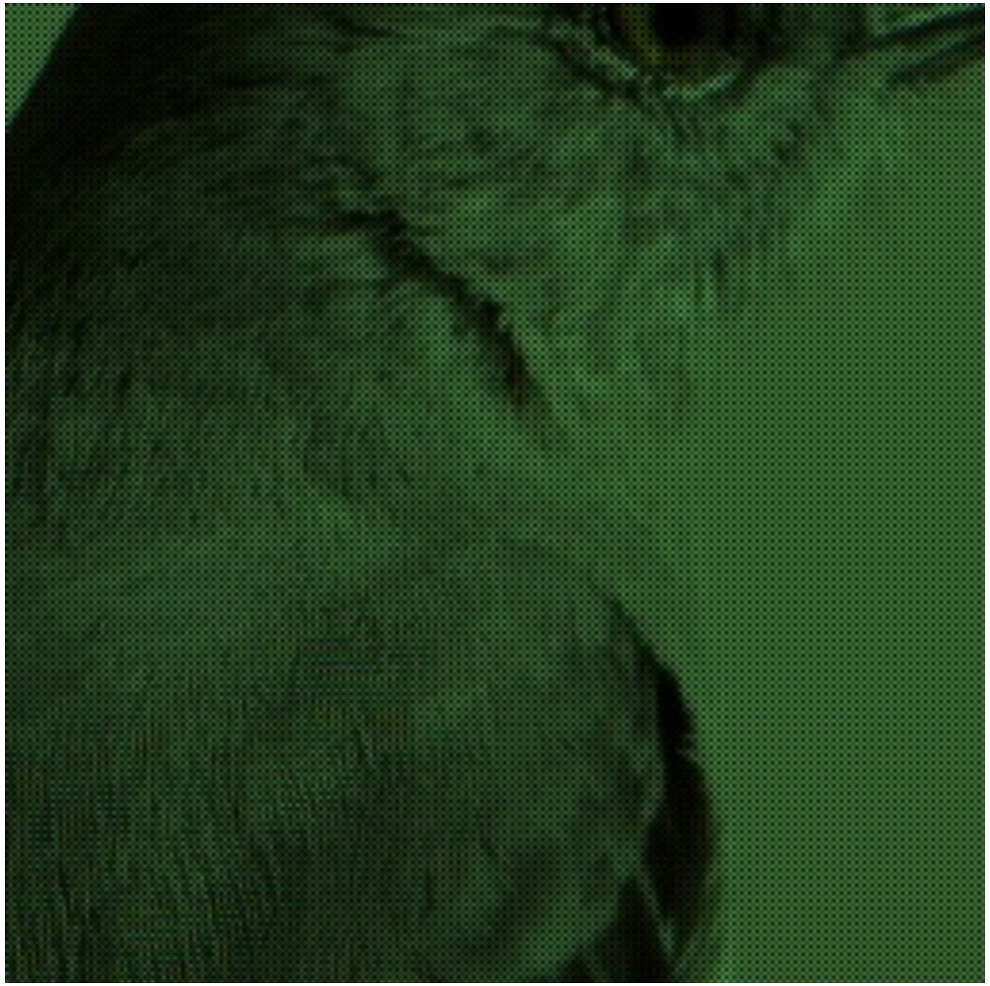


Industrial panel

3D Reconstruction from satellite imagery

Industrial panel

Industrial panel
Demosaicking



Industrial panel
Demosaicking

Industrial panel
Denoising

Industrial panel
Denoising

Industrial panel
Denoising

Cognitech
The Best Before & After

Industrial panel
Multi-exposure image fusion

Cognitech
The Best Before & After

Industrial panel
Multi-exposure image fusion

Industrial panel

Cognitech

The Best Before & After

Industrial panel
Video Enhancement

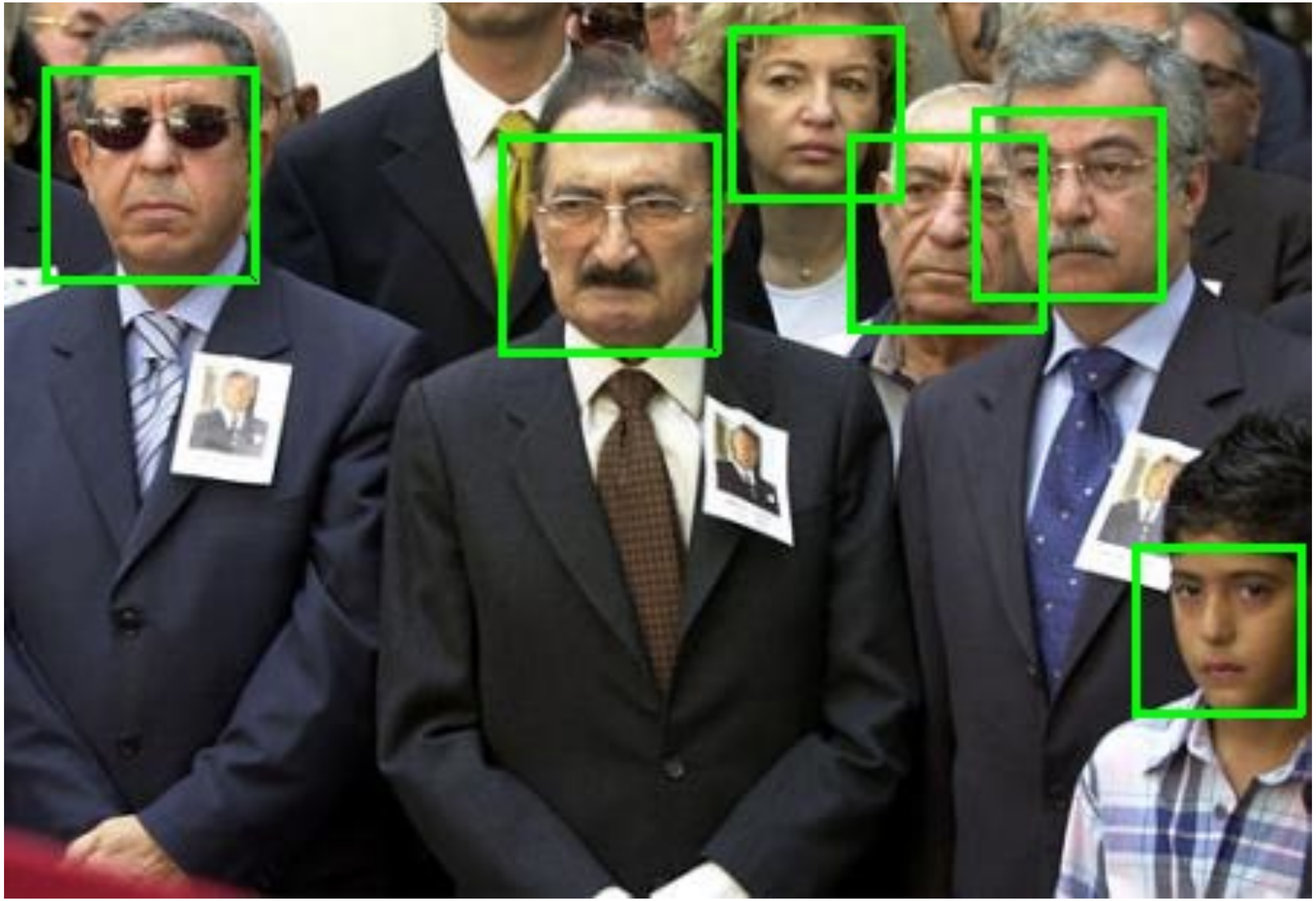
Cognitech
The Best Before & After

Industrial panel
Video Enhancement

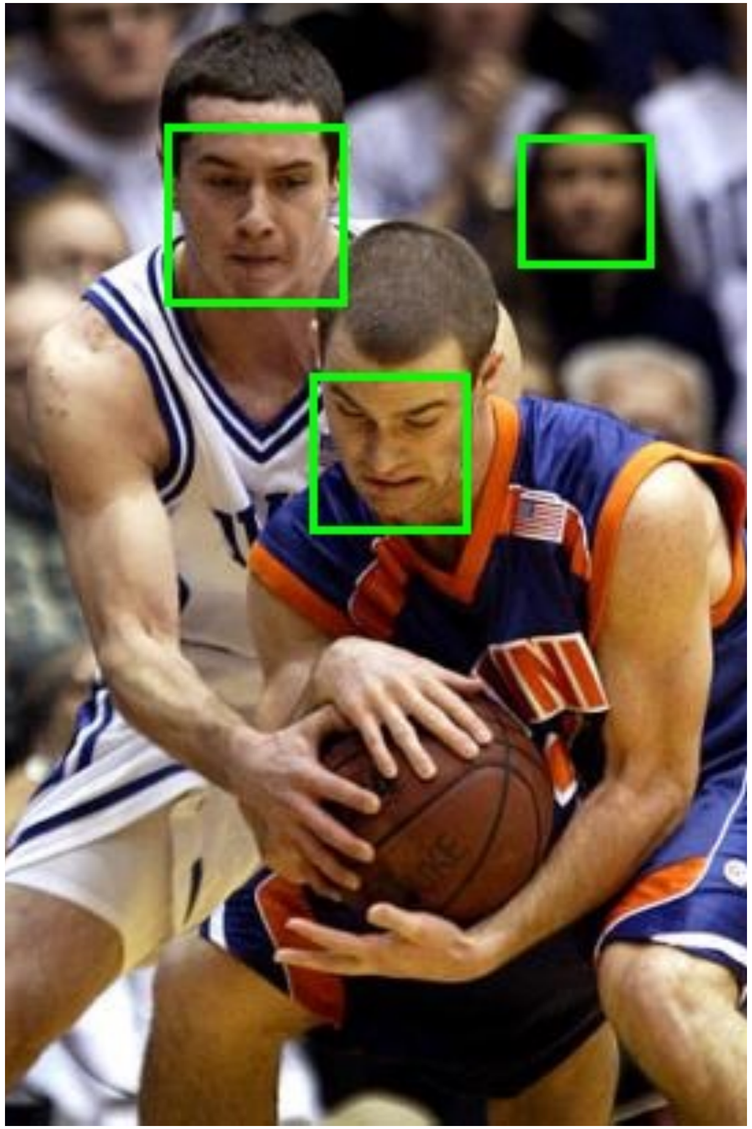
Modeling and Deep Learning

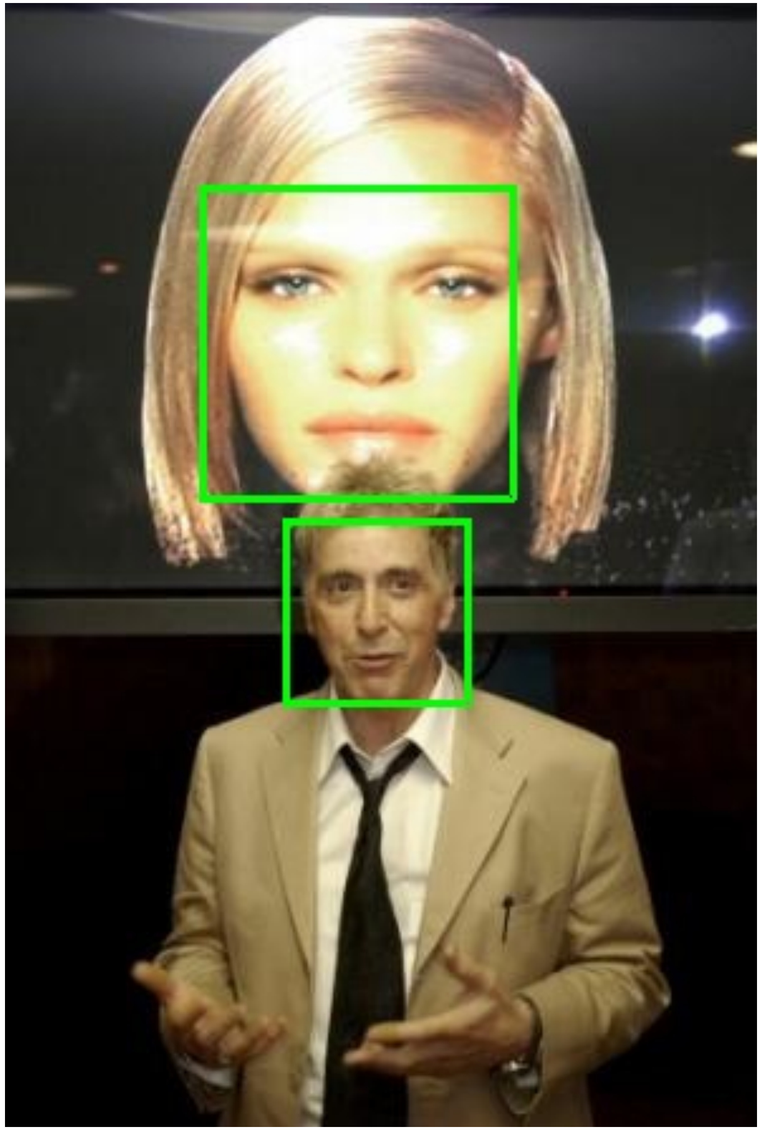
Actions and future projects

Learning based actions
Face detection



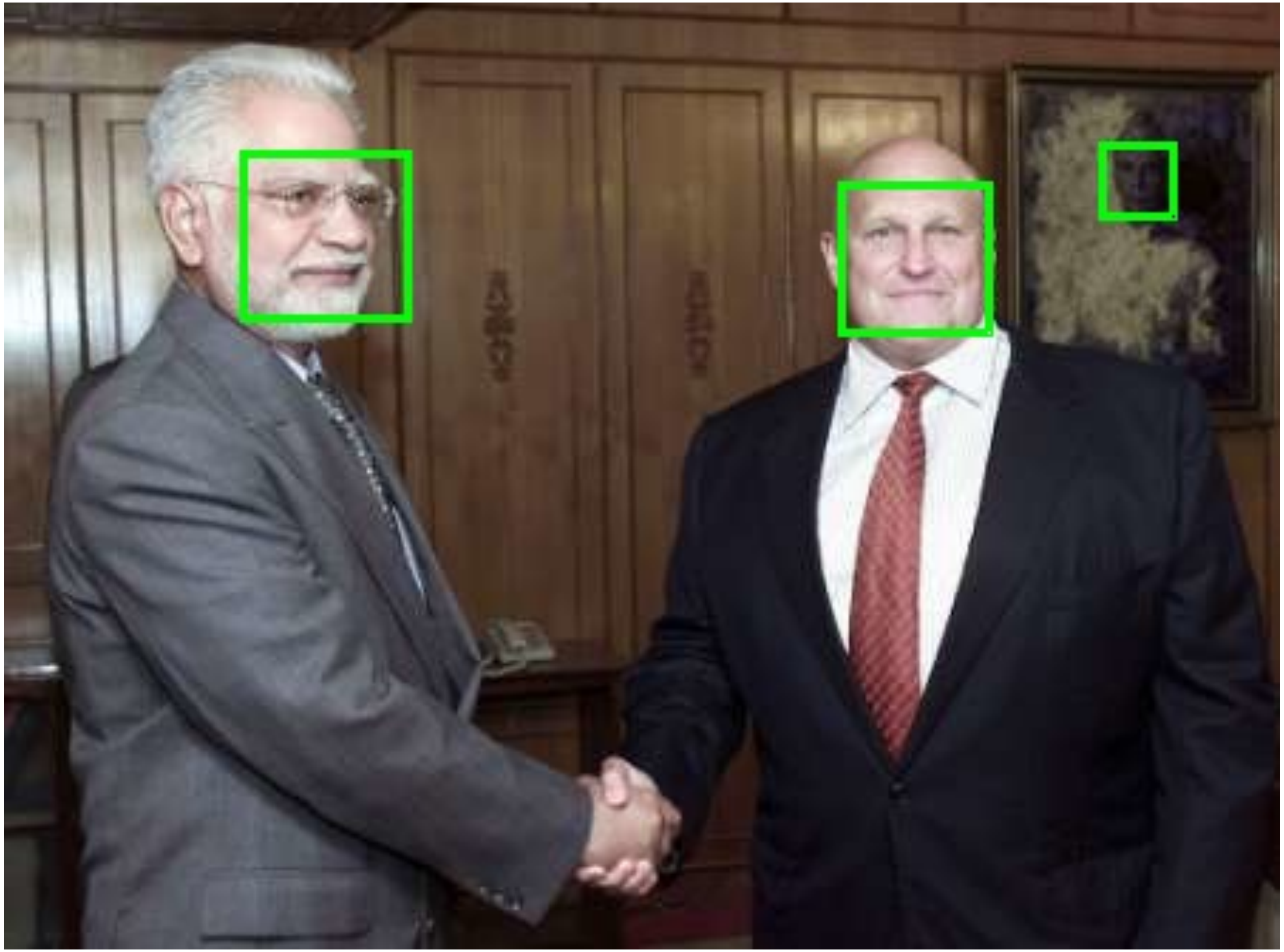
Learning based actions
Face detection





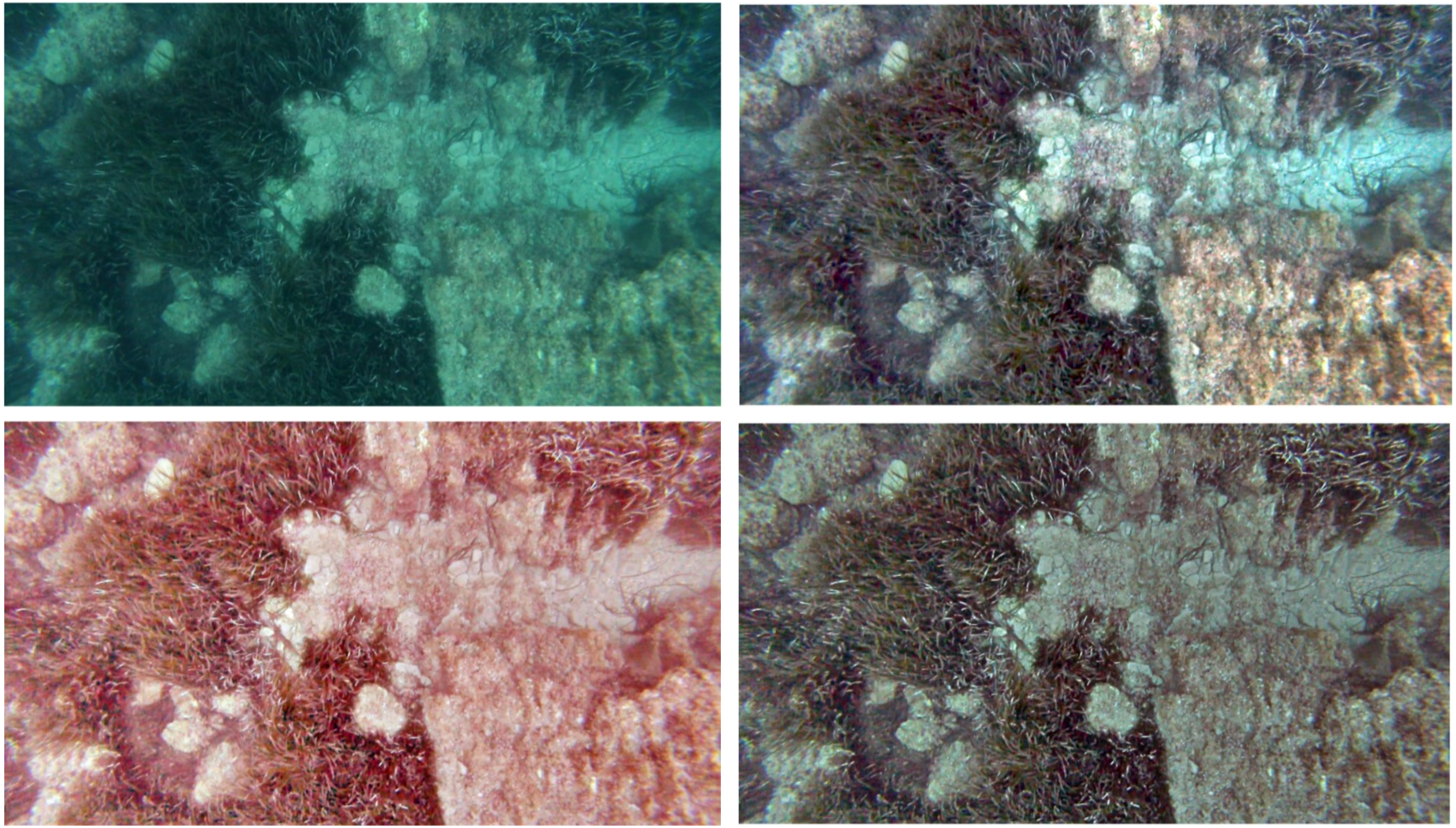
Learning based actions

Face detection



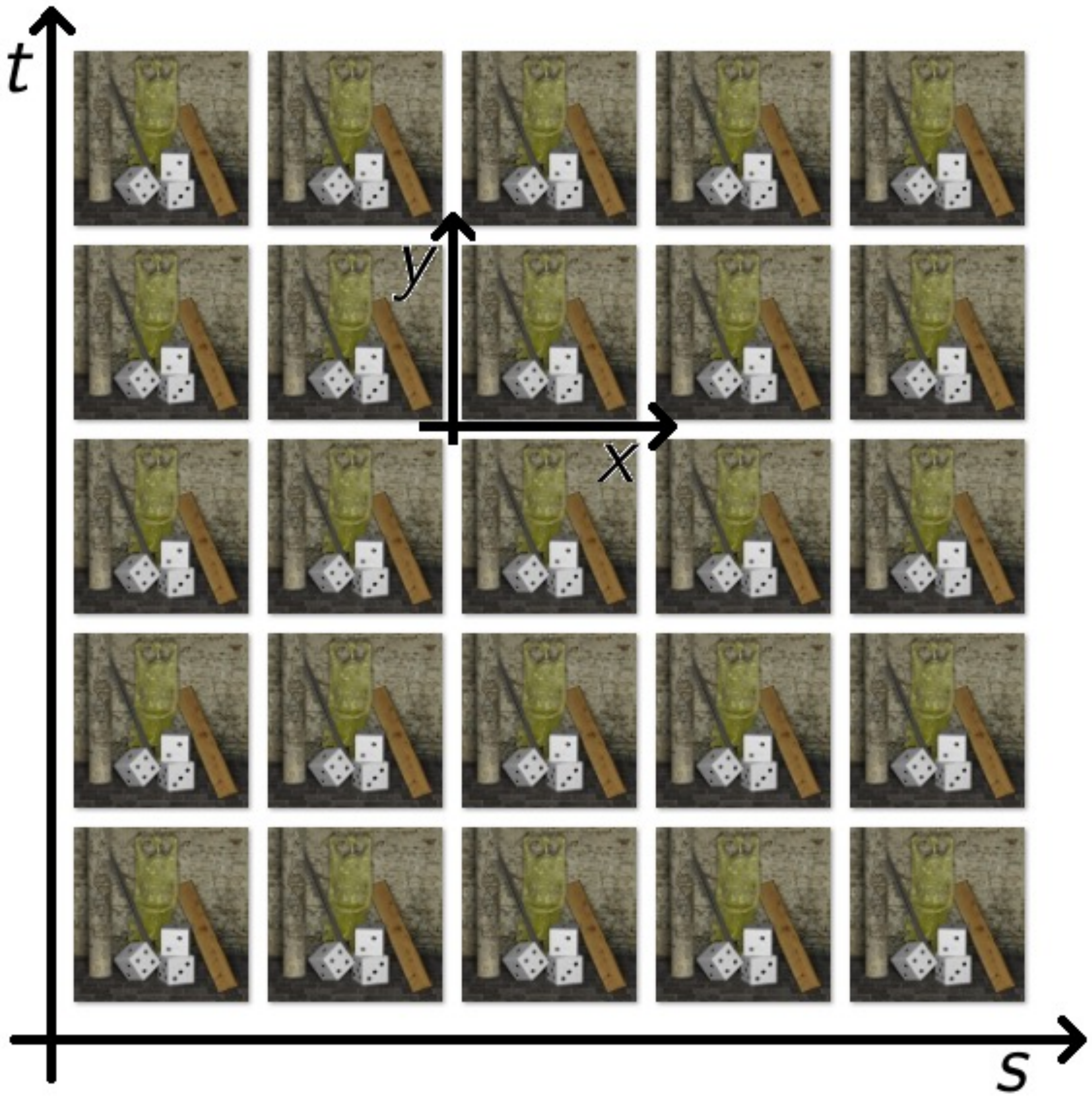
Learning based actions

Sea grass detection in underwater areas



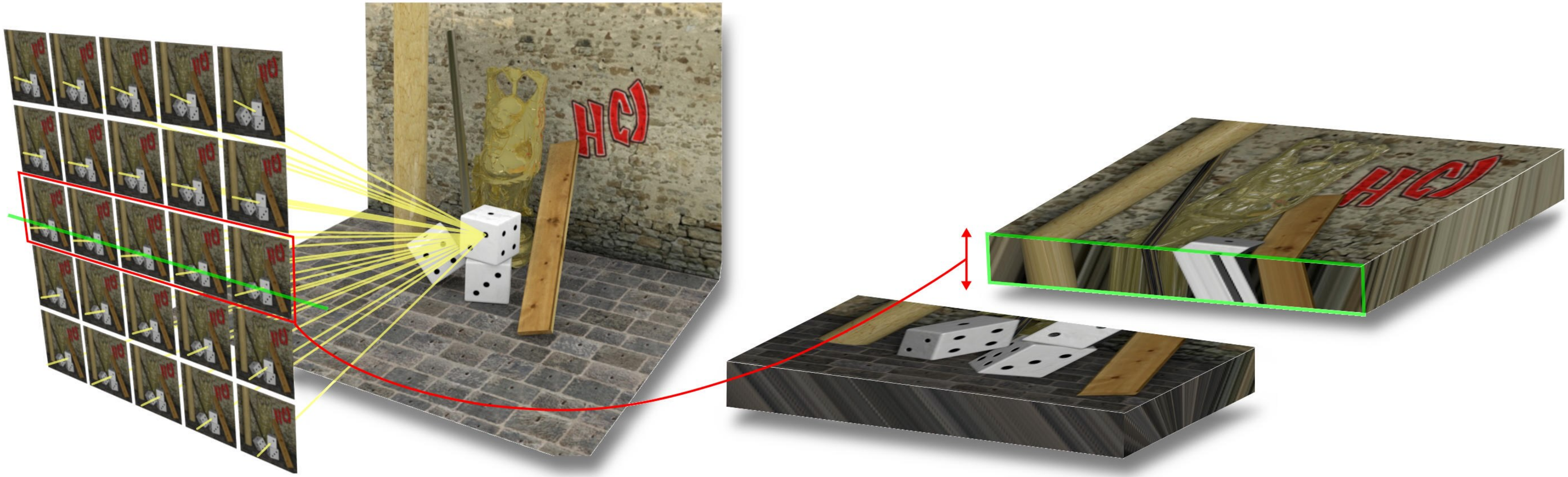
Learning based actions

Light-field images



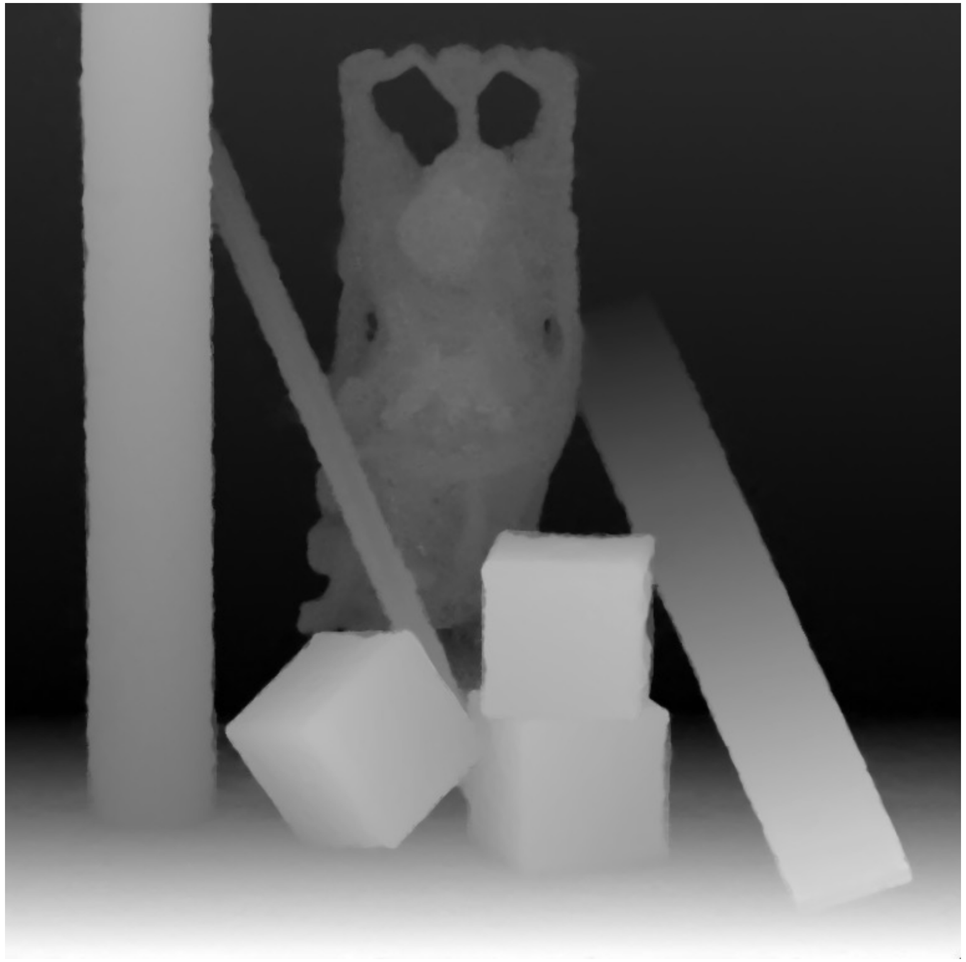
Learning based actions

Light-field images



Learning based actions

Light-field images



Learning based actions
Light-field images

technicolor

- Deep learning for depth estimation in light-field capturing devices and multi-camera datasets.
- Deep learning for view synthesis.
- 3D scene reconstruction using depth and image views.

Learning based actions
Geometric deep learning

- Deep learning techniques applied to non grid-like structure:
  - Graphs in social networks and particle physics.
  - Manifolds in computer graphics.
- Define deep neural models for this kind of non-Euclidean data.
- Theoretical inside and applications.

Some Proposals

What makes Paris look like Paris?

SIGGRAPH 2012

What Makes Paris Look like Paris?

Carl Doersch[1], Saurabh Singh[1], Abhinav Gupta[1],
Josef Sivic[2] and Alexei A. Efros[1,2]

[1] Carnegie Mellon University [2] INRIA / École Normale Supérieure, Paris

Some Proposals

- Hotel recognition, category classification, hotel location.
- Image restoration: contrast enhancement, noise and blur removal, anomaly detection, super-resolution.
- 3D reconstruction of rooms given depth and scene views.
- Detect and discard images not providing relevant information to users.
- Identify repeated images and discard them.
- Classify and detect facilities from hotel images.
- Identify priorities from user interactions and select displayed hotel images accordingly.

Deep Learning
Last trends and applications

Joan Duran Grimalt

Tractament i Anàlisi Matemàtica d'Imatges (TAMI)

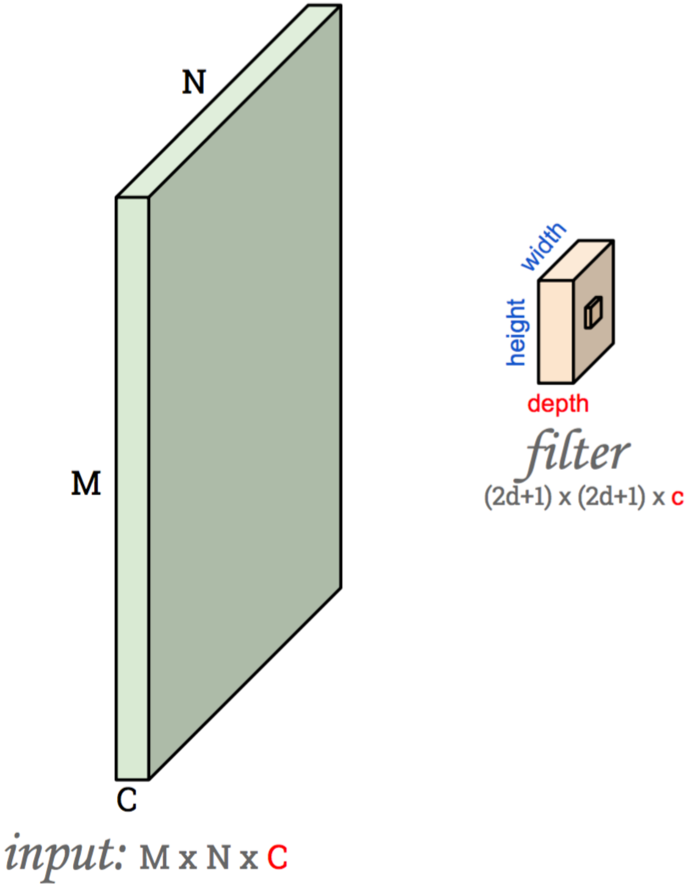
Departament de Ciències Matemàtiques i Informàtica

joan.duran@uib.es joandg @joan_dg

Universitat de les Illes Balears

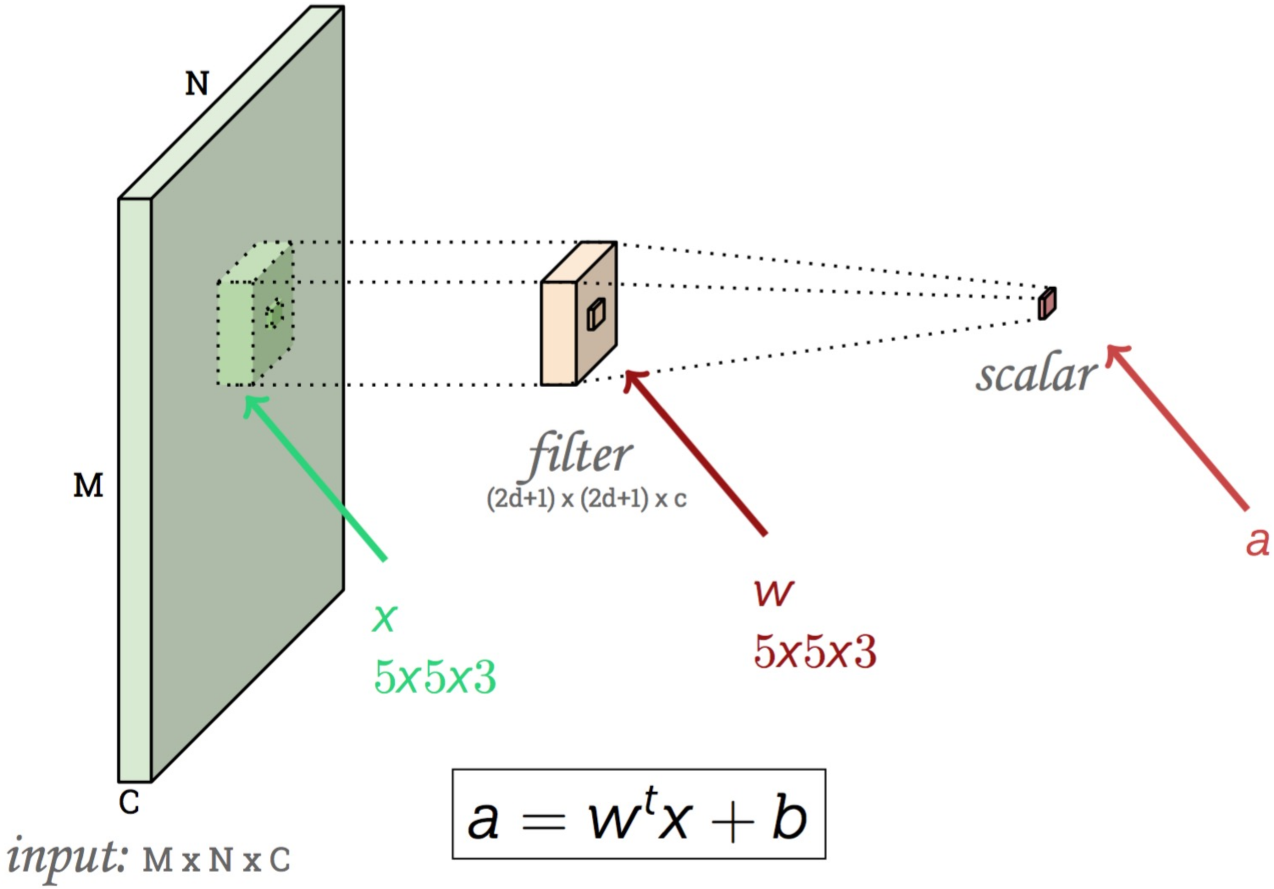
Convolutional Neural Networks

- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.



Convolutional Neural Networks

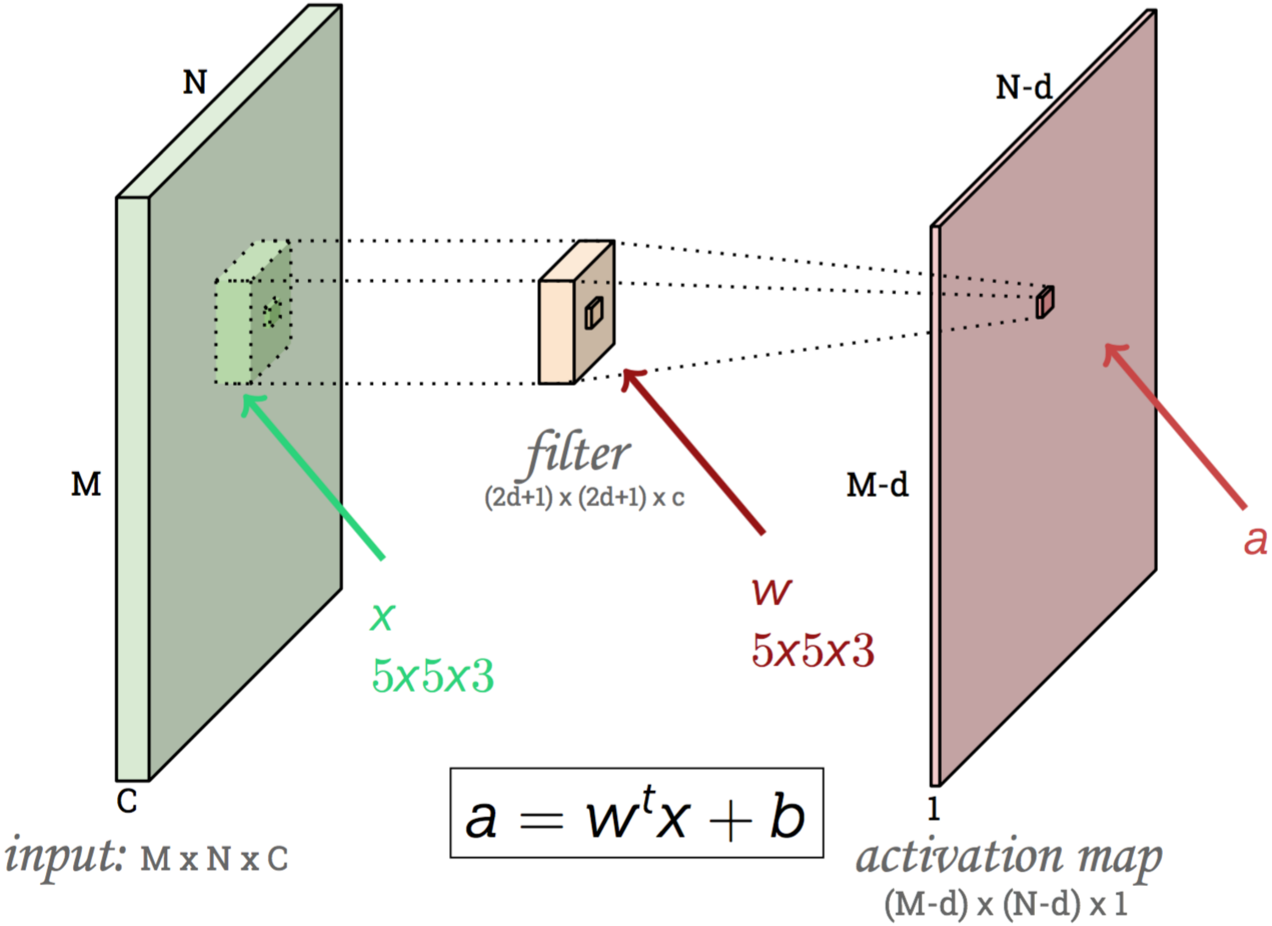
- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
    1. Perform several convolutions in parallel to produce a set of linear activations.



$$a = w^t x + b$$

Convolutional Neural Networks

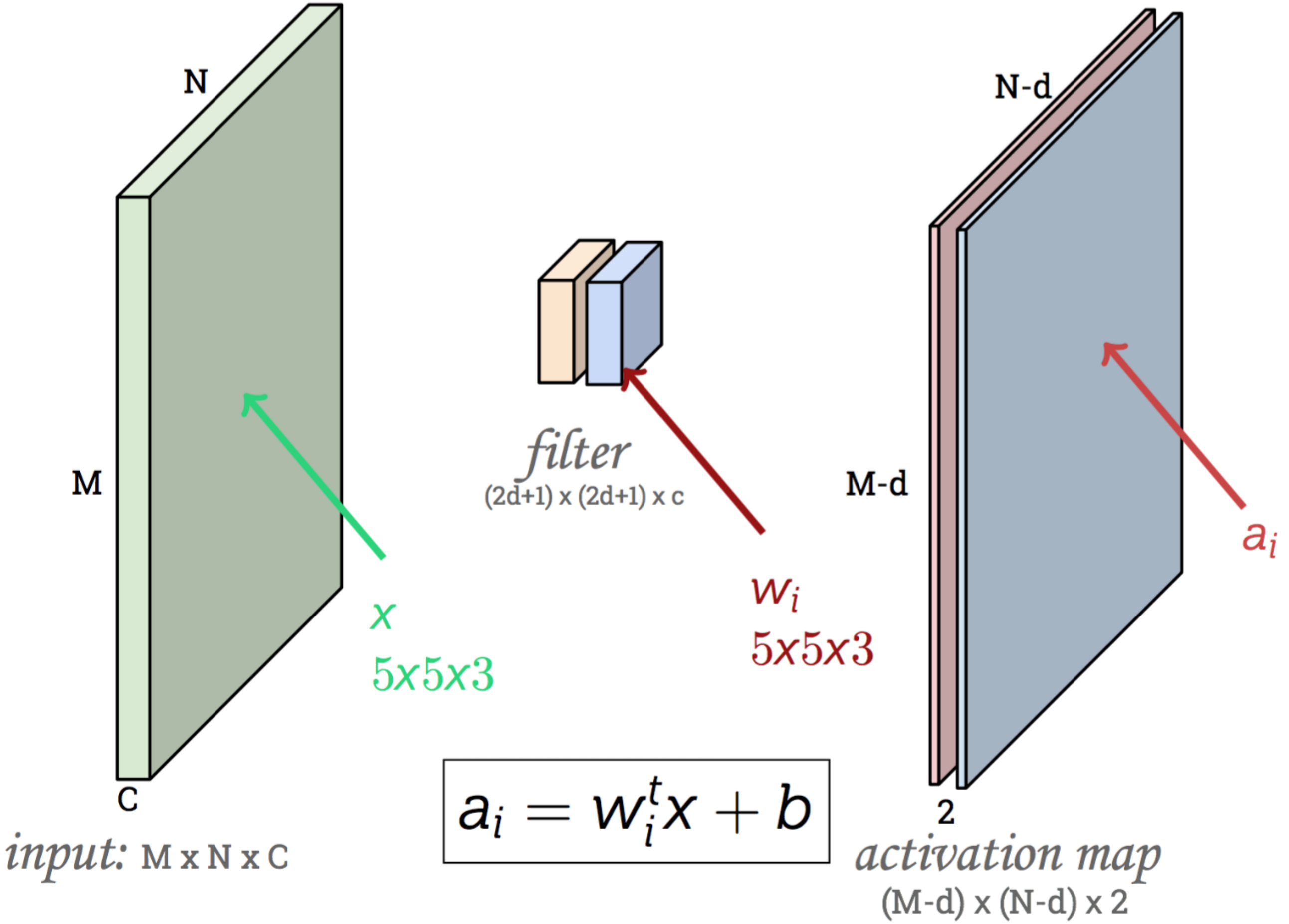
- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.



$$a = w^t x + b$$

Convolutional Neural Networks

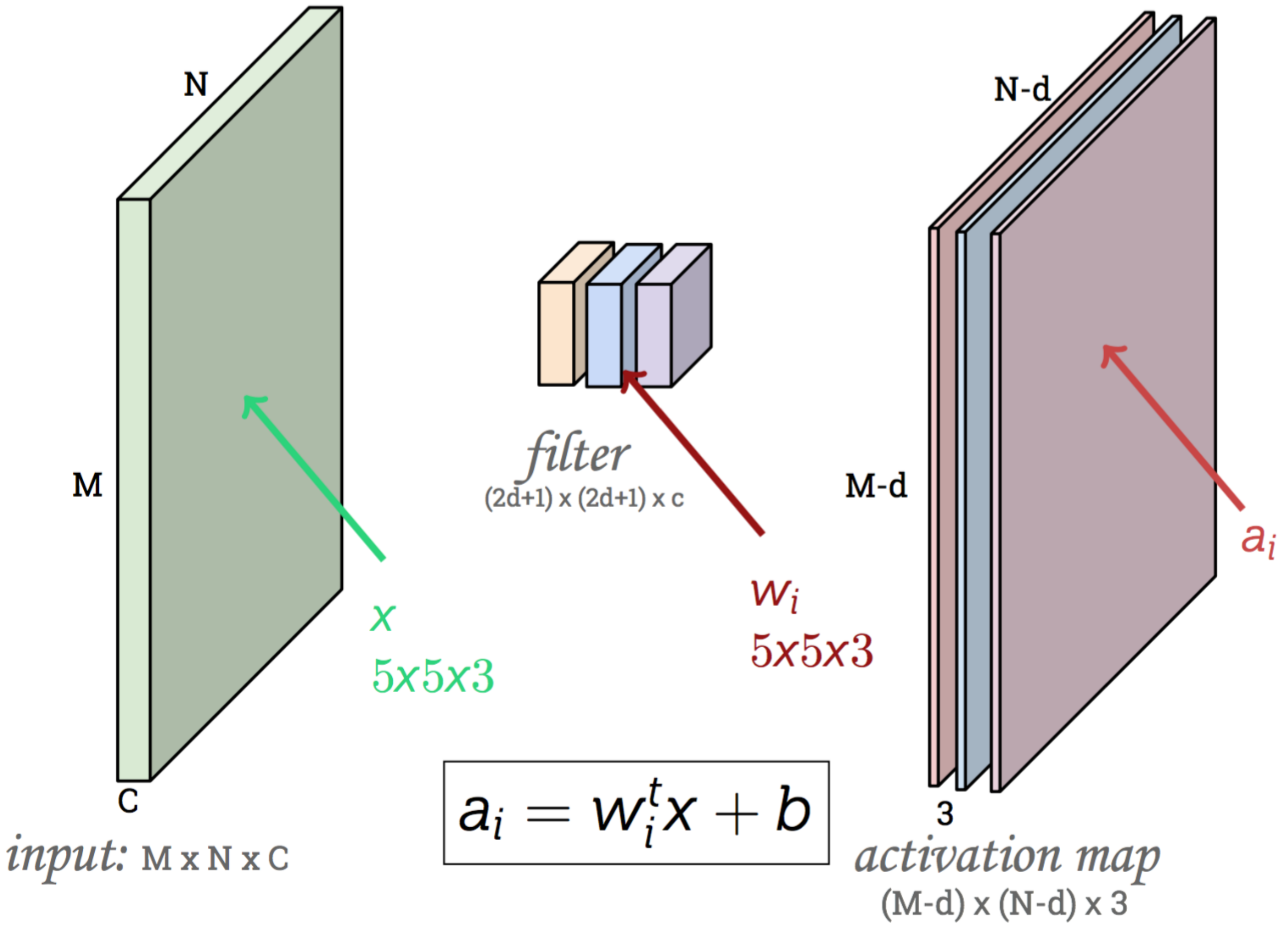
- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.



$$a_i = w_i^t x + b$$

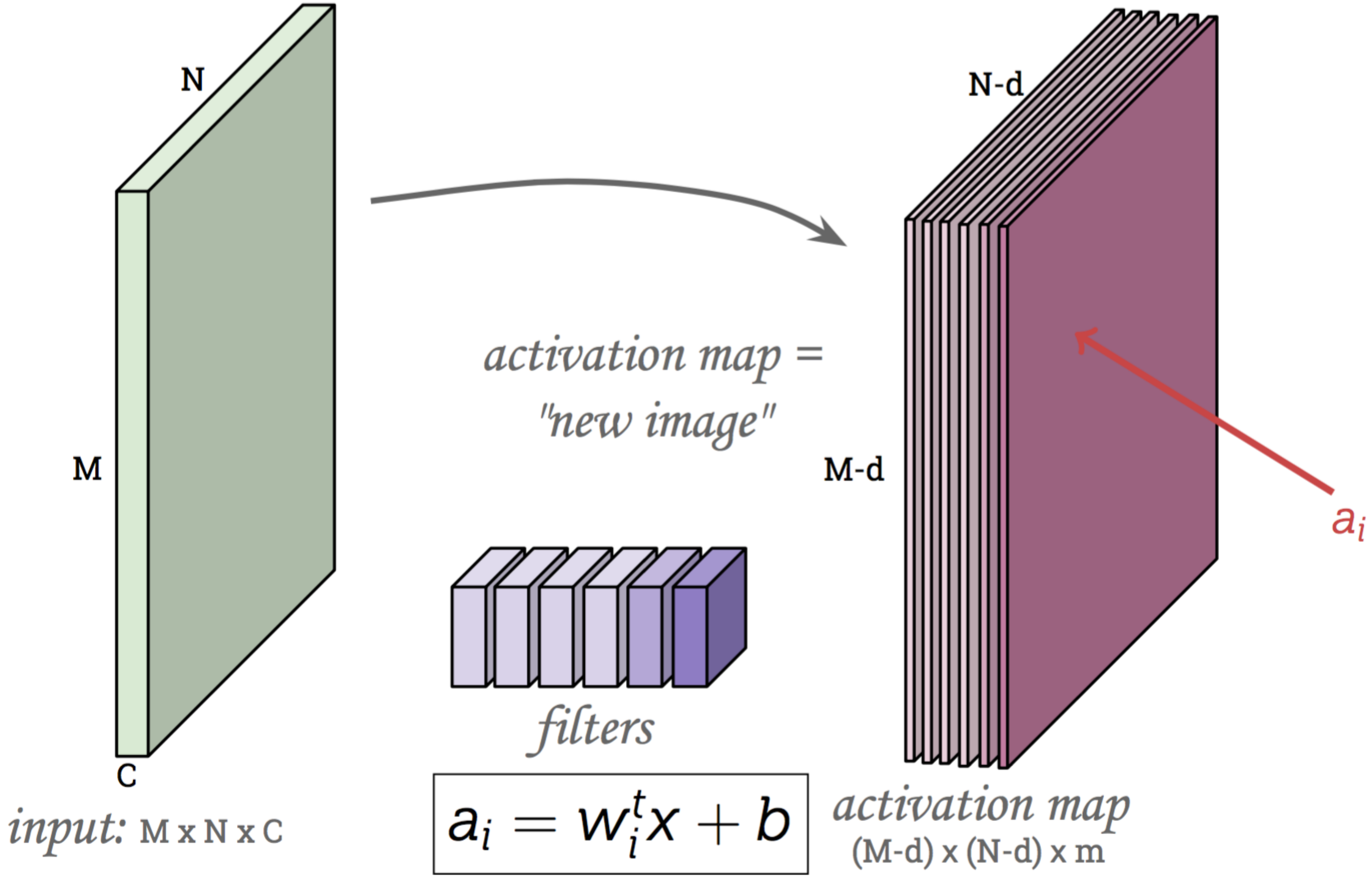
Convolutional Neural Networks

- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.



Convolutional Neural Networks

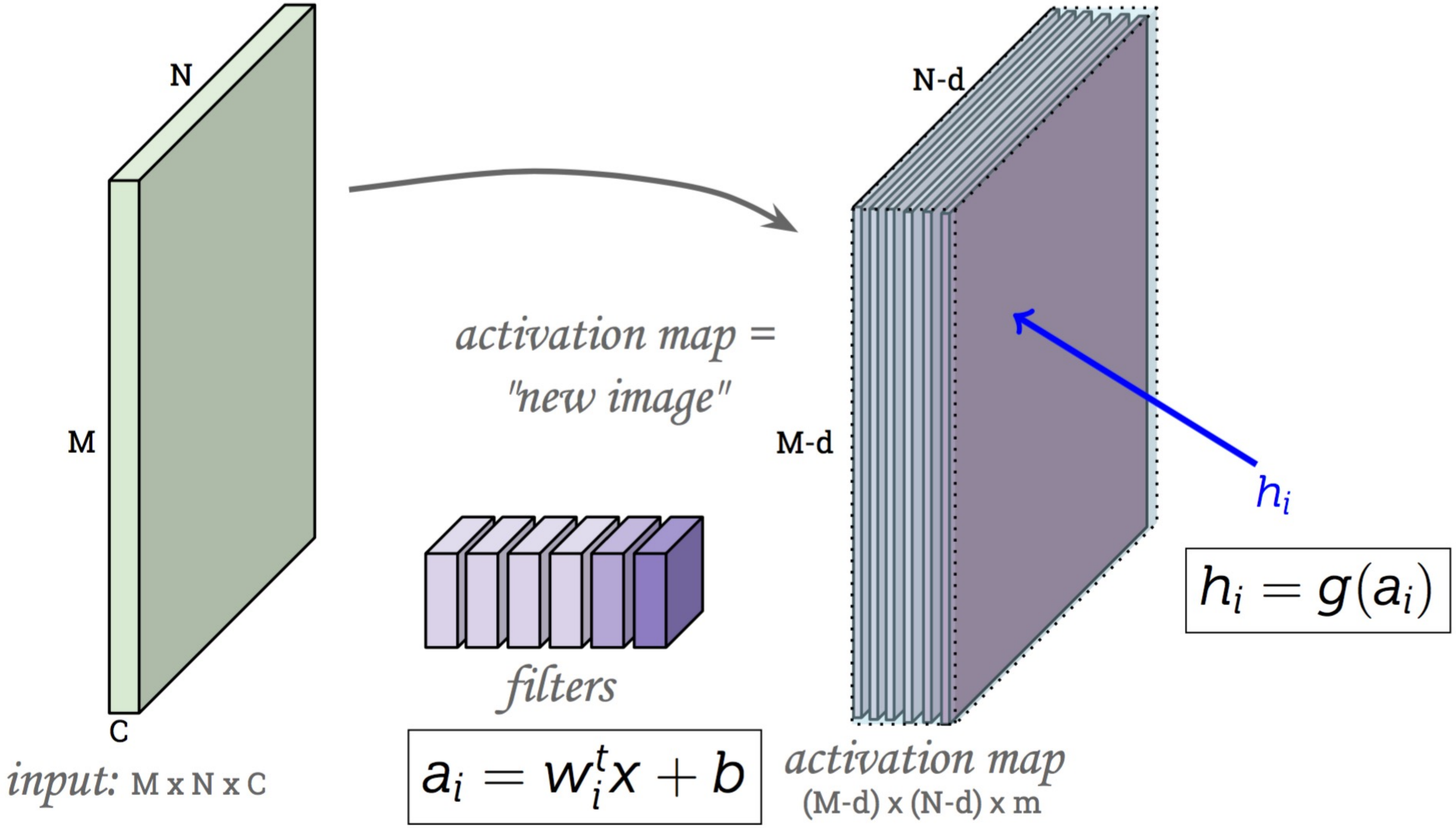
- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.



$$\boxed{a_i = w_i^t x + b}$$

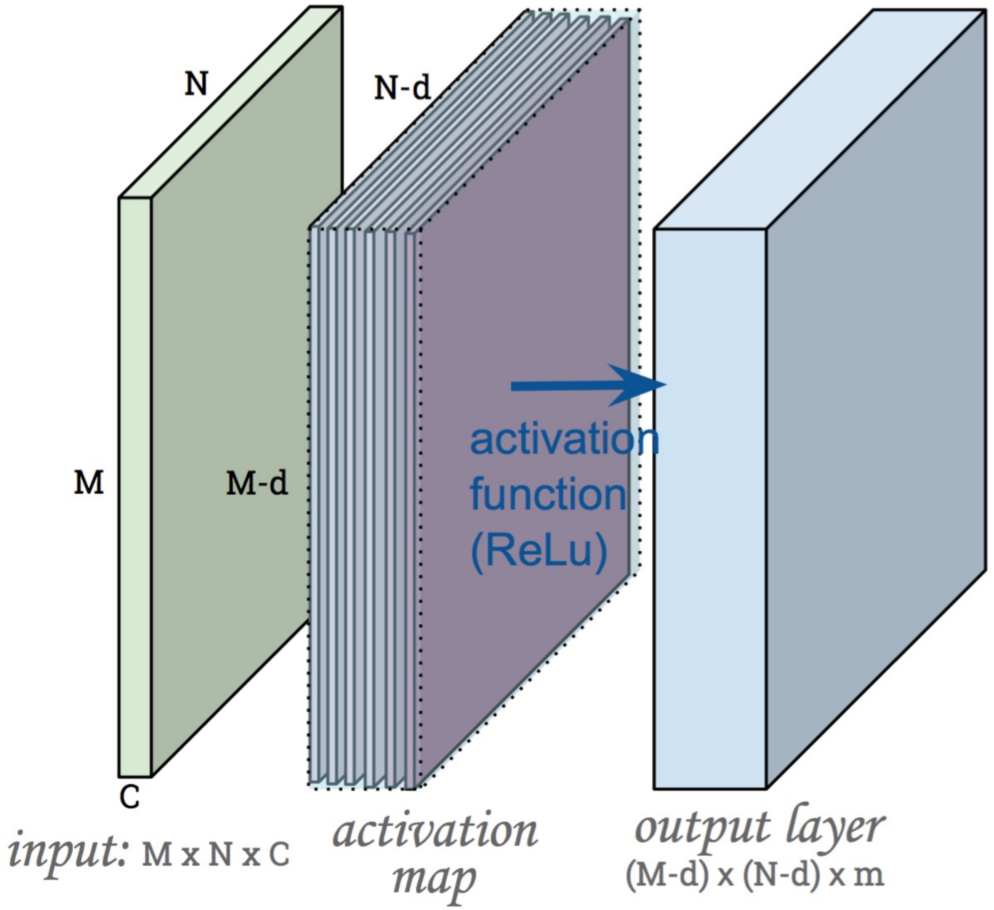
Convolutional Neural Networks

- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.



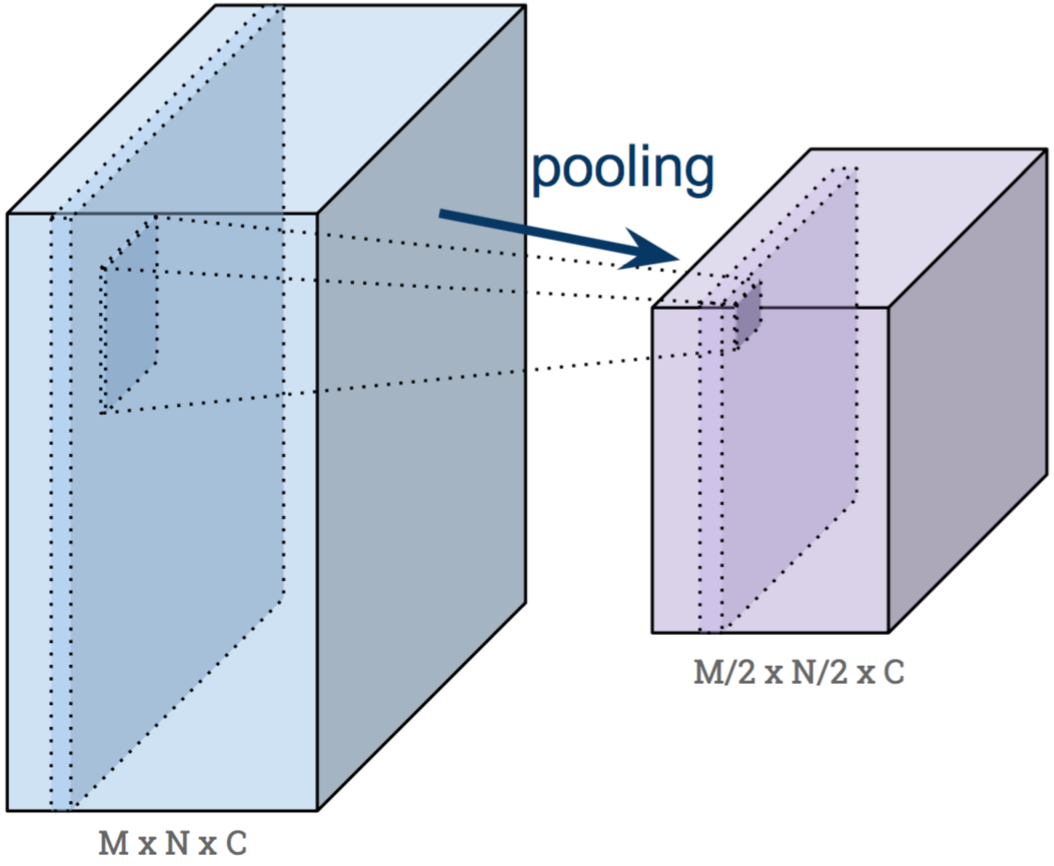
Convolutional Neural Networks

- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  2. Apply a nonlinear activation function g (sigmoid, tanh, ReLU) to each linear activation map a_i.



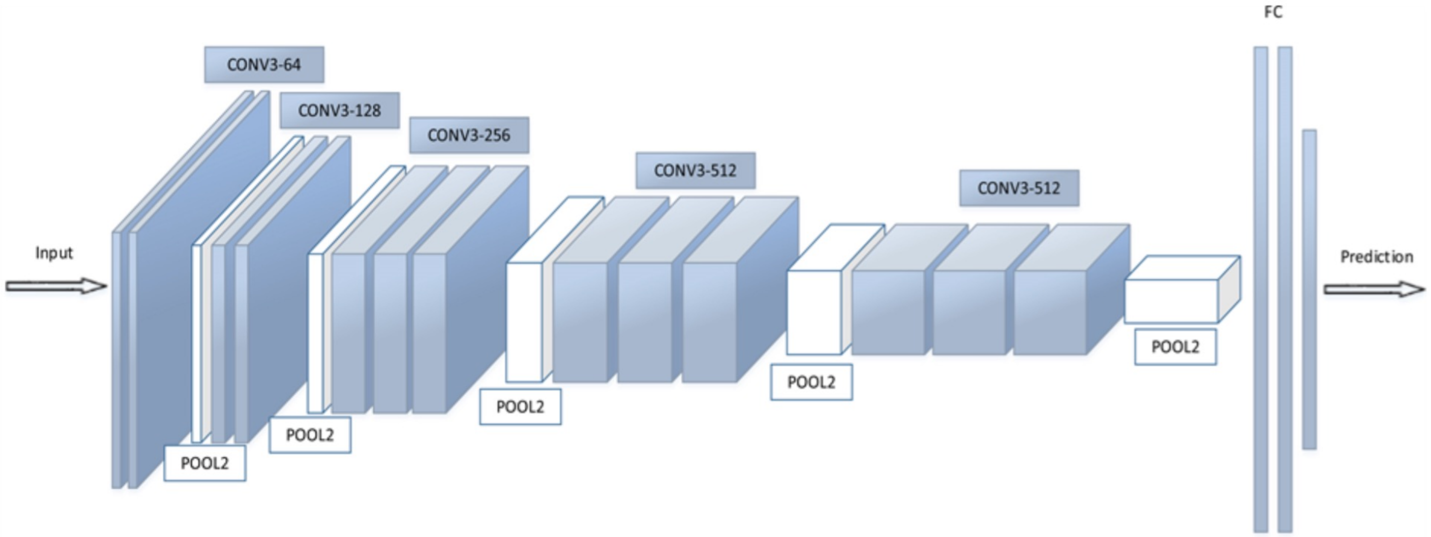
Convolutional Neural Networks

- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  3. Apply a pooling function to each activation map (subsampling), simplifying the representation.



Convolutional Neural Networks

- CNNs use **convolution operators** to process data with known grid-like topology and possibly of variable size, such as images.
- Stages of a CNNs layer:
  1. Perform several convolutions in parallel to produce a set of linear activations.
  2. Apply a nonlinear activation function to each linear activation map.
  3. Apply a pooling function to each activation map (subsampling), simplifying the representation.
- CNNs leverages **sparse connectivity**, **parameter sharing** and **equivariant representations**.
  - Image statistics are invariant to translations.
  - Low-level features (edges) are local, so local connectivity is imposed through kernel support.
  - High-level features are expected to be thick, so pooling is used as long as CNNs depth increases.



Applications in Computer Vision

Image Enhancement

$$u(x) = f\left(\frac{I(x)}{I * \kappa(x)}\right)$$

Applications in Computer Vision
Image Enhancement

$$u(x) = f\left(\frac{I(x)}{I * \kappa(x)}\right)$$

Applications in Computer Vision

Image Enhancement

Applications in Computer Vision

Image Enhancement