Foundations of Image Classification and Convolutional Neural Networks

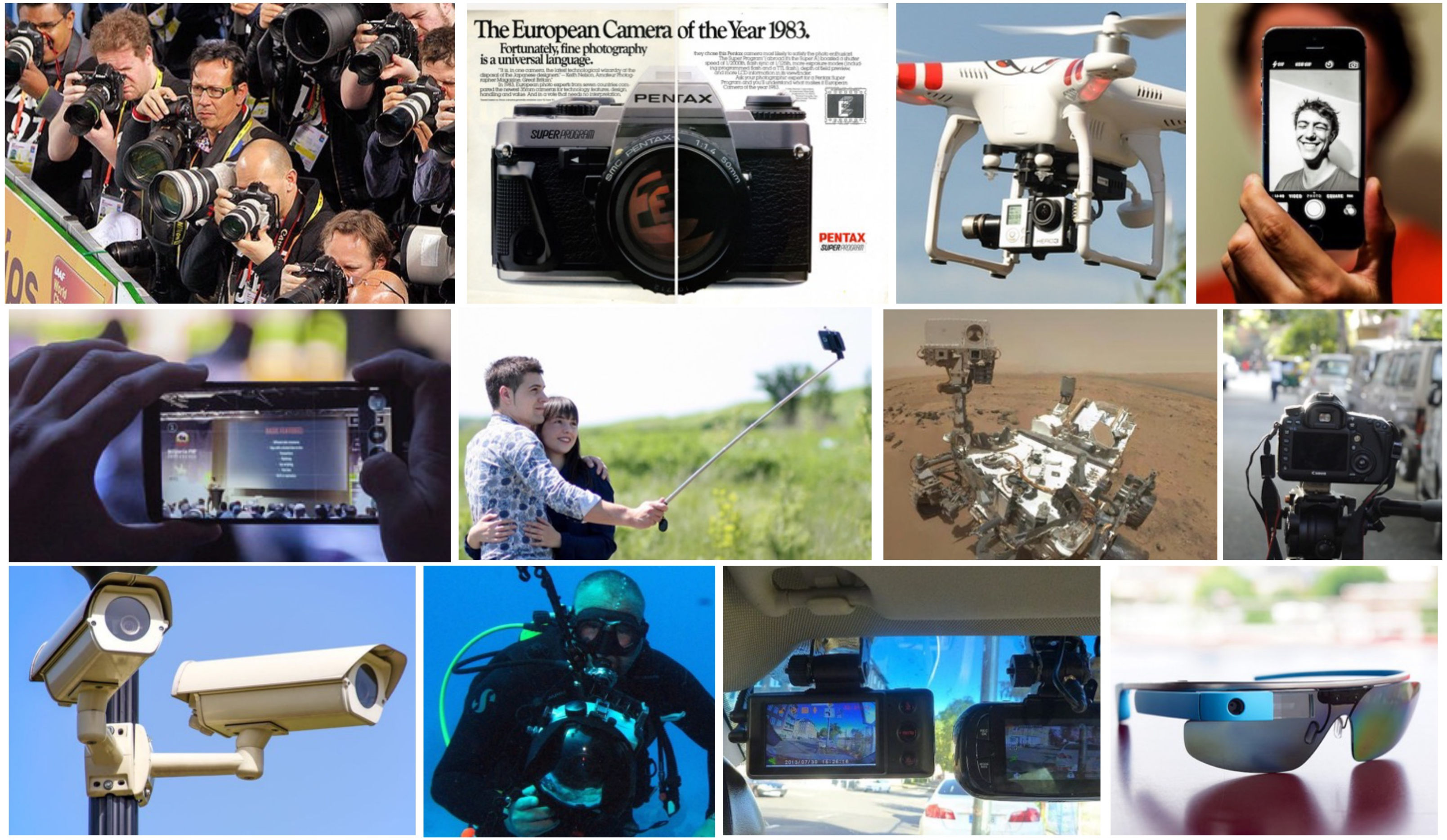
Saeid Nahavandi
Distinguished Professor
snahavandi@swin.edu.au

Parham Kebria
Senior Research Scientist
parhamkebria@ieee.org

Advanced Study Institute: Artificial Intelligence for Disaster Management
Orlando, November 17 – 25, 2025

This activity is supported by: The NATO Science for Peace and Security Programme

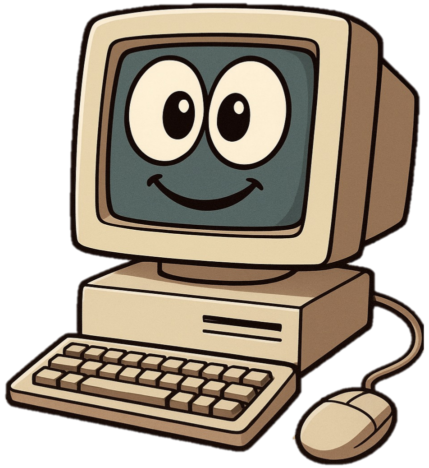
Why Image Processing/Computer Vision?



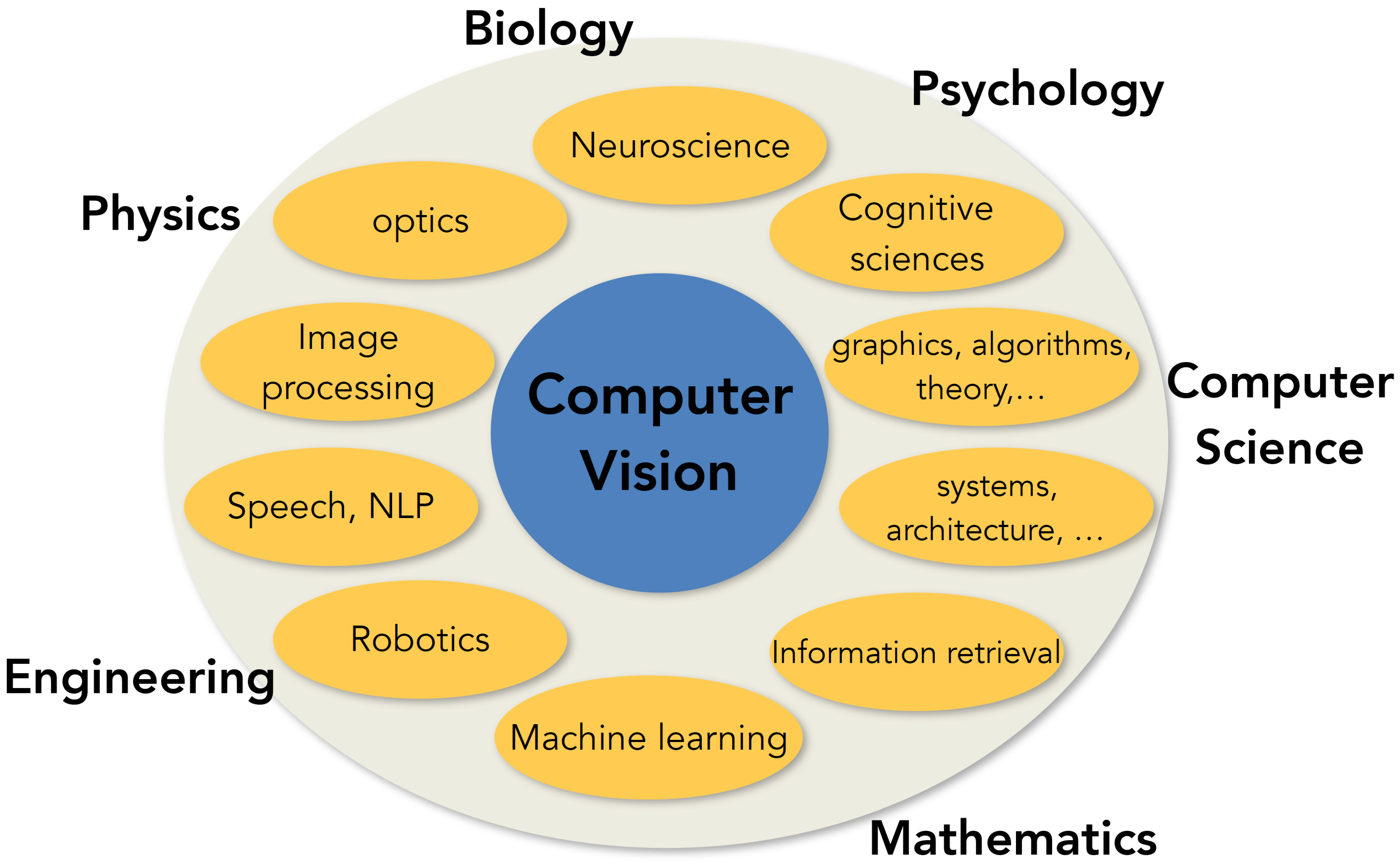
Top row, left to right:
Image by Augustas Didžgalvis; licensed under CC BY-SA 3.0; changes made
Image by Nesster is licensed under CC BY-SA 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain

Middle row, left to right
Image by BGPHP Conference is licensed under CC BY 2.0; changes made
Image is CC0 1.0 public domain
Image by NASA is licensed under CC BY 2.0; chang
Image is CC0 1.0 public domain

Bottom row, left to right
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image by Ted Eytan is licensed under CC BY-SA 2.0; changes made



Is it important?



Evolution's Big Bang

This image is licensed under CC-BY 2.5

This image is licensed under CC-BY 2.5

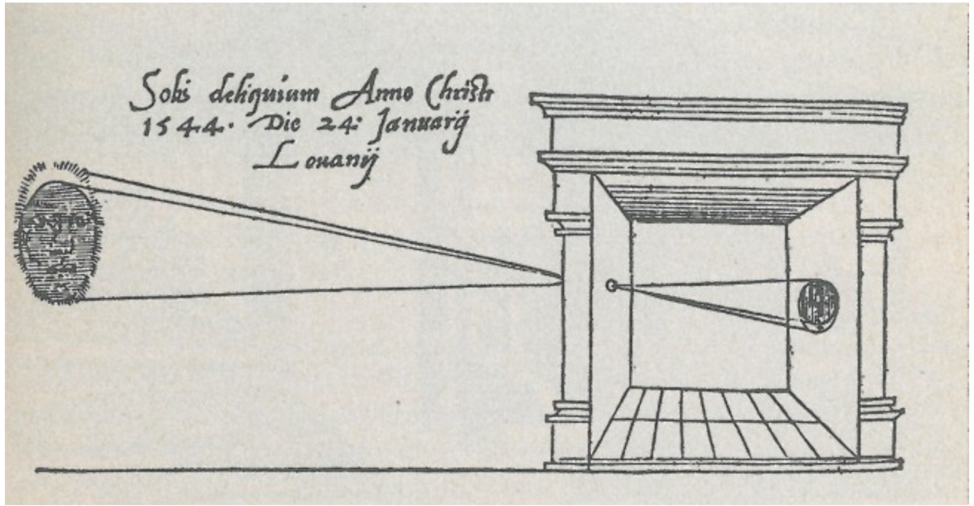
This image is licensed under CC-BY 3.0

543million years, B.C.

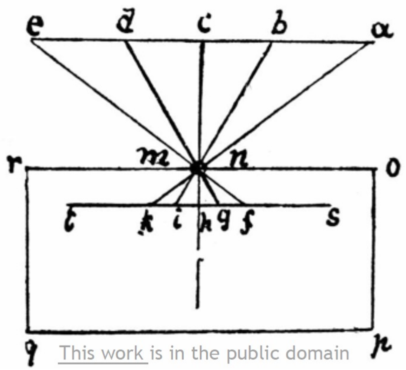
Evolution’s Big Bang

543million years, B.C.

Gemma Frisius, 1545



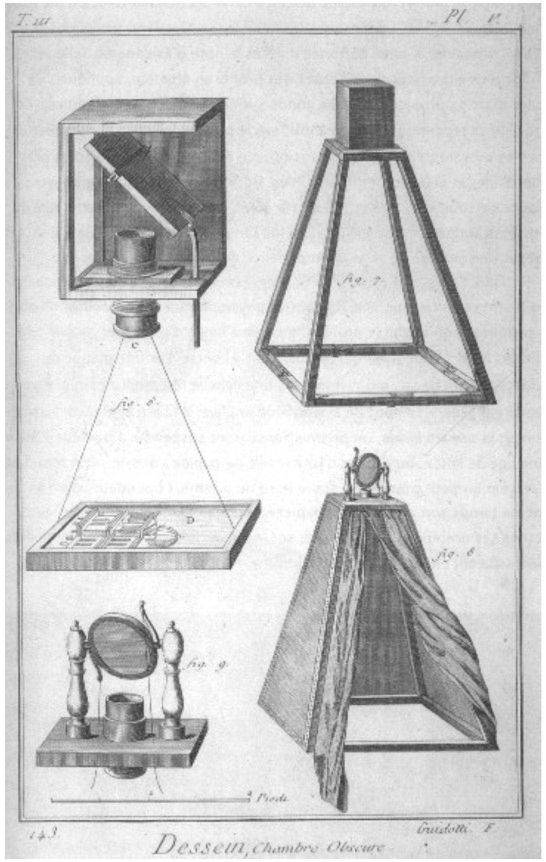
This work is in the public domain



This work is in the public domain

Leonardo da Vinci, 16th Century AD

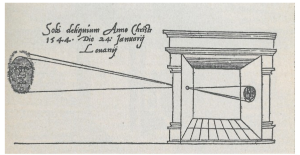
Encyclopedie, 18th Century



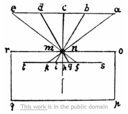
This work is in the public domain

2025

Gemma Frisius, 1545



This work is in the public domain

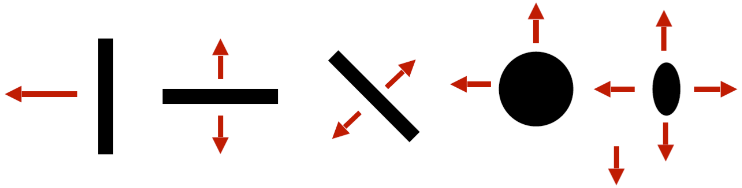


Leonardo da Vinci, 16th Century AD

This work is in the public domain

Evolution's Big Bang

543million years, B.C.

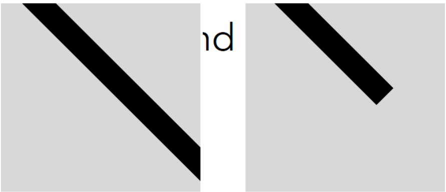


Hubel & Wiesel, 1959

Simple cells:
Response to light orientation

Complex cells:
Response to light orientation and movement

Hypercomplex cells:
r...nse to mc...nt with
...nd

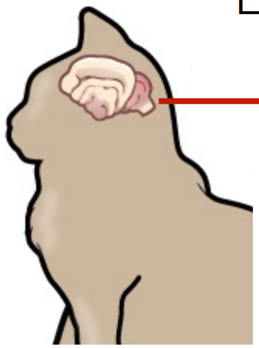


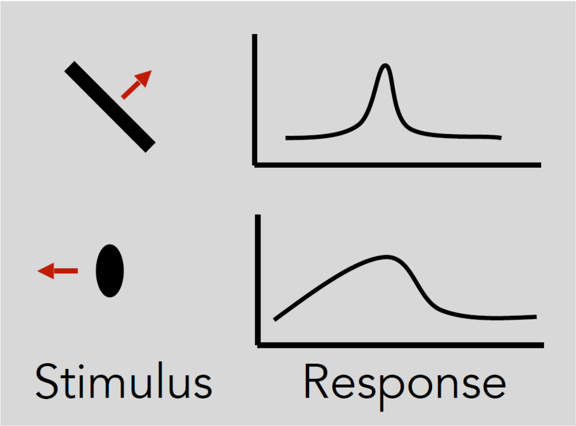
No response

Response (end point)

Stimulus

Electrical signal from brain



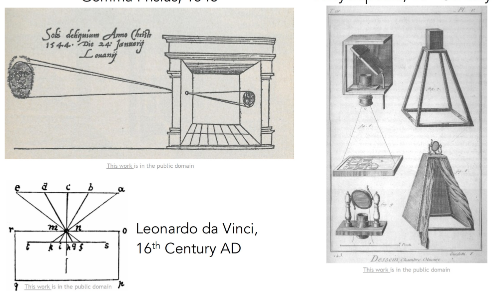


2025

Cat image by CNX OpenStax is licensed under CC BY 4.0; changes made

Gemma Frisius, 1545

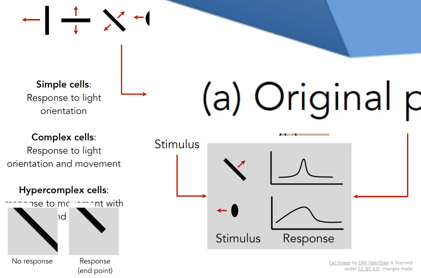
Encyclopedie, 18th Century



Leonardo da Vinci, 16th Century AD

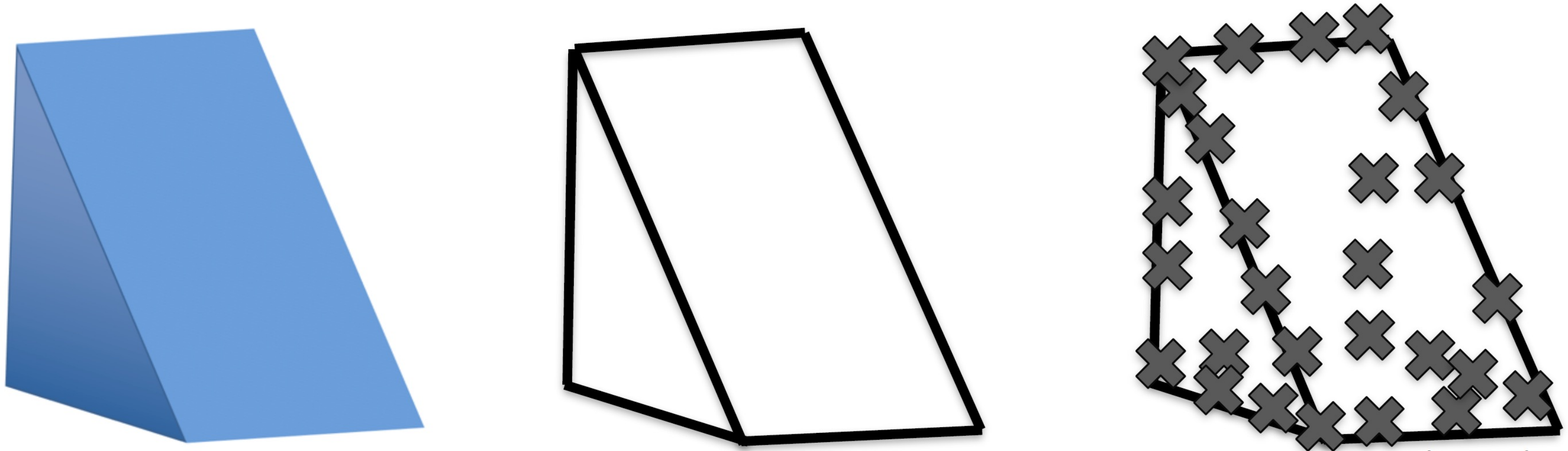
Evolution's Big Bang

543million years, B.C.



Block world

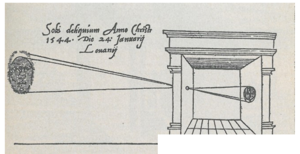
Larry Roberts, 1963



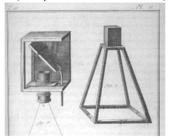
(a) Original picture (b) Differentiated picture (c) Feature points selected

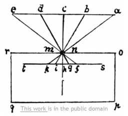
2025

Gemma Frisius, 1545



Encyclopedie, 18th Century

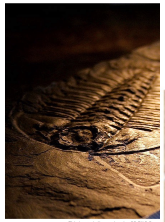




Lec
16th

- MIT Summer Vision Project, Seymor Papert, July 7, 1966
- Vision, Stages of Visual Representation, David Marr, 1970

Evolution's Big Ba

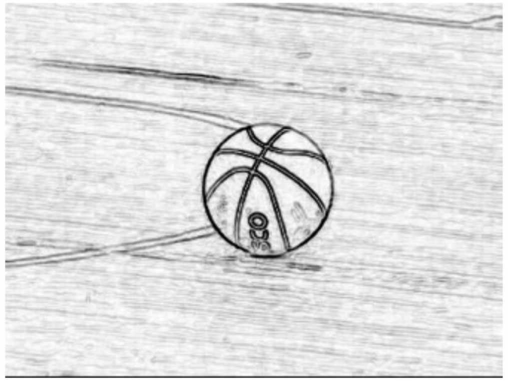


543mi

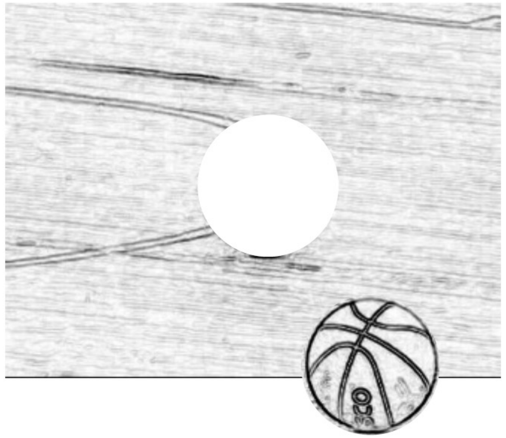
Input image

This image is CC0 1.0 public domain

Edge image

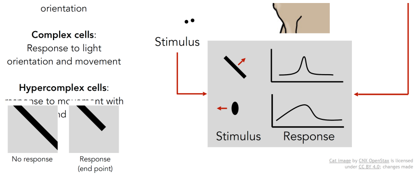


2 ½-D sketch



3-D model

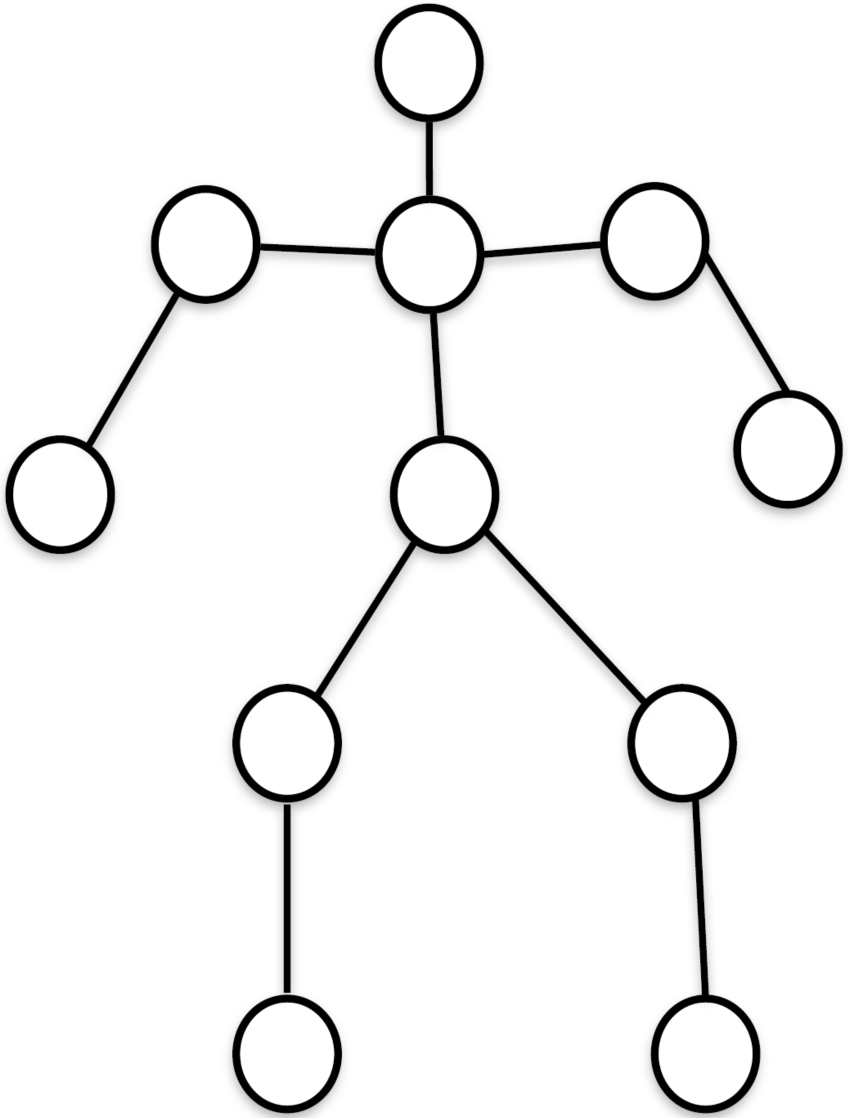
This image is CC0 1.0 public domain



2025

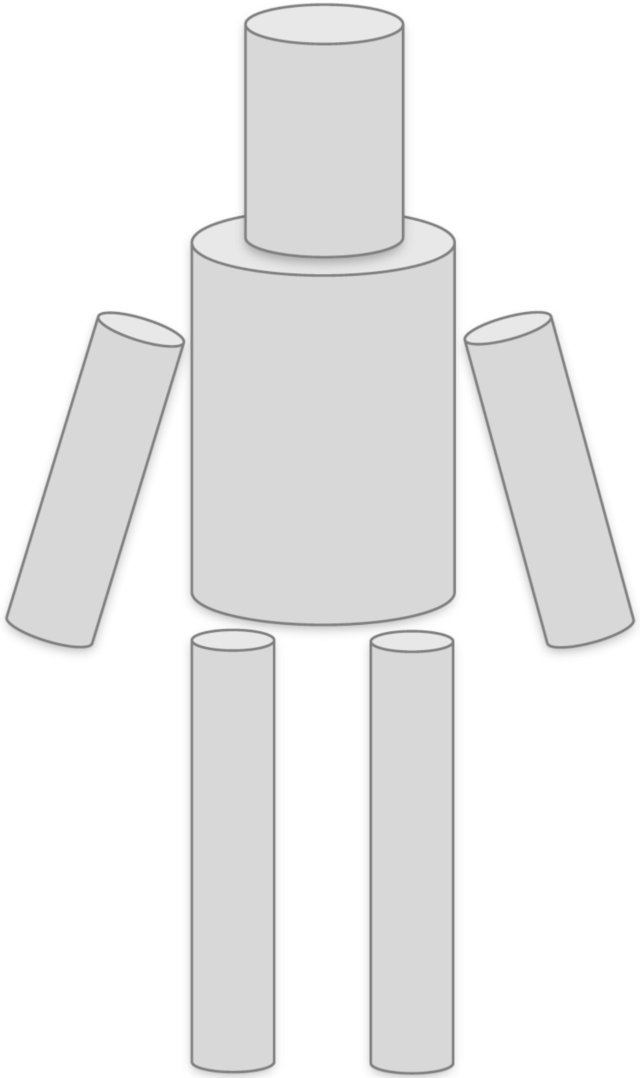
Pictorial Structure

Fischler and Elschlager, 1973

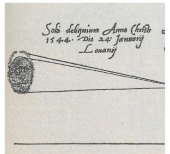


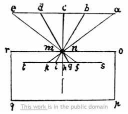
Generalized Cylinder

Brooks & Binford, 1979



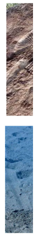
Gemma Frisi





Leor
16th

Evolution's Big Bang



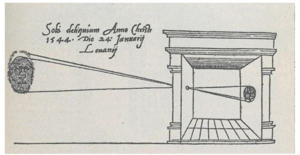
543million years, E

2025

Pictorial Structure
Fischler and Elschlager, 1973

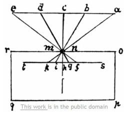
Generalized Cylinder
Brooks & Binford, 1979

Gemma Frisius, 1545



Encyclopedie, 18th Century

Normalized Cut (Shi & Malik, 1997)



Leonardo da Vinci, 16th Century AD

Evolution's Big Bang

543million years, B.C.

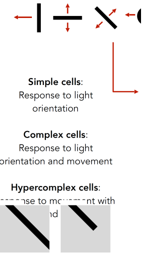


Image is CC BY 3.0
Image is public domain
Image is CC-BY SA 3.0

2025

Pictorial Structure
Fischler and Elschlager, 1973

Generalized Cylinder
Brooks & Binford, 1979

Evolutio

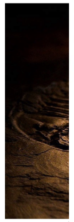


Image is public domain

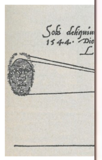
Image is CC BY-SA 2.0

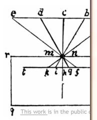
2025

"SIFT" & Object Recognition, David Lowe, 1999

Pictorial Structure
Fischler and Elschlager, 1973

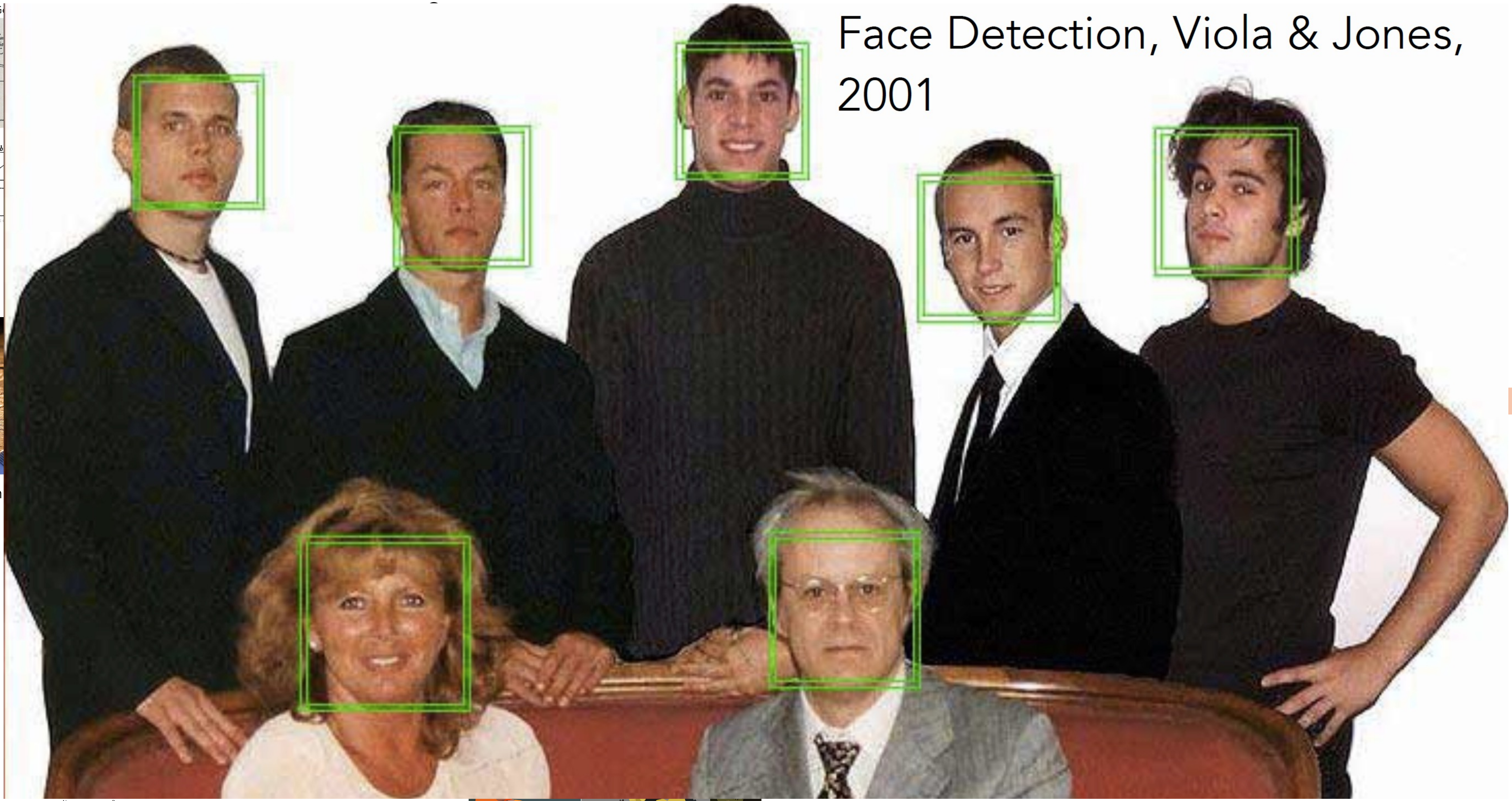
Generalized Cylinder
Brooks & Binford, 1979





Evolution's Big Bang

543million



No response

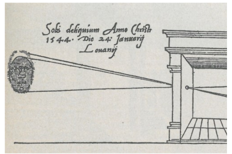
Response (end point)

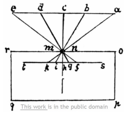
2025

Pictorial Structure
Fischler and Elschlager, 1973

Generalized Cylinder
Brooks & Binford, 1979

Gemma Frisius, 1545





Leonardo da V
16th Century A

PASCAL Visual Object Challenge (20 object categories)

[Everingham et al. 2006-2012]

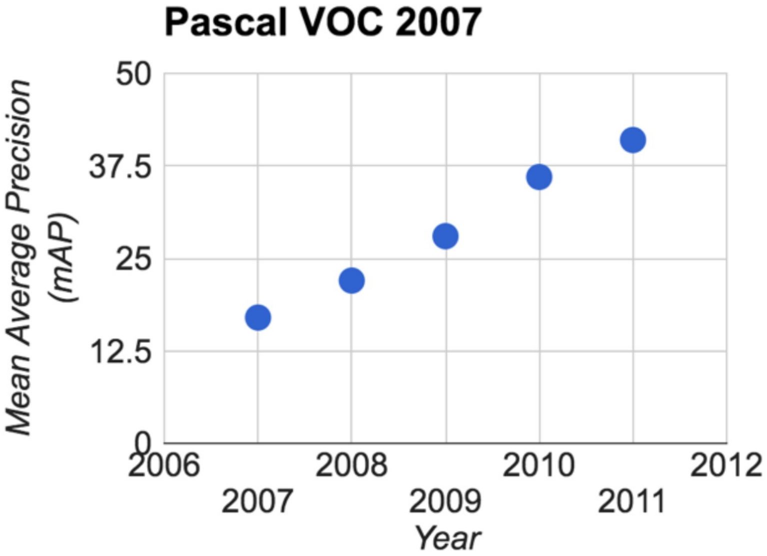
Evolution's Big Bang

543million years, B.C.

Image is CC BY-SA 3.0

Image is CC0 1.0 public domain

This image is licensed under CC BY-SA 2.0; changes made



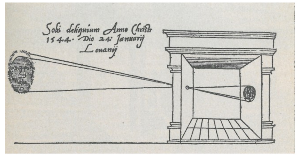
2025

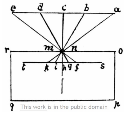
PASCAL Visual Object Challenge
(20 object categories)
[Everingham et al. 2006-2012]

Pictorial Structure

Generalized Cylinder

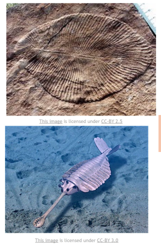
Gemma Frisius, 1545



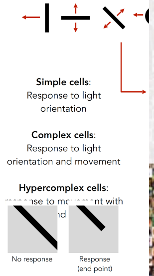


Leonardo da Vinci, 16th Century AD

Evolution's Big Bang



543million years, B.C.



IMAGENET

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

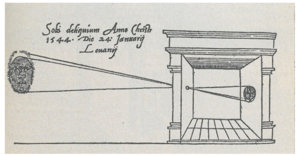
2025

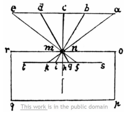
PASCAL Visual Object Challenge
(20 object categories)

Pictorial Structure

Generalized Cylinder

Gemma Frisius, 1545

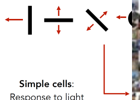




Leonardo da Vinci, 16th Century AD

Evolution's Big Bang

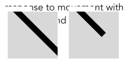
543million years, B.C.



Simple cells: Response to light orientation

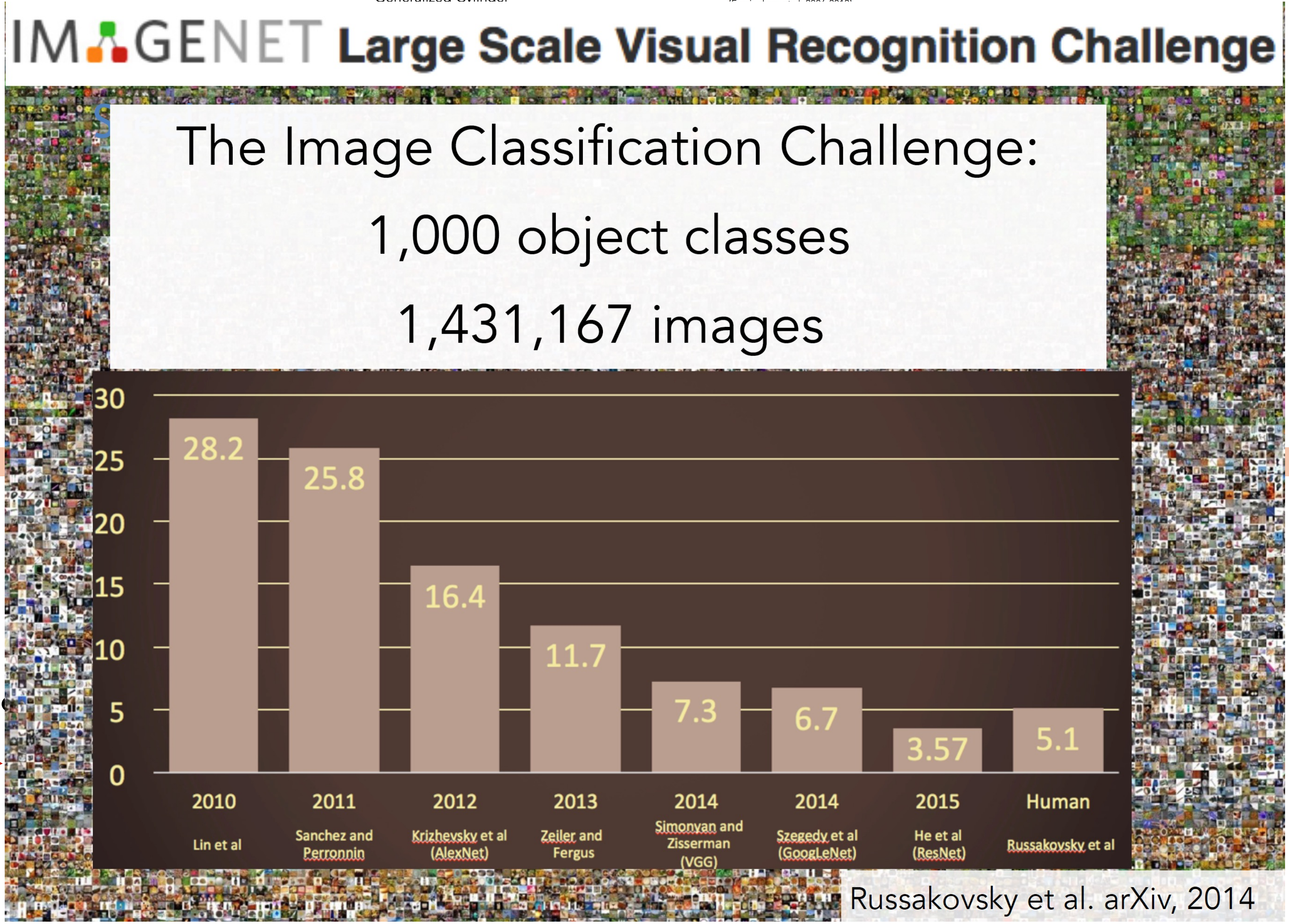
Complex cells: Response to light orientation and movement

Hypercomplex cells:



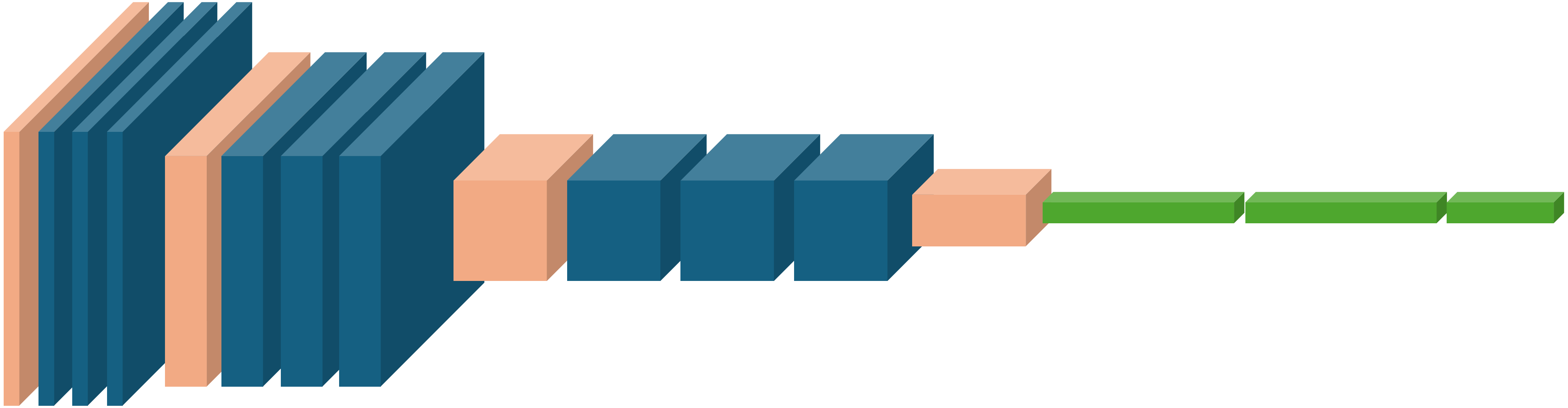
No response

Response (end point)



2025

Deep Convolutional Neural Networks (CNN),
AlexNet (2012), VGG (2014), ResNet (2015), DenseNet (2017), ...

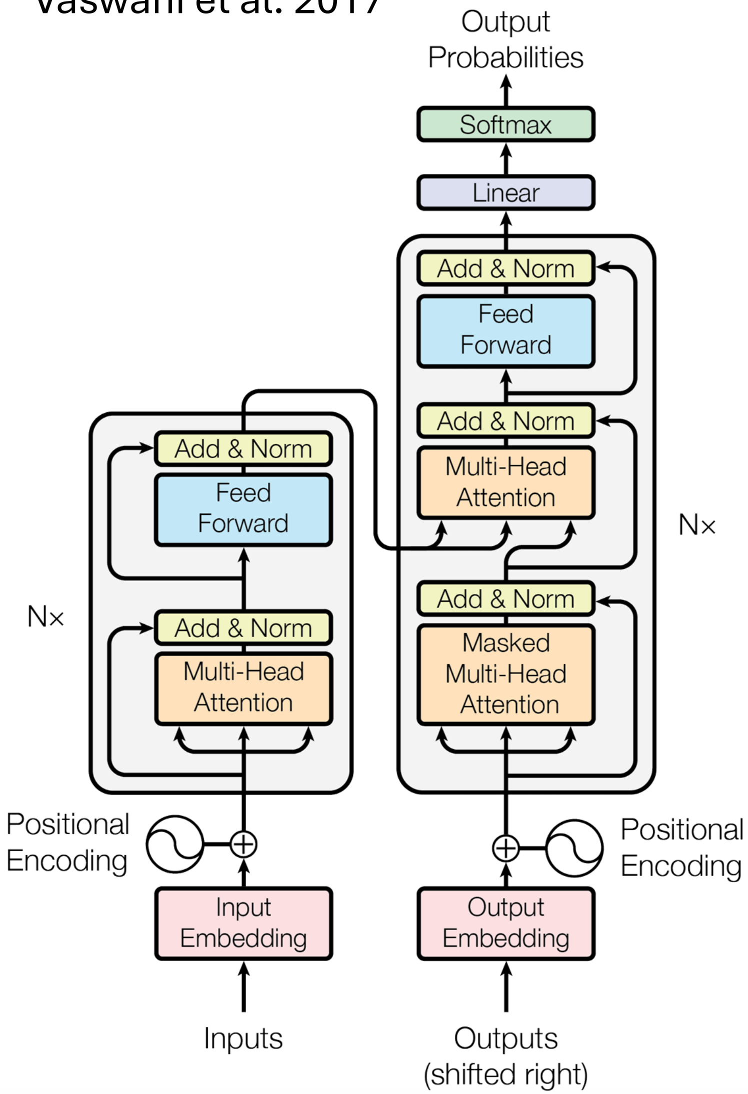


Evolut

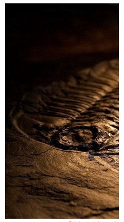
2025

Attention Is All We Need

Vaswani et al. 2017

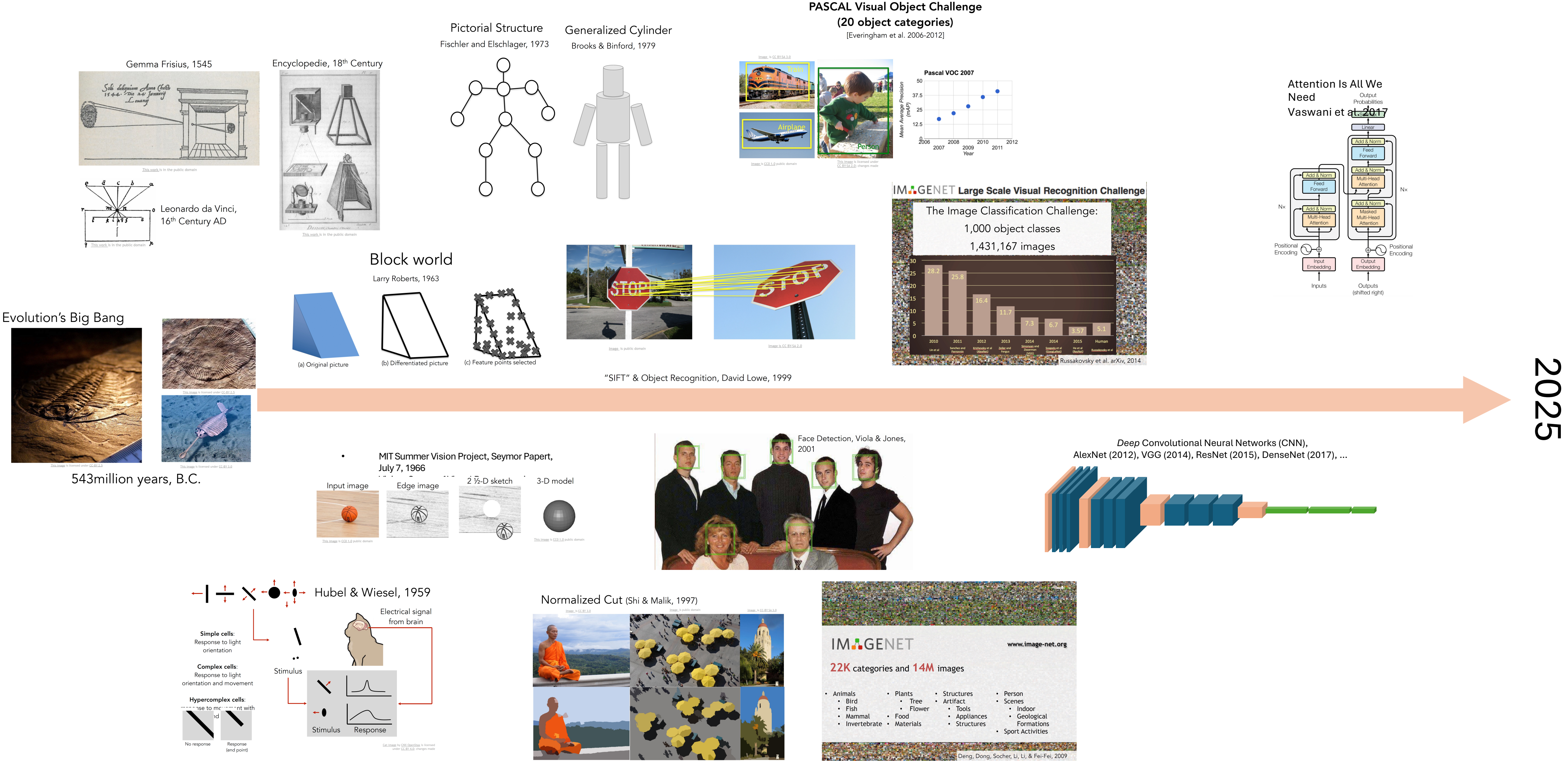


Evolution's Bi



54

2025



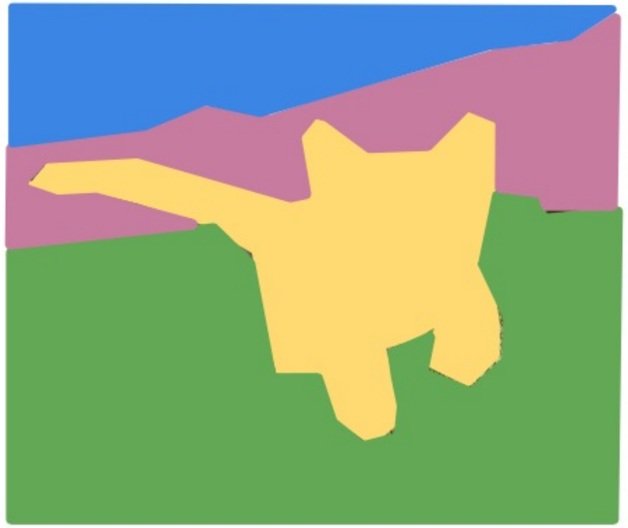
Gemma Frisius, 1545
Encyclopedie, 18th Century
Pictorial Structure
Fischler and Elschlager, 1973
Generalized Cylinder
Brooks & Binford, 1979
PASCAL Visual Object Challenge
(20 object categories)
[Everingham et al. 2006-2012]
Train
Airplane
Person
Pascal VOC 2007
Mean Average Precision (mAP)
Year
Attention Is All We Need
Vaswani et al. 2017
Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Multi-Head Attention
Masked Multi-Head Attention
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)
Nx
Leonardo da Vinci, 16th Century AD
Block world
Larry Roberts, 1963
(a) Original picture
(b) Differentiated picture
(c) Feature points selected
IMAGENET Large Scale Visual Recognition Challenge
The Image Classification Challenge:
1,000 object classes
1,431,167 images
28.2
25.8
16.4
11.7
7.3
6.7
3.57
5.1
2010
2011
2012
2013
2014
2014
2015
Human
Russakovsky et al. arXiv, 2014
Evolution's Big Bang
543million years, B.C.
"SIFT" & Object Recognition, David Lowe, 1999
2025
MIT Summer Vision Project, Seymor Papert, July 7, 1966
Input image
Edge image
2 ½-D sketch
3-D model
Face Detection, Viola & Jones, 2001
Deep Convolutional Neural Networks (CNN),
AlexNet (2012), VGG (2014), ResNet (2015), DenseNet (2017), ...
Hubel & Wiesel, 1959
Simple cells: Response to light orientation
Complex cells: Response to light orientation and movement
Hypercomplex cells:
No response
Response (end point)
Electrical signal from brain
Stimulus
Response
Normalized Cut (Shi & Malik, 1997)
IMAGENET
www.image-net.org
22K categories and 14M images
Animals
Bird
Fish
Mammal
Invertebrate
Plants
Tree
Flower
Food
Materials
Structures
Artifact
Tools
Appliances
Structures
Person
Scenes
Indoor
Geological Formations
Sport Activities
Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

More Advanced Image Processing

Classification

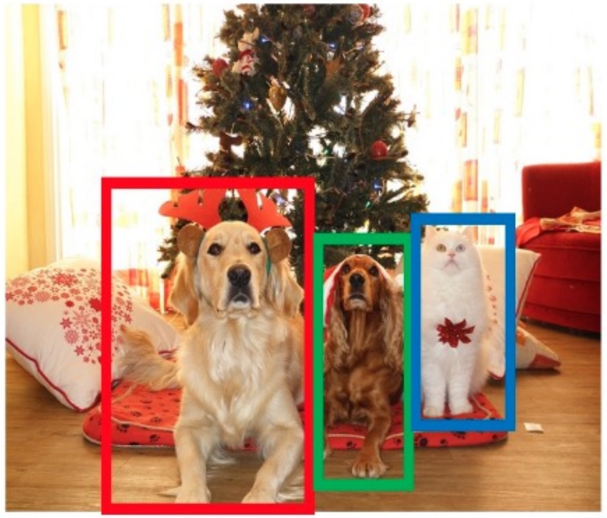
CAT

Semantic Segmentation



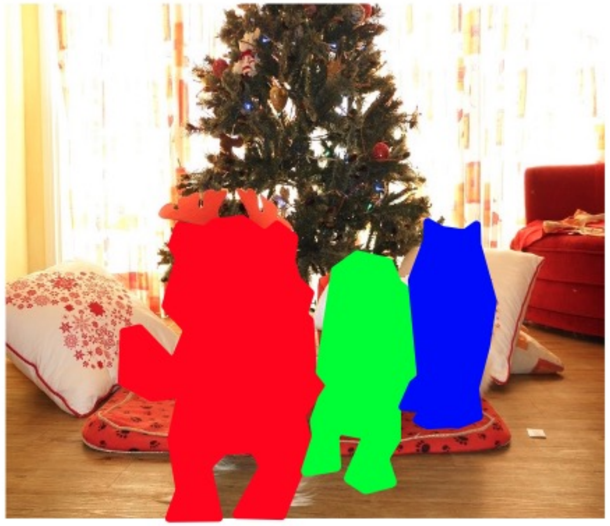
GRASS, CAT, TREE, SKY

Object Detection



DOG, DOG, CAT

Instance Segmentation



DOG, DOG, CAT

This image is CC public domain

Image Classification (Large Scale)

Image from www.image-net.org

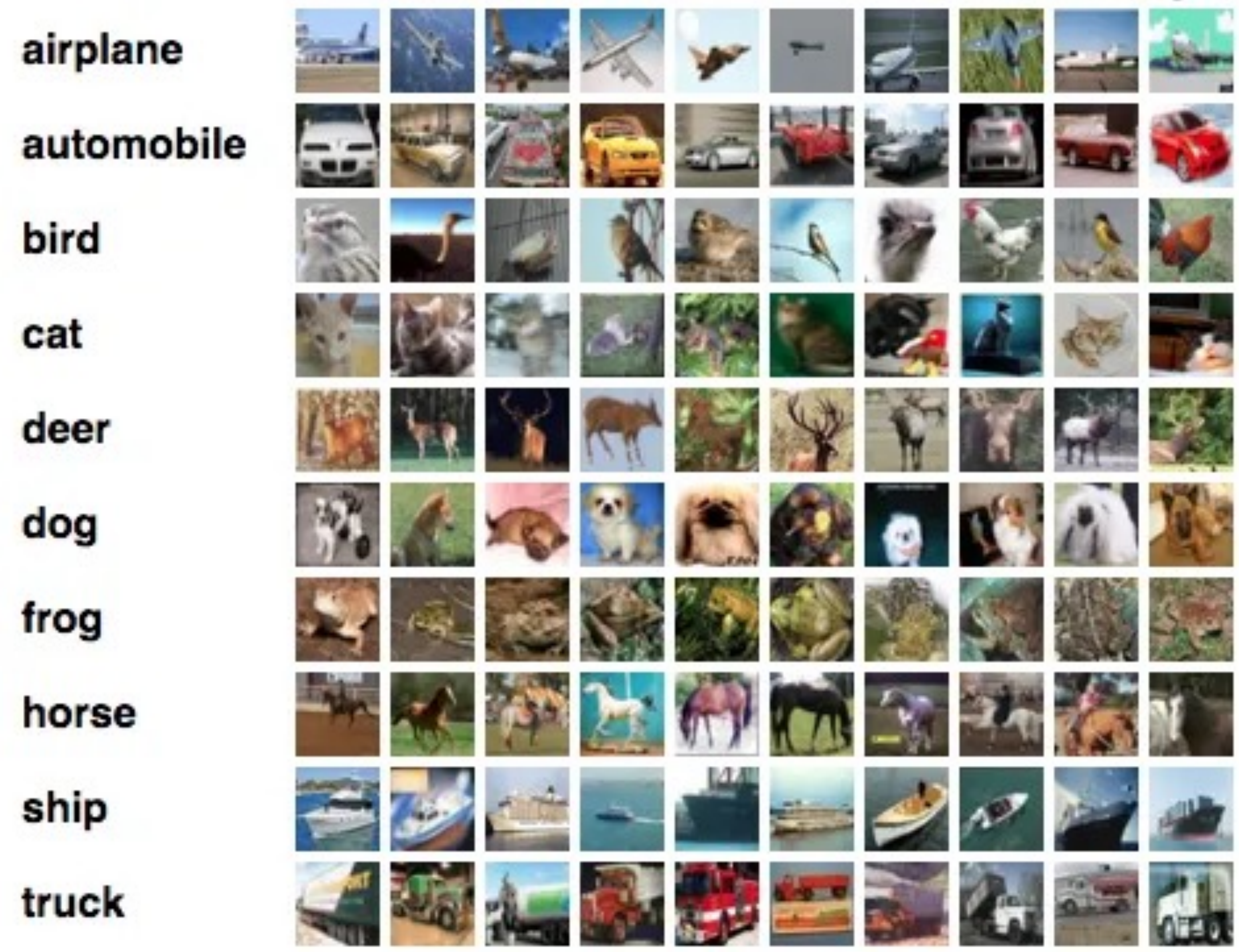


Image from cs.toronto.edu

Source: https://papercopilot.com/statistics/cvpr-statistics/

Image Processing

How Computers See

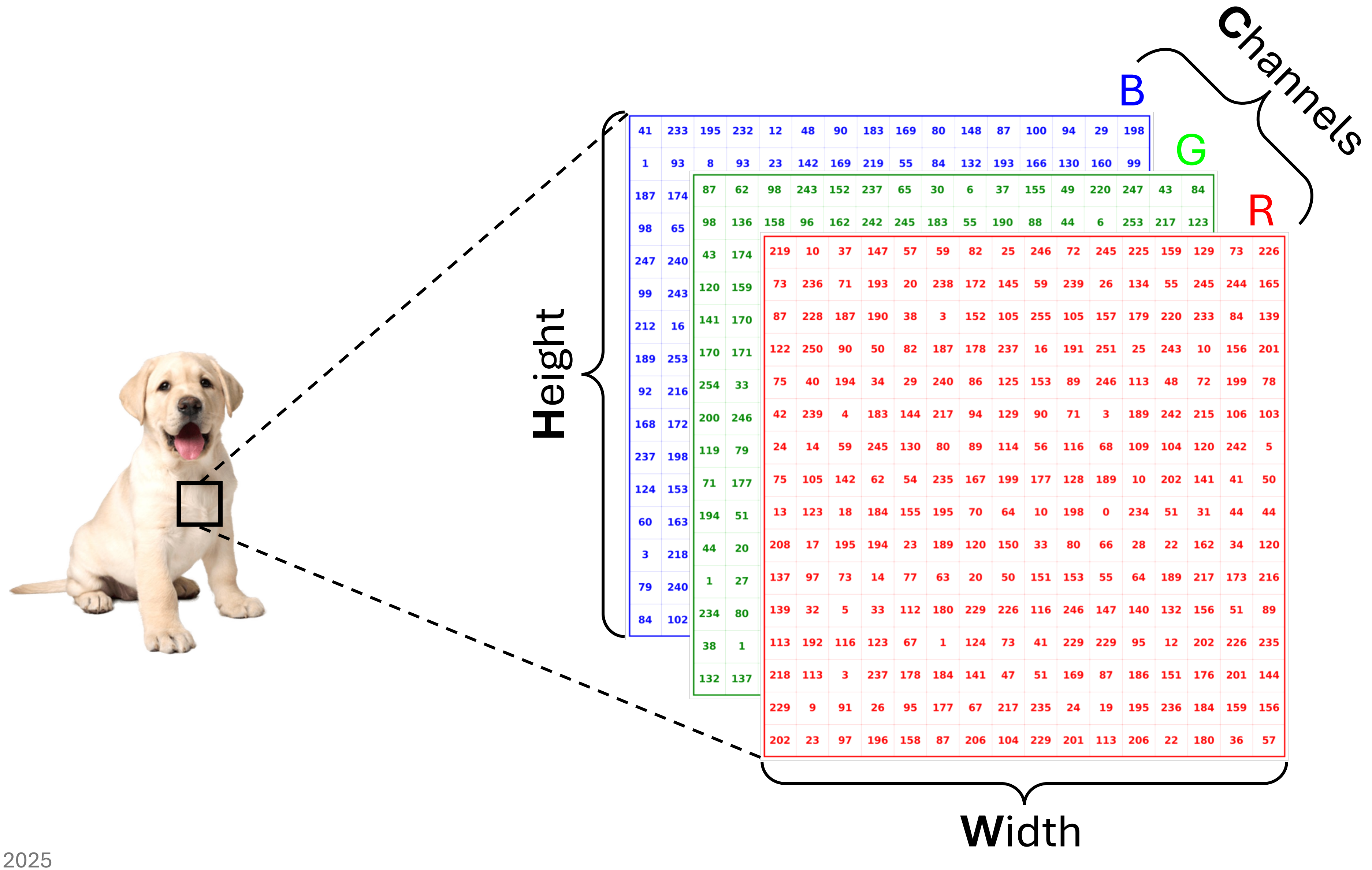


Image Processing

How Computers See

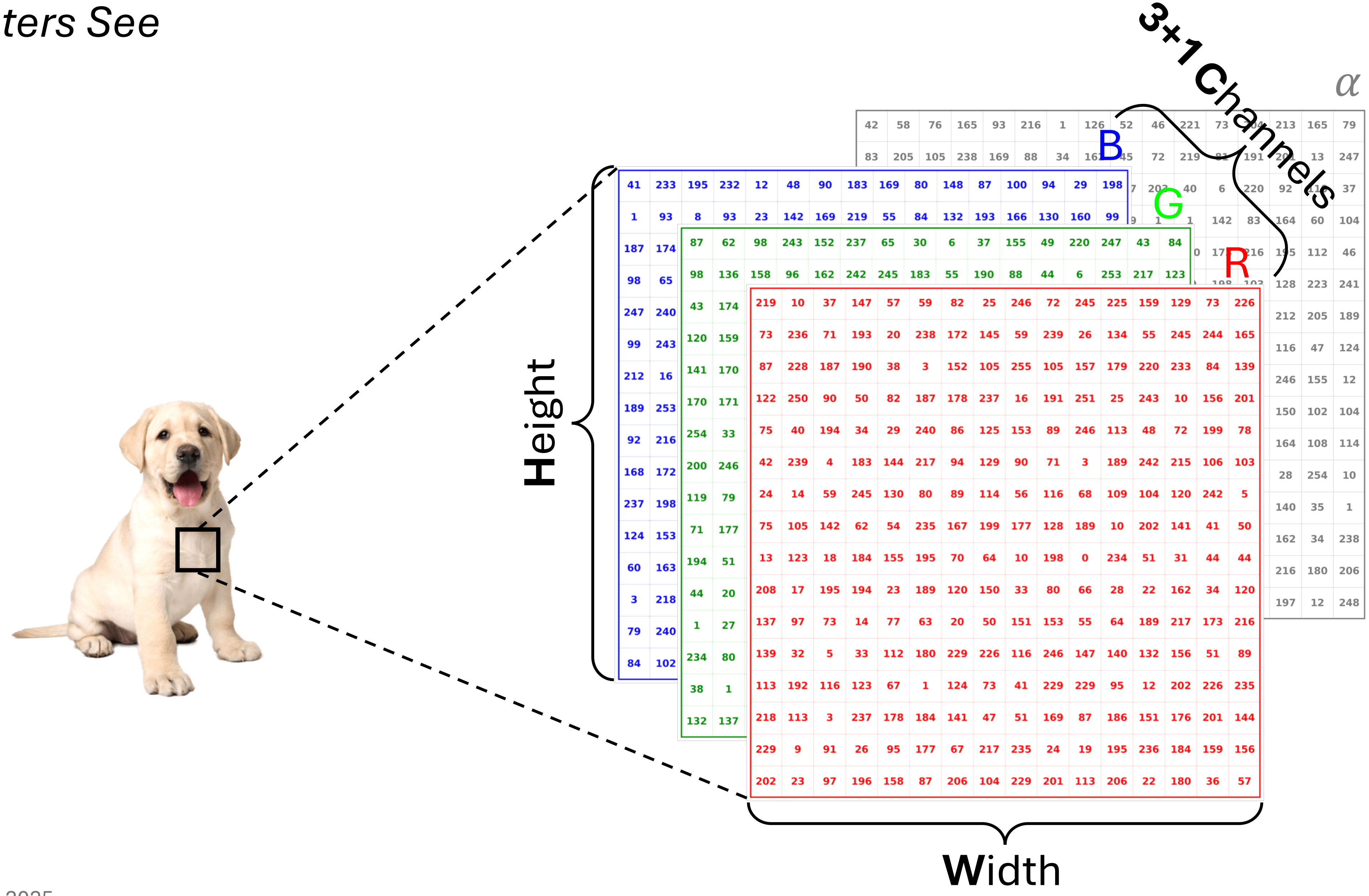


Image Processing

How Computers See

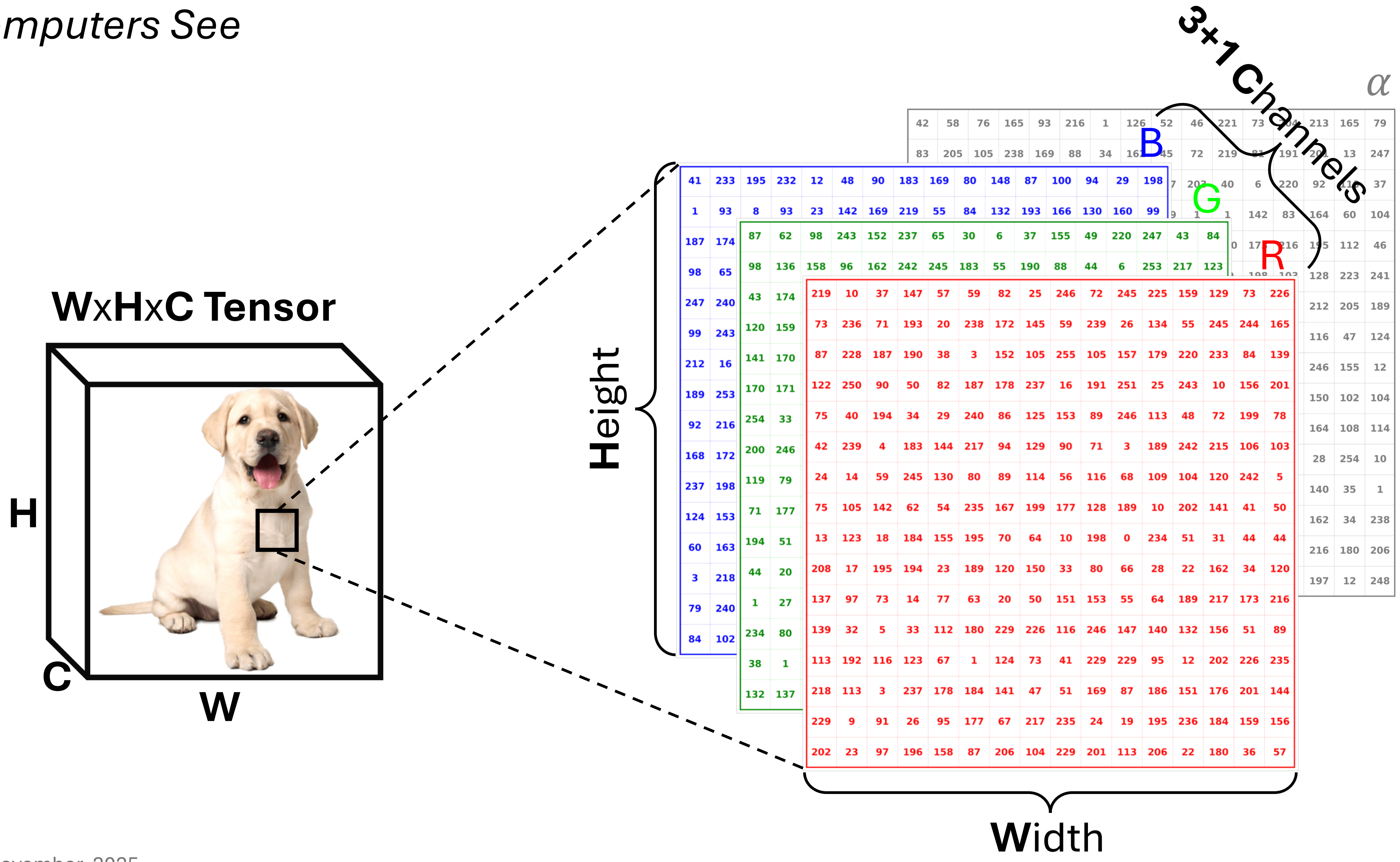
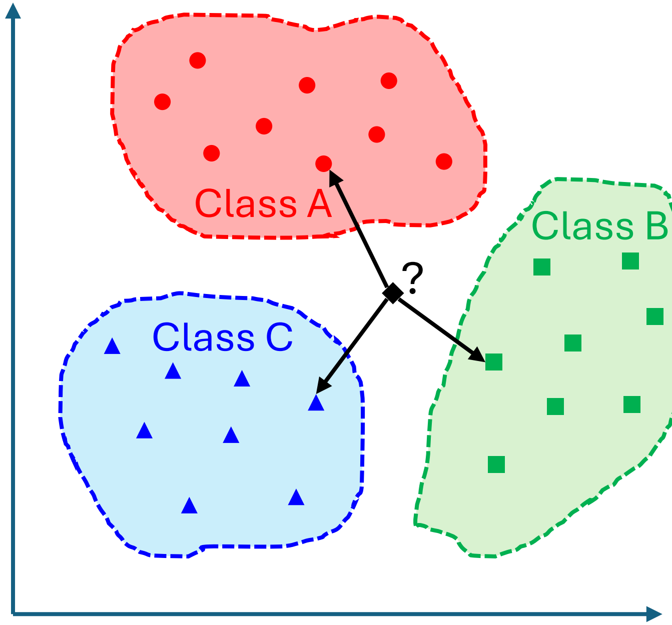


Image Classification

→ Dog (?)

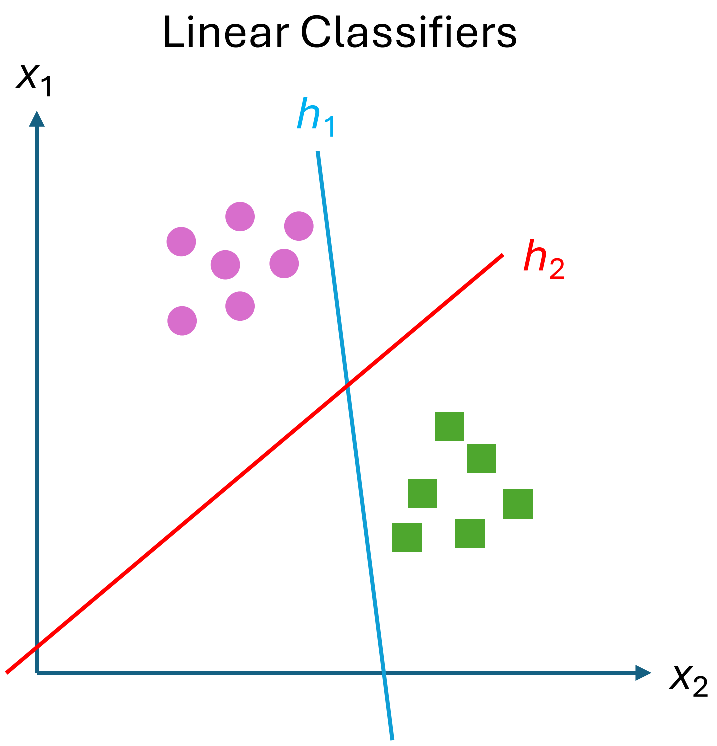
Image Classification

K Nearest Neighbors



Dog (?)

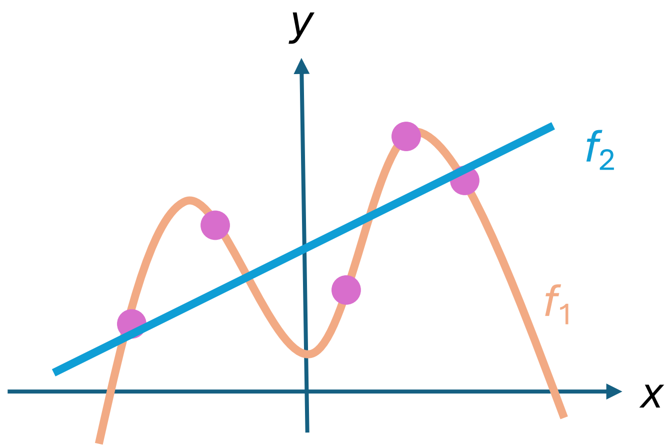
Image Classification



Dog (?)

Image Classification

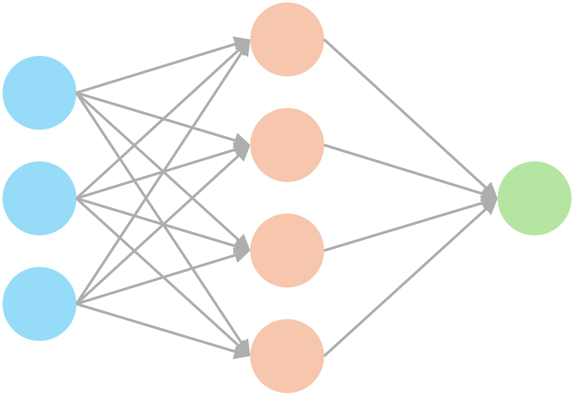
Regularization / Optimization



Dog (?)

Image Classification

Neural Networks



Dog (?)

Image Classification

Linear Approaches

$$f(x,W) = Wx$$

x (input image): 32x32x3=3072

W (weights/parameters): 10x3072

f (class scores): 10x1

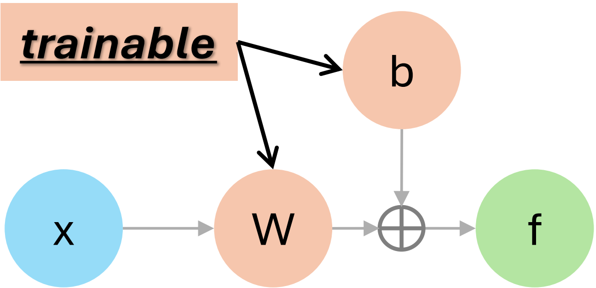
(CIFAR-10)

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

Image Classification

Linear Approaches

Linear Classifier



$$f(x,W) = Wx + b$$

x (input image): 32x32x3=3072

W (weights/parameters): 10x3072

f (class scores): 10x1

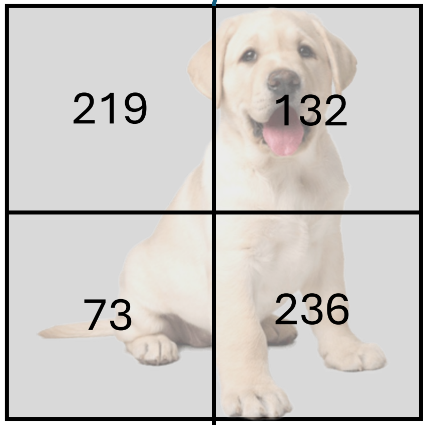
b (bias): 10x1

(CIFAR-10)

- airplane
- automobile
- bird
- cat
- deer
- dog
- frog
- horse
- ship
- truck

Image Classification

Linear Classifiers



Flatten tensor into a vector

$$\underbrace{\begin{bmatrix} 0.2 & 1.5 & -0.8 & 0.2 \\ -0.5 & 0.6 & 0.0 & -1.3 \\ 0.1 & 2.1 & -1.1 & 1.2 \end{bmatrix}}_{W} \times \underbrace{\begin{bmatrix} 219 \\ 132 \\ 73 \\ 236 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} -0.1 \\ 0.2 \\ 1.1 \end{bmatrix}}_{b} = \begin{bmatrix} 230.5 \\ -336.9 \\ 503.1 \end{bmatrix}$$

Bird
Cat
Dog

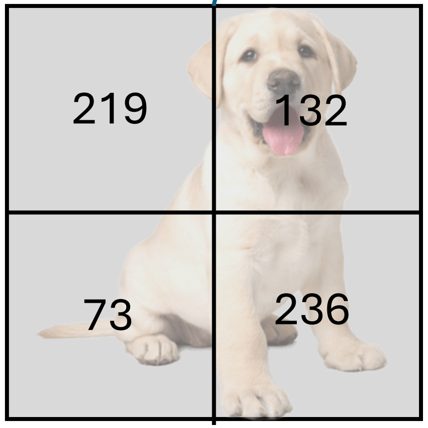
Image Classification

Linear Classifiers

```python
import torch

x = [219, 132, 73, 236]
W = [[0.2,  1.5, -0.8, 0.2], [-0.5, 0.6, 0.0, -1.3], [0.1, 2.1, -1.1, 1.2]]
b = [-0.1, 0.2, 1.1]

x_tensor = torch.tensor(x, dtype=torch.float32)
W_tensor = torch.tensor(W, dtype=torch.float32)
b_tensor = torch.tensor(b, dtype=torch.float32)

y = torch.matmul(W_tensor, x_tensor) + b_tensor
print(y)
```

```
tensor([ 230.5000, -336.9000,  503.1000])
```

Flatten tensor into a vector

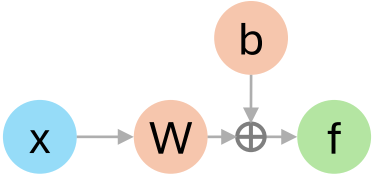


$$\underbrace{\begin{bmatrix} 0.2 & 1.5 & -0.8 & 0.2 \\ -0.5 & 0.6 & 0.0 & -1.3 \\ 0.1 & 2.1 & -1.1 & 1.2 \end{bmatrix}}_{W} \times \underbrace{\begin{bmatrix} 219 \\ 132 \\ 73 \\ 236 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} -0.1 \\ 0.2 \\ 1.1 \end{bmatrix}}_{b} = \begin{bmatrix} 230.5 \\ -336.9 \\ 503.1 \end{bmatrix}$$

Bird
Cat
Dog

Image Classification

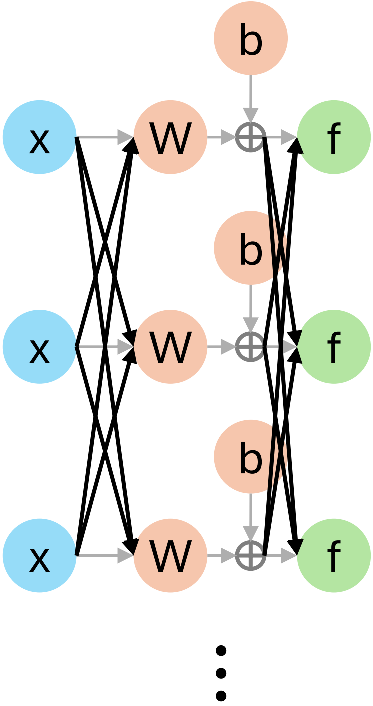
Neural Networks



- airplane
- automobile
- bird
- cat
- deer
- dog
- frog
- horse
- ship
- truck

Image Classification

Neural Networks

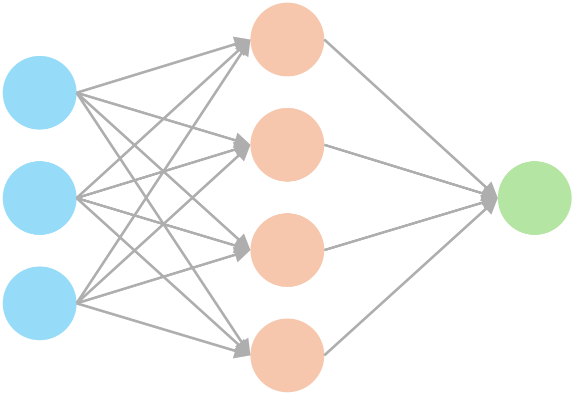


- airplane
- automobile
- bird
- cat
- deer
- dog
- frog
- horse
- ship
- truck

Image Classification

Neural Networks

Neural Networks



- airplane
- automobile
- bird
- cat
- deer
- dog
- frog
- horse
- ship
- truck

Image Classification

Loss functions

Softmax:

- Likelihood of classes
- Entropy (KL-divergence)
- Maximizing probability of the correct class

$$\mathcal{L}_i = -\log\left(\frac{p(\hat{y}_i)}{\sum_j p(\hat{y}_j)}\right)$$

SVM (multiclass):

- Scores to each class
- Difference between them

$$\mathcal{L}_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

Regularizations

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_i(f(x_i, W), \hat{y}_i) + \boldsymbol{\lambda R(W)}$$

$\boldsymbol{R(W)} =$

- L_1 regularization: $\sum_k \sum_l |W_{k,l}|$
- L_2 regularization: $\sum_k \sum_l W_{k,l}^2$
- Elastic: $\sum_k \sum_l |W_{k,l}| + \sum_k \sum_l W_{k,l}^2$
- Dropout
- Batch normalization
- Stochastic depth

λ: regularization coefficient (hyperparameter)

Optimizations

$$\nabla_W \mathcal{L} = \frac{\partial\left(\frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_i(f(x_i, W), \hat{y}_i) + \lambda R(W)\right)}{\partial W}$$

Numerical gradient:

- Approximate, slow, easy to write

Analytic gradient:

- Exact, fast, prone to errors

<u>Note</u>: Always use analytic but check with numerical; "*gradient check*"!

Image Classification

Optimization: Gradient Descent

$$\nabla_W \mathcal{L} = \frac{\partial \left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i(f(x_i, W), \hat{y}_i) + \lambda R(W) \right)}{\partial W}$$

```
# Gradient Descent

TRAINING = True
while TRAINING:
    weights_gradient = evaluate_gradient(loss_function, data, weights)
    weights += learning_rate * weights_gradient ######## update parameters
```

$$W_l^{t+1} = W_l^t + \eta \frac{\partial \mathcal{L}}{\partial W_l} |_{W_l^t}$$

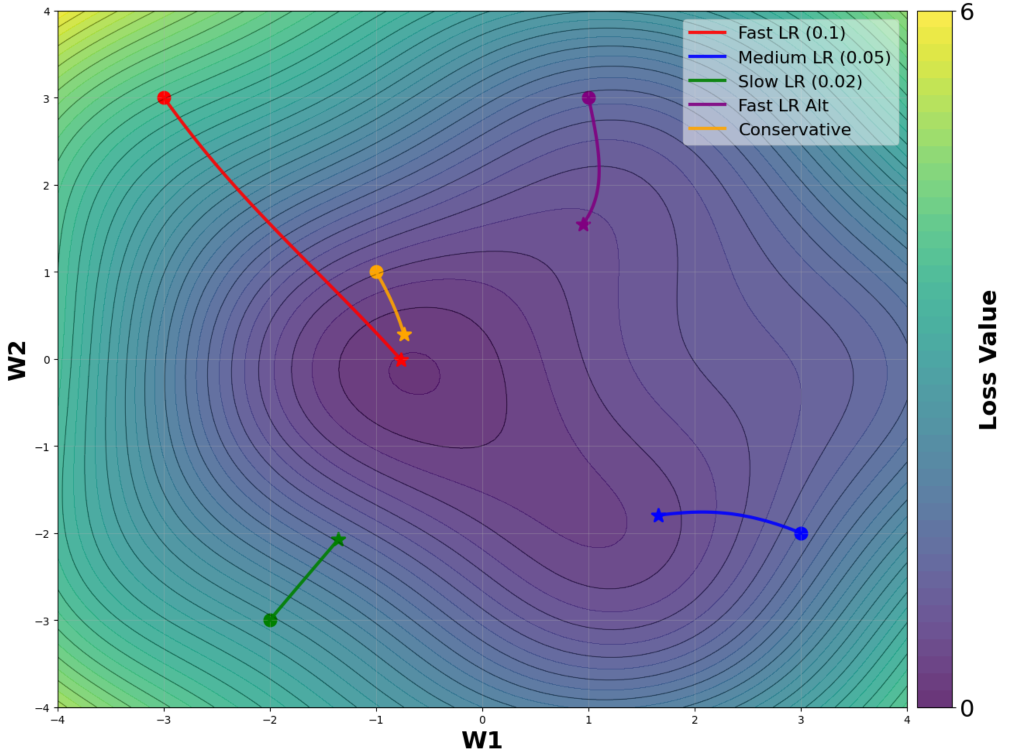
Image Classification

Optimization: Gradient Descent

$$\nabla_W \mathcal{L} = \frac{\partial\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i(f(x_i, W), \hat{y}_i) + \lambda R(W)\right)}{\partial W}$$

```
# Gradient Descent

TRAINING = True
while TRAINING:
    weights_gradient = evaluate_gradient(loss_function, data, weights)
    weights += learning_rate * weights_gradient ######## update parameters
```

$$W_l^{t+1} = W_l^t + \eta \frac{\partial \mathcal{L}}{\partial W_l}\Big|_{W_l^t}$$



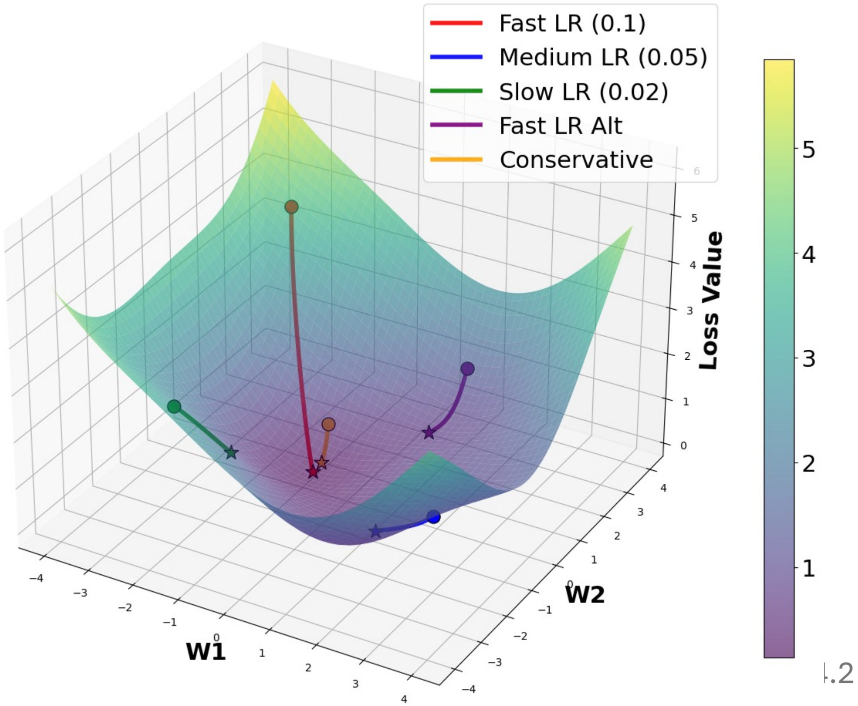


Image Classification

Backpropagation

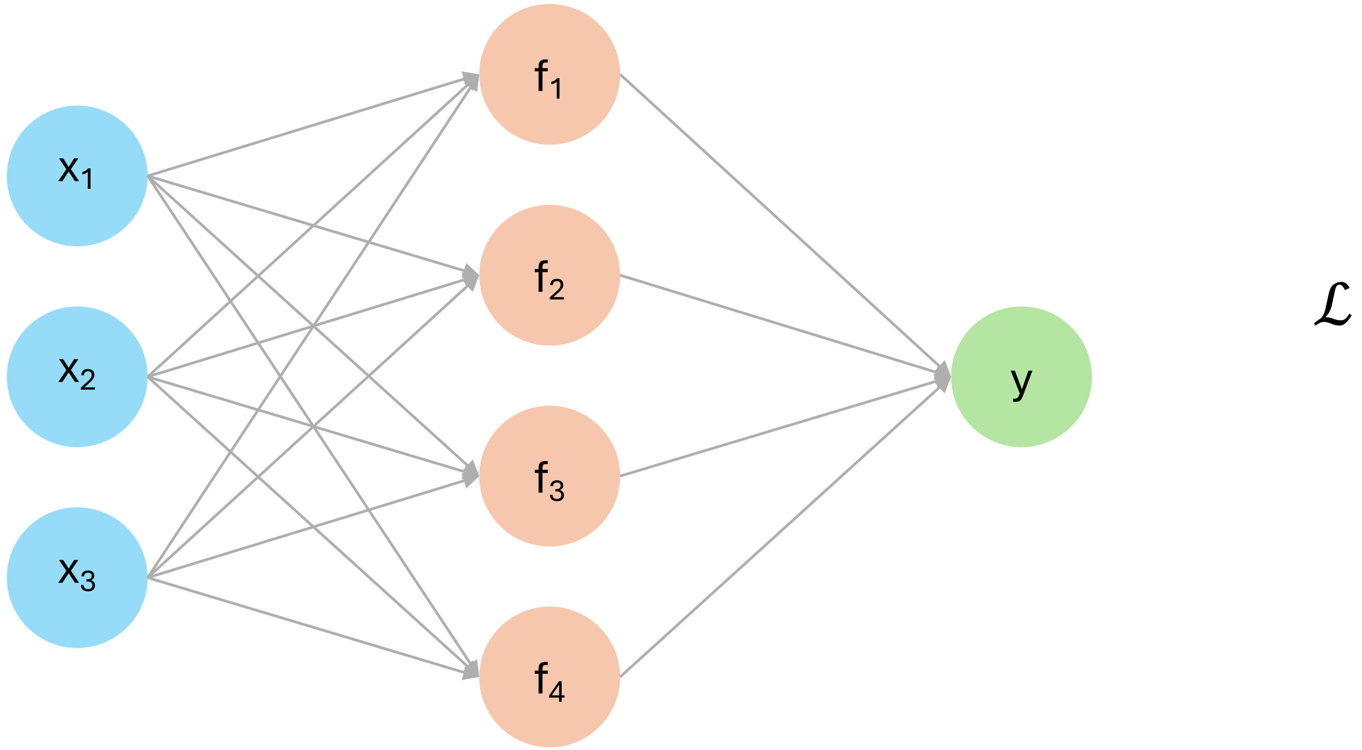


Image Classification

Backpropagation

$$\boxed{\frac{\partial \mathcal{L}}{\partial f_1} = \frac{\partial \mathcal{L}}{\partial y}\frac{\partial y}{\partial f_1}}$$

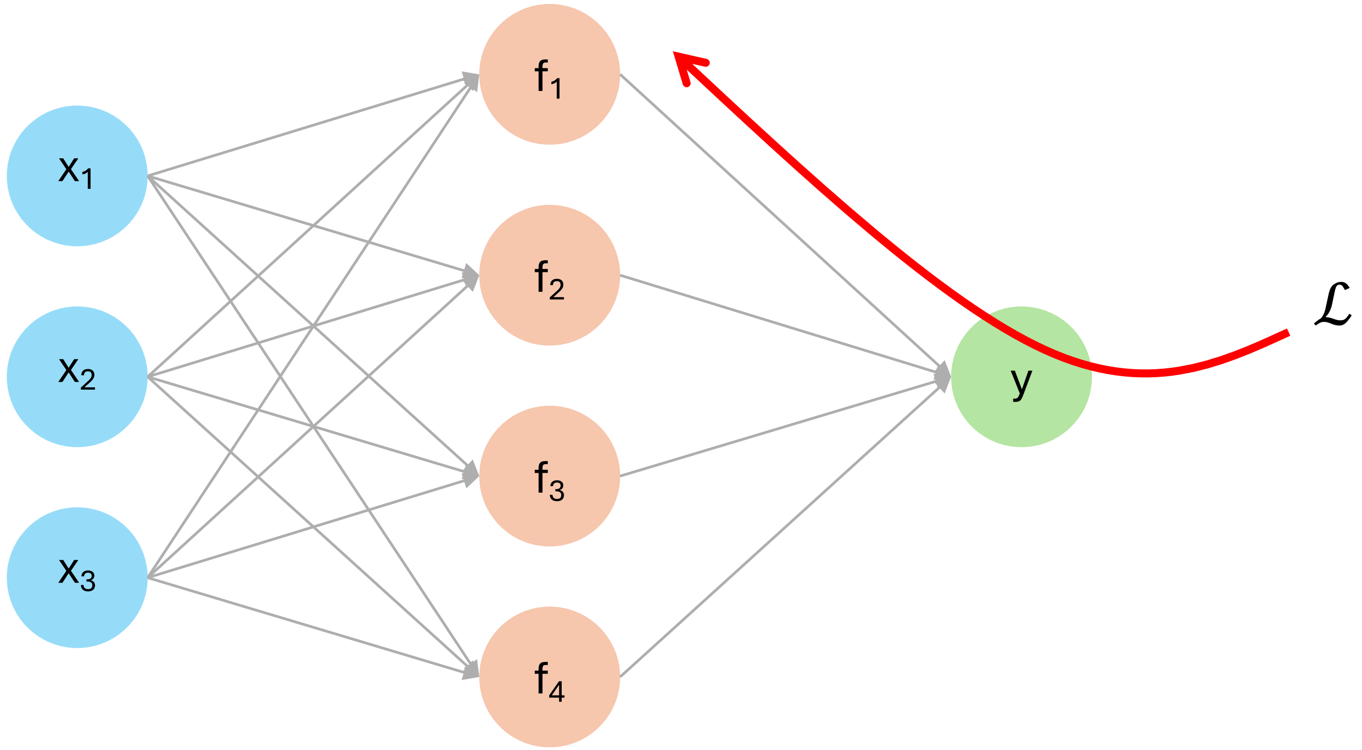
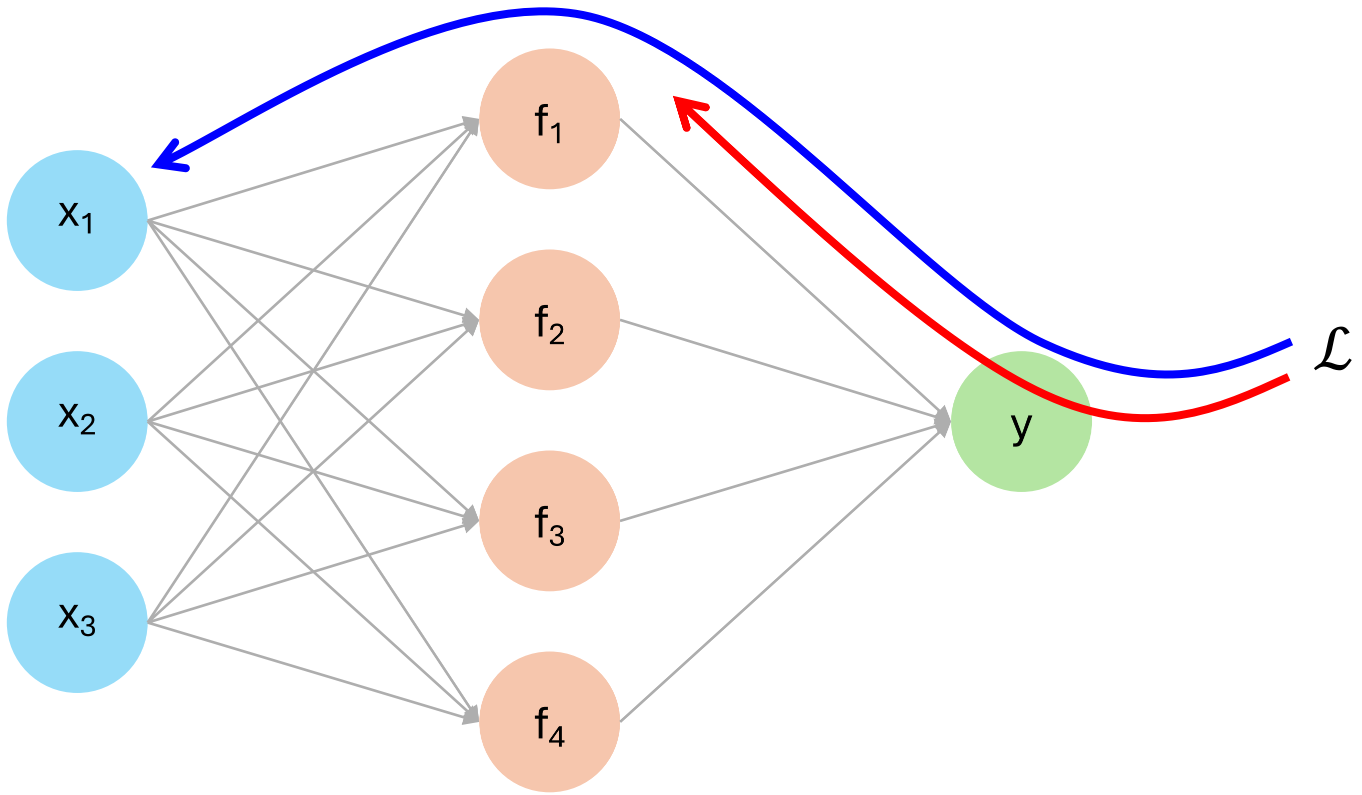


Image Classification

Backpropagation

$$\frac{\partial \mathcal{L}}{\partial f_1} = \frac{\partial \mathcal{L}}{\partial y}\frac{\partial y}{\partial f_1}$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial \mathcal{L}}{\partial y}\frac{\partial y}{\partial f_1}\frac{\partial f_1}{\partial x_1}$$



Let’s Code!

<\>

Break

(Q&A)

Image Classification

Using Machine Learning: *data-driven*

Image Classification

Using Machine Learning: *data-driven*

1. A dataset of images and labels

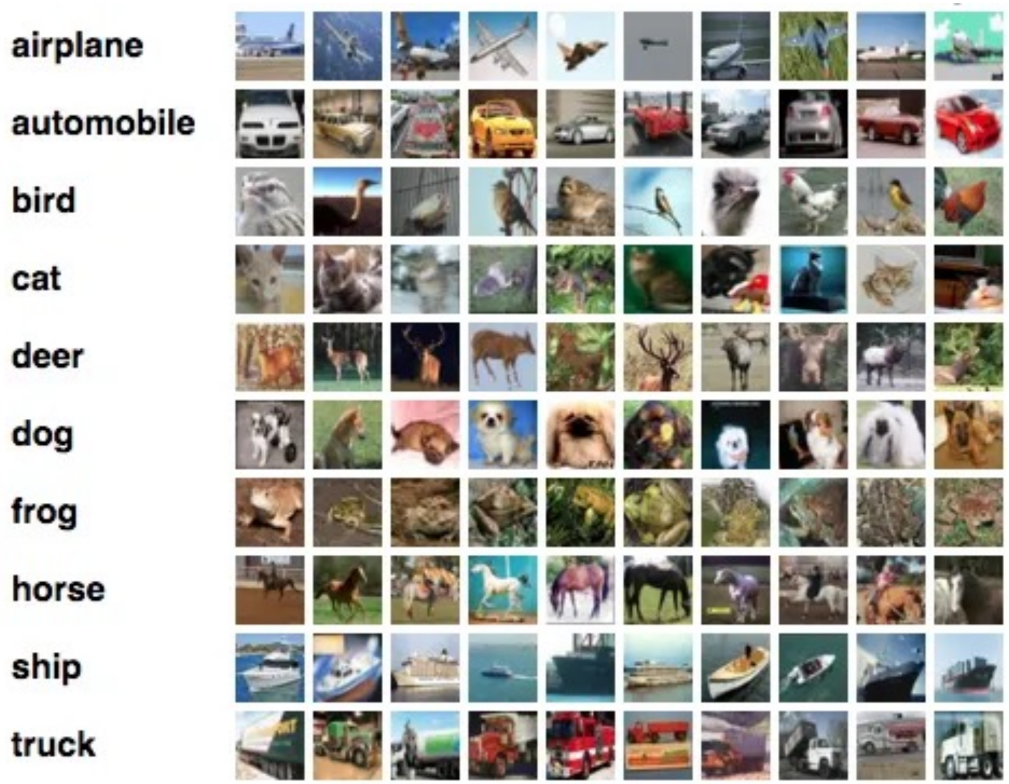


Image from cs.toronto.edu

Image Classification

Using Machine Learning: *data-driven*

1. A dataset of images and labels

2. Train a Machine Learning classifier

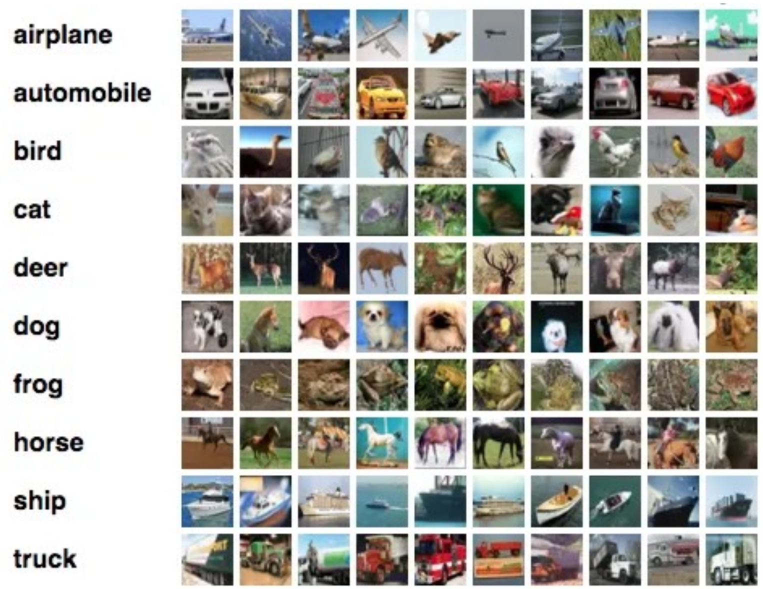


Image from cs.toronto.edu

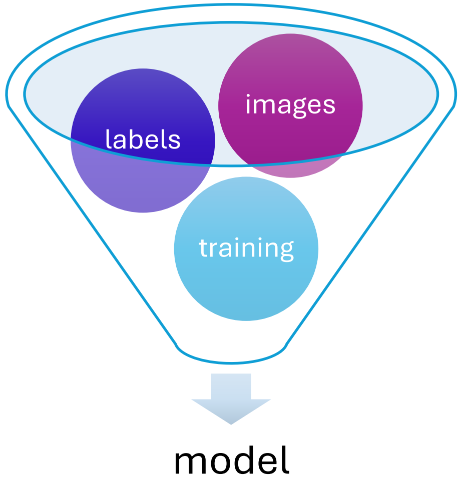


Image Classification

Using Machine Learning: *data-driven*

1. A dataset of images and labels

2. Train a Machine Learning classifier

3. Test and evaluate on new images

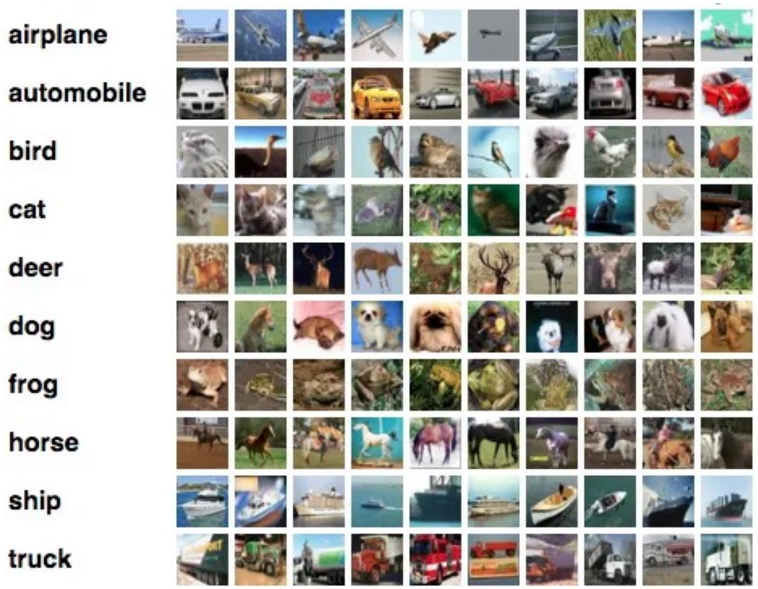
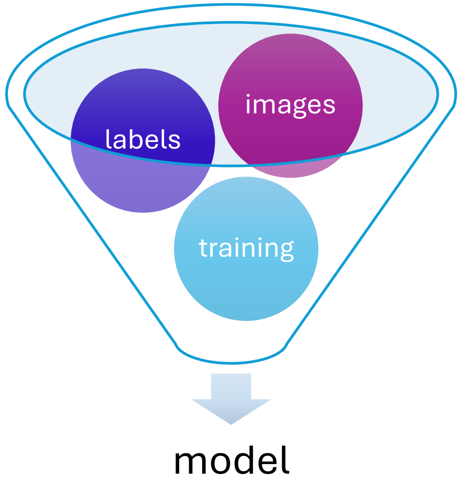
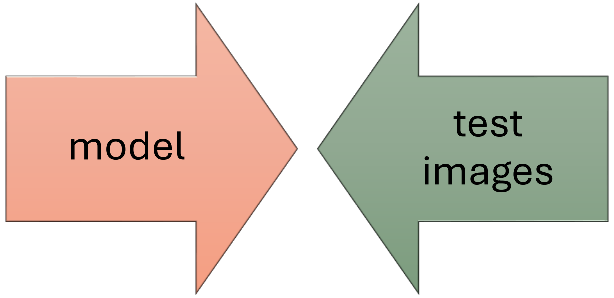


Image from cs.toronto.edu





Challenges and Limitations:

Challenges and Limitations:

- *Background Clutter*

Images from: https://www.the-sun.com/news/671403/dog-background-hidden-spot-them/

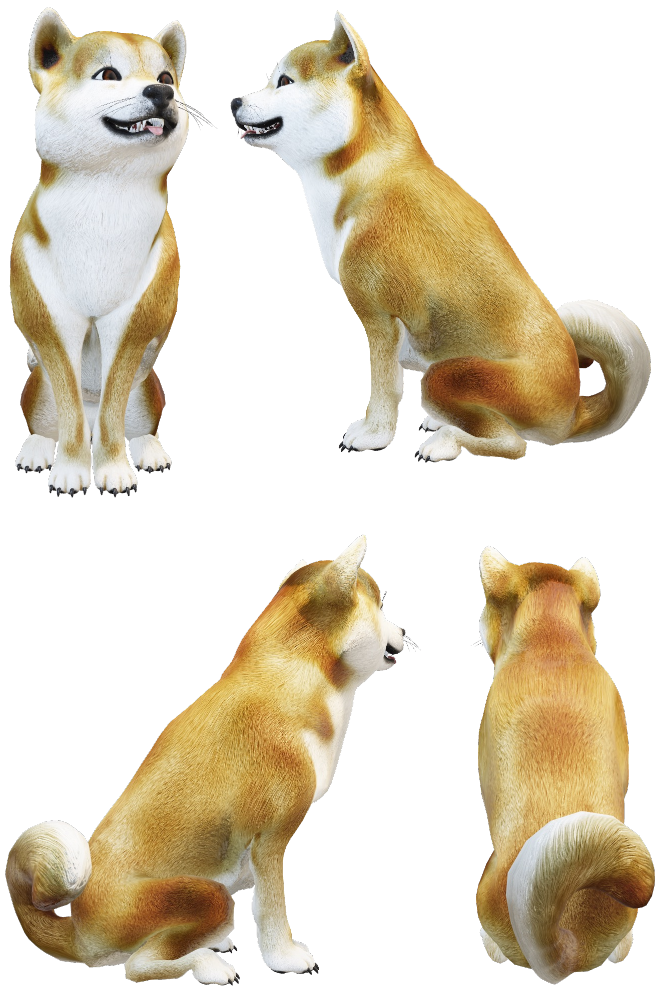
Challenges and Limitations:

- *Background Clutter*
- *Illumination*

Images CC public domain

Challenges and Limitations:

- *Background Clutter*
- *Illumination*
- *Viewpoint*



Challenges and Limitations:

- *Background Clutter*
- *Illumination*
- *Viewpoint*
- *Occlusion*


Images CC public domain

Challenges and Limitations:

- *Background Clutter*
- *Illumination*
- *Viewpoint*
- *Occlusion*
- *Intraclass Variations*

Challenges and Limitations:

- *Background Clutter*
- *Illumination*
- *Viewpoint*
- *Occlusion*
- *Intraclass Variations*
- *Deformation*

Images CC public domain

Challenges and Limitations:

- *Background Clutter*
- *Illumination*
- *Viewpoint*
- *Occlusion*
- *Intraclass Variations*
- *Deformation*
- *Contextual*

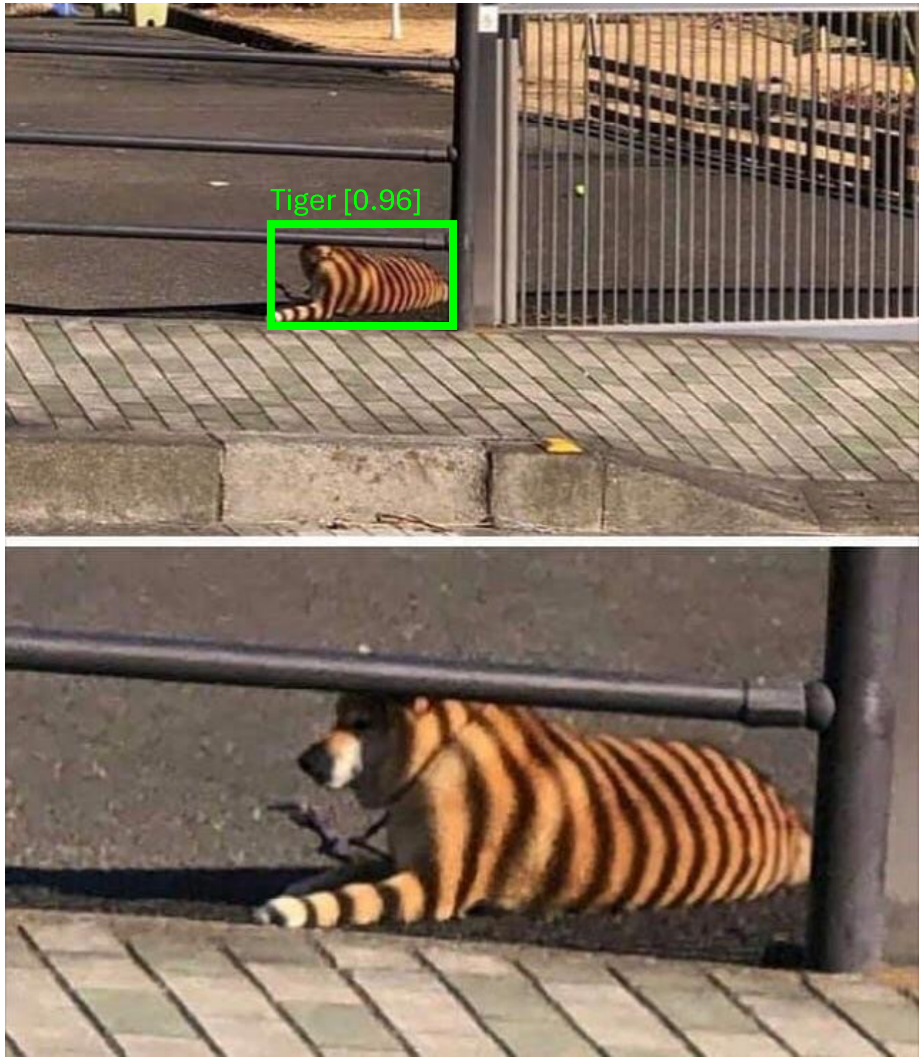


Image from: https://www.reddit.com/r/confusing_perspective/comments/af97sy/is_that_a_tig_oh_never_mind/

Convolutional Neural Networks (CNN)

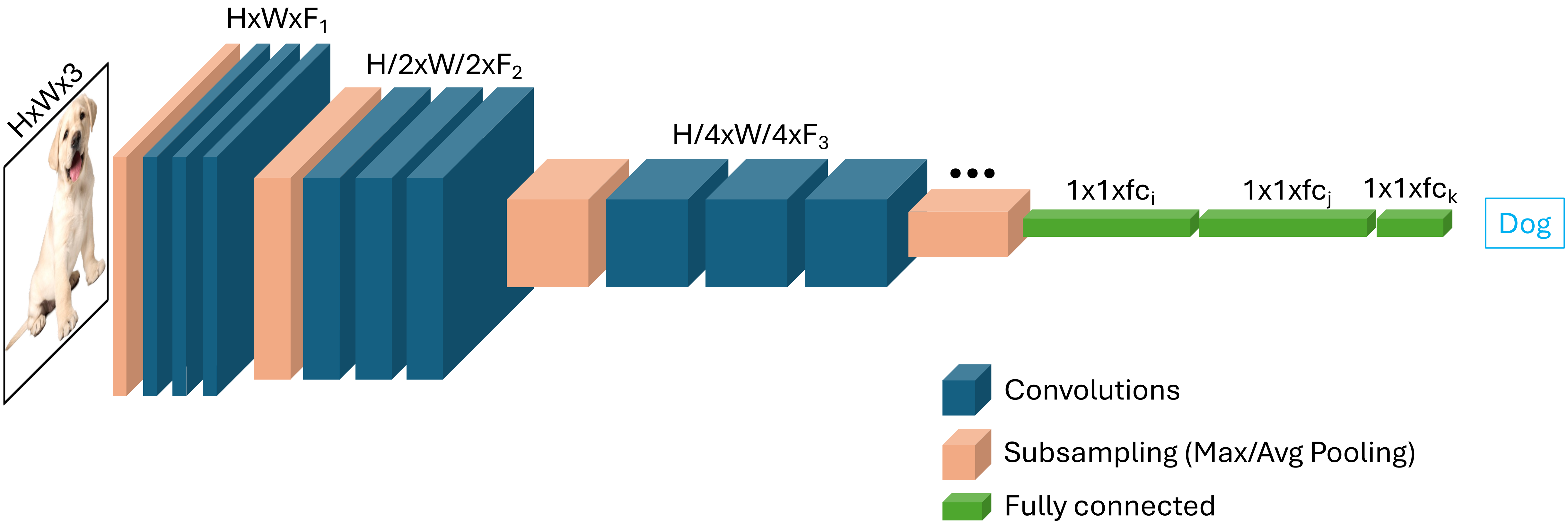


Image from Kebria et al. 2018

Image Features and CNN

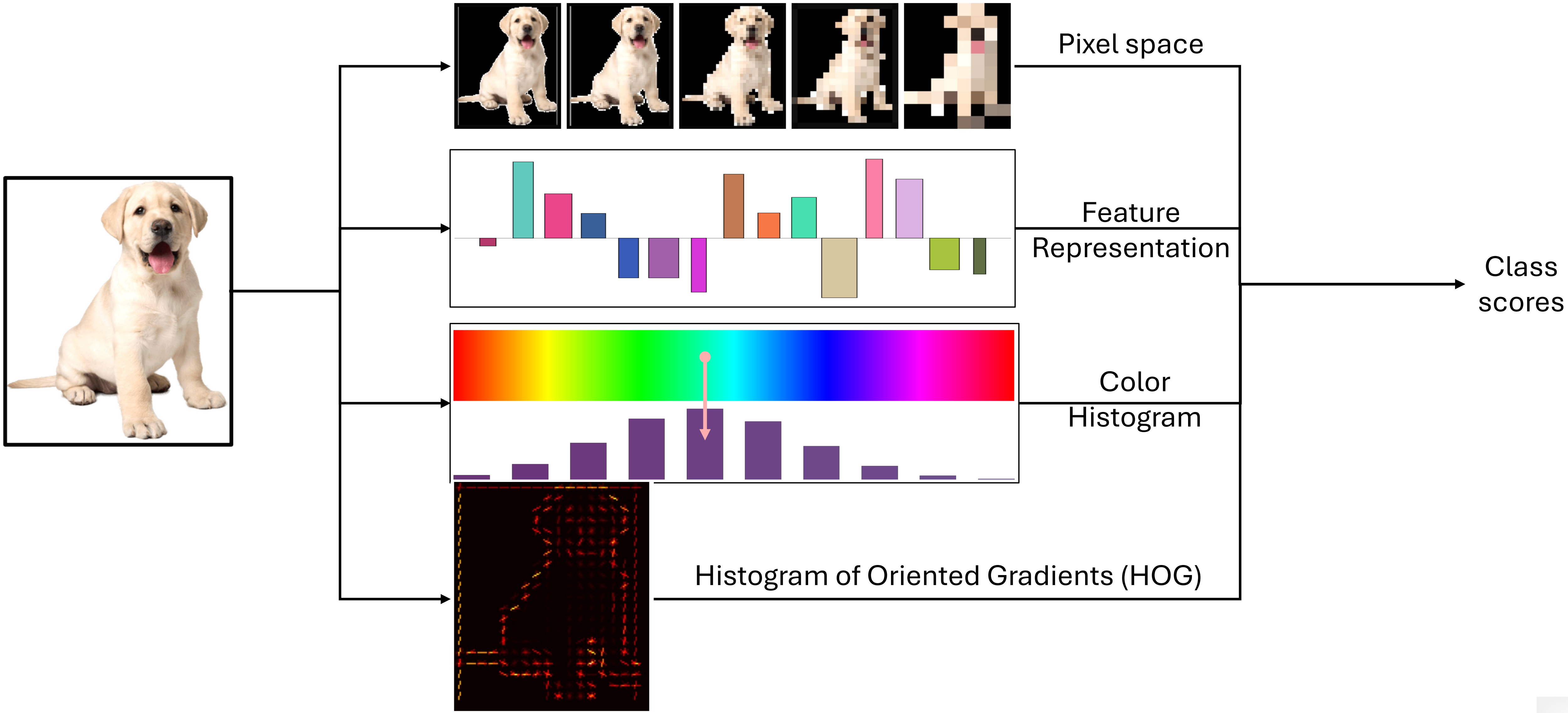


Image Features and CNN

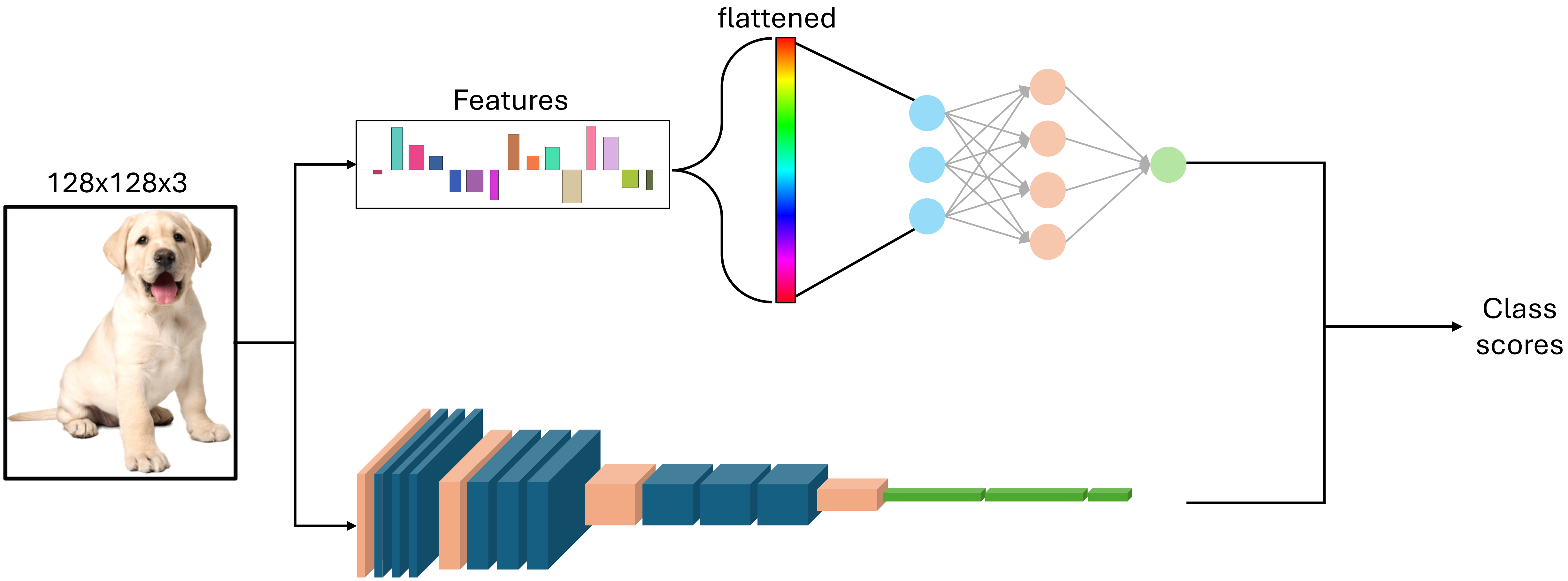


Image Features and CNN

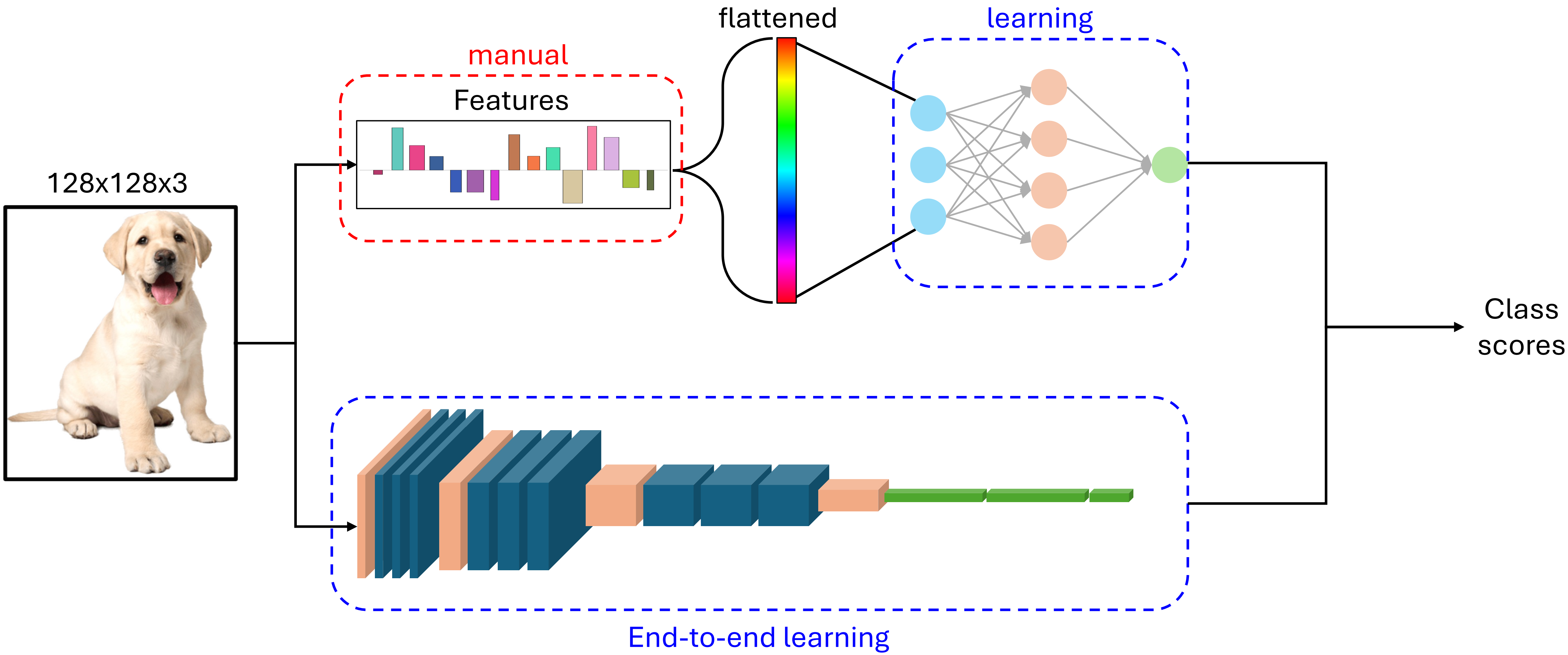
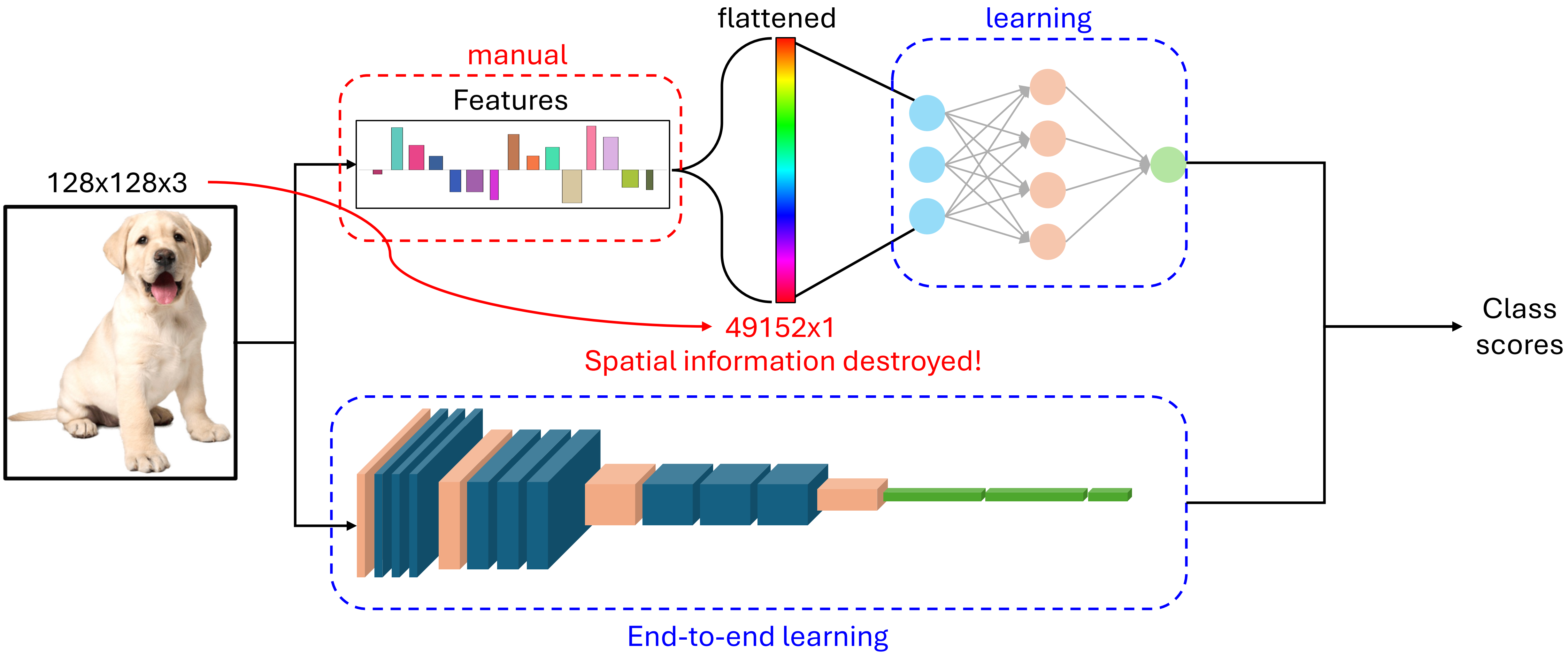
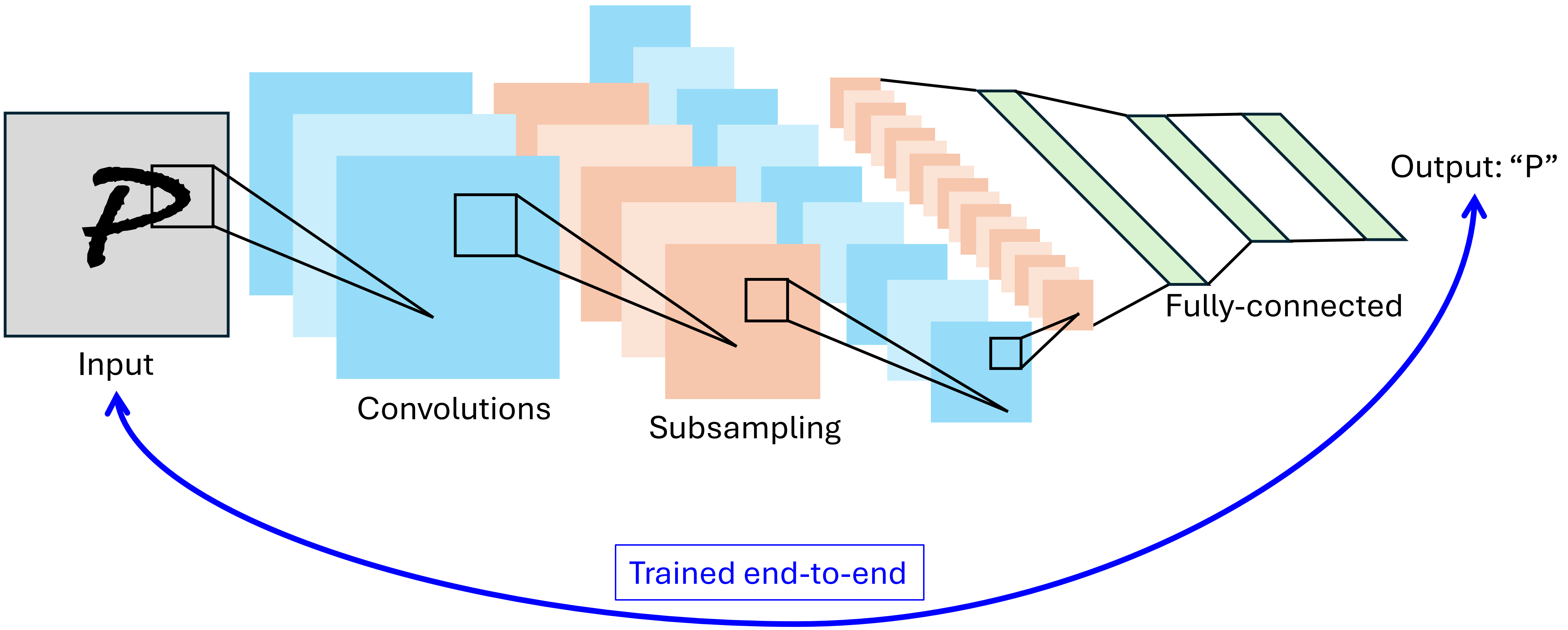


Image Features and CNN



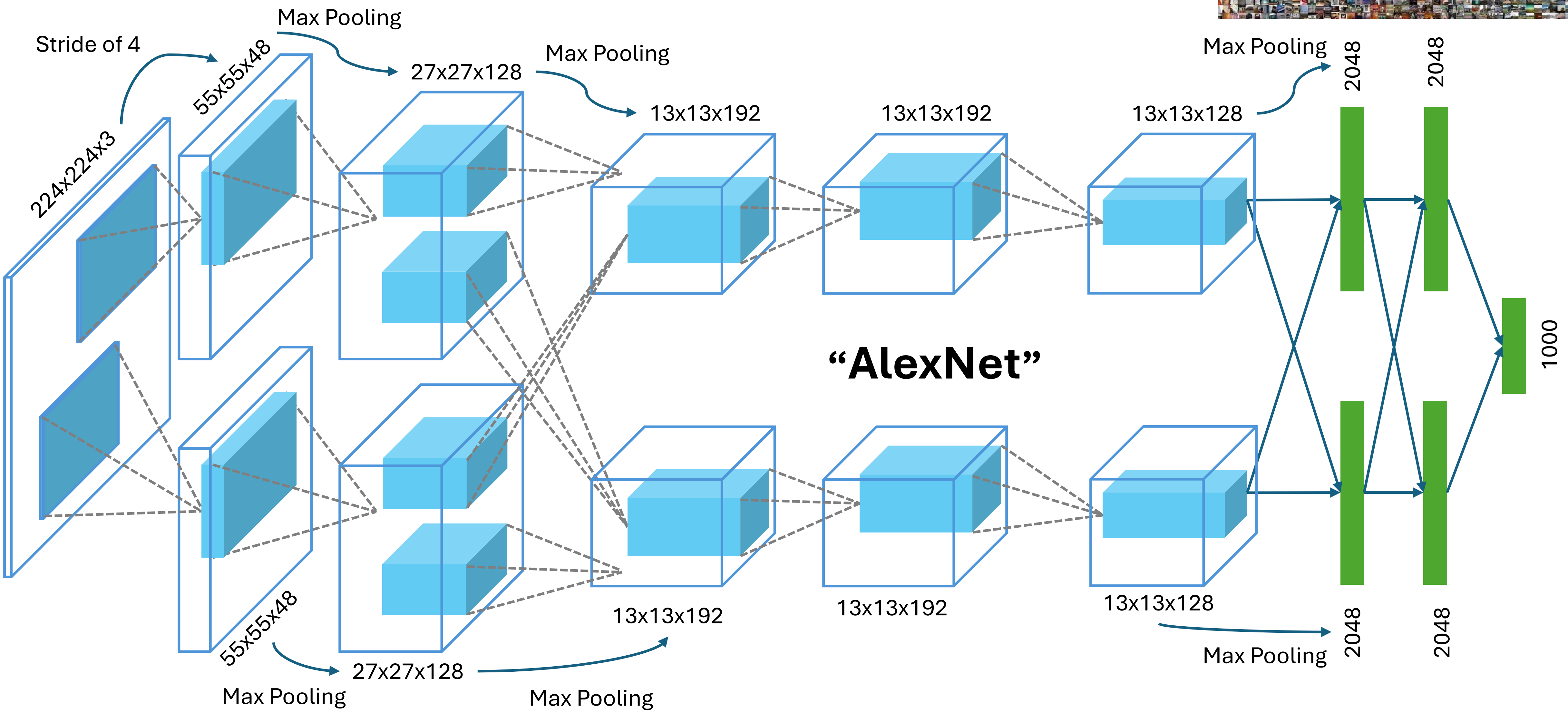
Convolutional Neural Networks (CNN)

History (Document Recognition, LeCun et al. 1998)



Convolutional Neural Networks (CNN)

History (ImageNet Classification, Krizhevsky et al. 2012)



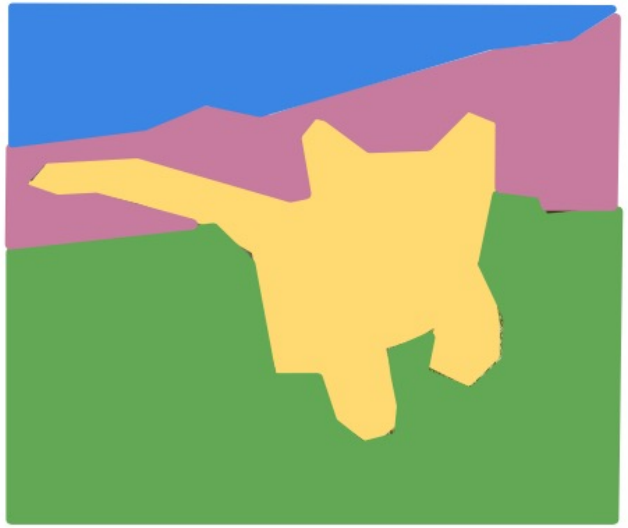
Convolutional Neural Networks (CNN)

2012-2020 CNNs dominated computer vision: ***YOLO, SegNet, PixNet, VGG, ResNet, ...***

Classification

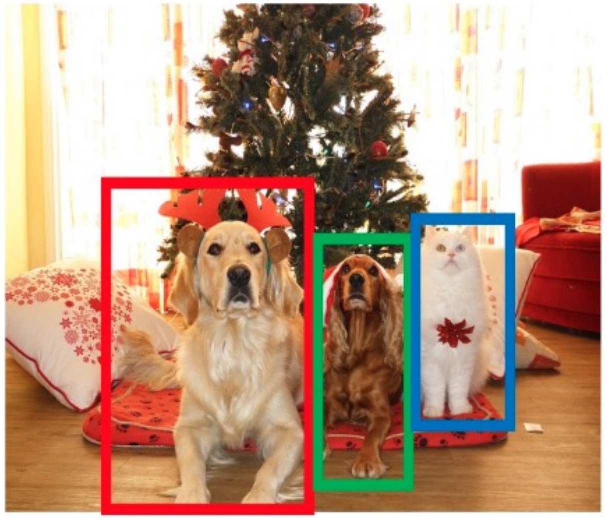
CAT

Semantic Segmentation



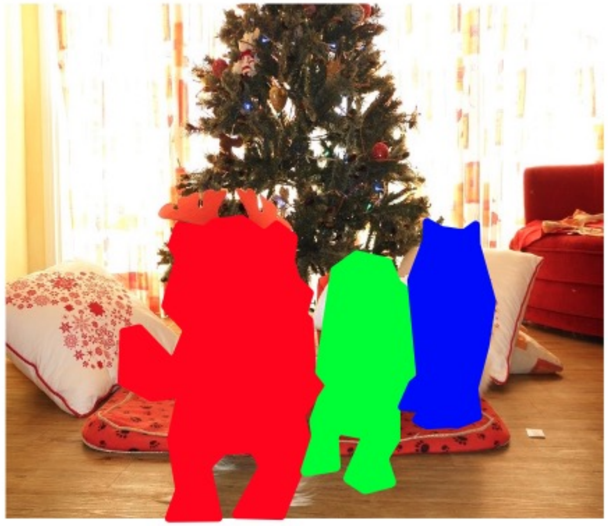
GRASS, CAT, TREE, SKY

Object Detection



DOG, DOG, CAT

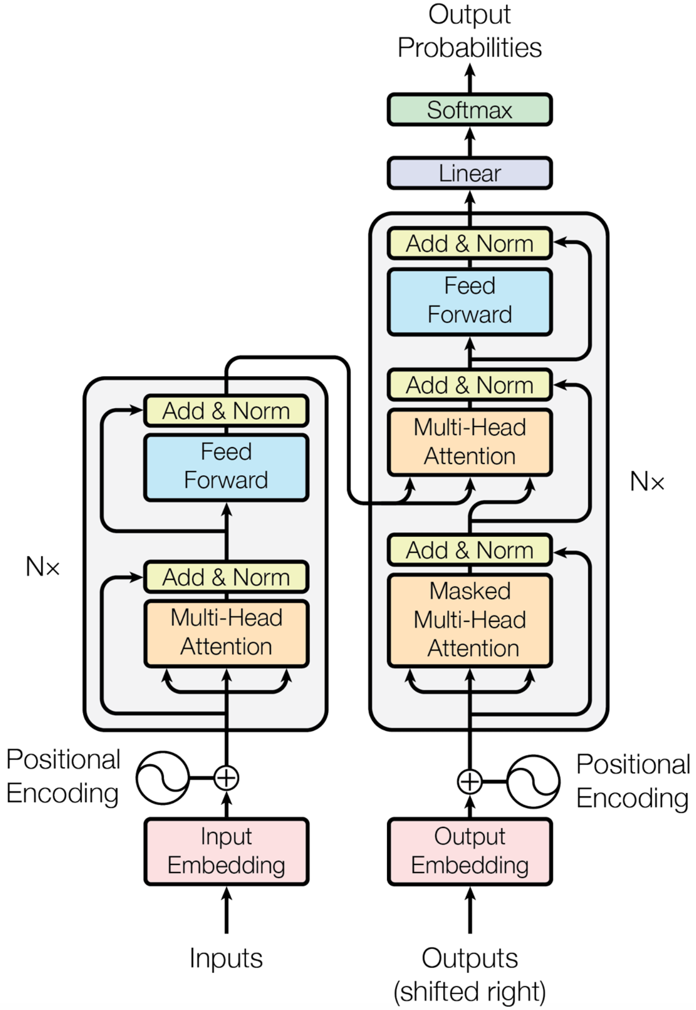
Instance Segmentation



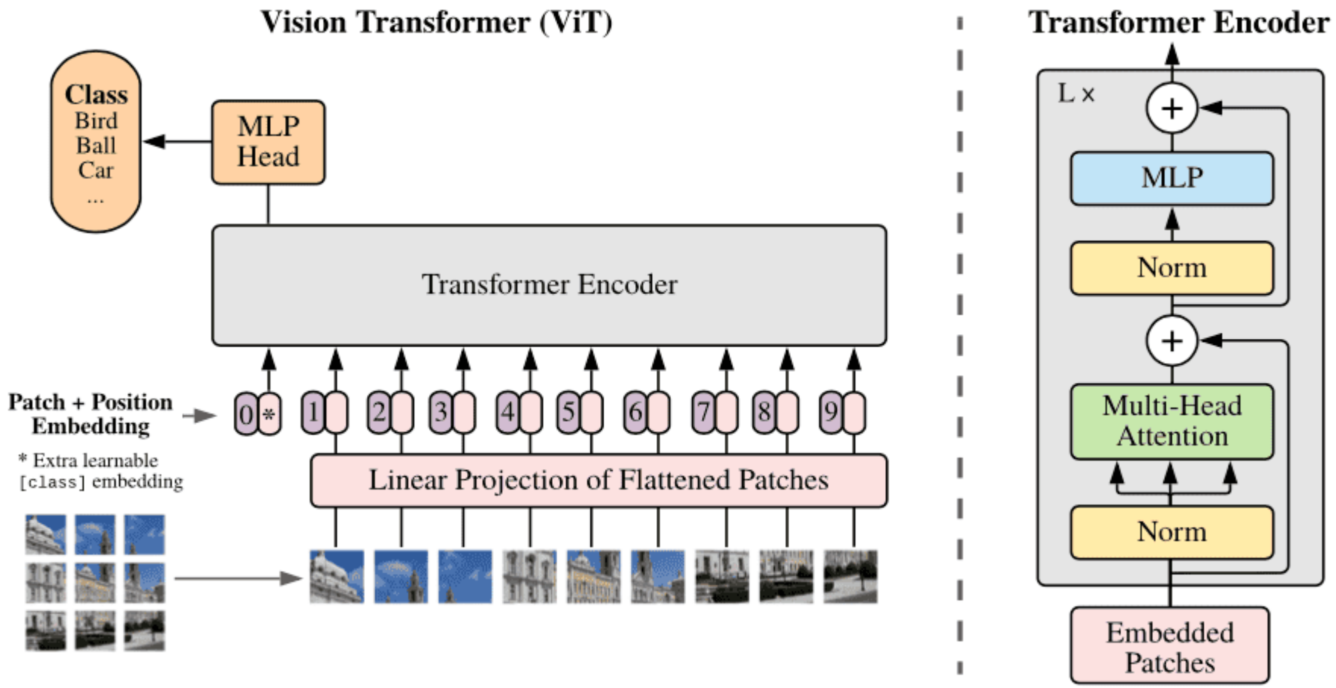
DOG, DOG, CAT

Transformers *2021 – present*

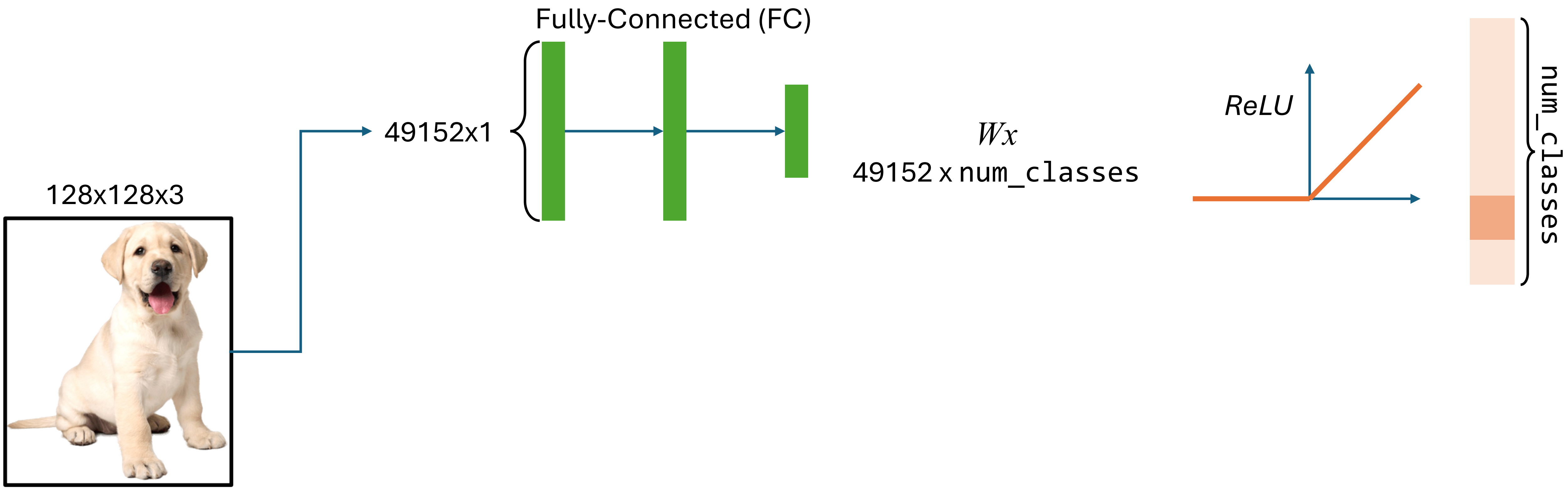
Large Language Models (LLM)
Vaswani et al. NIPS 2017



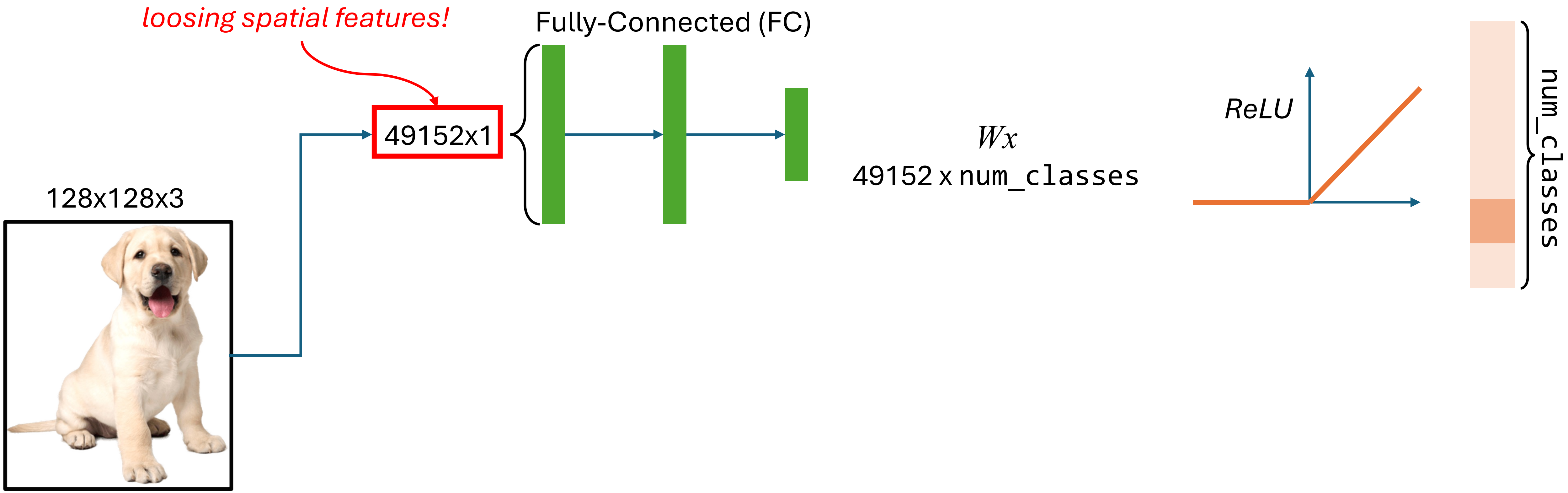
Vision Transformers (ViT)
Dosovitskiy et al. ICLR 2021



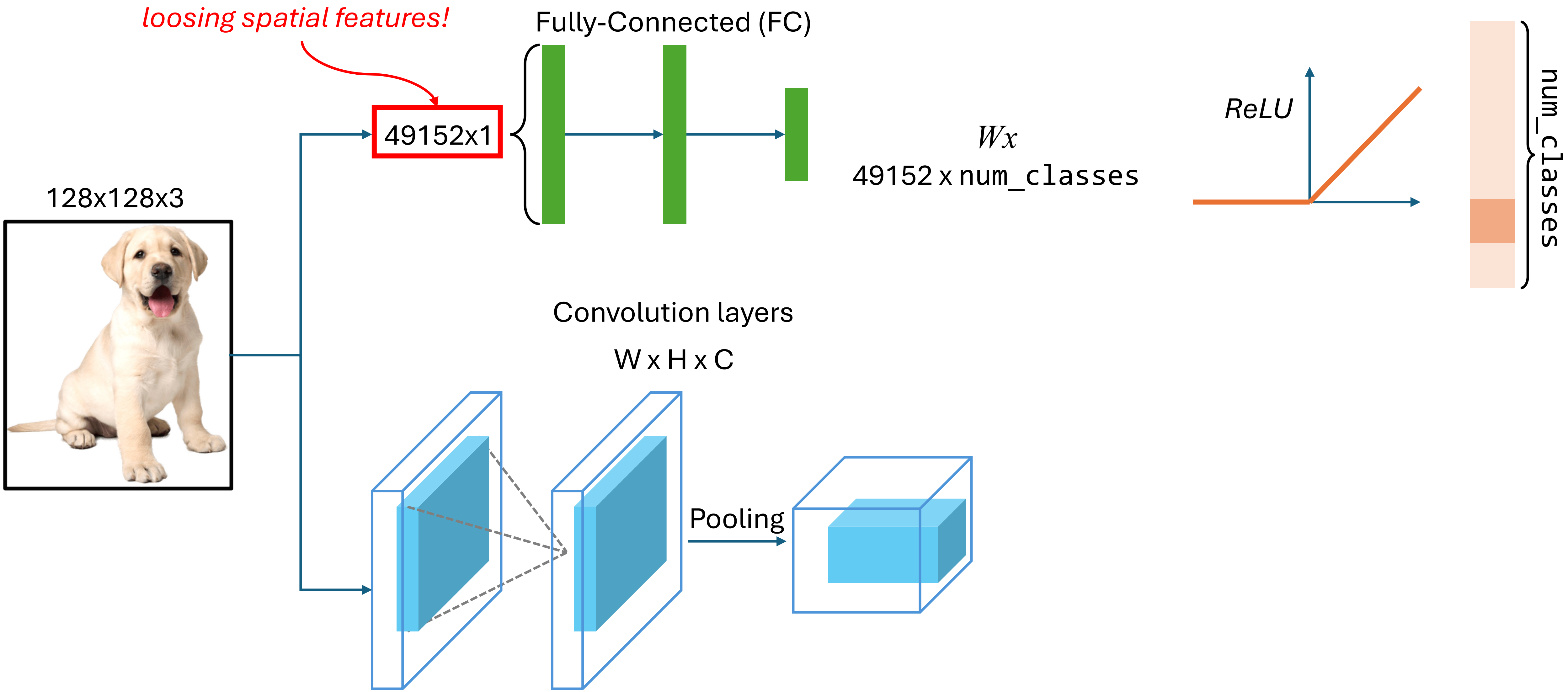
Convolutional Neural Networks (CNN)



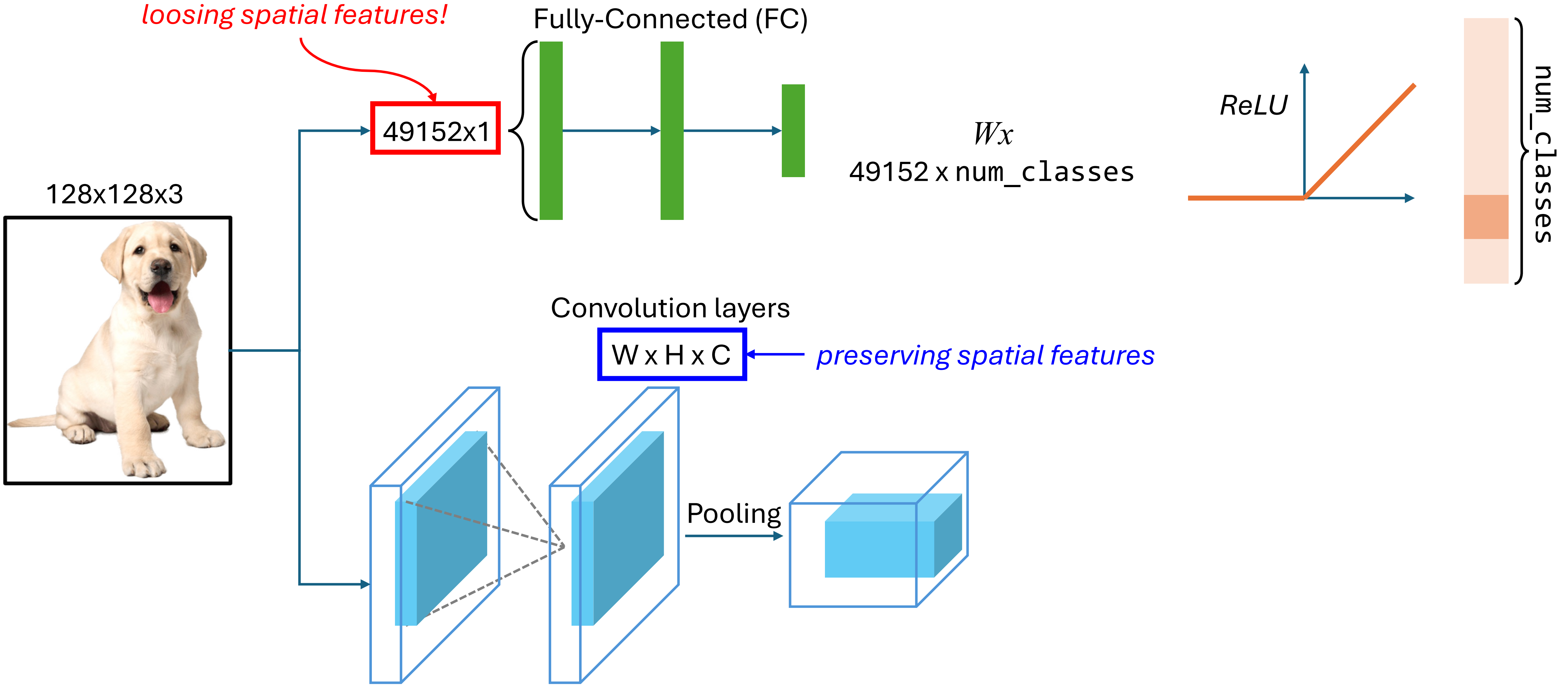
Convolutional Neural Networks (CNN)



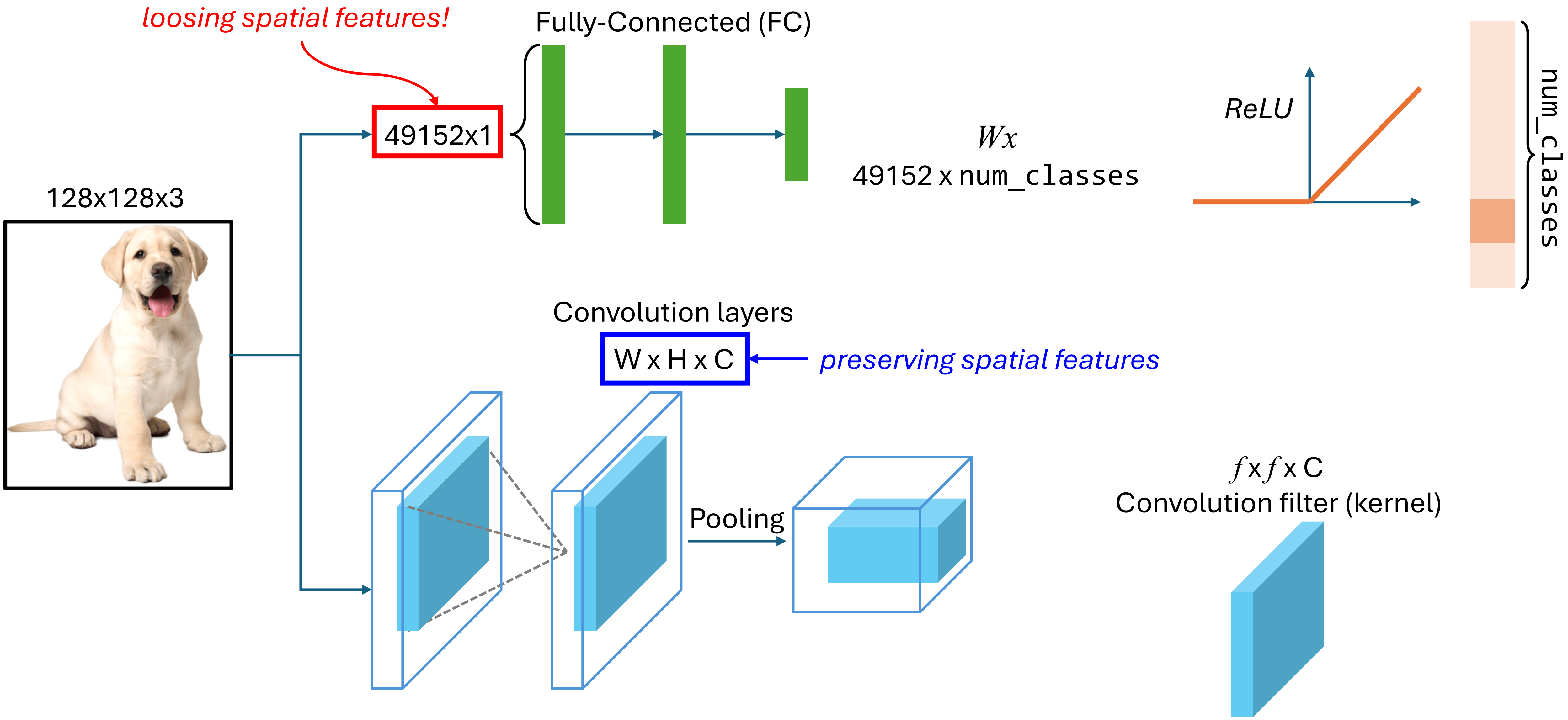
Convolutional Neural Networks (CNN)



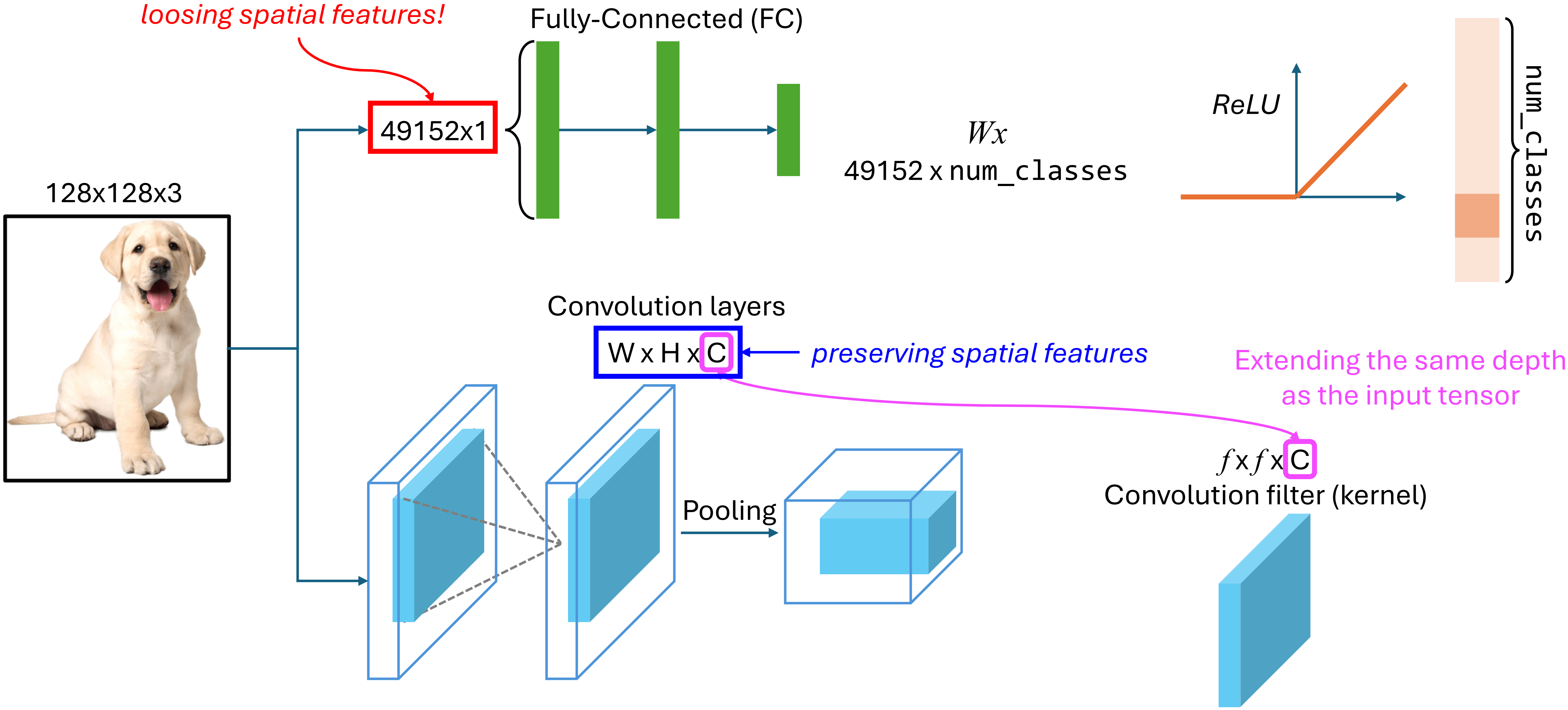
Convolutional Neural Networks (CNN)



Convolutional Neural Networks (CNN)



Convolutional Neural Networks (CNN)

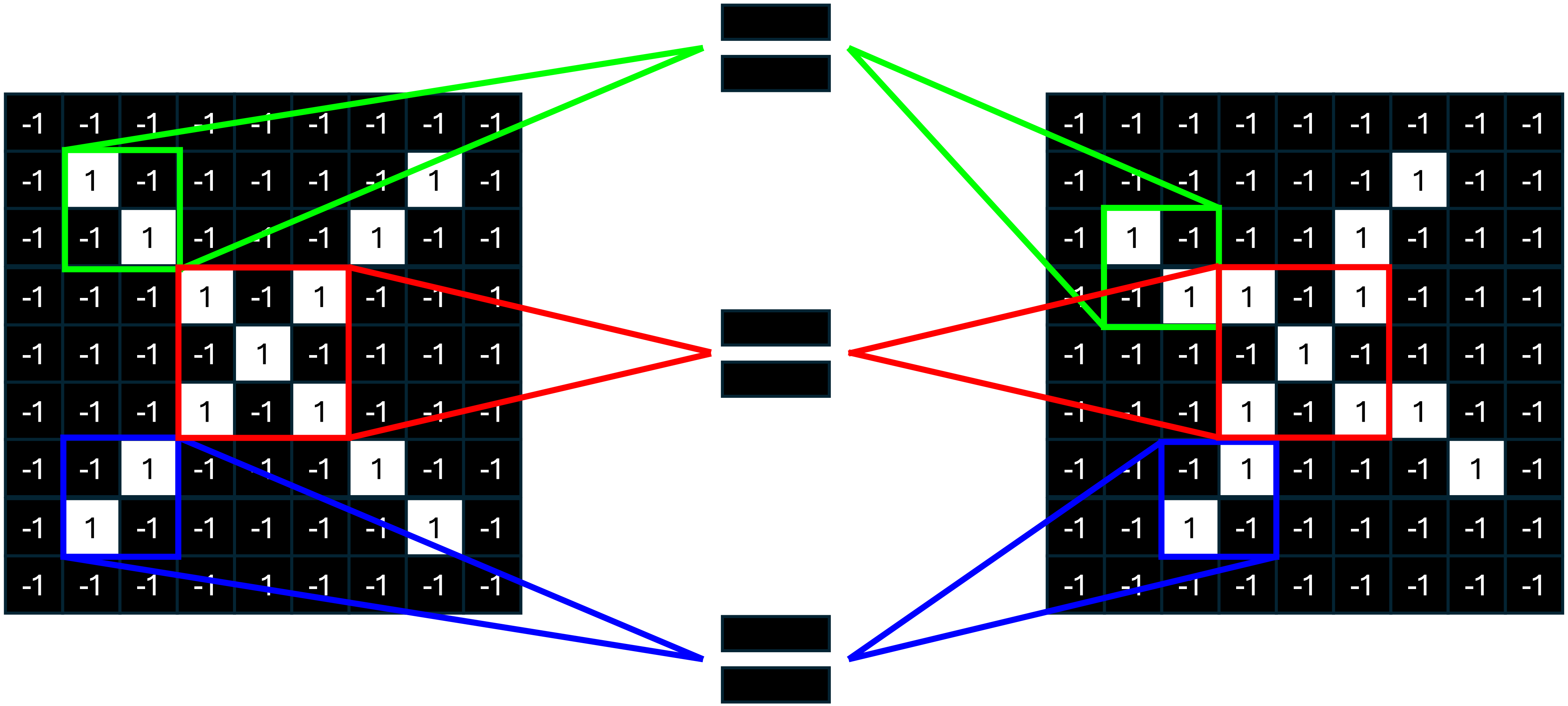


-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

$$\stackrel{?}{=}$$

-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	1	-1	-1	-1
-1	-1	1	1	-1	1	-1	-1	-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	1	-1	-1
-1	-1	-1	1	-1	-1	-1	1	-1
-1	-1	1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1

Convolutional Neural Networks (CNN)



Convolutional Neural Networks (CNN)

Looking for ***patterns***

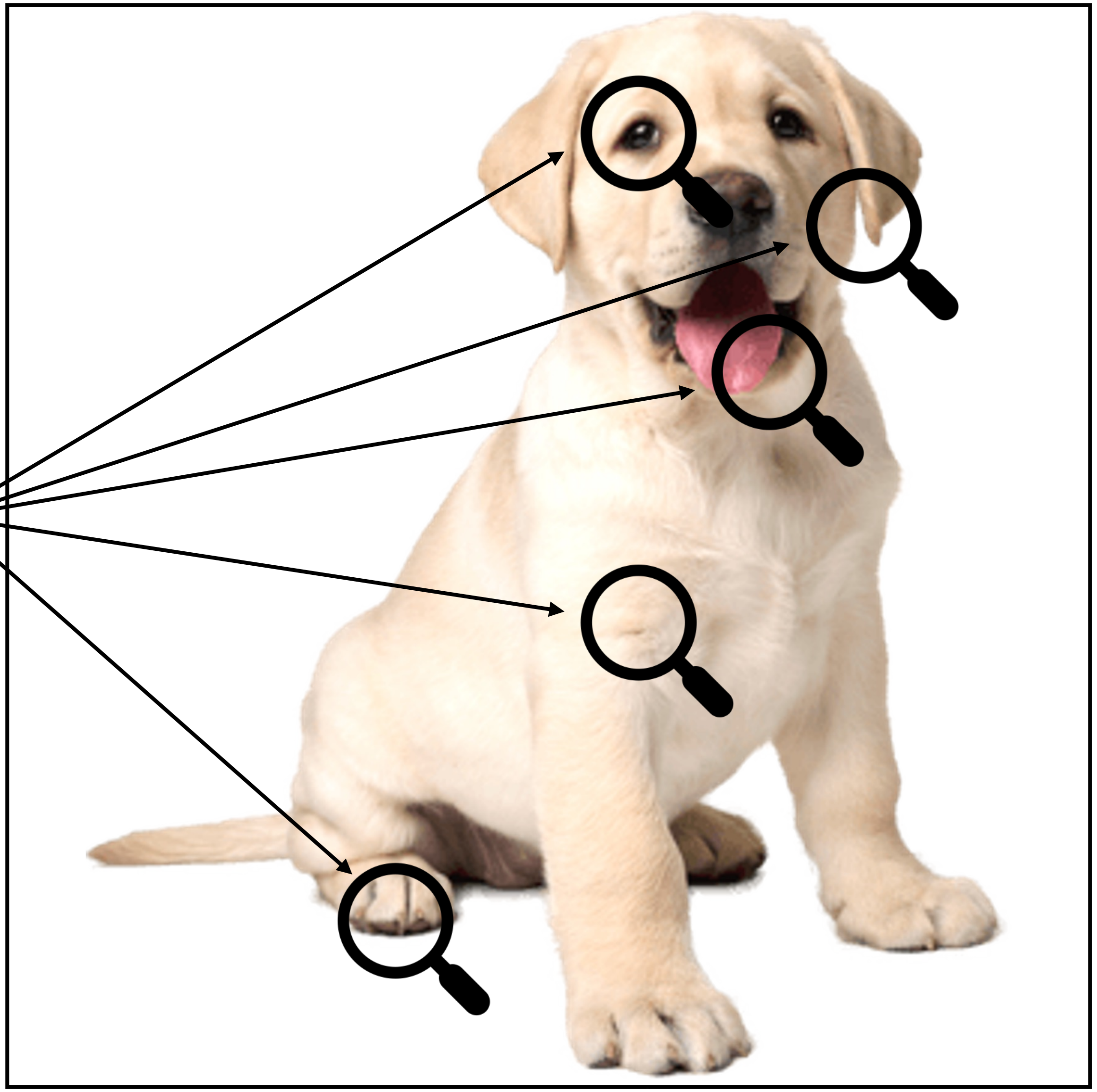


Image Classification

Quiz time!

1. Which of the following best describes a *linear model* for image classification?
   A) A model that applies a nonlinear kernel to image pixels
   B) A model that computes output as a weighted sum of input features
   C) A model that uses hierarchical feature extraction
   D) A model that randomly assigns labels

2. The weight vector (w) in a linear classifier determines:
   A) The bias of the dataset
   B) The importance of each input feature for classification
   C) The number of classes
   D) The learning rate of the model

3. A key limitation of linear models in image classification is:
   A) They require large datasets
   B) They cannot capture complex, nonlinear relationships in image data
   C) They are too computationally expensive
   D) They always overfit

Image Classification

Quiz time!

1. Which of the following best describes a *linear model* for image classification?
   - A) A model that applies a nonlinear kernel to image pixels
   - ✅ B) A model that computes output as a weighted sum of input features
   - C) A model that uses hierarchical feature extraction
   - D) A model that randomly assigns labels

2. The weight vector (w) in a linear classifier determines:
   - A) The bias of the dataset
   - ✅ B) The importance of each input feature for classification
   - C) The number of classes
   - D) The learning rate of the model

3. A key limitation of linear models in image classification is:
   - A) They require large datasets
   - B) They cannot capture complex, nonlinear relationships in image data
   - ✅ C) They are too computationally expensive
   - D) They always overfit

Image Classification

(Q&A)

Parham Kebria
parhamkebria@ieee.org