MAI4CAREU
Master programmes in Artificial Intelligence 4 Careers in Europe

University of Cyprus

University of Cyprus

MAI645 - Machine Learning for Graphics and Computer Vision

Andreas Aristidou, PhD

Spring Semester 2024

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Let's get to know each other: *The team*

Andreas Aristidou
Assistant Professor

Office: FST01, Room B113

Office hours: Only after appointment

email: andarist@ucy.ac.cy

Research Interests:

Machine Learning, Deep Learning and its applications in Computer Graphics and Character Animation, Virtual/Augmented Reality, Digital Heritage

https://www.cs.ucy.ac.cy/~andarist

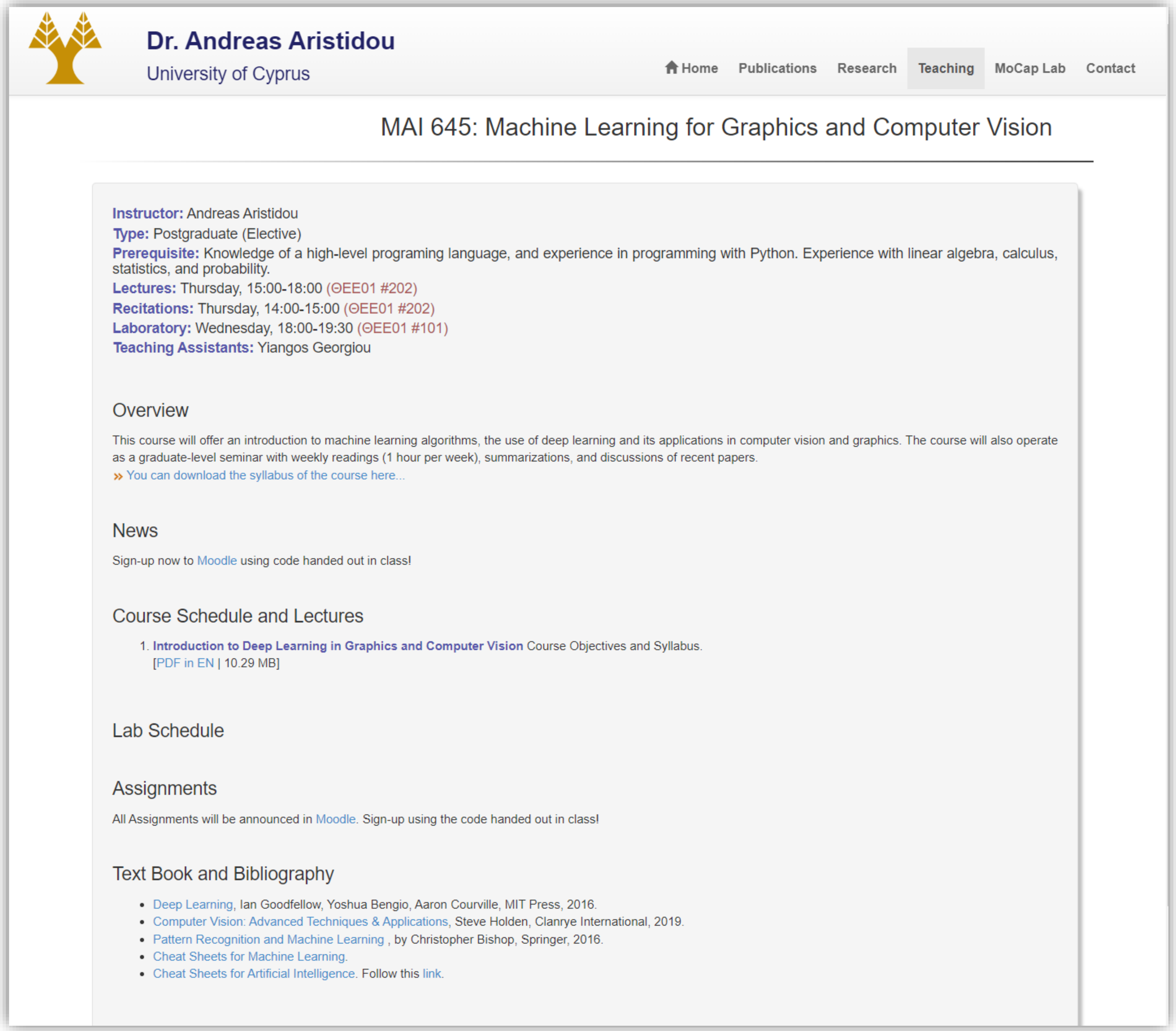
Yiangos Georgiou
Teaching Assistant

Office hours: Only after appointment

email: georgiou.yiangos@ucy.ac.cy

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Let's get to know each other: *Module website*

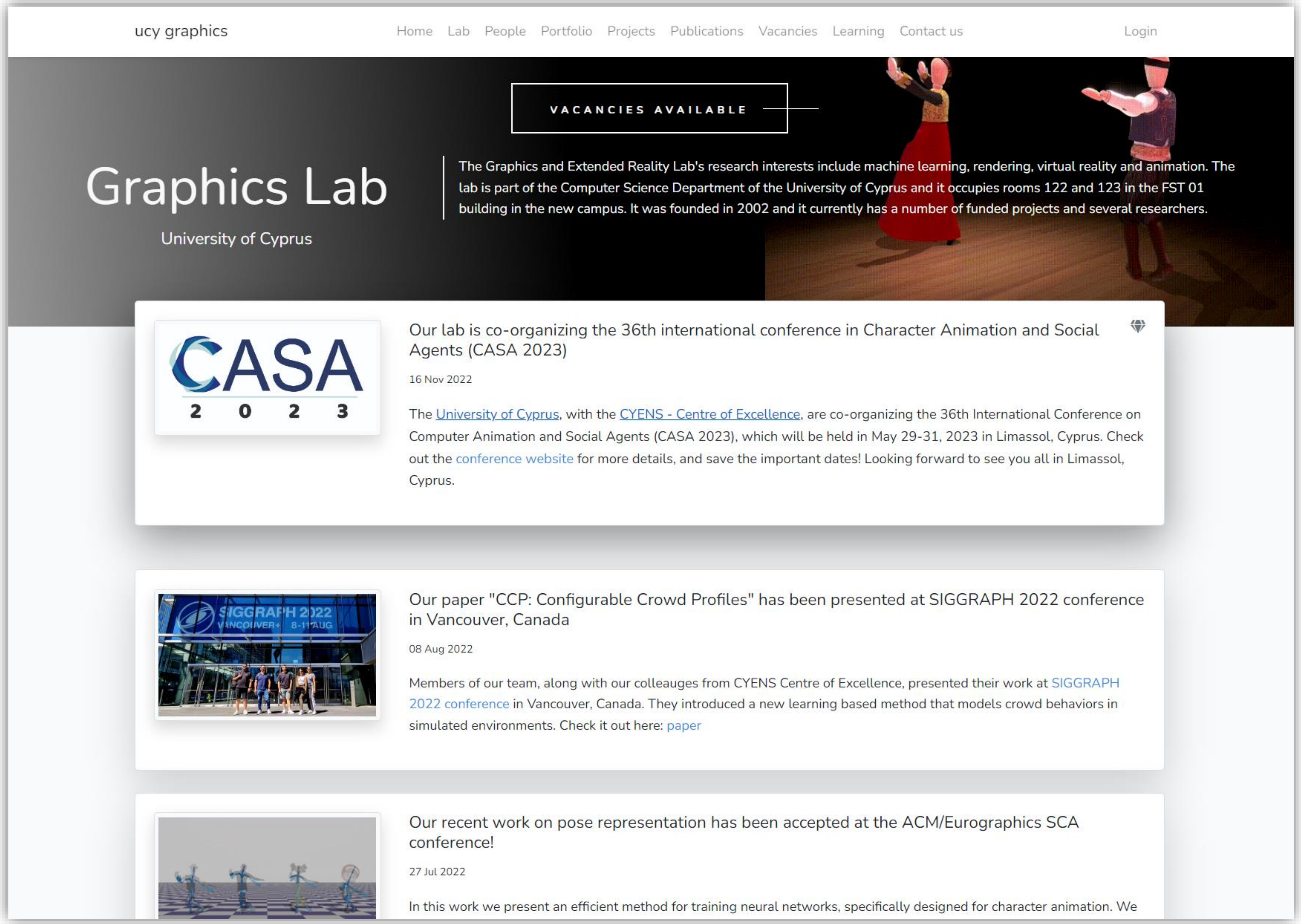


The **MAI 645** course webpage:

- All notes.
- Course Schedule
- Syllabus

Website:
https://www.cs.ucy.ac.cy/courses/MAI645/

Let's get to know each other: *Virtual & Extended Reality Lab*



The **Graphics and Extended Reality Lab** at the University of Cyprus, part of the Computer Science Department, conducts research in areas such as machine learning, rendering, virtual reality and animation.

Founded in 2002, the lab is located in rooms 122 and 123 of the FST 01 building on the university's new campus, and is staffed by two faculty members and twelve research associates. It also has several active, funded projects.

Website: https://graphics.cs.ucy.ac.cy/

MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Course purpose and objectives

This course will offer an introduction to machine learning algorithms focusing on computer vision and graphics, and the use of deep learning and its applications in the domain. The course will also operate as a graduate-level seminar with weekly readings, summarizations, and discussions of recent papers.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Learning outcomes

Participants will explore the latest developments in neural network research and deep learning models that are enabling highly accurate and intelligent computer vision and graphics systems. By the end of the course, participants will:

- Be familiar with fundamental concepts and applications in computer vision and graphics.
- Grasp the principles of state-of-the art deep neural networks.
- Gain knowledge of high-level vision tasks, such as object recognition, scene recognition, face detection and human motion categorization.
- Gain knowledge of high-level graphics tasks, such as composite image generation, style transfer, motion reconstruction, and motion synthesis.
- Develop practical skills necessary to build highly-accurate, advanced computer vision and graphics applications

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Syllabus (Computer Vision)

L02 Learning from Images: Deep learning for image classification and object detection.

L03 Learning from Videos: Deep learning for video classification.

L04 Feature Extraction: Deep learning for feature extraction and face recognition.

L05 Semantic Understanding: Deep learning for semantic segmentation, visualizing networks, impainting, saliency detection.

L06: Creative applications: Photo collections: style and enhancement, Ambiguity and style, style transfer.

L07: Vision->Graphics: Computer vision as inverse computer graphics, Novel image synthesis – compositional image generation.

Syllabus (Computer Graphics)

L08 From 2D to 3D: 3D meshes and point clouds.

L09 Inverse graphics in practice: Generation (Audio synthesis - Audio/2D/3D images and pixel processing: Load and stored images/3D data, audio in the coding environment they use. 3D labeling/classification

L10 Creative Applications: Generative networks, generating faces, landscapes, portraits, Sketches, denoising.

L11 Motion: Motion capture, character animation, and synthesis (style transfer, retargeting, control), deep reinforcement learning for physics-based animation and authoring.

L12 (neural) rendering, physics, materials, virtual reality.

L13 Advanced Topics in Deep Learning: Levels of supervision; Adversarial training, open problems.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Teaching methodology

Weekly lectures, discussions of practical examples and (unsupervised) lab activities where the active learning element is encouraged and supported. Weekly or bi-weekly recitation and discussion on recent papers, with students' presentations. Students would be strongly guided to view all topics presented and discussed with a critical eye.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Prerequisite

Knowledge of a high-level programing language, and experience in programming with Python.

Attended the MAI612 Machine Learning course would be very helpful.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Assessment

- Student paper presentations (15%);
- Programming assignments (35%);
- Final course project and presentation (50%)

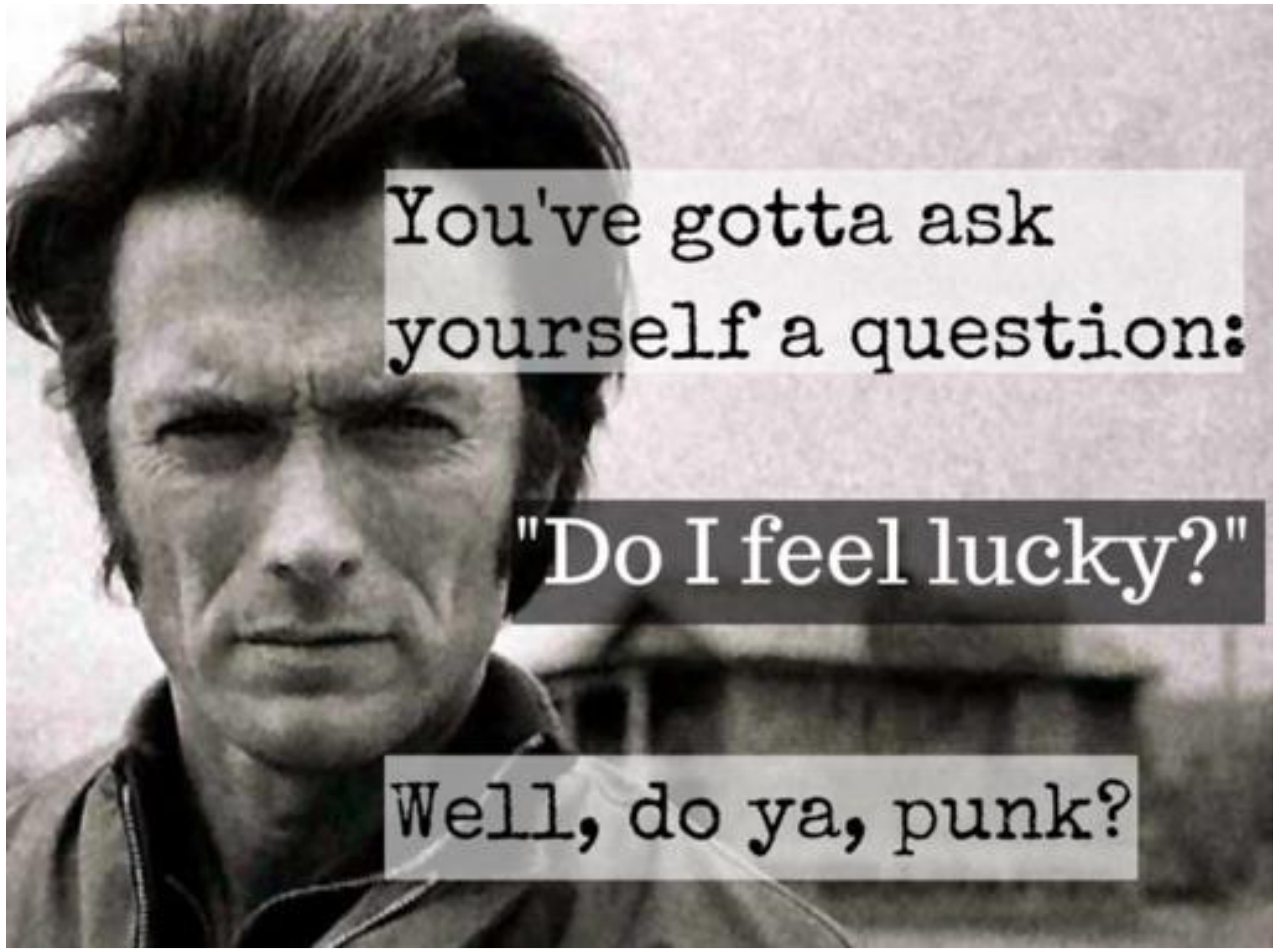
To qualify one must:

- Hand in all assignments
- Achieve at least 50% overall

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

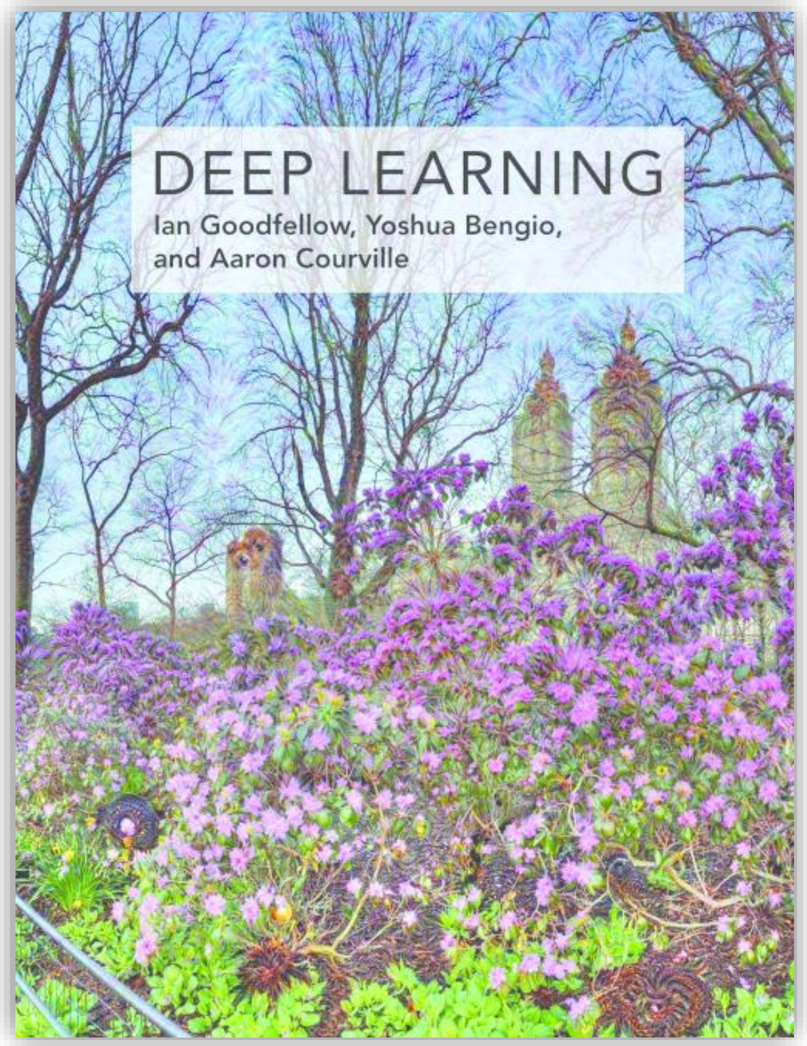
MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)



Let's keep it simple: if you are caught cheating, you will get a **zero** for the entire course (not just the assignment).

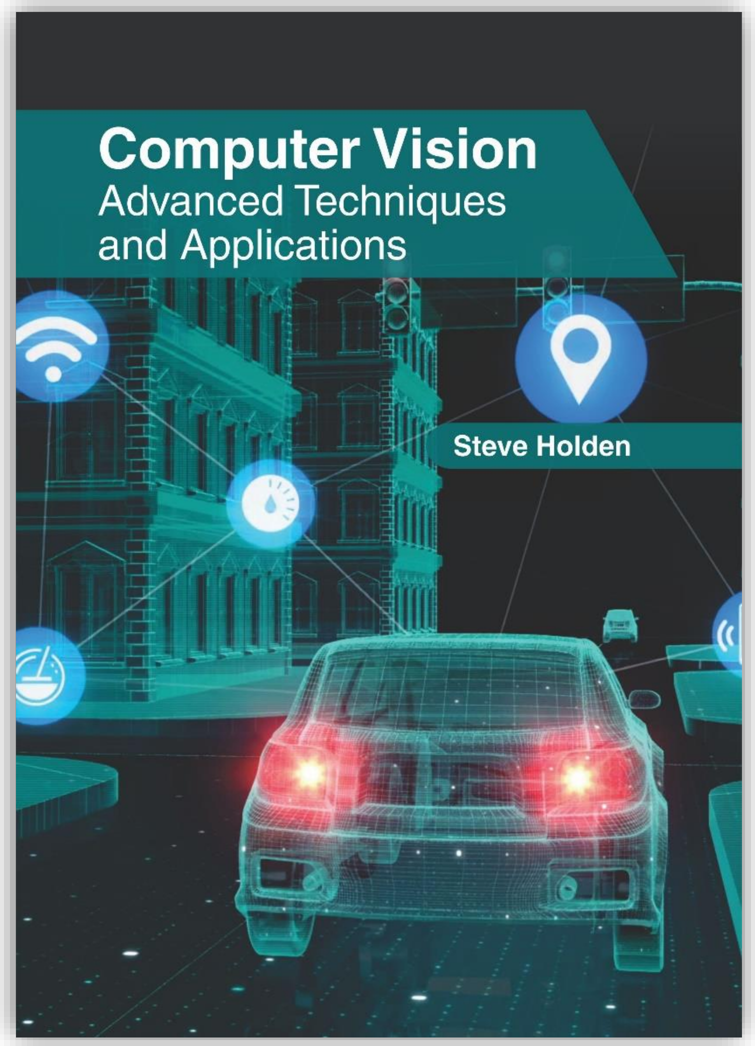
MAI645 - Machine Learning for Graphics and Computer Vision (8 ECTS)

Bibliography



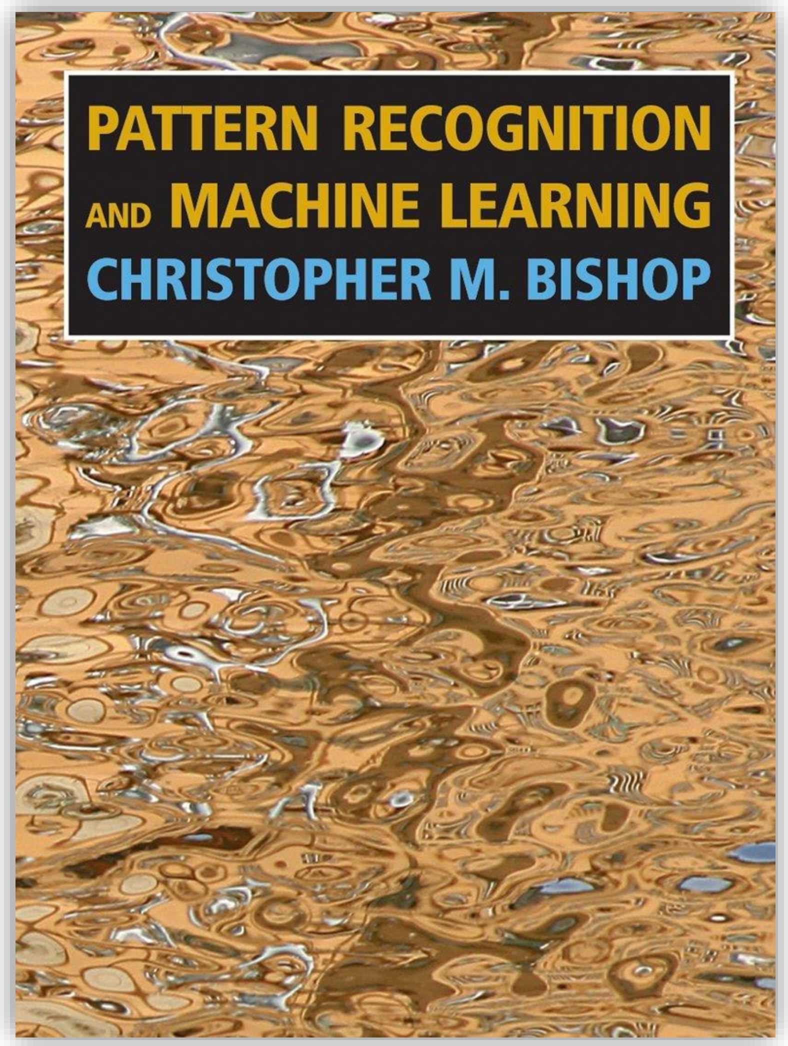
Deep Learning

by Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press, 2016



Computer Vision: Advanced Techniques & Applications

by Steve Holden, Clanrye International, 2019



Pattern Recognition and Machine Learning

by Christopher Bishop, Springer, 2016

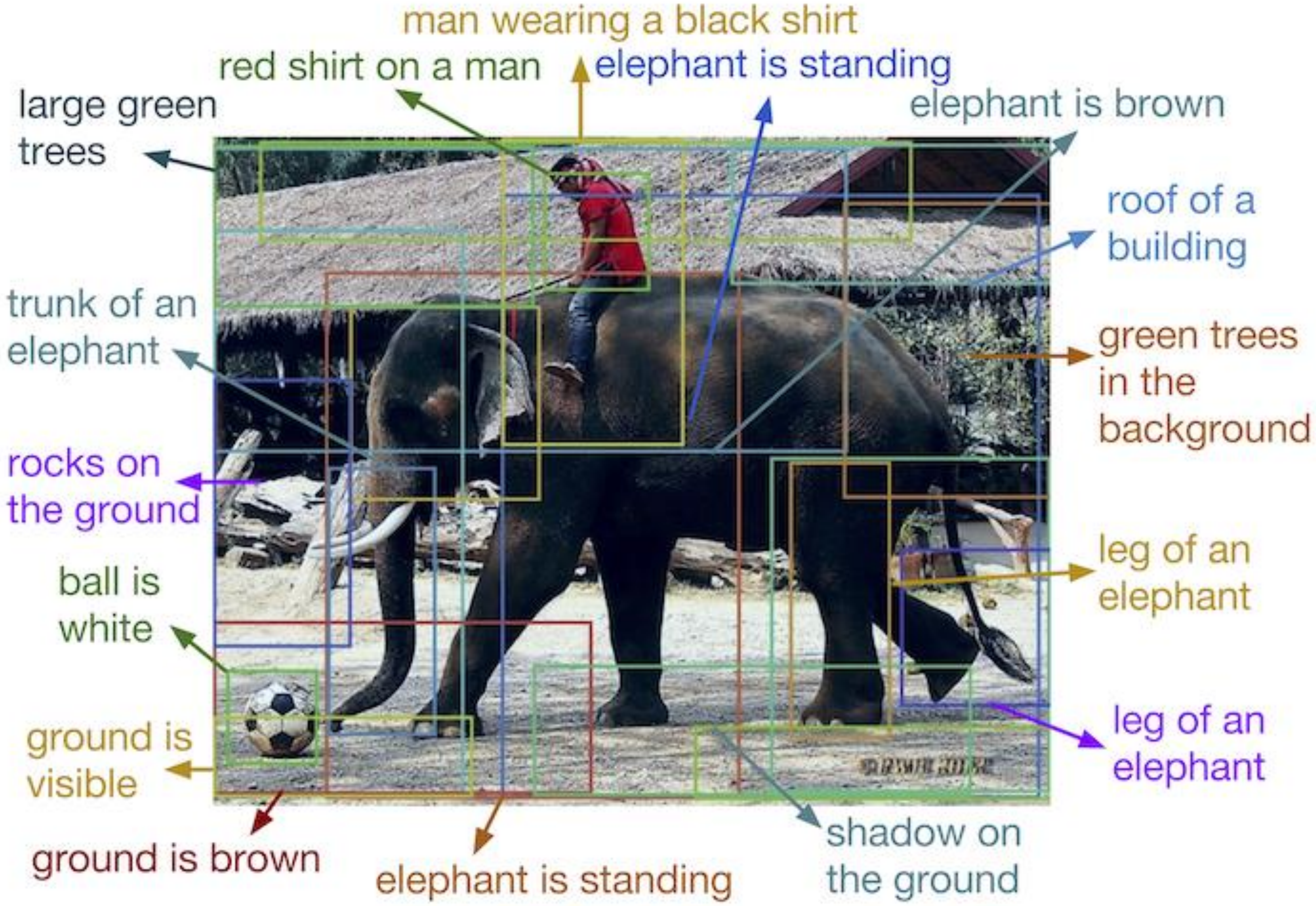
Computer Vision: *Introduction*

com•put•er vi•sion /kəmˈpyo͞odər vɪʒn/ *n.* seeks to understand and automate tasks that the **human visual system** can do.

It is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos.

It is the **center block** of robotic intelligence!

Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding.



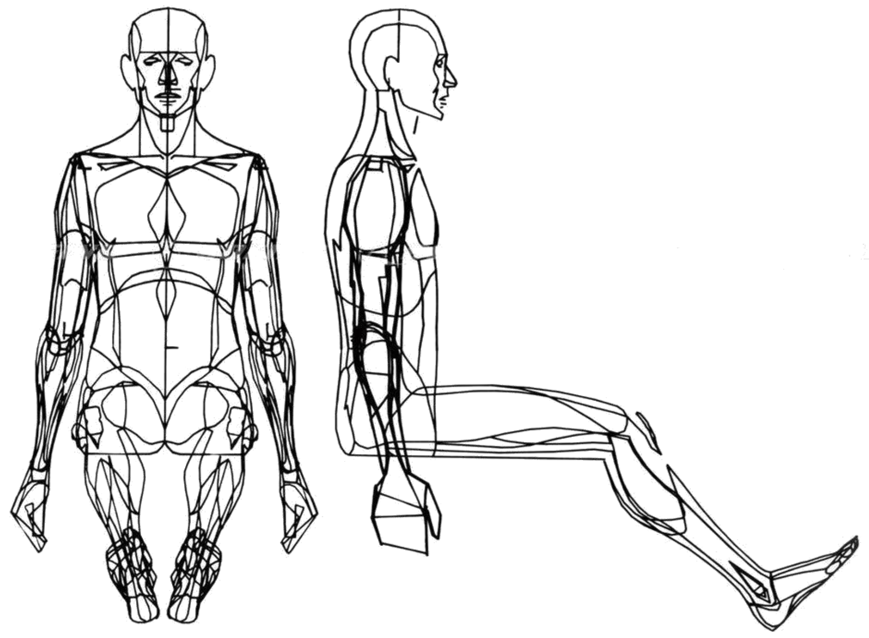
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Computer Graphics: *Introduction*

com•put•er graph•ics /kəmˈpyo͞odər ˈgrafiks/ *n.* The use of computers to synthesize and manipulate **visual information**.

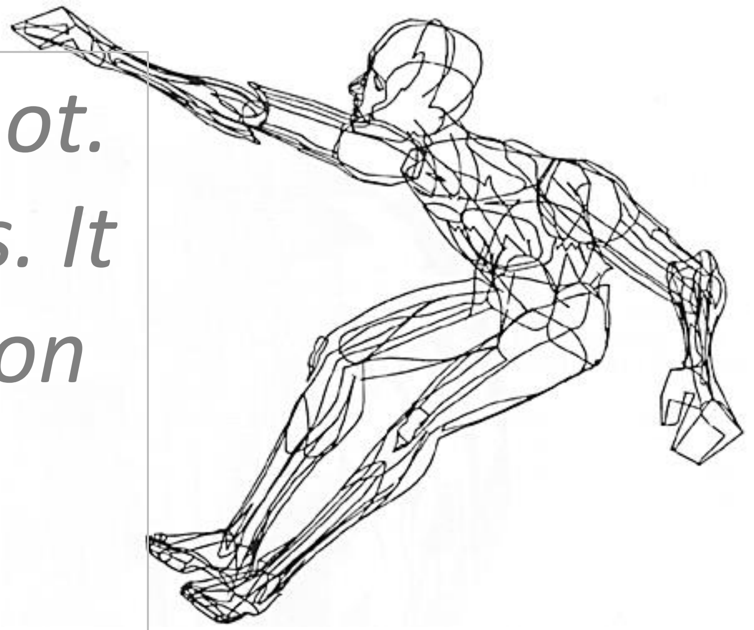
The creation, storage and manipulation of models and images. Such models come from diverse and expanding set of fields including physical, biological, mathematical, artistic, and conceptual/abstract structures.

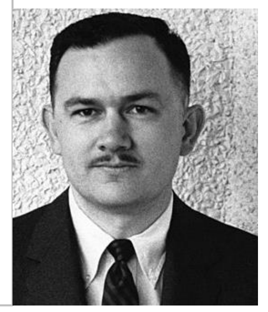


Perhaps the best way to define computer graphics is to find out what it is not. It is not a machine. It is not a computer, nor a group of computer programs. It is not the know-how of a graphic designer, a programmer, a writer, a motion picture specialist, or a reproduction specialist.

Computer graphics is all these – a consciously managed and documented technology directed toward ***communicating information*** *accurately and descriptively."*

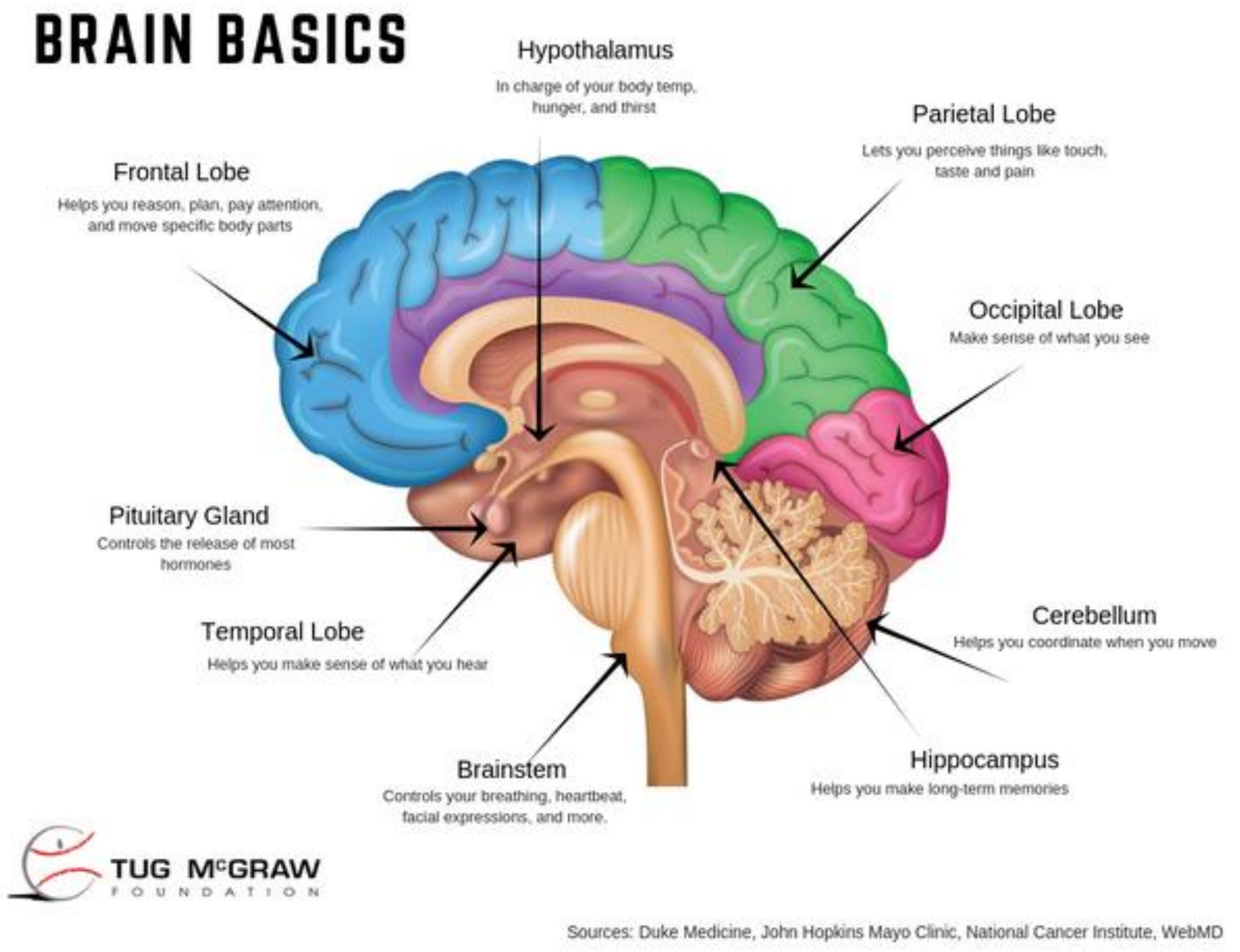
Computer Graphics, by William A. Fetter, 1966





Computer Vision: *Why Visual Information?*

- About 30% of brain dedicated to visual processing...



- ...eyes are highest-bandwidth port into the head!

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Visual Computing

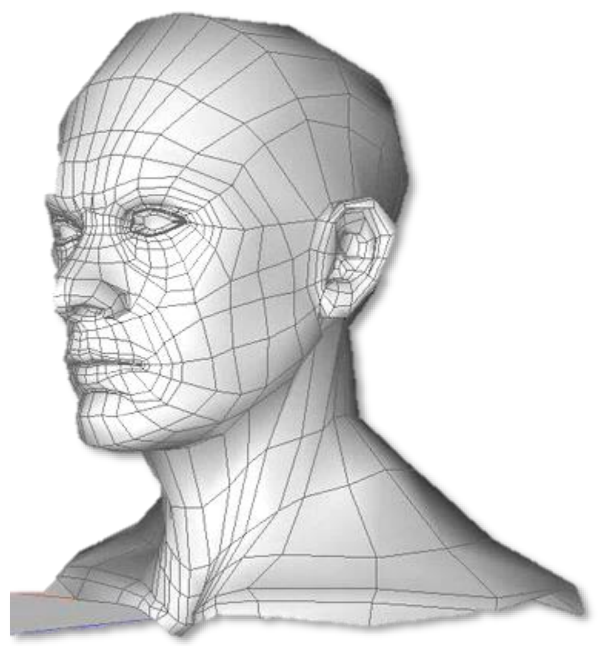
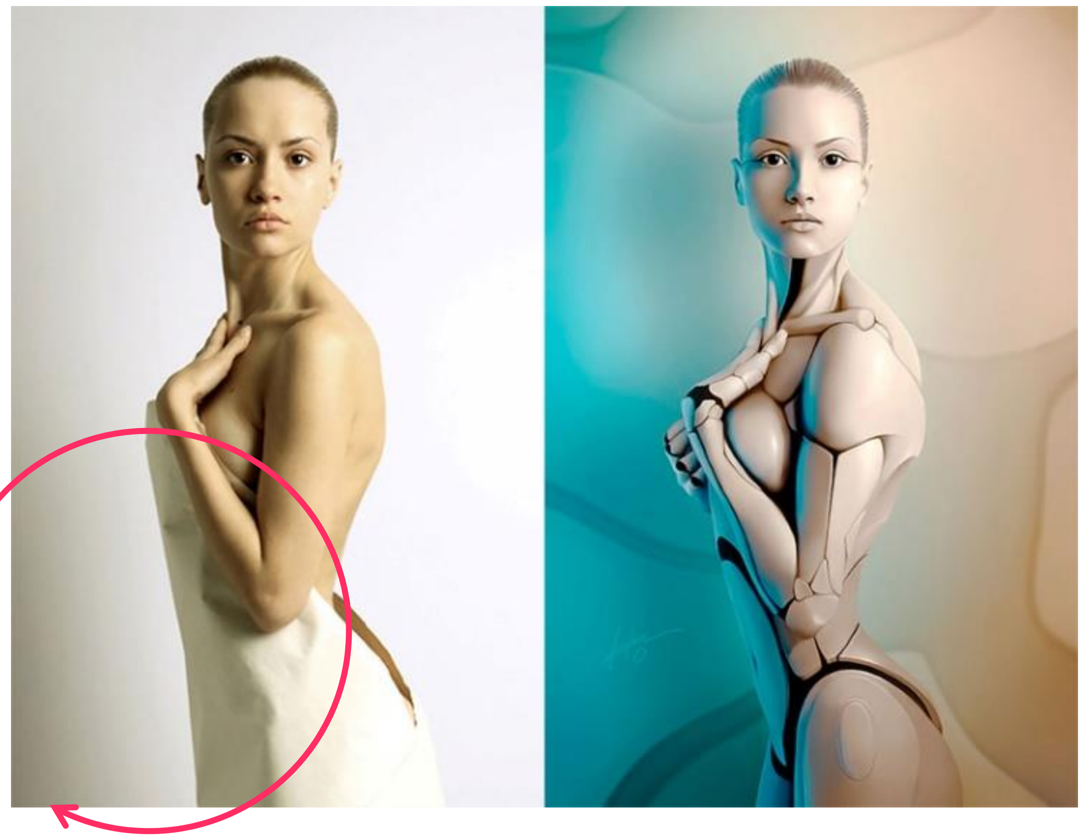


Image Processing



Objects (Real, imaginary, mathematical)

Geometric Modeling →

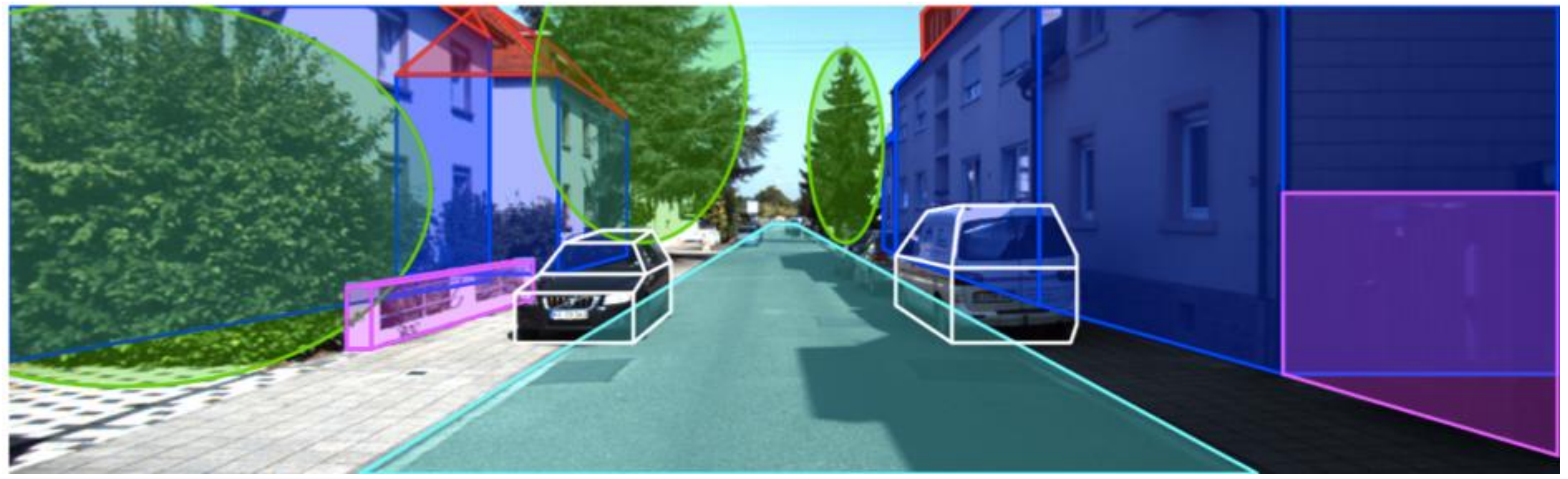
← *Fabrication*

3D Scene (Model)

Computer Graphics →

← *Computer Vision*

Image



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Problems *in computer vision and graphics*

How do we gain insights from the visual data?

- **Visualization:** One of the simplest and most effective ways to gain insights from visual data is to create visual representations of the data, such as plots, charts, and maps. This allows you to see patterns, trends, and outliers in the data that might not be apparent from looking at the raw data. → **Manually**
- **Image Processing:** For image data, techniques such as image segmentation, feature extraction, and pattern recognition can be used to extract meaningful information from the images. → **Automatic, handcrafted feature extraction, supervised**
- **Machine Learning:** Machine learning algorithms can be used to analyze visual data and identify patterns, classify images, detect objects, or generate new images. Some examples of machine learning techniques applied to visual data are convolutional neural networks (CNNs), object detection, and image recognition. → **Automatic, supervised/unsupervised**
- **Human-in-the-loop:** Human interpretation is still important when it comes to visual data, and it's always a good idea to have a human expert validate or interpret the results, especially when dealing with complex or novel problems.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Hubel and Wiesel

- David Hubel and Torsten Wiesel were neurobiologists from Harvard Medical School
- They conducted an experiment that revealed several secrets of the human vision system. In particular, the experiment was conducted in the 1950s and 1960s, and it involved recording the activity of individual neurons in the primary visual cortex of cats and monkeys. They used a slide projector to show specific patterns and noted that specific patterns stimulated activity in specific parts of the brain. They systematically created a map of the visual cortex with these experiments.
- This experiment was very important in the field of computer vision because it provided crucial insights into the way the brain processes visual information.
- The researchers found that these neurons responded selectively to specific features in the visual scene, such as edges and lines, and that they were organized into a complex, hierarchical network.
- This research laid the foundation for the development of models of visual processing and computer vision
- Won 2 Nobel prizes

Major Results: *Visual cortex cells are sensitive to the orientation of edges but insensitive to their position*

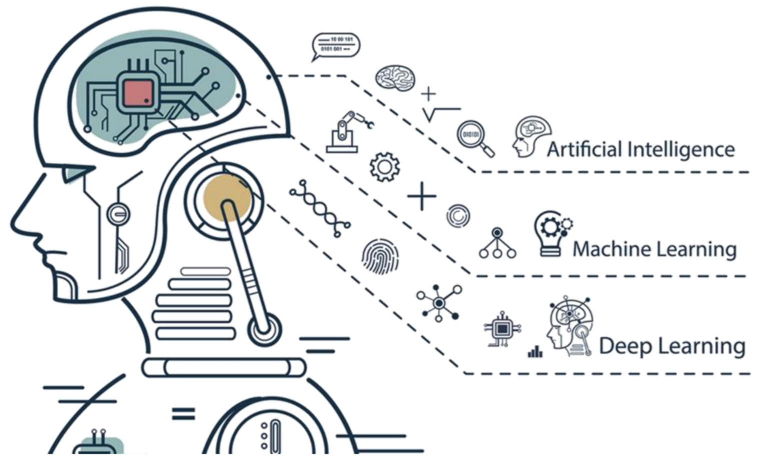
Few Decades Later…

Mathematics

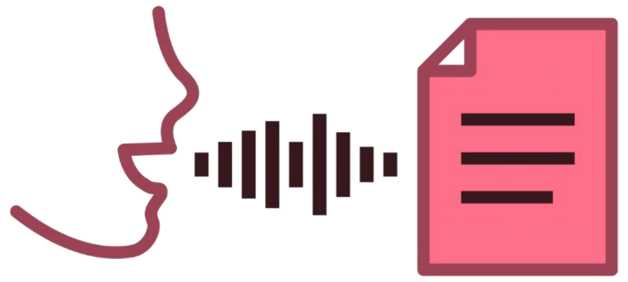
Engineering

Computer science

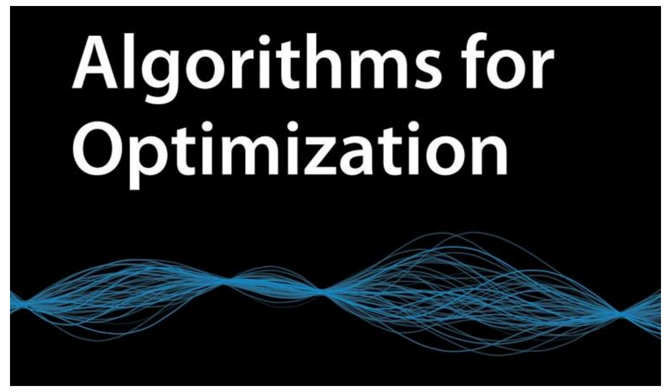
Robotics



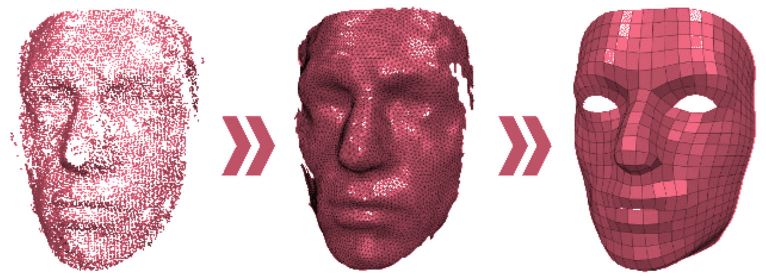
Artificial Intelligence



NLP Speech



Algorithms Optimization



Optics Image processing

Computer Vision

Neuroscience

Physics

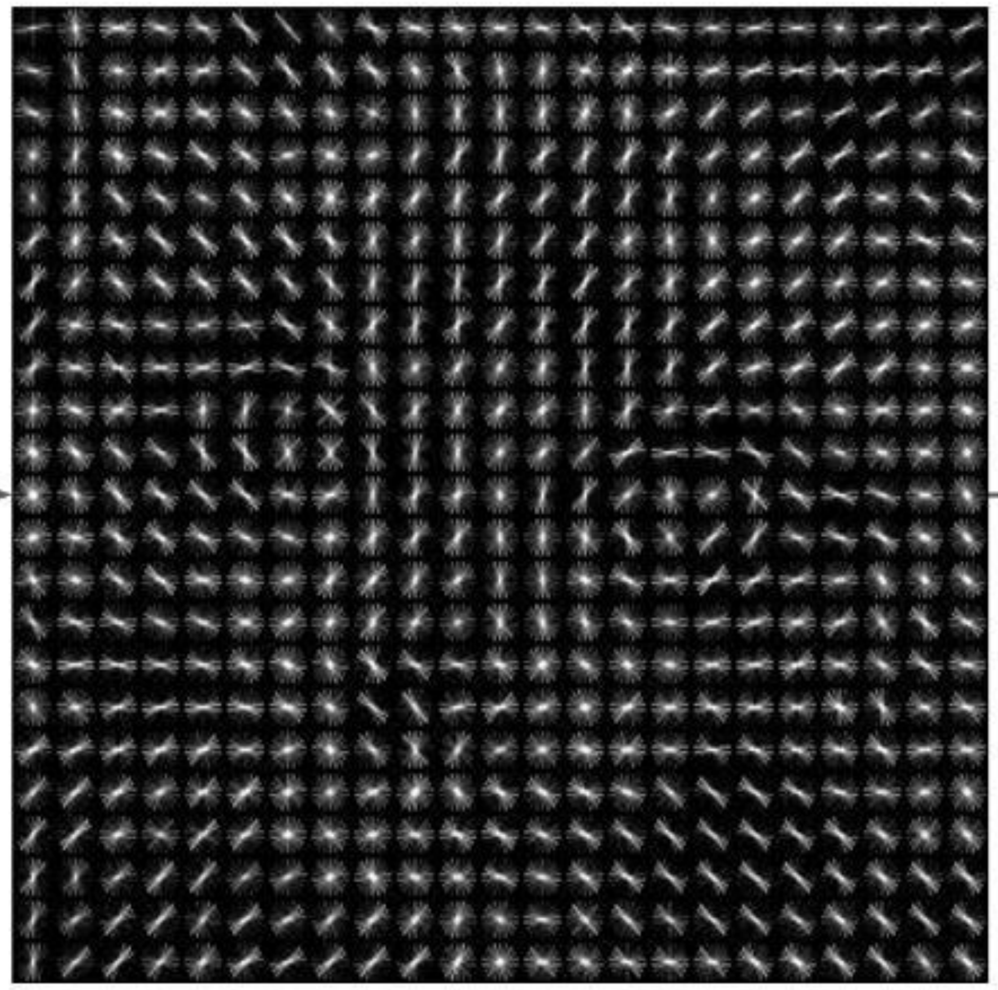
Psychology

Biology

Computer Vision: *Pre 2012*

Input image

Preprocessing



Features : HAAR, HOG, SIFT, SURF

Learning Algorithm

SVM, Random Forests, ANN

Label Assignment

Cat or Background

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Computer Vision: *Modern techniques*

Input image

Deep Learning

Label Assignment

Cat or Background

Open the box

Become the Magician

Goal: Learn a Parametric Function

$$f_\theta : \mathbb{X} \longrightarrow \mathbb{Y}$$

θ: function parameters, these are learned

$\mathbb{X}$: source domain

$\mathbb{Y}$: target domain

Examples:

Image Classification: $f_\theta : \mathbb{R}^{w \times h \times c} \longrightarrow \{0, 1, \ldots, k-1\}$

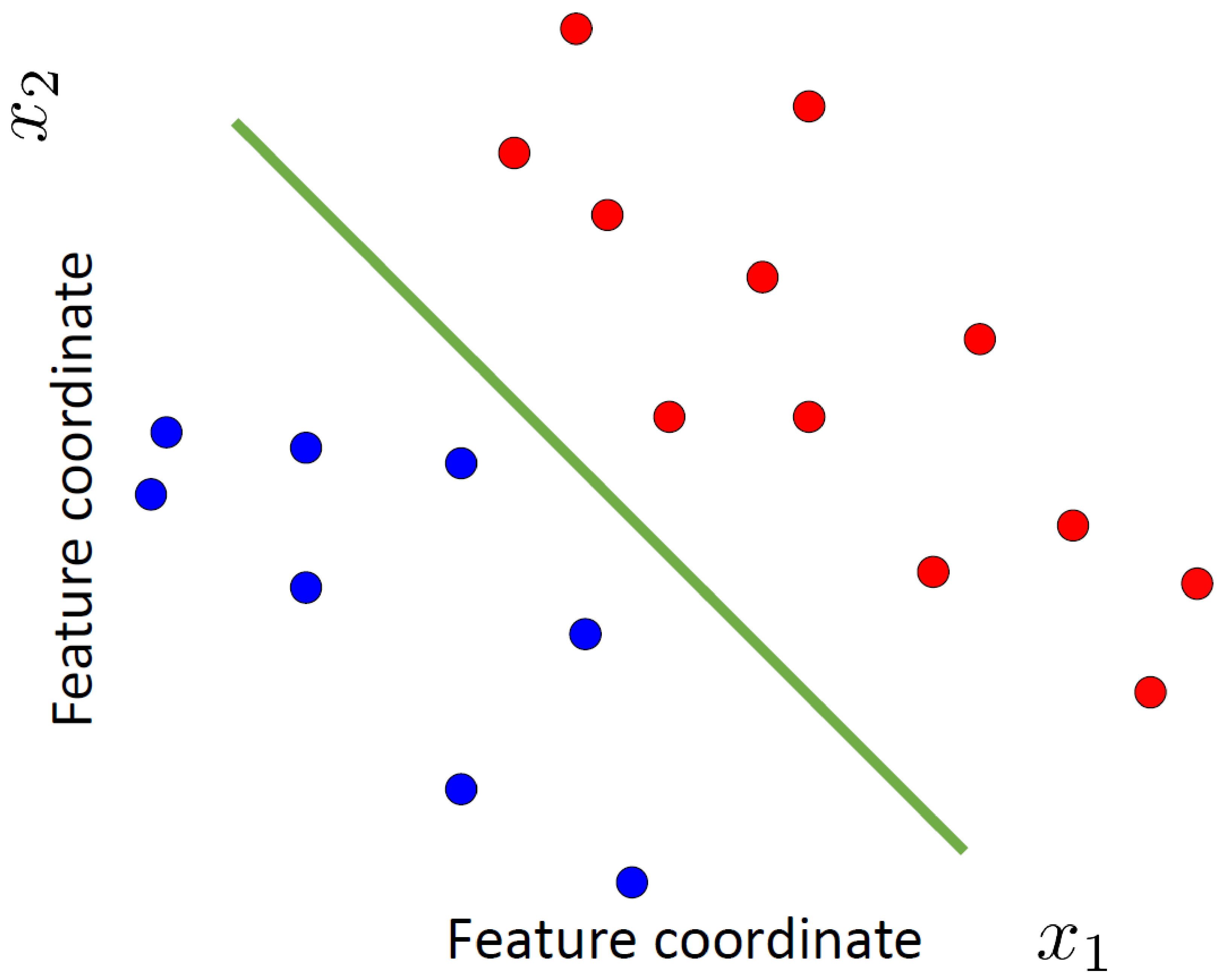
$w \times h \times c$: image dimensions k: class count

Image Synthesis: $f_\theta : \mathbb{R}^n \longrightarrow \mathbb{R}^{w \times h \times c}$

n : latent variable count $w \times h \times c$: image dimensions

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Goal: Learn a Parametric Function



$$f_\theta : \mathbb{R}^n \longrightarrow \{0,1\}$$

$$f_\theta(x) = \begin{cases} 1 & \text{if } wx + b \geq 0 \\ 0 & \text{if } wx + b < 0 \end{cases}$$

$$\theta = \{w, b\}$$

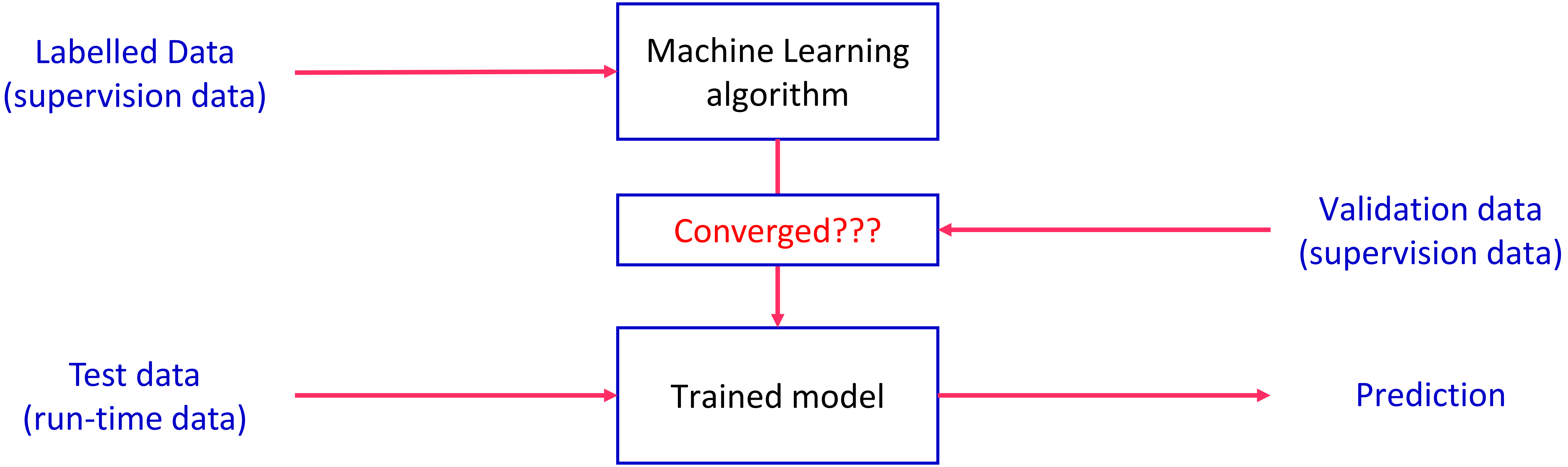
Each data point has a class label:

$$y^i = \begin{cases} 1 & (\color{red}{\bullet}) \\ 0 & (\color{blue}{\bullet}) \end{cases}$$

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Data-driven Algorithms (supervised)



Implementation Practice: Training: 70%; Validation: 15%; Test 15%

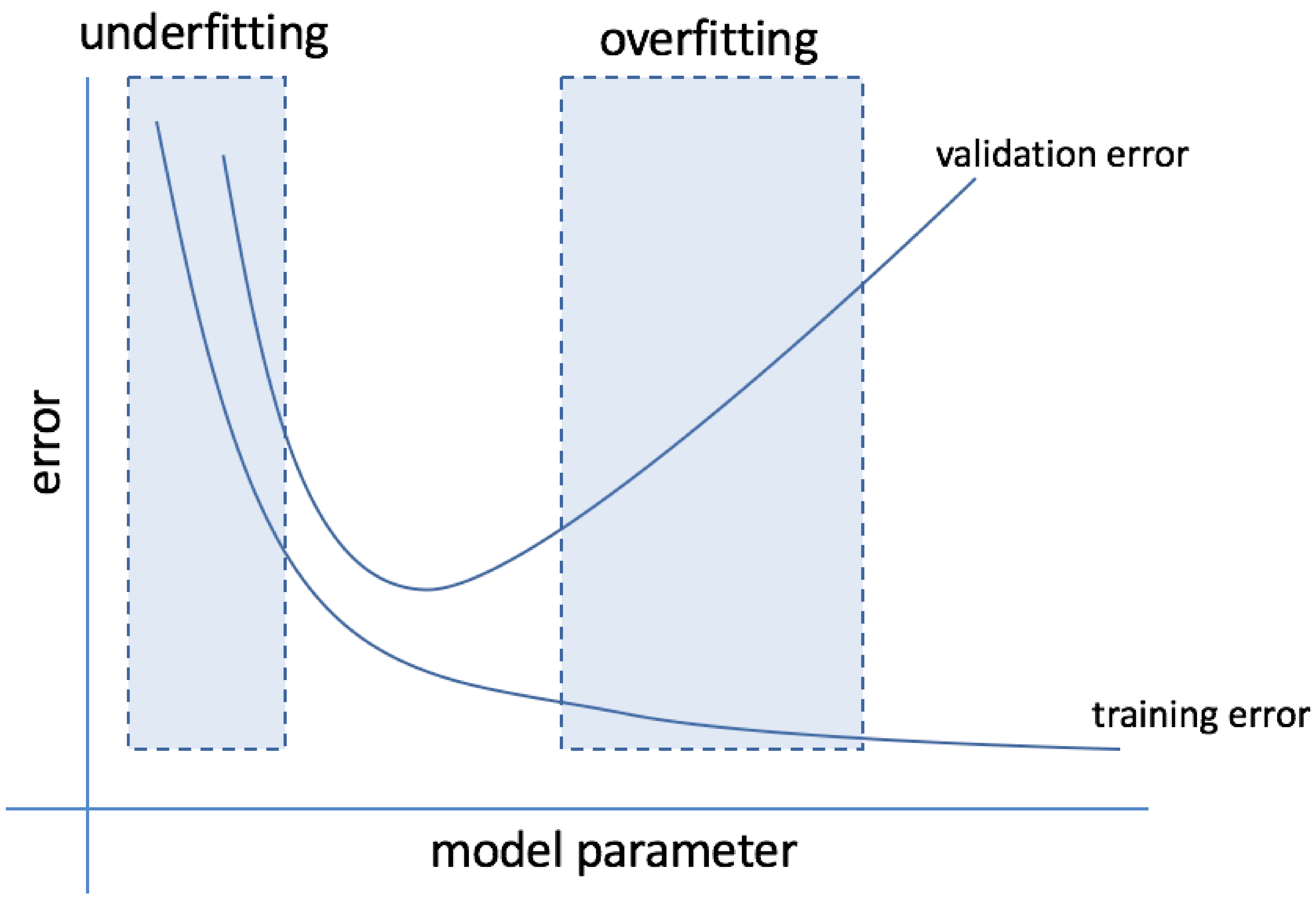
Data-driven Algorithms (supervised)

One way to check if a model is converging is to monitor the performance of the model on the validation dataset during training.

- One common approach is to plot the performance metric (e.g. accuracy, mean squared error) on the validation dataset after each iteration or epoch of training. If the model is converging, you would expect to see the performance metric improve as training progresses, and eventually plateau or converge to a stable value.
- Another way to check for convergence is to check the change in the parameters of the model, if the change is below a certain threshold, it's a good indication that the model has converged.
- Another way to check convergence is to check the change in the loss function if it's below a certain threshold, it's a good indication that the model has converged.
- It's worth noting that an overfitting model will have high accuracy on the training set but low accuracy on validation set, so it's important to use both training and validation set to check for convergence.

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Training versus Validation Loss/Accuracy



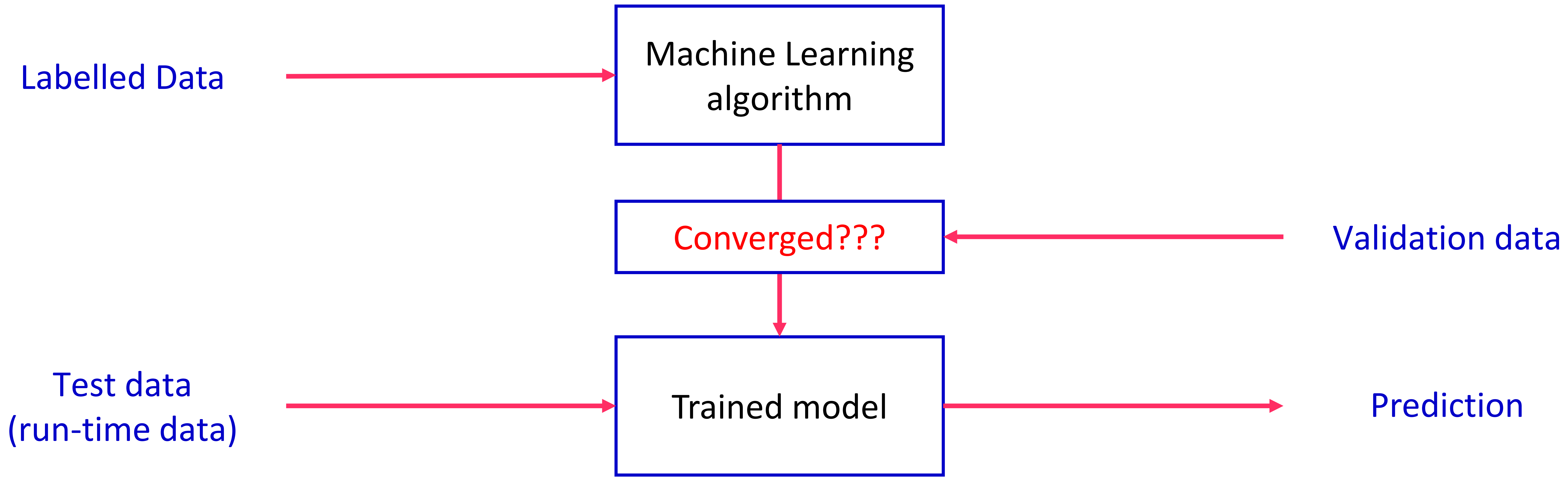
In machine learning, underfitting and overfitting refer to the cases where a model is not able to capture the underlying patterns in the data correctly.

Underfitting occurs when a model is too simple and cannot capture the complexity of the data. This results in a model that has poor performance on both the training and validation datasets. It can be identified by having a high bias and low variance.

Overfitting, occurs when a model is too complex and fits the noise in the data, rather than the underlying patterns. This results in a model that performs well on the training data but poorly on new, unseen data (validation or test set). It can be identified by having a low bias and high variance.

One way to mitigate overfitting is to use regularization techniques (such as L1 or L2 regularization) to constrain the model's complexity and prevent it from fitting the noise in the data. Another way is to use techniques such as early stopping, dropout, or bagging to reduce the variance of the model

Data-driven Algorithms (unsupervised)

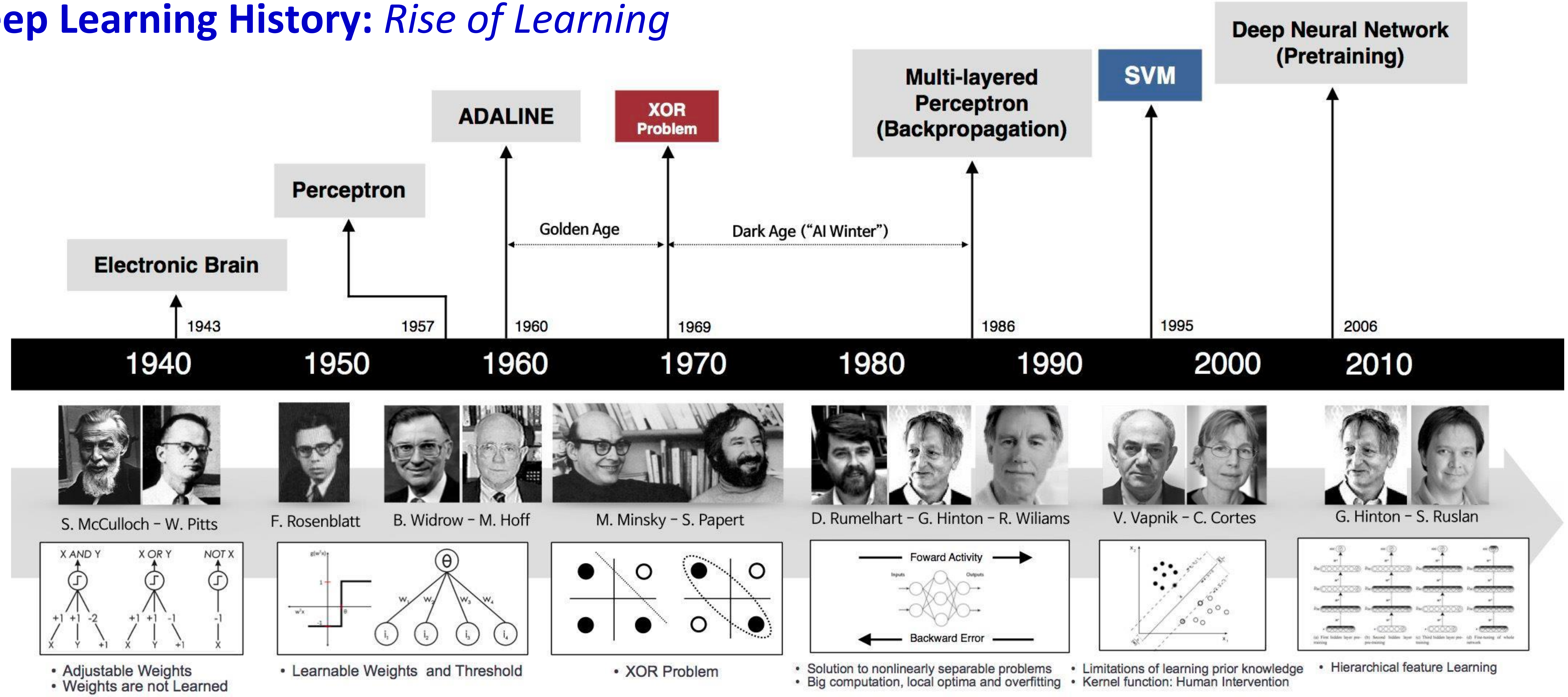


Implementation Practice: Training: 70%; Validation: 15%; Test 15%

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning History: *Rise of Learning*



http://beamlab.org/deeplearning/2017/02/23/deep_learning_101_part1.html

What's changed? Why Now?

Many of the core concepts for deep learning were in place by the 80s or 90s, so what happened in the past 6-8 years that changed things?

Though there are many factors, the two most crucial components appear to be availability of **massive labeled data sets** and **GPU computing**. Here's a run down of factors that seem to have had a role in the deep learning revolution:

- Appearance of large, high-quality labeled datasets - *Data along with GPUs probably explains most of the improvements we've seen. Deep learning is a furnace that needs a lot of fuel to keep burning, and we finally have enough fuel.*
- Massively parallel computing with GPUs - *It turns out that neural nets are actually just a bunch of floating point calculations that you can do in parallel. It also turns out that GPUs are great at doing these types of calculations. The transition from CPU-based training to GPU-based has resulted in massive speed ups for these models, and as a result, allowed us to go bigger and deeper, and with more data.*
- Backprop-friendly activation functions - *The transition away from saturating activation functions like tanh and the logistic function to things like ReLU have alleviated the vanishing gradient problem*

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What's changed? Why Now?

Many of the core concepts for deep learning were in place by the 80s or 90s, so what happened in the past 6-8 years that changed things?

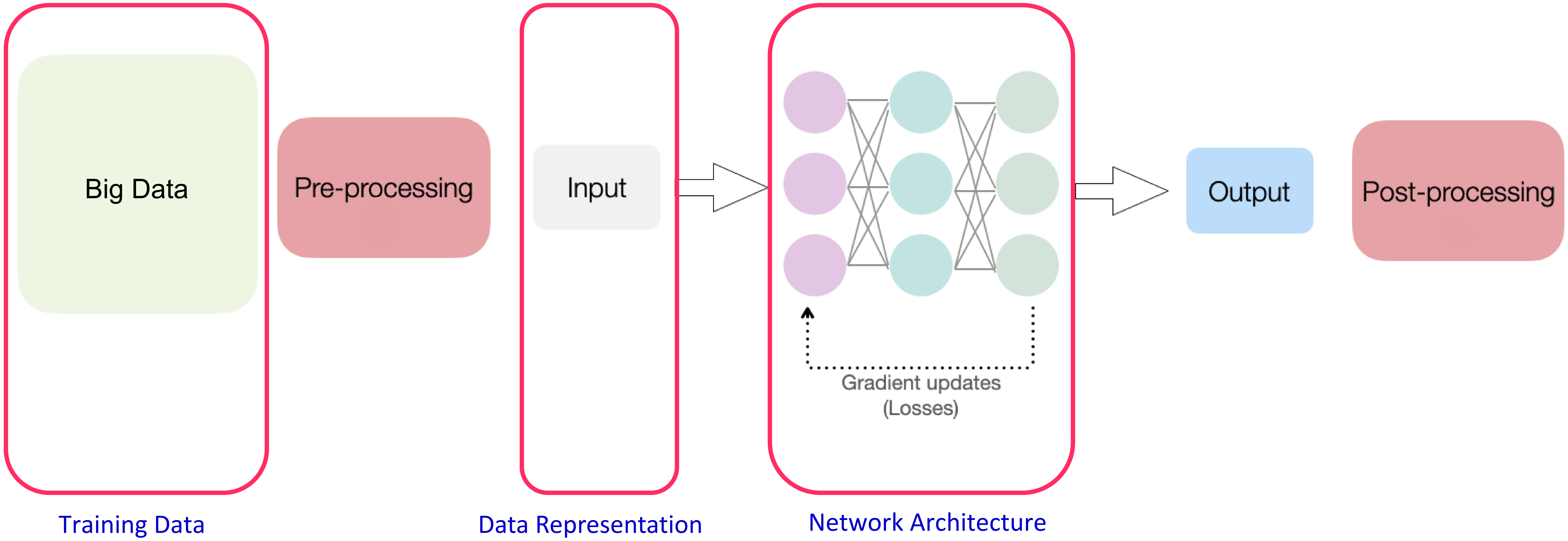
Though there are many factors, the two most crucial components appear to be availability of **massive labeled data sets** and **GPU computing**. Here's a run down of factors that seem to have had a role in the deep learning revolution:

- Improved architectures - *Resnets, inception modules, and Highway networks keep the gradients flowing smoothly, and let us increase the depth and flexibility of the network*
- Software platforms - *Frameworks like tensorflow, theano, chainer, and mxnet that provide automatic differentiation allow for seamless GPU computing and make protoyping faster and less error-prone. They let you focus on your model structure without having to worry about low-level details like gradients and GPU management.*
- New regularization techniques - *Techniques like dropout, batch normalization, and data-augmentation allow us to train larger and larger networks without (or with less) overfitting*
- Robust optimizers - *Modifications of the SGD procedure including momentum, RMSprop, and ADAM have helped eek out every last percentage of your loss function.*

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A common machine (deep) learning architecture



Data is the New Currency

Synthetic data

- Generative model + photo-realistic rendering
- Object geometry + physical simulation
- Object geometry + synthetic materials + realistic simulations

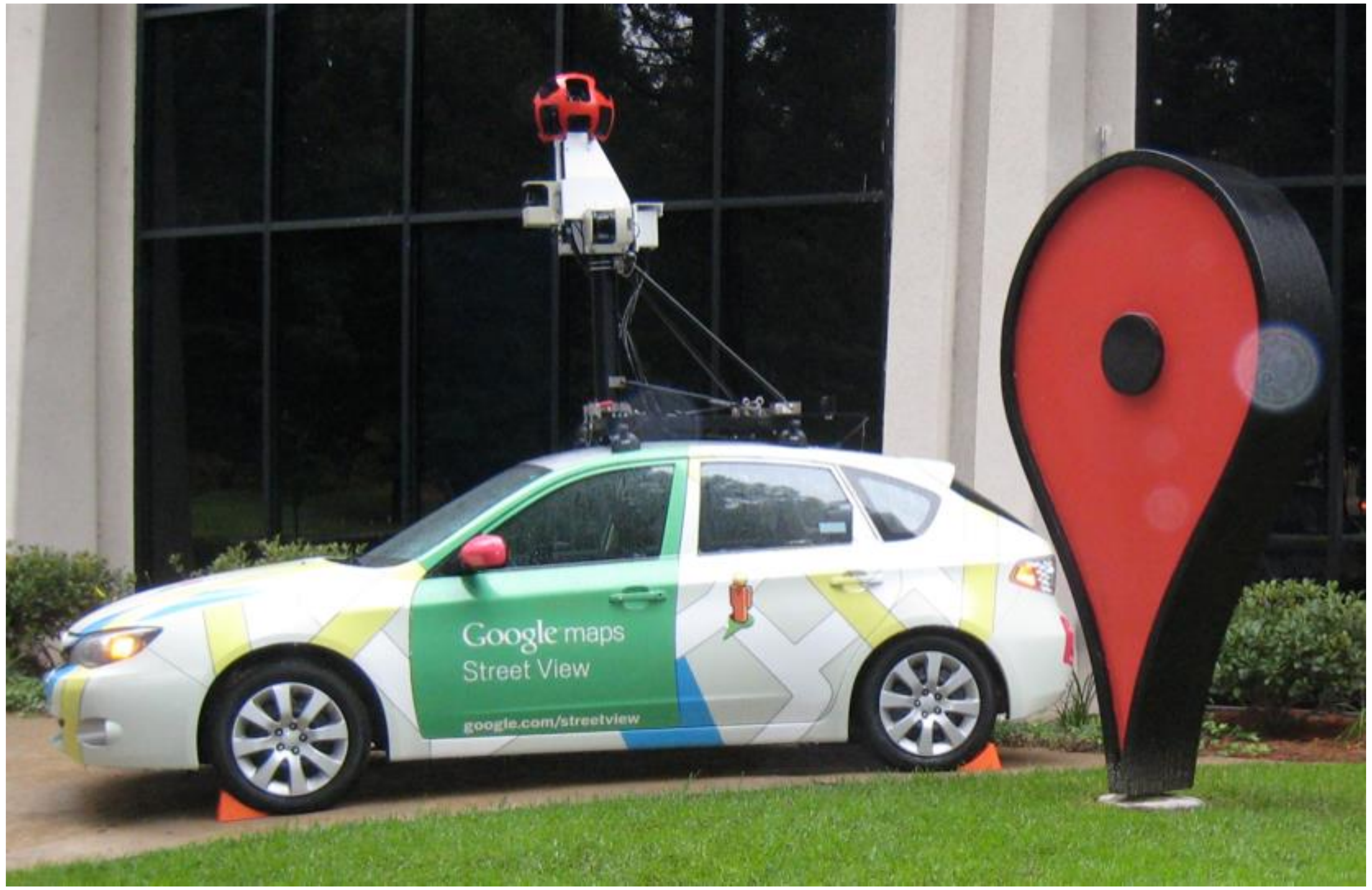
Real data

- Collected from images, scans, mocap sessions
- Collected using specialized equipments (e.g., light-field, pressure gloves)

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

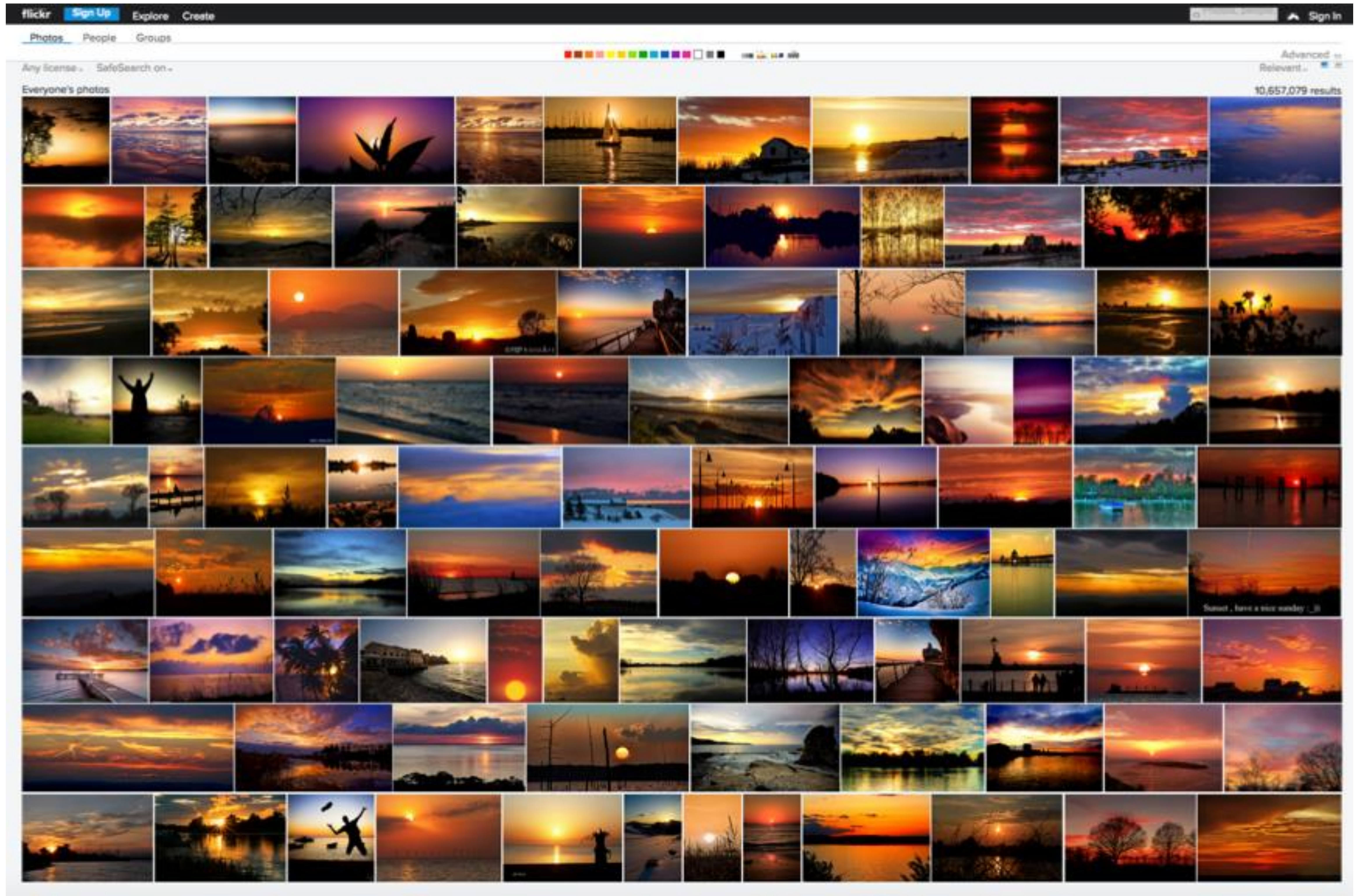
Big Visual Data

Billions of photos: Google street view



Big Visual Data

Billions of photos: Flickr



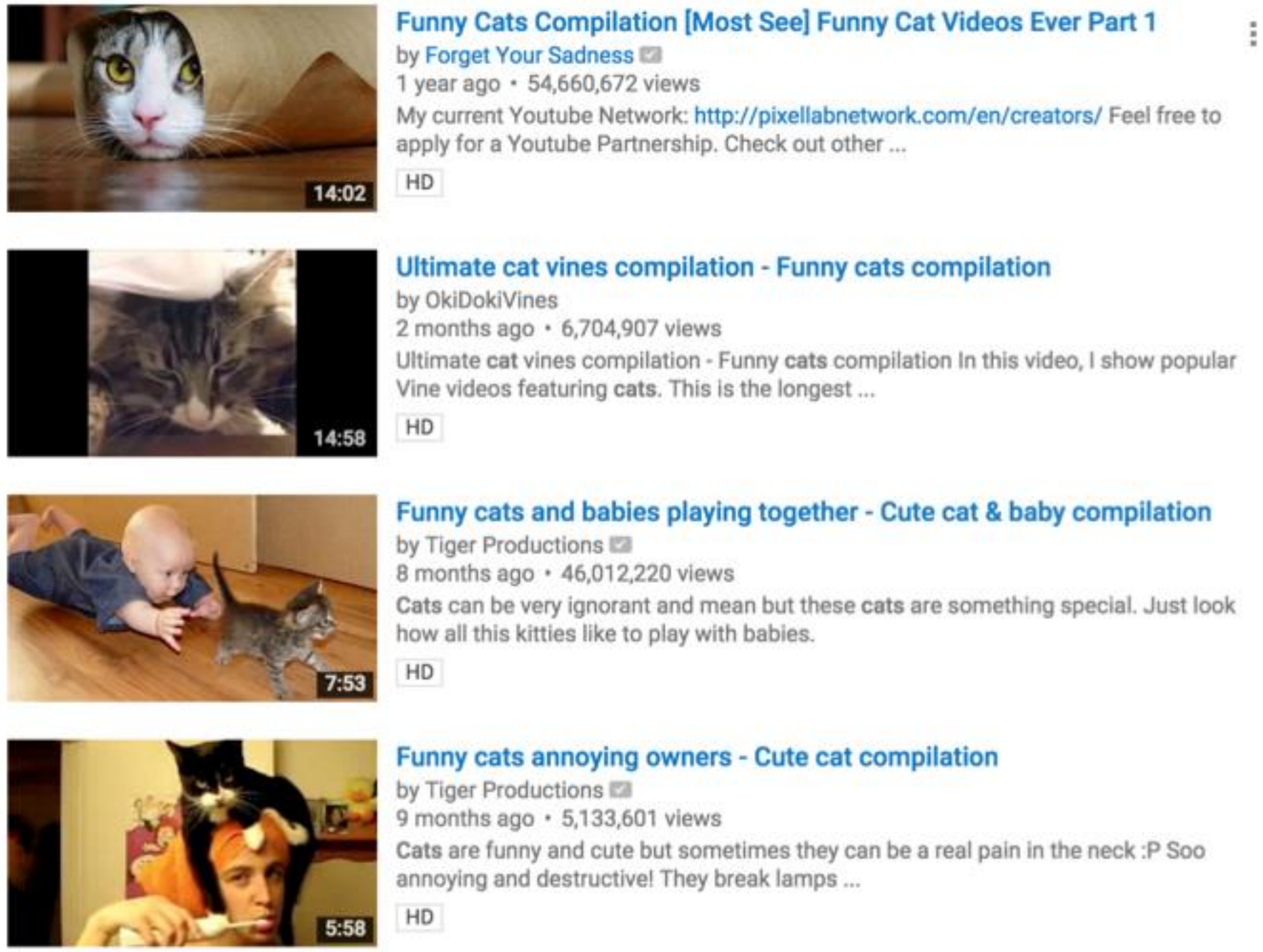
Big Visual Data

Billions of photos: Facebook

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

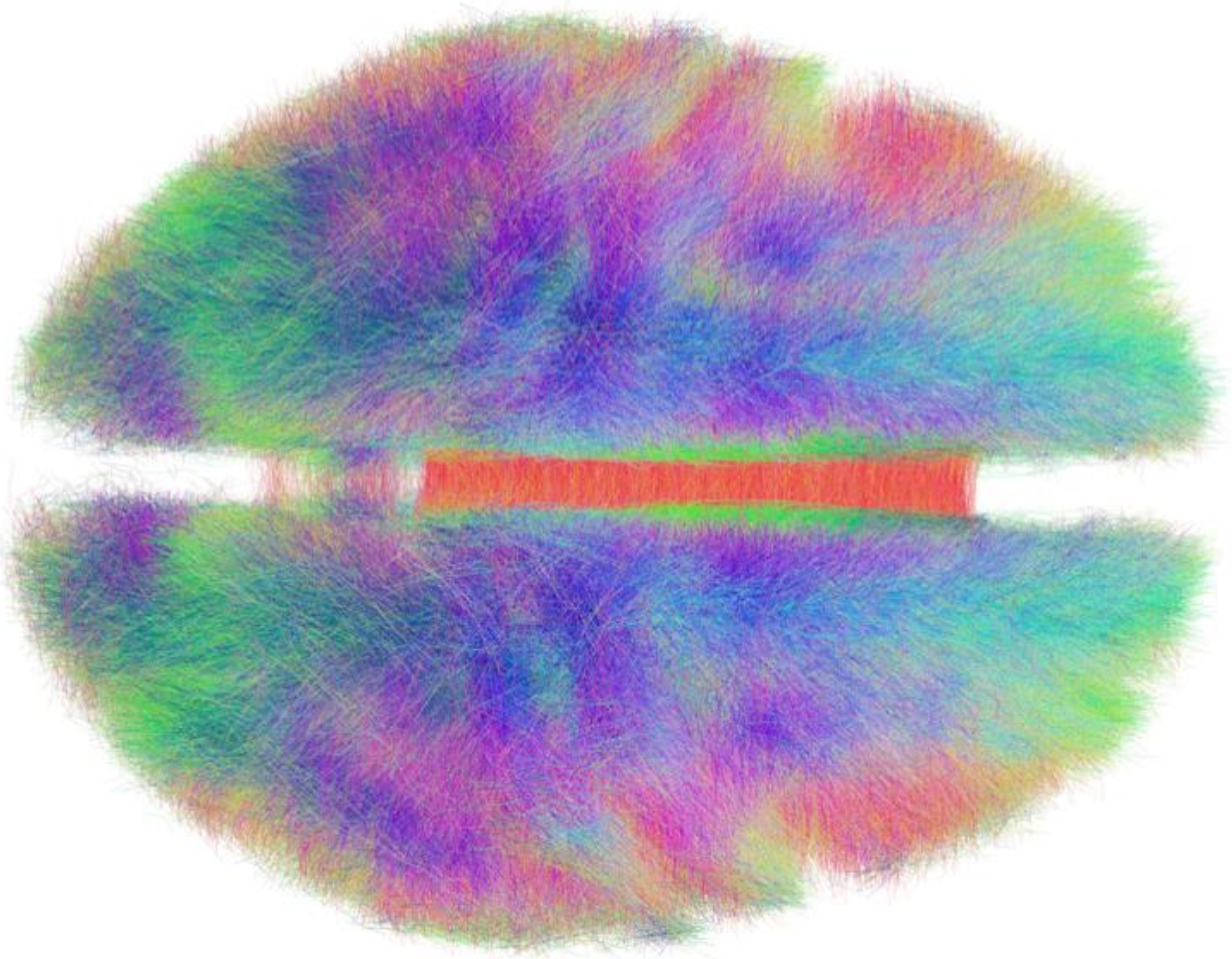
Big Visual Data

Billions of videos: YouTube



Big Visual Data

Medical imaging: photos or volumetric data

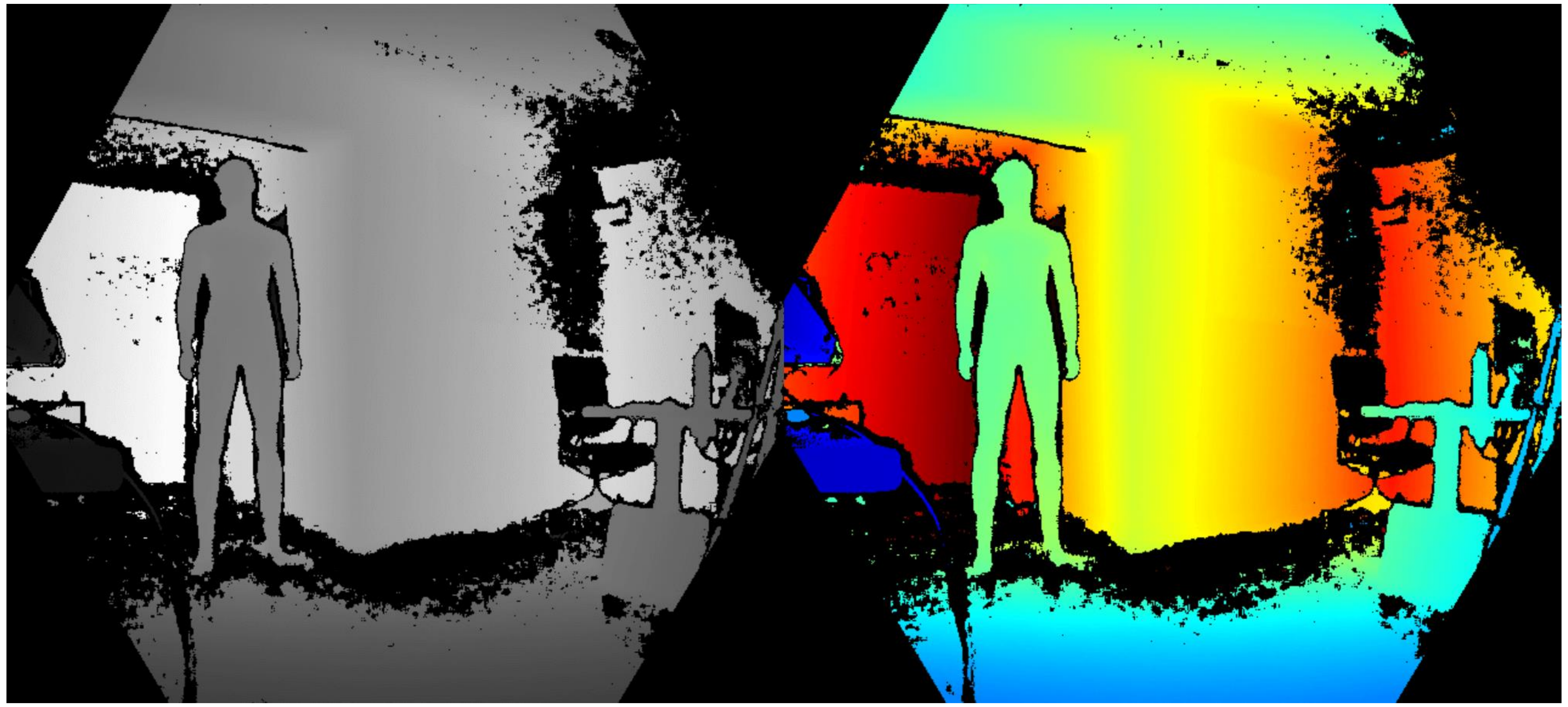


Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Big Visual Data

New sensors: depth (Kinect), stereo cameras



New displays: VR headsets (Oculus Rift, Microsoft Hololens)

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Big Visual Data

Autonomous devices: cars, quadcopters

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Problems *in computer vision and graphics*

How do we process all the data efficiently?

- **Data pre-processing:** Before the data can be processed, it should be pre-processed to remove outliers, noise, and missing values. It should also be normalized or standardized to ensure that all features are on the same scale. This can help to reduce the amount of data that needs to be processed and increase the accuracy of the results.
- **Data augmentation:** Data augmentation is a technique that creates new training examples by applying random transformations to existing images. This can increase the size of the training dataset and reduce overfitting.
- **Batch Processing:** Batch processing is the practice of processing a large dataset in smaller chunks, or "batches", rather than all at once. This can help to reduce the memory requirements and computational time required for processing the data.
- **Parallel Processing:** Modern computer vision algorithms can be computationally intensive, and parallel processing can help to speed up the processing time. This can be done by using multiple CPU cores, GPUs or cloud services.
- **Transfer Learning:** Transfer learning is a technique where a pre-trained model is fine-tuned on a new dataset, rather than training a new model from scratch. It reduces the amount of data and computational resources required to train a new model.
- **Compression:** Compression techniques can be applied to the data to reduce the file size and make it easier to transfer and store. This can be done by using lossless or lossy compression.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which data is considered good for training a deep network?

Data that is considered good for training a deep network typically has the following characteristics:

- **Large quantity:** Training a deep network requires a large dataset to ensure that the network has enough examples to learn from.
- **High quality:** The data should be accurate, relevant, and unbiased. The training data should also be representative of the problem you're trying to solve.
- **Diversified:** The data should be diverse and cover a wide range of variations of the problem (near the boundary edges). This is especially important if you are dealing with images, videos, or audio data.
- **Labeled:** The data should be labeled, meaning that it has been annotated with the correct output or class, to train the deep network.
- **Pre-processed:** The data should be pre-processed to remove outliers, noise, and missing values. It should also be normalized or standardized to ensure that all features are on the same scale.
- **Balanced:** The data should be balanced, meaning that it should have roughly the same number of examples for each class. If the data is imbalanced, it can lead to the network being biased towards the majority class.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

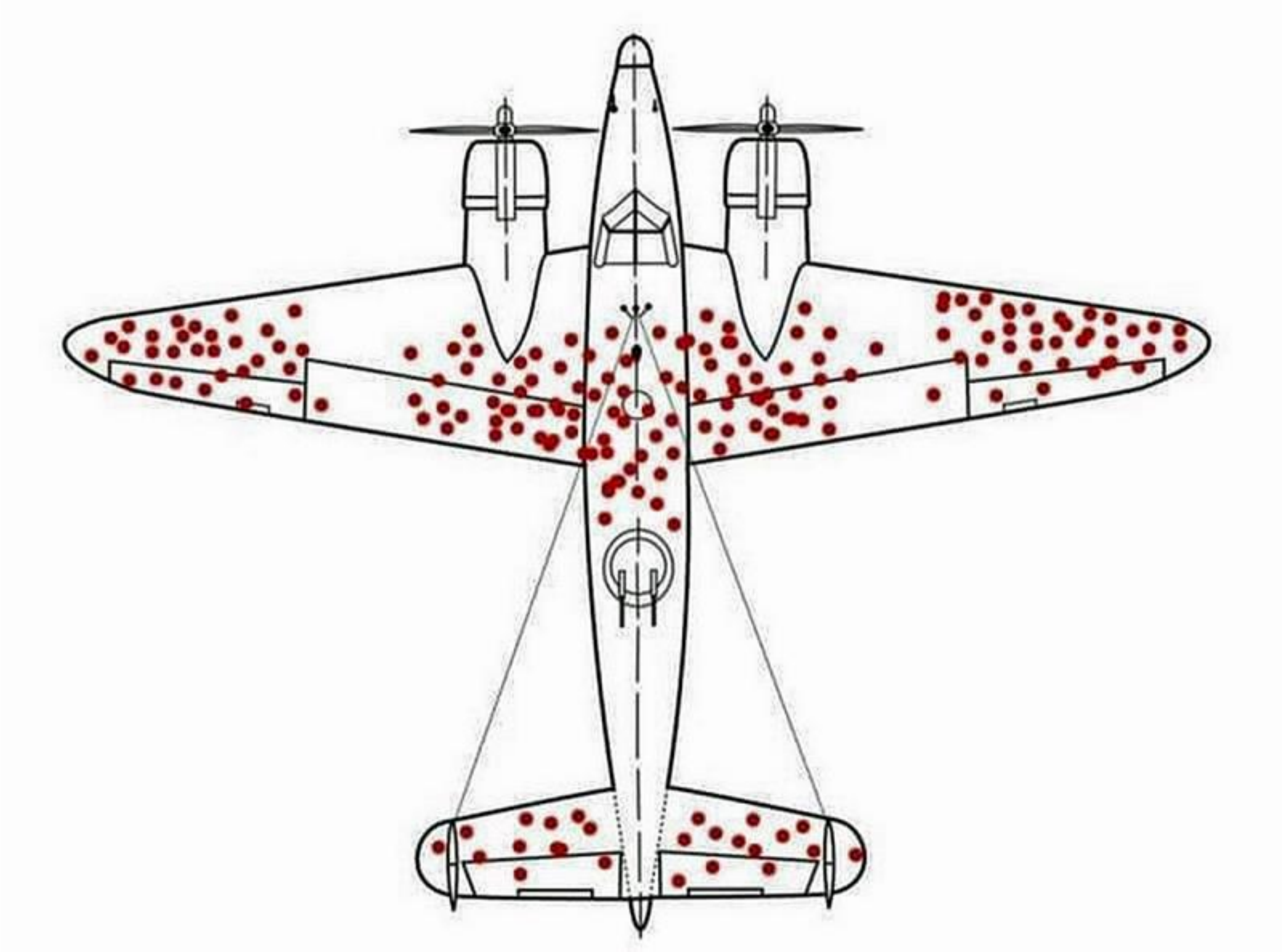
Data Importance

During World War II, fighter planes would come back from battle with bullet holes. The Allies initially sought to strengthen the most commonly damaged parts of the planes to increase combat survivability.

A mathematician, Abraham Wald, pointed out that perhaps the reason certain areas of the planes weren't covered in bullet holes was that planes that were shot in certain critical areas did not return. This insight led to the armor being re-enforced on the parts of returning planes where there were no bullet holes.

The story behind the data is arguably more important than the data itself. This shows that the reasons why we are missing certain data may be more meaningful than the available data, itself. In questions of aircraft design, don't only listen to what the evidence says, listen also to what is not being said.

Wallis, W. Allen (1980). "The Statistical Research Group, 1942-1945: Rejoinder". Journal of the American Statistical Association.



Learned Features

Before

- Handcrafted feature extraction, e.g., edges or corners (hand-crafted)
- Mostly with linear models (PCA)

Now

- End-to-end
- Move away from hand-crafted representations

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Learned Loss

Before

- Evaluation came after
- It was a bit optional
  - You might still have a good algorithm without a good way of quantifying it

Now

- It is essential and build-in
- If the loss is not good, the result is not good
- (Extensive) Evaluation happens automatically

While still much is left to do, this makes graphics much more reproducible

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

What is Special about Computer Vision and Graphics?

1. **Regular data structure** and easy to parallelize (e.g., image translation)
2. Many sources of input data — **model building** (e.g., images, scanners, motion capture)
3. Many sources of **synthetic data** — can serve as supervision data (e.g., rendering, animation)
4. Many problems in **generative models** and need for **user-control**

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Main Challenges and Scope for Innovation

1. **Representation:** How is the data organized and structured?
2. **Training data:** Is it synthetic or real, or mixed?
3. **User control:** End-to-end or in small steps?
4. **Loss functions:** Hand-crafted or learned from data?

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Convolutional Neural Networks (CNNs):** These are widely used for image classification, object detection, and semantic segmentation tasks. CNNs are designed to process data with a grid-like topology, such as an image.

> - Convolutional Neural Networks (CNNs) are a type of deep learning algorithm that are particularly well-suited for image recognition and processing. They are inspired by the structure of the visual cortex in the human brain and are designed to mimic the way the brain processes visual information.
> - CNNs consist of multiple layers of interconnected nodes, called neurons, that are each connected to a small subset of the input image. The neurons in each layer extract different features of the image, and the final output is a label or class that corresponds to the image.
> - CNNs are widely used in **computer vision** tasks such as object detection, image classification, and image generation.
> - They are also used in **graphics** for tasks such as super-resolution, style transfer, and image-to-image translation.

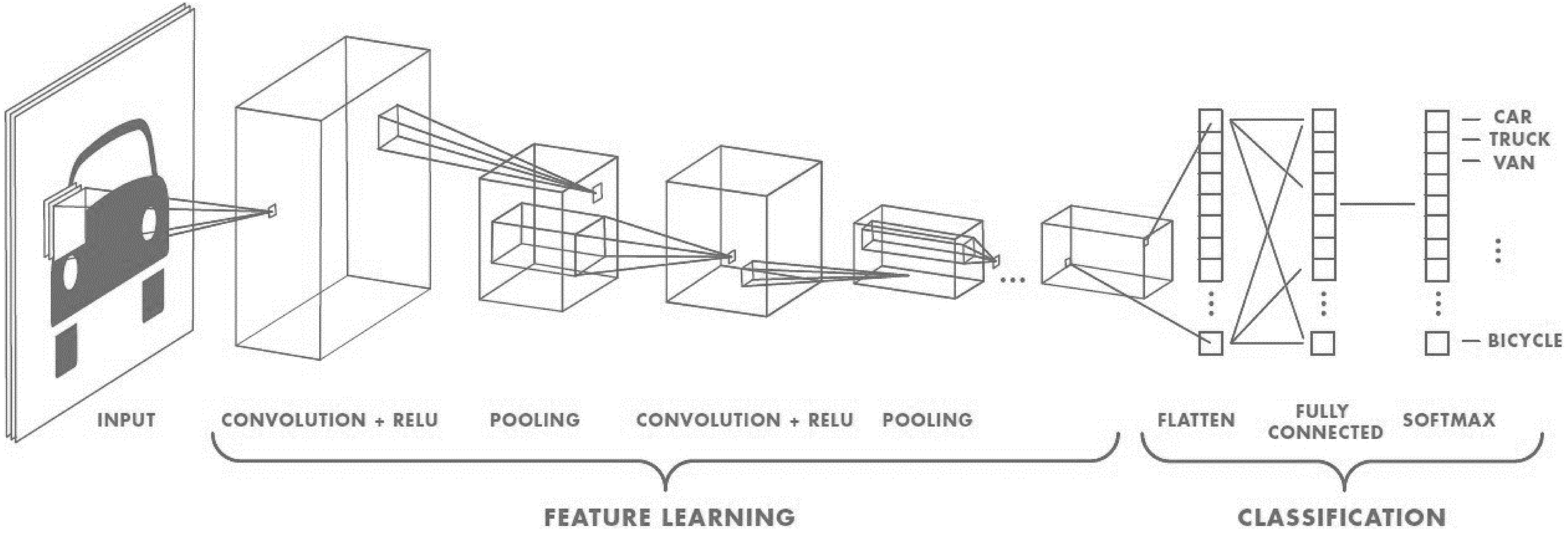
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Convolutional Neural Networks (CNNs):** These are widely used for image classification, object detection, and semantic segmentation tasks. CNNs are designed to process data with a grid-like topology, such as an image.



Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Recurrent Neural Networks (RNNs):** These are used for tasks that involve sequential data, such as video analysis and natural language processing. RNNs can process sequential data by looping the network to allow information to flow from one step of the sequence to the next.

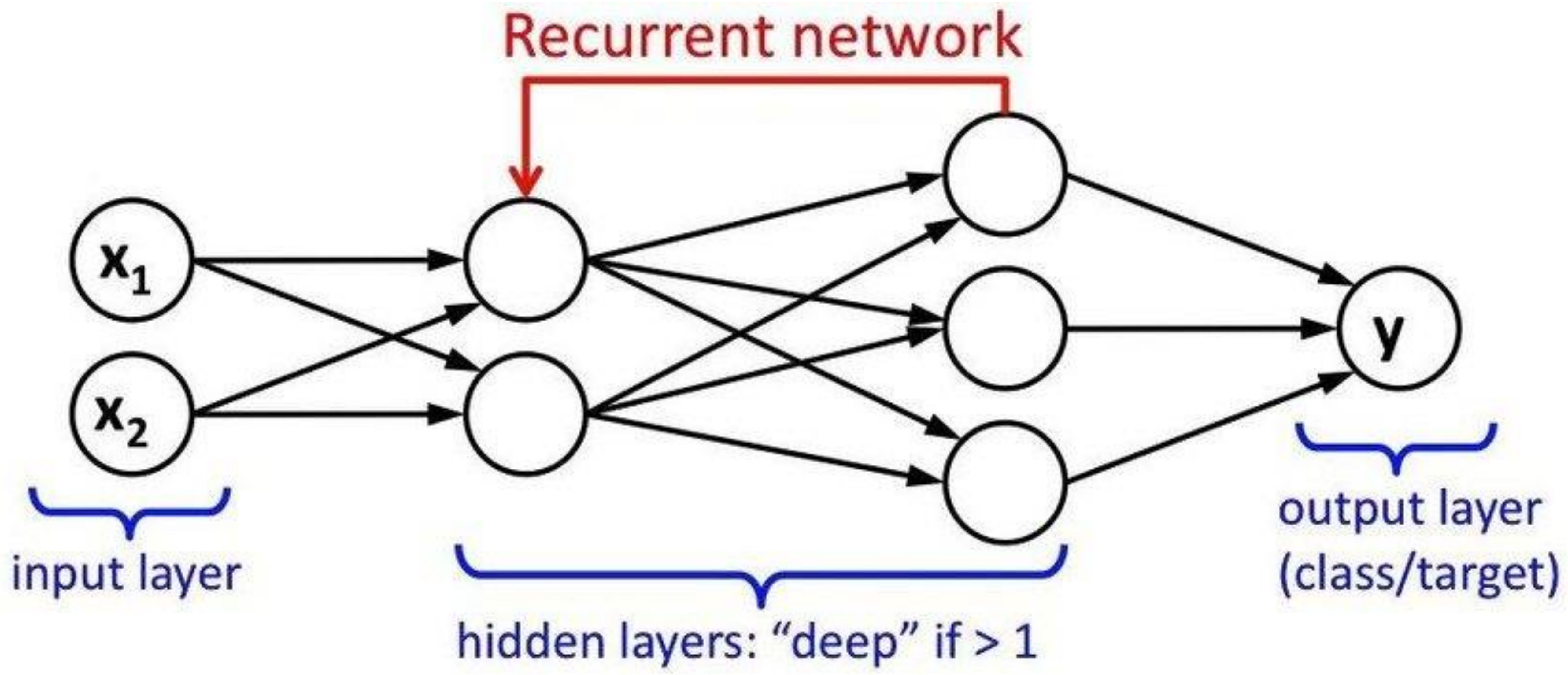
- Recurrent Neural Networks (RNNs) are a type of deep learning algorithm that are particularly well-suited for processing sequential data, such as time series or natural language. They differ from traditional feedforward neural networks in that they have a "memory" component, allowing them to maintain information about previous inputs. This makes them useful for tasks that involve understanding context or relationships between different parts of the input.
- In **computer vision**, RNNs are used in tasks such as video action recognition, where the network needs to understand the temporal context of the video frames in order to correctly identify the action being performed. They are also used in image captioning, where the network generates a natural language description of an image based on the objects and scenes it contains.
- In **graphics**, RNNs are used to generate sequential data, such as video frames or text, by modeling the underlying probability distributions of the data. For example, they can be used to generate realistic animations or to compose music.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Recurrent Neural Networks (RNNs):** These are used for tasks that involve sequential data, such as video analysis and natural language processing. RNNs can process sequential data by looping the network to allow information to flow from one step of the sequence to the next.



Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Generative Adversarial Networks (GANs):** GANs are a class of neural networks that are used for generative tasks such as image synthesis, super-resolution and style transfer. They consist of two neural networks, a generator and a discriminator, that work together to generate new, realistic images.

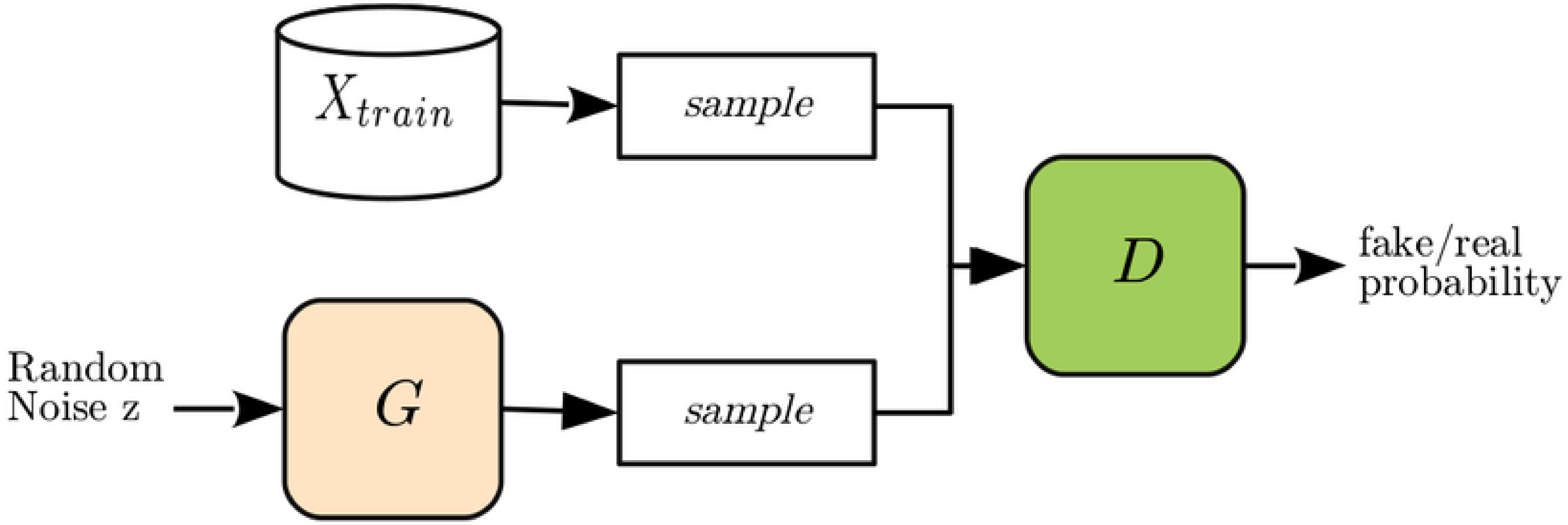
- GANs are a type of deep learning algorithm that are used to generate new, previously unseen data that is similar to a given dataset. They consist of two main components: a generator network and a discriminator network. The generator network generates new data, while the discriminator network attempts to distinguish the generated data from the original dataset. The two networks are trained together in an adversarial manner, with the generator network trying to produce data that can fool the discriminator, and the discriminator trying to correctly identify the generated data.
- In **computer vision**, GANs are used to generate new images, such as realistic photos of faces or objects that do not exist in the real world. They can also be used for tasks such as super-resolution, where the network is trained to generate high-resolution versions of low-resolution images.
- In **graphics**, GANs are used to generate realistic images, videos, and animations. They have been applied in various areas such as creating realistic 3D models, generating realistic textures, creating realistic virtual environments, and realistic characters.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Generative Adversarial Networks (GANs):** GANs are a class of neural networks that are used for generative tasks such as image synthesis, super-resolution and style transfer. They consist of two neural networks, a generator and a discriminator, that work together to generate new, realistic images.



Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Autoencoder:** Autoencoder are a type of neural network that are used for unsupervised learning tasks, such as dimensionality reduction, feature learning, and anomaly detection. Autoencoder can learn a compact representation of the input data in an unsupervised manner.

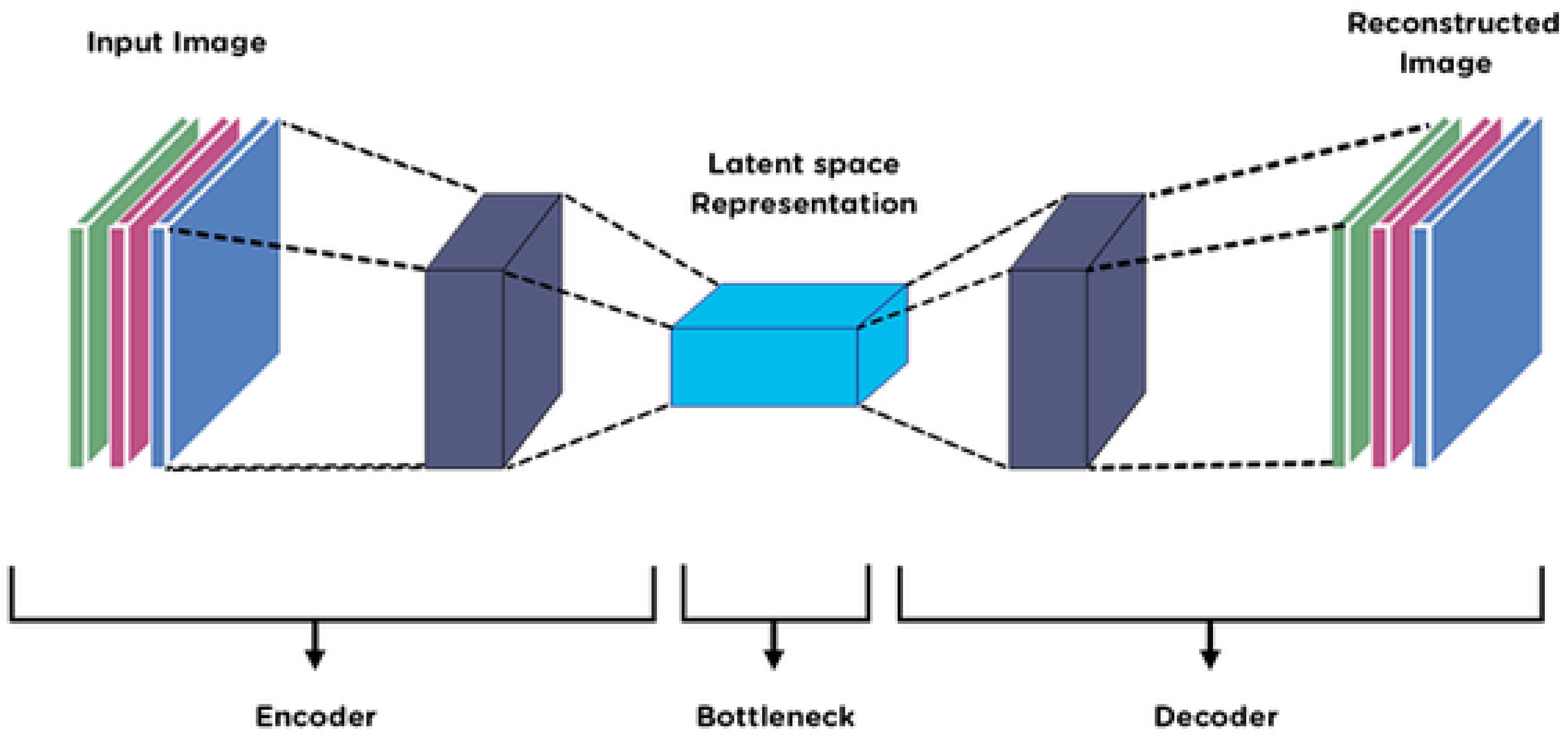
- Autoencoders are a type of neural network that is designed to learn a compact, efficient representation of the input data. They consist of two main components: an encoder, which maps the input data to a lower-dimensional representation, and a decoder, which maps the lower-dimensional representation back to the original input. The goal of training an autoencoder is to learn a representation that captures the most important features of the input data while discarding the less important information.
- In **computer vision**, autoencoders are used to learn a compact representation of images that can be used for tasks such as image compression, denoising, and anomaly detection. They can also be used as a pre-training step for other computer vision tasks, such as object recognition or segmentation.
- In **graphics**, autoencoders are used to learn a compact representation of 3D models, textures, and animations, which can be used for tasks such as compression, style transfer, and data-driven animation. They can also be used to learn a compact representation of images that can be used for tasks such as image-to-image translation.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **Autoencoder:** Autoencoder are a type of neural network that are used for unsupervised learning tasks, such as dimensionality reduction, feature learning, and anomaly detection. Autoencoder can learn a compact representation of the input data in an unsupervised manner.



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **ResNet:** ResNet stands for Residual Network, it's a CNN architecture that addresses the problem of vanishing gradients in deep neural networks. ResNet is widely used for image classification and object detection tasks.

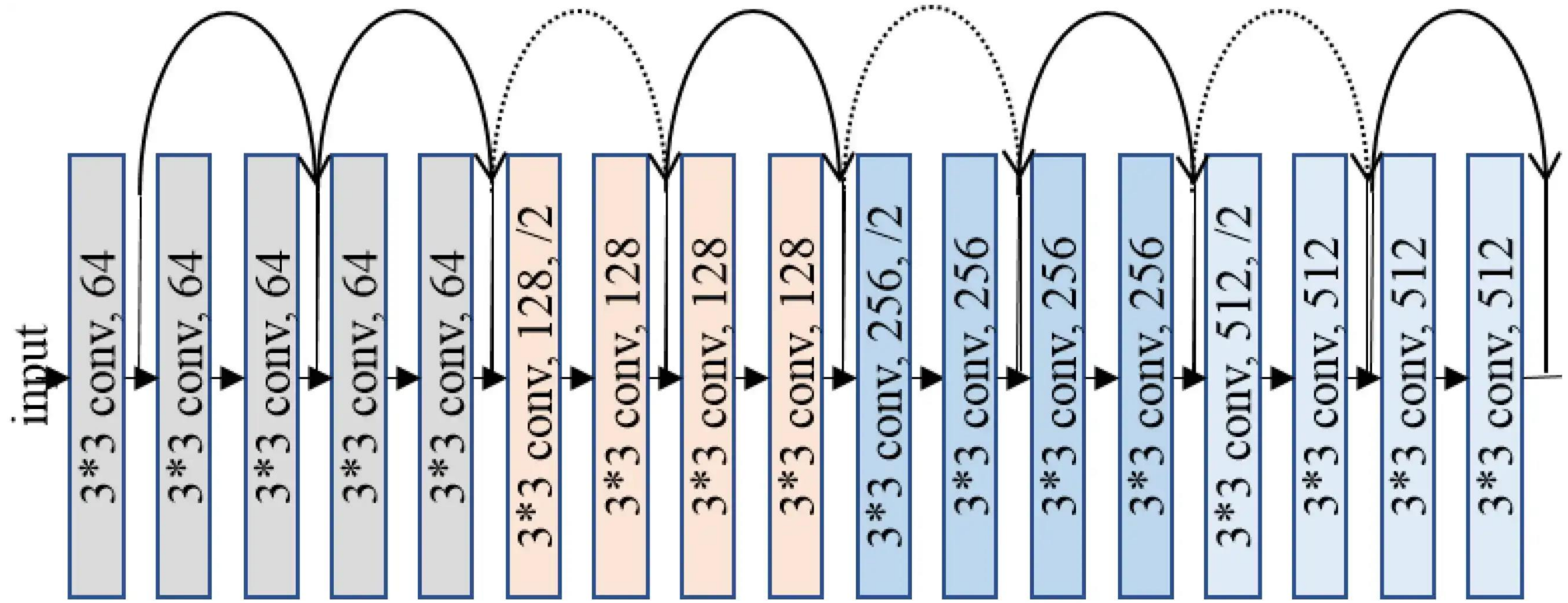
- Residual Networks (ResNets) are a type of convolutional neural network that were first introduced in 2015. They were designed to address the problem of vanishing gradients, which occurs when training deep neural networks. In a traditional neural network, the error signal that is propagated through the network during training can become very small as it flows through many layers, making it difficult to train the network effectively. ResNets address this problem by introducing a new architecture called the "residual block," which allows the error signal to bypass one or more layers and flow directly to deeper layers in the network.
- In **computer vision**, ResNets have been widely used for image classification, object detection, and semantic segmentation tasks. They have shown to achieve state-of-the-art performance on many benchmark datasets.
- In **graphics**, ResNets are used to generate realistic images, videos and animations. They are used for tasks such as image-to-image translation, style transfer and super-resolution. They have been also used in generative models to generate 3D models, textures, and animations, and they have been applied in video game industry to create more realistic characters and environments.

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Which are the most popular deep learning networks in computer vision and graphics?

There are several popular deep learning networks that are used in computer vision and graphics, e.g.,:

- **ResNet:** ResNet stands for Residual Network, it's a CNN architecture that addresses the problem of vanishing gradients in deep neural networks. ResNet is widely used for image classification and object detection tasks.



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Other very popular deep learning networks used in computer vision

U-Net: U-Net is a fully convolutional network (FCN) that is used for image segmentation tasks. It is widely used in medical imaging, where the goal is to segment specific structures within an image.

YOLO: YOLO (You Only Look Once) is a real-time object detection system. It's a single neural network that can detect multiple objects in an image, and it's widely used in self-driving cars, security systems, and other real-time applications.

Mask R-CNN: Mask R-CNN is a variant of Faster R-CNN, an object detection network. Mask R-CNN adds an additional branch for predicting an object mask in parallel with the existing branch for bounding box recognition.

SPADE: SPADE (Semantic Image Synthesis with Spatially-Adaptive Normalization) is a GAN architecture that can generate images with high-level semantic control.

FPN: FPN stands for Feature Pyramid Network, it's a CNN architecture that is used for object detection task, it helps to detect object at different scales by building a pyramid of features with different resolutions.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep learning in character animation

Deep learning has been used in various ways to create realistic and expressive character animation. Some popular deep learning networks used in character animation include:

- Motion capture is a technique that uses sensors to track the motion of a person and translate it to a digital character. Deep learning algorithms can be used to improve the accuracy of motion capture by analyzing the sensor data and inferring the movement of the body parts.
- **Autoencoder-based models** can be used to generate realistic human motion. They can be trained to encode the motion data of a human performer and then decode it to generate new, natural-looking motion for a digital character.
- Generative models such as **GANs** or **Variational Autoencoder (VAE)** can be used to generate realistic animations of characters or creatures. They can learn the distribution of the motion data and can generate new motion data with high level of realism.
- Inverse kinematics is a technique used to calculate the movements of a character's joints to achieve a specific end effector position. Deep learning models can be used to predict the inverse kinematics of a character, making animation more natural and efficient.
- **Neural rendering** is a technique that uses deep learning to generate photorealistic images of characters. It can be used to create high-quality animation, video game characters, and other digital content.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today


A cat sitting on a suitcase on the floor


A cat is sitting on a tree branch


A woman is holding a cat in her hand

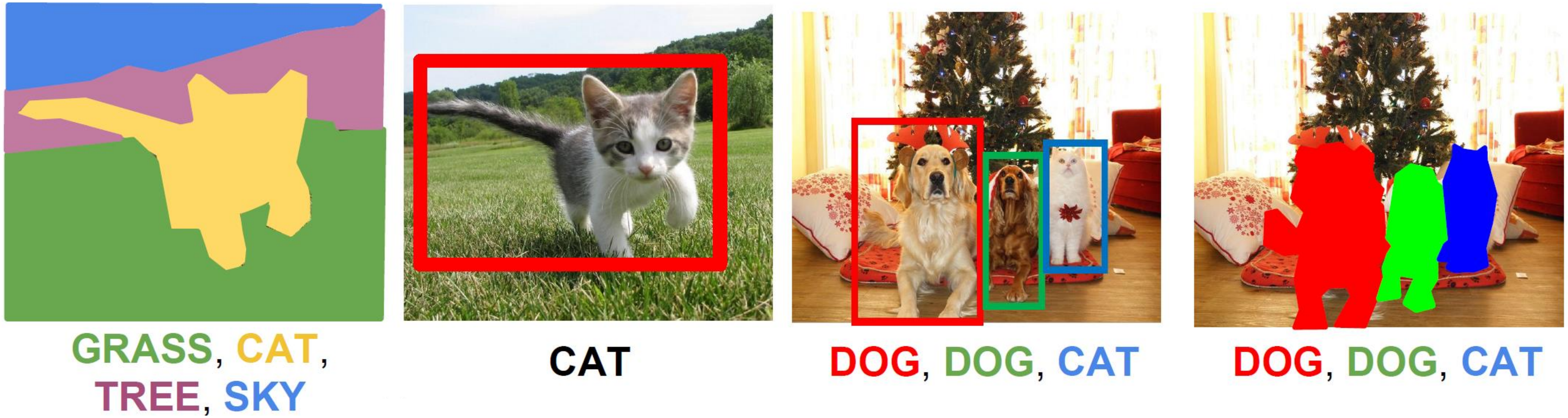

Two people walking on the beach with surfboards


A tennis player in action on the court


A person holding a computer mouse on a desk

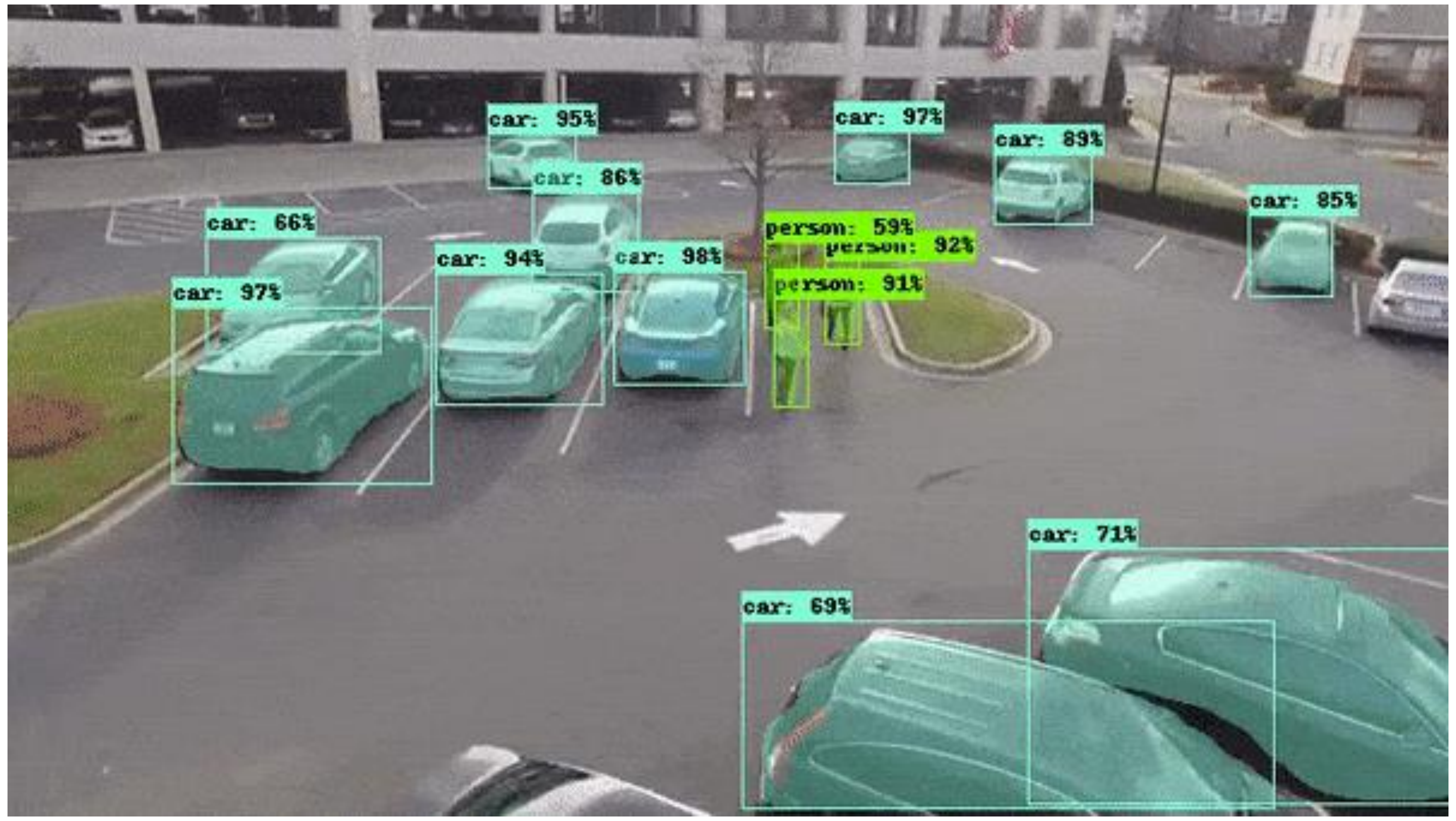
Scene understanding

Deep Learning Today



Semantic Segmentation

Deep Learning Today

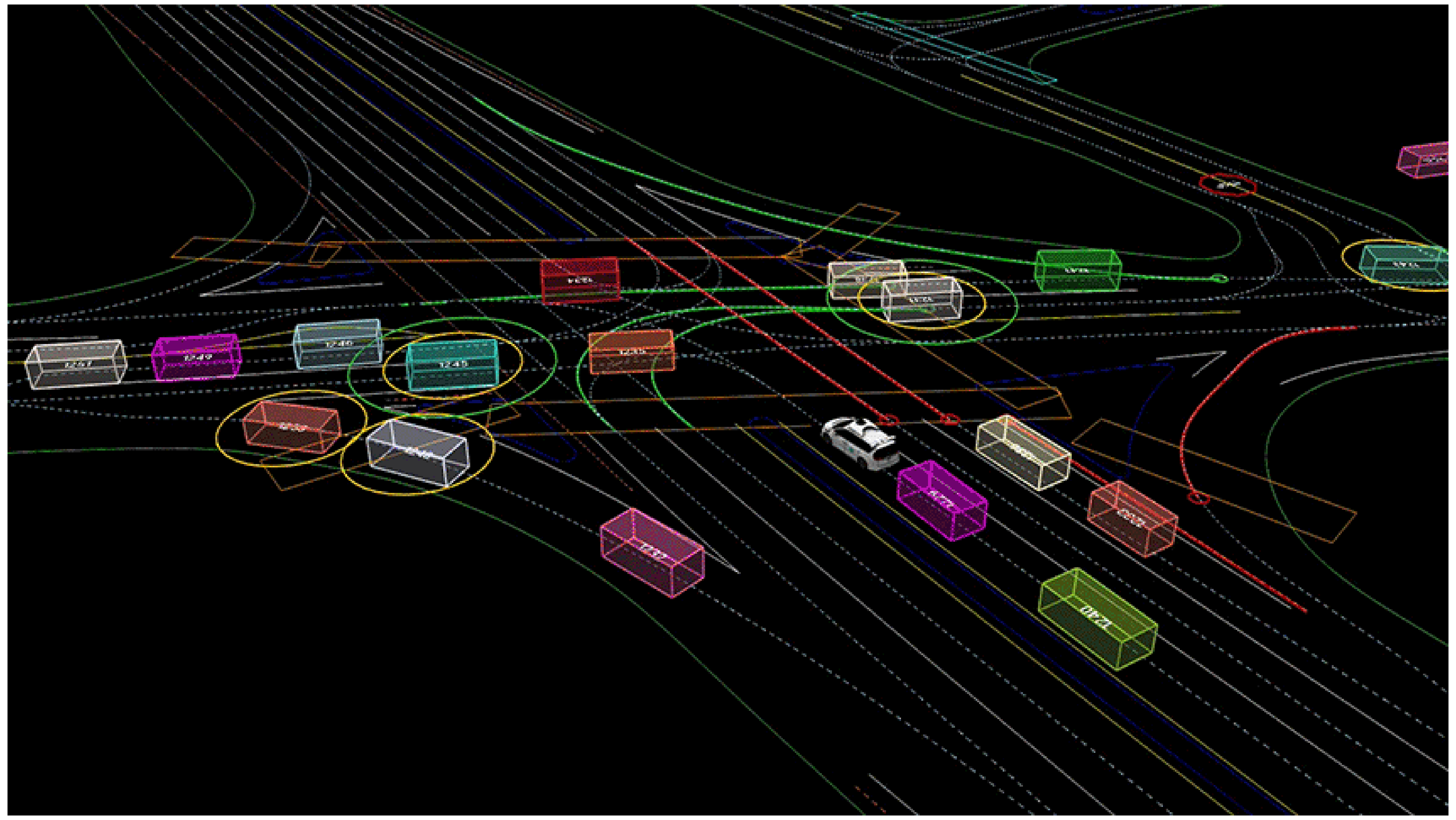


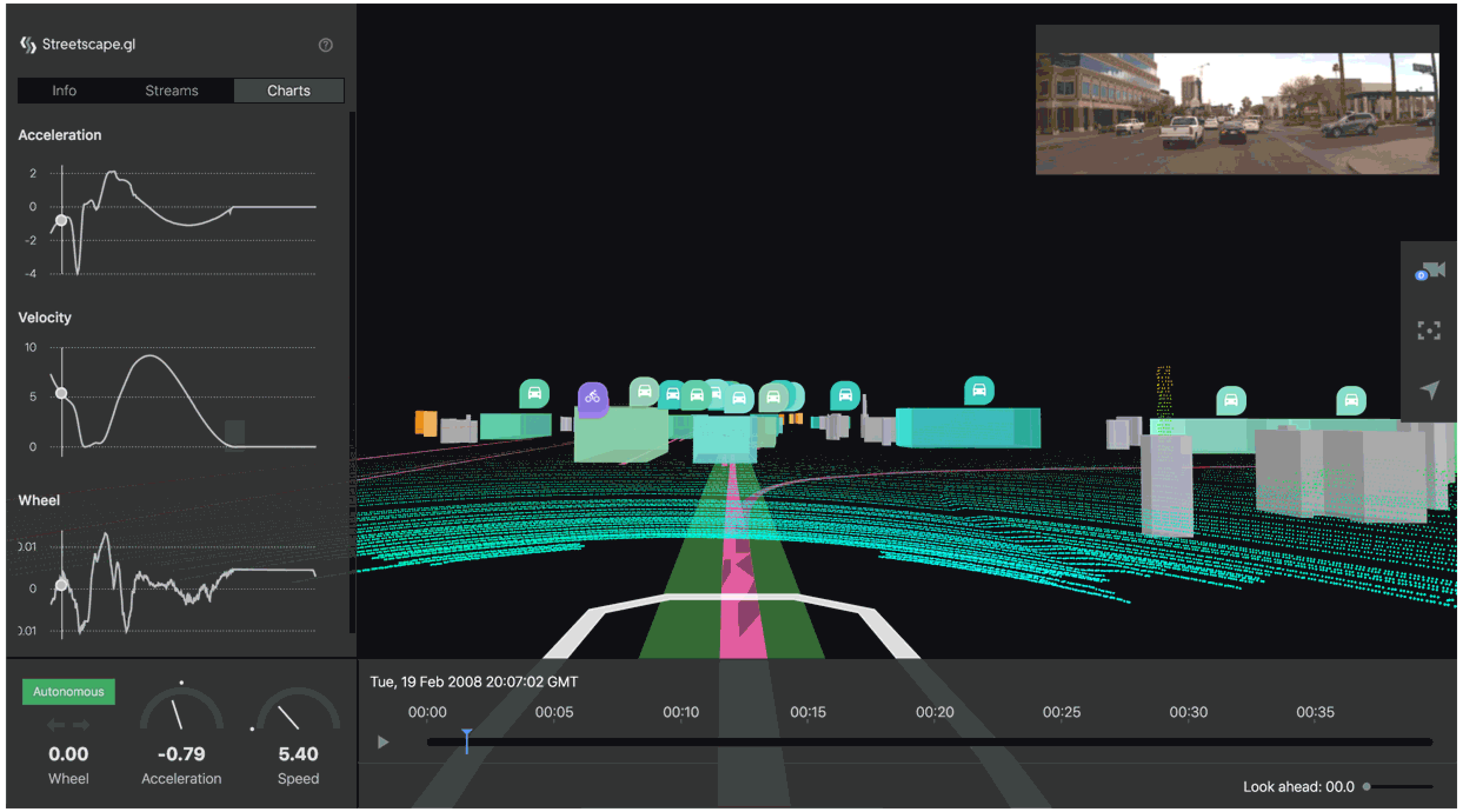
Object detection

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today



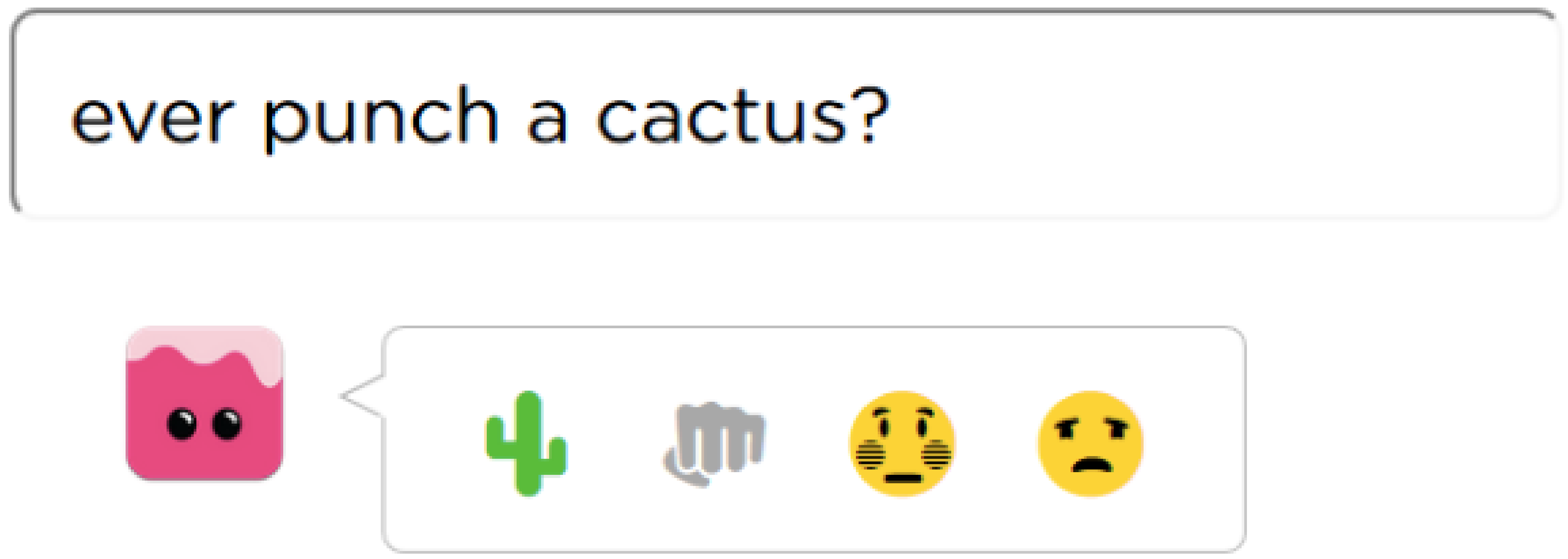


Self-driving cars

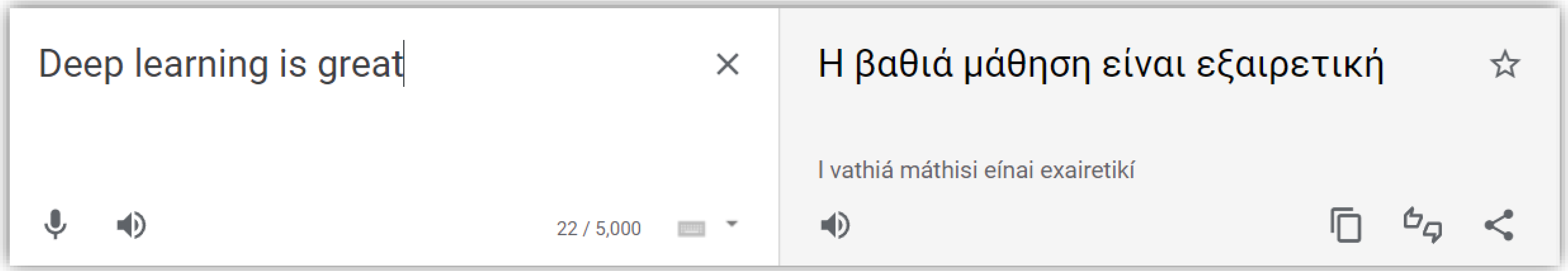
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today

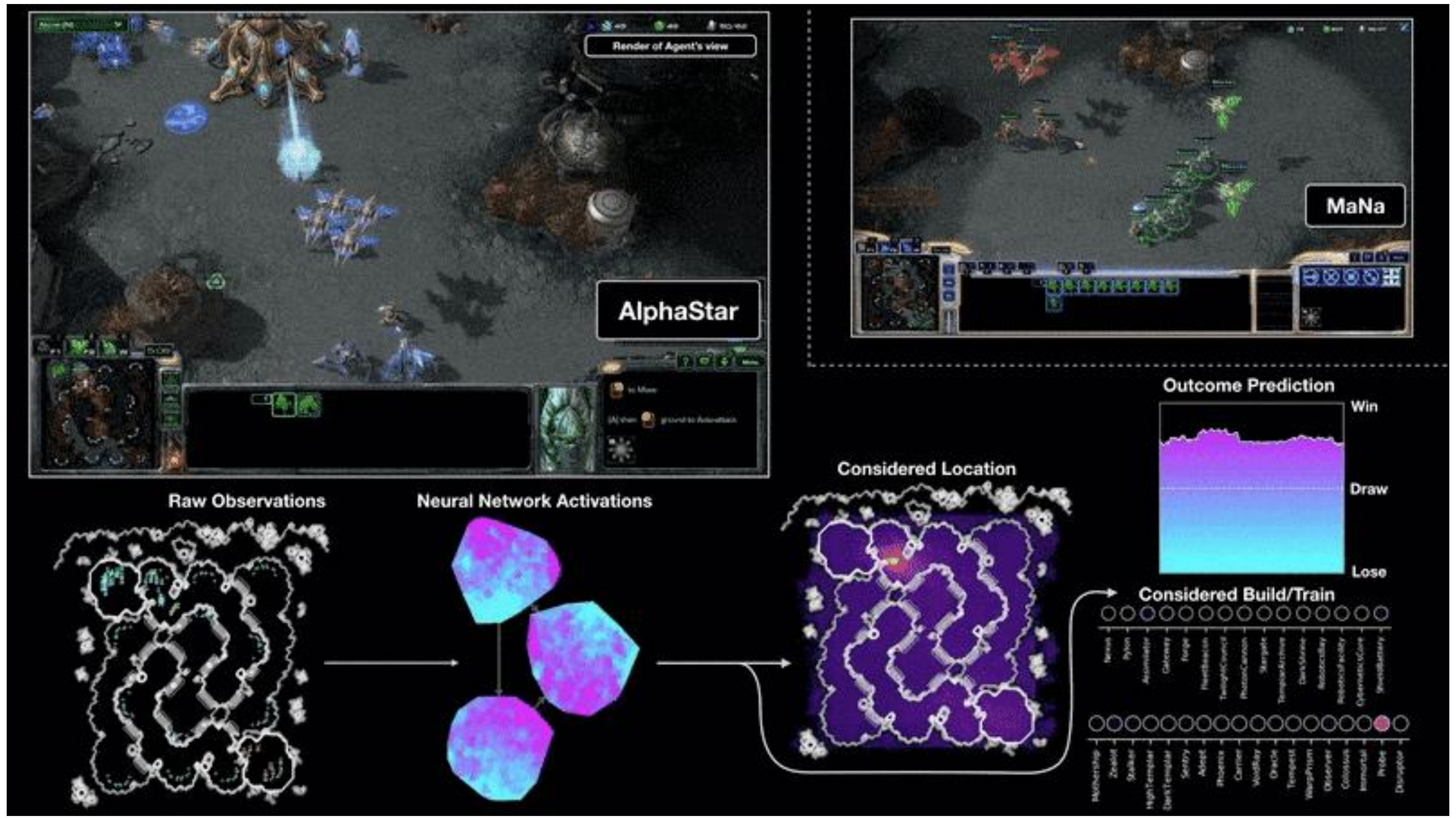


Emoticon suggestion



Machine translation

Deep Learning Today

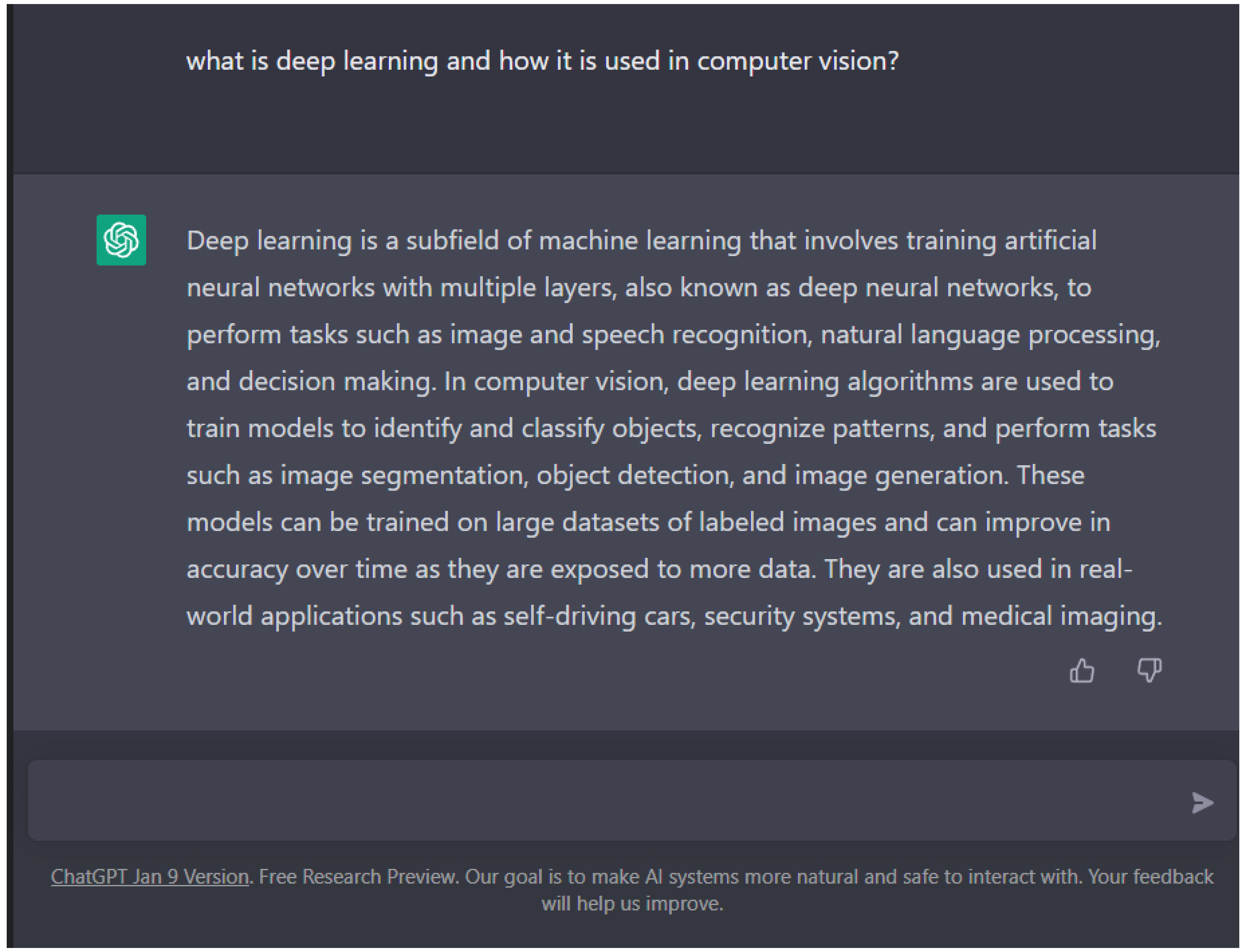


Alpha Star

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today



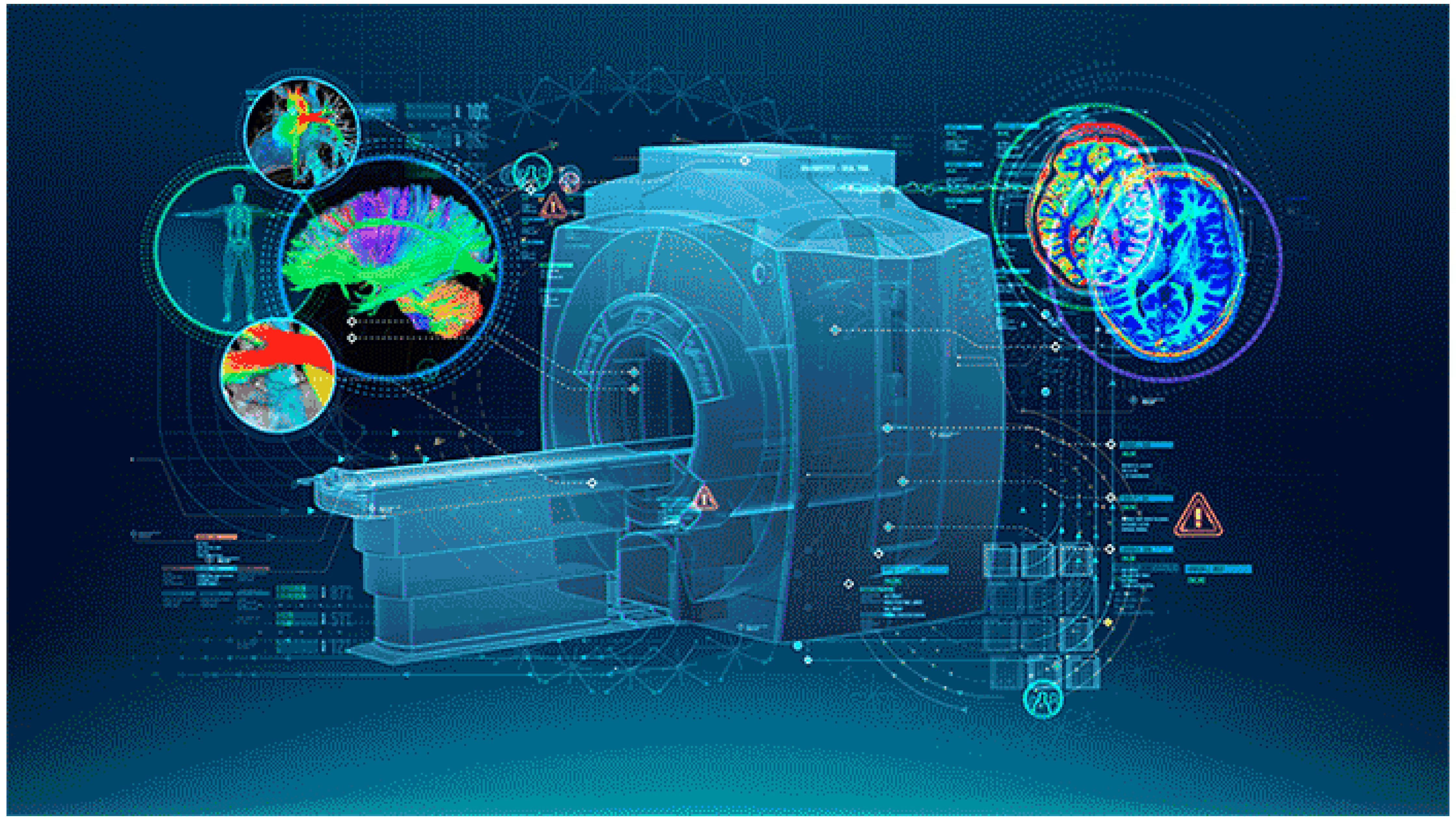
Automated Text Generation (ChatGPT)

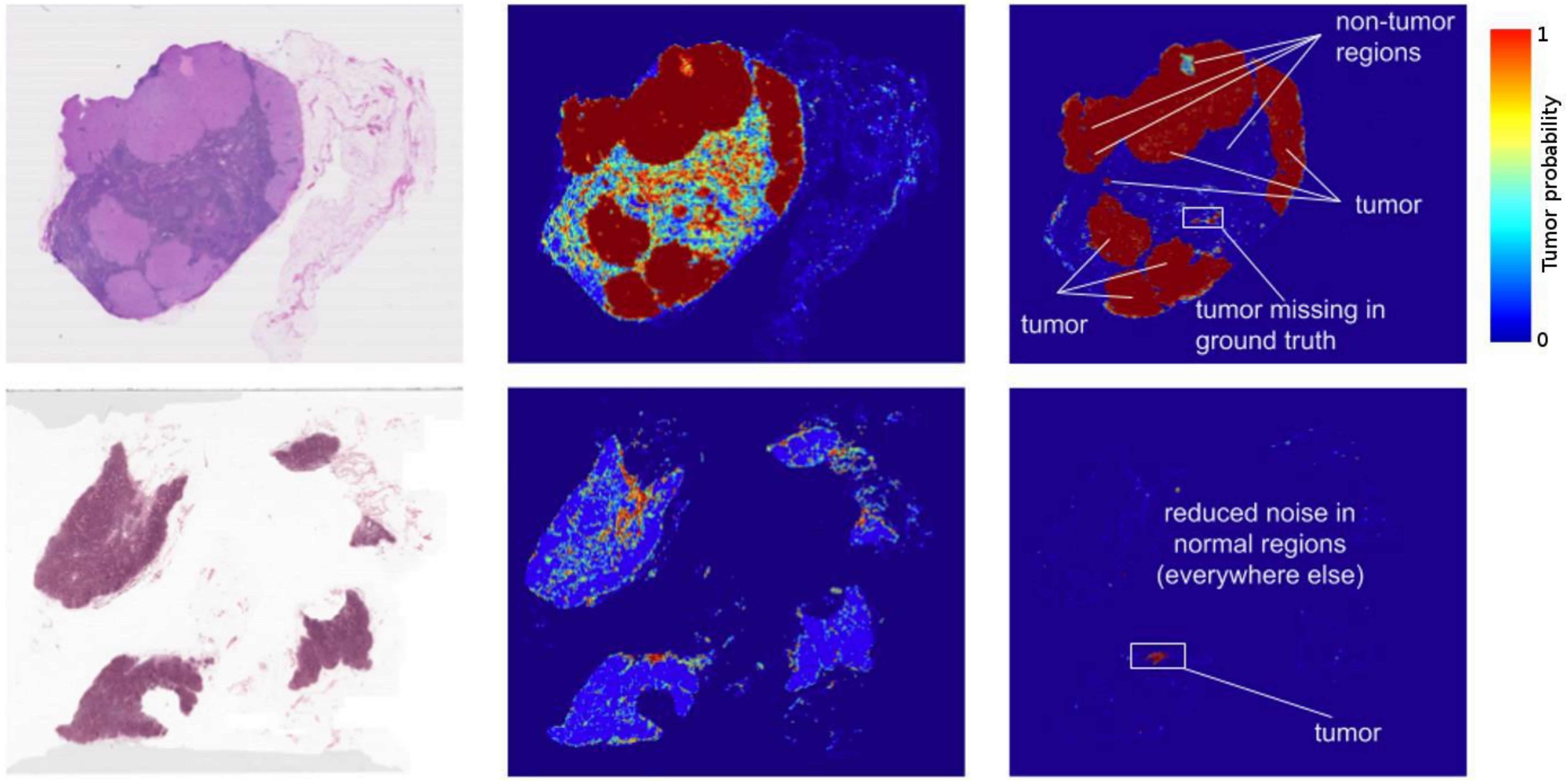
Google

Google Assistance

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today





Healthcare, cancer detection

Deep Learning Today: *Applications*

Image style transfer

Gatys et al. 2016. Image Style Transfer Using Convolutional Neural Networks. Proc. CVPR 2016.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today: *Applications*

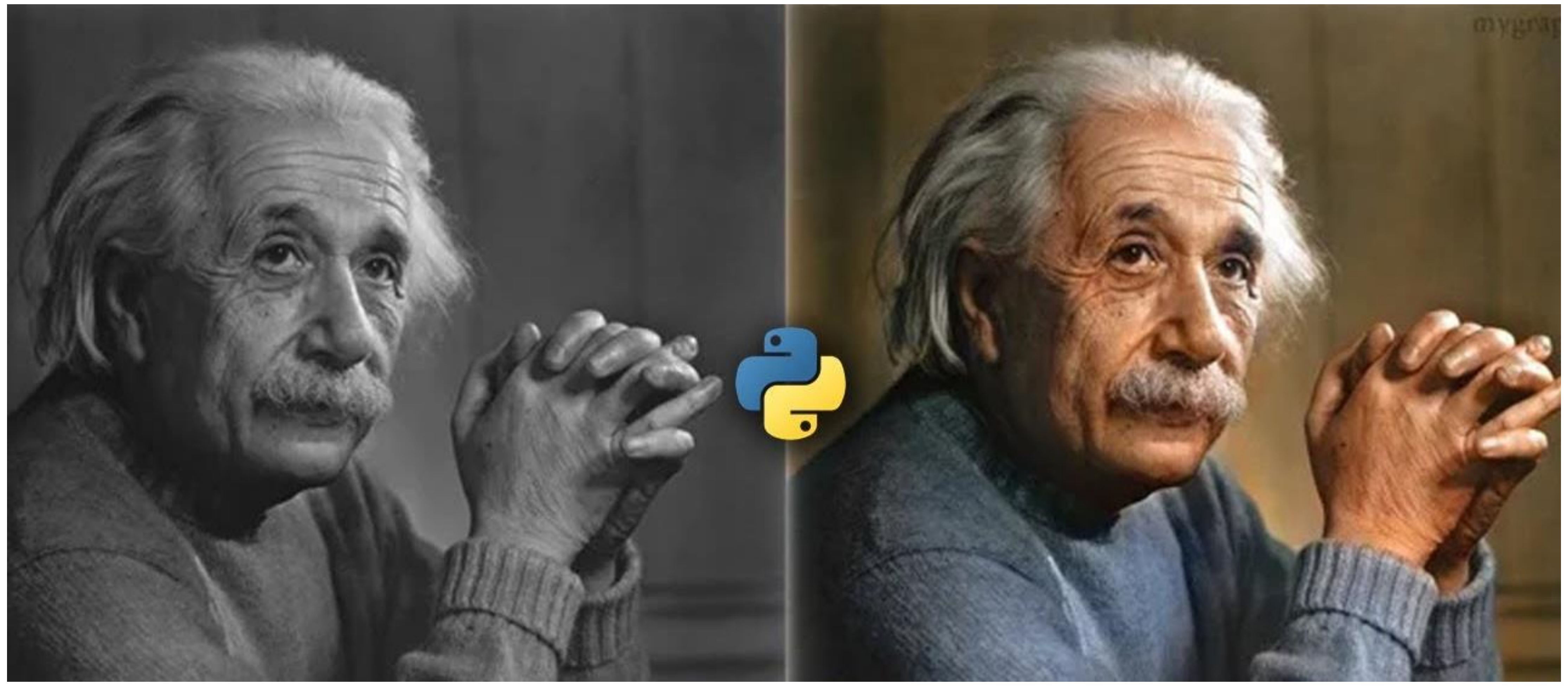
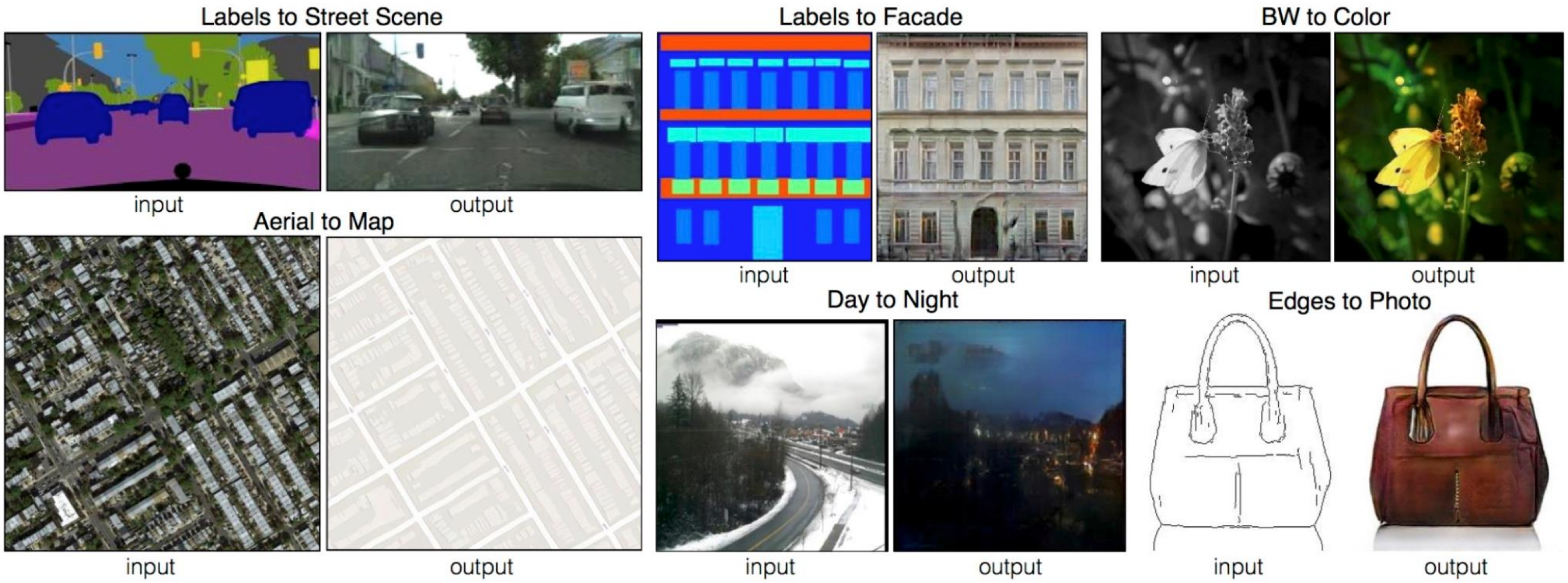


Image coloring

Deep Learning Today: *Applications*



This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today: *Applications*

DeepSketch2Face

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today: *Applications*

Denoising,
increase image resolution

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

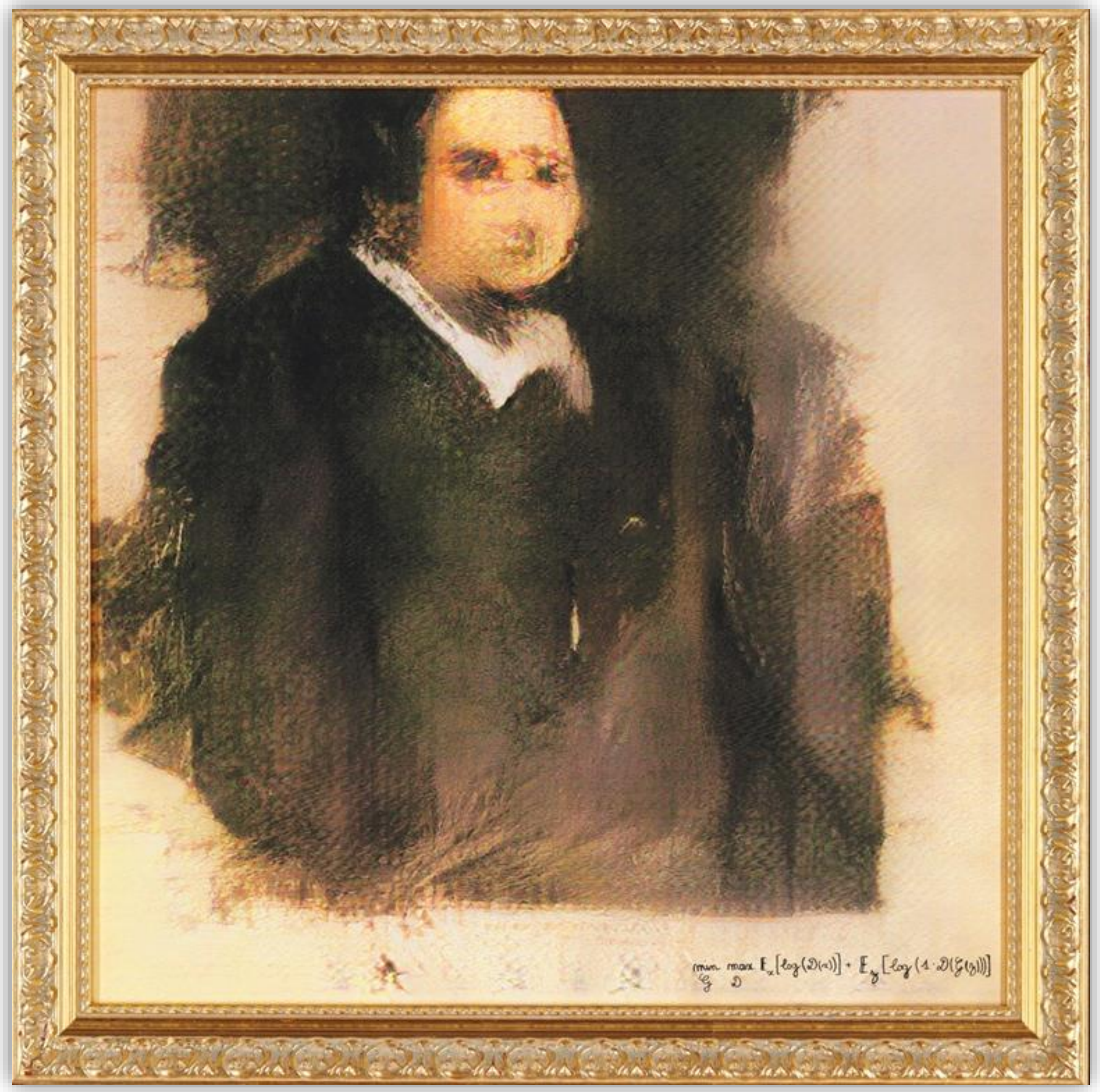
Deep Learning Today: *Applications*

Generative networks

Deep Learning Today: *Applications*

Obvious Art, “Katsuwaka of the Dawn Lagoon”, Image credits: Sotheby's.

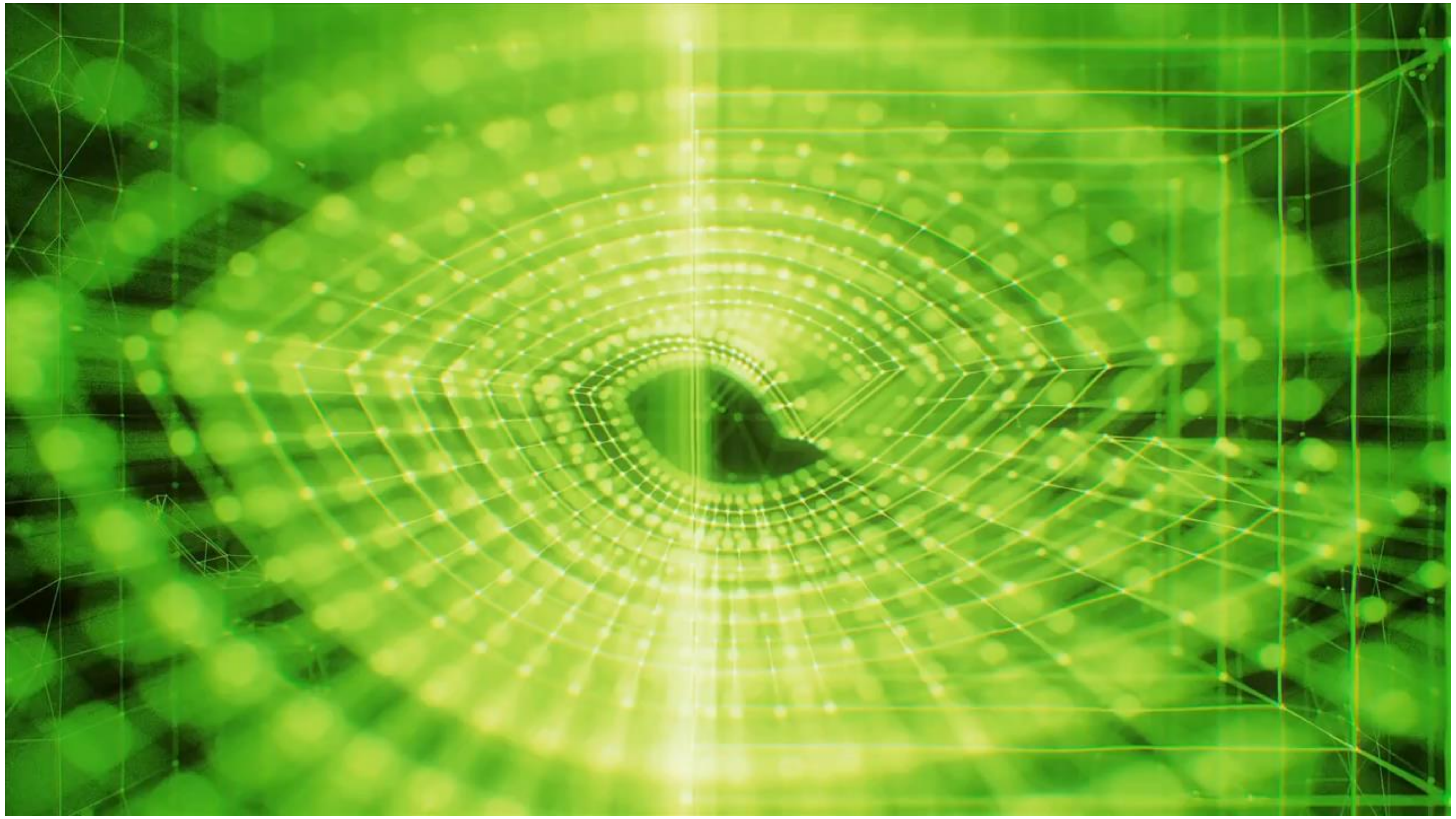
Portrait Le Baron De Belamy, Image credits: Sotheby's.



Portrait of Edmond Belamy, 2018, created by GANs (Generative Adversarial Networks). Sold for $432,500 on 25 October at Christie’s in New York. Image © Obvious

GAN involves feeding thousands of images of the same style into a computer until the machine concludes that it has created a new portrait that it thinks accurately reflects that style.

Deep Learning Today: *Applications*



https://thispersondoesnotexist.com/

Generative style transfer

Deep Learning Today: *Timeline of images generated by artificial intelligence*

2014

Goodfellow et al. (2014) – Generative Adversarial Networks

2015

Radford, Metz, and Chintala (2015) – Unsupervised Representation Learning with Deep Convolutional GANs

2016

Liu and Tuzel (2016) – Coupled GANs

2017

Karras et al. (2017) – Progressive Growing of GANs for Improved Quality, Stability, and Variation

2018

Karras, Laine, and Aila (2018) – A Style-Based Generator Architecture for Generative Adversarial Networks

2019

Karras et al. (2019) – Analyzing and Improving the Image Quality of StyleGAN

2020

Ho, Jain, & Abbeel (2020) – Denoising Diffusion Probabilistic Models

2021 Image generated with the prompt: *"a couple of people are sitting on a wood bench"*

Ramesh et al. (2021) – Zero-Shot Text-to-Image Generation (OpenAI's DALL-E 1)

2022 Image generated with the prompt: *"A Pomeranian is sitting on the King's throne wearing a crown. Two tiger soldiers are standing next to the throne."*

Saharia et al. (2022) – Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding (Google's Imagen)

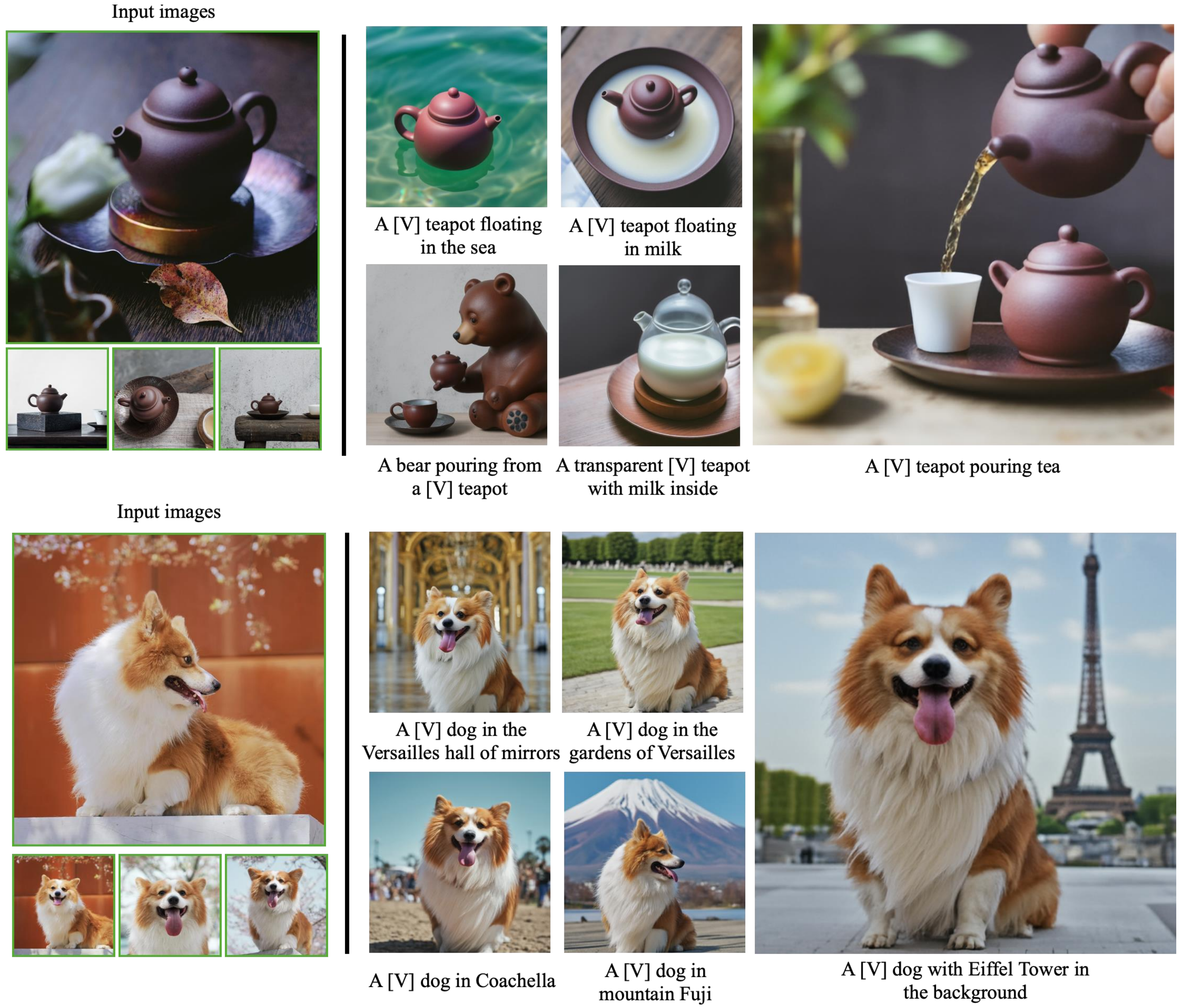
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today: *Applications*

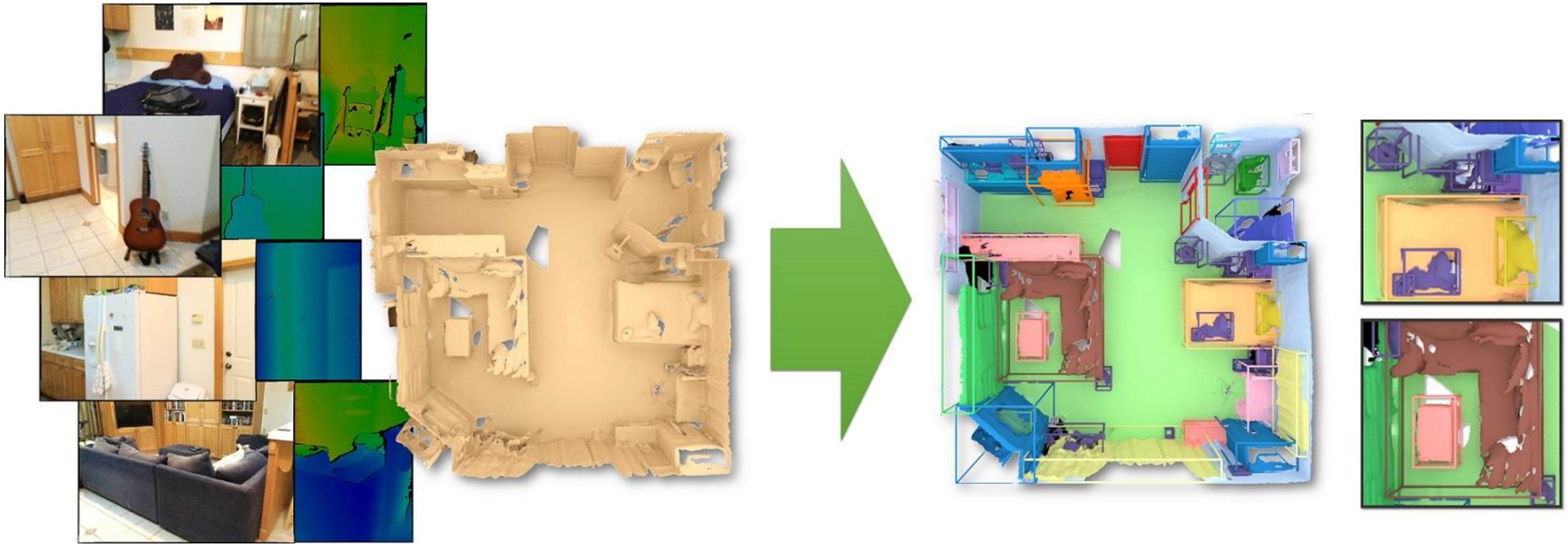
Sketch-Guided Text-to-Image

Deep Learning Today: *Applications*



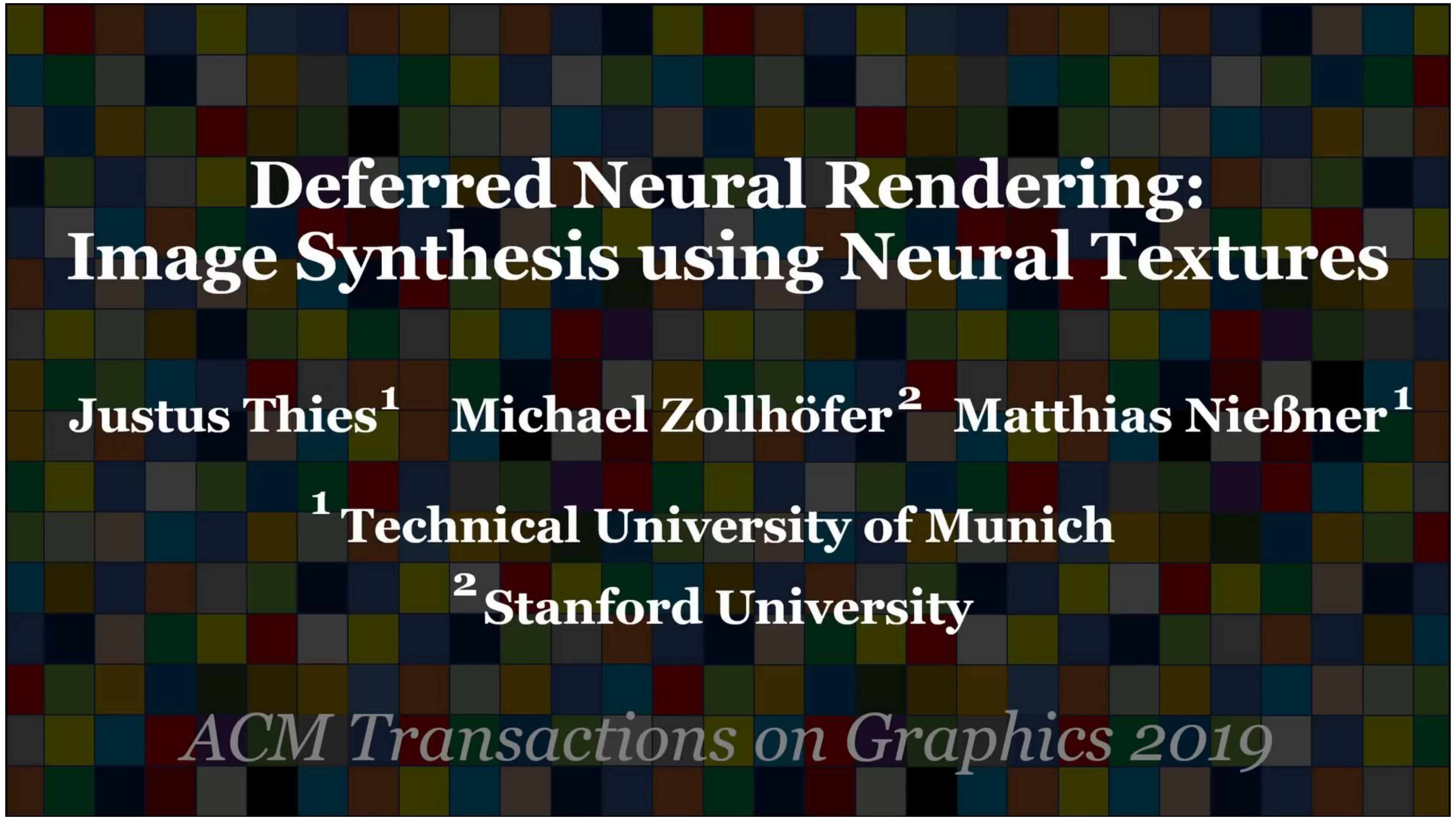
Text-driven content image generation [DreamBooth]

Deep Learning Today: *Applications*



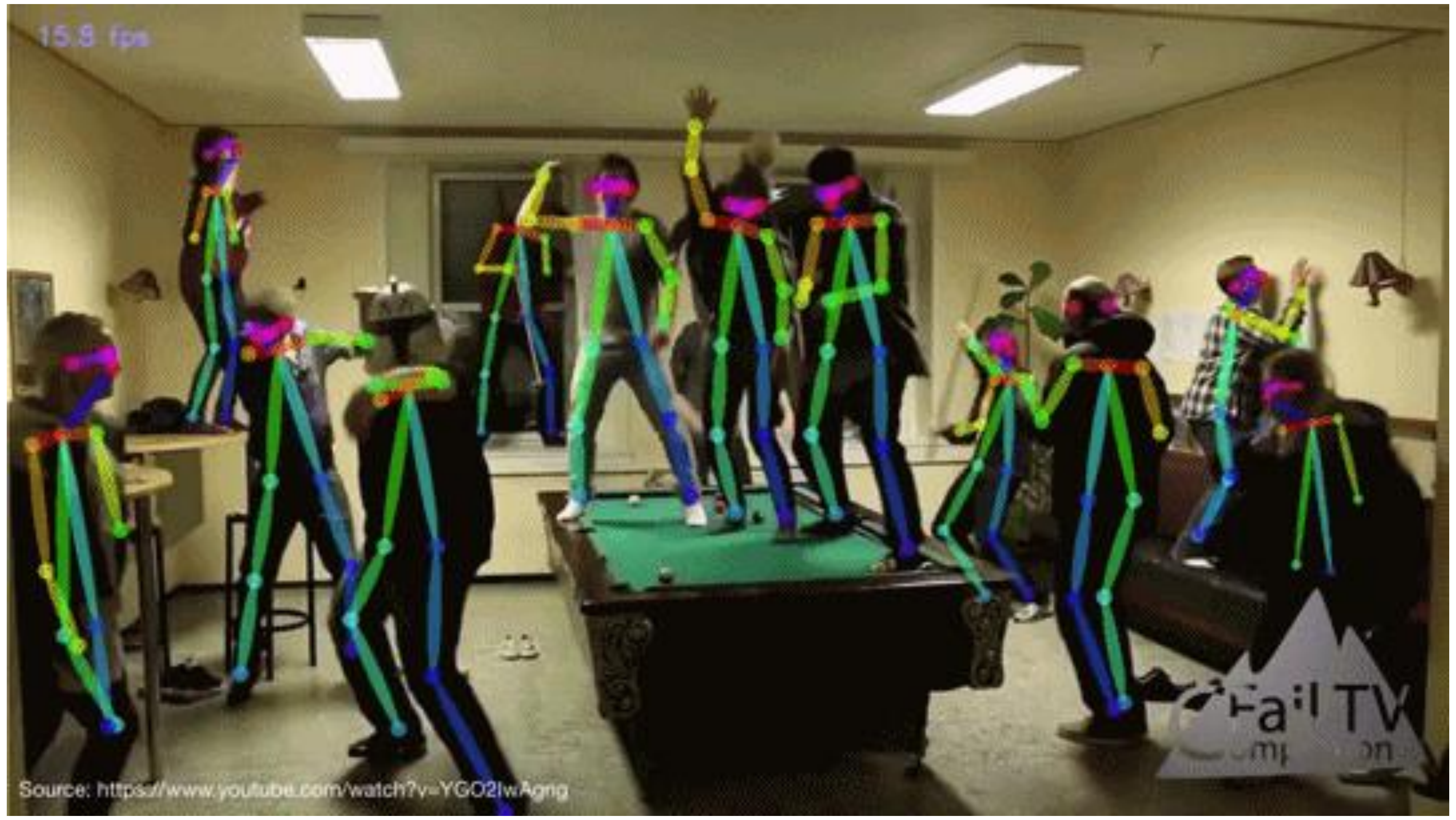
Scan and scene reconstruction

Deep Learning Today: *Applications*



Facial animation synthesis

Deep Learning Today: *Applications*



Pose detection in 2D

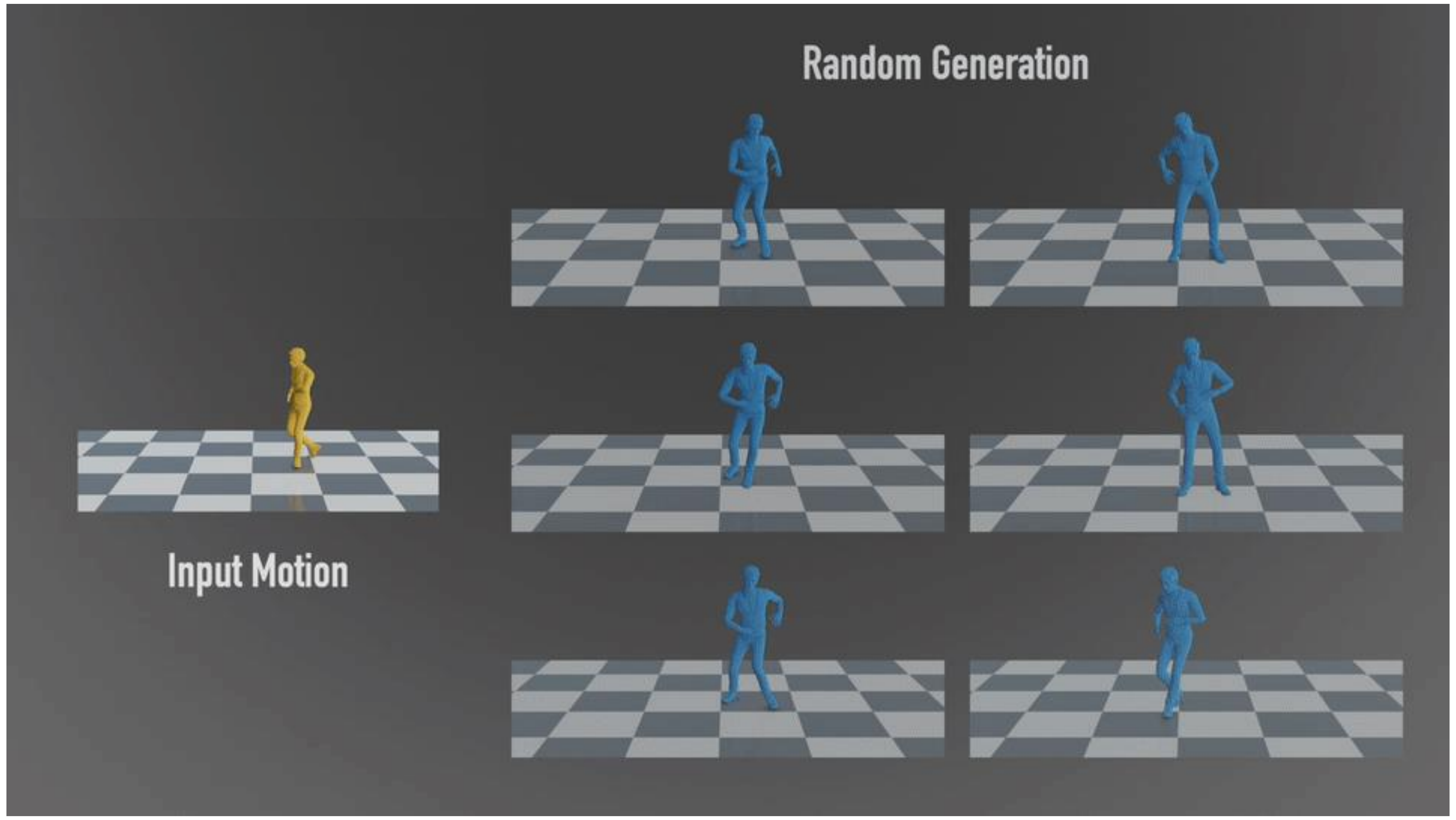
Co-financed by the European Union
Connecting Europe Facility

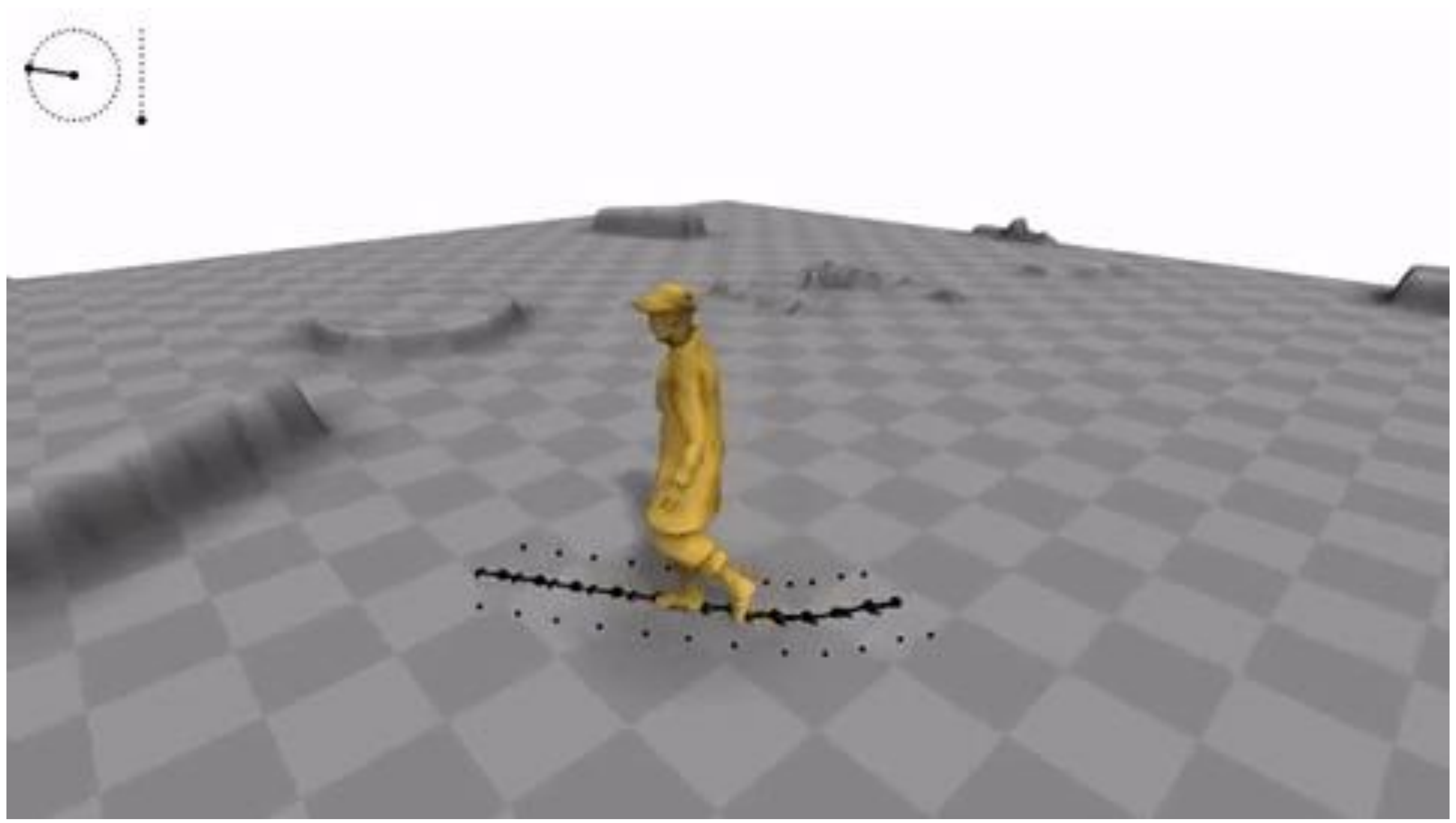
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today

Pose detection in 3D

Deep Learning Today





Motion Synthesis

Deep Learning Today

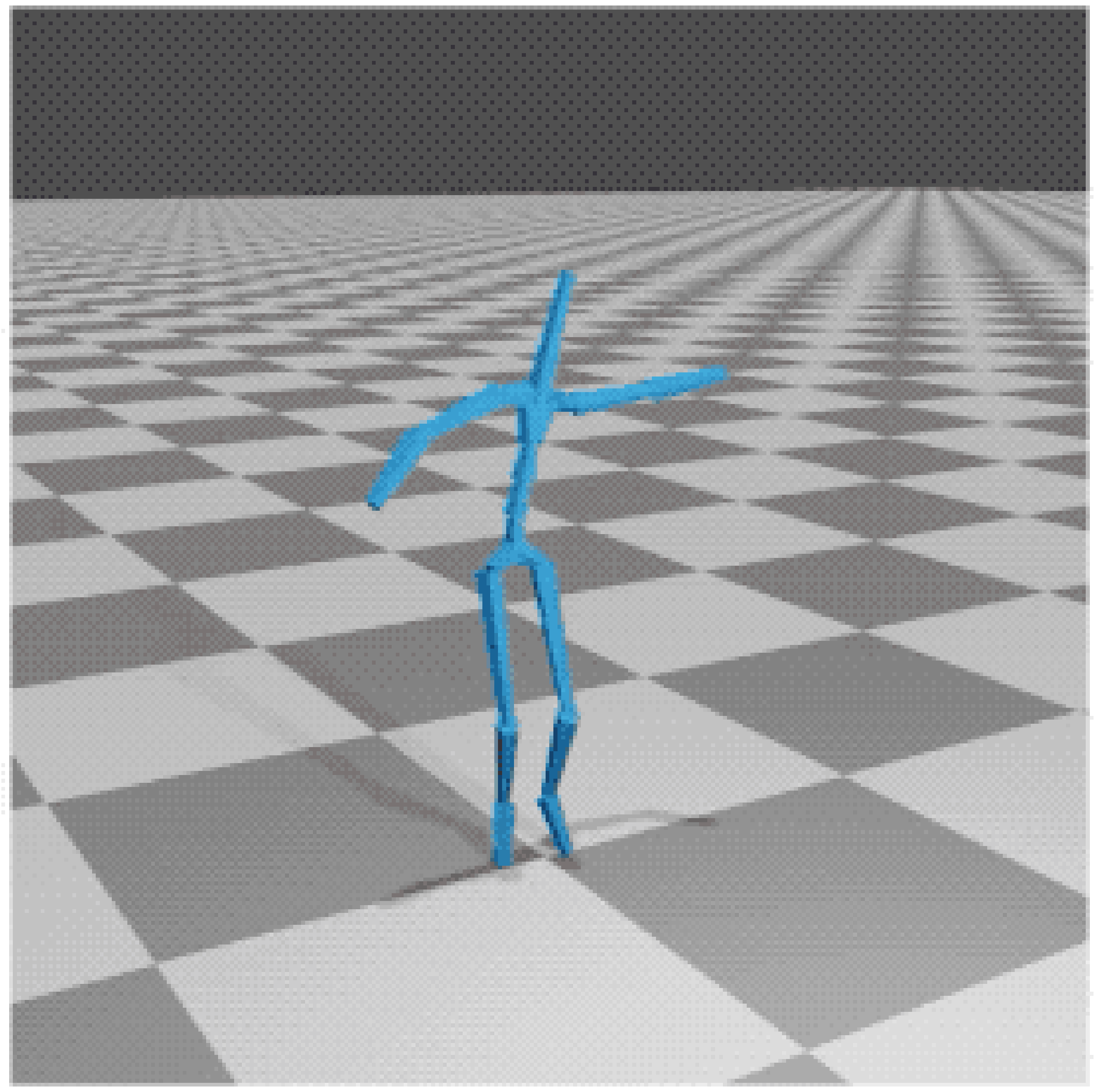
Motion Synthesis

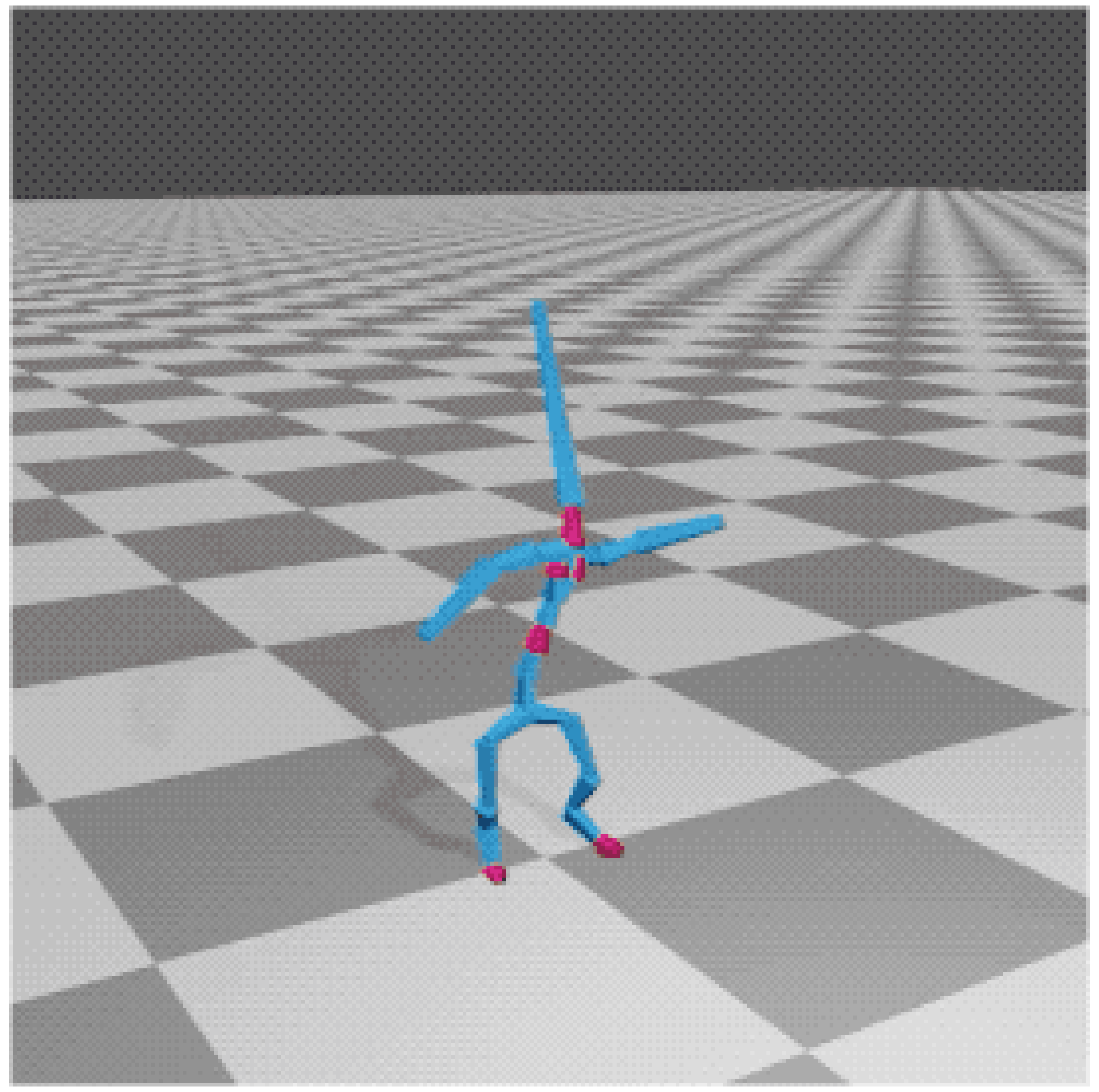
Deep Learning Today

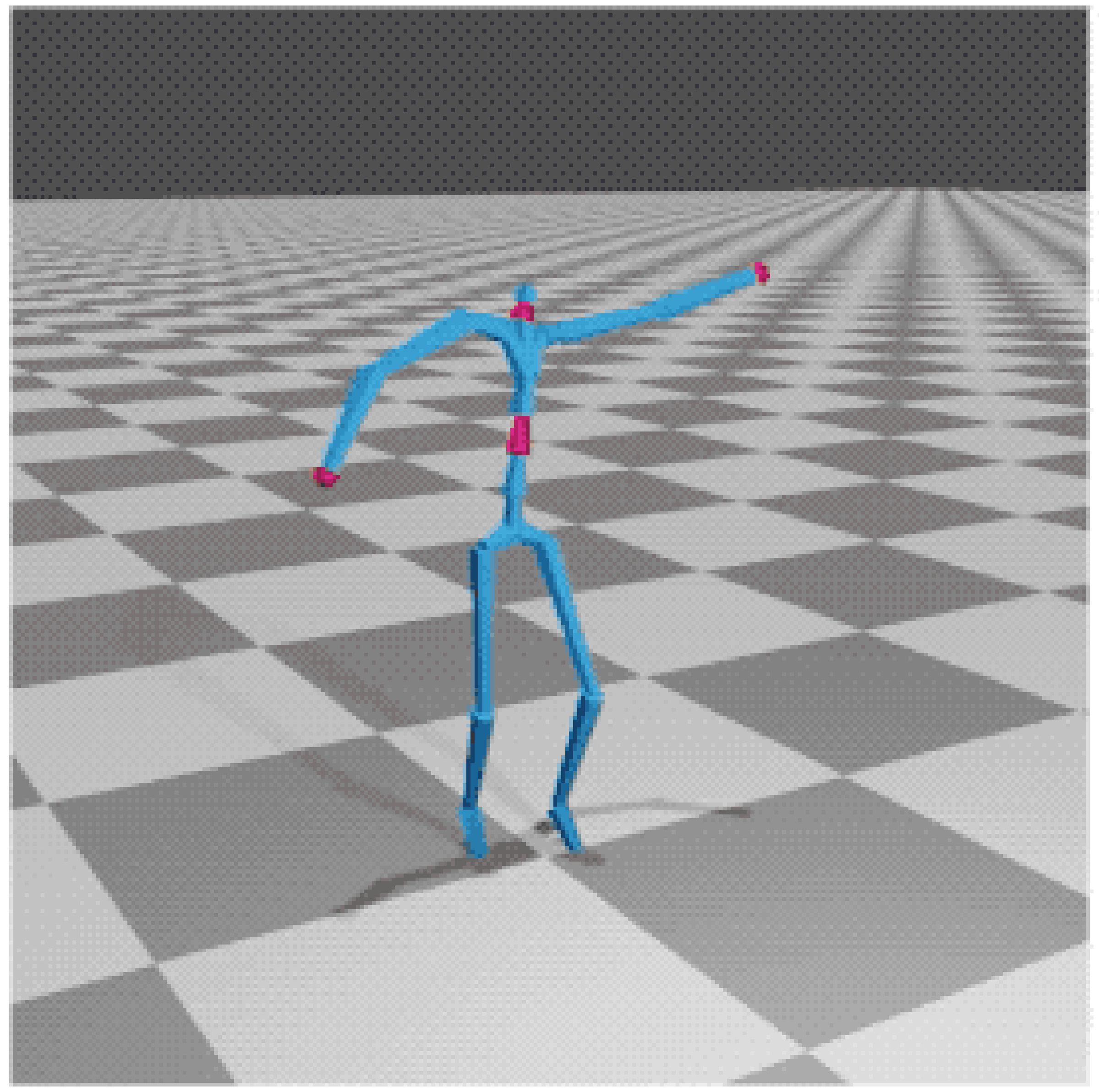



Motion style transfer

Deep Learning Today





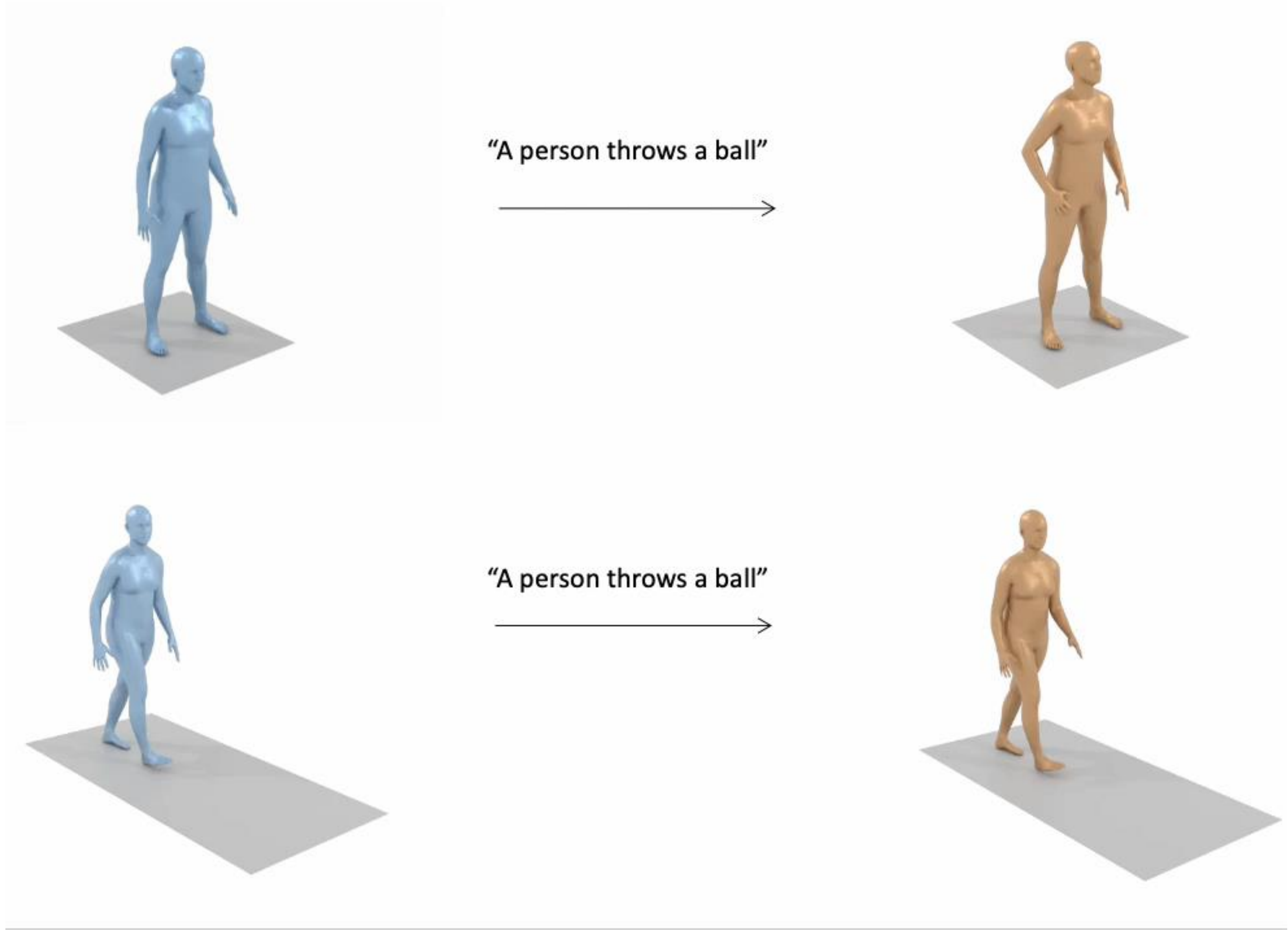


Motion Retargeting

Co-financed by the European Union
Connecting Europe Facility

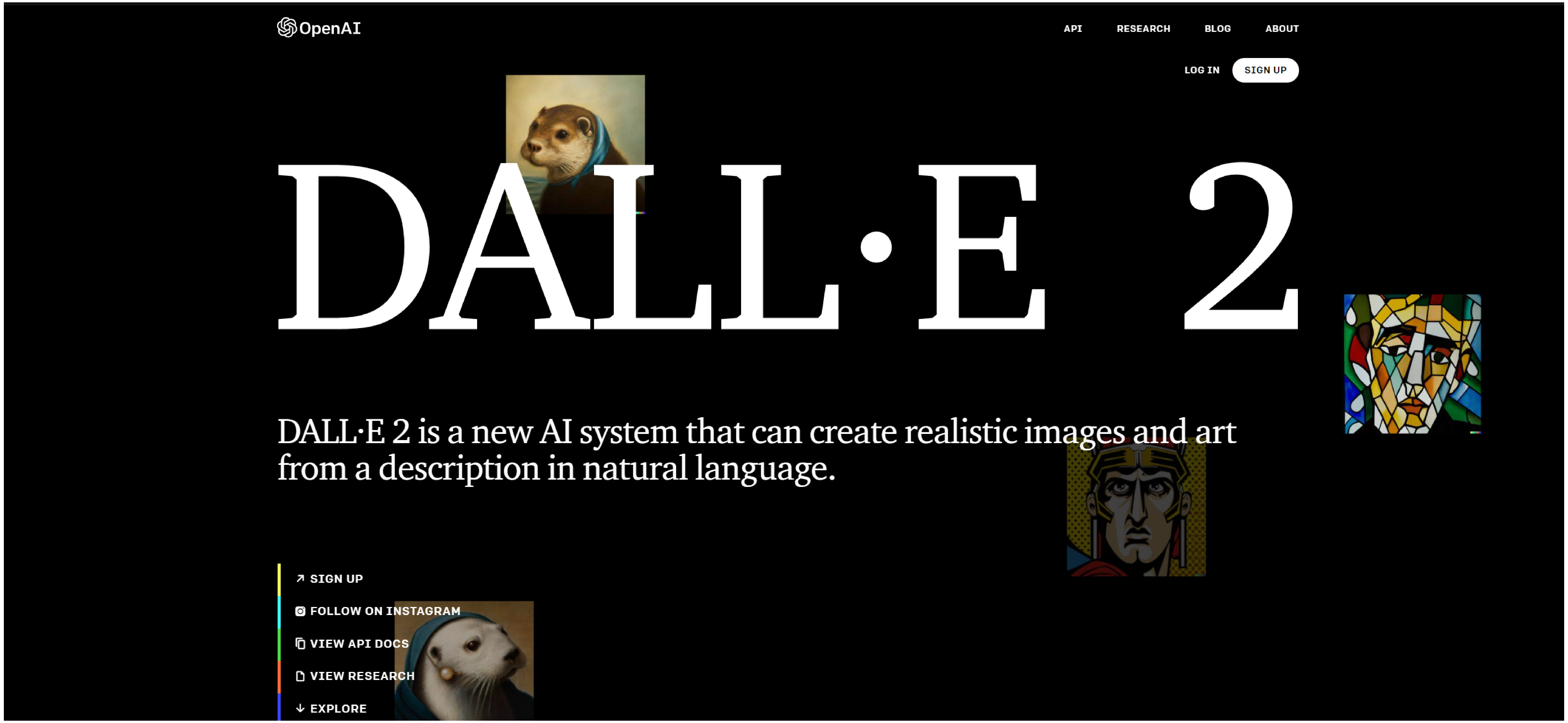
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Today



Text-to-Animation

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423



https://openai.com/dall-e-2/

Deep Learning Market

Deep learning market (in USD Billions)

Deep Learning Job Perspective

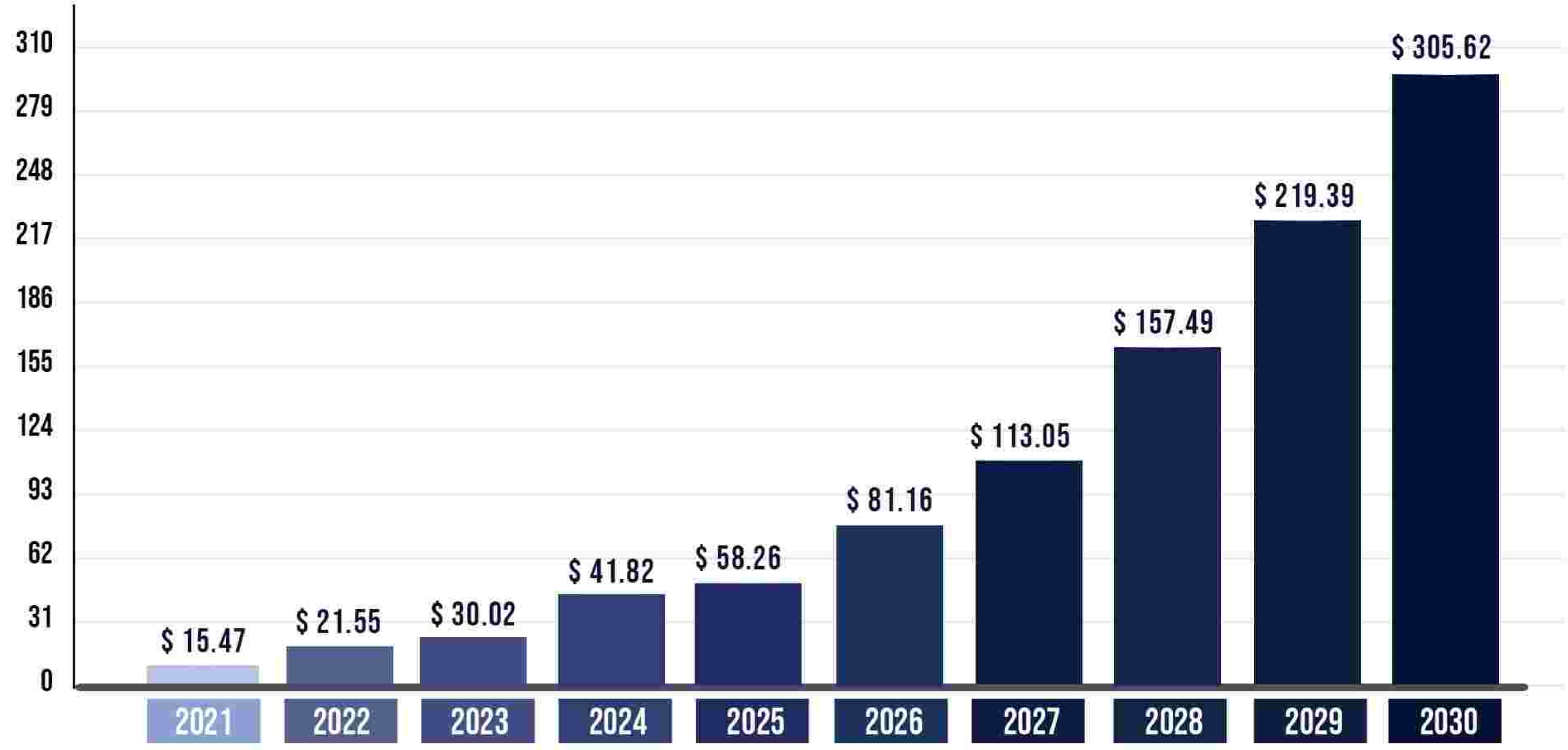
Excellent Job Perspectives!

- Automation requires ML/DL -> growth!
- Top-notch companies will gladly hire you!

Many industries now:

- IT-Companies
- Cars, Logistic, Health Care, etc…
- Manufacturing / Robotics, etc…

Deep Learning Market



Deep learning market (in US$ Billions)

But: Also Challenging!

High level understanding is not enough

- Need proper theory background
- Need proper practical skillsets

Can be competitive!

- Many good people
- Downloading scripts / running code not enough
- Deeper understanding often requires PhDs

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

How can we stay current?

This field moves fast, and papers get old on a timescale of *months*, so keeping your finger on the pulse can be tough.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Deep Learning Memes

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank you!

See you next week

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423