AI Seminars: 2022

Computer Vision: Deep Learning Power and Geometry Wisdom

Giacomo Boracchi, Luca Magri, Federica Arrigoni

DEIB- Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano
December 6th 2022

Expressive photo of a 3D scene of a building in daylight low angle, extreme wide shot, high detail.
Created by DALL-E 2

POLITECNICO MILANO 1863

ellis European Laboratory for Learning and Intelligent Systems

AIRIC Artificial Intelligence Research and Innovation Center

Cefriel

Laya

The Team

3 faculties, 7 PhD students, 2 Research Assistants

Past Students

Giacomo Boracchi

Luca Magri (Researcher)

Federica Arrigoni (Researcher)

Antonino Rizzo

Michele Craighero

Giuseppe Bertolini

Diego Carrera

Filippo Leveni

Diego Stucchi

Loris Giulivi

Andrea Porfiri Dal Cin

Riccardo Margheritti

Edoardo Peretti

Luca Frittoli

..on top of 20+ MSc students

Computer Vision

An interdisciplinary scientific field that deals with ***how computers can*** *be made to* ***gain high-level understanding*** *from digital images or videos*

Visual Recognition: Understanding Image Semantic

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

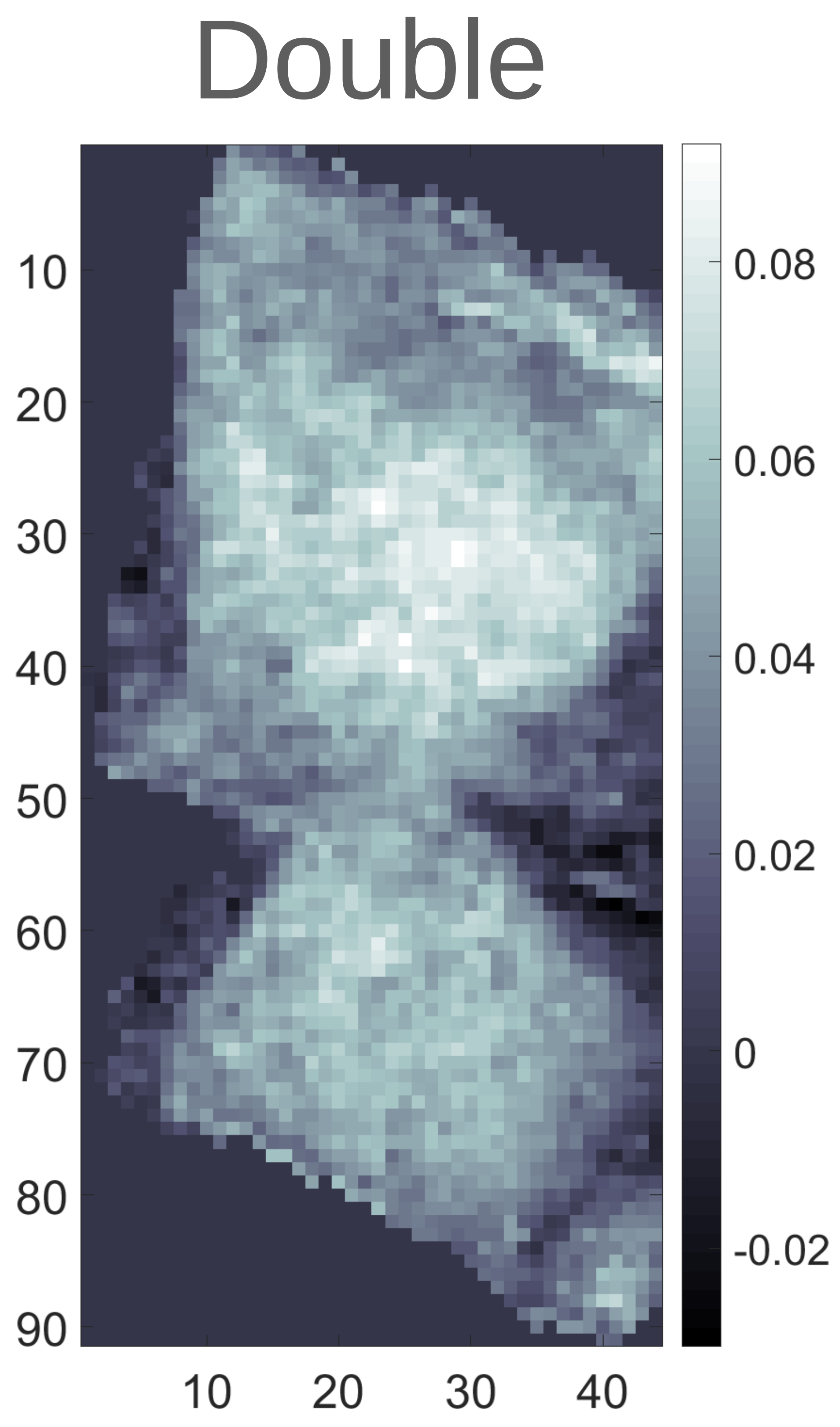
Image Classification

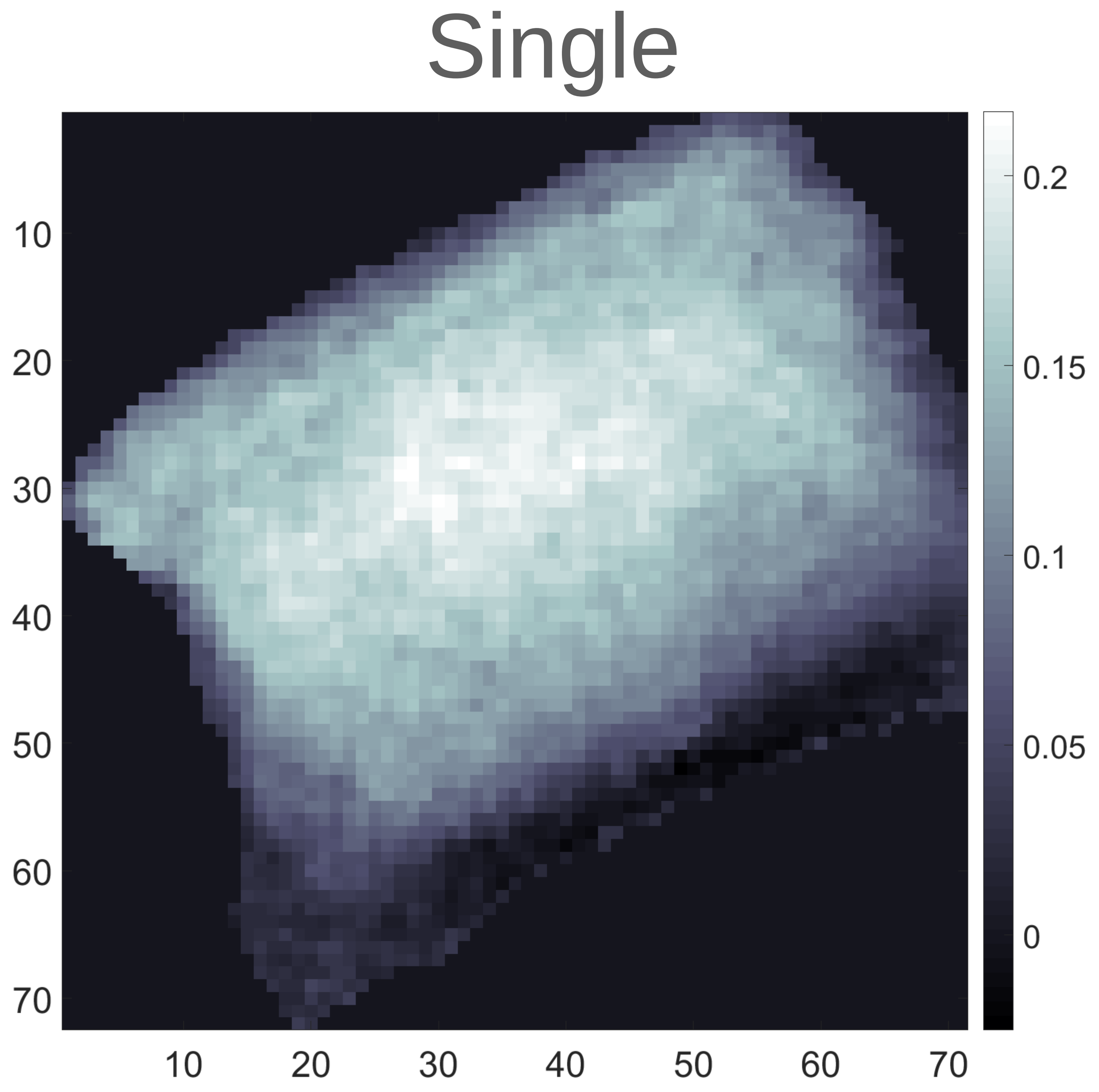
Before The Deep Learning Revolution

An Illustrative Example: Classification of Depth Images

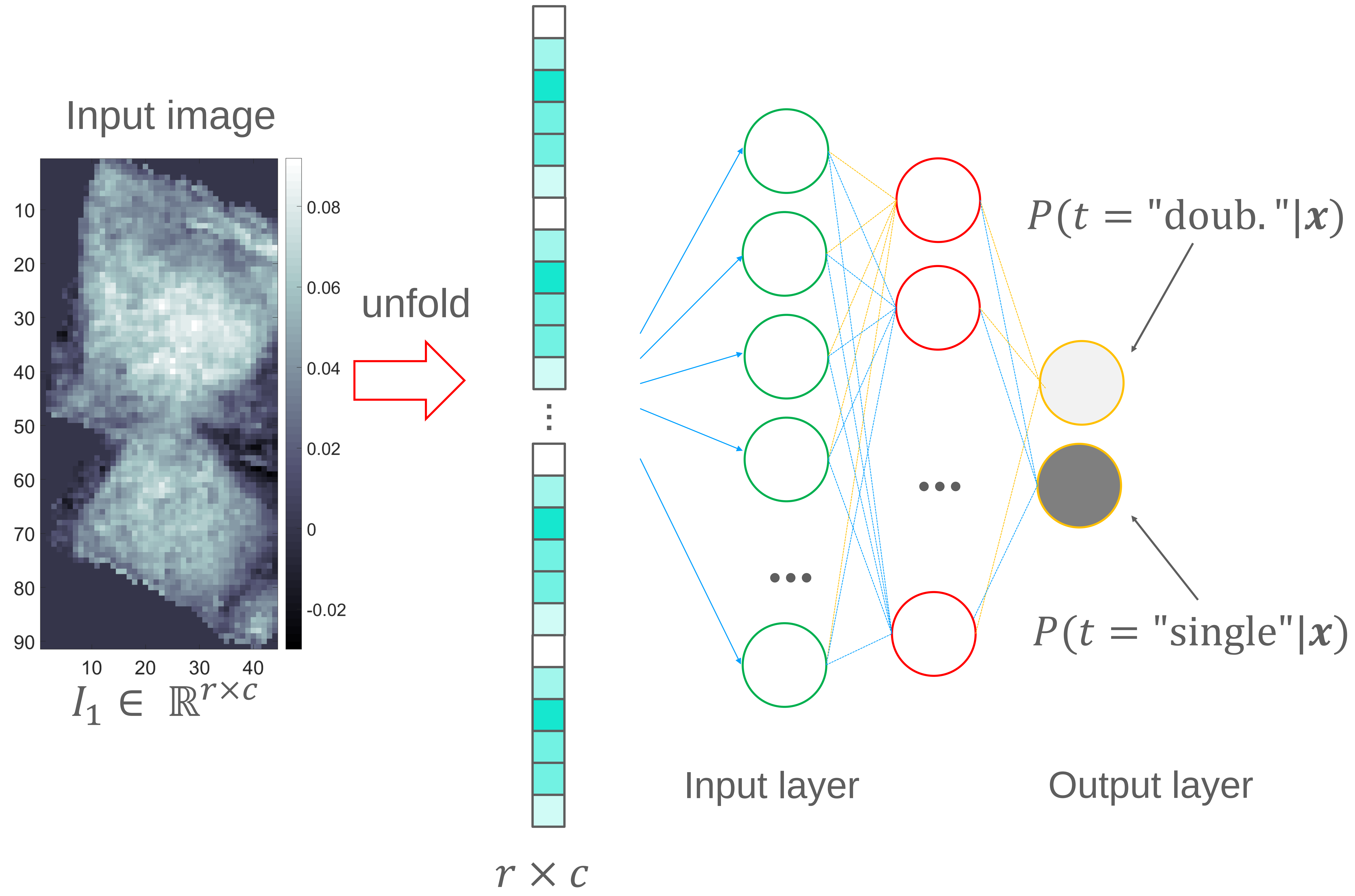
The system should automatically take corrective actions when there are two objects.

This is a **binary classification** problem!

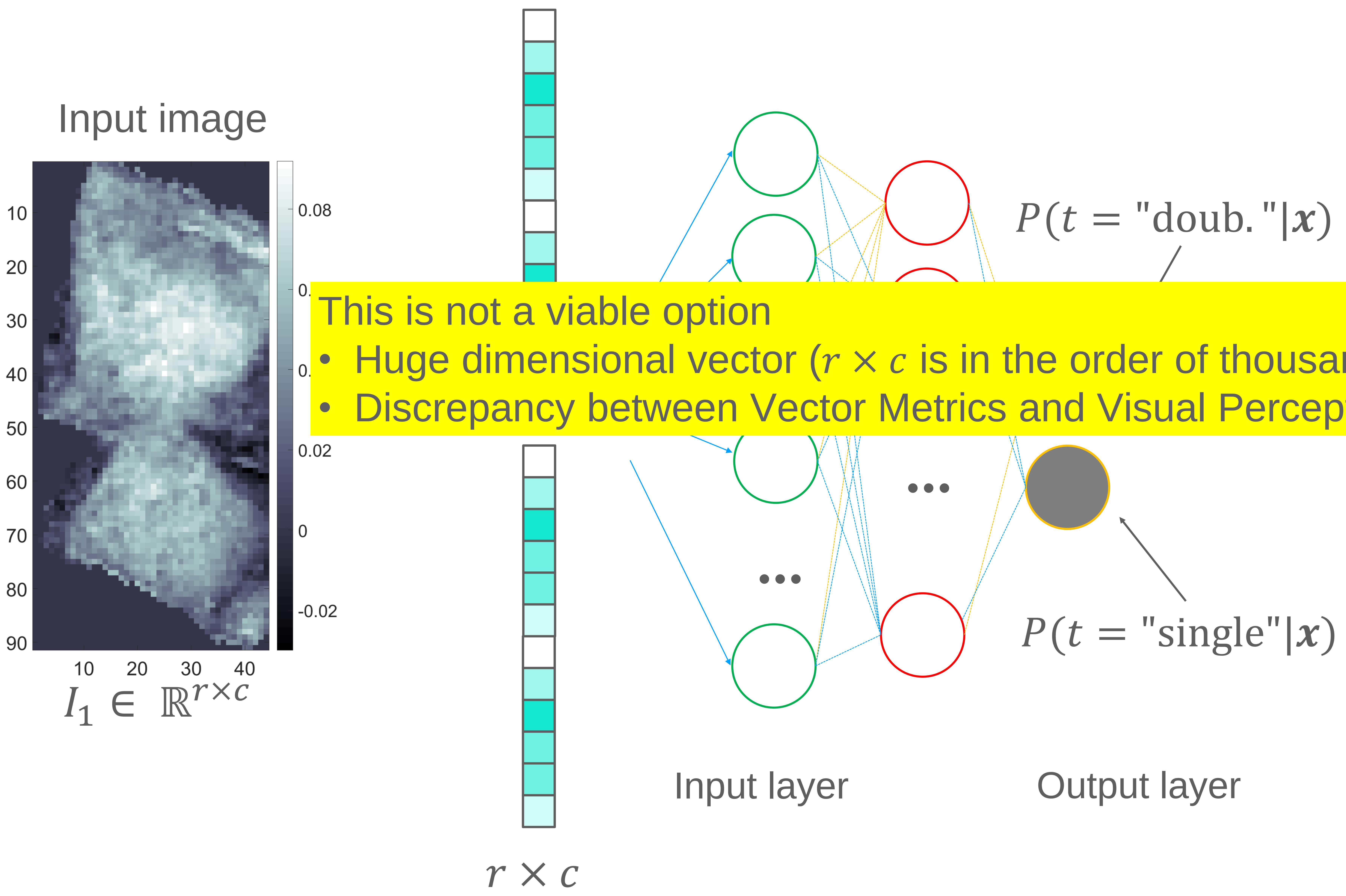




Images: a Difficult Input for a Traditional Classifier



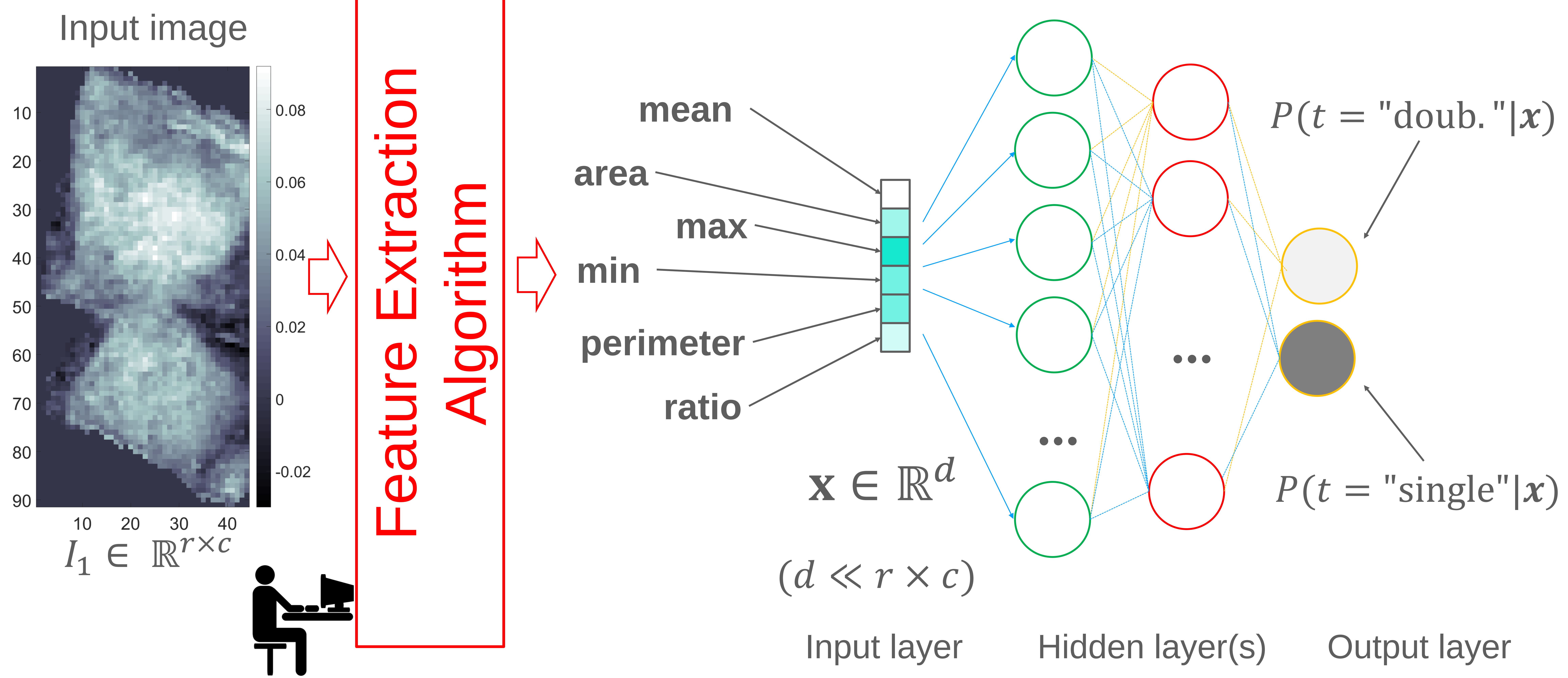
Images: a Difficult Input for a Traditional Classifier



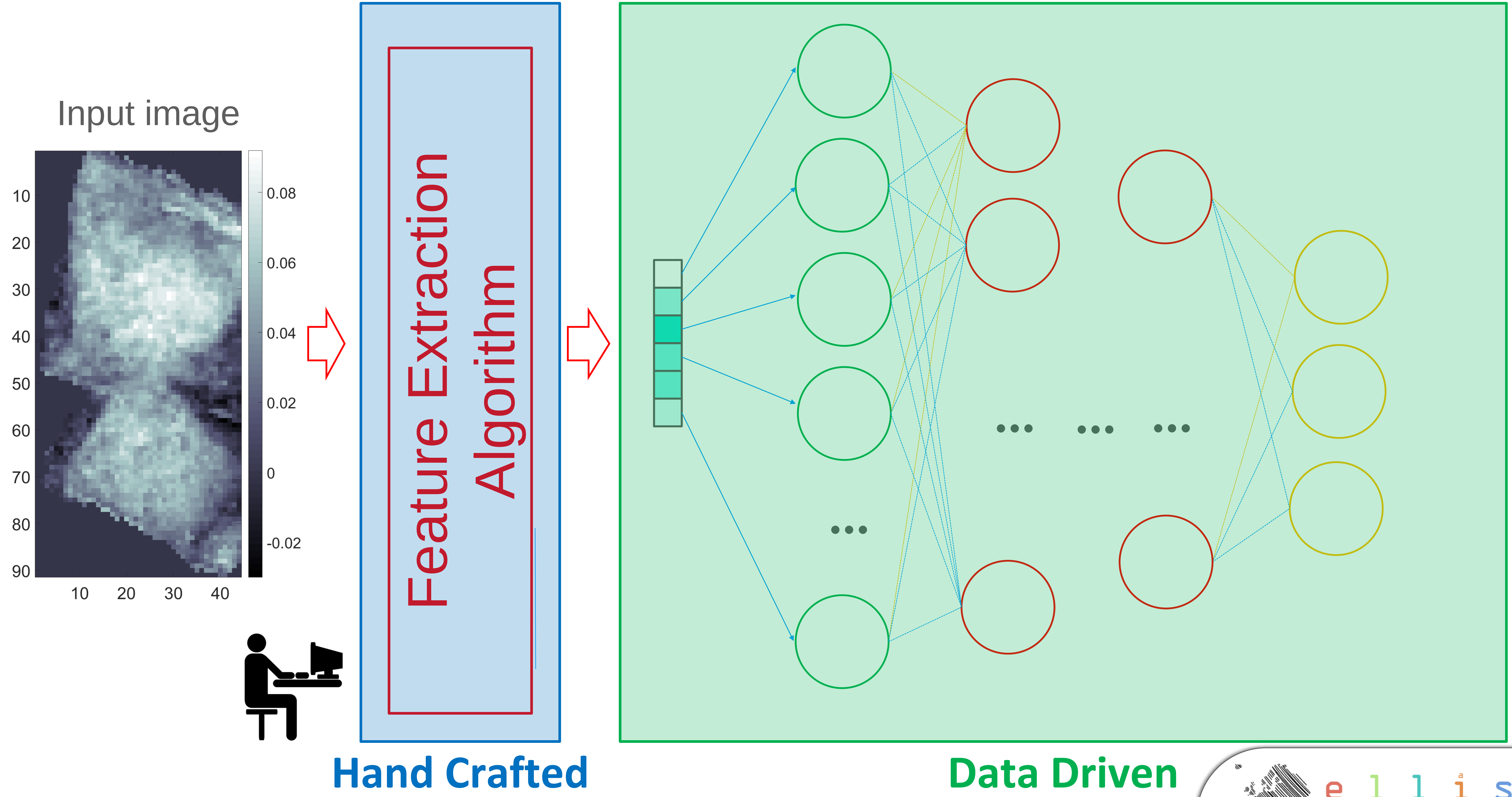
This is not a viable option

- Huge dimensional vector ($r \times c$ is in the order of thousands / Million samples)
- Discrepancy between Vector Metrics and Visual Perception

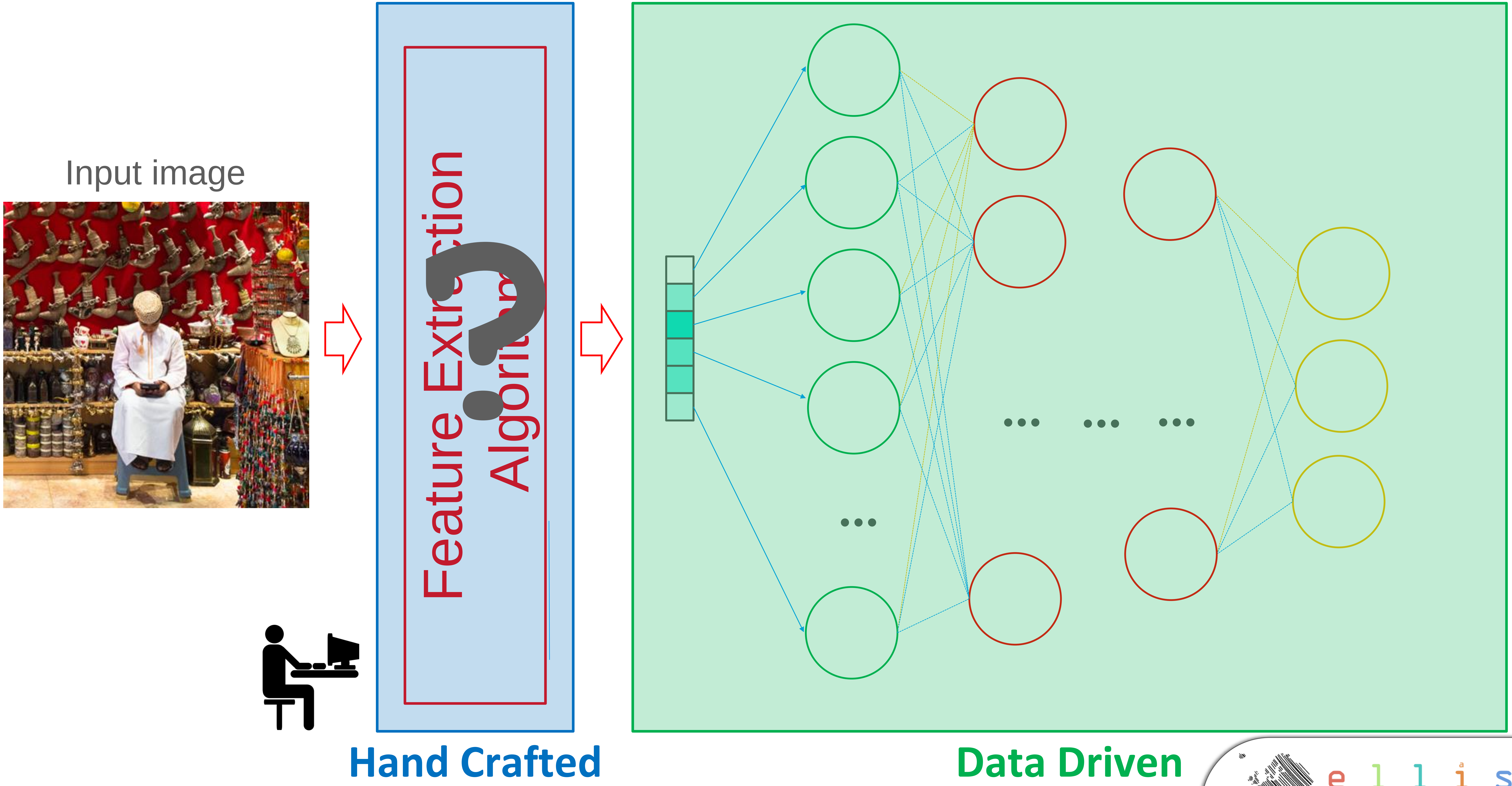
Hand Crafted Features



A Viable Approach in Very Controlled Scenarios

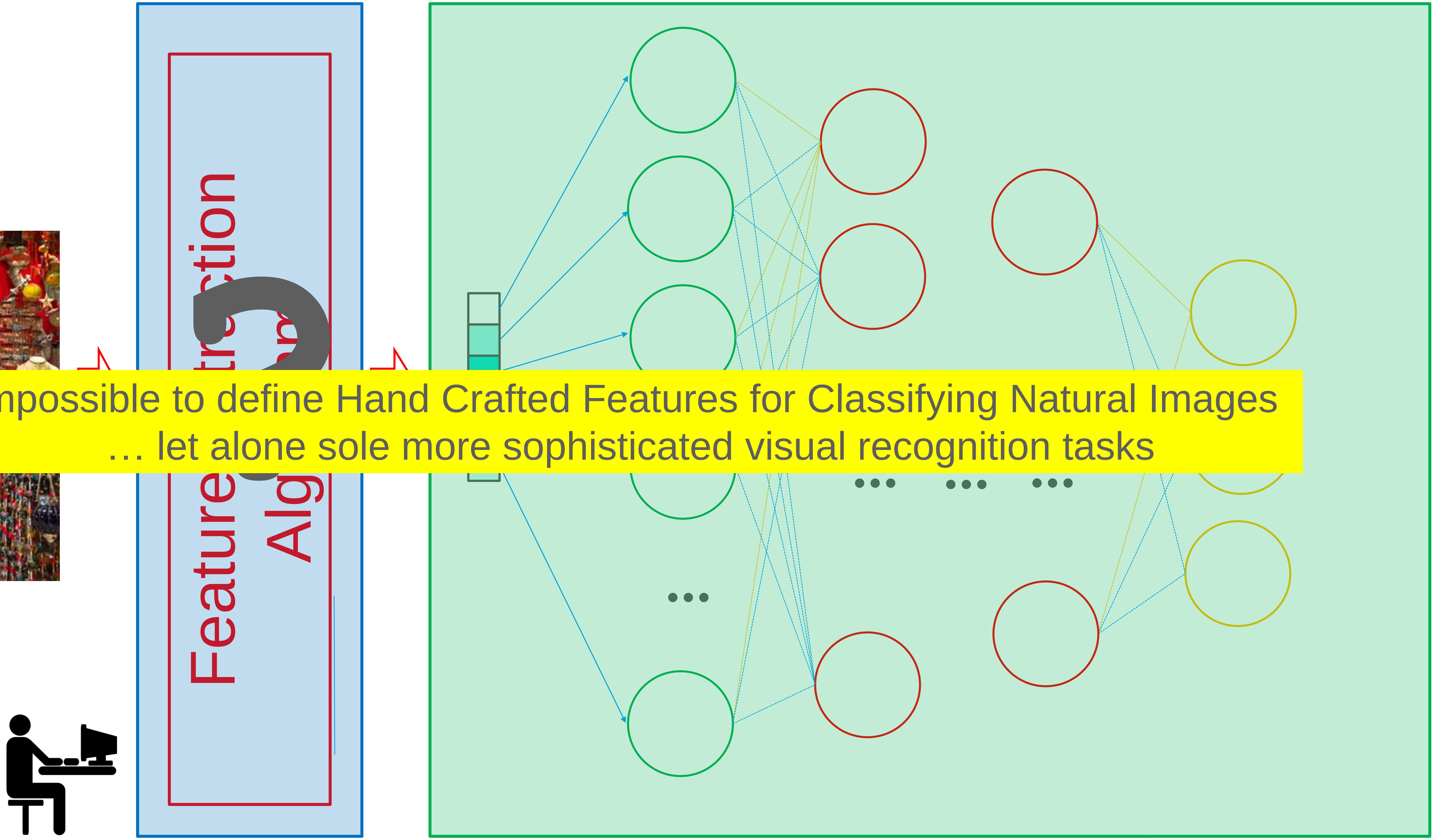


Hand-Crafted Features fall Short on Natural Images…



Hand-Crafted Features fall Short on Natural Images…

Input image



Impossible to define Hand Crafted Features for Classifying Natural Images
… let alone sole more sophisticated visual recognition tasks

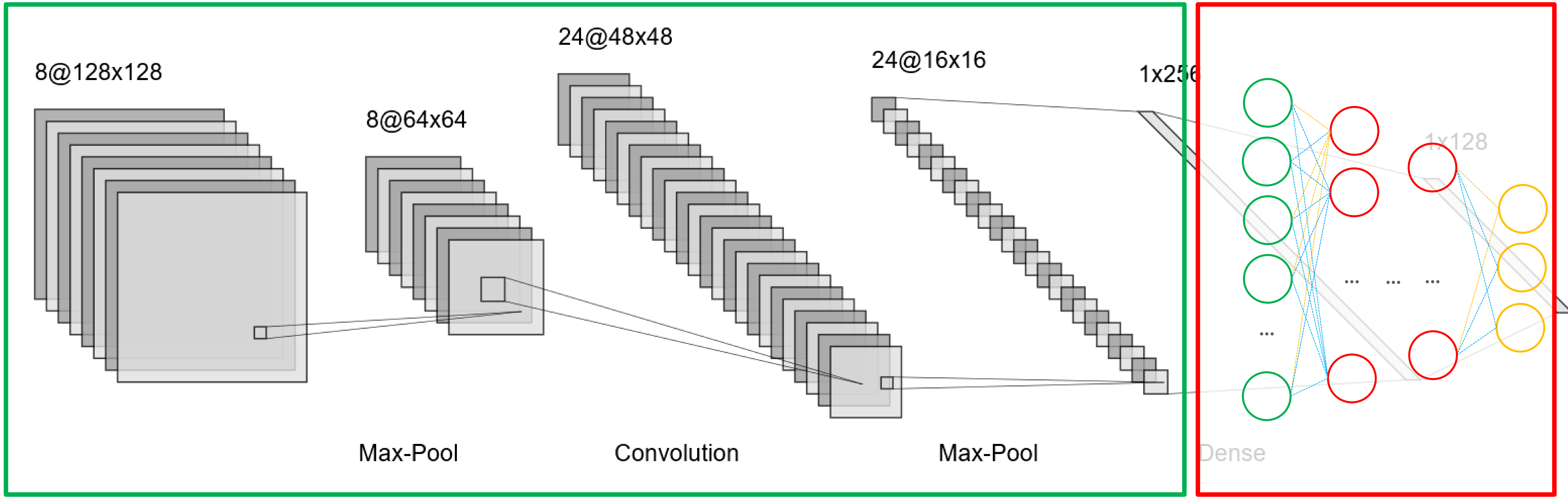
Hand Crafted **Data Driven**

The Deep Learning Revolution

… a new perspective for handling images

The typical architecture of a CNN

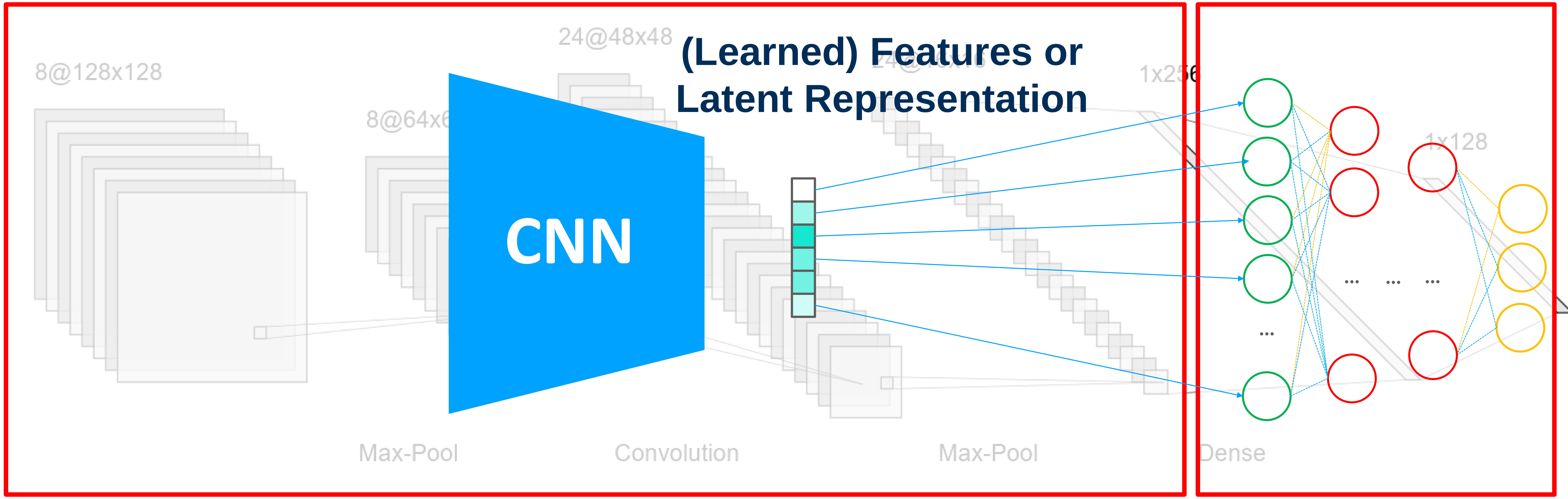
The typical architecture of a CNN



Data-Driven Feature extraction

Feature Classification

The typical architecture of a CNN



Data-Driven Feature extraction

Feature Classification

Typically, to learn meaningful representations, many layers are required

The network becomes deep

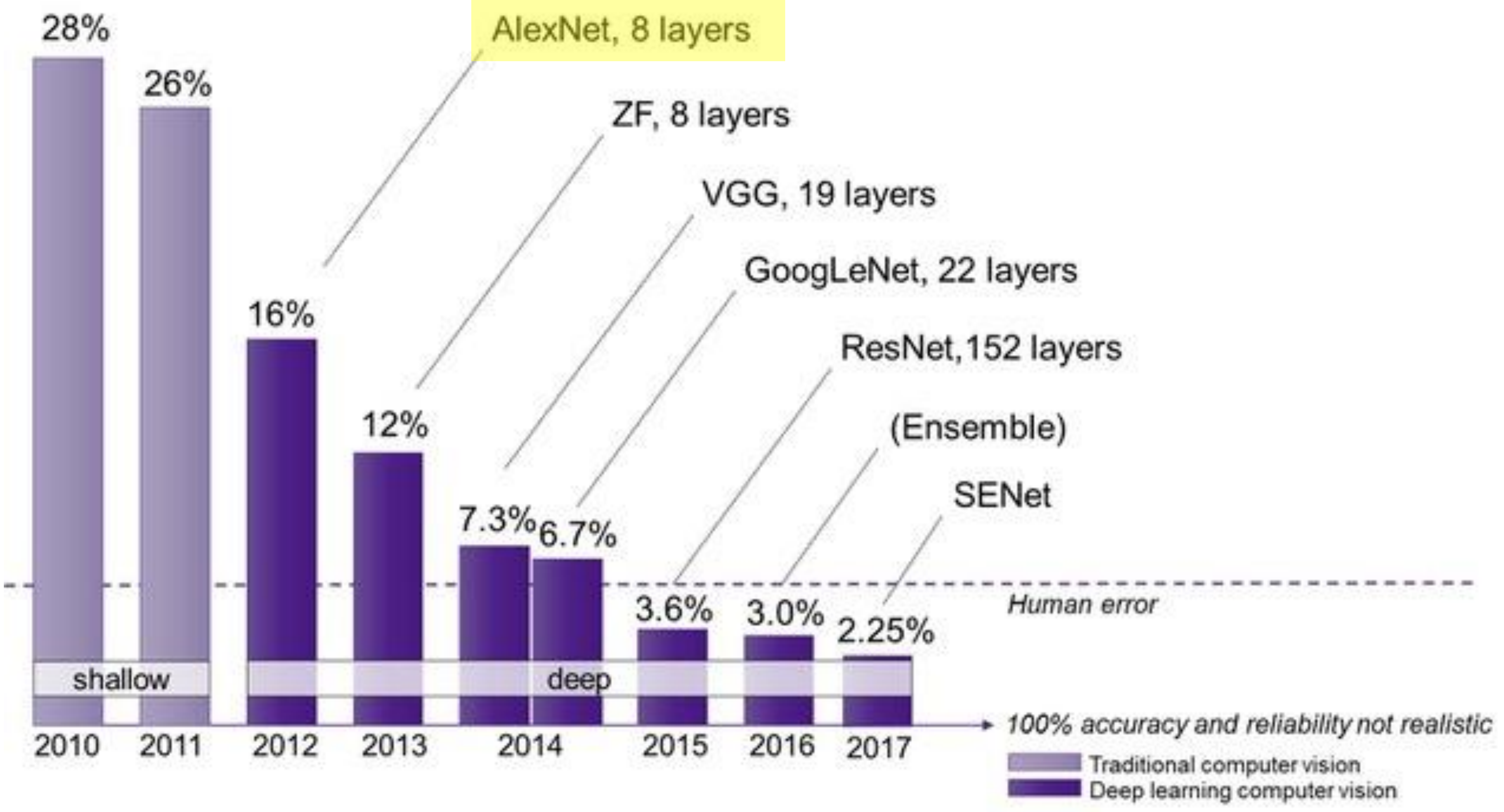
Deep Learning Potential (… and power)

Key advantages:

- **Everything** (feature extraction and classification) **is optimized** for improving the task at hand.
- **End-to-end trainable** solutions require no experts, just annotated data.
- Plenty of high-level frameworks (Keras, Tensorflow, PyTorch, TensorFlow Lite) that allows solving complex visual recognition by simply **programming black-boxes.**
- Democratisation of Computer Vision!
- Very effective…

The impact of Deep Learning in Visual Recognition

Classification accuracy on **ILSVRC**



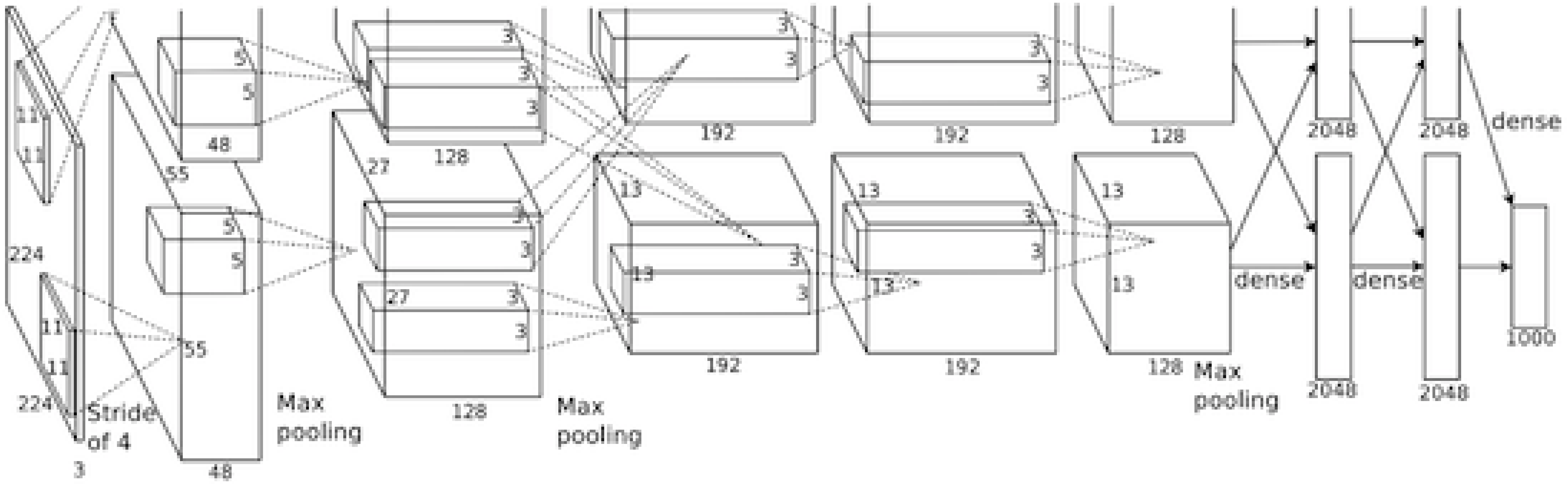
ILSVCR: ImageNet Large Scale Visual Recognition Challenge

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



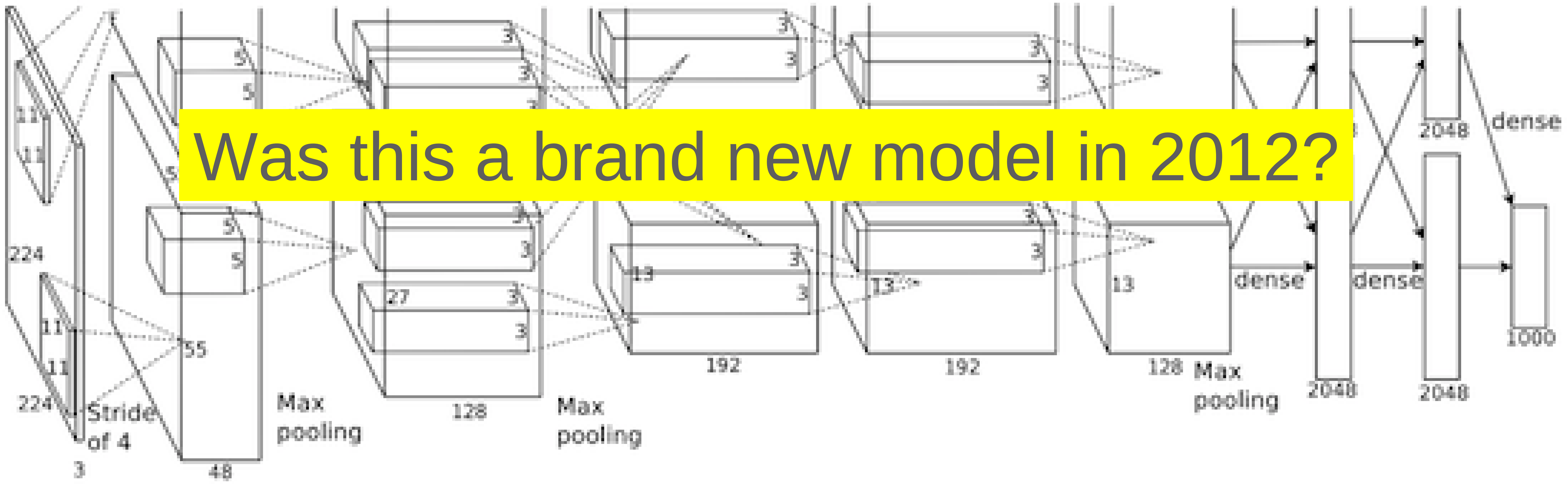
Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." NIPS 2012.

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

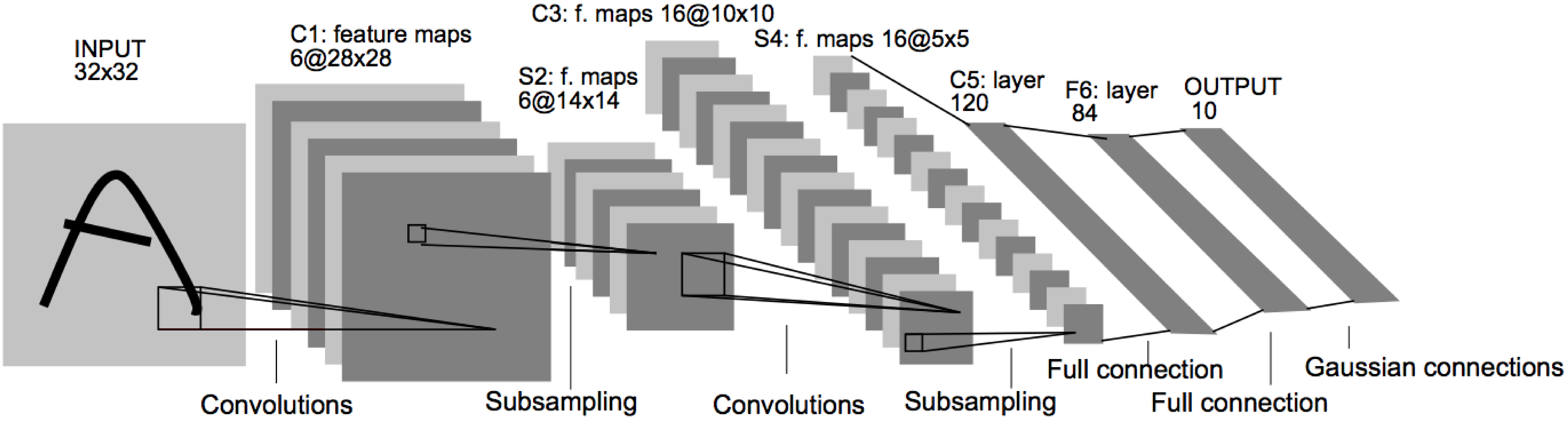


Was this a brand new model in 2012?

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." NIPS 2012.

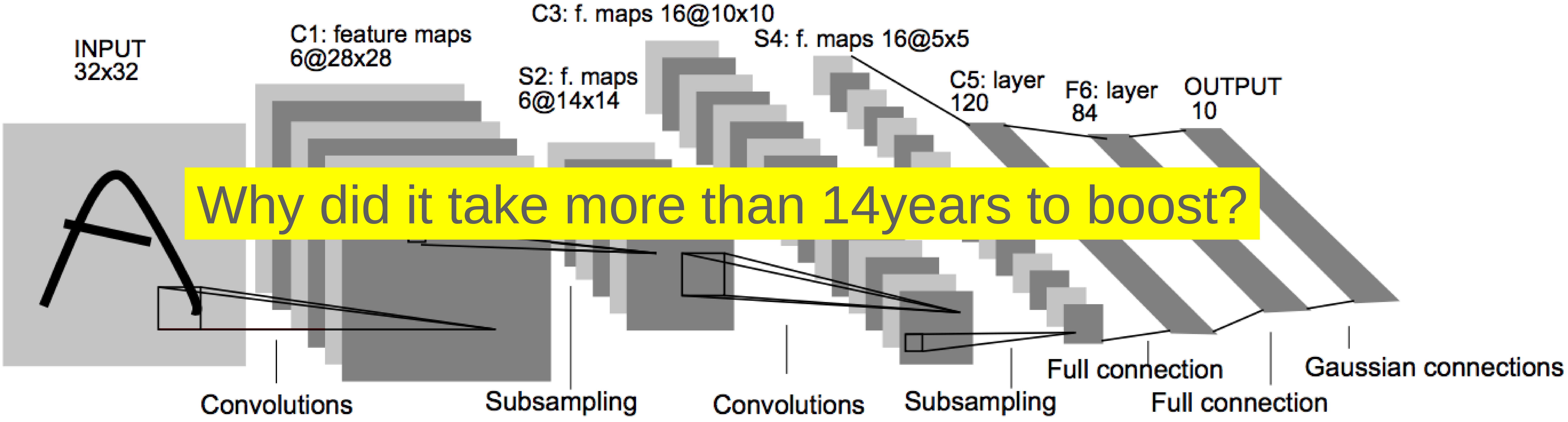
Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

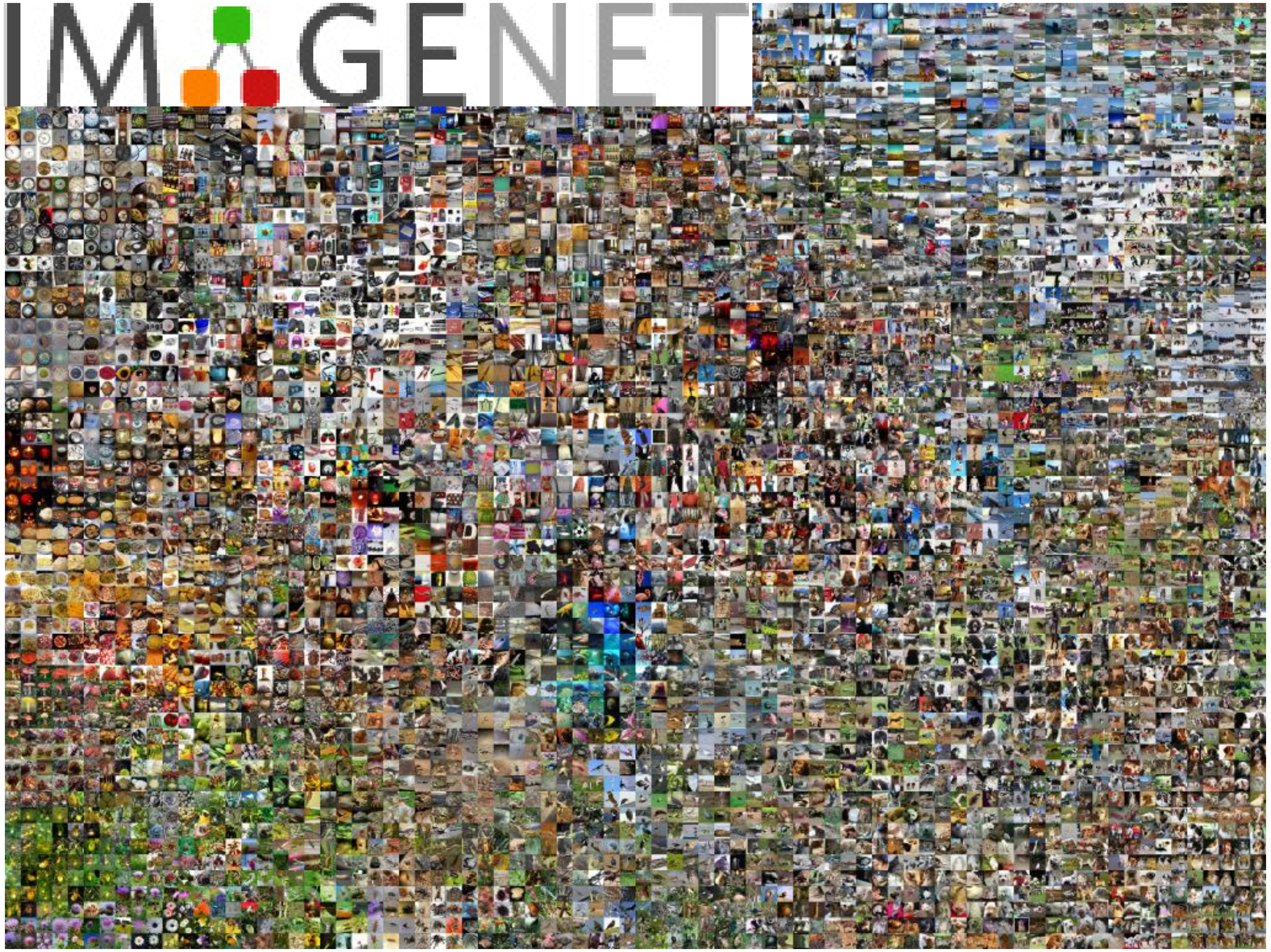


Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner



Large Collections of Annotated Data & Parallel Computing



https://www.flickr.com/photos/nvidia/34686550412

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. *CVPR, 2009.*

A Steadily Increasing Interest (and Power)

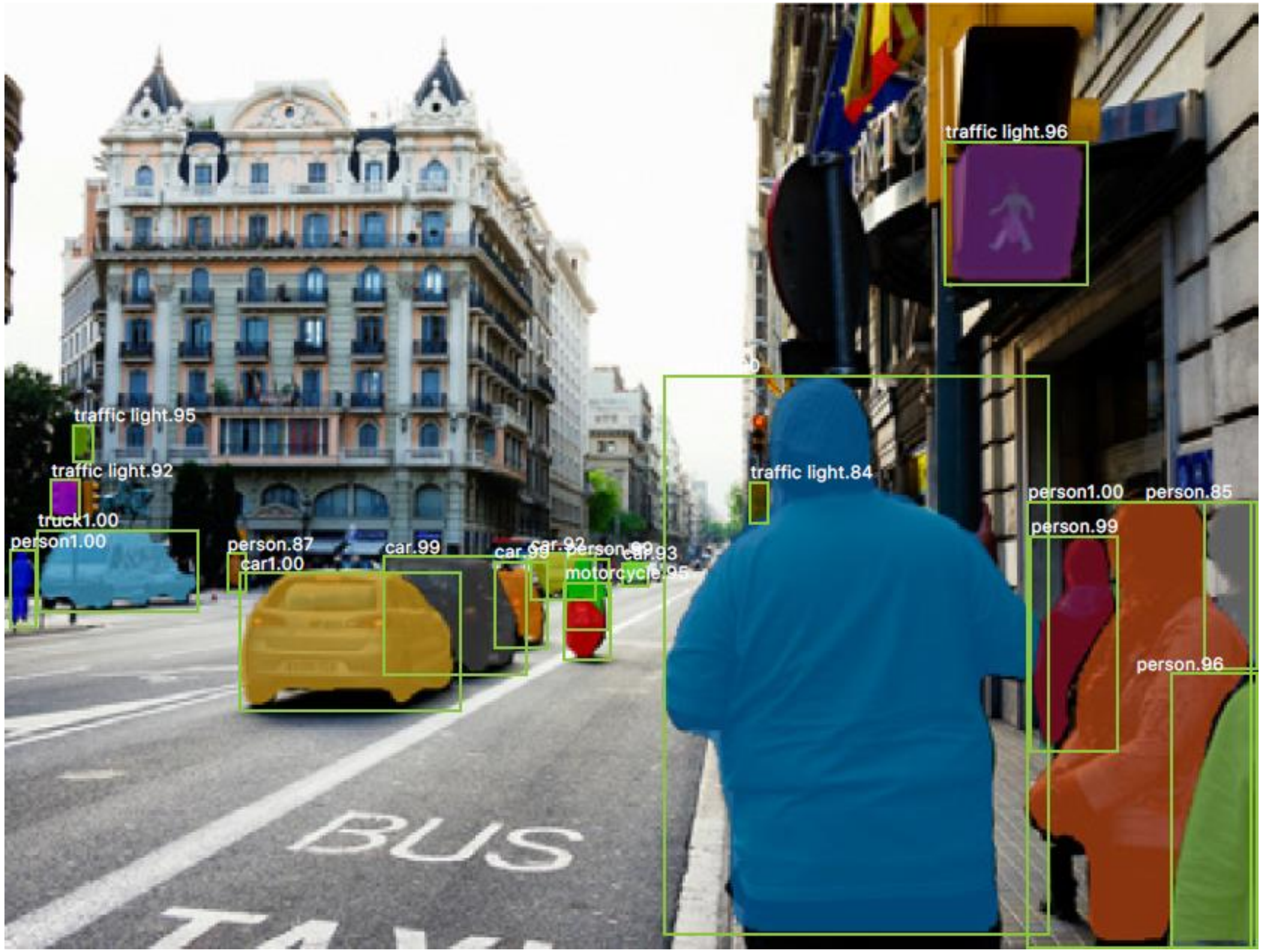
From Academia, Industries, Technology Enthusiastic….

Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award

Bengio, Hinton and LeCun Ushered in Major Breakthroughs in Artificial Intelligence

https://awards.acm.org/about/2018-turing

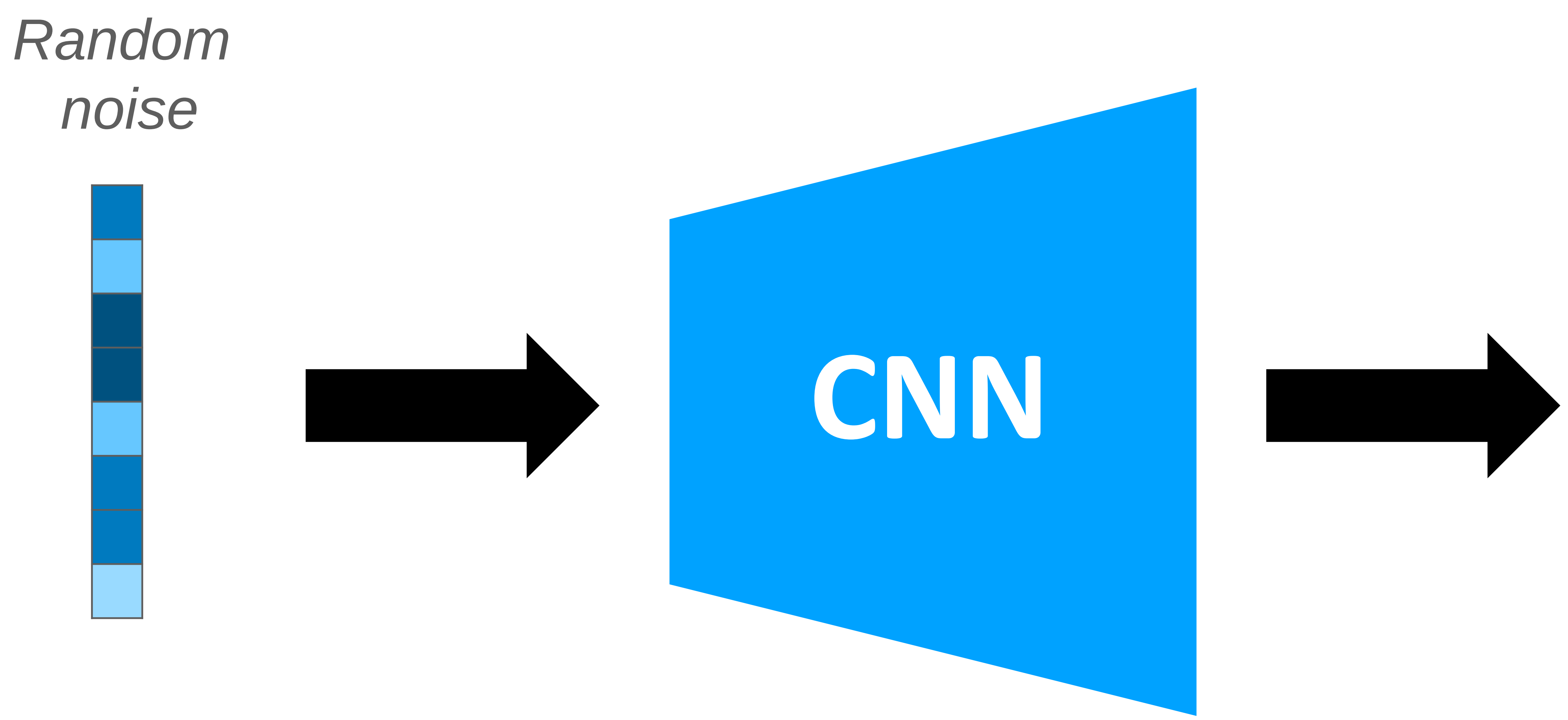
Instance Segmentation / Human Pose Estimation



He, K., Gkioxari, G., Dollár, P., & Girshick, R. “*Mask R-CNN*”. ICCV 2017

GAN: Generative Adversarial Networks

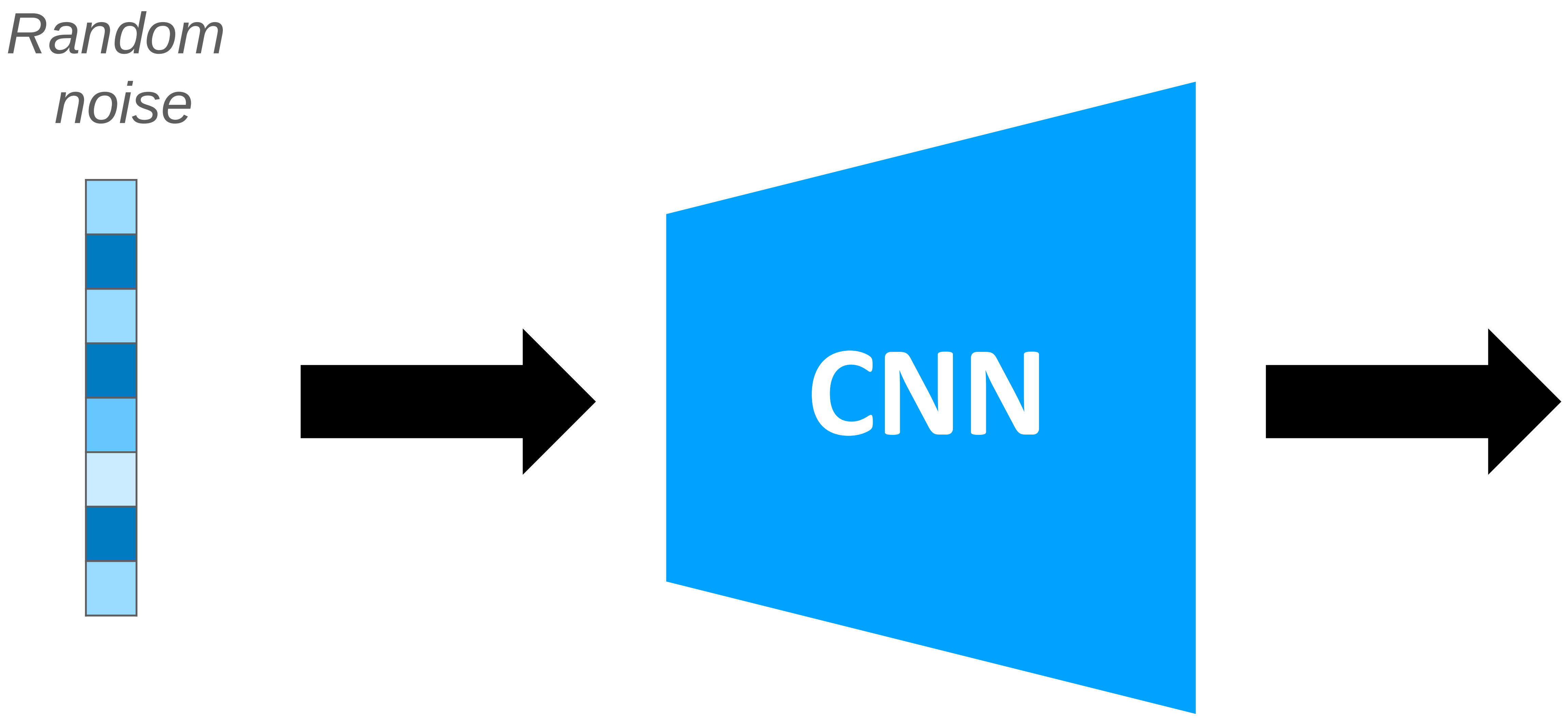
https://thispersondoesnotexist.com



Tero Karras, Samuli Laine, Timo Aila «A Style-Based Generator Architecture for Generative Adversarial Networks” CVPR 2019

GAN: Generative Adversarial Networks

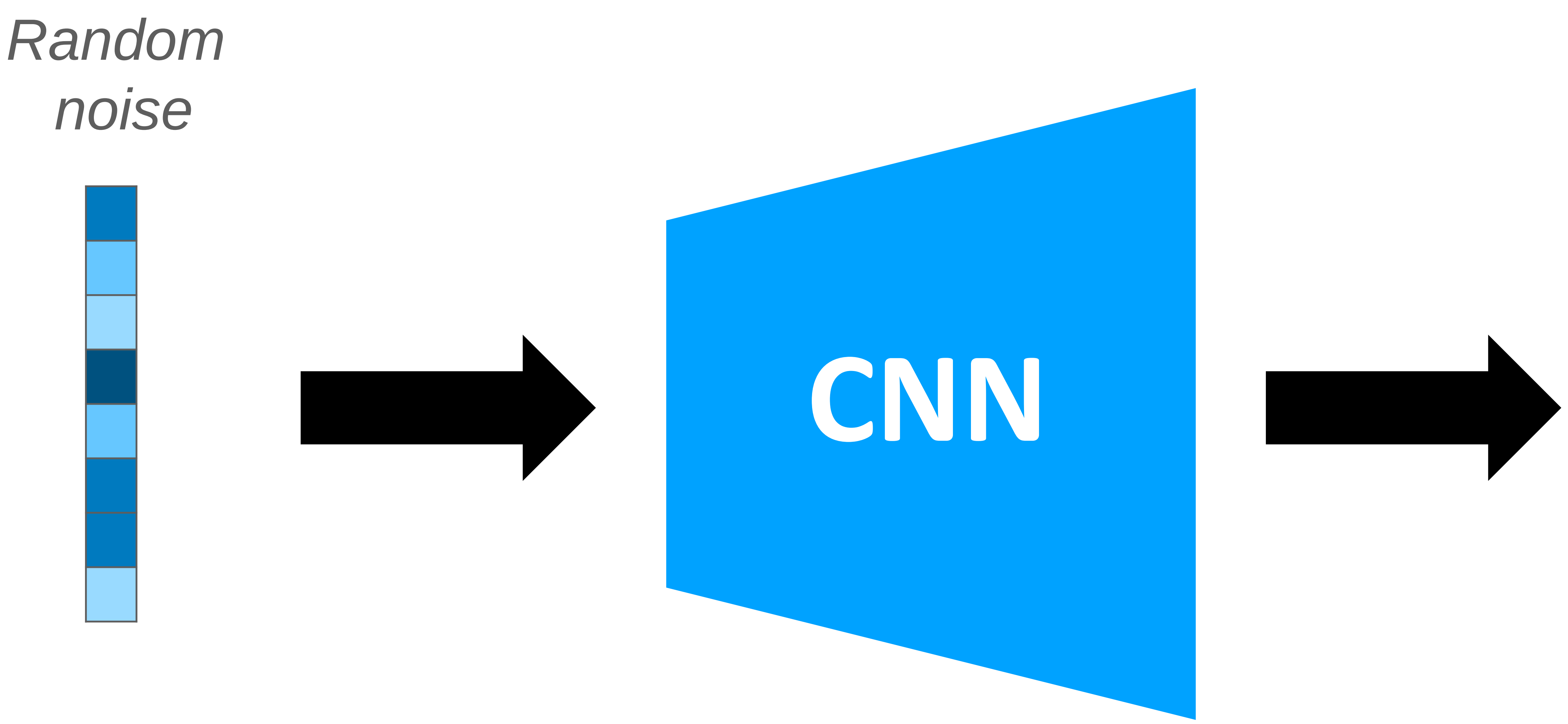
https://thispersondoesnotexist.com



Tero Karras, Samuli Laine, Timo Aila «A Style-Based Generator Architecture for Generative Adversarial Networks” CVPR 2019

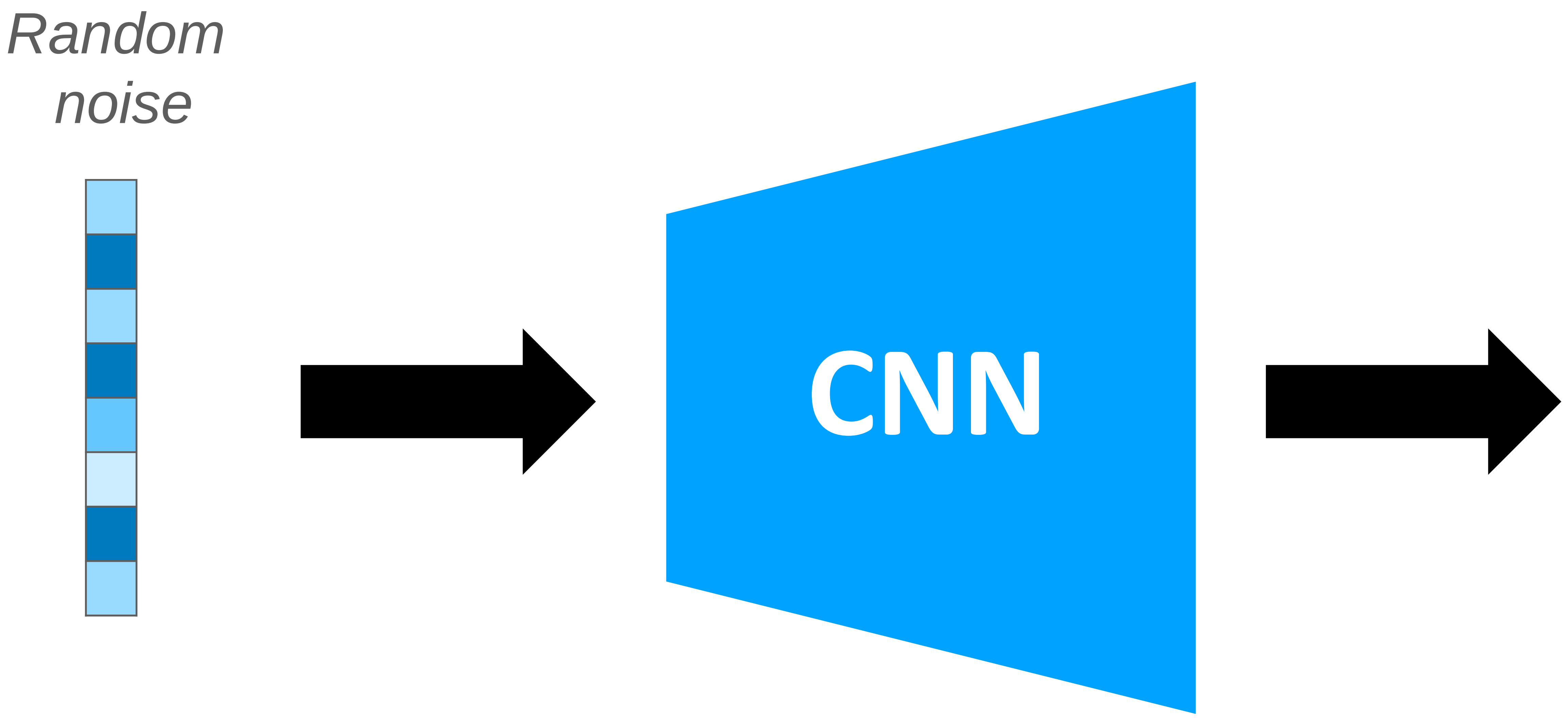
GAN: Generative Adversarial Networks

https://thispersondoesnotexist.com



Tero Karras, Samuli Laine, Timo Aila «A Style-Based Generator Architecture for Generative Adversarial Networks” CVPR 2019

GAN: Generative Adversarial Networks

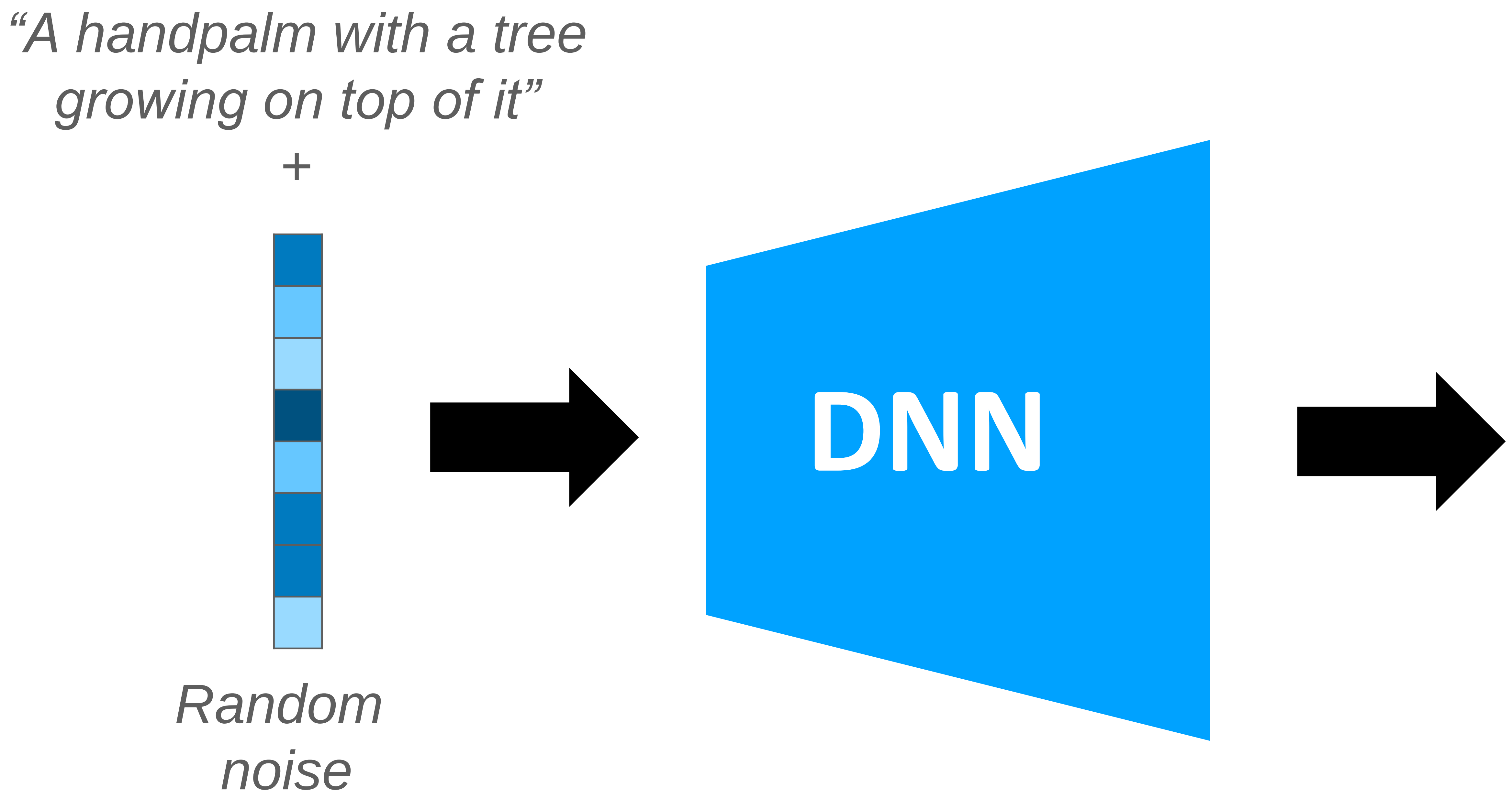


https://thispersondoesnotexist.com

Tero Karras, Samuli Laine, Timo Aila «A Style-Based Generator Architecture for Generative Adversarial Networks" CVPR 2019

DALL-E: Image Generation from Text

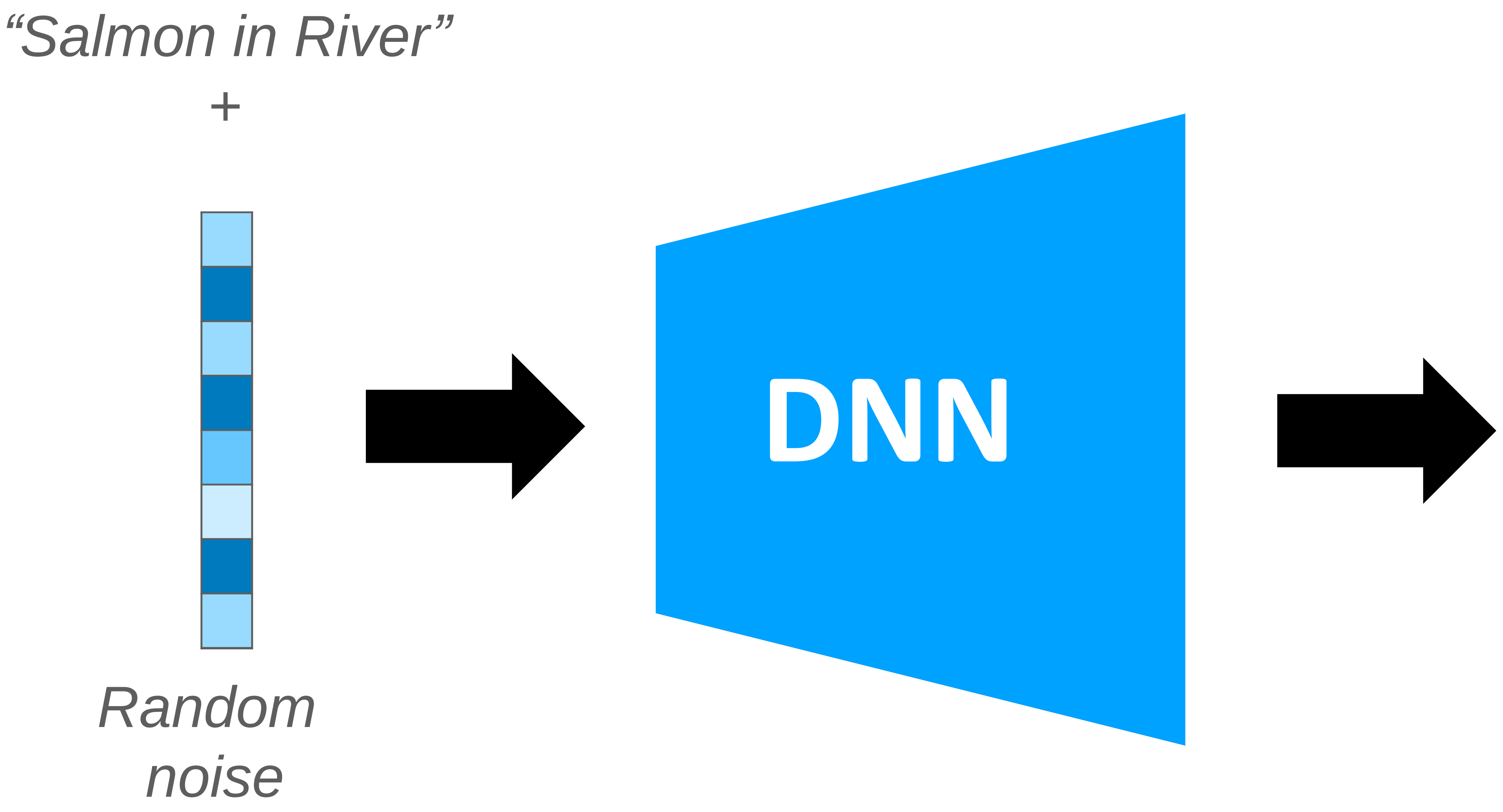
https://openai.com/dall-e-2/



Ramesh, Aditya, et al. "Zero-shot text-to-image generation." International Conference on Machine Learning. PMLR, 2021.

DALL-E: Image Generation from Text

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



Ramesh, Aditya, et al. "Zero-shot text-to-image generation." International Conference on Machine Learning. PMLR, 2021.

… what about here?

Our initiatives related to deep learning

Artificial Neural Networks and Deep Learning@Polimi

Teaching initiatives from Prof. Matteucci and Prof. Boracchi

AY 2017/2018 PhD Courses «*Image Classification: Modern Approaches*» and «*Deep Learning*» (150+)

AY 2018/2019 PhD Course: «*Advances In Deep Learning with Appl. in Text and Image Processing*» (71)

AY 2019/2020 PhD Course on «*Machine Learning for Non Matrix Data*» (43)

AY 2020/2021 MSc Course (CSE + MTM) «*Artificial Neural Networks and Deep Learning*» **AN2DL** (493)

+ Advanced PhD Course

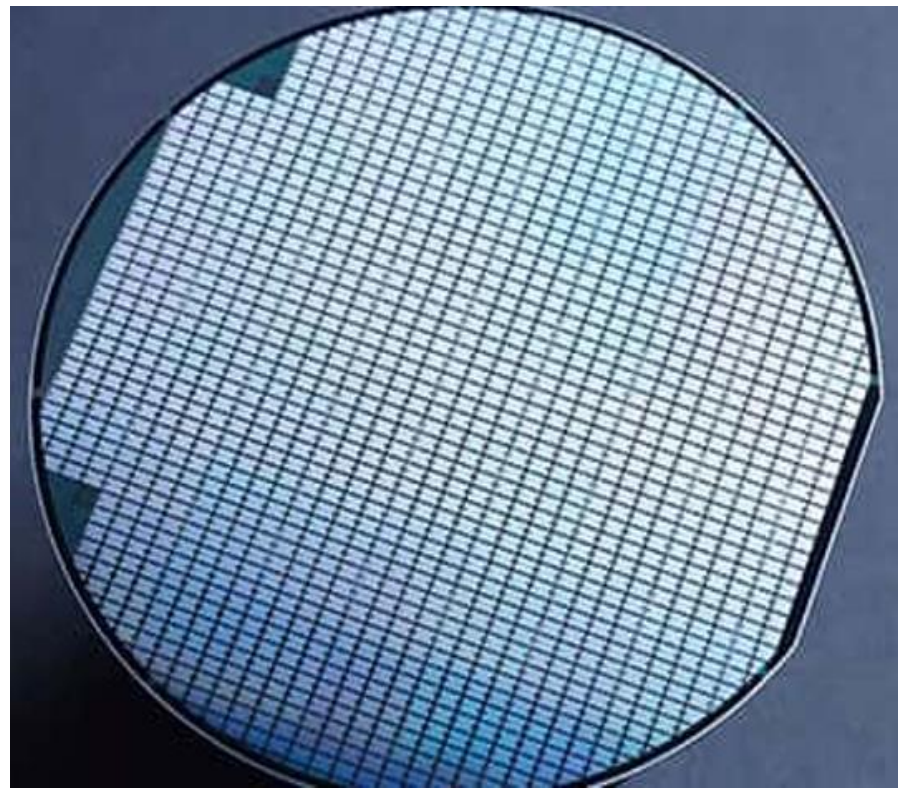
AY 2021/2022 **AN2DL** opened to BIO students (overall 572) + Advanced PhD Course

AY 2022/2023 **AN2DL CSE + BIO + MTM** (overall 731) + Advanced PhD Course

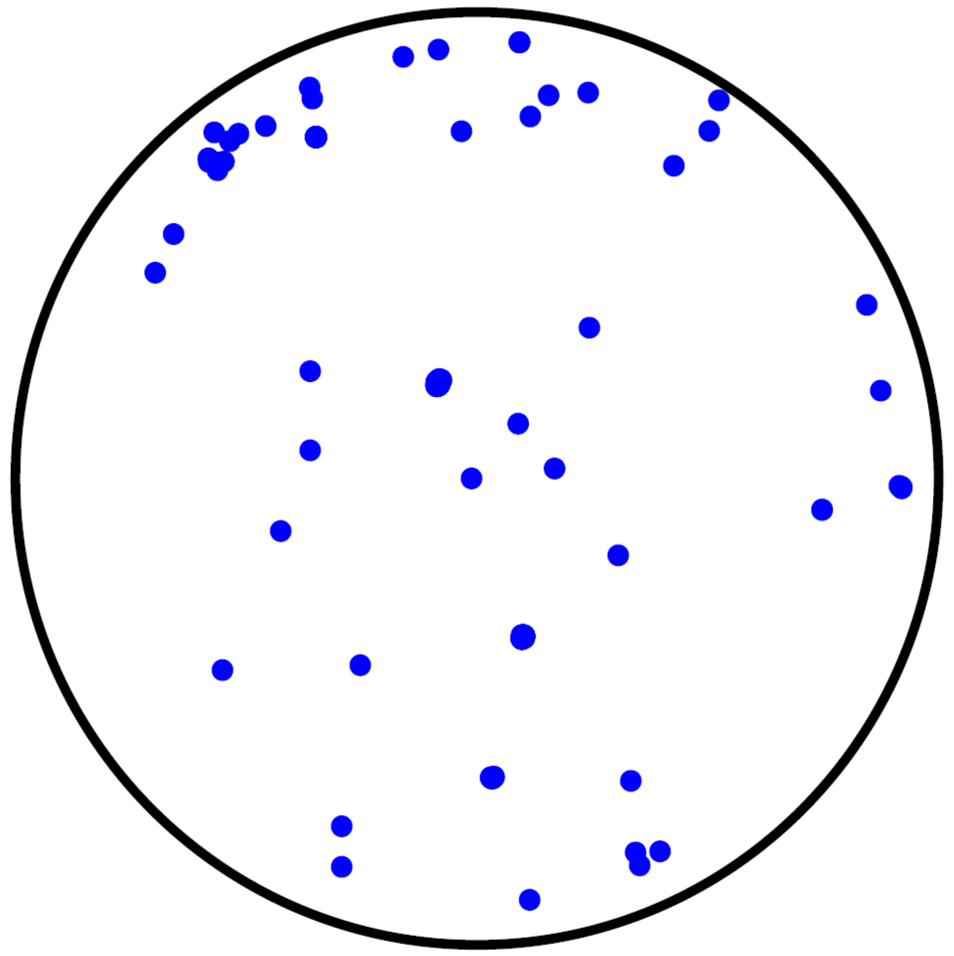
… and courses being offered in companies as well

Successful Industrial Projects

Silicon Wafer



Inspection Tool



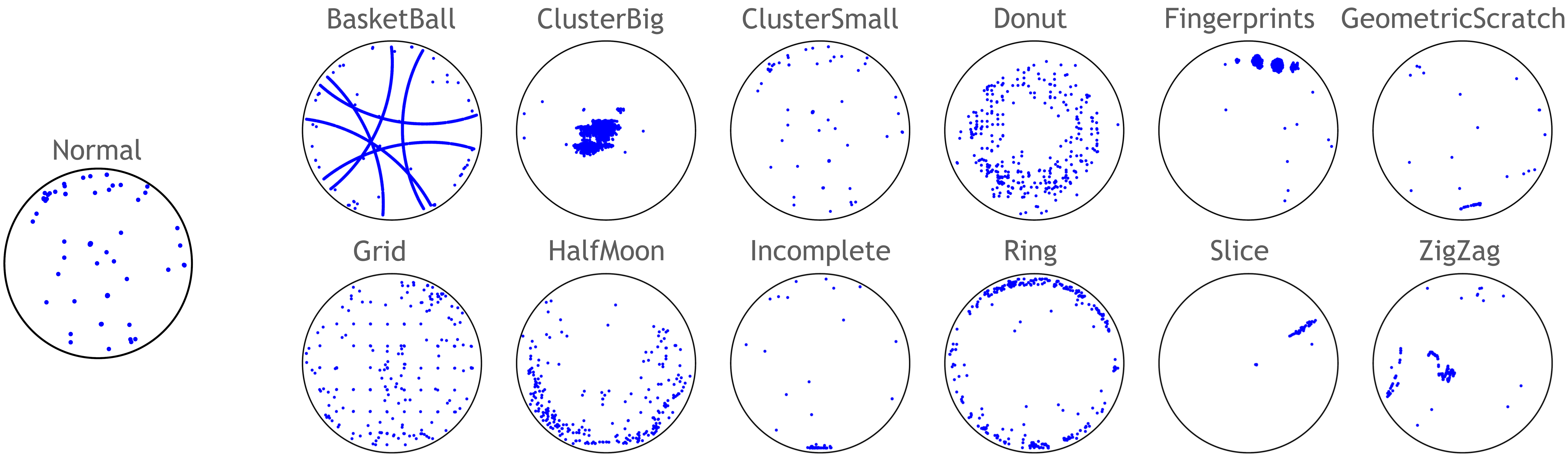
Wafer Defect Map (WDM)

In collaboration with

life.augmented

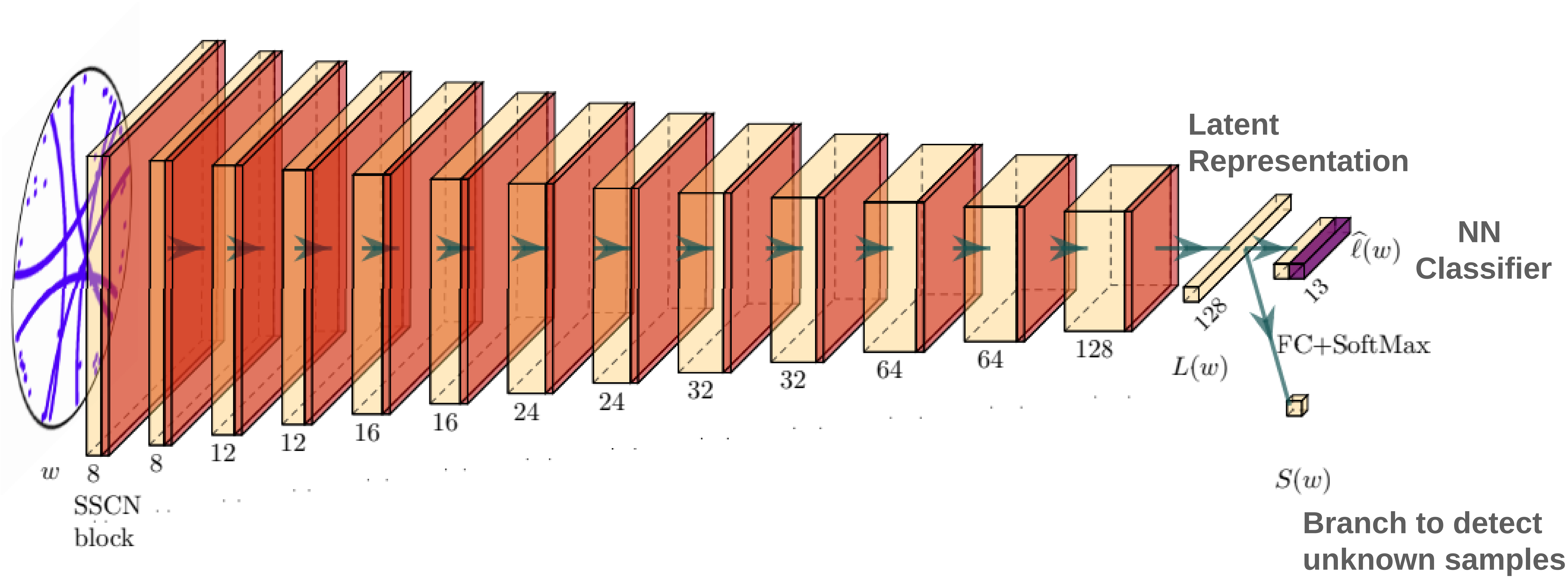
Classes of WDM Defective Patterns

Specific **patterns** in WDMs might indicate **problems** in the production line



Classify WDM to raise prompt alerts

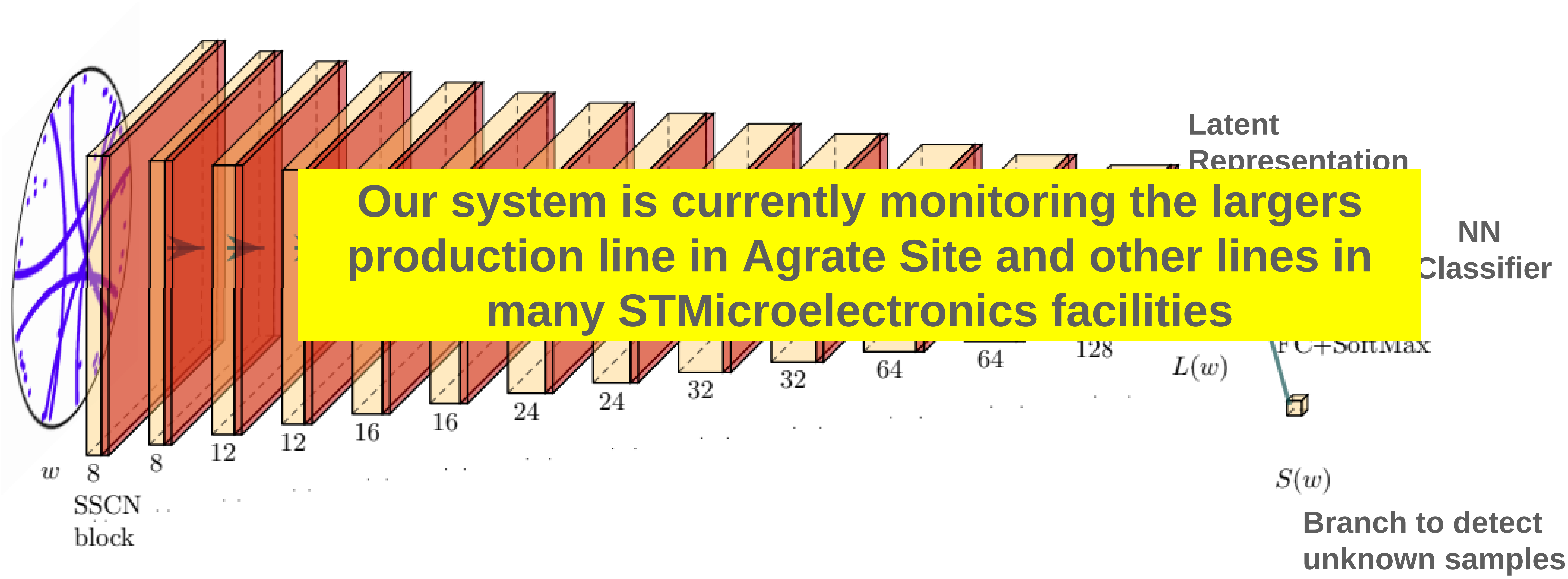
Our CNN



Frittoli, L., Carrera, D., Rossi, B., Fragneto, P., & Boracchi, G. (2022). «Deep open-set recognition for silicon wafer production monitoring». Pattern Recognition, 124, 108488.

L. Moioli, P. Fragneto, B. Rossi, D. Carrera, G. Boracchi, M. Fumagalli, E. Tagliabue, P. Giugni, A. Aurigemma “Wafer Manufacturing System, Device And Method” US Patent Application [US10922807] – 2018-10-29, Granted.

Our CNN



Our system is currently monitoring the largers production line in Agrate Site and other lines in many STMicroelectronics facilities

Frittoli, L., Carrera, D., Rossi, B., Fragneto, P., & Boracchi, G. (2022). «Deep open-set recognition for silicon wafer production monitoring». Pattern Recognition, 124, 108488.

L. Moioli, P. Fragneto, B. Rossi, D. Carrera, G. Boracchi, M. Fumagalli, E. Tagliabue, P. Giugni, A. Aurigemma "Wafer Manufacturing System, Device And Method" US Patent Application [US10922807] – 2018-10-29, Granted.

Automated Prohibited Item Detection System

Advanced **X-ray systems for hand luggage inspection**, equipped with algorithms to **automatically detect and localize threat items** inside each bag, such as firearms, sharps and blunt weapons, based on their shape.

https://www.pointfwd.com/news/tag/APIDS

https://www.gilardoni.it ARGO

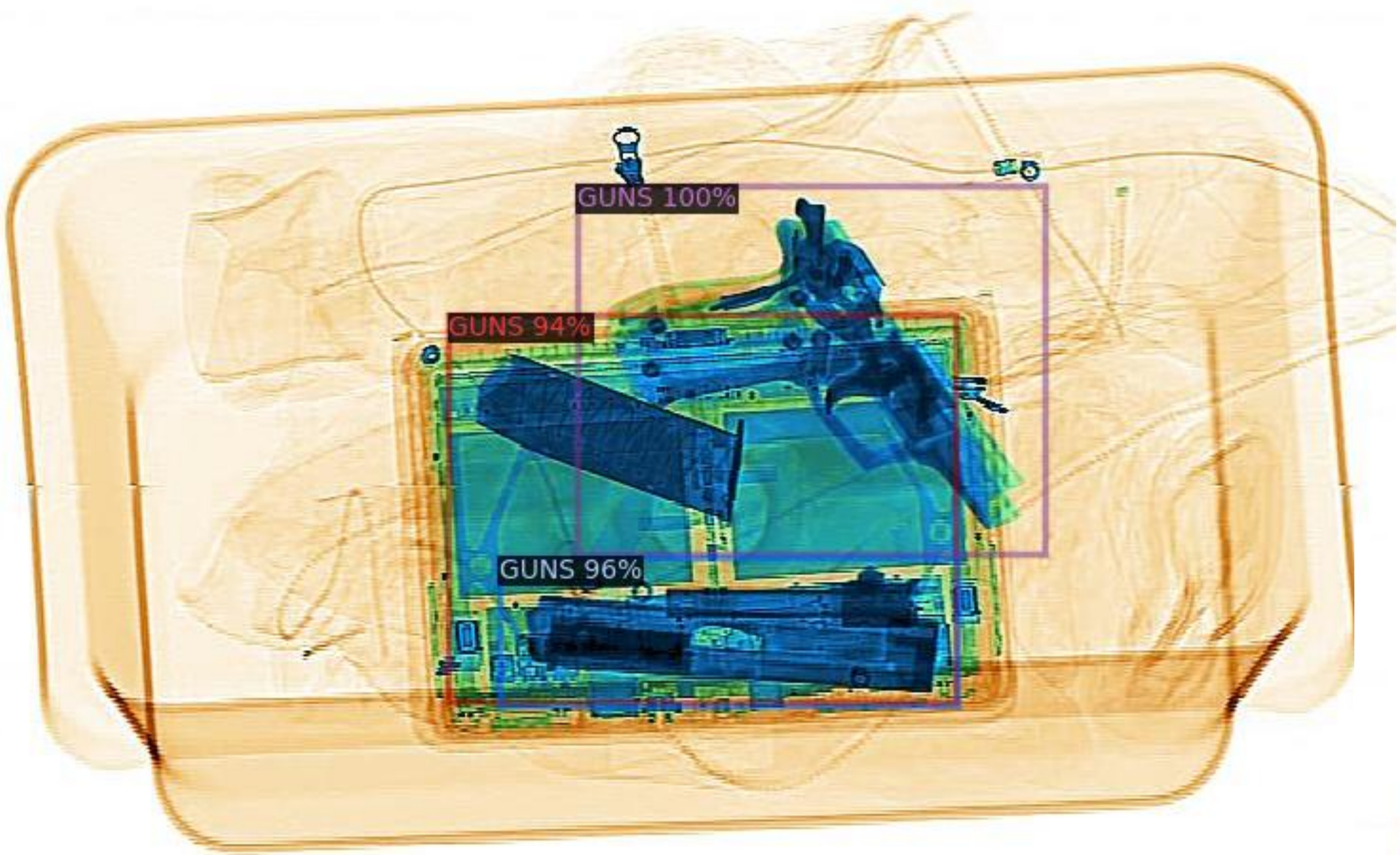
In collaboration with

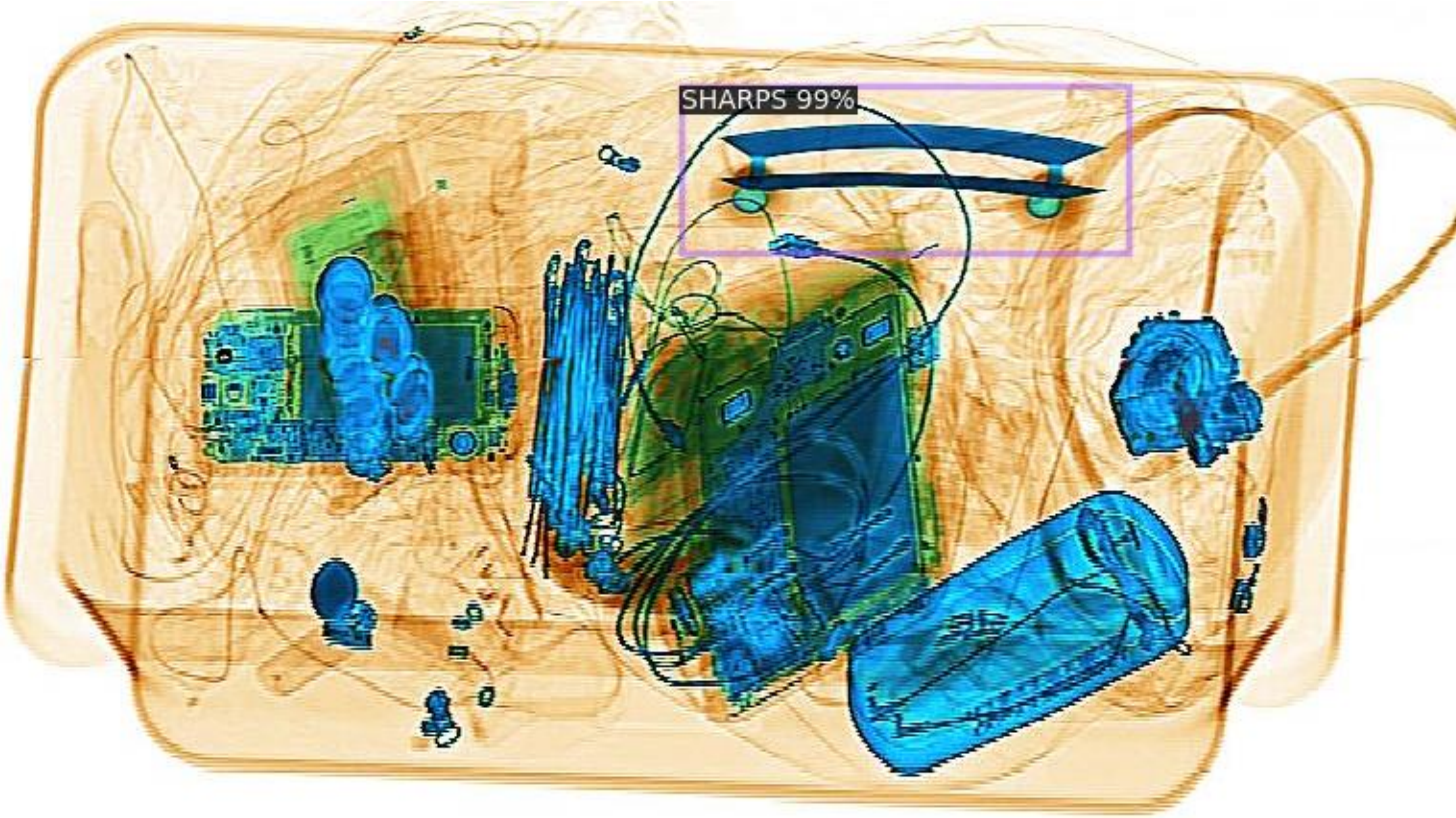
GILARDONI

ellis European Laboratory for Learning and Intelligent Systems

POLITECNICO MILANO 1863

Automated Prohibited Item Detection System





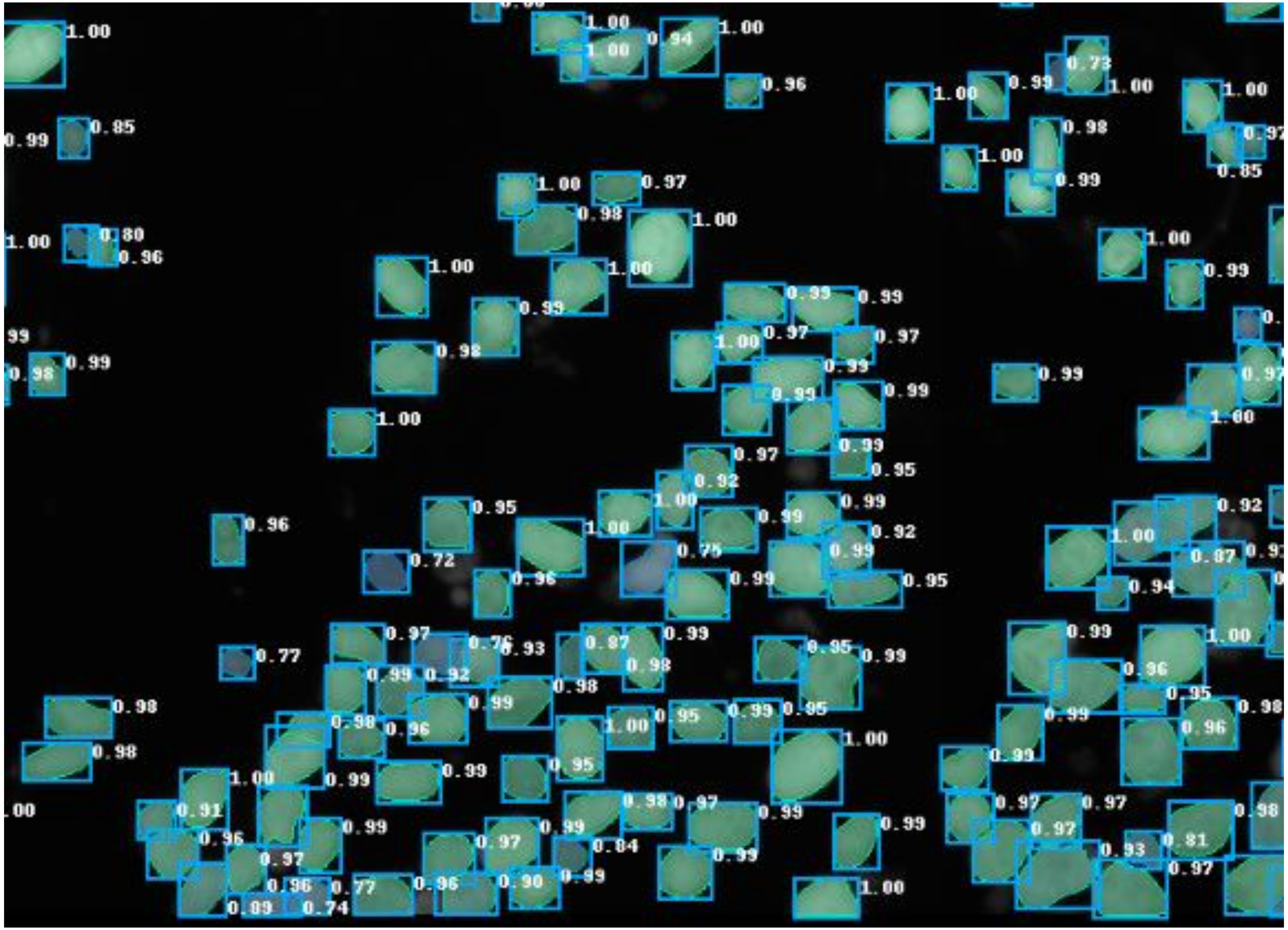
In collaboration with

GILARDONI

Isolating Cell Nuclei in Lung Tissues

The network provides, for each identified nucleus,

- Its class with prediction confidence,
- its segmentation mask,
- its bounding box.



In collaboration with

ikonisys

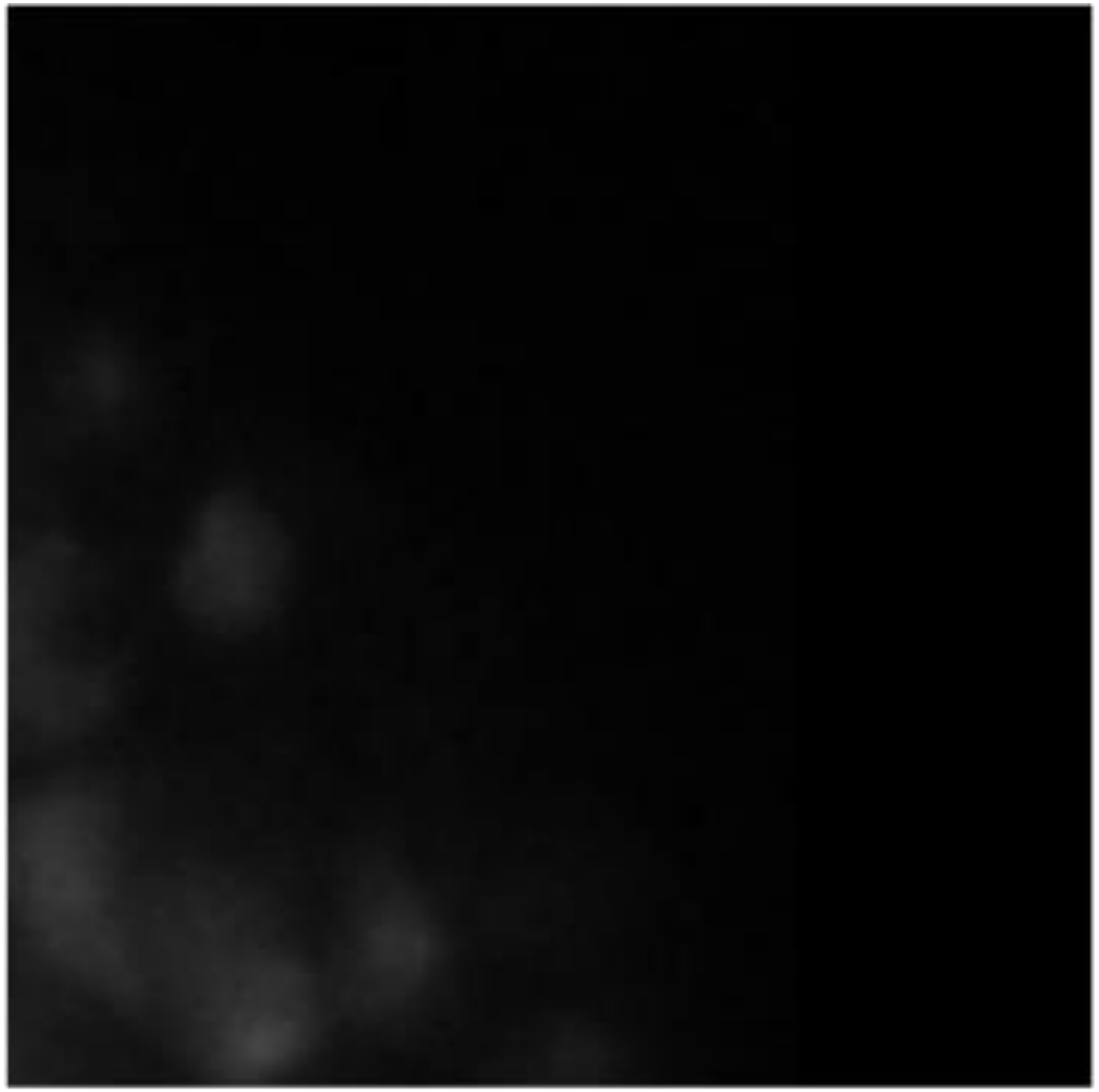
FINDING THE CELLS THAT MATTER

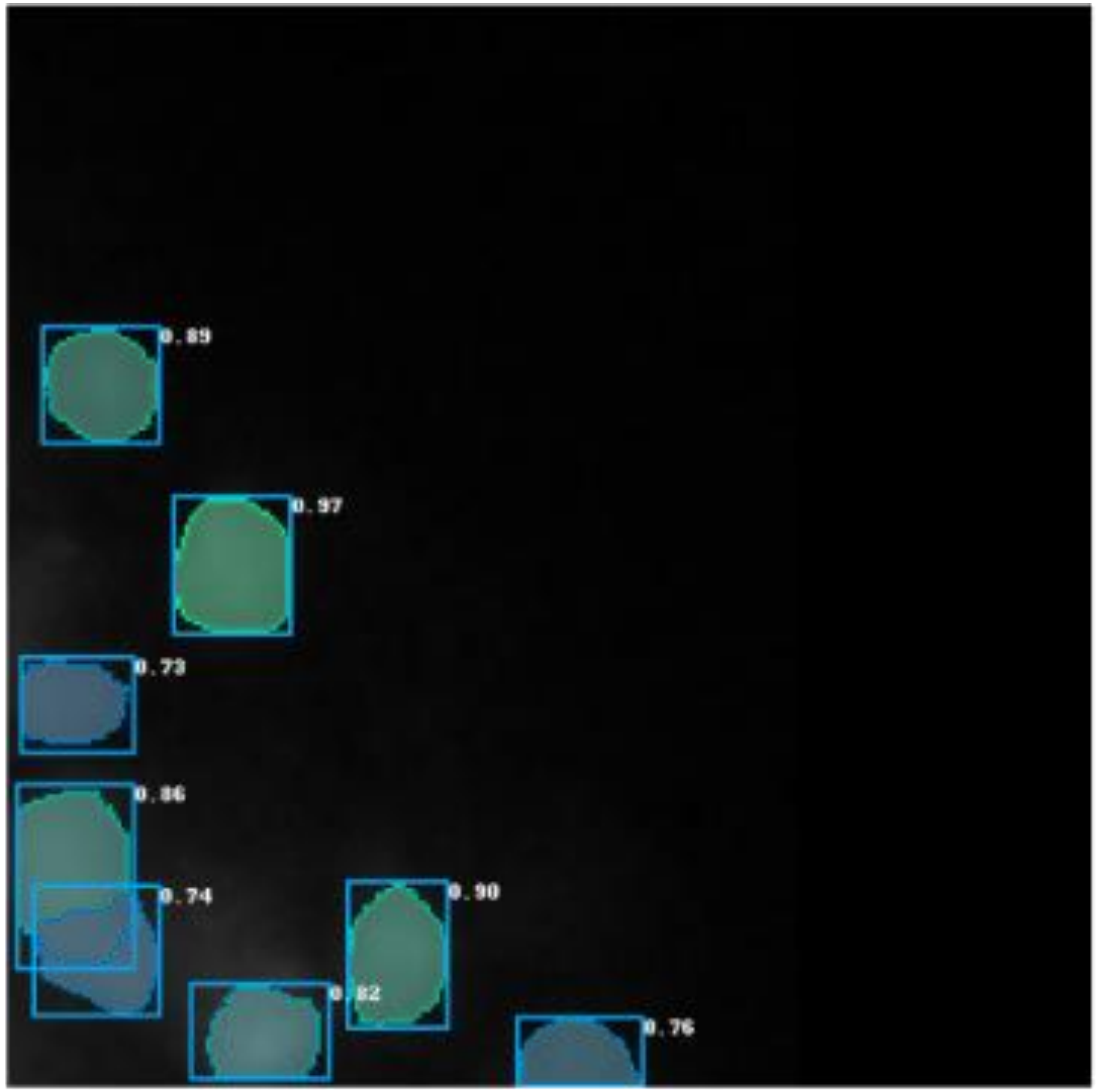
ellis

European Laboratory for Learning and Intelligent Systems

POLITECNICO MILANO 1863

Isolating Cell Nuclei in Lung Tissues





In collaboration with

ikonisys
FINDING THE CELLS THAT MATTER

ellis
European Laboratory for Learning and Intelligent Systems

POLITECNICO
MILANO 1863

Understanding Neural Networks

..to rely on their decisions in *critical* tasks

Understanding Deep Neural Networks

Deep Neural Networks have **Million parameters:** their inner functioning is **totally obscure**.

Healthy scepticism to resort to NN decision in critical tasks (e.g. medical domain) or even services (e.g., blocking credit cards).

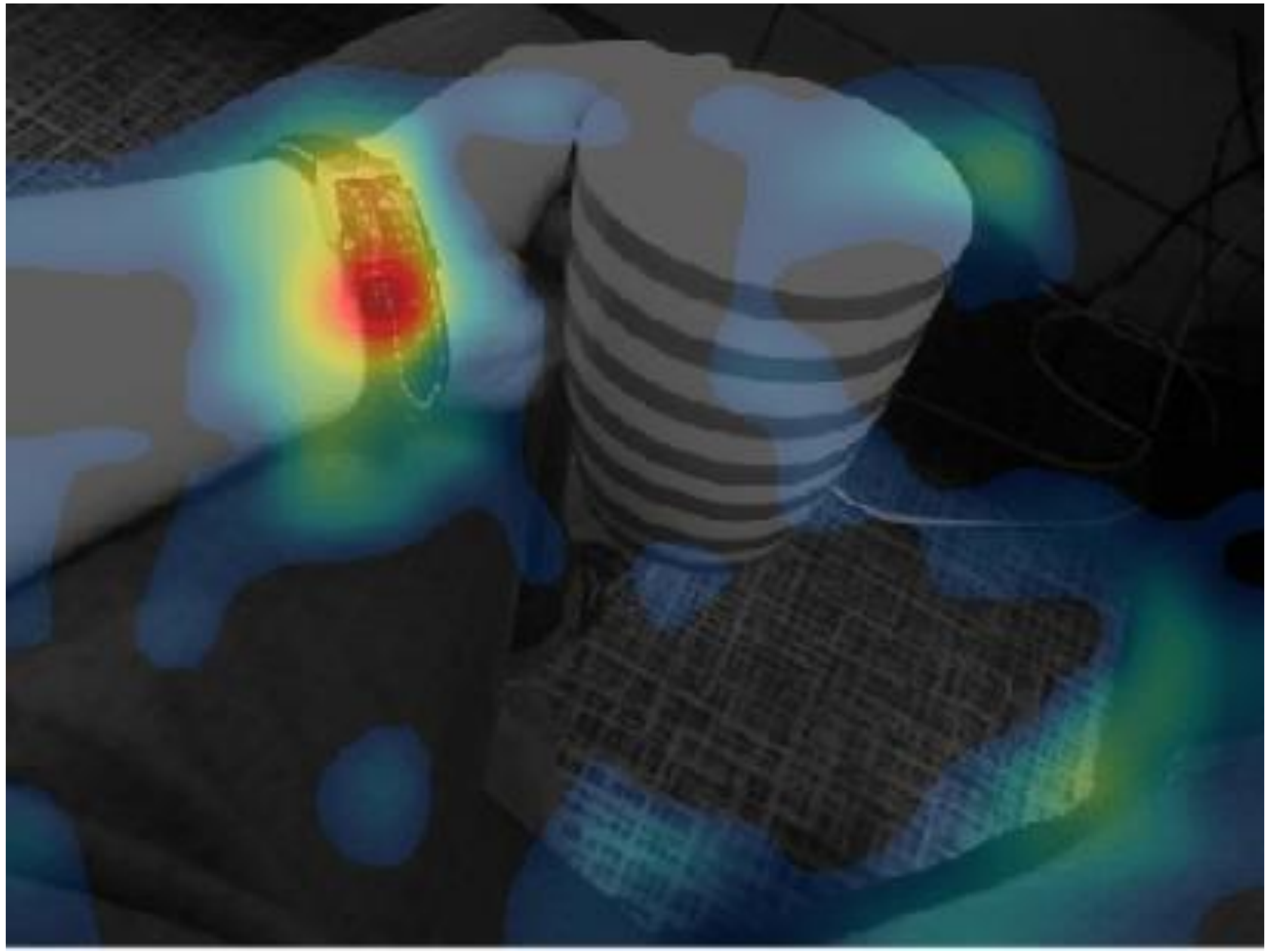
Vivid research activity around **gaining an understanding of Neural Network decision.**

Mispredicted as “buckle”

Saliency Maps to Understand Model Mistakes

Make sense of model mistakes

Mispredicted as “buckle”

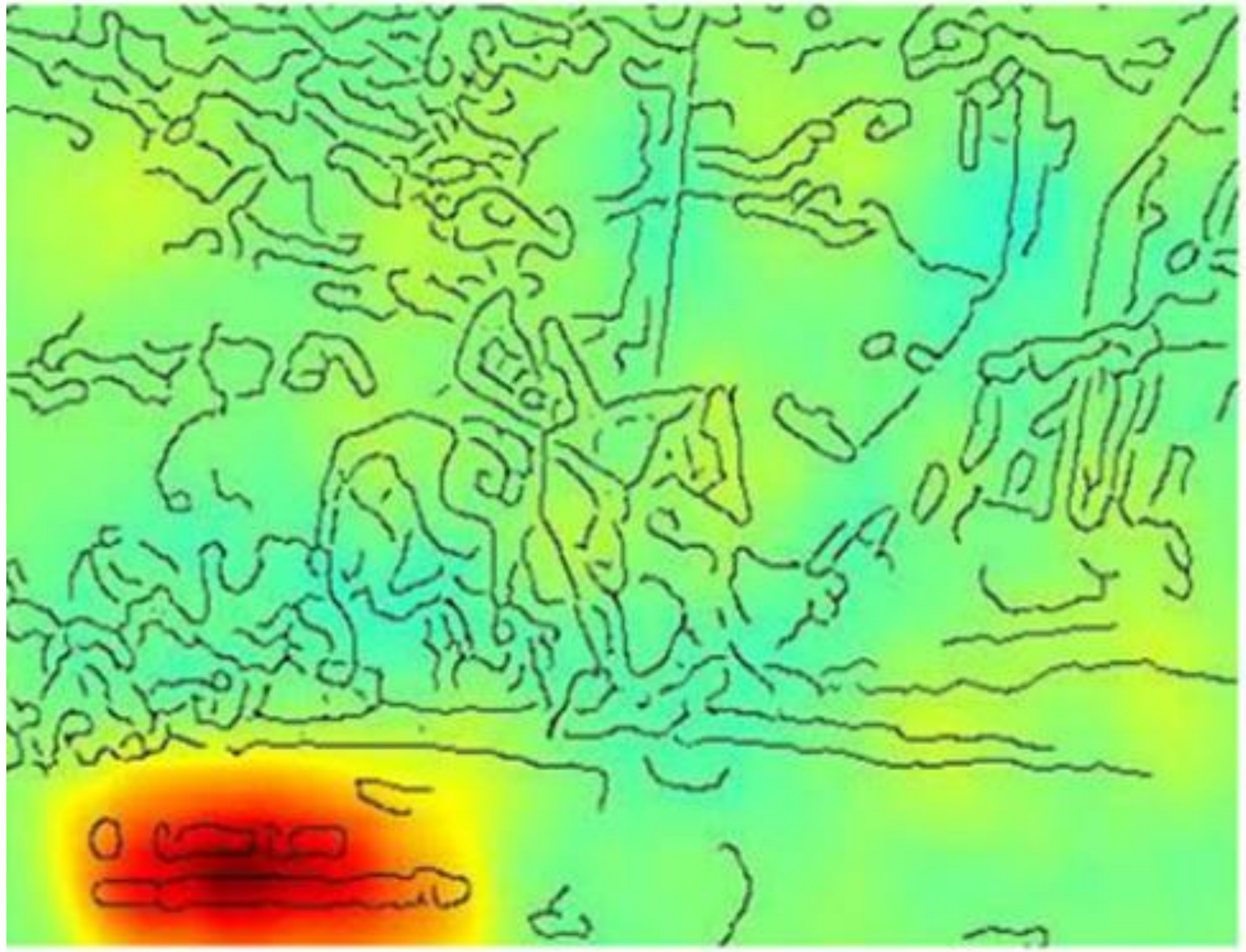


Saliency shows why

Saliency Maps to Discover Systematic Errors

Highlight clever Hans phenomena

Correctly classified as "horse"



But for the wrong reason

Augmented Grad-CAM: Improving Saliency Maps Resolutions

Mastiff Class

(a) Grad-CAM.

Mastiff Class

(b) Augmented Grad-CAM.

Morbidelli, P., Carrera, D., Rossi, B., Fragneto, P., & Boracchi, G. (2020, May). Augmented Grad-CAM: Heat-Maps Super Resolution Through Augmentation. In *ICASSP 2020*

Perception Visualization

Misclassified as “boat”

«where»

Saliency doesn’t say much

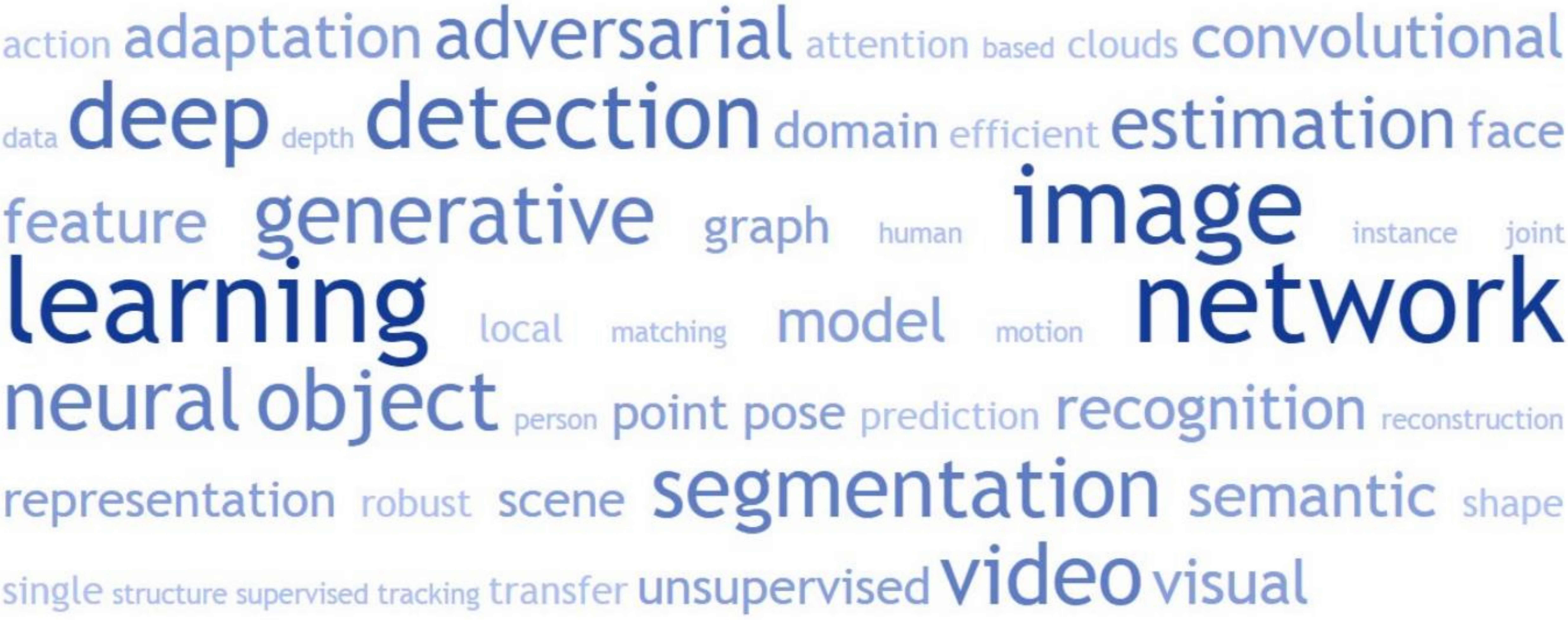
«where and what»

PV shows why

L. Giulivi M.J. Carman G. Boracchi “Perception Visualization: Seeing Through the Eyes of a DNN” BMVC 2021

… should then Everything go Deep?

CVPR 2019 Keywords

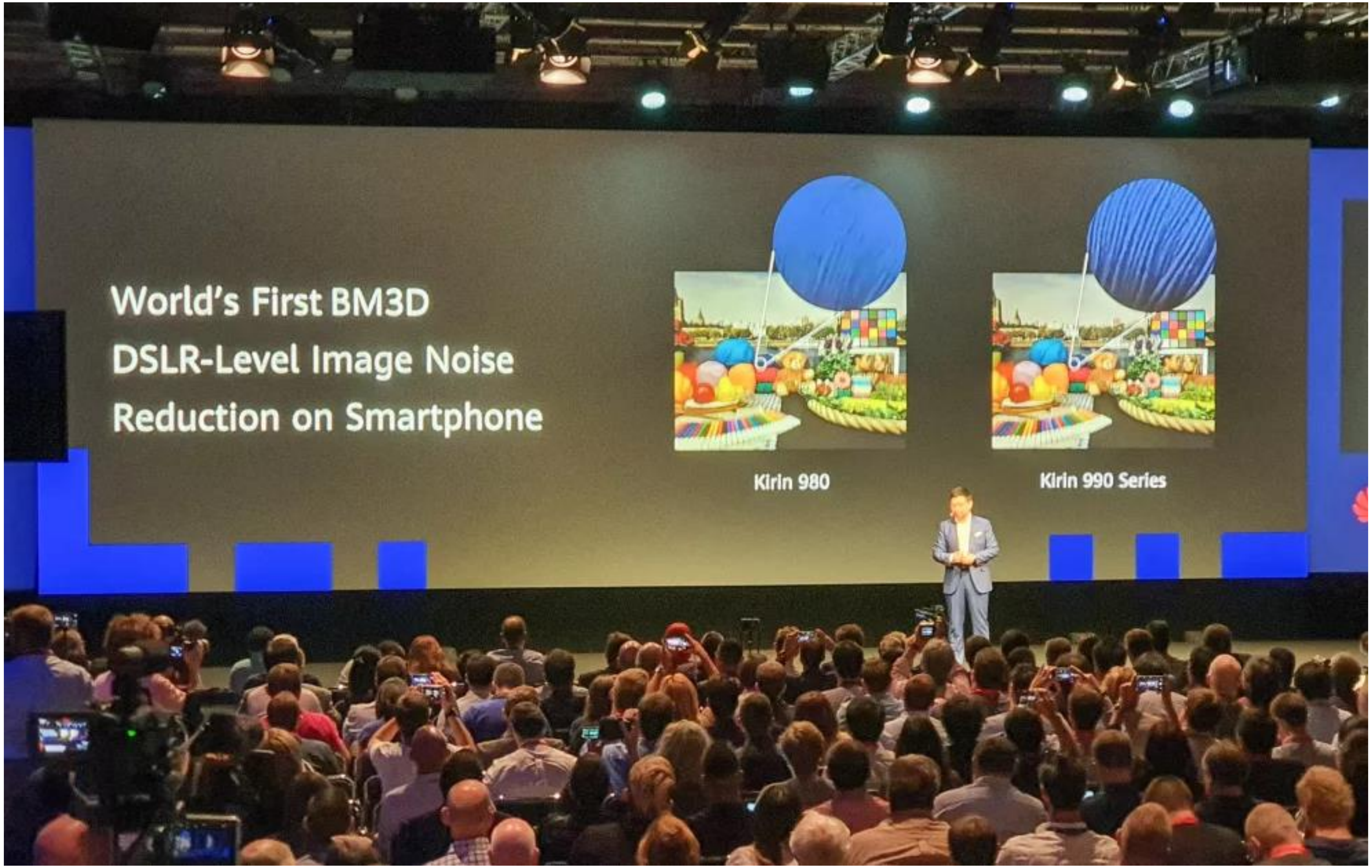


source: CVPR 2019 Welcome slides

2019, well after deep learning boom

Image Denoising: a fundamental ingredient in any image processing pipeline

In 2019 **Huawei** has announced **integrating BM3D** in is new chipset series Kirin 990



(Image credit: Basil Kronfli/Digital Camera World)

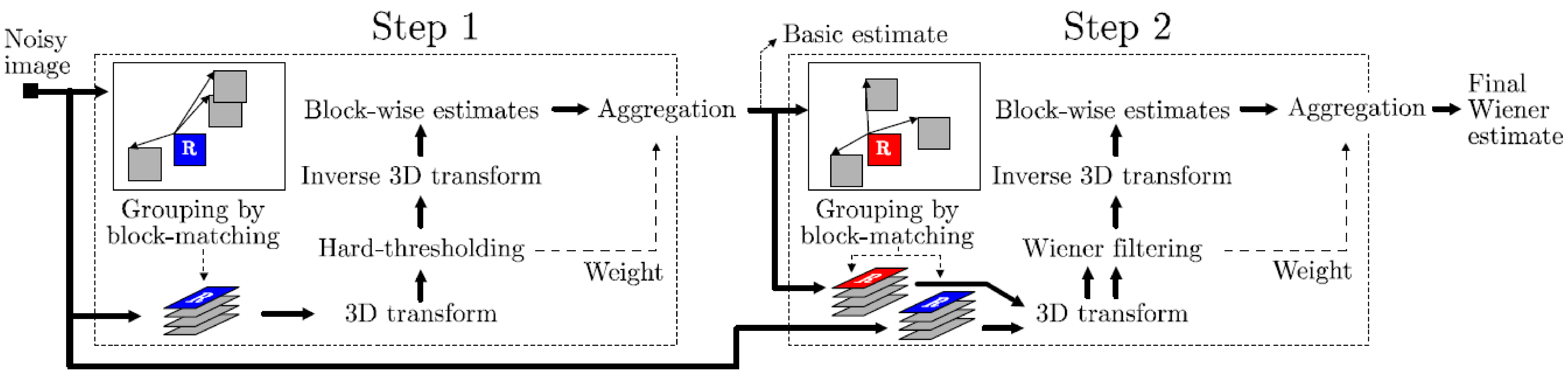
BM3D Image Denoising

A Breakthrough in 2007 introducing a new paradigm to handle natural images.

The resulting algorithm is based on:

- an **explicit prior** describing patches of natural images
- an **explicit modelling** of **sensor noise**

… overall, a bunch of parameters to tune



K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080-2095, August 2007.


Noisy 16.10 dB

Restored 28.49 dB

State of the art in video denoising in 2012

M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, “Video Denoising, Deblocking and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms”, IEEE Transactions on Image Processing

… and so we do in our projects

X-ray imaging systems often need to operate at low signal-to-noise ratio conditions, due to:

- very high absorbing materials
- dose limits (e.g. medicine)
- short exposures (e.g. video fluoroscopy)

Noise is from many known sources and need to be suppressed

In collaboration with

GILARDONI

In collaboration with

digitec
DIGITAL INNOVATION


Noisy
Denoised


Enhanced
Low-Res

… and Why Not Deep Learning?

Deep Neural Networks for Image Denoising:

- Can leverage supervision, which was not possible in previous methods.
- … after almost 10 years have outperformed BM3D in terms of **image restoration quality**

Phone manufacturers are still relying on algorithms based on an explicit model

- Hand-crafted / mathematical priors for images
- Statistical modelling of the noise distribution

Good Reasons to do so:

- **Computational complexity / hardware requirements of image restoration CNN** is way larger than BM3D.
- These CNN are impossible to run on a high-resolution images on a phone.
- Better control on the process.

We'll see more contexts that are not dominated by DL…

The geometric wisdom

Luca Magri

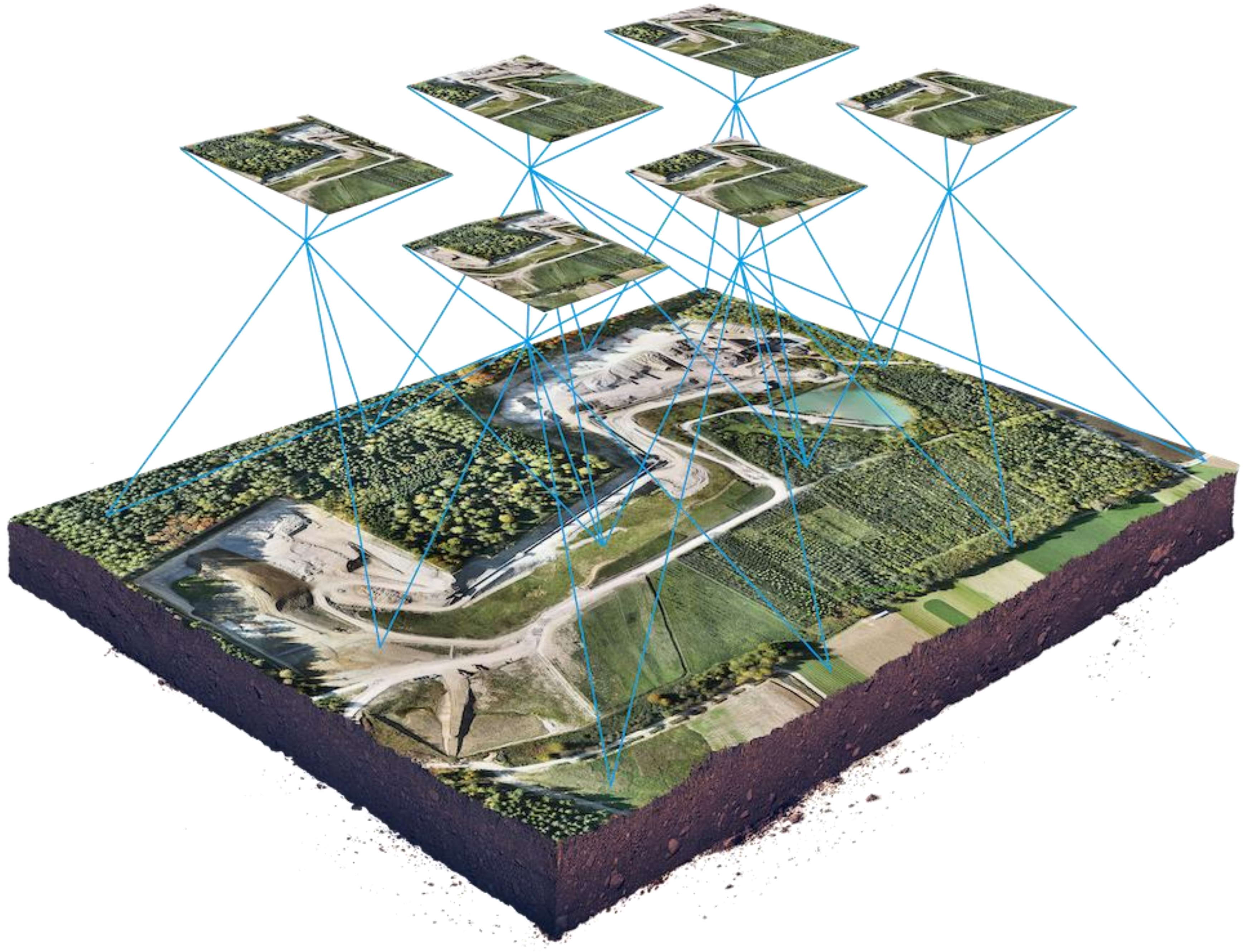
Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)

Politecnico di Milano

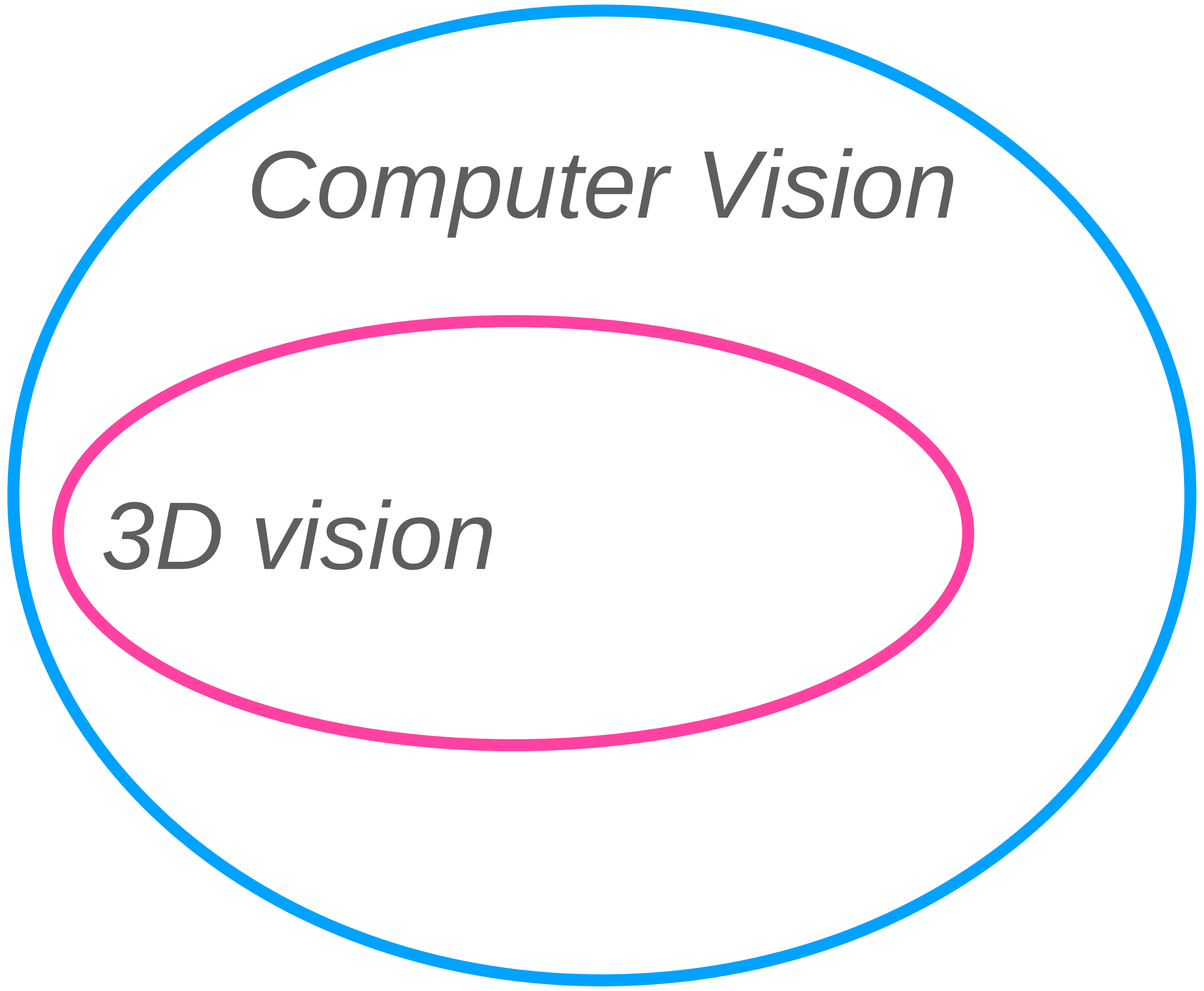
An ancient wisdom

Geometry

Photogrammetry



AI



~ 3000 - 300 BC

measure the farmlands flooded by the Nile

~ 1850

taking measurements from images

~ Since ‘70

Now

3D vision: from 2D to 3D

Input: unordered 2D images

Output: 3D point cloud

From 2D to 3D

Stereopsis: eyes perceive a pair of images processed in the brain to yield depth perception.

Similarly, geometric reasoning allow to:

1. Determine pairs of corresponding points
2. Calculate the position of the camera
3. Triangulate the 3D structure of the scene

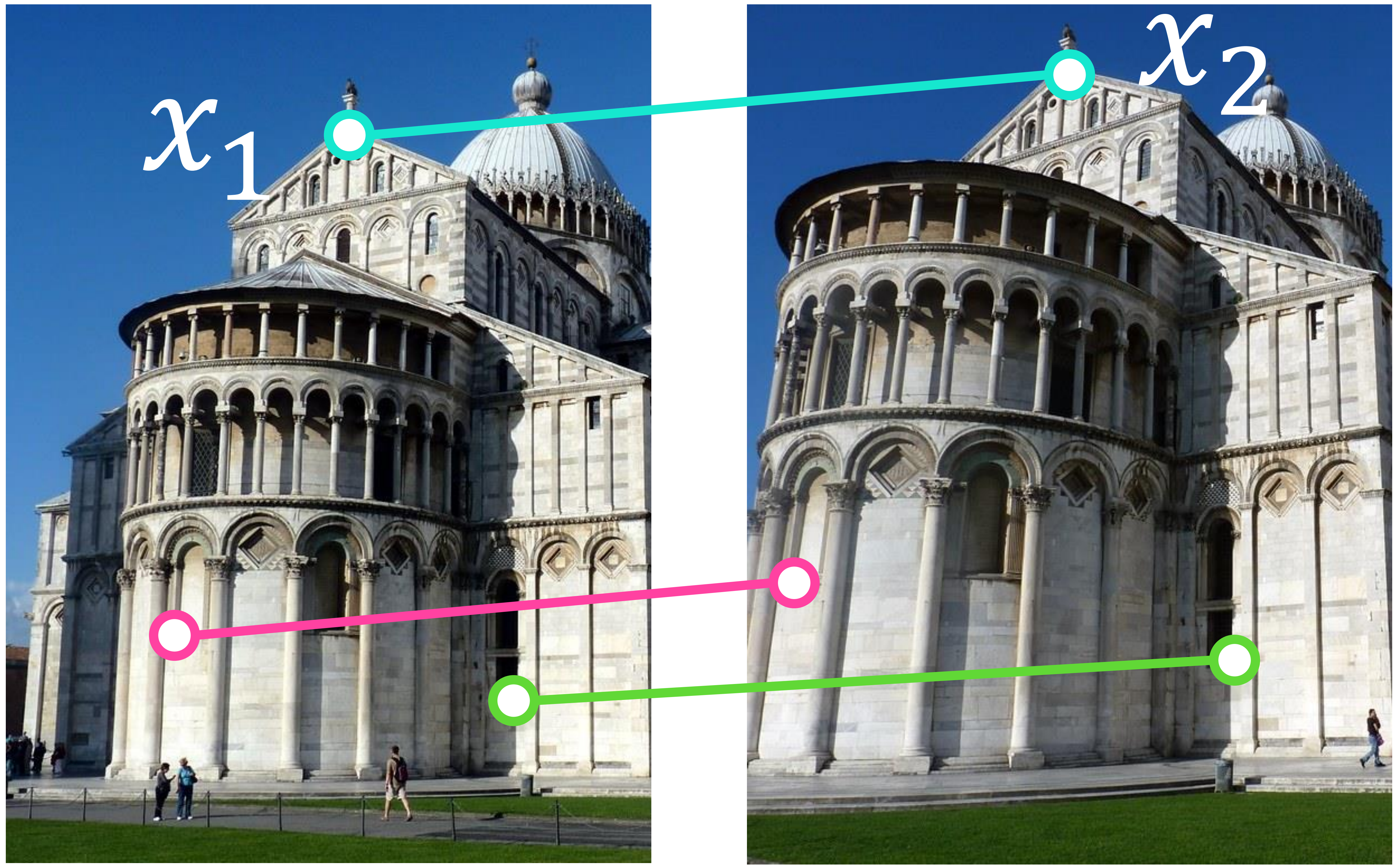
From 2D to 3D

Stereopsis: eyes perceive a pair of images processed in the brain to yield depth perception.

Similarly, geometric reasoning allow to:

1. Determine pairs of corresponding points
2. Calculate the position of the camera
3. Triangulate the 3D structure of the scene

Stereo Matching



$$x_2^\top F x_1 = 0$$

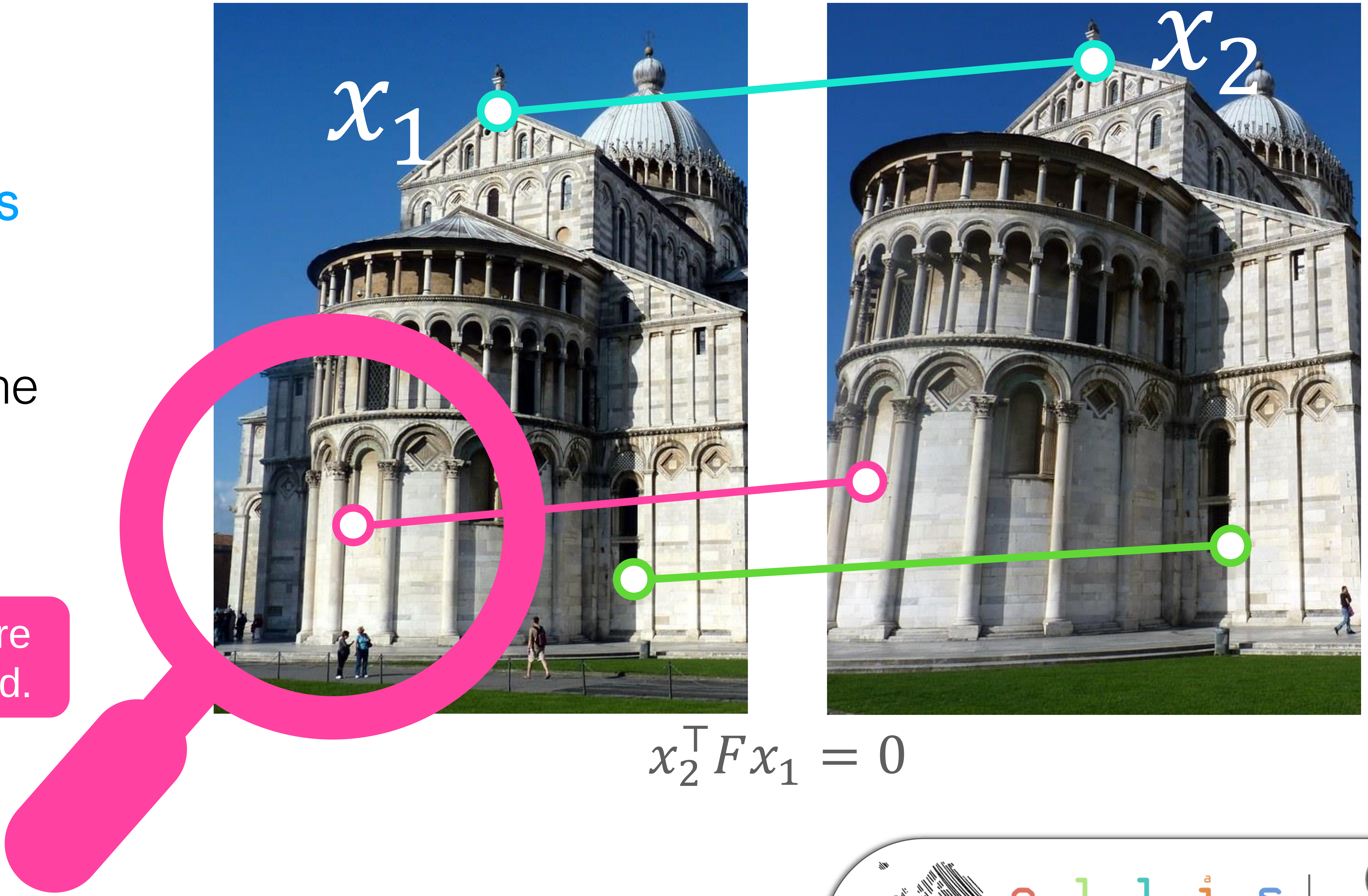
From 2D to 3D

Stereopsis: eyes perceive a pair of images processed in the brain to yield depth perception.

Similarly, geometric reasoning allow to:

1. Determine pairs of corresponding points
2. Calculate the position of the camera
3. Triangulate the 3D structure of the scene

Stereo Matching



$$x_2^\top F x_1 = 0$$

Hand-crafted descriptor & more recently learned ones are used.

From 2D to 3D

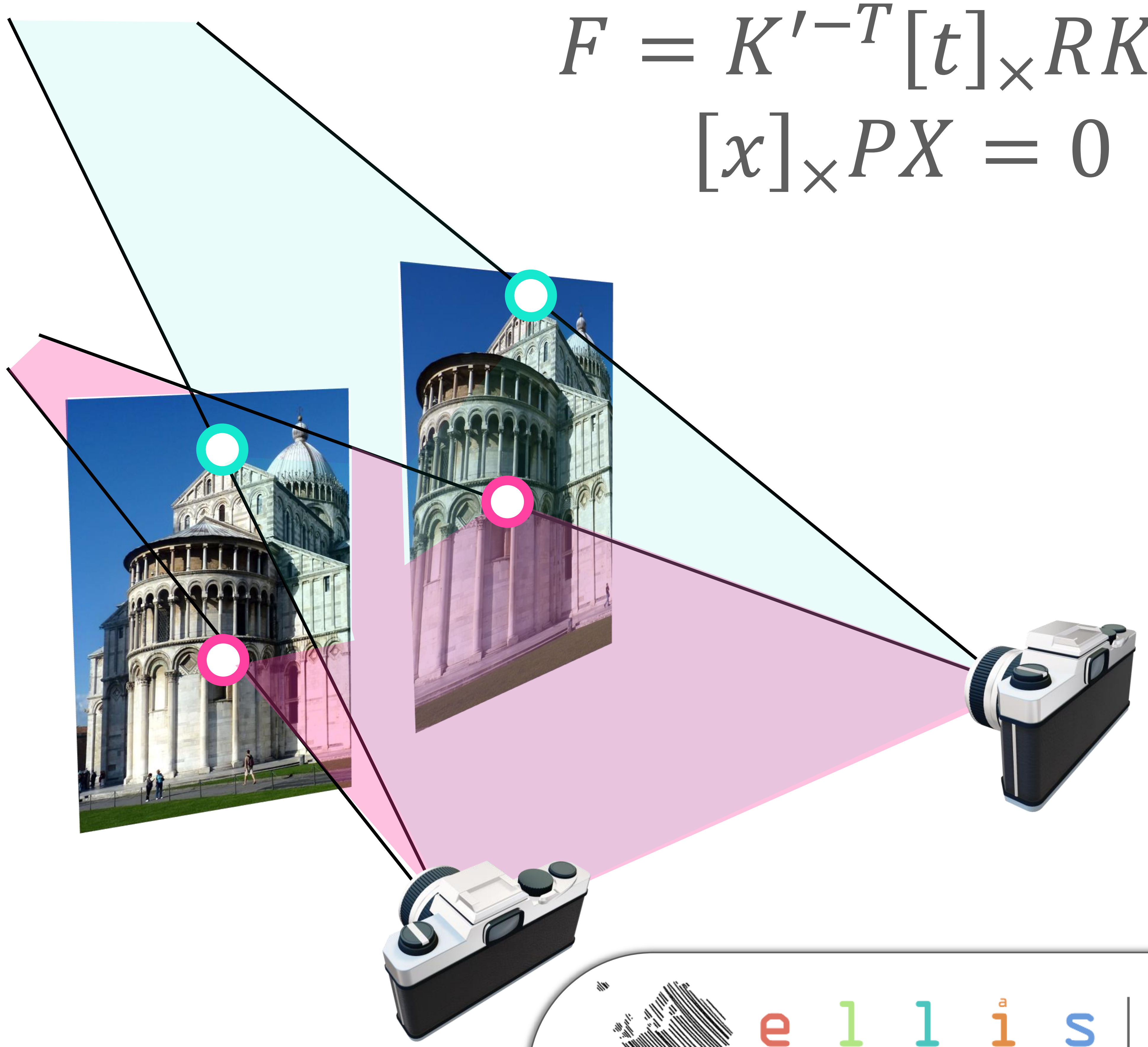
Stereopsis: eyes perceive a pair of images processed in the brain to yield depth perception.

Similarly, geometric reasoning allow to:

1. Determine pairs of corresponding points
2. Calculate the position of the camera
3. Triangulate the 3D structure of the scene

Camera localization

$$F = K'^{-T}[t]_{\times}RK^{-1}$$
$$[x]_{\times}PX = 0$$

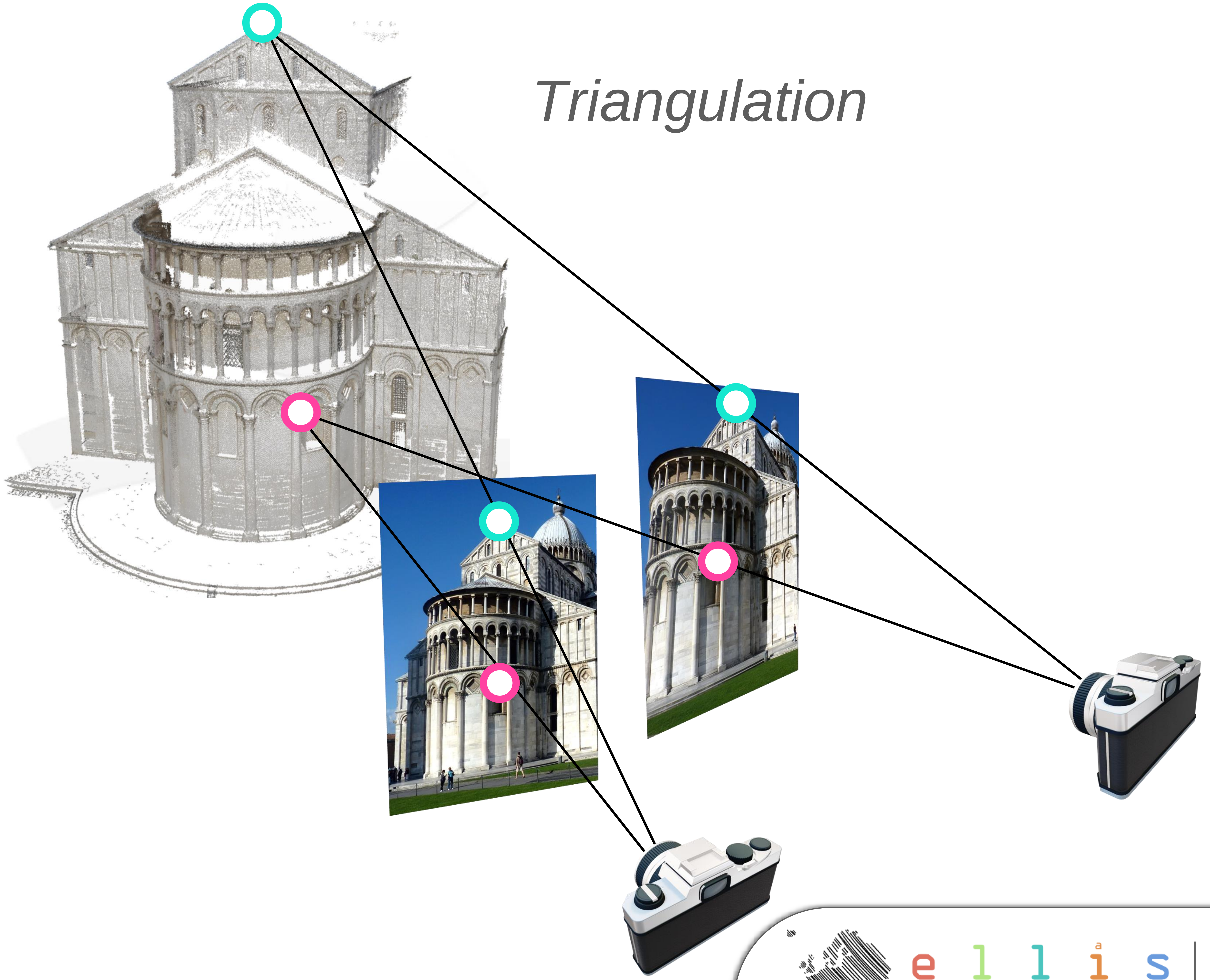


From 2D to 3D

Stereopsis: eyes perceive a pair of images processed in the brain to yield depth perception.

Similarly, geometric reasoning allow to:

1. Determine pairs of corresponding points
2. Calculate the position of the camera
3. Triangulate the 3D structure of the scene

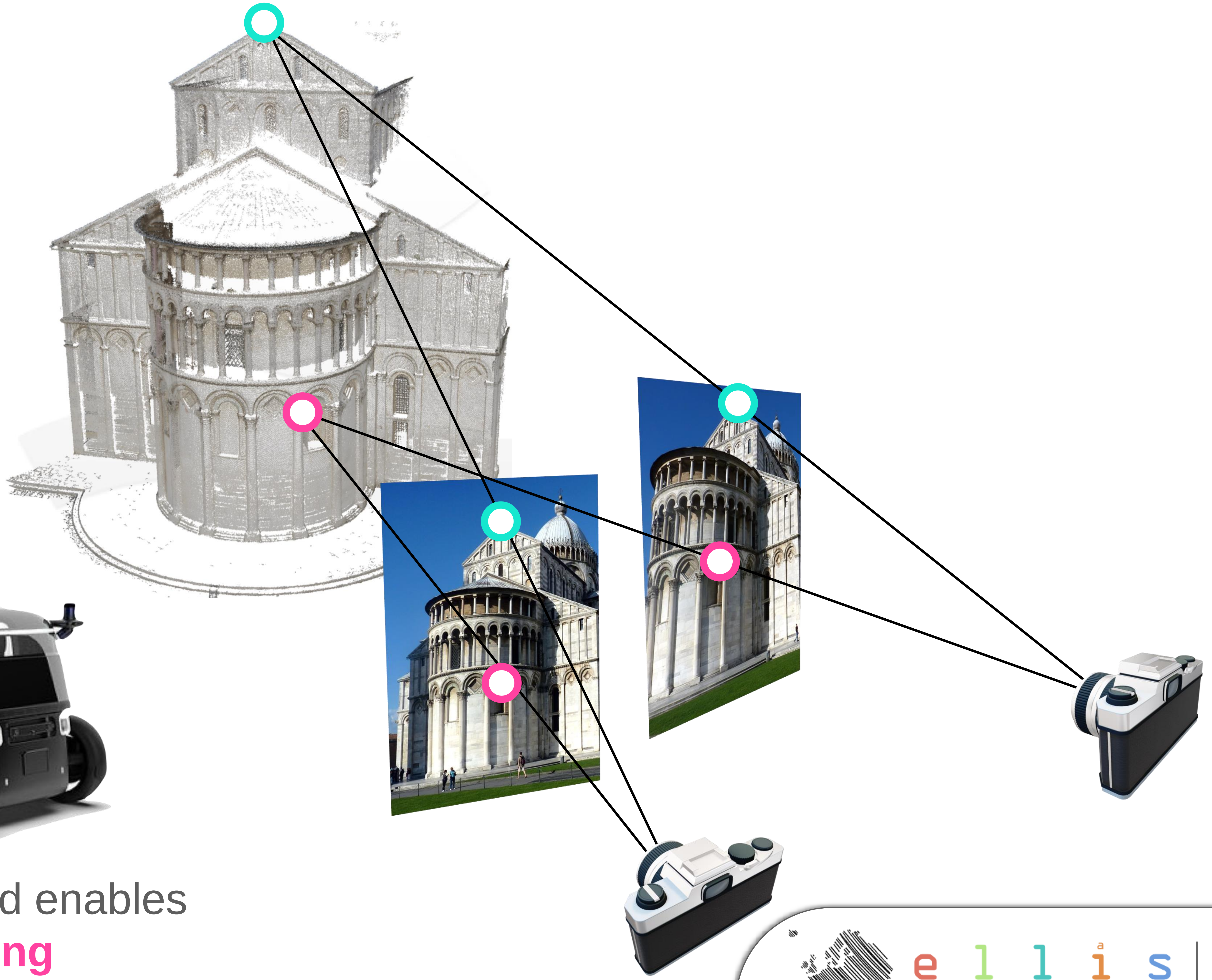


From 2D to 3D

Stereopsis: eyes perceive a pair of images processed in the brain to yield depth perception.

Similarly, geometric reasoning allow to:

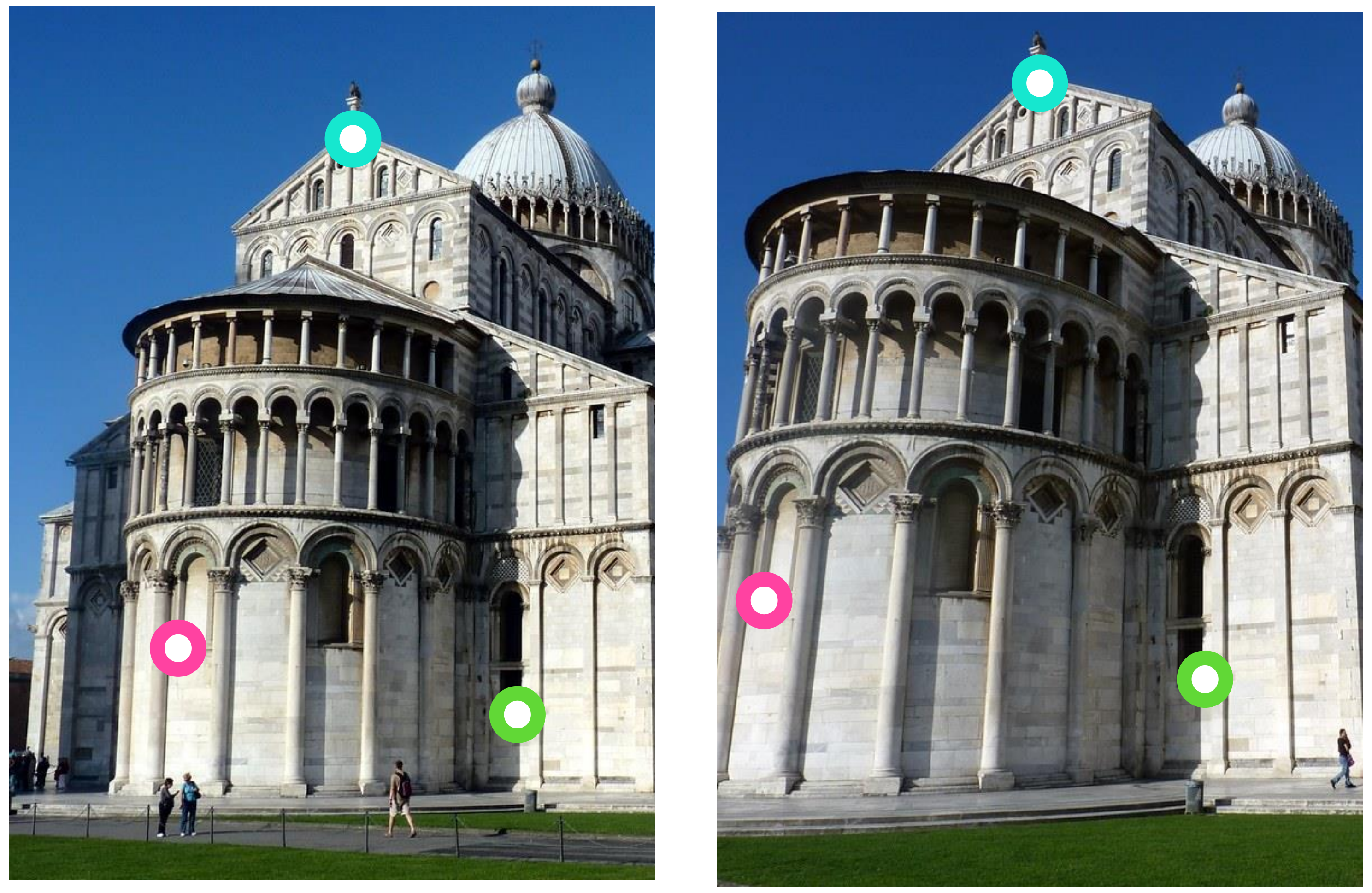
1. Determine pairs of corresponding points
2. Calculate the position of the camera
3. Triangulate the 3D structure of the scene



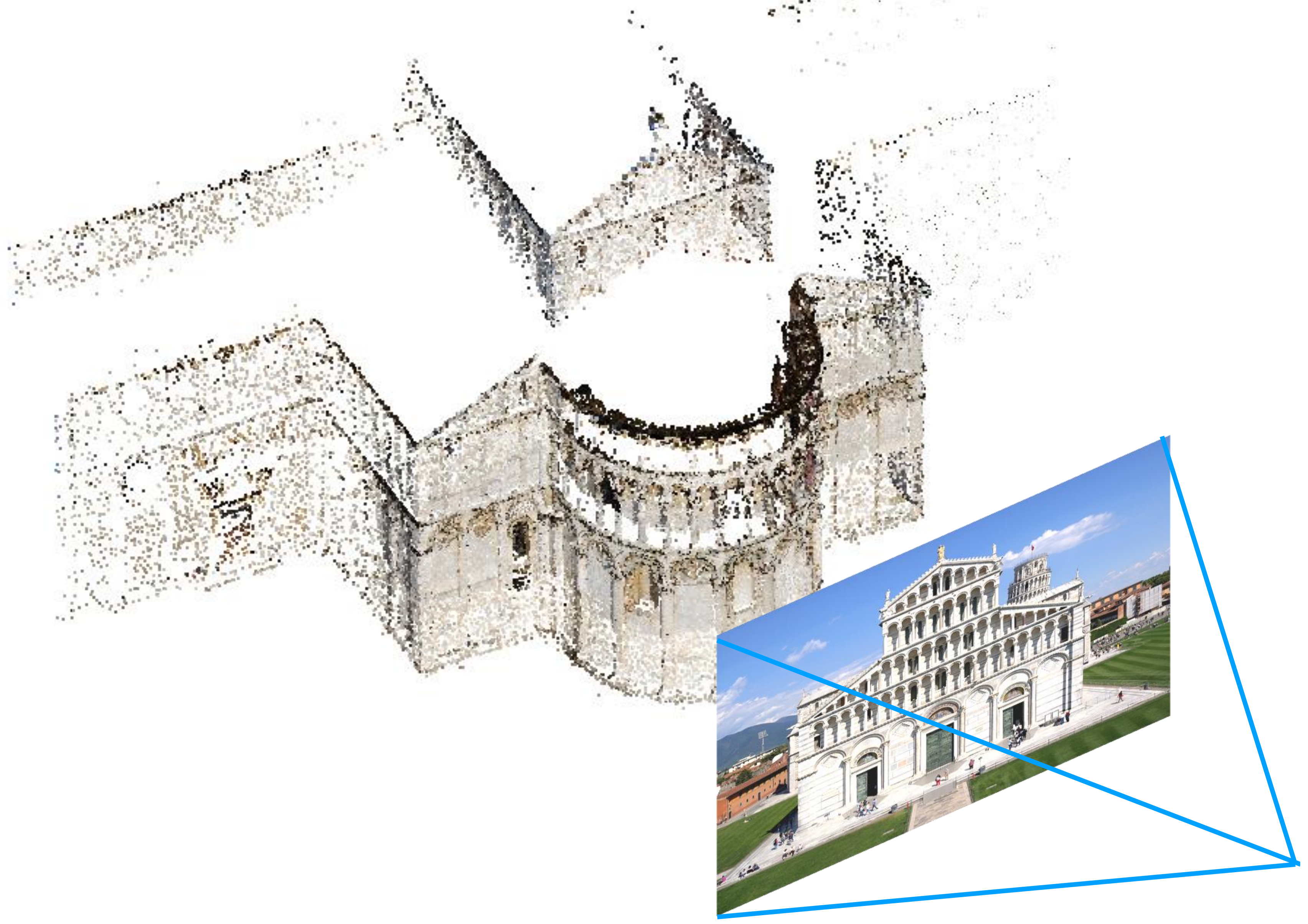
Geometry is employed by most **3D devices** and enables applications such as **autonomous driving**

Fitting geometric models

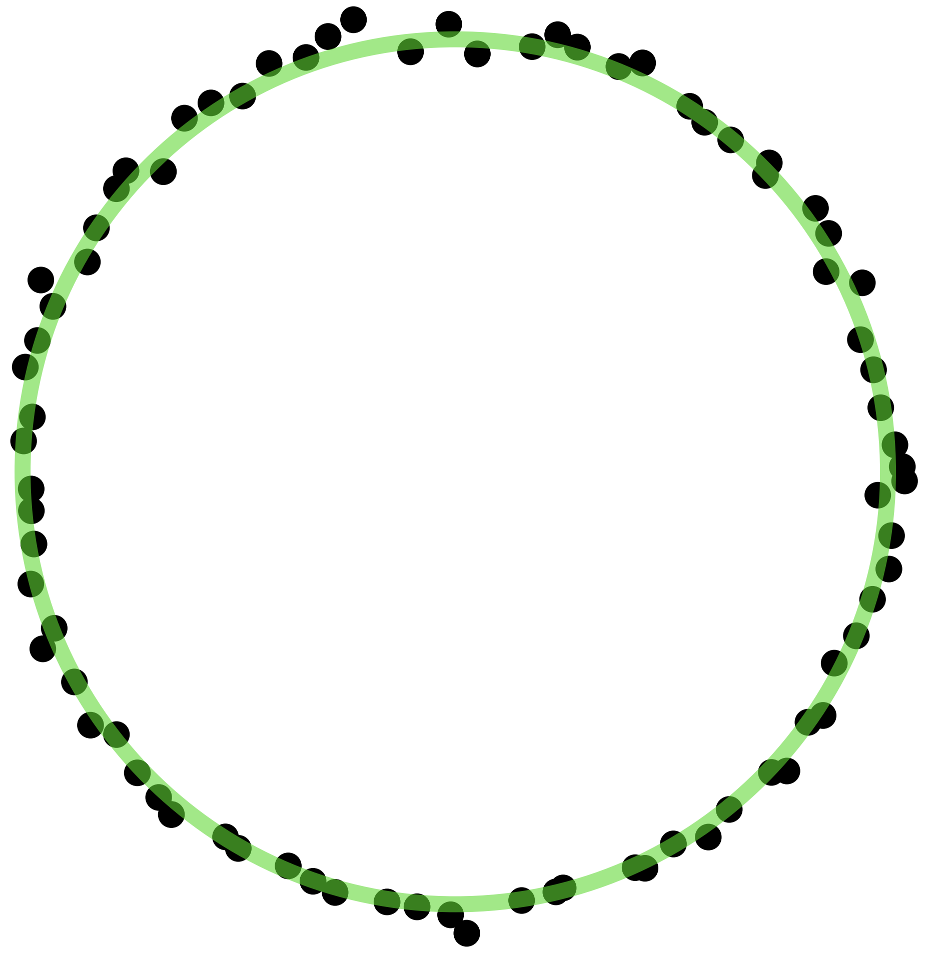
Stereo Matching



Camera localization



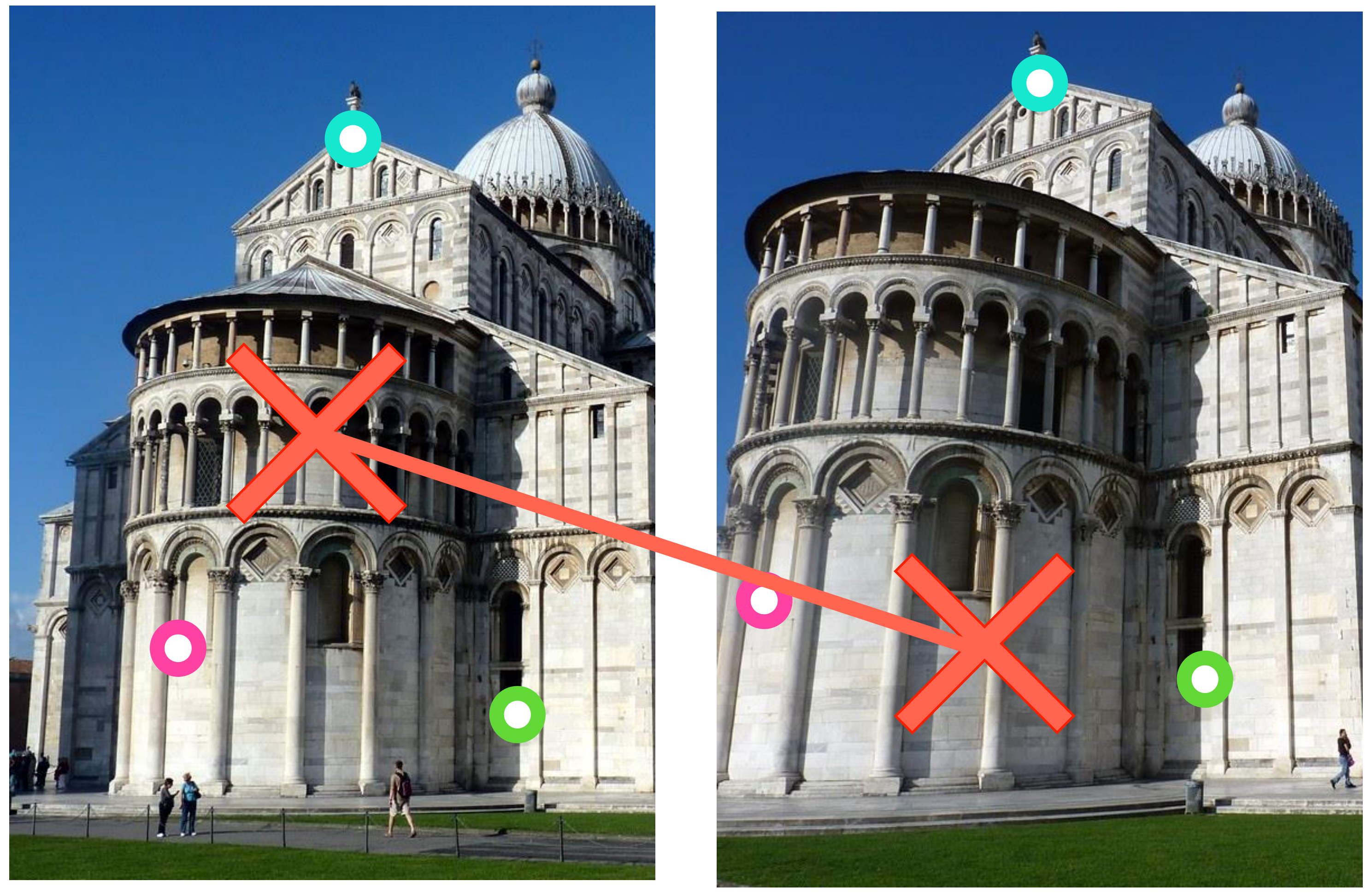
$$\hat{\theta} = \arg\min_{\theta} \ell(x, \theta)$$

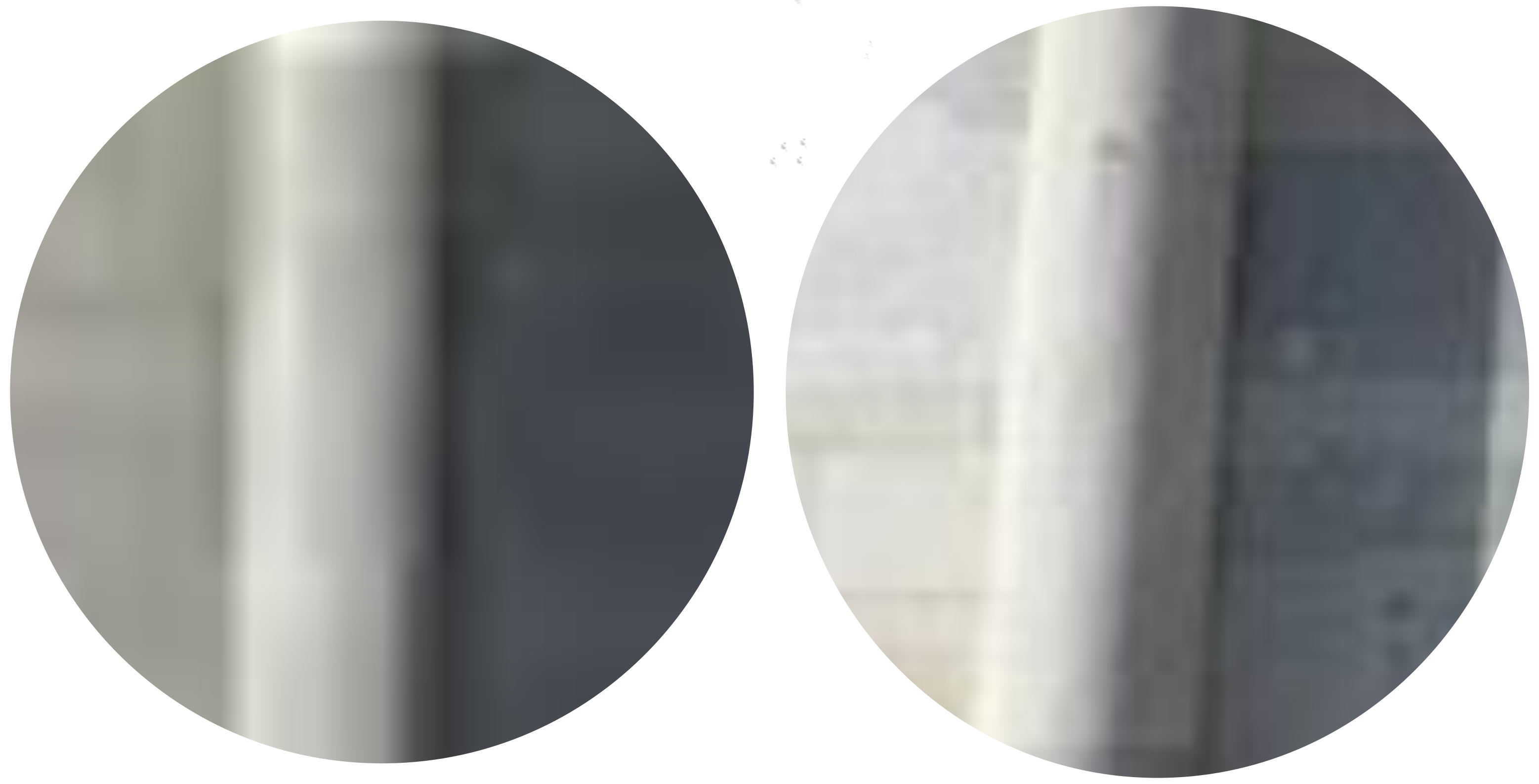


Geometric models	***Neural models***
Few parameters	Tons of parameters
Equations with clear geometric meaning	Often difficult to explain
Often admit closed-form solution	Require training

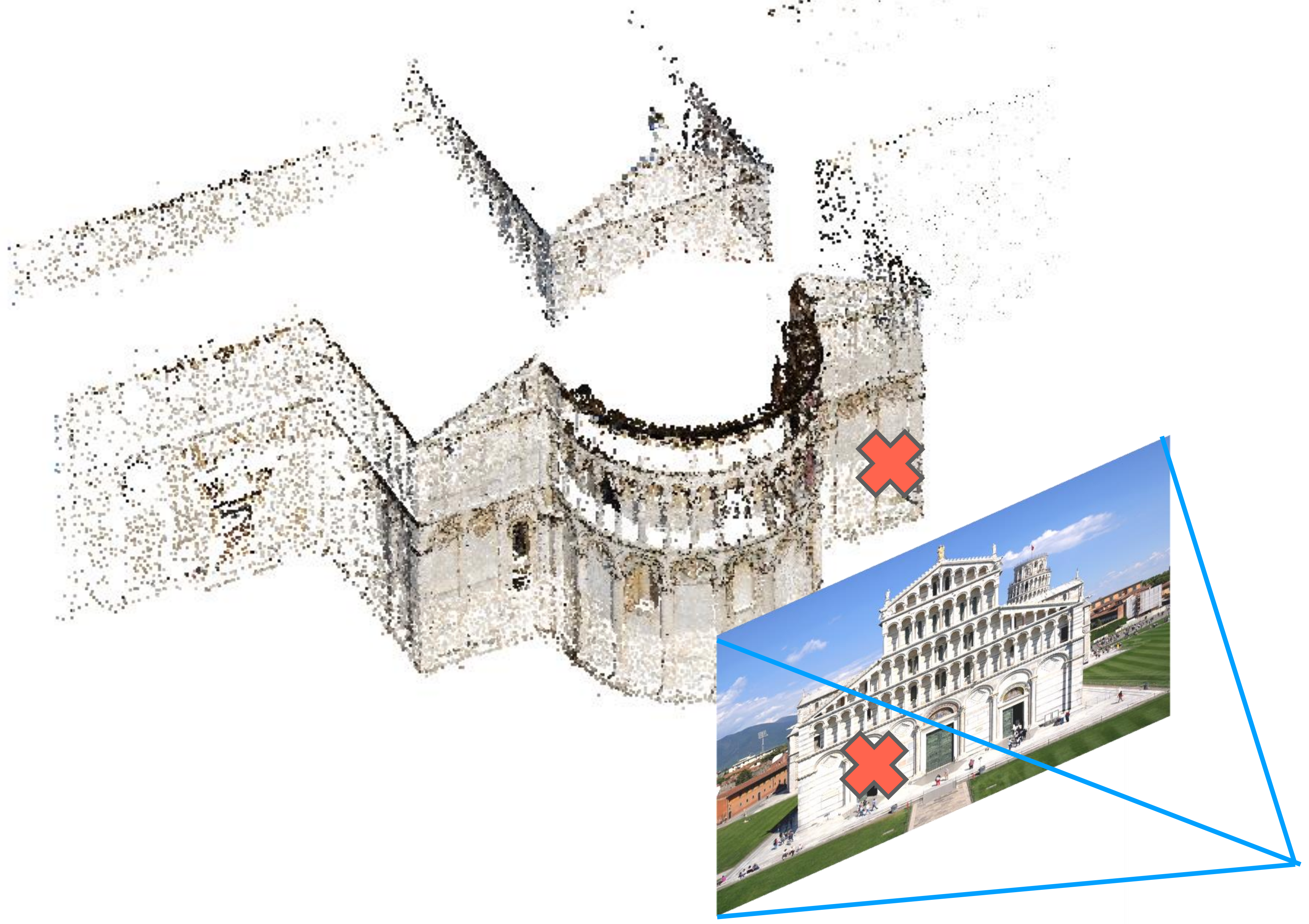
The need of robust fitting

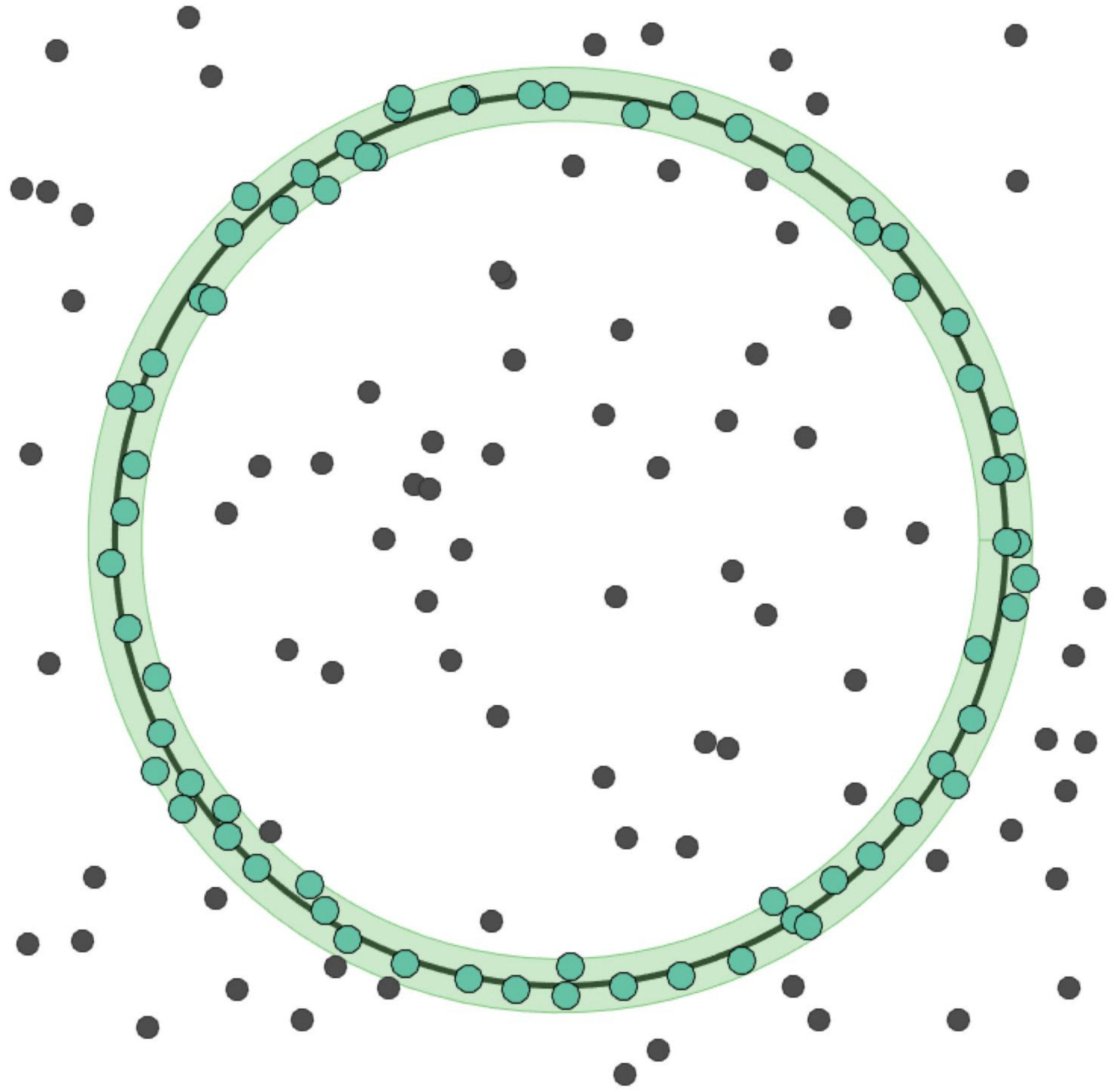
Stereo Matching





Camera localization



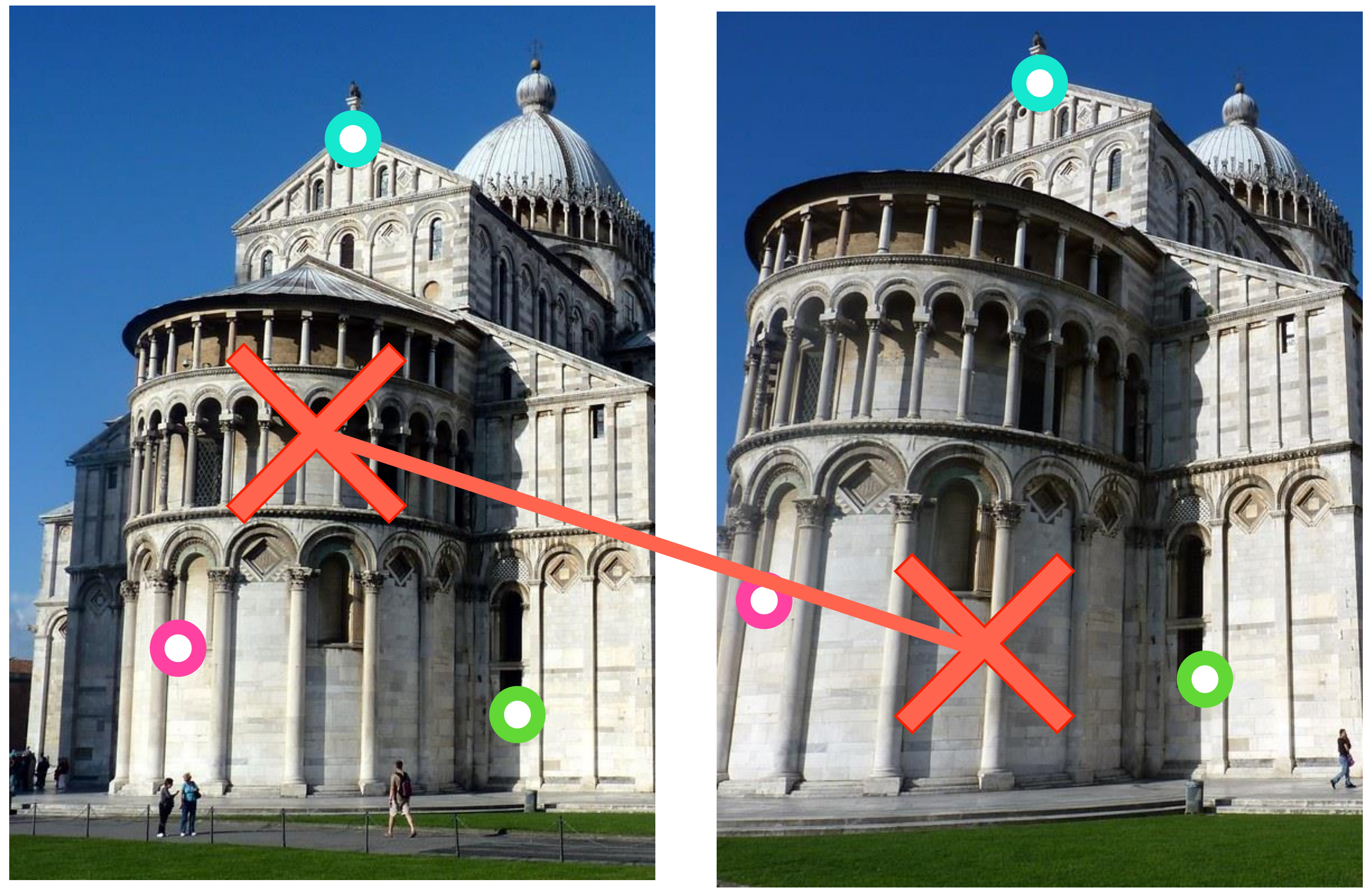


These estimation problems are **not trivial**

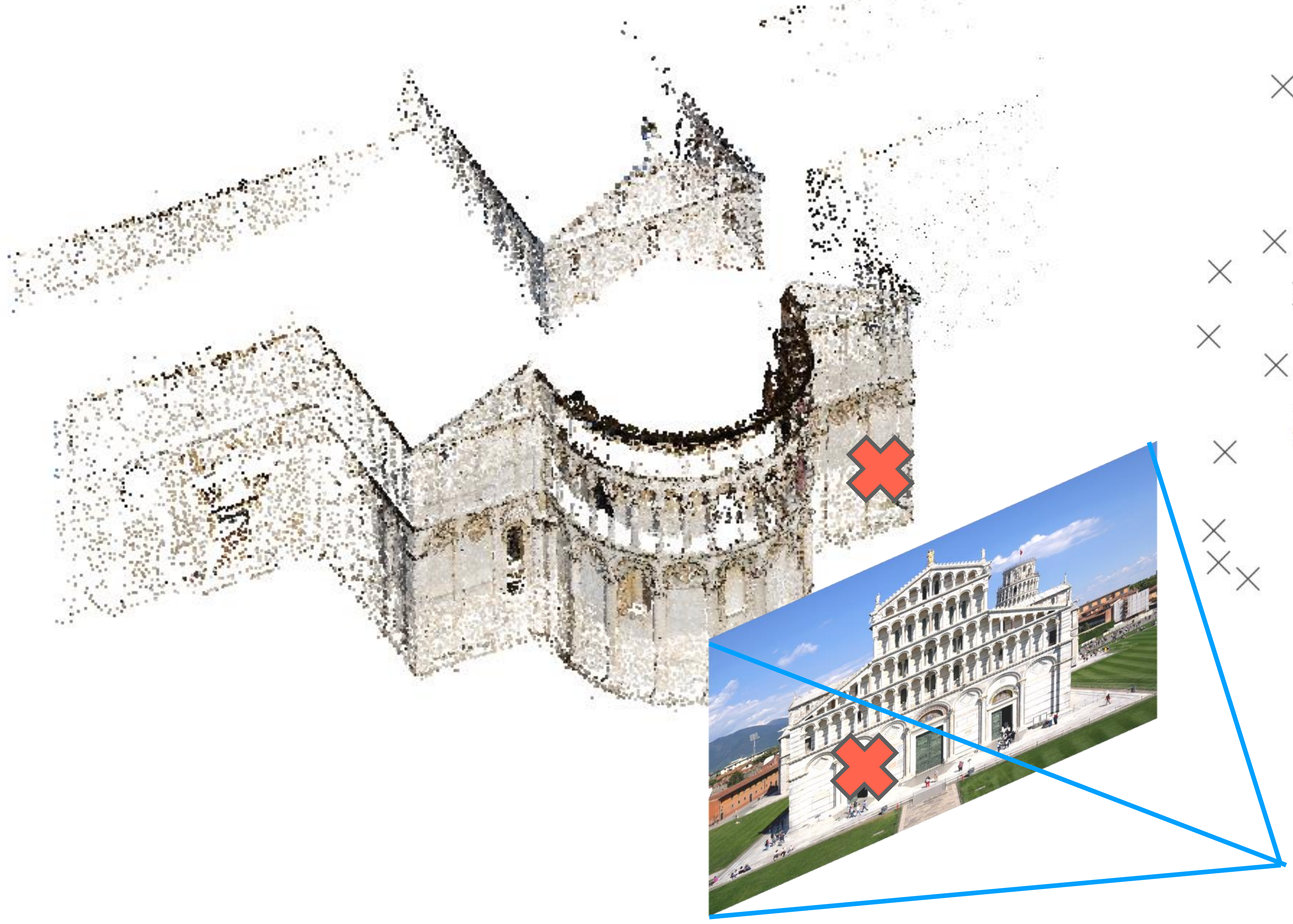
Large number of outliers (e.g., due to repeated structures) hinder geometric estimation.

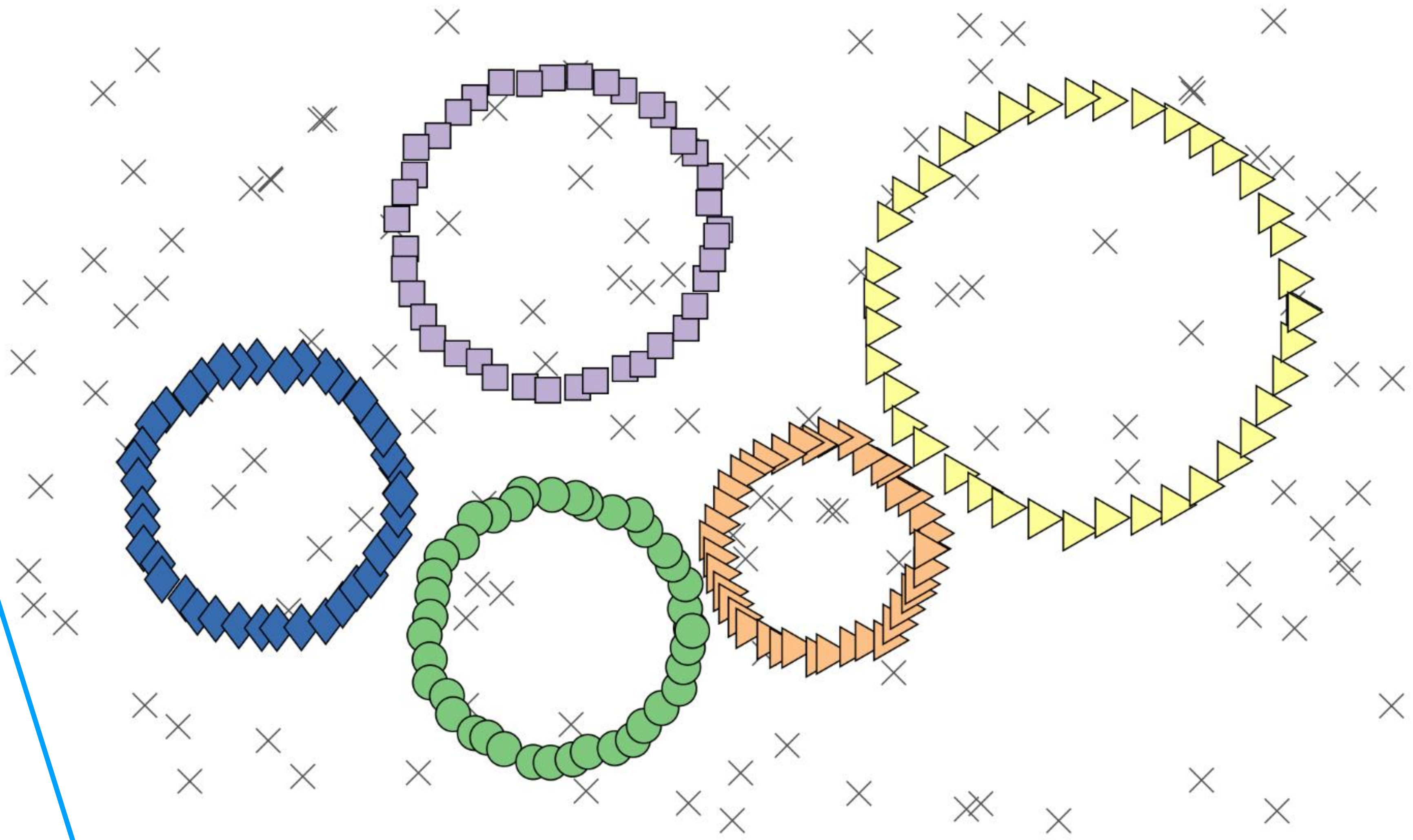
The need of robust fitting

Stereo Matching

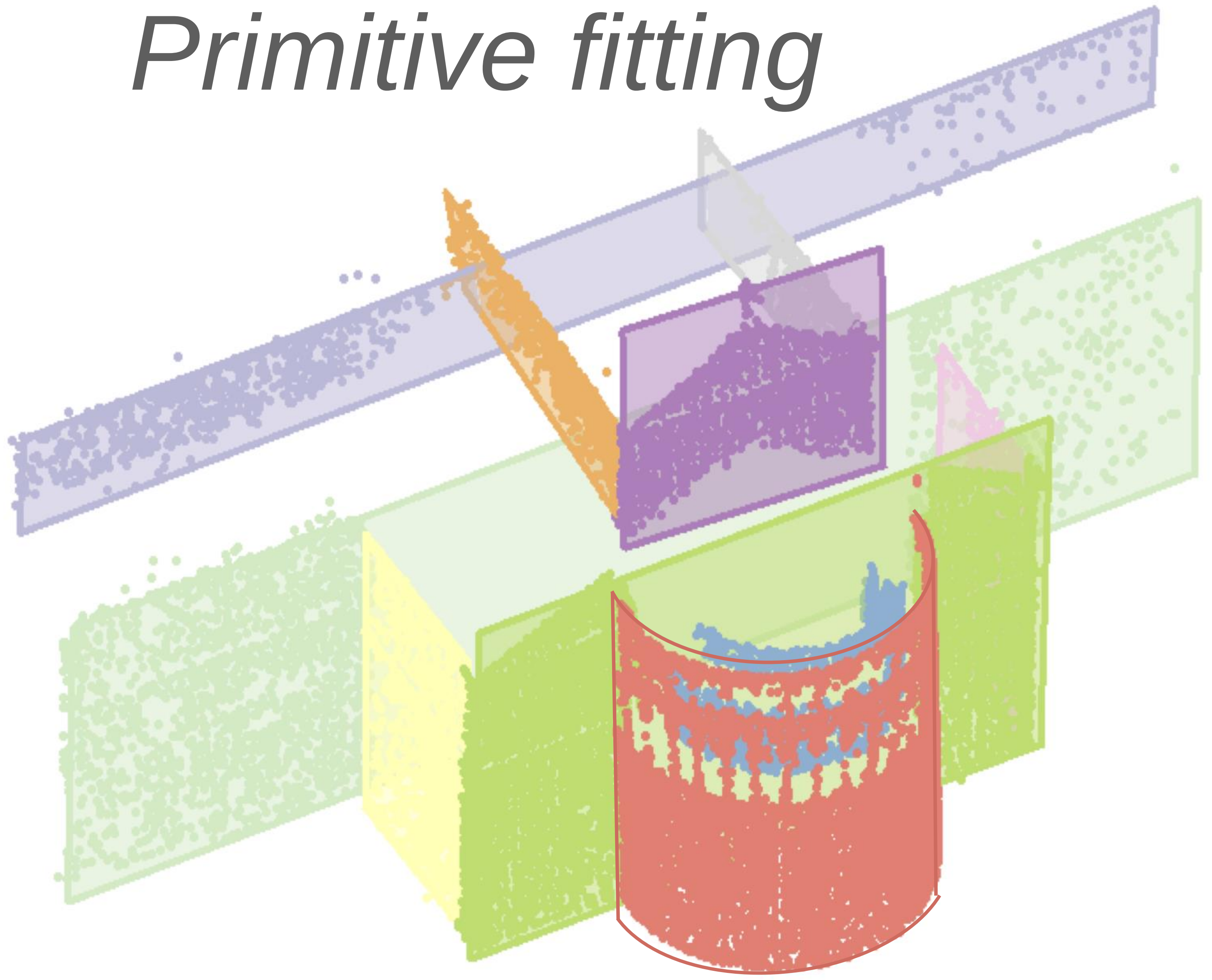


Camera localization





Primitive fitting



These estimation problems are **not trivial**

Large number of outliers (e.g., due to repeated structures) hinder geometric estimation.

Multi-model fitting has additional challenges:

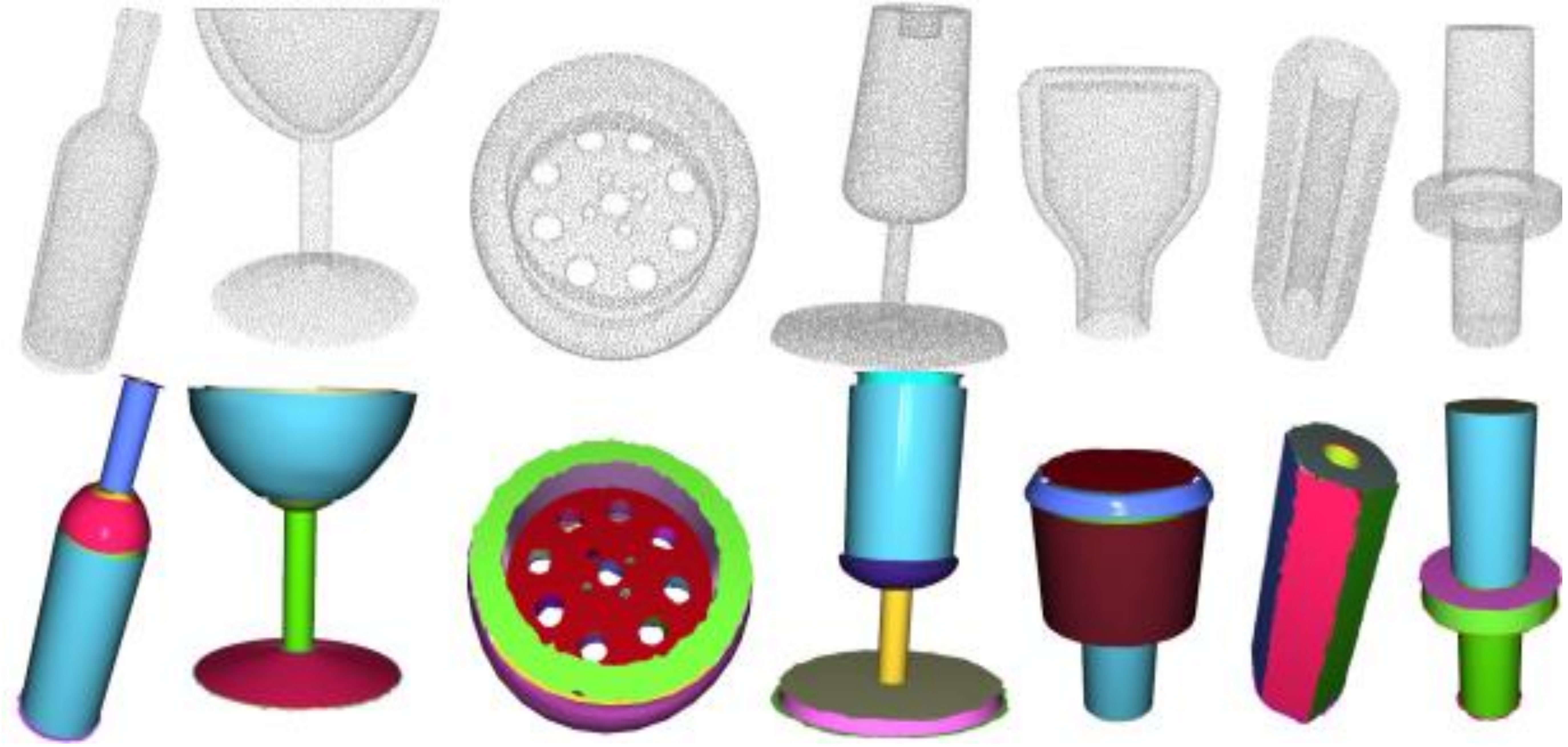
- Inliers to the other structure act as outliers
- Chicken & egg dilemma
- Ill-posedness: #models, type of models

A ubiquitous problem

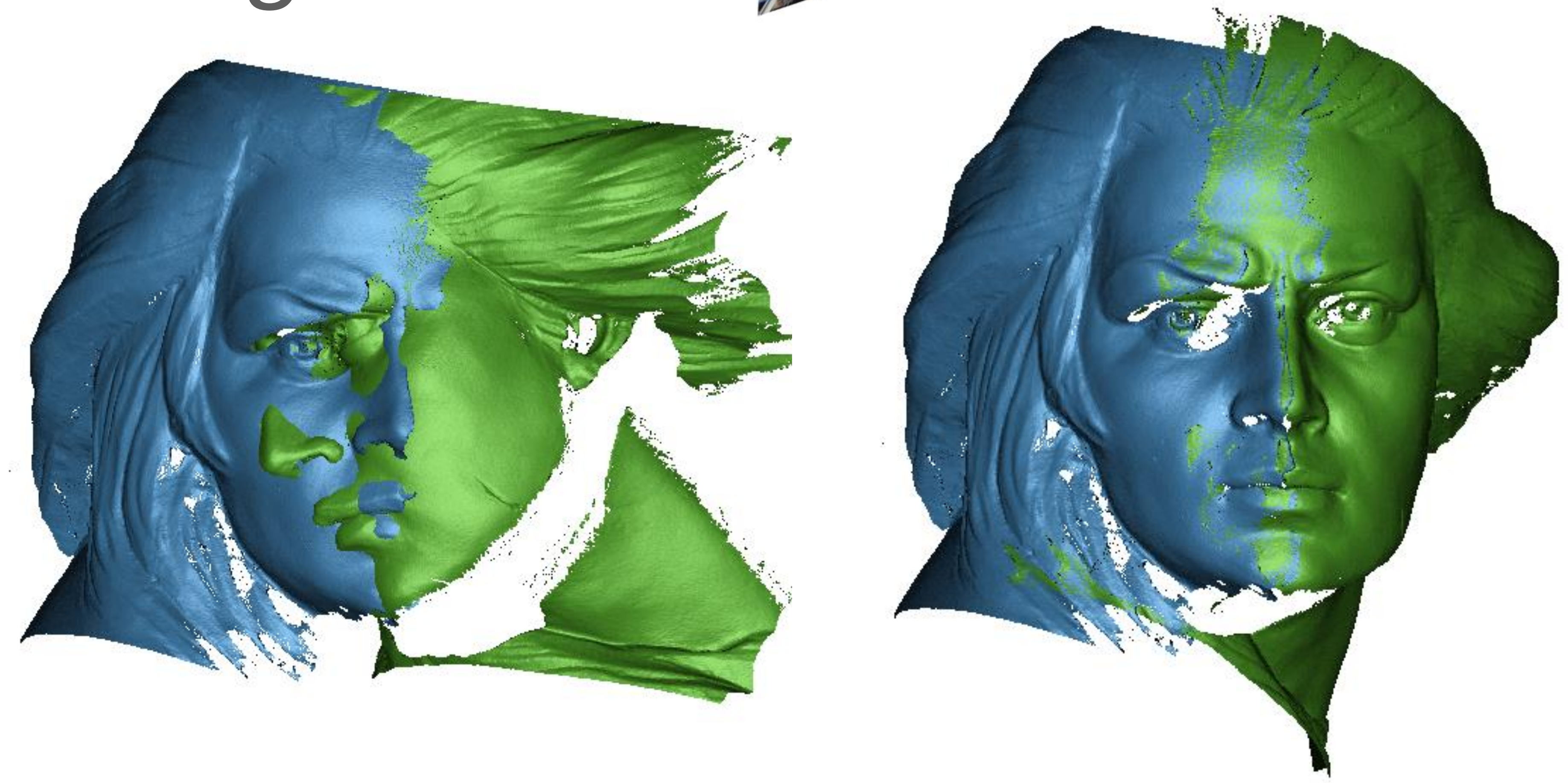
Motion segmentation

Mosaicking

Primitive decomposition



3D registration



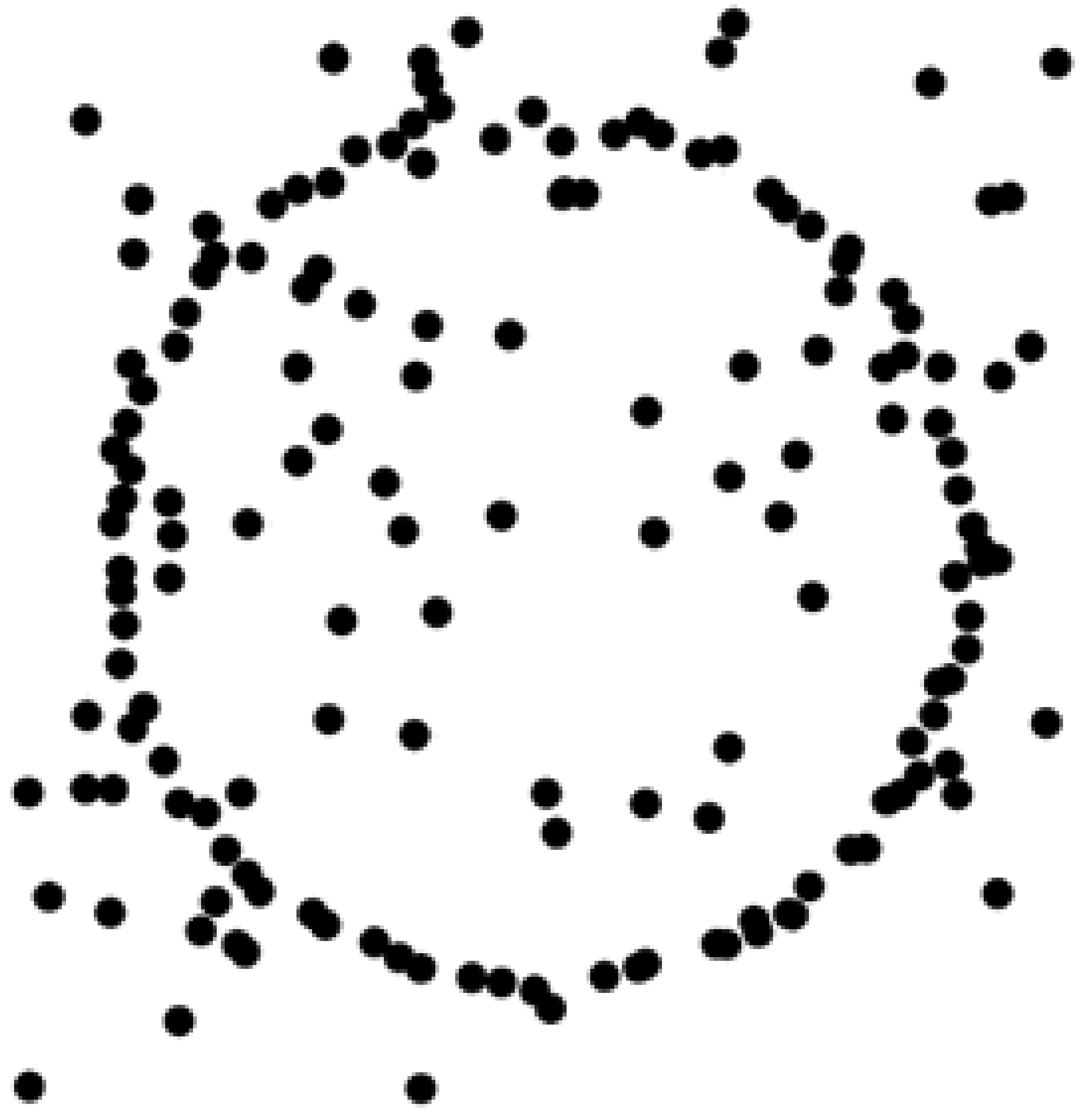
Consensus maximization

Randomize Sample Consensus: samples the space of possible models and, given an inlier threshold ϵ, keeps the model with the highest number of inliers (high consensus).

✓ Very popular approach (citations ~23K)

✓ General and versatile

x Does not cope well with multiple models:



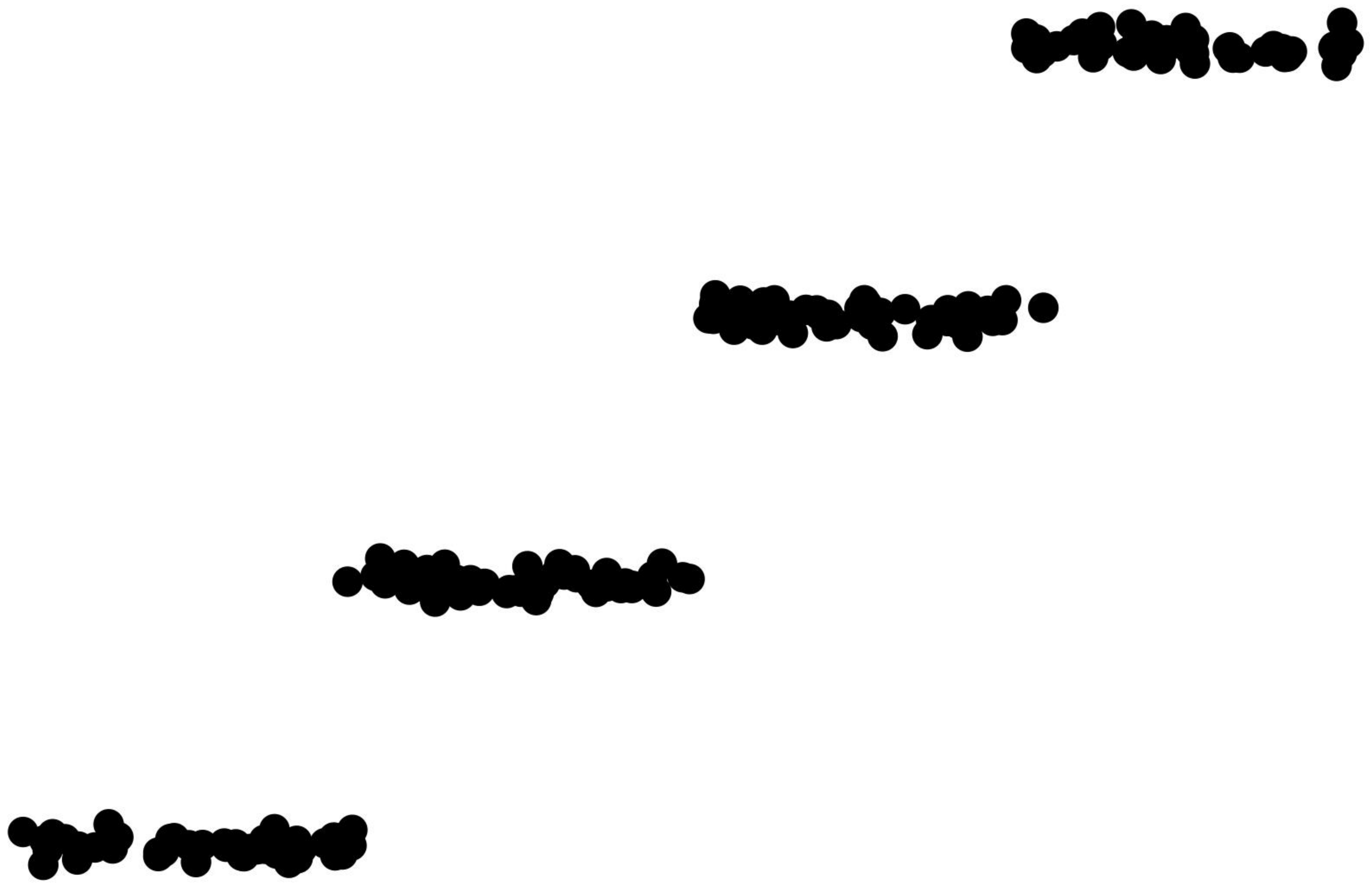
Consensus maximization

Randomize Sample Consensus: samples the space of possible models and, given an inlier threshold ϵ, keeps the model with the highest number of inliers (high consensus).

✓ Very popular approach (citations ~23K)

✓ General and versatile

x Does not cope well with multiple models:



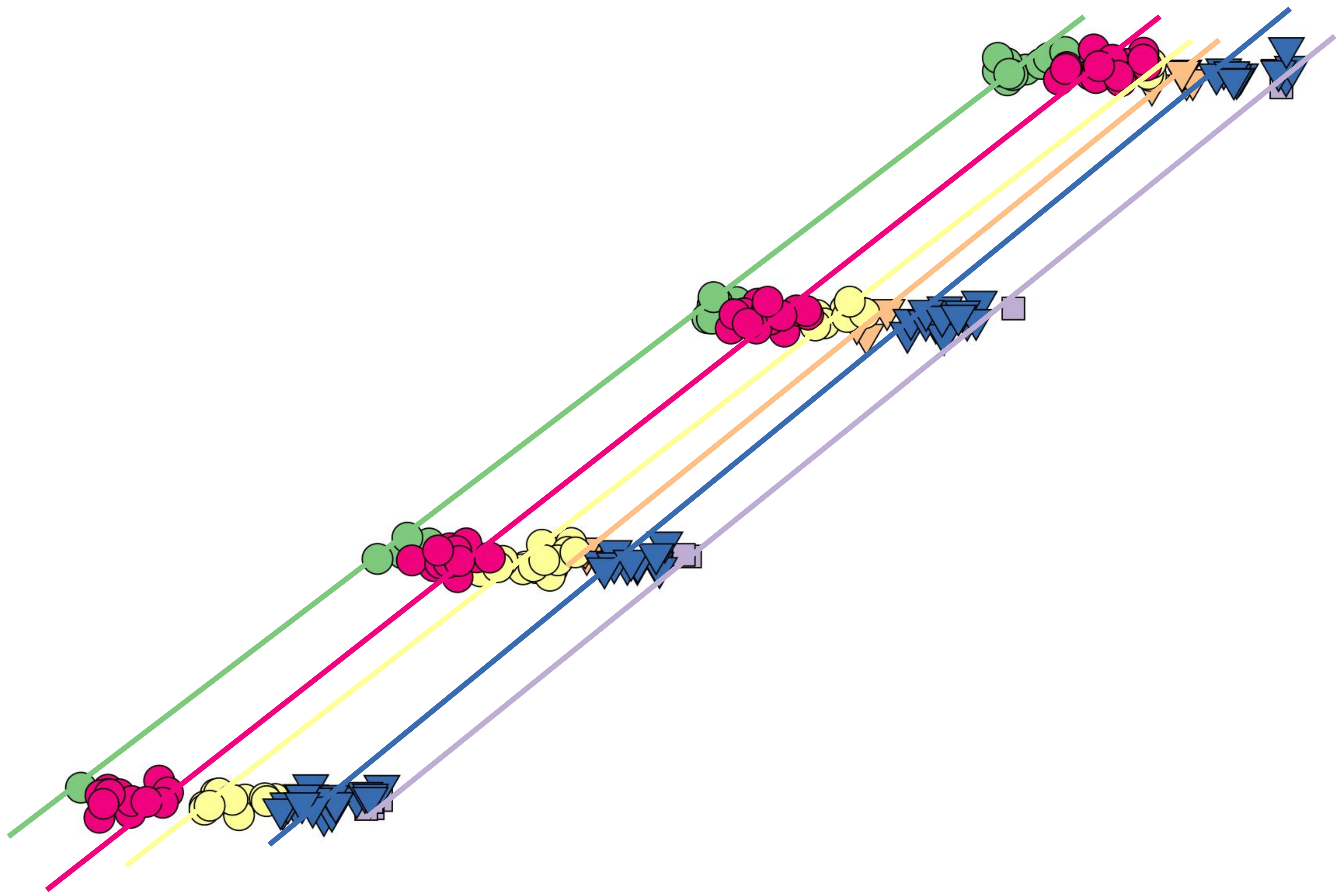
Consensus maximization

Randomize Sample Consensus: samples the space of possible models and, given an inlier threshold ϵ, keeps the model with the highest number of inliers (high consensus).

✓ Very popular approach (citations ~23K)

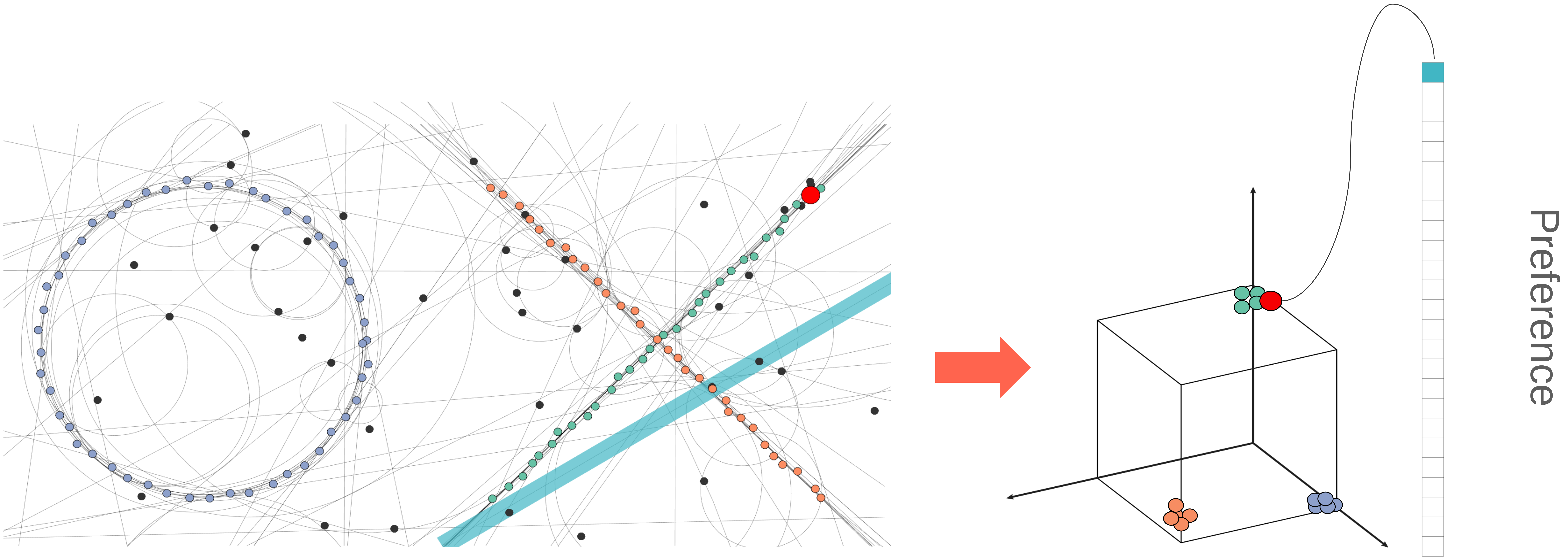
✓ General and versatile

x Does not cope well with multiple models:



Preference embedding

Every point is represented as the vector of probabilities of belonging to each of the sampled models.



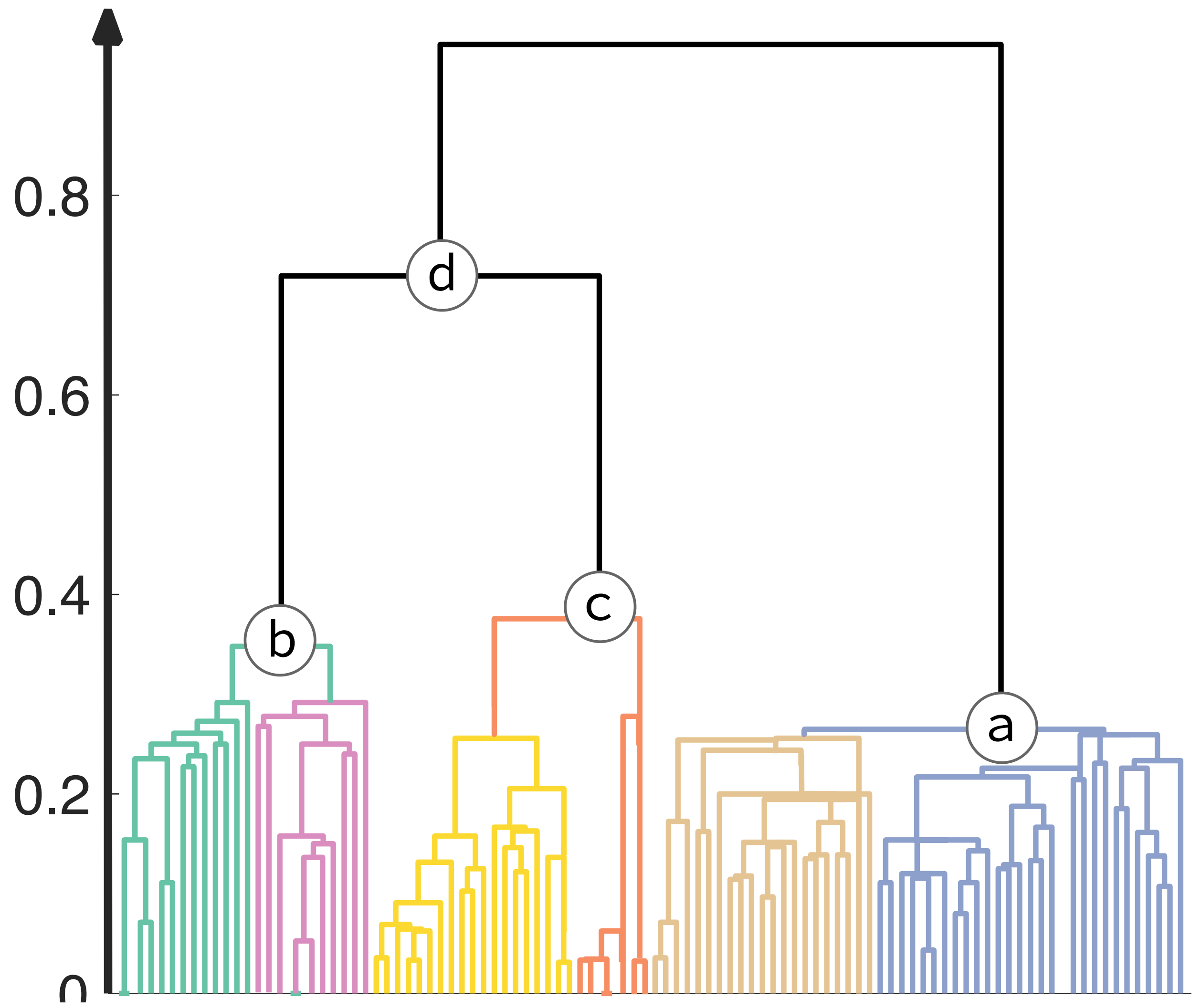
References

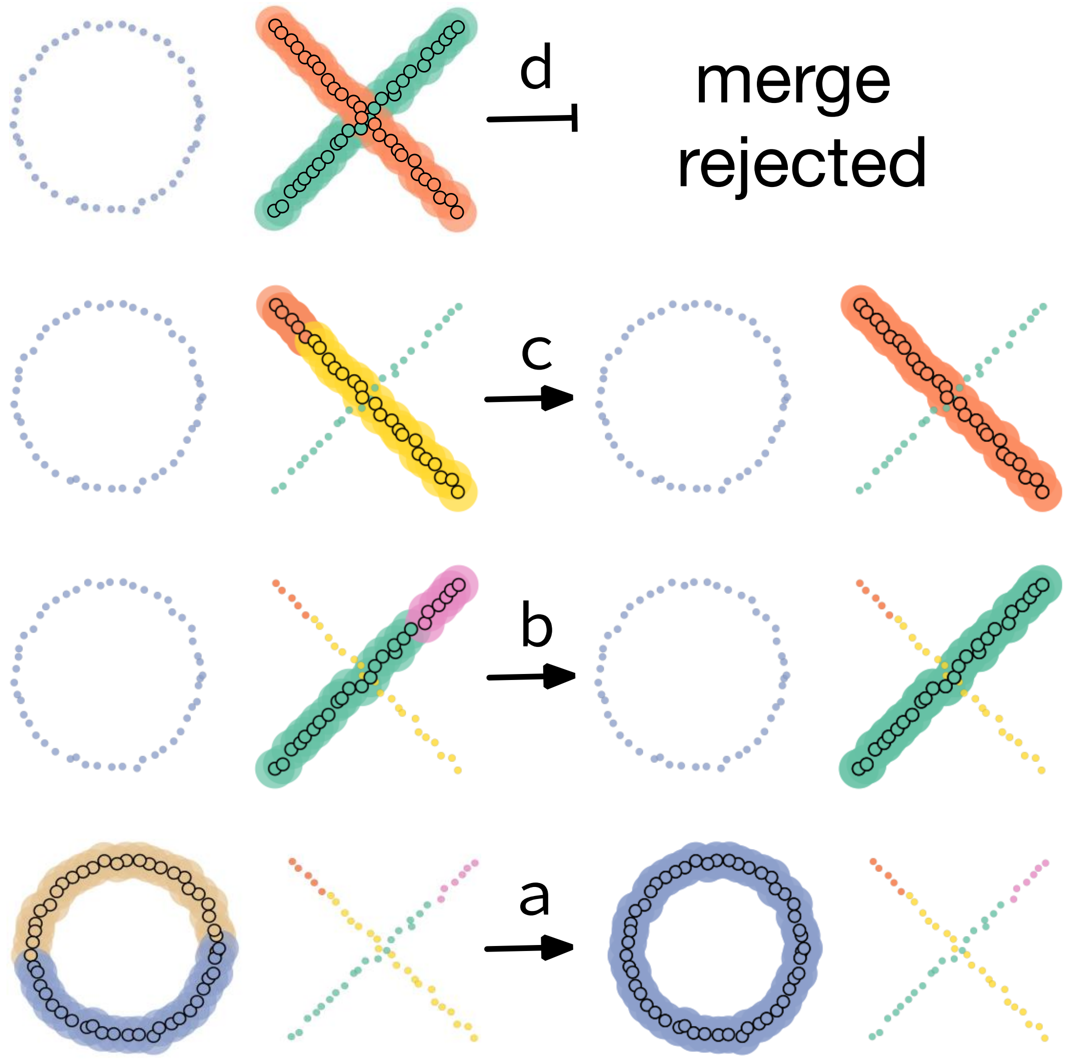
- Magri and Fusiello. ***T-linkage: A continuous relaxation of j-linkage for multi- model fitting***. CVPR 14
- Magri and Fusiello. ***Fitting Multiple Heterogeneous Models by Multi-Class Cascaded T-Linkage***. CVPR 19

Preference clustering

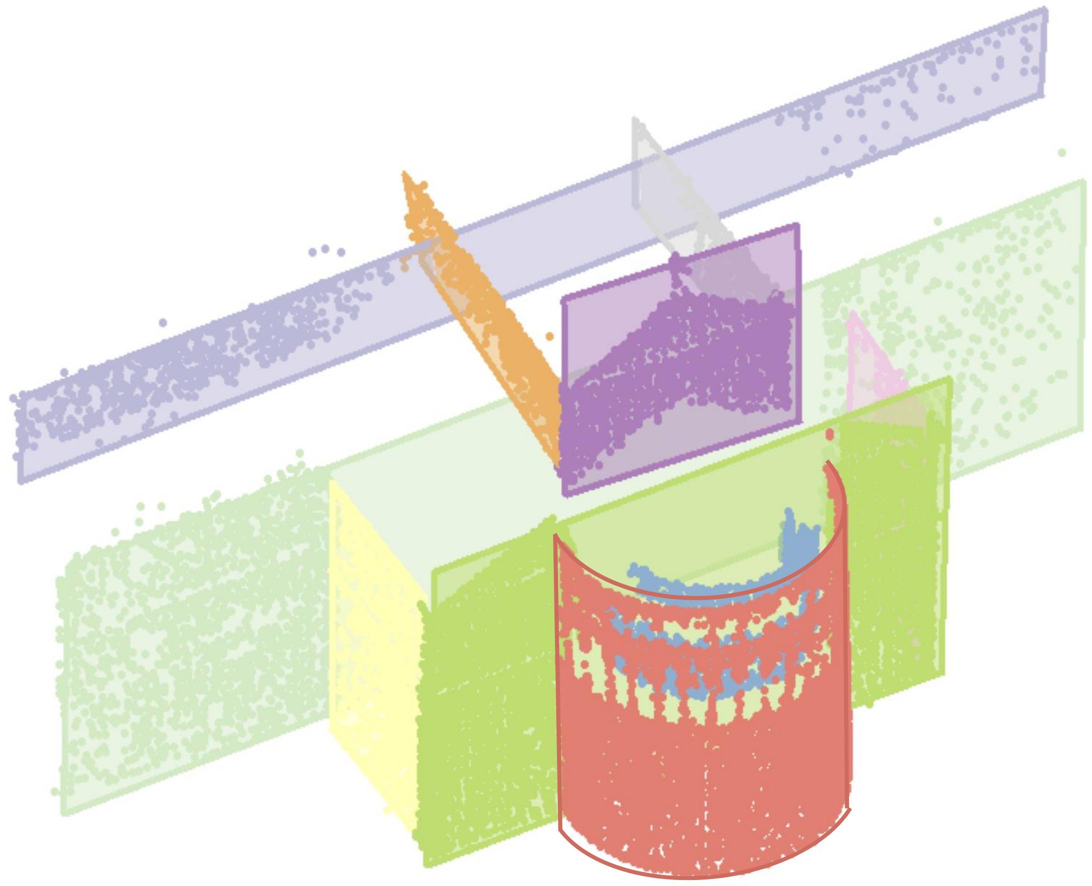
The more similar the preferences, the closer the points are.

We investigated different clustering formulations coupled with model selection criteria

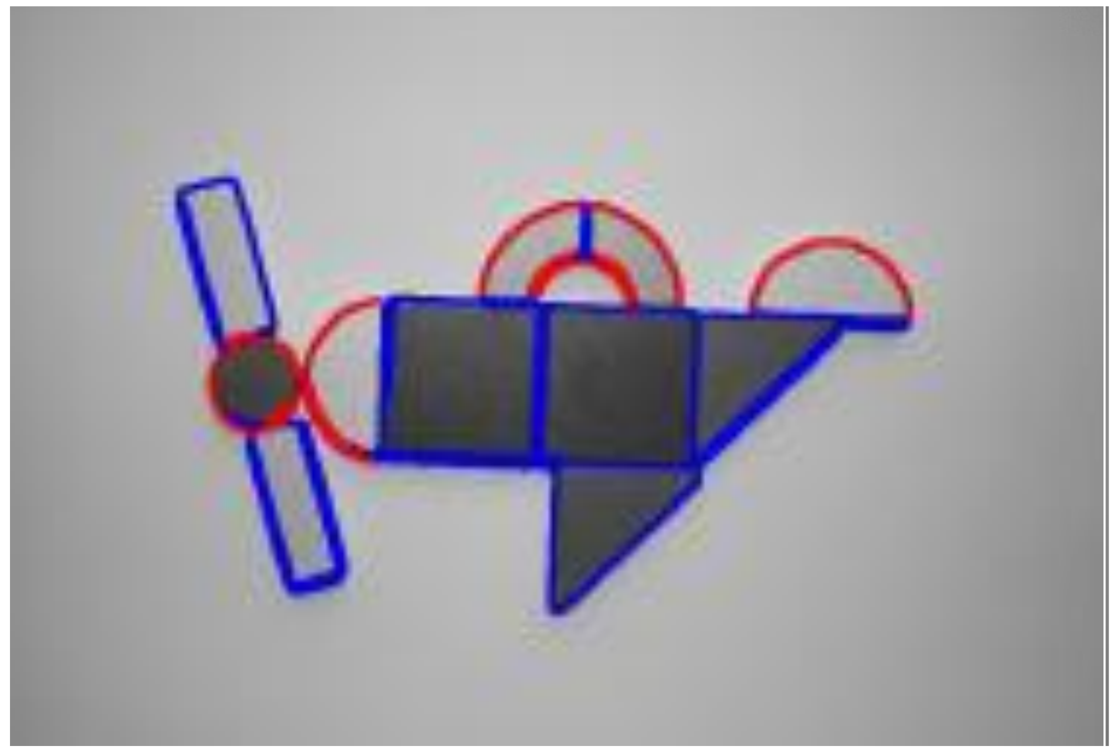




Plane or cylinder?



circle or line?

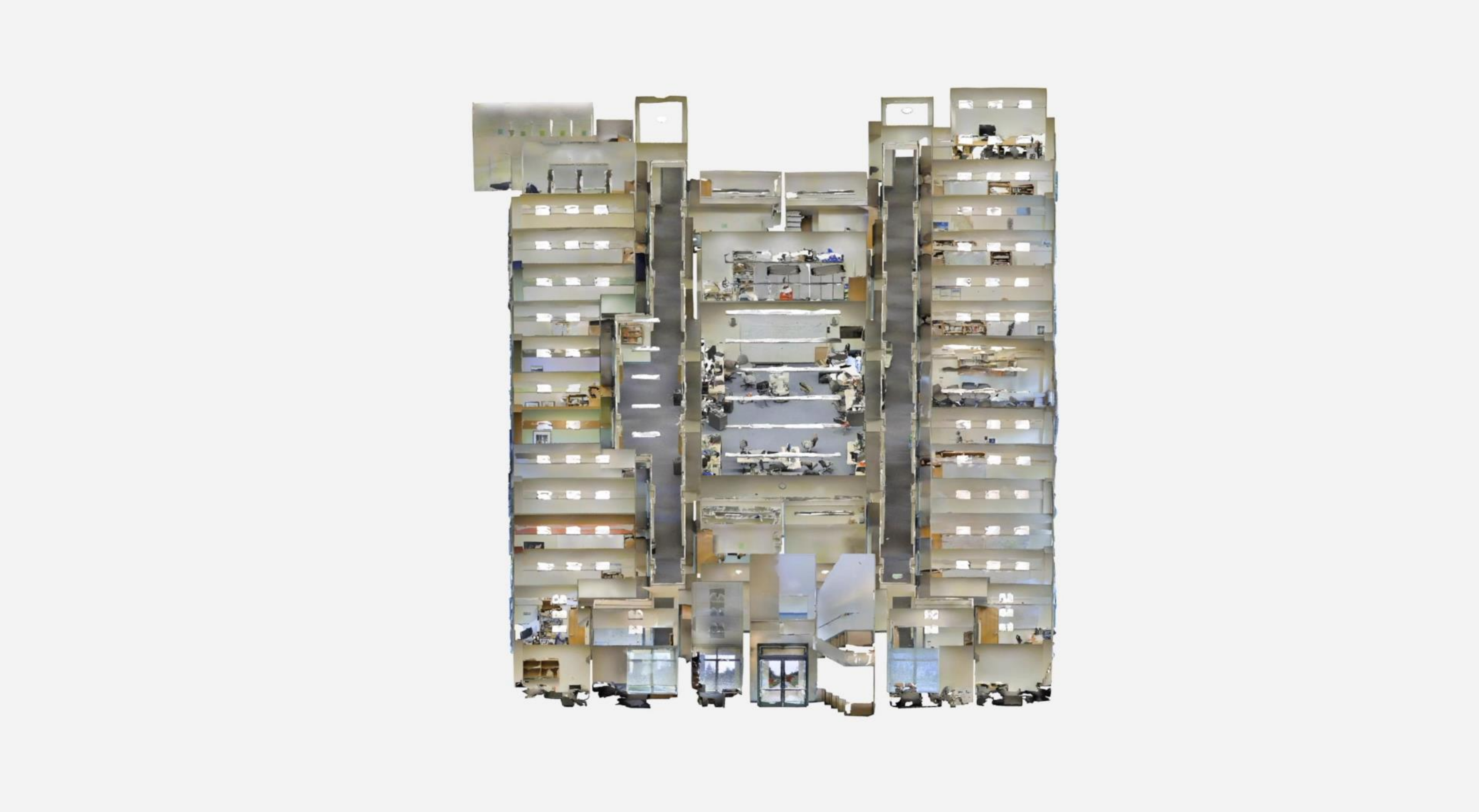


References

- Magri, Leveni, and Boracchi. ***MultiLink: Multi-class Structure Recovery via Agglomerative Clustering and Model Selection.*** CVPR 21
- Magri and Fusiello. ***Multi-model fitting as a Set Coverage problem***. CVPR 16

Industrial projects: Scan2bim

Goal: Automatically extract, from a scanned point cloud of an indoor space, building information such that you would typically find in a floor-plan.



References

- Magri and Fusiello. ***Reconstruction of interior walls from point cloud data with min-hashed J-linkage***. 3DV18

Industrial project: Template Matching

In collaboration with

T&O ITALIA

Industrial project: Template Matching

In collaboration with

T&O ITALIA

Passing on the geometric wisdom

Image Analysis and Computer Vision @Polimi (Prof. Vincenzo Caglioti)

- Around 20 years teaching CSE students
- AY 2022/2023 178 CSE enrolled students

Short courses (tutorials) (Federica Arrigoni & Luca Magri)

- *Inside Plato's door: a tour in multi-view geometry* (**CVPR 2022**)
- *Synchronization & Cycle Consistency in Computer Vision* (**CVPR 2020**)
- *Synchronization: a general framework for mosaicking, 3D reconstruction, matching & segmentation problems* (**ICPR 2020**)
- *Parametric model fitting* (**ICPR 2020**)

PhD Course – **Geometric Computer Vision: from images to 3D models** (Federica Arrigoni & Luca Magri)

- University of Trento (2021)
- Sapienza University of Rome (2022)

The geometric wisdom

Federica Arrigoni

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)

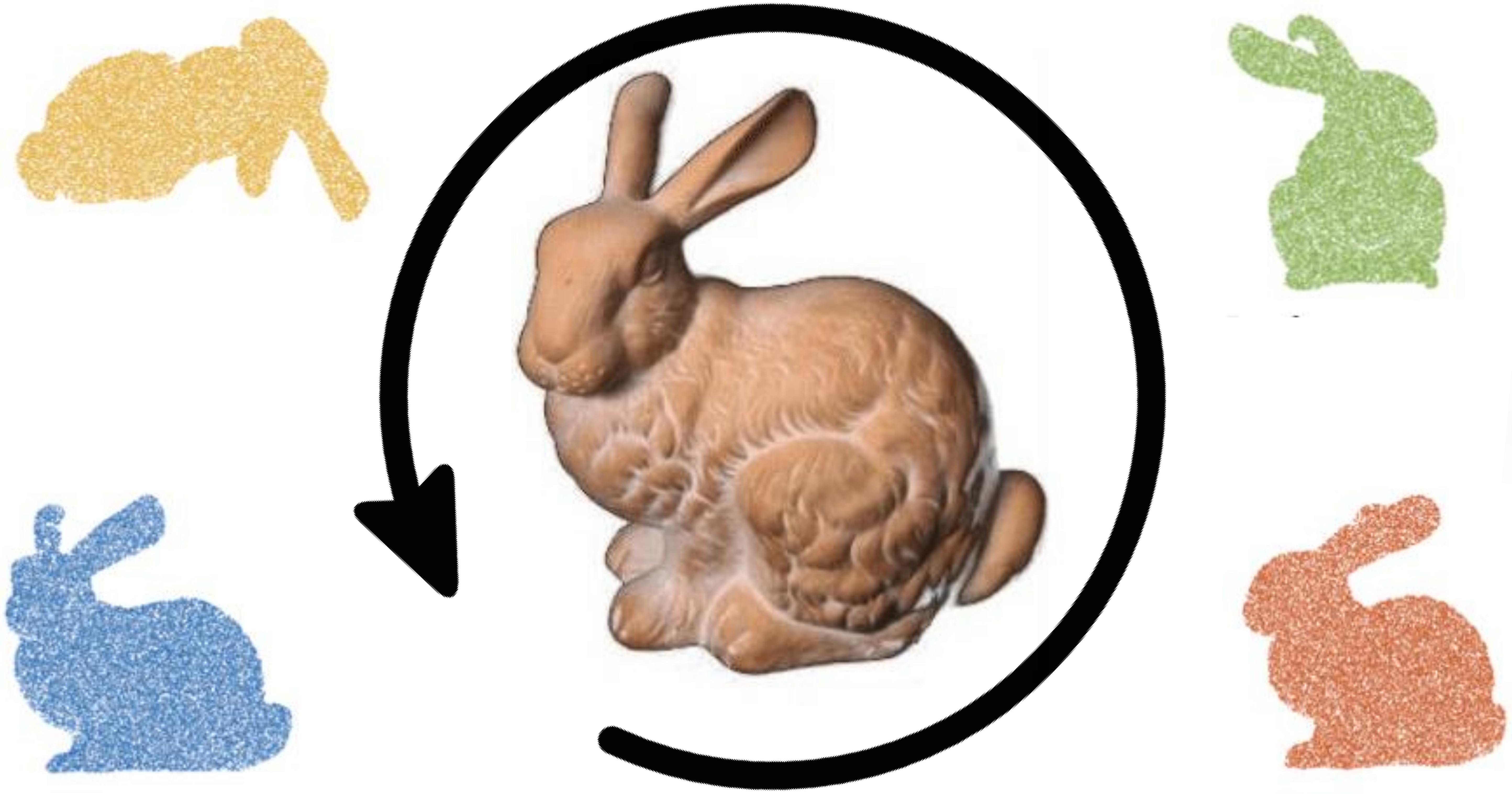
Politecnico di Milano

3D Registration

🤔 Often we have to manage multiple views!

TASK: bring multiple 3D point clouds into alignment.

CHALLENGE: each point cloud is a partial representation expressed in a different coordinate system!



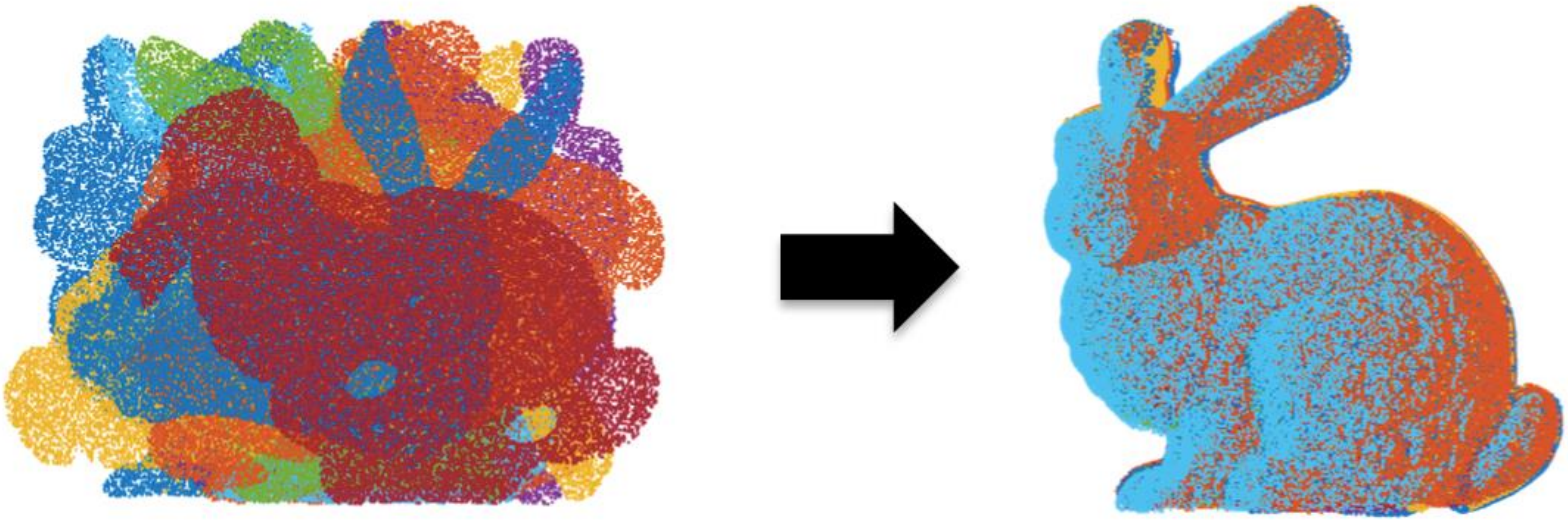
3D scanner

3D Registration

🤔 Often we have to manage multiple views!

TASK: bring multiple 3D point clouds into alignment.

CHALLENGE: each point cloud is a partial representation expressed in a different coordinate system!



3D Registration

🤔 Often we have to manage multiple views!

🧰 GRAPH REPRESENTATION:

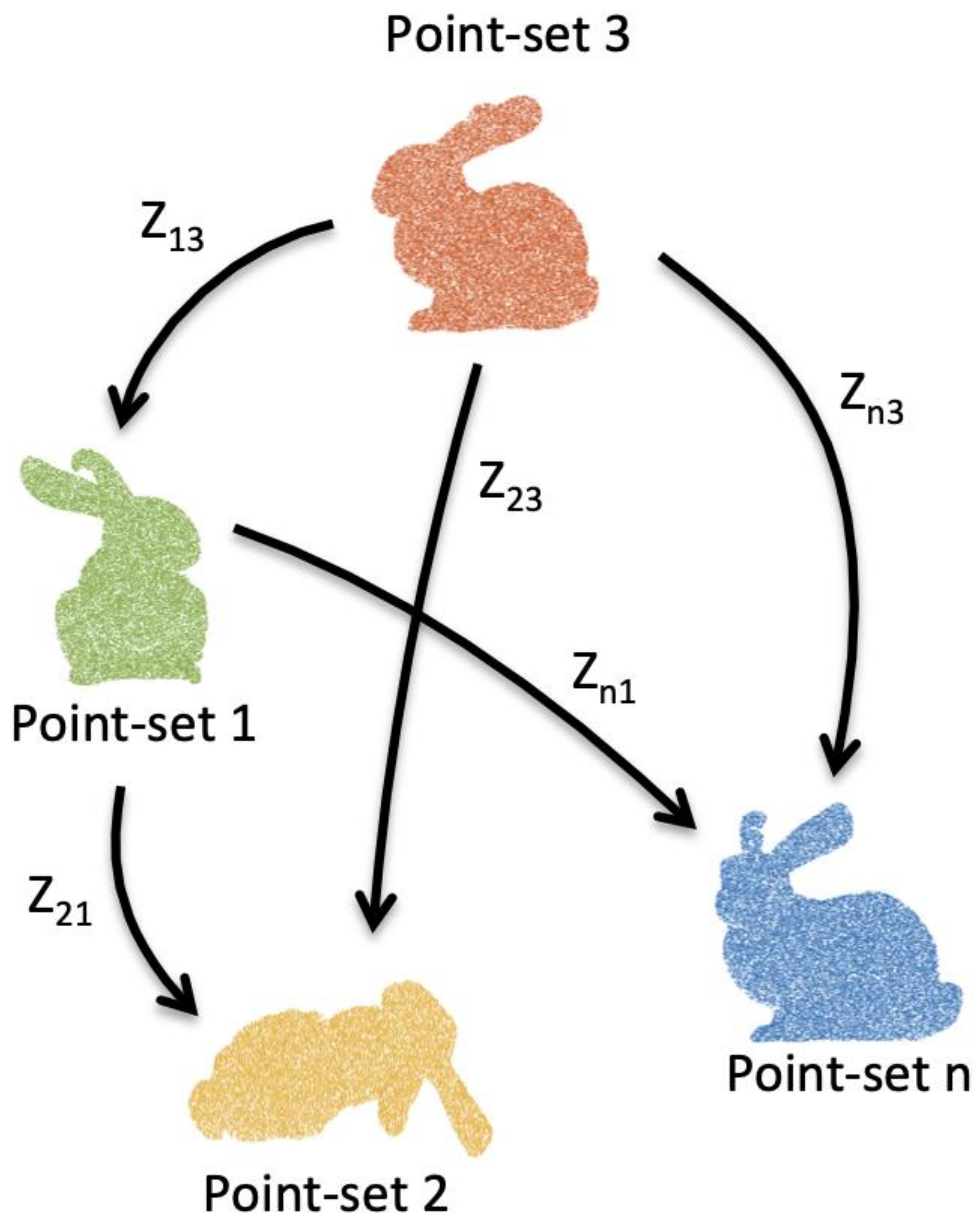
- nodes = point clouds
- edges = overlap between pairs of point clouds

💡 APPROACH:

1. Compute pairwise (local) transformations → *noisy*
2. Enforce global coherence (i.e. cycle consistency) → *error compensation*

✅ ADVANTAGES:

- Exploit two-view tools (e.g., closed form solutions) → smaller problems
- Very compact representation → transformations



3D Registration

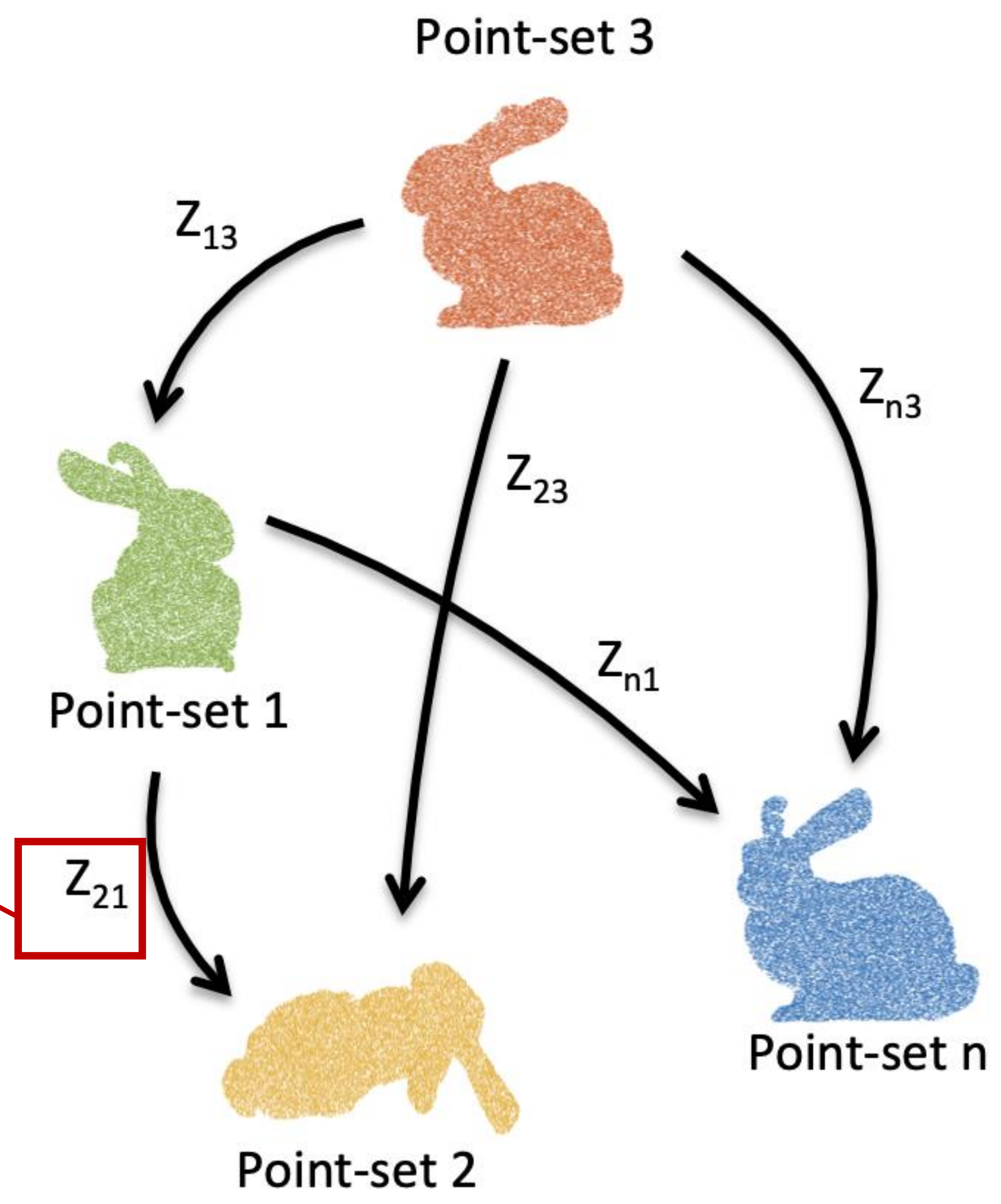
Often we have to manage multiple views!

SOLUTION:

- Each transformation is encoded in a 4x4 matrix
- All matrices are collected in a big block matrix
- It can be proved that the solution can be derived from the eigenvectors

$$SE(3) = \left\{ M = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}, \text{ such that } \underbrace{R \in SO(3)}_{\text{Rotation}},\ \underbrace{\mathbf{t} \in \mathbb{R}^3}_{\text{Translation}} \right\}$$

$$Z = \begin{bmatrix} I_d & Z_{12} & \dots & Z_{1n} \\ \boxed{Z_{21}} & I_d & \dots & Z_{2n} \\ \dots & & & \dots \\ Z_{n1} & Z_{n2} & \dots & I_d \end{bmatrix}$$



References

- Huang, Wang, Birdal, Sung, **Arrigoni**, Hu & Guibas: *MultiBodySync: Multi-Body Segmentation and Motion Estimation via 3D Scan Synchronization.* CVPR 2021
- **Arrigoni**, Rossi & Fusiello: *Global registration of 3D point sets via LRS decomposition*. ECCV 2016

A General Framework: Synchronization

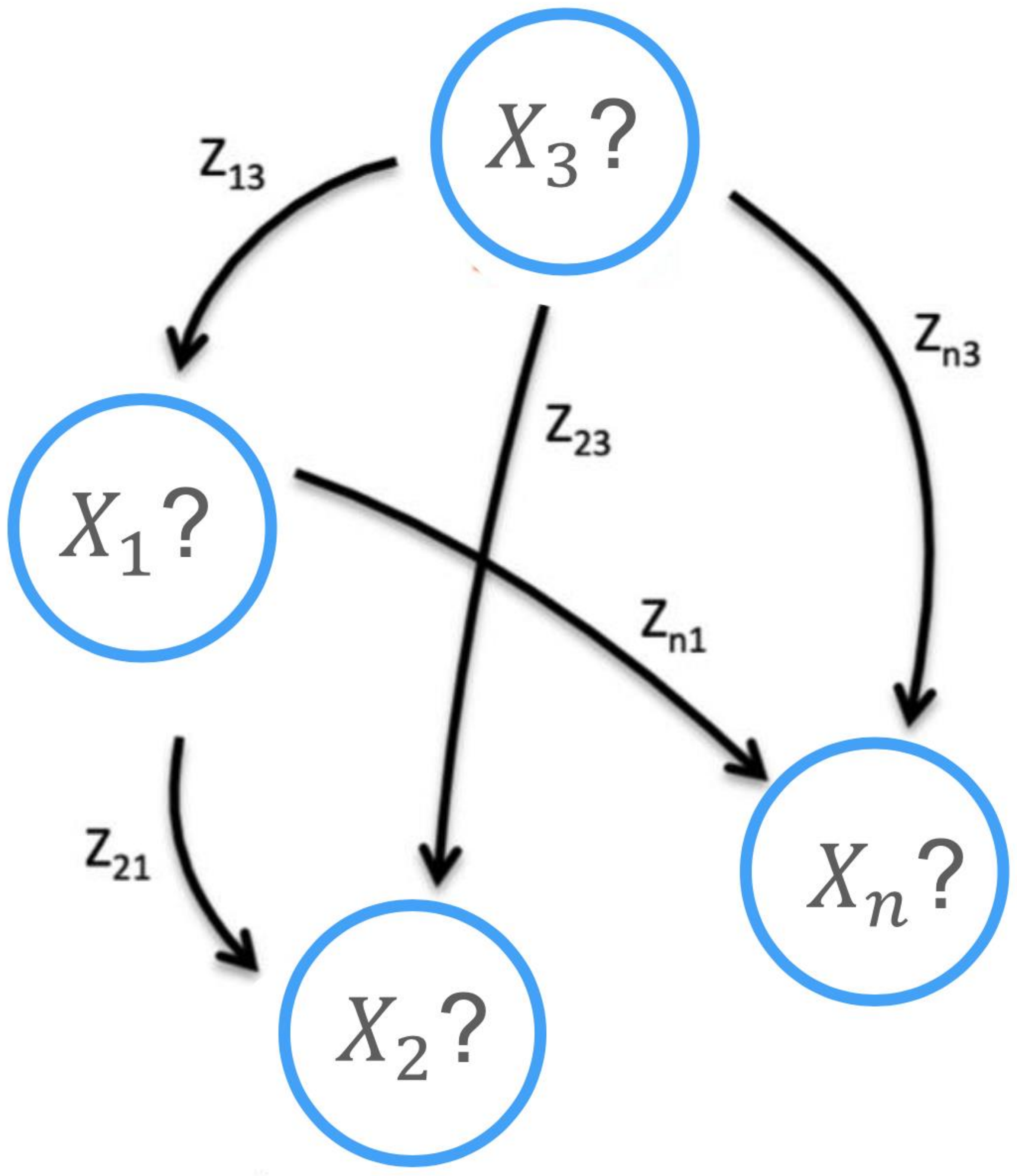
💡 3D registration is an example of a general framework: synchronization

TASK: compute absolute/global quantities starting from pairwise/relative measures.

NAME: generalization of clock synchronization

🧰 GRAPH REPRESENTATION:

- nodes = unknowns
- edges = measures



References

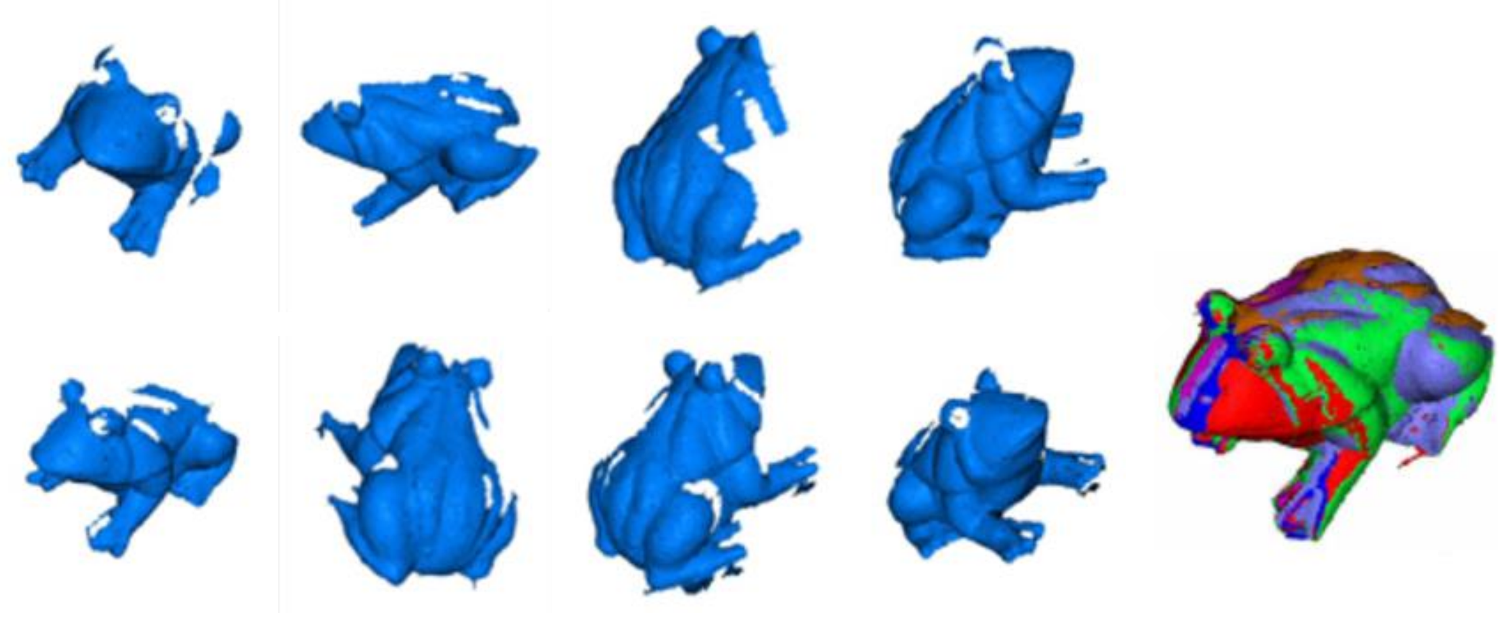
- **Arrigoni** & Fusiello: *Synchronization problems in Computer Vision with closed-form solutions* International Journal of Computer Vision 2020

A General Framework: Synchronization

The framework of synchronization can be used for several applications 😃

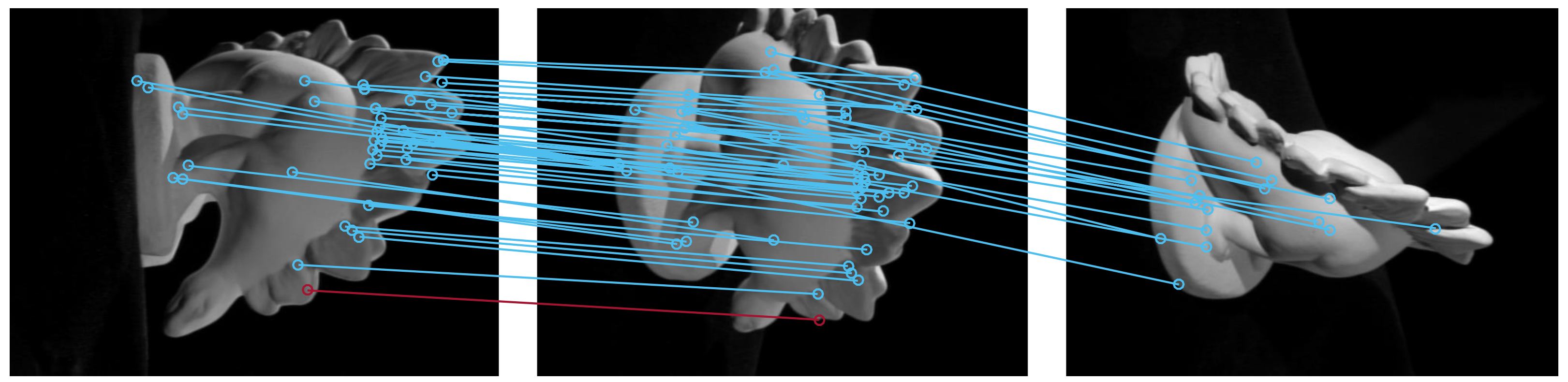
1 SOLUTION TO RULE THEM ALL: matrix representation + Linear Algebra

3D registration

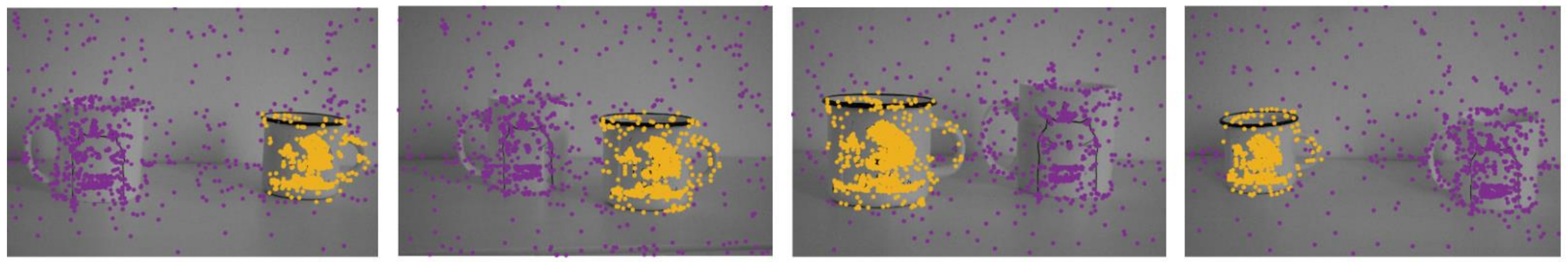


Structure from motion

Multi-view matching



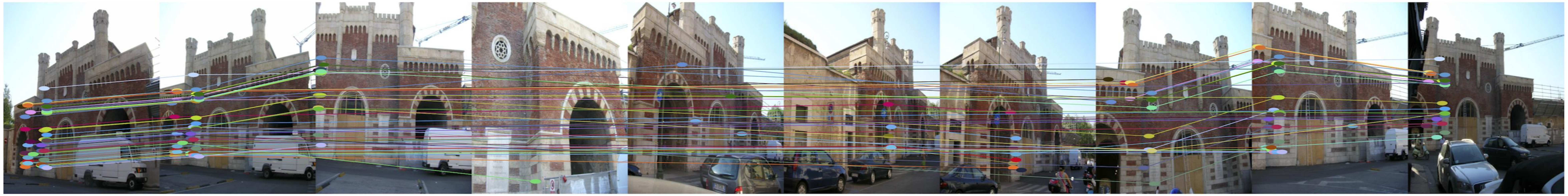
Motion segmentation



Multi-Image Matching

Multi-image matching is an example of synchronization.

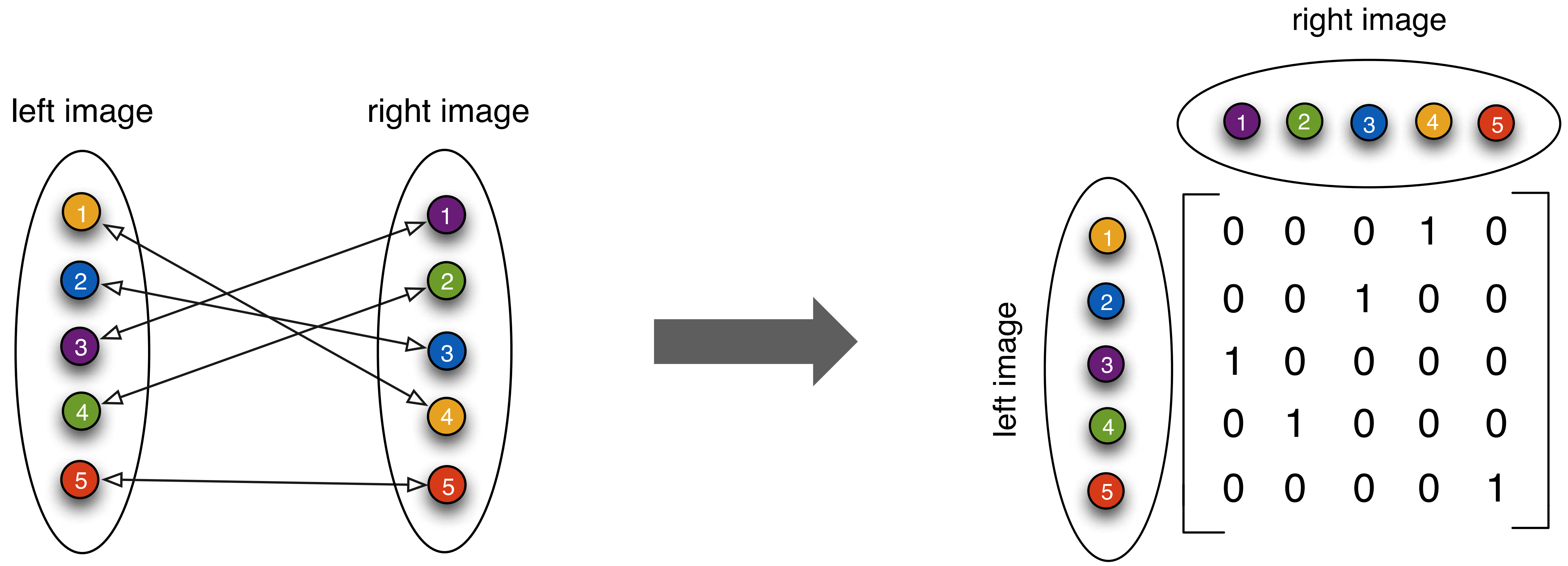
TASK: find correspondences between key-points in multiple images



MATRIX REPRESENTATION:

Matches can be encoded as permutation matrices:

$$Sym(d) = \{P \in \{0,1\}^{d \times d} \ \text{ such that } \ P\mathbf{1} = \mathbf{1}, \ \mathbf{1}P = \mathbf{1}\}$$



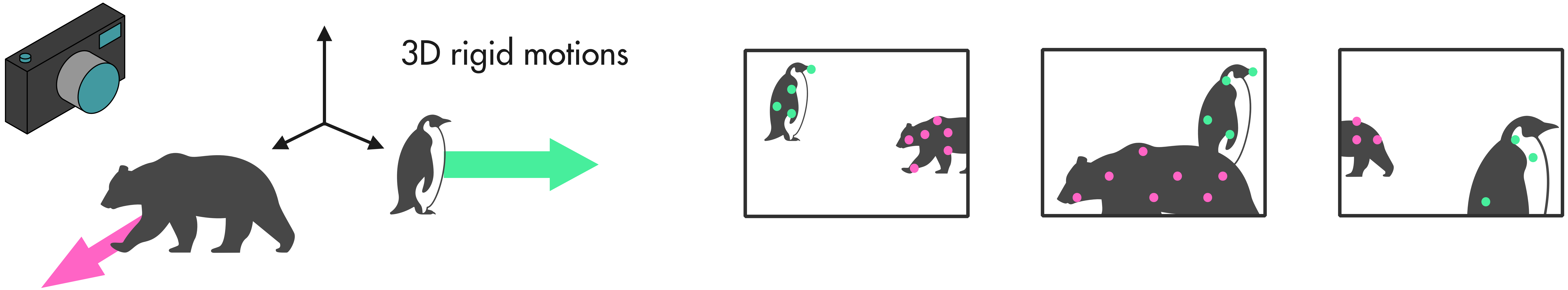
References

- Maset, **Arrigoni** & Fusiello: *Practical and efficient multi-view matching* ICCV 2017

Motion Segmentation

💡 Motion Segmentation is an example of synchronization.

TASK: classify points in multiple images (dynamic scene) based on the **moving** object they belong to.



🧰 **MATRIX REPRESENTATION:** Labels (i.e., clustering) can be represented as binary matrices.

References

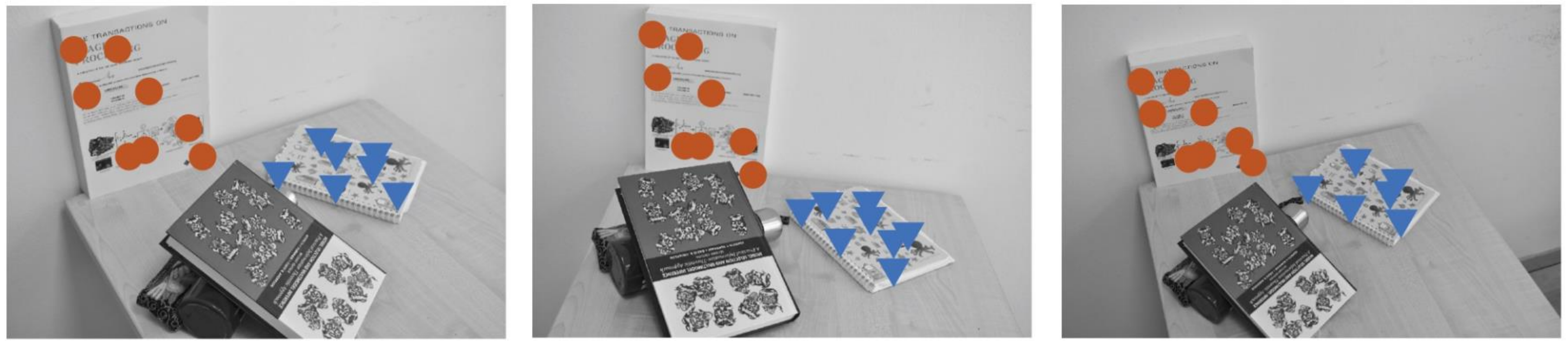
- **Arrigoni**, Ricci & Pajdla: *Multi-frame Motion Segmentation by Combining Two-Frame Results.* International Journal of Computer Vision 2022
- **Arrigoni**, **Magri** & Pajdla: *On the Usage of the Trifocal Tensor in Motion Segmentation*. ECCV 2020
- **Arrigoni** & Pajdla: *Robust motion segmentation from pairwise matches*. ICCV 2019

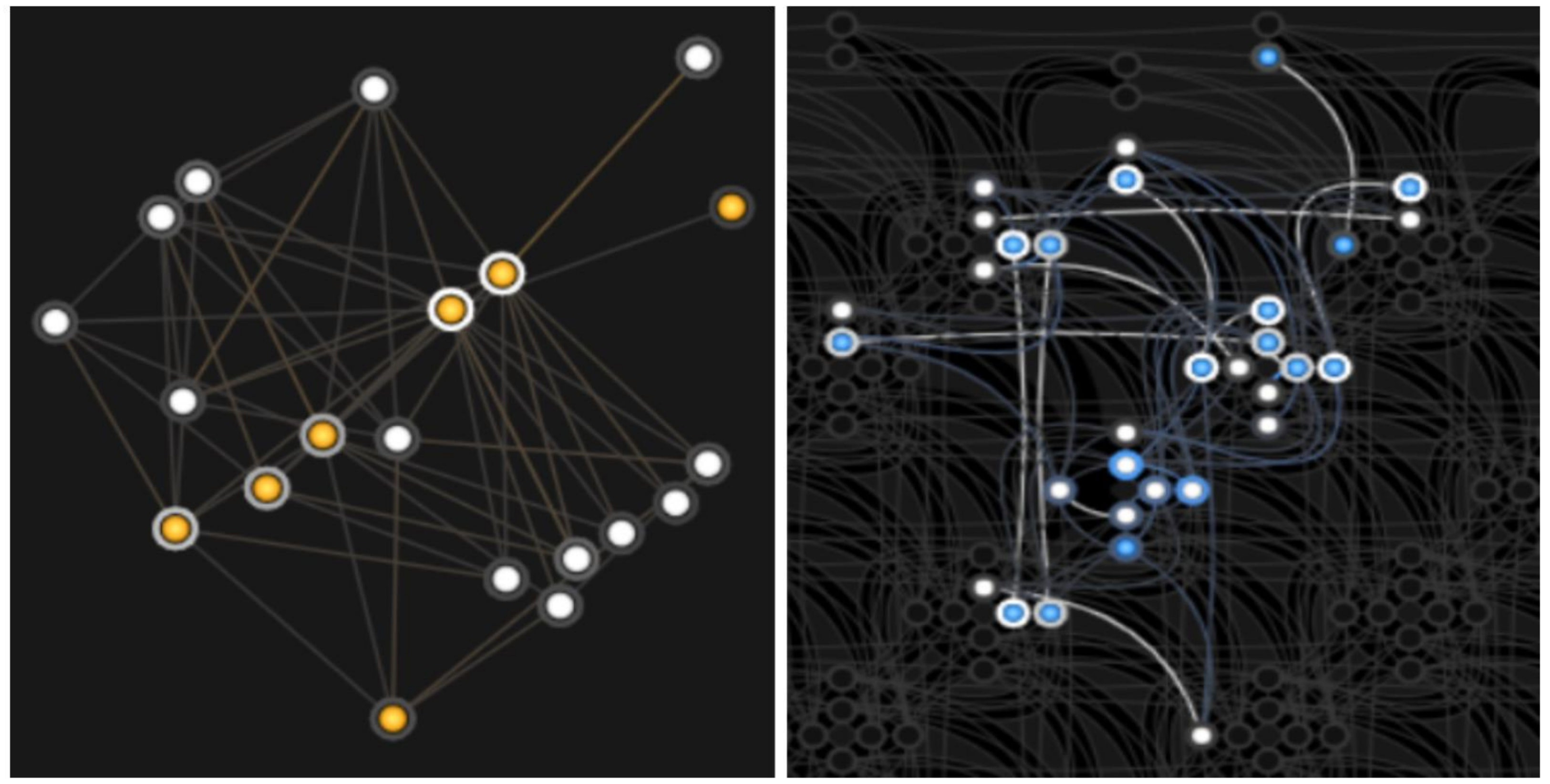
Motion Segmentation: Quantum Approach

💡 Motion Segmentation is an example of synchronization.

✅ ADVANTAGE:

It is possible to cast motion segmentation as a quadratic unconstrained binary optimization (QUBO)
→ adiabatic quantum computing





References

- **Arrigoni**, Menapace, Benkner, Ricci & Golyanik: *Quantum motion segmentation.* ECCV 2022
- Farina, **Magri**, Menapace, Ricci, Golyanik & **Arrigoni**: *Quantum multi-model fitting.* Under Review

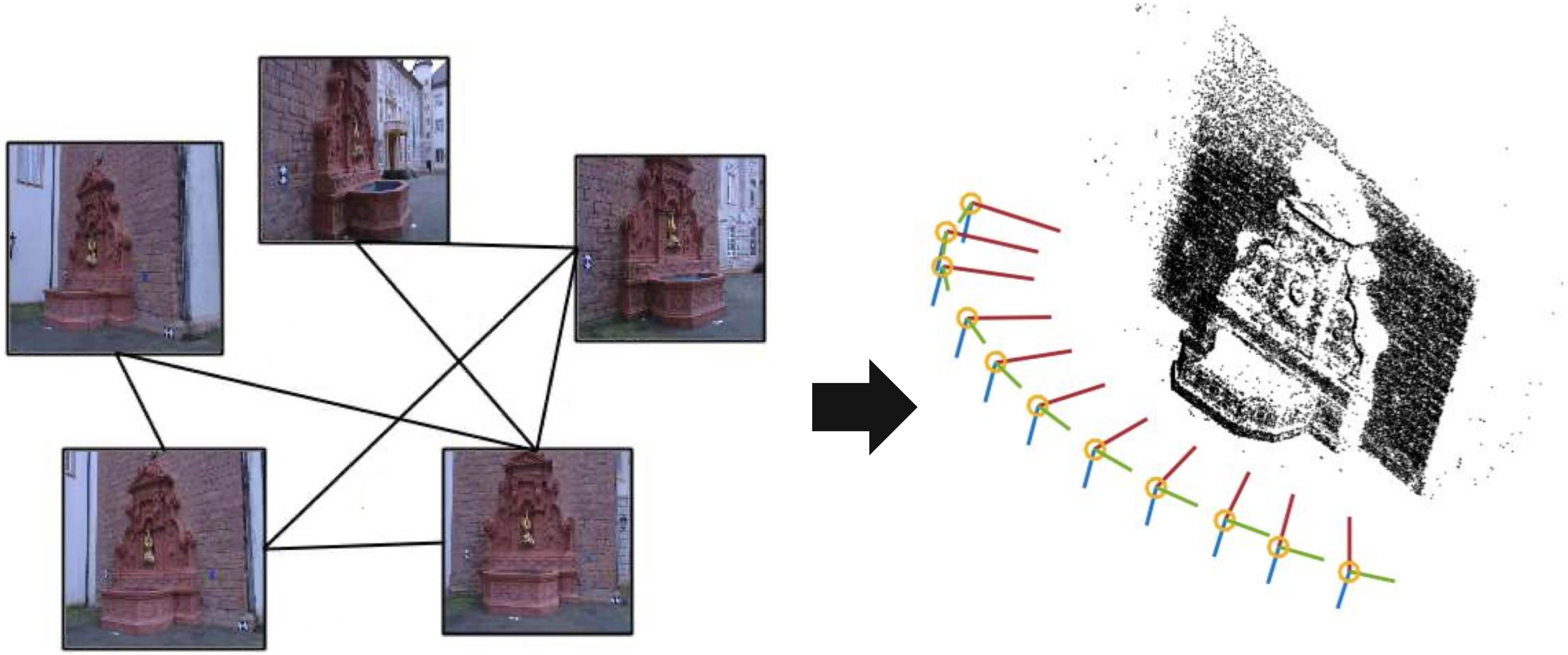
Structure from Motion

💡 Structure from motion is an example of synchronization.

TASK: compute camera motion (i.e., poses) and scene structure (i.e., 3D coordinates of points) starting from multiple images (static scene).

🧰 MATRIX REPRESENTATION:
Cameras can be represented as rotation matrices:

$$SO(3) = \{R \in \mathbb{R}^{3\times 3} \ \text{such that} \ R^{\mathsf{T}}R = I_3 = RR^{\mathsf{T}}, \ \det(R) = 1$$



References

- **Arrigoni** & Fusiello: *Bearing-based network localizability: a unifying view.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2019

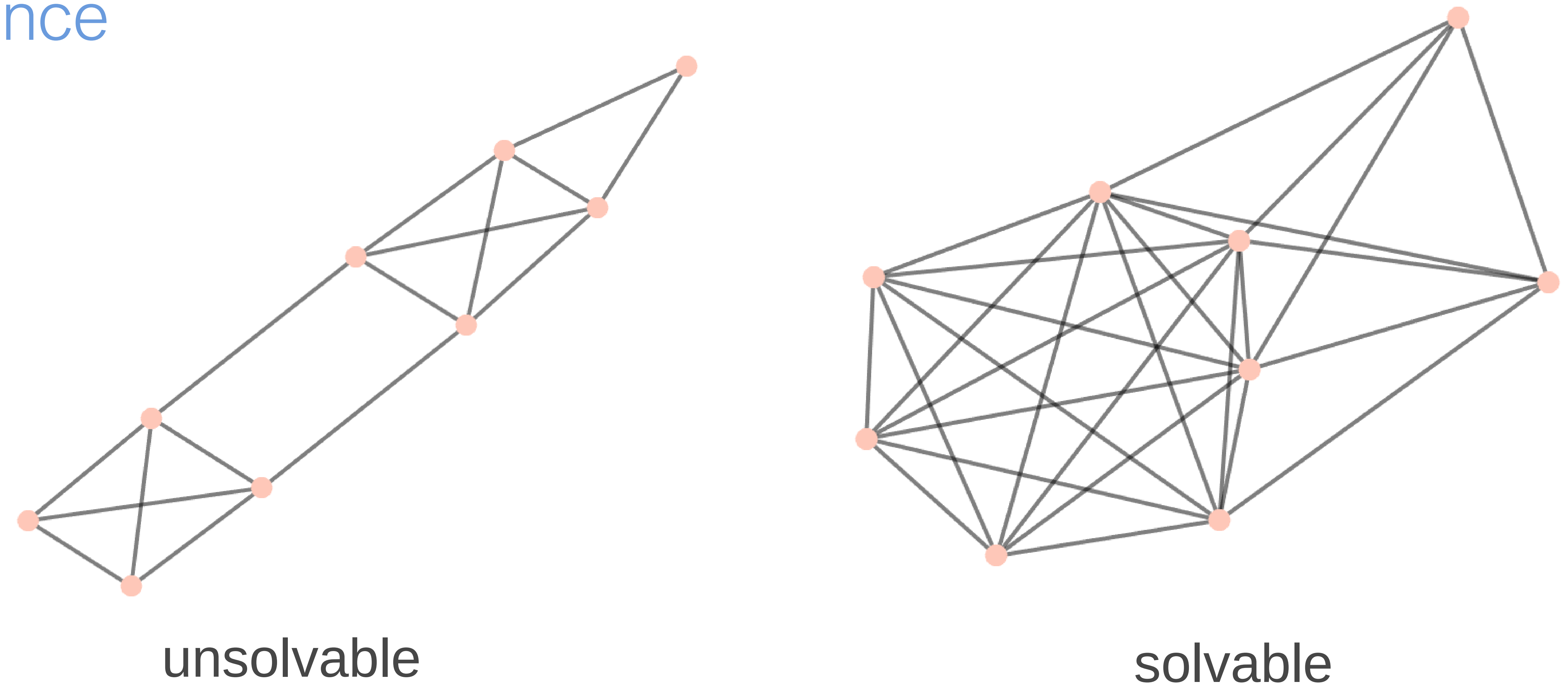
Structure from Motion: Theoretical Analysis

When is 3D reconstruction well-posed?

MOTIVATION: any reconstruction method will fail if the problem is ill-posed

TASK: classify solvable and unsolvable cases in advance

SOLUTION:
Algebraic Geometry (symbolic computation)



References

- **Arrigoni**, Fusiello, Rizzi, Ricci, & Pajdla: *Revisiting viewing graph solvability: an effective approach based on cycle consistency.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2022 – **special section on the best papers of ICCV**
- **Arrigoni**, Fusiello, Ricci & Pajdla: *Viewing graph solvability via cycle consistency*. ICCV 2021 – **Best Paper Honorable Mention**

Conclusion

Deep Learning **has revolutionized Computer Vision**

- Solved problems that were believed to be impossible
- Impressive results in a variety of visual recognition tasks
- Very powerful, comfortable, tool

Geometric Wisdom **plays a fundamental role in Computer Vision**

- No data annotation, low-amount of resources
- Clear understanding/interpretability, enable theoretical analysis
- Robust & accurate solutions
- No need to use Deep Learning when **explicit models** are available

In many cases monocultures are bad: **take the best of both worlds!**

- Machine Learning Research should be aware of Computer Vision to solve advanced tasks
- Computer Vision Research cannot ignore Deep Learning

Thank you for your attention!

Role: Assistant Professor (RTDb) – *since June 2022*

Research topics: Computer Vision

- 3D Vision
- Geometry
- Theory
- Quantum Computer Vision

Contact: federica.arrigoni@polimi.it

Selected References

- Arrigoni, Menapace, Benkner, Ricci & Golyanik: *Quantum motion segmentation.* ECCV 2022
- Arrigoni, Ricci & Pajdla: *Multi-frame Motion Segmentation by Combining Two-Frame Results.* International Journal of Computer Vision 2022
- Arrigoni, Fusiello, Rizzi, Ricci, & Pajdla: *Revisiting viewing graph solvability: an effective approach based on cycle consistency.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2022
- Arrigoni, Fusiello, Ricci & Pajdla: *Viewing graph solvability via cycle consistency*. ICCV 2021 – **Best Paper Honorable Mention**
- Huang, Wang, Birdal, Sung, Arrigoni, Hu & Guibas: *MultiBodySync: Multi-Body Segmentation and Motion Estimation via 3D Scan Synchronization.* CVPR 2021
- Arrigoni, Magri & Pajdla: *On the Usage of the Trifocal Tensor in Motion Segmentation*. ECCV 2020
- Arrigoni & Fusiello: *Synchronization problems in Computer Vision with closed-form solutions* International Journal of Computer Vision 2020
- Arrigoni & Pajdla: *Robust motion segmentation from pairwise matches*. ICCV 2019
- Arrigoni & Fusiello: *Bearing-based network localizability: a unifying view.* IEEE Transactions on Pattern Analysis and Machine Intelligence 2019

Luca Magri
Role: Assistant Professor (RTD-b)

Research topics:

- 3D vision
- Robust fitting
- Pattern Recognition
- Clustering

Contacts: luca.magri@polimi.it. Site: https://magrilu.github.io

References

- Magri and Fusiello. ***T-linkage: A continuous relaxation of j-linkage for multi- model fitting***. CVPR 14
- Magri and Fusiello. ***Multi-model fitting as a Set Coverage problem***. CVPR 16
- Magri and Fusiello. ***Fitting Multiple Heterogeneous Models by Multi-Class Cascaded T-Linkage***. CVPR 19
- Magri, Leveni, and Boracchi. ***MultiLink: Multi-class Structure Recovery via Agglomerative Clustering and Model Selection.*** CVPR 21
- Magri and Fusiello. ***Reconstruction of interior walls from point cloud data with min-hashed J-linkage***. 3DV18

Giacomo Boracchi

Role: Associate Professor

Research topics:

- Image Processing
- Deep Learning
- Unsupervised Learning,
- Change / Anomaly Detection

Contacts: giacomo.boracchi@polimi.it https://boracchi.faculty.polimi.it/

References

- M. Maggioni, G. Boracchi, A. Foi and K. Egiazarian *Video Denoising, Deblocking and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms* IEEE TRANSACTIONS ON IMAGE PROCESSING, 2012
- Foi A. and Boracchi G. *Foveated Nonlocal Self-Similarity*, INTERNATIONAL JOURNAL ON COMPUER VISION, Springer, 2016
- P. Morbidelli, D. Carrera, B. Rossi, P. Fragneto, G. Boracchi, *Augmented Grad-CAM: heat-maps super resolution through augmentation* ICASSP 2020
- L. Giulivi, M. Carman, G. Boracchi *Perception Visualization: Seeing Throughthe Eyes of a DNN* BMVC 2021
- Frittoli, L., Carrera, D., Rossi, B., Fragneto, P., & Boracchi, G. *Deep open-set recognition for silicon wafer production monitoring*. PATTERN RECOGNITION Elsevier, 2022 vol 124.