



ReInHerit



European  
Commission

# ReInHerit webinar

“Artificial Intelligence and Computer Vision  
for Cultural Heritage”

**18 November 2022 | 3:00-4:00 pm (CET)**

Marco Bertini and Paolo Mazzanti - MICC Florence IT

**MICC** - *Media Integration and Communication Center* - was established by the Italian Ministry for Education, University and Research at the University of Florence in 2001 as a **Center of Excellence** in the area of new media.

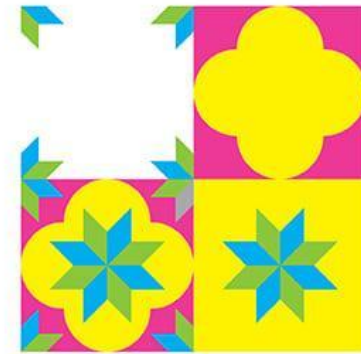
MICC works as an **interdisciplinary center** for advanced research in the fields of **computer vision, multimedia technologies** applied to smart environments, natural interaction, **Internet Based Applications** and **collective intelligence**.

[www.micc.unifi.it](http://www.micc.unifi.it)

**NEMECH** - *New Media for Cultural Heritage* - is the **Competence Centre** on Cultural Heritage established by the Region of Tuscany and activated at MICC – University of Florence. NEMECH connects research centers and institutions: it promotes the transfer of know-how of research from university labs to the cultural spaces NEMECH develops **digital tools and new media for cultural heritage**.

- *Research labs*
- *Interdisciplinary programs and trainings*
- *Demonstration space - results and projects*

[www.nemech.unifi.it](http://www.nemech.unifi.it)



**ReInHerit**  
Redefining the Future  
of Cultural Heritage

**#make it  
your own**

*This project has received funding from the European Union's Horizon 2020  
research and innovation programme under grant agreement  
No 101004545*





### *ReInHerit's Mission*

to disrupt the current status quo of communication, collaboration and innovation exchange between museums and cultural heritage sites



#ReInHerit #MakeItYourOwn

# Innovation Disruption Sustainability

---

**WP1** Project Management

---

**WP2** CH Needs Analysis & Guidelines

---

**WP3** ReInHerit Toolkit

---

**WP4** Digital Hub

---

**WP5** ReInHerit Best Practices on CH Management

---

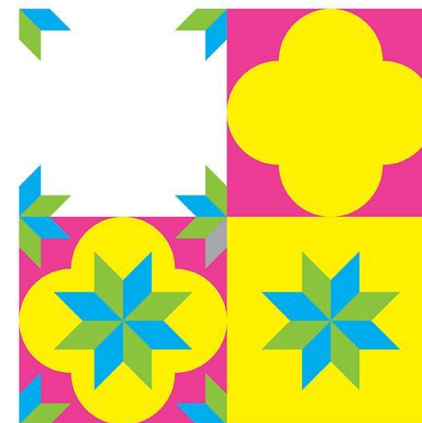
**WP6** Pilot Phase tested through Digital & Travelling  
Exhibition

---

**WP7** Dissemination, Exploitation & Communication

---

**WP8** Ethics requirements



**ReInHerit**  
Redefining the Future  
of Cultural Heritage

## Primary research (WP2)

Target



**visitors**

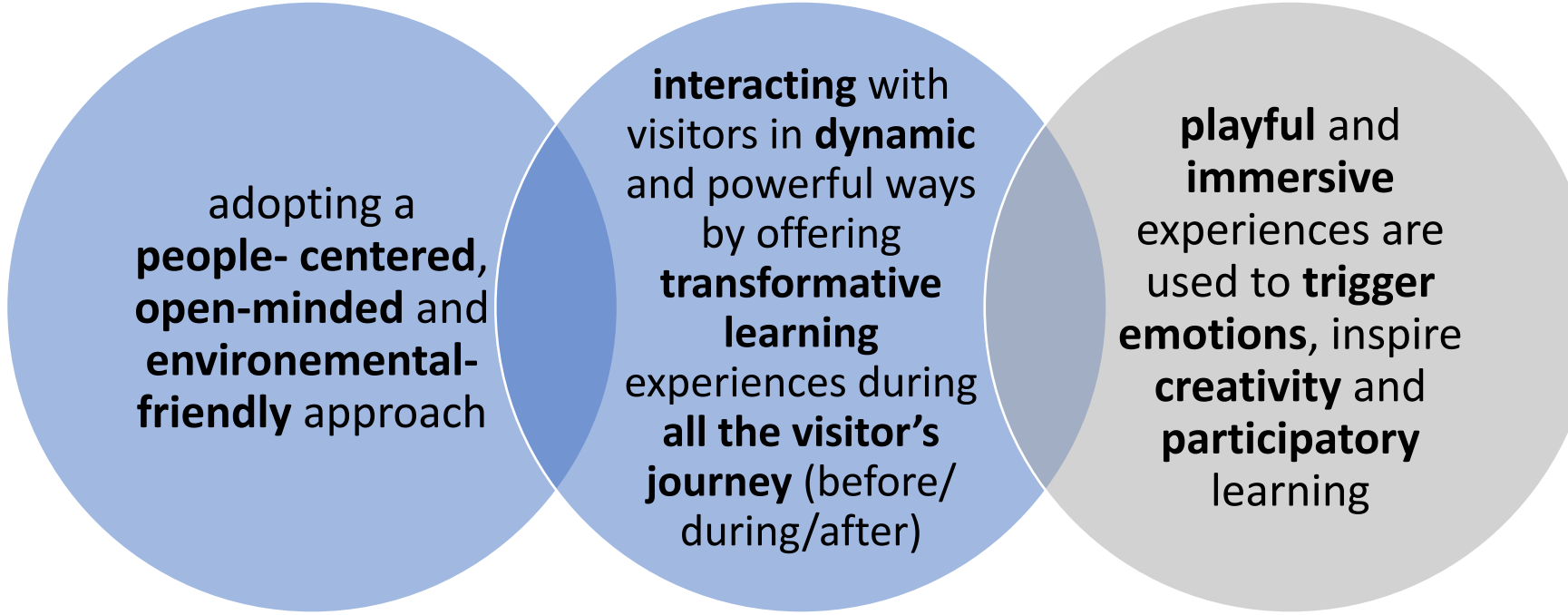
**Young museum visitors (18-29)** are more likely to use digital tools in a museum environment and, thus, be more positively inclined towards using the digital tools the ReInherit project will be offering and the digital hub.

**professionals**

**Small and mid-sized museums** have less capacity to use digital services for cultural heritage management

## Secondary research (WP2)

### Approach



adopting a **people-centered, open-minded and environmental-friendly** approach

**interacting** with visitors in **dynamic** and powerful ways by offering **transformative learning** experiences during **all the visitor's journey** (before/ during/after)

**playful** and **immersive** experiences are used to **trigger emotions**, inspire **creativity** and **participatory** learning



Edited by **Paolo Mazzanti**  
in collaboration with **Margherita Sani**

### Table of Contents:

■ Introduction - Margherita Sani	3
■ Emotions inside/out museums - Paolo Mazzanti	6
■ Emotions in digital - Alberto Del Bimbo	24
■ The place of emotions in museums: the scenographer's point of view - Lorenzo Greppi	31
■ Emotions, stories and storytelling for audience engagement strategies - Antonia Silvaggi	38
■ Use-case: transforming a museum network into an emotional approach - Elisa Bruttini	45
■ Connected Audience 2019 – conference overview - Judith Koke	52
■ The role of emotions in museum-going - John H. Falk	55
■ Planning for emotions in museums - Tom Owen	61
■ Fostering empathy through the visual arts - Karleen V. Gardner	66
■ Emotions in the history museum - Sheila Watson	74
■ BIOGRAPHIES:	78

**EMOTIONS**  
and **LEARNING**  
in **MUSEUMS**



Network of European  
Museum Organisations

A **NEMO** Report  
by **LEM** – The Learning Museum Working Group

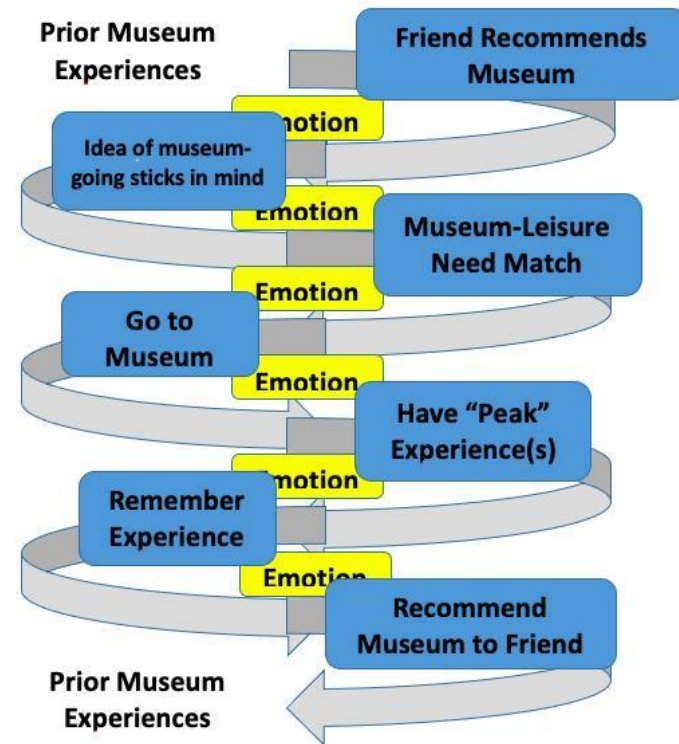
### The LEM Working Group

The working groups of NEMO (The Network of European Museum Organisations) offer their members a European perspective on different topics that are important to museums. The Learning Museum Working Group (LEM WG) explores topics relating to the fields of museum education, audience development, intercultural dialogue and lifelong learning. It started out as a continuation of *LEM - The Learning Museum*, a network project funded by the EU and carried out between 2010 and 2013, in which NEMO was a partner. Carrying on the legacy of LEM, the working group today supports the exchange of information and learning among museum professionals in Europe through study visits to various museums in Europe for its members, as well as through studies and reports produced by the group.

Cover photo by P. Mazzanti: Tate Modern, London, UK

«The museum experience turns out to be **not linear but cyclical**, with **emotions** playing a critical role at EVERY stage»

*The role of emotions in museum-going* - John H. Falk

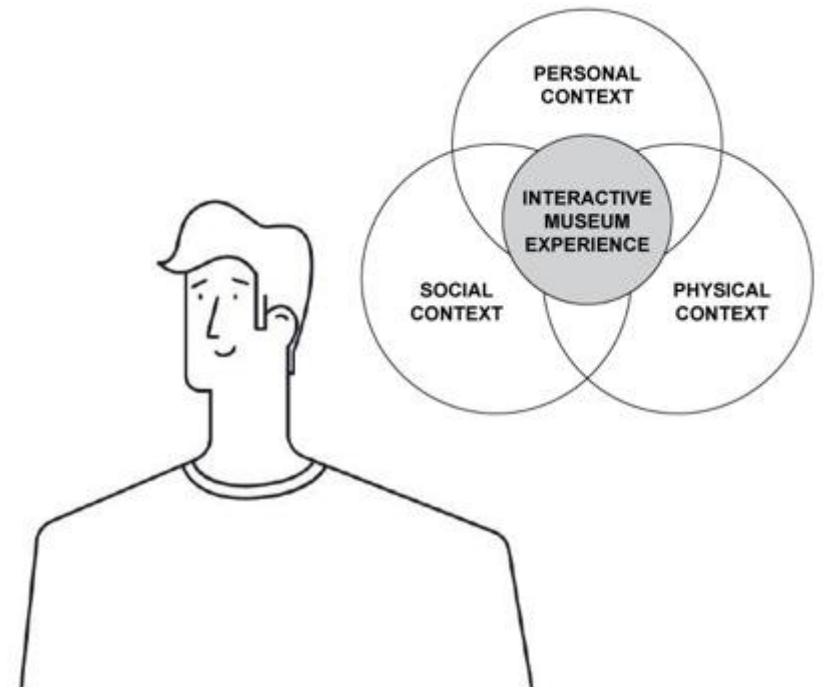




«Digital technologies have offered the opportunity to **extend** museum visits into sessions of **experiential education** by expanding the modes of visitor engagements. A well-founded question arises on whether, in doing this, digital technologies might also contribute to establishing a **stronger relationship between the artwork and the visitor**: whether they are able to **create emotions**.» [Emotions at Digital- A. Del Bimbo 2021]

- ❑ **PHYGITAL MUSEUMS**  
(MIXED/EXTENDED EXPERIENCE)
- ❑ **EMPATHETIC MUSEUMS**
- ❑ **RELATIONAL MUSEUMS**
- ❑ **MUSEUMS «OUT OF THE BOX»**

# Inter\_actions



**innovative trend**

AI/CV based tools  
in museums and CH contexts

«What about **your experience?**»

---

# Artificial Intelligence & Computer Vision Tools in museums

Museums are using AI technology mainly to engage audiences and personalize visitor experiences

User/Human Centered Approach

New points of view & perspectives about collections

Interactive & Memorable Experiences

Phygital Experiences

Emotional Engagement for Learning Motivation

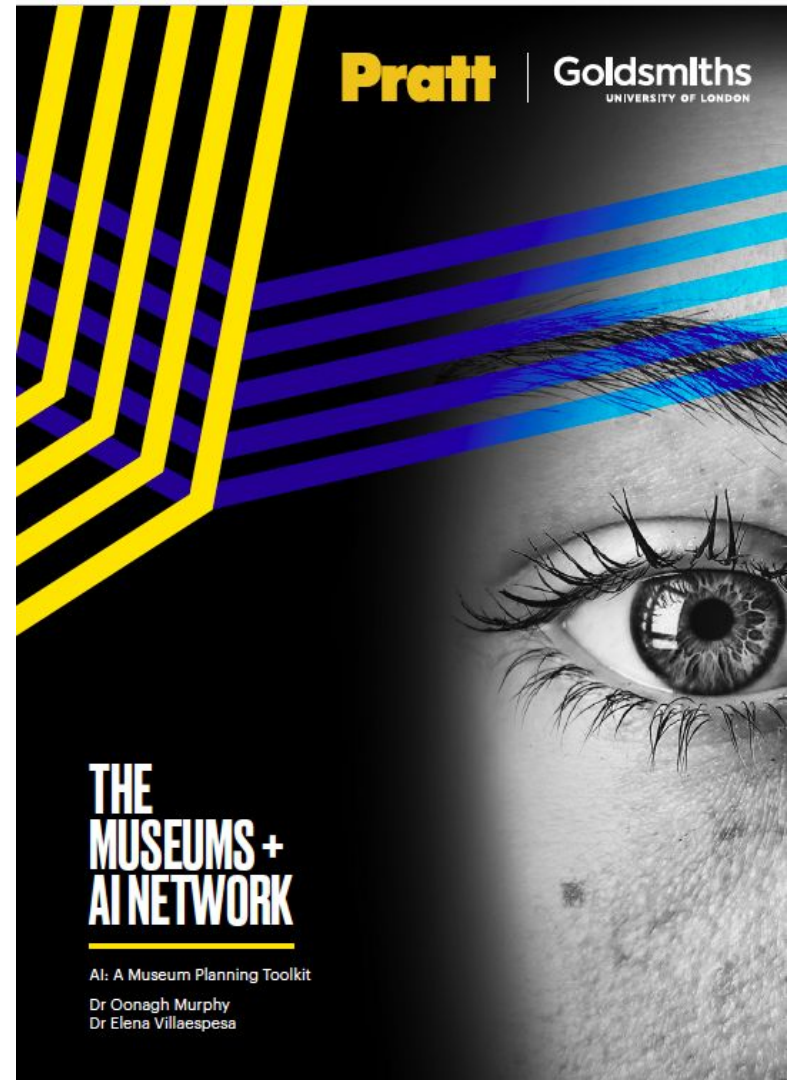
Storytelling & Social Media Sharing

Users Generated Narratives & Contents

# museum innovation barometer

## 2021

by MUSEUM BOOSTER



<https://themuseumsai.network/toolkit/>

«Computer vision can **help visitors to engage with collections** in **new ways**, and help curators to develop **new insights** into objects that they may not have had the time or resources to research in an analogue manner. It is for these reasons that computer vision is fast becoming a potential instrument **to enrich museum collections data** in a diverse range of ways, which ultimately can have an important **impact on the user experience.**»

[E. Villaespesa & O. Murphy 2021]

**Pratt**

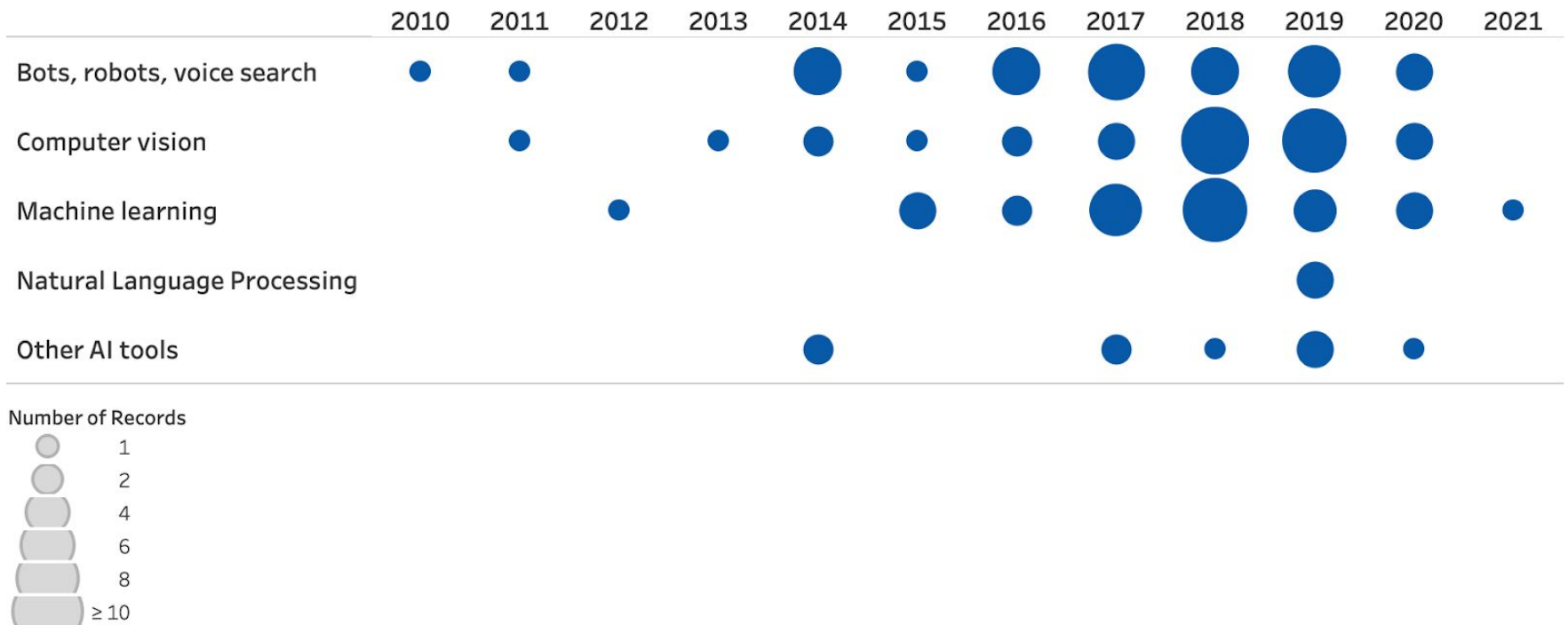
**Goldsmith**  
UNIVERSITY OF LONDON

**THE  
MUSEUMS +  
AI NETWORK**

... A Museum Planning Toolkit

... onagh Murphy  
... Villaespesa

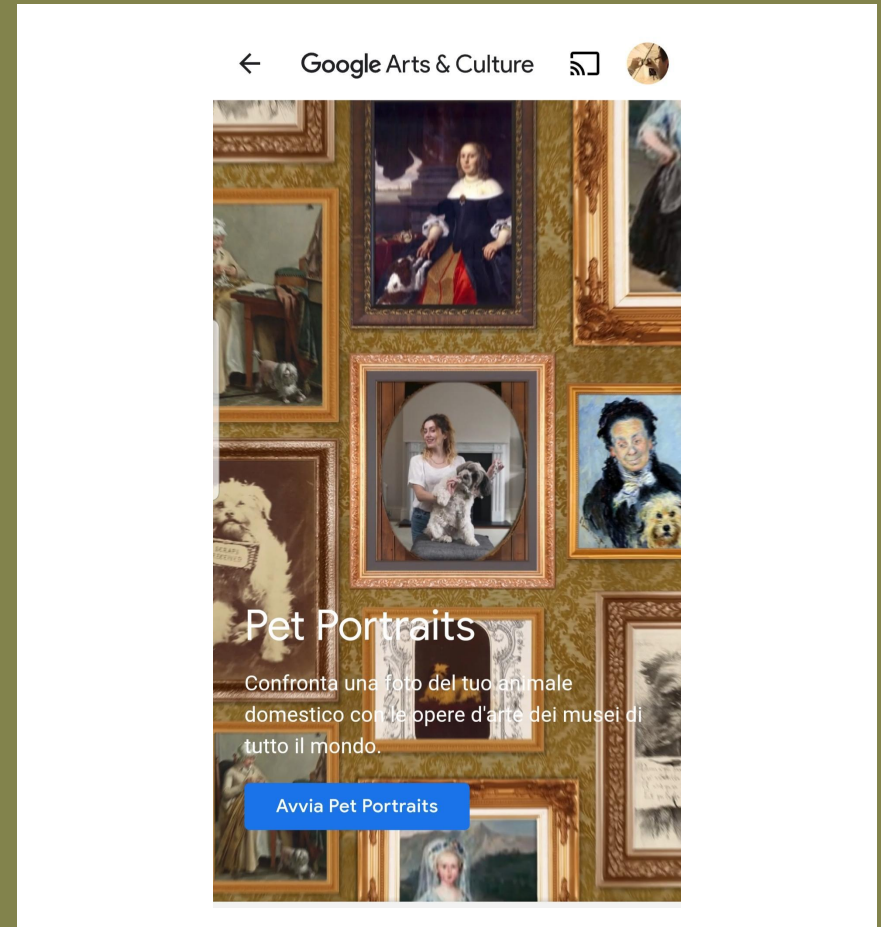
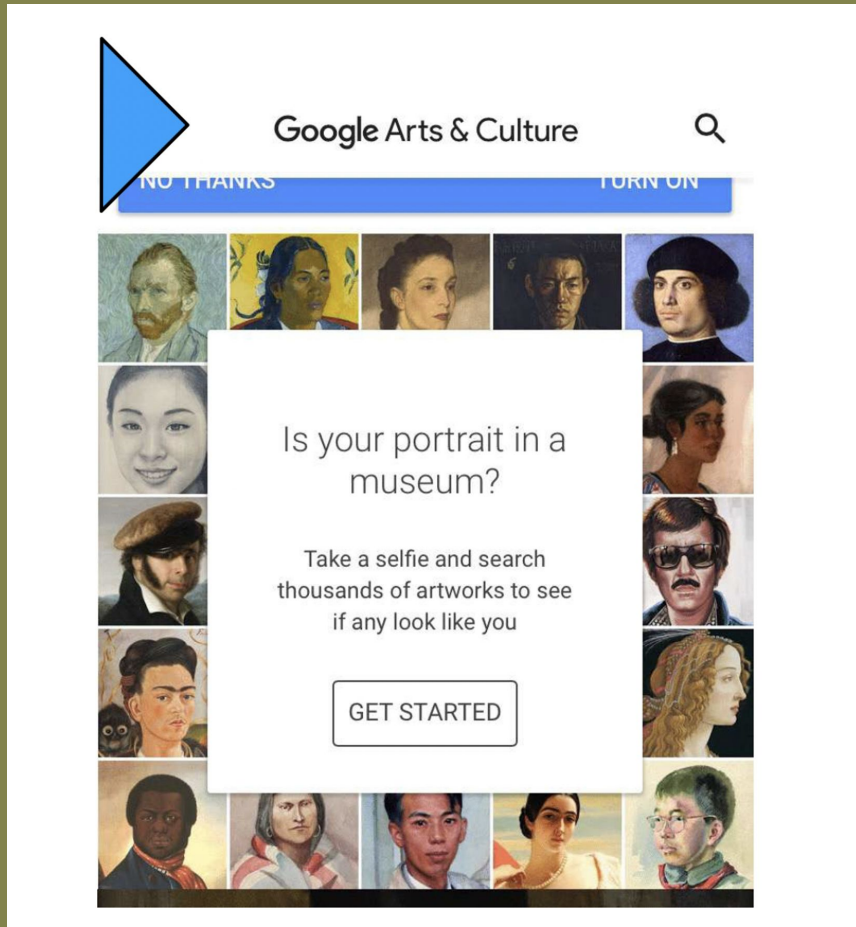
# Number of AI initiatives in museums (2010-2021)



Source: <https://www.artsmetrics.com/en/list-of-artificial-intelligence-ai-initiatives-in-museums/>

# AI – image recognition to share experiences

## Art Selfie – Google Arts & Culture



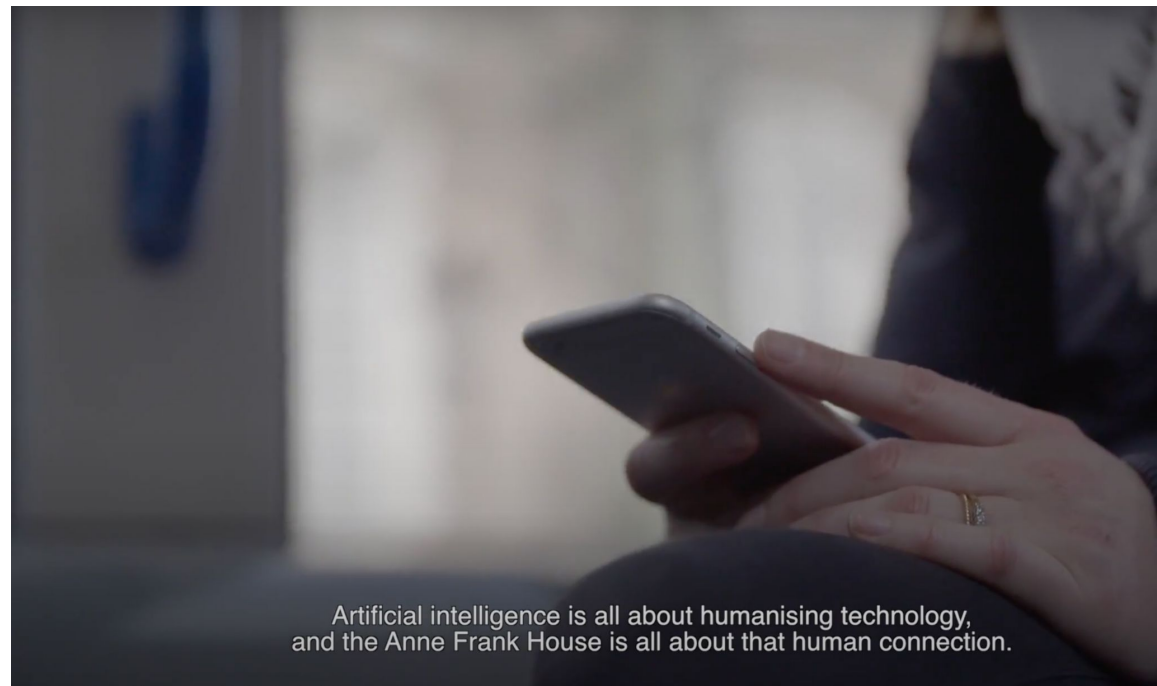
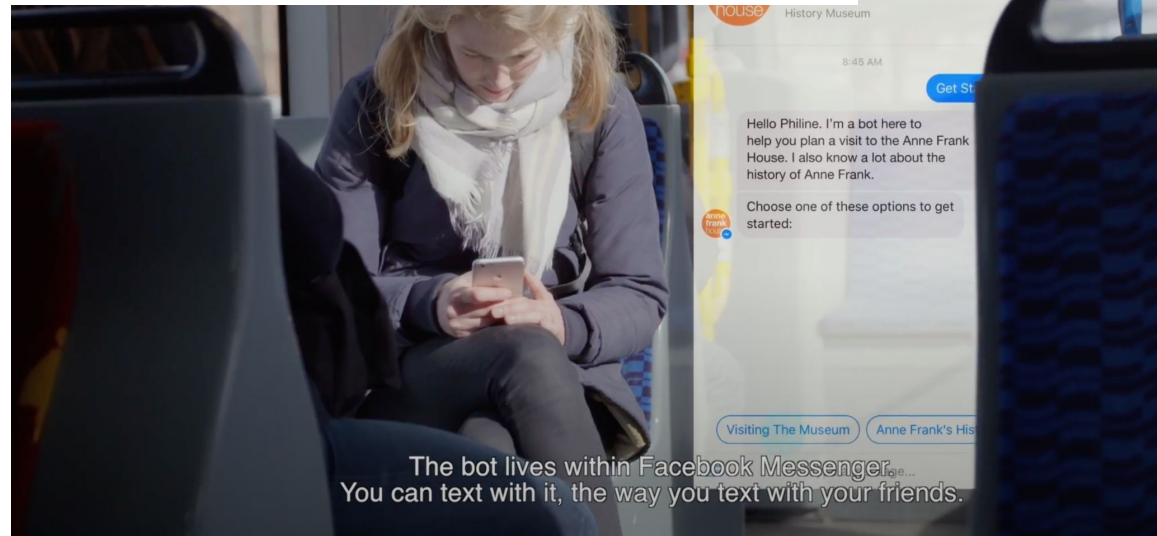
# AI – engaging museum visitors with Chatbots

Anne Frank Museum  
Amsterdam NL

## «Messenger Bot»

Anne Frank House museum uses **Facebook Messenger chatbot** to answer common visitor questions about visiting, with a **AI computer program** that is designed to mimic a human interaction in a text based conversation.

<https://youtu.be/YPH4vUWcN2U>



# AI – image recognition to search similarities

Google's X Degrees of Separation – Google Arts & Culture

<https://artsexperiments.withgoogle.com/xdegrees/>

## X Degrees of Separation

What is the connection between a 4000 year old clay figure and Van Gogh's Starry Night?

How do you get from Bruegel's Tower of Babel to the street art of Rio de Janeiro?

What links an African mask to a Japanese wood cut?

What visual similarities can a computer vision algorithm find to connect a sculpture with a drawing?



Daniel Chester French, 1922  
Benediction  
Amon Carter Museum of American Art



Frederic Remington, 1906  
The Outlaw  
Amon Carter Museum of American Art



Mme. Beer  
Giri's Hat  
The Frick Pittsburgh



Unknown  
The Metropolitan Museum of Art



Unknown  
The Metropolitan Museum of Art



Unknown  
The Metropolitan Museum of Art



George Seideneck, ca. 1937  
Copper Kettle  
National Gallery of Art, Washington DC



Sydney Roberts, ca. 1941  
Mug  
National Gallery of Art, Washington DC



Charles Goodwin, 1941  
Shaker Sugar Jar  
National Gallery of Art, Washington DC

[Click here to find your own paths through art space](#)

# AI – image recognition to search similarities

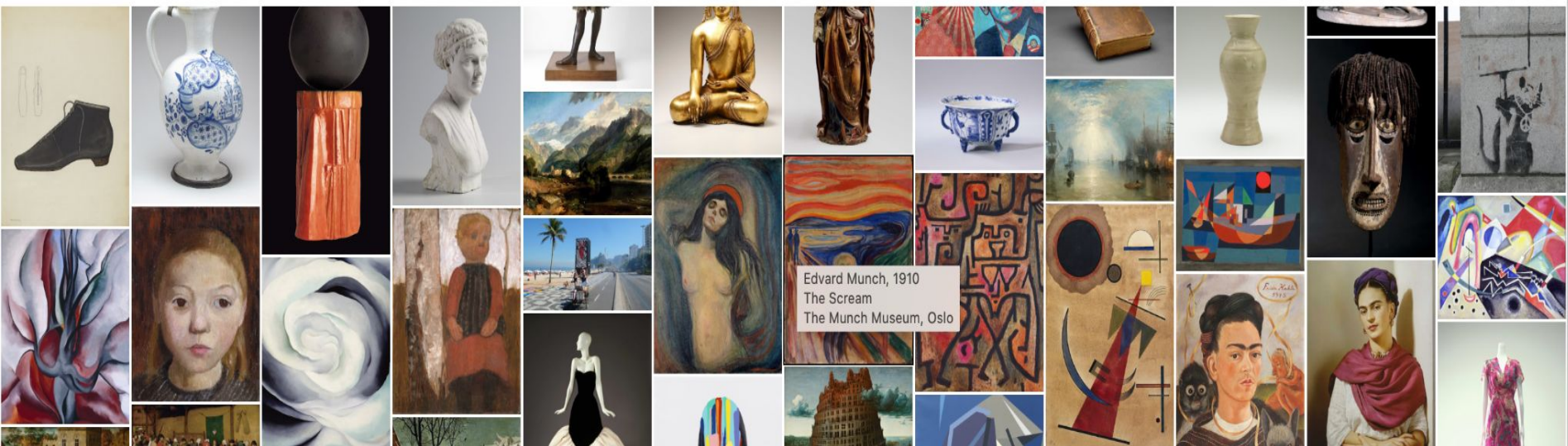
Google's X Degrees of Separation – Google Arts & Culture

<https://artsexperiments.withgoogle.com/xdegrees/>



A random selection of artifacts - select two images and discover their visual connection:

Q Try searching for "Van Gogh" or "Moma"



# AI - to make the collection more accessible

Metropolitan Museums of Art US

## «Providing access to the Collection»

The Museum is developing new ways to document, and interpret the museum's collection in a way that will allow it to **become searchable and browsable online.**

MET worked on the **generation of tags** manually and testing with computer vision. The goals of tagging the museum collection are **to increase user engagement, improve search** and discovery of the collection, make the collection accessible to the widest possible audience and explore **using tags as training data for AI models.**

Implementing tags both into the collection management system and on the website user interface.



Visually Similar Art



An artificial intelligence feature in Art Explorer produces visually similar works of art from The Met Open Access collection.  
Above: Vincent van Gogh. *Olive Trees*, 1889. Below, left to right: Unknown Artist. *Demons Fighting over an Animal Limb*, late 17th century, India. Robert Frederick Blum. *View from the Artist's Window, Grove Street*, ca. 1900. Vincent van Gogh. *Cypresses*, 1889

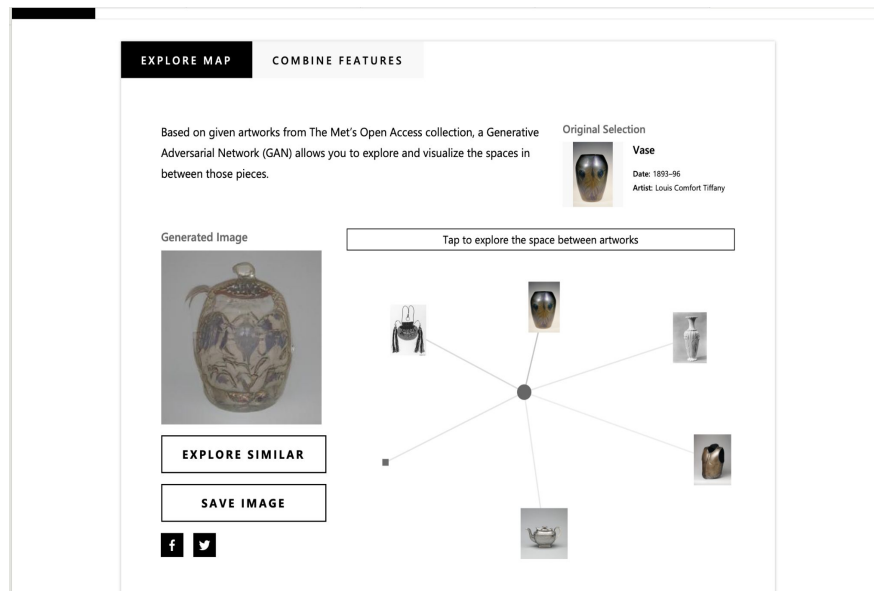
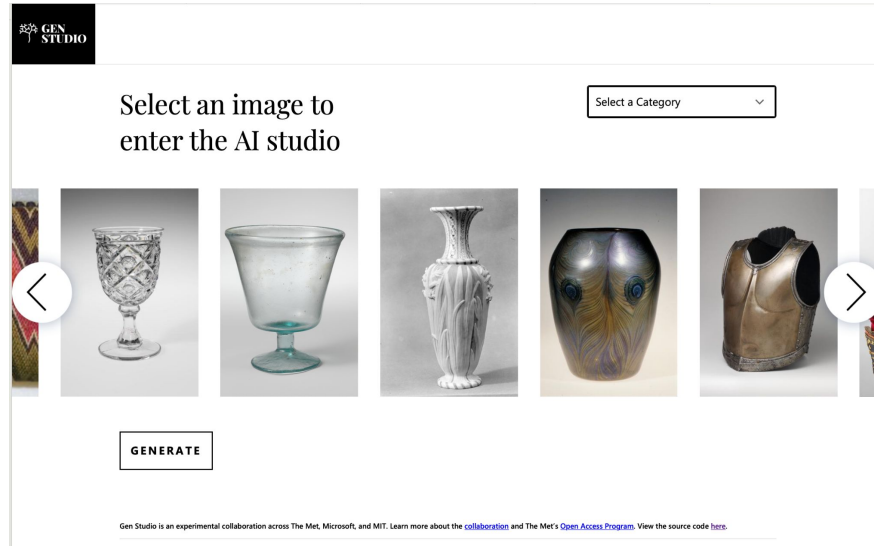
# AI – image recognition to explore collections

a collaboration between MIT, Microsoft and The Met

<https://gen.studio>

## “Gen Studio”

The project consists of images created using AI on artworks from **The Met’s Open Access collection**. The generated images enable users to **explore and visualise possible artworks between selected pieces** from the collections. For example, user can see what an object between a vase and a cup could look like.



# AI – a puzzle of relationships

Tate Modern UK

## «RECOGNITION»

Re-cognition, winner of IK Prize 2016 for digital innovation, is an **artificial intelligence program** that compares up-to-the-minute photojournalism with **British art** from the Tate **collection** - virtual gallery at Tate Britain.

<http://recognition.tate.org.uk/archive/>





☰ Context recognition is a process which analyses the titles, dates, tags, and descriptions associated with each image.

LEFT 13/09/2016  
A boy immerses an idol of Hindu god Ganesh, the deity of prosperity, into the Sabarmati river during the 10-day-long Ganesh Chaturthi festival, in Ahmedabad  
© AMIT DAVE / REUTERS

WATER, BOAT, OUTDOOR, MAN, RIDING, POND, BAYOU, RIVER, CREEK, RAFT

RIGHT 1893-4  
August Blue  
© HENRY SCOTT TUKE / TATE

WATER, SPORT, OUTDOOR, MAN, RIDING, RIVER



△ Composition recognition is a process for identifying prominent shapes and structures, visual layout, and colours.

LEFT 18/10/2016  
File photo of farmers collecting corn for a cargo at a farm in Gaocheng  
© KIM KYUNG HOON / REUTERS

COMPOSITION 80%

RIGHT 1980  
Portrait of V.I. Lenin with cap, in the style of Jackson Pollock III  
© ART & LANGUAGE (MICHAEL BALDWIN, MEL RAMSDEN) / TATE

COMPOSITION 80%

# AI for visitor engagement

Cleveland Museum of Art - USA

## ArtLens Exhibition

### «Strike a Pose»

<https://www.clevelandart.org/artlens-gallery/artlens-exhibition>

Visitors was asked **to imitate the pose of a sculpture**, and provided feedback relating to the accuracy of their pose.

Visitors were able to share their poses and view others' poses, in addition to trying another pose.



# AI for visitor engagement

## Cleveland Museum of Art - USA

### ArtLens APP «Make a Face»

Facial recognition software matches visitors' facial expressions with one of artworks in the museum's collection. Visitors are shown a portrait to interpret the figure's emotion, then their facial expression is matched with another portrait. Visitors will see how meaning is created through facial expression in an artwork.



Visitors can "search" the collection through facial recognition, matching their own facial gesture to paintings and sculptures spanning centuries. This never before created interface connects visitors with the collection on a human and emotional level, creating snapshots that can be saved and shared through social media.

SEARCH THE COLLECTION BY  
MAKING DIFFERENT FACES

LOCAL  
PROJECTS



# AI to inspire immersion and engagement

The Salvador Dalí Museum  
St. Petersburg, USA

## «YOUR PORTRAIT»

Visitors to The Dalí may further their understanding of the genre of Cubism through the Museum's exclusive **new artificial intelligence experience**. Guests **can have their photo transformed into a one-of-a-kind Cubist work of art**. While their unique portrait is being generated, they will gain insight into Cubist imagery.



# AI to inspire immersion and emotional engagement

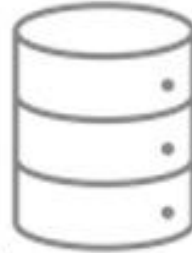
The Salvador Dalí Museum  
St. Petersburg, USA

## «DALÍ LIVES» interactive installation

Dalí Lives is an (AI) experience that **employs machine learning** to **create a version of Dalí's likeness** on a series of screens throughout the Museum, providing Museum visitors an opportunity **to learn more about Dalí's life** from the person who knew him best: the artist himself.

<https://youtu.be/BIDaxl4xqJ4>





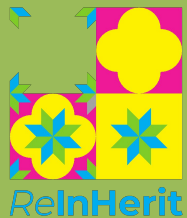
**Toolkit Strategy: Smart Apps and AI/CV Tools for phy-gital interaction**

# Analysis

## National Surveys and ICT Reports

### “Innovation as a process ”

Digital Transformation and **Innovation Process** in ReInHerit

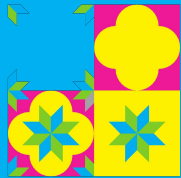


## STRATEGY

**Digital Technologies**  
“user-centered” approach  
to heritage management

### D3.1

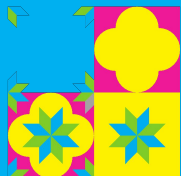
- Develop **phygital interaction** and **expanded user-experience**
- Promote **immersivity** and **interactive** experiences
- Create **narrative contents** via **participative storytelling**
- Use **gamification** and **playful** solutions
- Understand the user's **movement** and degree of **interest**
- Provide appropriate **information**
- Extend museum visits into sessions of **experiential education**
- Expand the modes of **visitor engagements**
- Trigger **emotions**, before, during and after the visit with **multisensory** engagement
- Develop **dynamic exhibits** that make visitors feel part of what is happening (enjoy).
- Create stronger **relationship** between the CH and the visitor



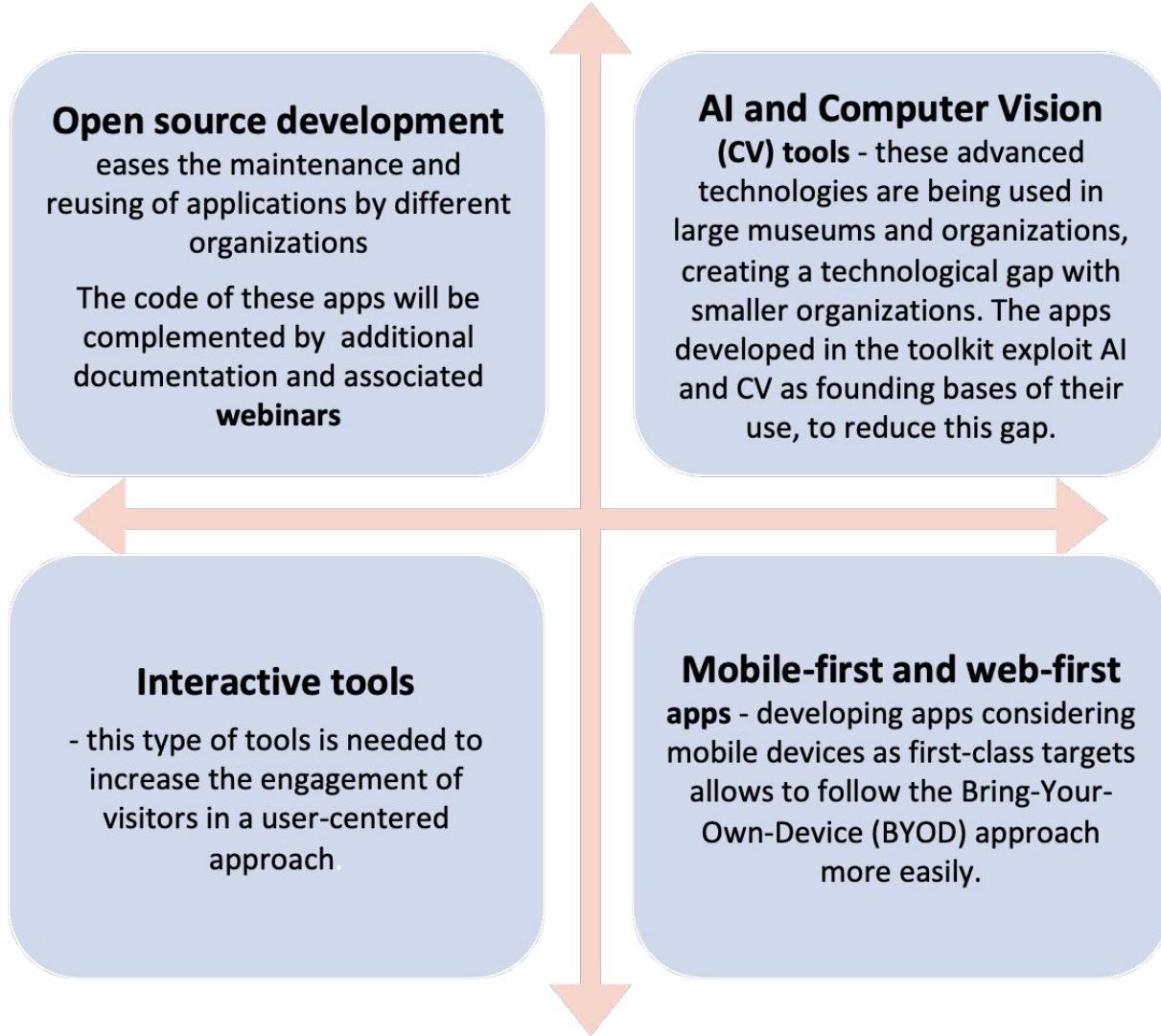
ReInHerit

# The toolkit development strategy

D3.2



ReInHerit



# Toolkit and Digital Hub

D3.2

Toolkit Strategy



## The Digital Hub

The ReInHerit Digital Hub is designed as the **interactive and dynamic space** to collect and share resources (webinars, tools and documentations) for cultural heritage professionals.

## The ReInHerit Toolkit

The ReInHerit Toolkit is a set of **applications, instructions, webinars** that provide guidelines, prototypes for developing technology-assisted immersive performances, digital exhibitions, and educational and smart tourism applications.

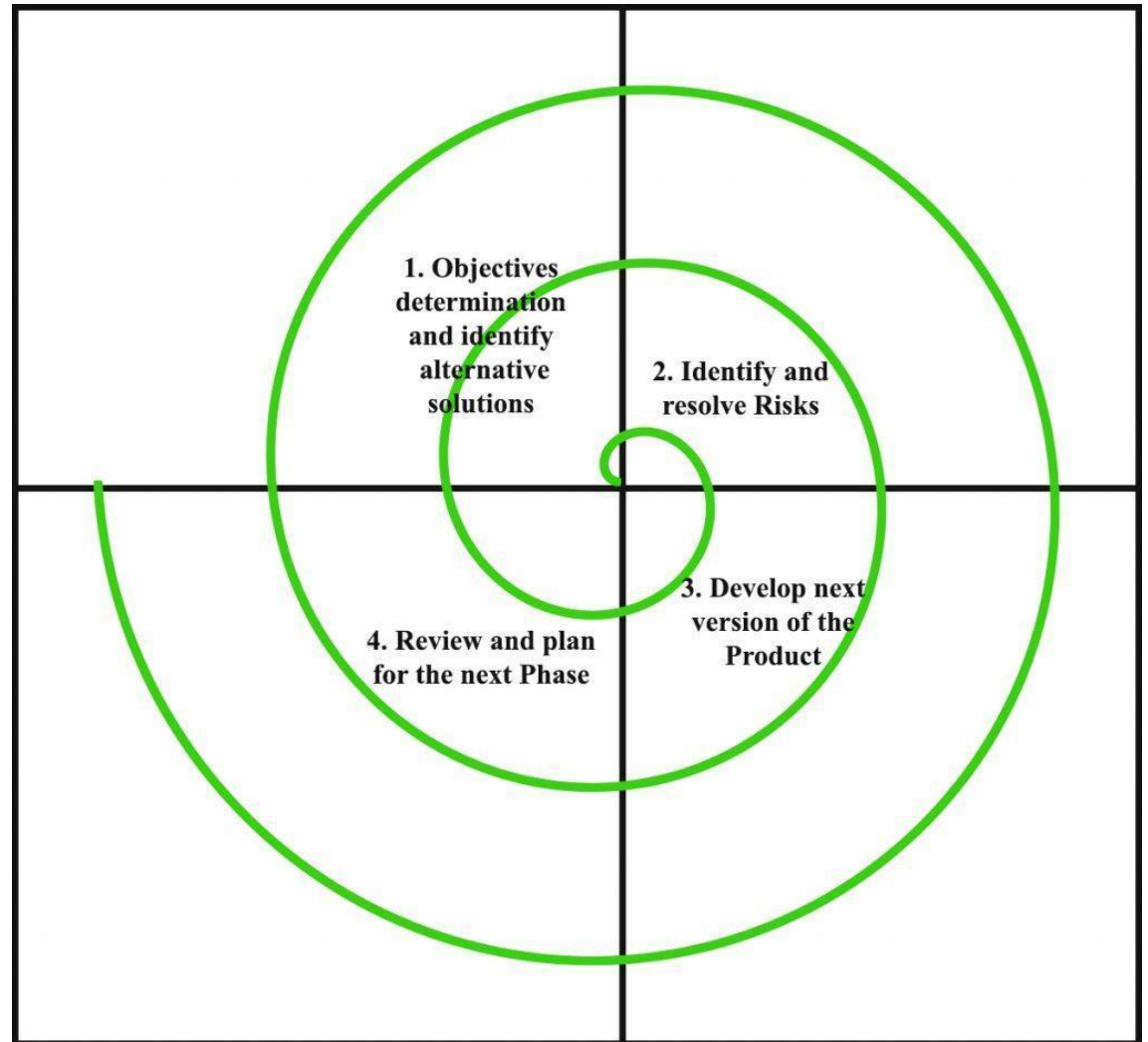
# The PROCESS

The **development** of all the apps follows a **spiral model** to manage the development risks (especially considering the *technological risks*)

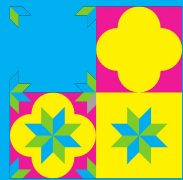
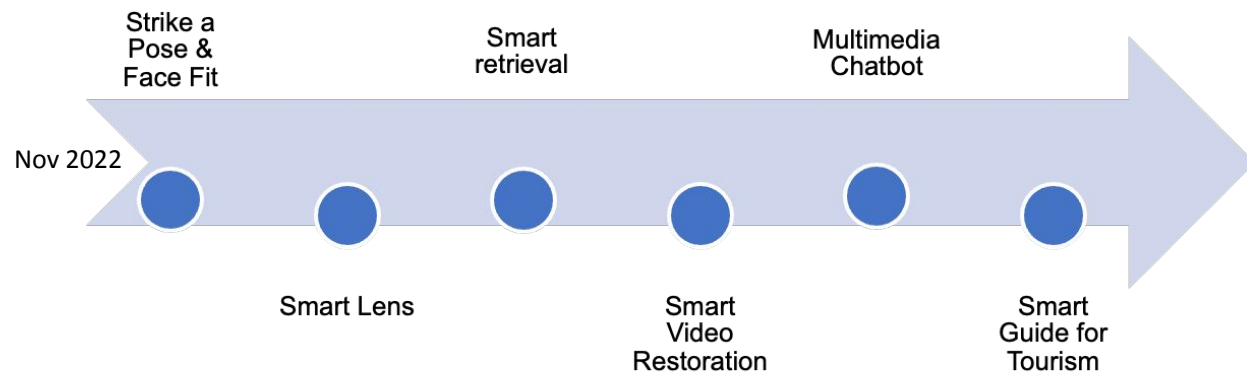
## DONE and TO BE DONE :

Activities related to other WPs

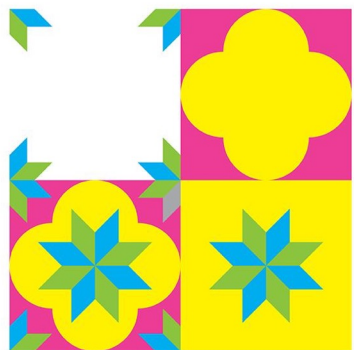
- Workshop
- Testing Usability
- Hackathons
- Webinars & Training



# Apps and Codes on Digital Hub



ReInHerit



**ReInHerit**  
Redefining the Future  
of Cultural Heritage



# Artificial Intelligence and Computer Vision for Cultural Heritage

Overview and example applications



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE  
**MICC**  
Centro per la Comunicazione  
e l'Integrazione dei Media



## Artificial Intelligence

---

- Services we use everyday to stream our favorite music, buy products online, or get a ride across town all use *Artificial Intelligence*. But because many of us don't necessarily understand how AI works, we take it for granted
- When it comes to museums AI can be incorporated across the spectrum, from visitor experience to behind the scenes, and the technology can and has come in many forms

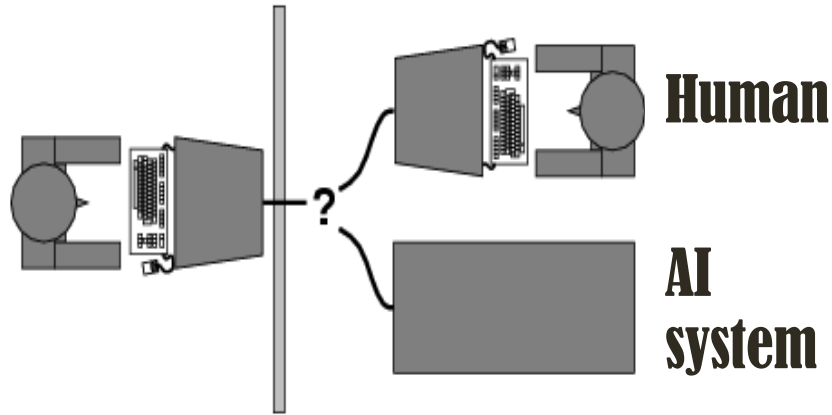


## A short introduction to AI

1950

*The Alan Turing's imitation game, an operational test for intelligent behavior of machines: Can machines think? Can machines behave intelligently?*

**Human  
interrogator**



*AI system passes  
if interrogator  
cannot tell which is the  
machine*

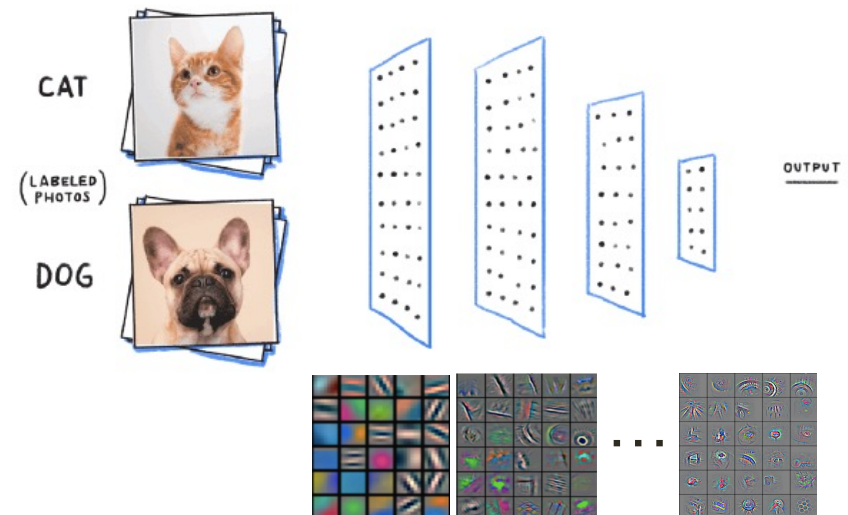
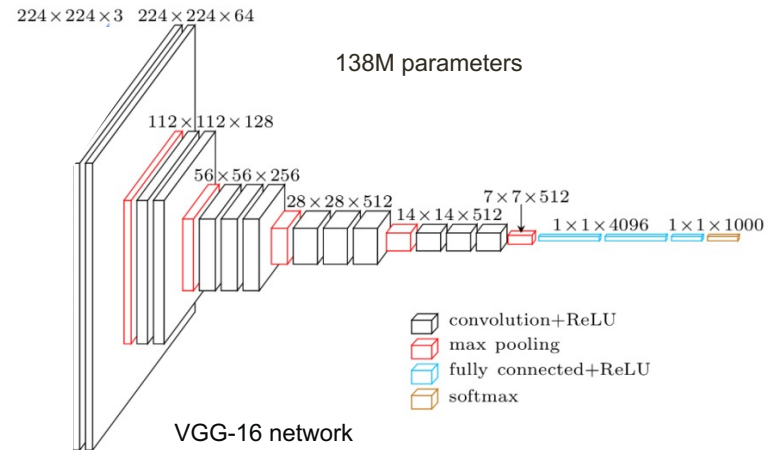
*interaction via written questions*

2022

*An intelligent machine is a system that is capable of extracting information from the environment, and sensing something, and understanding it and doing something in reaction to it*

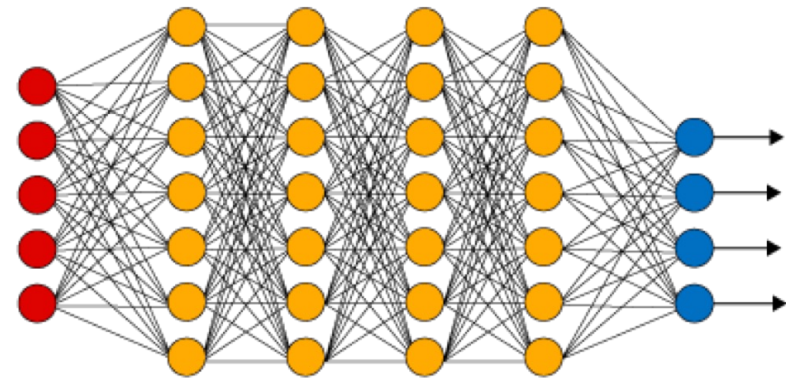
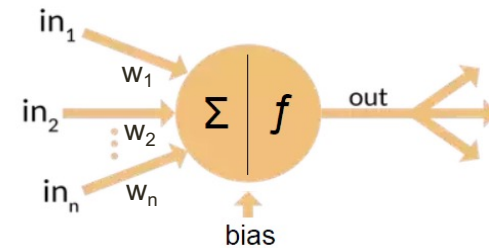
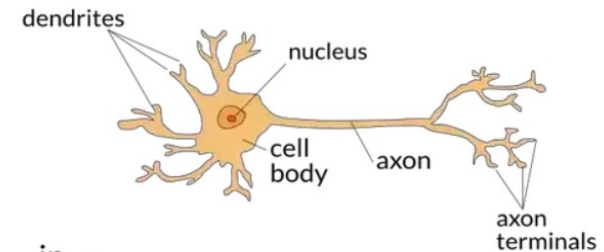
# Machine Learning and Deep Learning

- *Machine Learning* technology could learn, categorize, and make predictions on data
- *Deep Neural Networks* are computing systems loosely modeled on the human brain
- The building blocks are filters that are used to extract the relevant features from the input using the *convolution* operation
- The *deep* in deep learning stands for the idea of successive layers of representations. How many layers contribute to a model of the data is called the depth of the model.



## Training a Deep Network

- *Convolutions* are very simple operations, but the machine should learn thousands of weights to obtain meaningful descriptions of the entities to classify / recognize
- You must provide a huge number of correctly annotated examples and let the machine converge to a stable configuration of weights

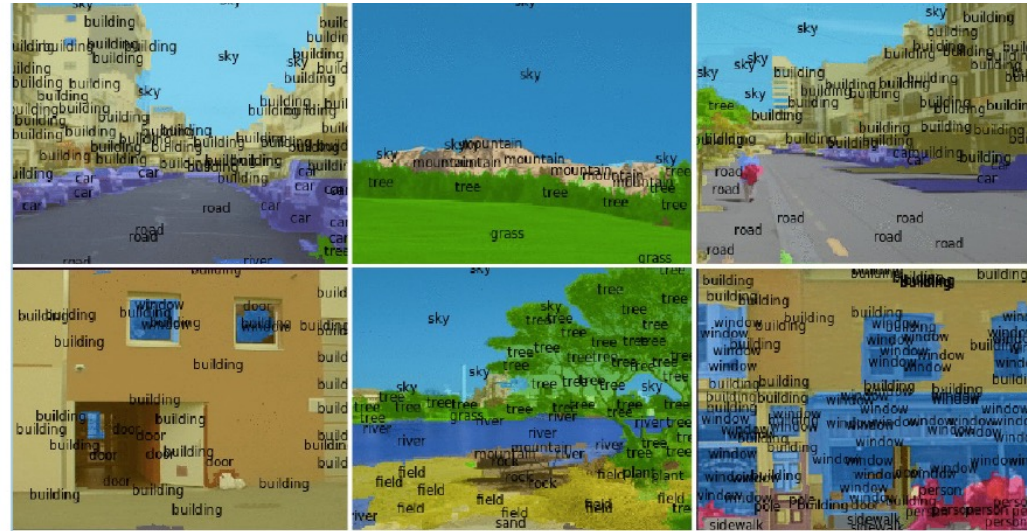


# Computer Vision as enabling technology

A powerful artificial sense to extract information from images: about places, objects, people....

Capable of understanding both contextual behaviors and situational conditions of people to provide the right information at the right time and place

- Fixed external cameras: to understand what visitors observe or do and determine their degree of interest...
- Mobile wearable cameras to augment the visitor experience providing the equivalent of multiple simultaneously active eyes....



# AI vs. ML vs. DL

- Artificial intelligence, machine learning, and deep learning: how do they relate to each other?
- AI: the effort to automate intellectual tasks normally performed by humans.
  - Includes ML and DL but includes also approaches that do not involve learning (e.g., using handwritten rules)
- ML: a machine-learning system is trained rather than explicitly programmed. It's presented with many examples relevant to a task, and it finds statistical structure in these examples that eventually allows the system to produce rules for automating the task.
- DL: a specific subfield of machine learning; aims at learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations.

# Why DL ?

- Around 2010 neural networks were almost completely shunned by the scientific community at large. A few people still working on them started to make important breakthroughs.
- In 2012 a team led by Alex Krizhevsky and advised by Geoffrey Hinton was able to achieve a top-five accuracy of 83.6% on the ImageNet large-scale image-classification challenge.
  - In 2011 the winning model reached 74.3%.
  - After 2012 everybody has started to use DL, until 2015 when an accuracy of 96.4% was achieved.
  - Nowadays deep convolutional neural networks (*convnets*) have become the go-to algorithm for all computer vision tasks, and in general for many perceptual-related tasks (e.g. speech recognition/translation)

## AI in museums

- Museums are utilizing AI technology mainly to engage audiences and personalize visitor experiences
- *What the AI technologies are used for audience engagement ?*
- *How are they implemented into museums?*
- *Are there any challenges or problems?*

## ARTIFICIAL INTELLIGENCE USED IN MUSEUMS AUDIENCE ENGAGEMENT

### AI

#### Chatbot

- Speech understanding and translation
- Customer services
- Connecting with the audience

#### Machine Vision

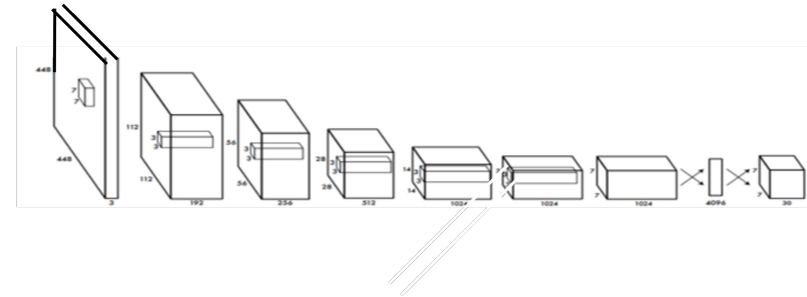
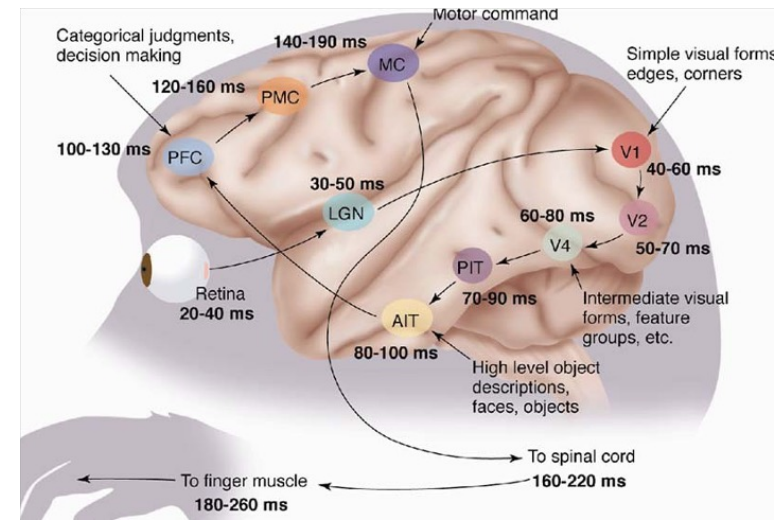
- Automatic tagging of images
- Track attendance for future exhibition strategy
- On-line recognition of artworks
- Assistance for visually impaired visitors

#### Machine Learning

- Near-human robots
- Deepfakes for visitor interactivity and engagement

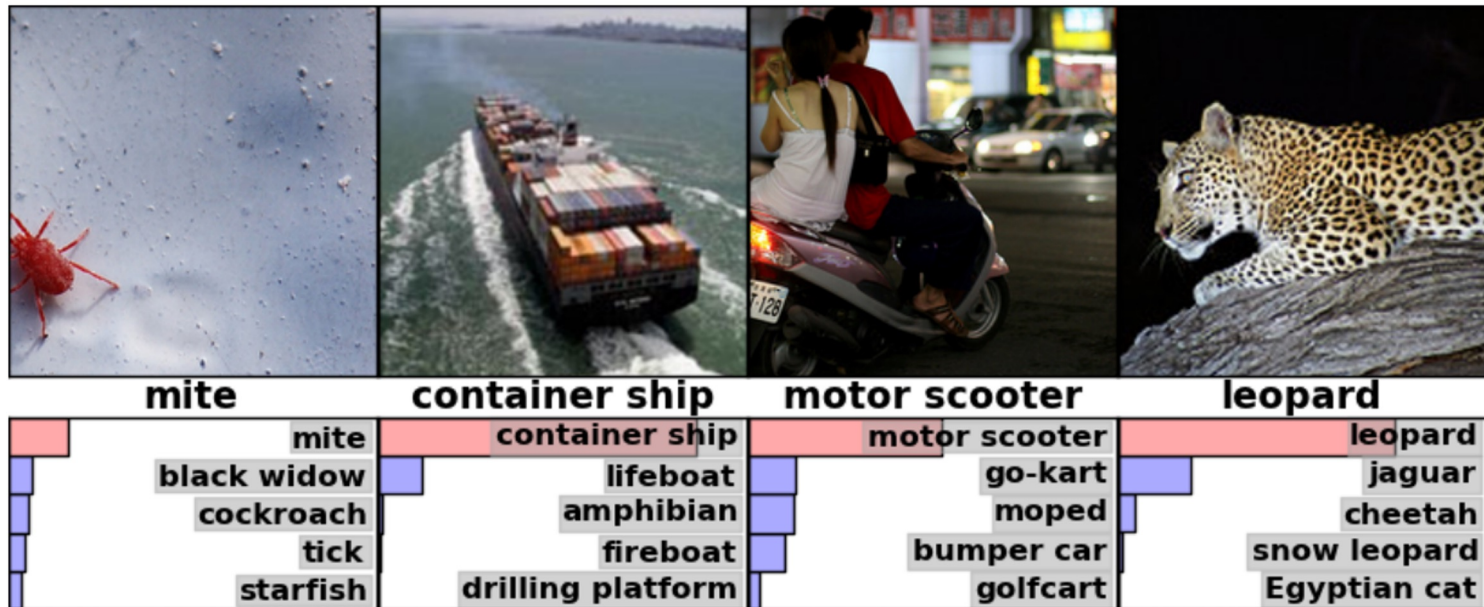
# Interesting Computer Vision tasks solved with AI

*Remind: if it's written in PowerPoint it's AI, if it's written in Python it's Machine (or Deep) Learning.*



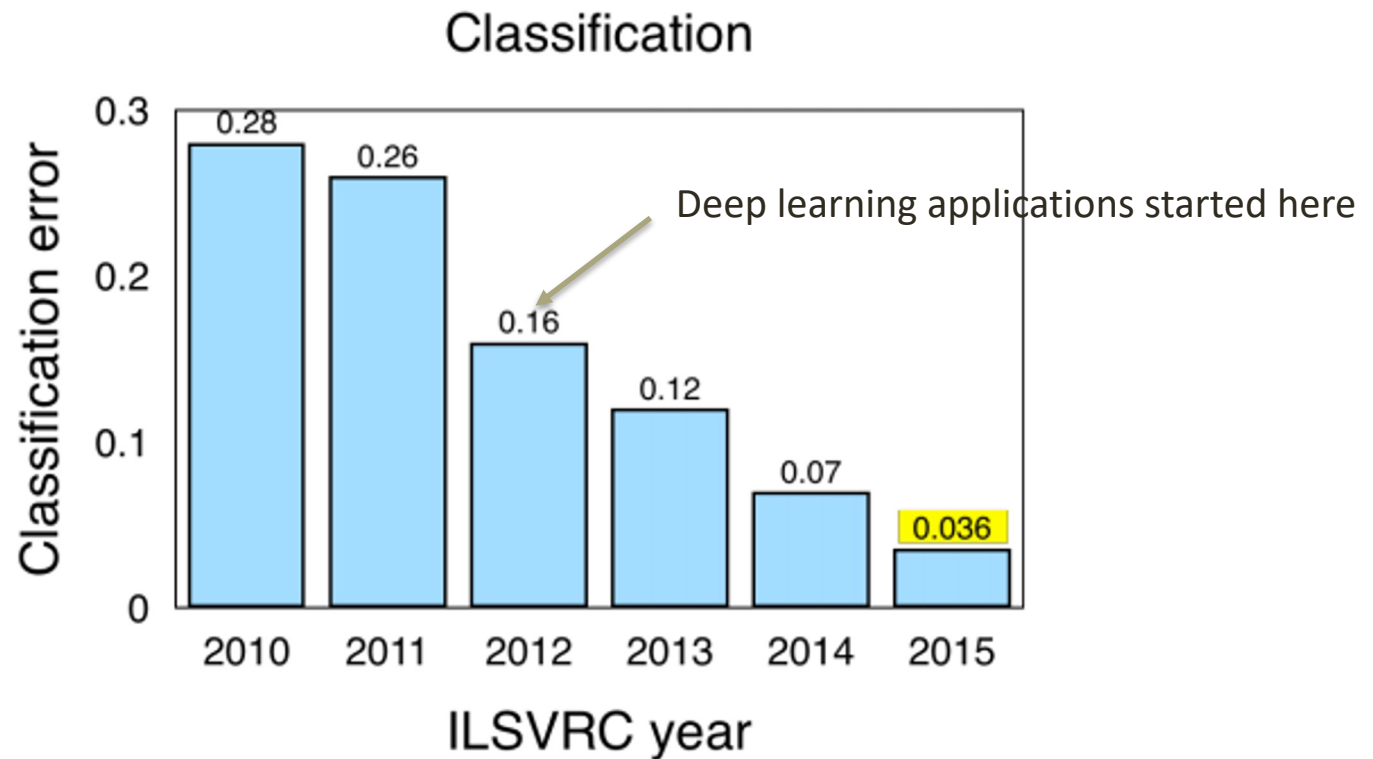
# Image classification

- The goal is to classify the content of an image.
- An important international competition (ILSVRC) uses ImageNet. Participants provide the most probable concepts (e.g. the 5 most probable)



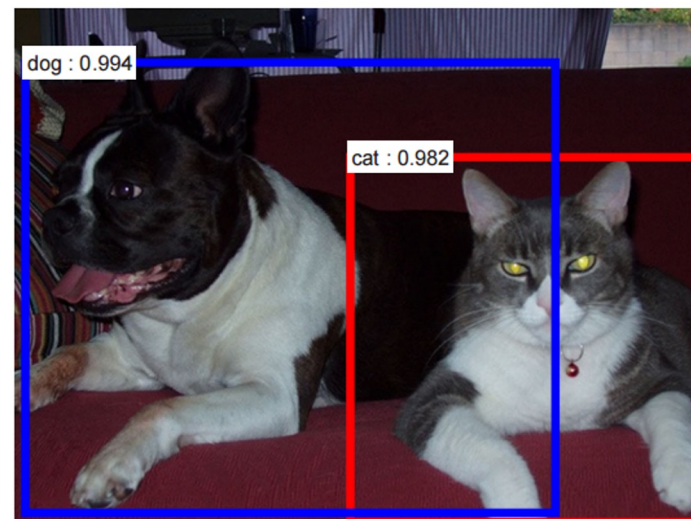
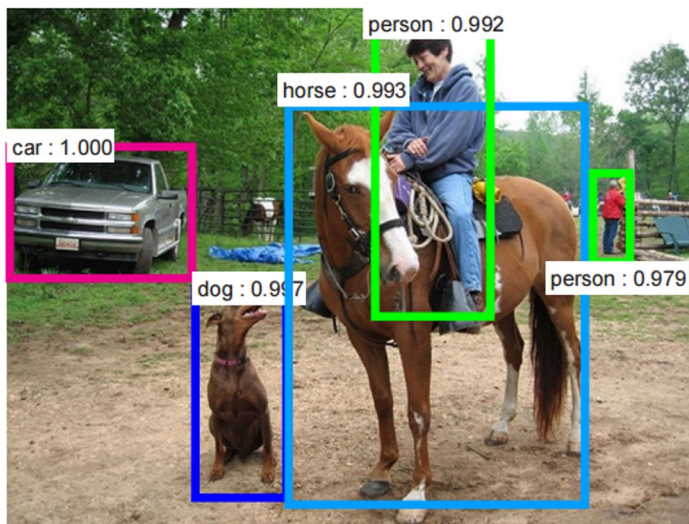
## ImageNet: ILSVRC results

- Result in ILSVRC (classification) over the years



# Object recognition

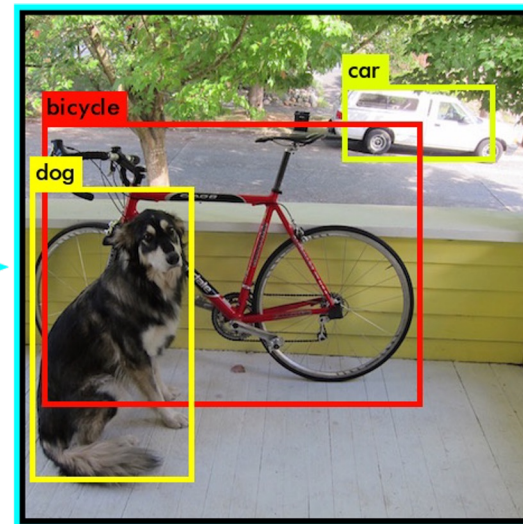
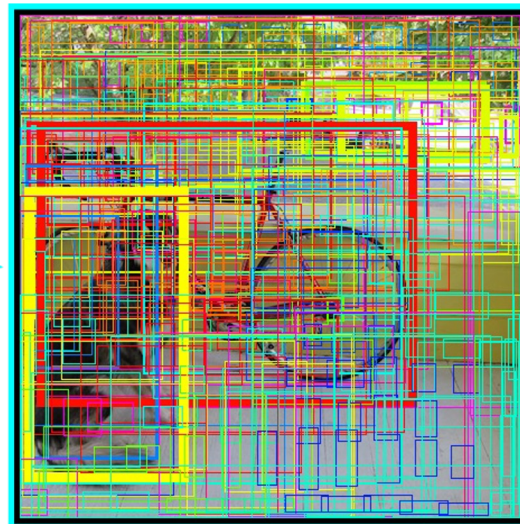
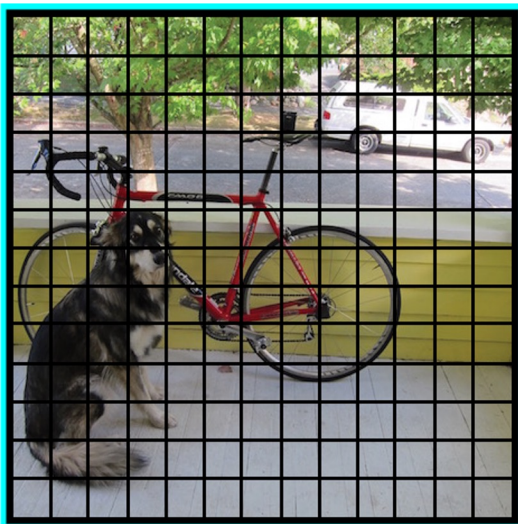
- The problem is to identify where an object that belongs to a specific class is. We represent the position as bounding box.
- Typically, it requires to move a sliding window over the image and then analyze its content to classify the object



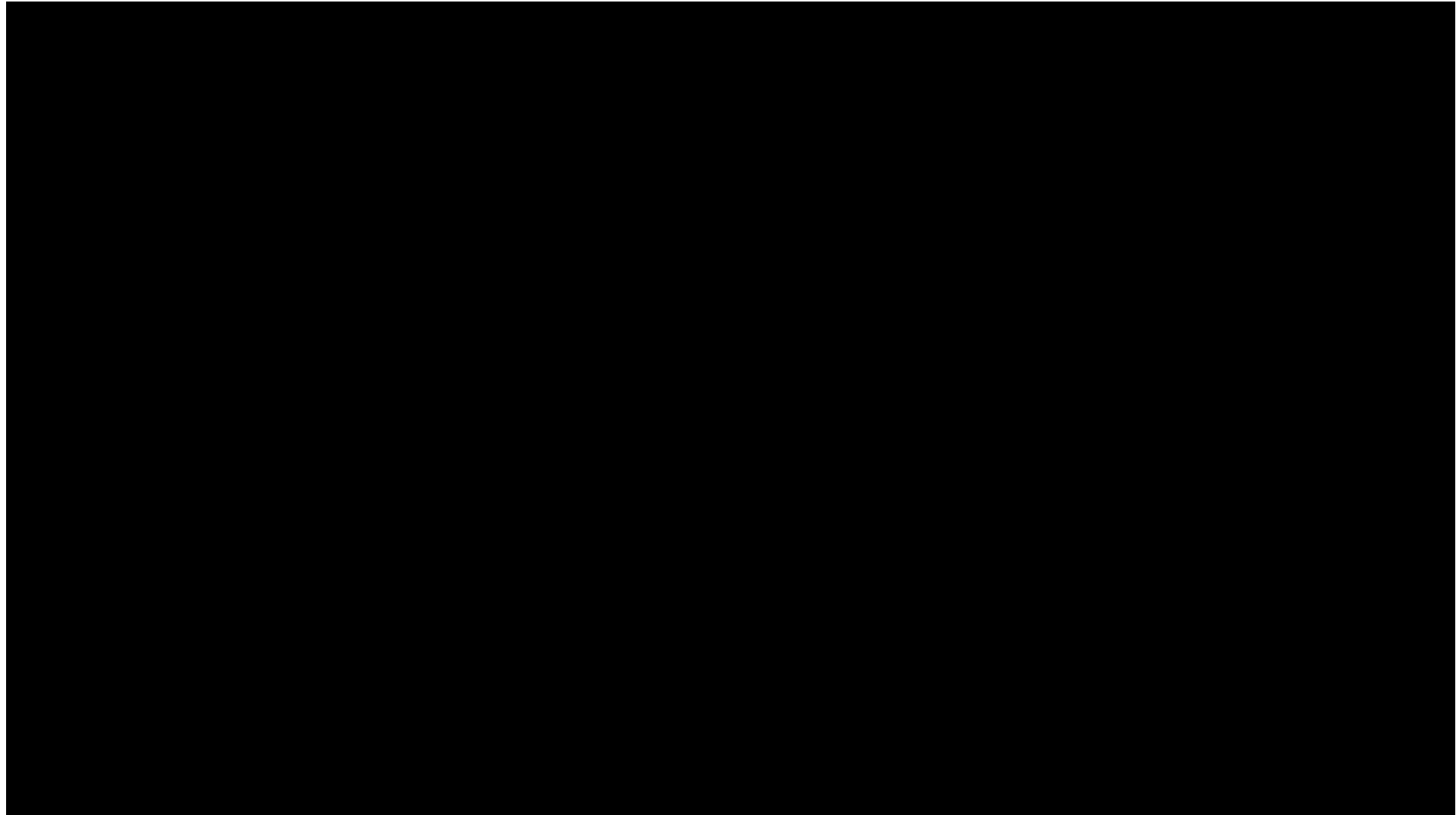
## Fast object recognition

- The current fastest approach apply a single neural network to the full image.

This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.



# Video object recognition



## Semantic segmentation

- It's a more complex case of object recognition
- Label each pixel with a class of objects (Car, Person, Dog, ...) and non-objects (Water, Sky, Road, ...)



## Semantic segmentation for autonomous driving



## Instance segmentation

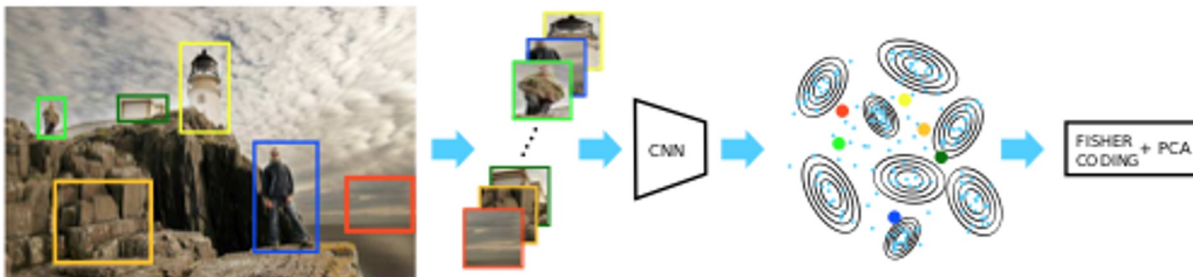
In Instance aware Segmentation we find out the individual instance of each object, i.e. different instances of the same object are segmented differently.



# Image retrieval

Content-based image retrieval can be performed using the whole image representation obtained from some layer of a CNN

- Or combining these representations, e.g. from different parts of the image



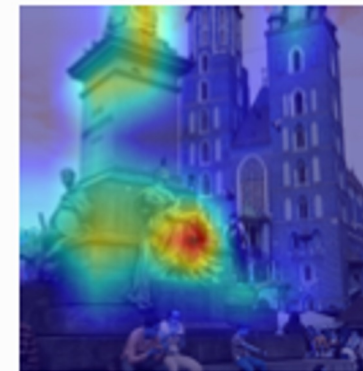
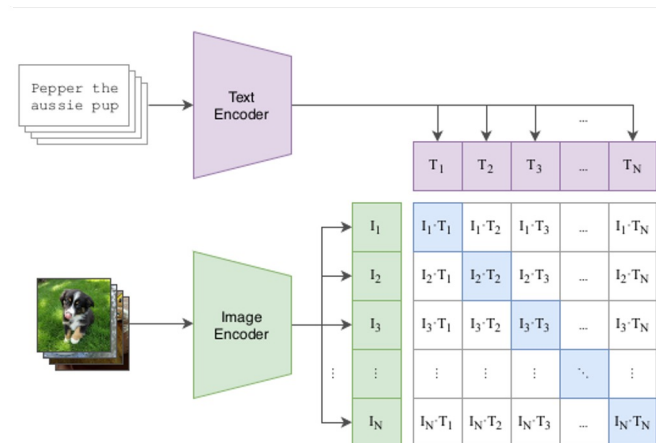
# Image retrieval

Content-Based Image Retrieval (CBIR) combining text and images

CBIR is the problem of searching for semantically relevant images by analyzing their visual content given a query that describes the user's needs

We investigate how textual supervision and multi-modal training techniques can be applied in content-based image retrieval

- This is an ongoing research activity, we use the recent CLIP neural network that combines text and image features



## Captioning

- Given an image we want to automatically generate a natural language description of this image and its regions.
- CNNs are coupled with RNNs/transformers that learn the language model and can generate new text.



man in black shirt is playing guitar.



construction worker in orange safety vest is working on road.



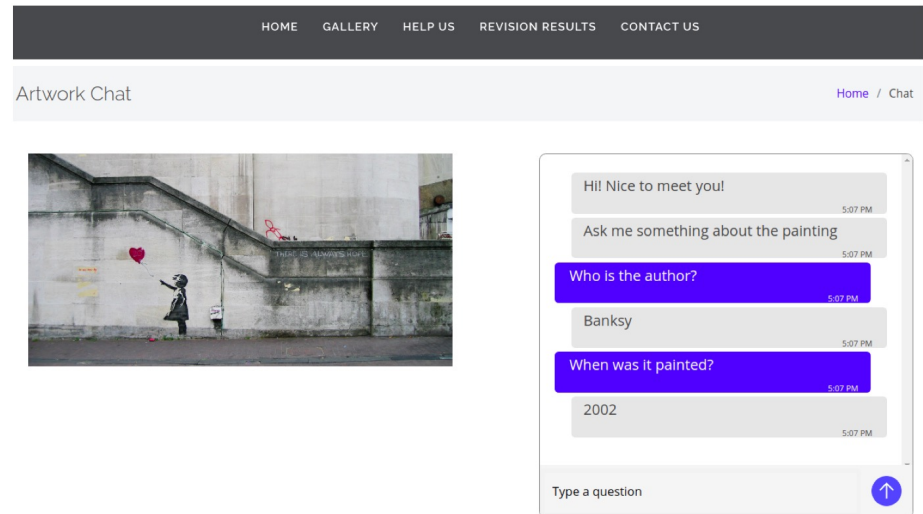
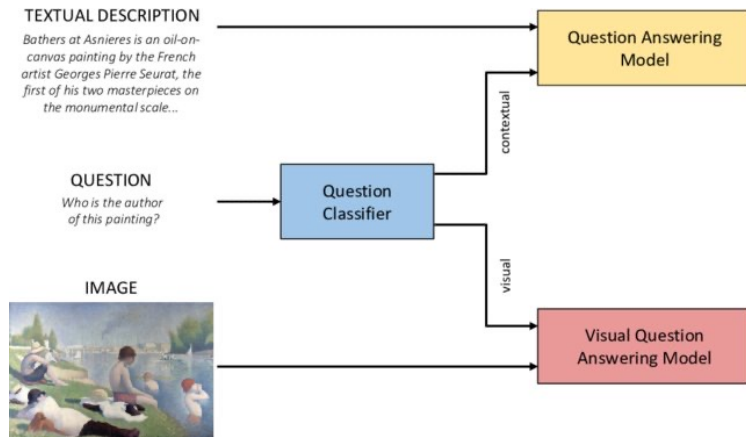
two young girls are playing with lego toy.

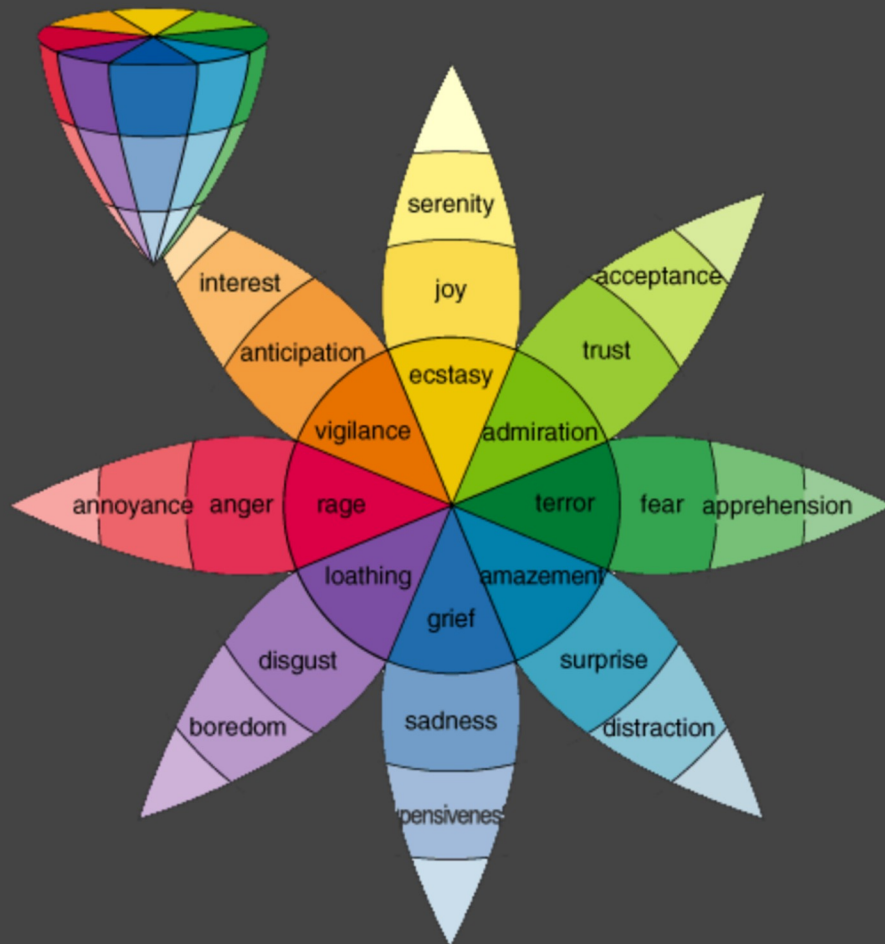


boy is doing backflip on wakeboard.

# Visual question answering

- In Visual Question Answering (VQA), users can interact with a neural network by posing questions in natural language and receiving answers about the visual content.
- An application of VQA for Cultural Heritage may consider two types of questions: *visual* if they refer to the content of the artwork and *contextual* if they refer to knowledge deductible only from an external source.





Selected Emotion:

◀ joy [589] ▶

Top ANPs	sentiment	emotion
1. ▶ happy smile	1.92	0.388
2. ▶ innocent smile	1.92	0.376
3. ▶ happy christmas	2.0	0.373
4. ▶ happy father	2.0	0.358
5. ▶ happy wedding	1.72	0.348
6. ▶ friendly smile	1.92	0.346
7. ▶ delicious cupcake	1.71	0.341
8. ▶ shy smile	0.62	0.340
9. ▶ charming smile	1.92	0.339
10. ▶ happy birthday	1.79	0.337
11. ▶ warm smile	1.92	0.336
12. ▶ happy mother	2.0	0.333
13. ▶ happy halloween	1.81	0.331
14. ▶ delicious drink	1.59	0.330
15. ▶ happy heart	2.0	0.329
16. ▶ happy kids	2.0	0.323
17. ▶ healthy food	1.69	0.312
18. ▶ happy guy	1.61	0.312
19. ▶ fresh food	1.59	0.307
20. ▶ delicious pie	1.76	0.296

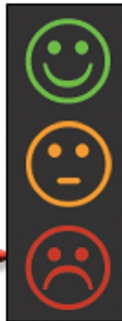


24 emotions

Sentiment Words

Visual Sentiment Ontology

SentiBank (1200 detectors)



Wheel of Emotion (Psychology)

Data-driven Discovery

Adj + Nouns = ANP's

Detector Training and Validation

Sentiment Prediction

# Sentiment analysis: example



**traditional tattoo**  
 little pony  
 sexy tattoo  
 dead skull  
 awesome tattoo  
 fresh tattoo  
 hot body  
 fancy hat  
 bright stars  
 gorgeous butterfly



pretty rose  
**beautiful rose**  
 golden flower  
 fresh rose  
 awesome blossom  
 stunning flower  
 cheerful flowers  
 fantastic flowers  
 happy mother  
 dying rose



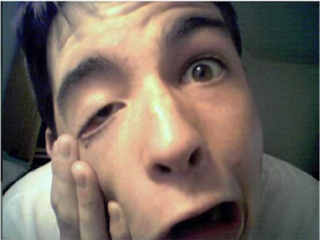
**colorful food**  
 healthy food  
 colorful flowers  
 amazing food  
 pretty flowers  
 tiny flower  
 fresh food  
 outdoor market  
 yummy food  
 natural food



**famous tower**  
 magnificent church  
 famous monument  
 famous castle  
 splendid church  
 magnificent castle  
 ancient monument  
 haunted castle  
 ancient church  
 famous church



**curious deer**  
 wild deer  
 young deer  
 graceful animals  
 dark winter  
 broken legs  
 cute animals  
 awesome animals  
 slender legs  
 shy dog



tired eyes  
**ugly face**  
 mad face  
 bad eye  
 chubby face  
 sexy lips  
 pretty eyes  
 silly face  
 innocent face  
 fat face



**abandoned car**  
 abandoned vehicle  
 derelict car  
 broken tree  
 abandoned tank  
 abandoned train  
 abandoned places  
 creepy house  
 abandoned boat  
 ancient farm



**cloudy moon**  
 cloudy night  
 super moon  
 bright moon  
 shiny moon  
 magical moon  
 stormy night  
 clear moon  
 dark places  
 clear night



**ugly fish**  
 weird fish  
 dark night  
 dead fish  
 fresh water  
 abandoned school  
 noisy bird  
 tiny teeth  
 wet leaves  
 dirty glass



haunted castle  
**creepy house**  
 abandoned cemetery  
 scary house  
 dark tower  
 holy cross  
 ancient castle  
 creepy shadow  
 abandoned house  
 fascinating places



**broken ice**  
 little island  
 lost coast  
 gentle waves  
 icy river  
 warm sun  
 smooth waves  
 rough waves  
 calm sea  
 cold morning



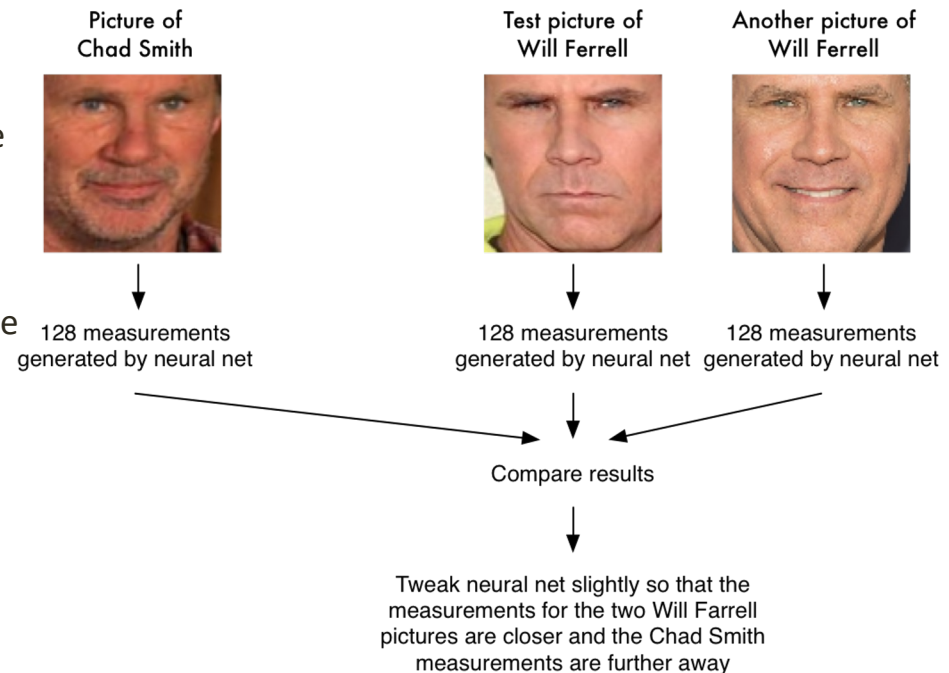
**muddy dog**  
 wet dog  
 dirty dog  
 playful dog  
 angry dog  
 happy dog  
 funny dog  
 adorable puppy  
 crazy dog  
 happy puppy

- Classification results: in red the terms used by the photographer to describe the content of the photo

# CNNs for face recognition

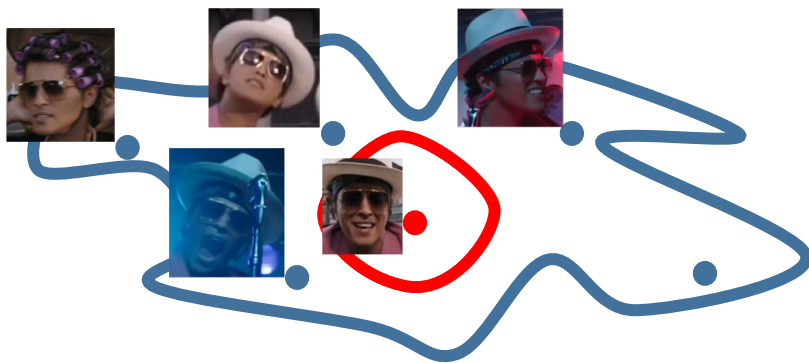
- A recent solution is to use CNNs that receive triplets of photos: two of the same person and one of a different person.
- The network is trained to minimize the difference between the photos of the same person and to differentiate from the wrong person
- The network will thus learn what measures can be used to differentiate between persons. The measures will be computed for each photo and the most similar person will be returned.
- can be seen as a problem of retrieval or classification.
- Related tasks: gender classification, age estimation, sentiment classification

## A single 'triplet' training step:



# Multiple Face Tracking by Recognition

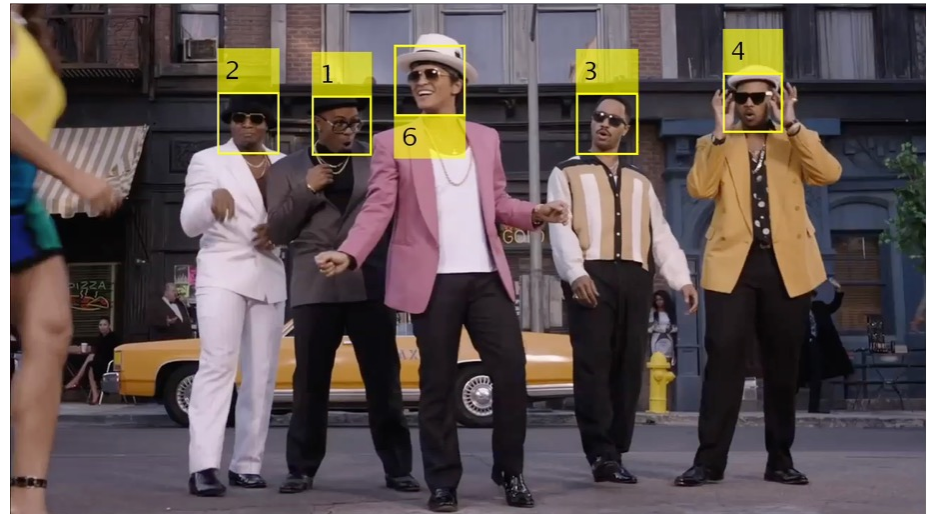
- Track multiple unknown face identities in unconstrained videos.
- Uses state-of-the-art CNN representation for face recognition as face appearance model.
- Performs incremental learning: extends the face appearance model using the video exemplars collected while tracking.



Appearance Learned Offline (i.e. VggFace Deep Learning )

The extended appearance learned from video

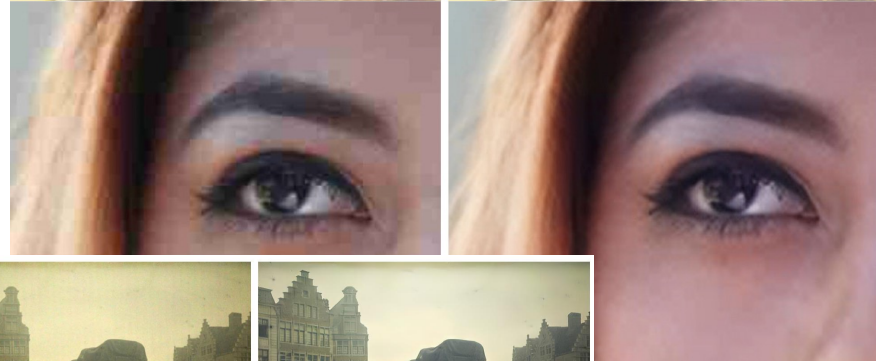
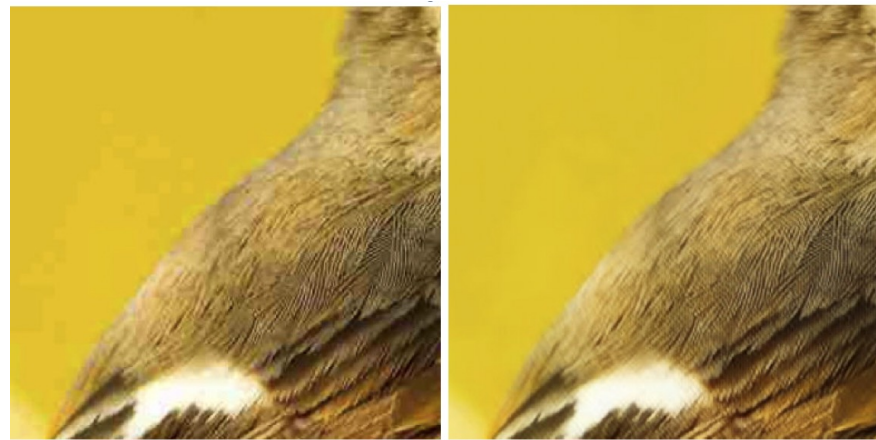
• Video data exemplars



# Style transfer

- Style transfer is the technique of recomposing images in the style of other images.
- Neural networks can learn the style (i.e. texture) of an artist from its paintings and restyle other images according to it.
- Both style and content are learned using CNNs.





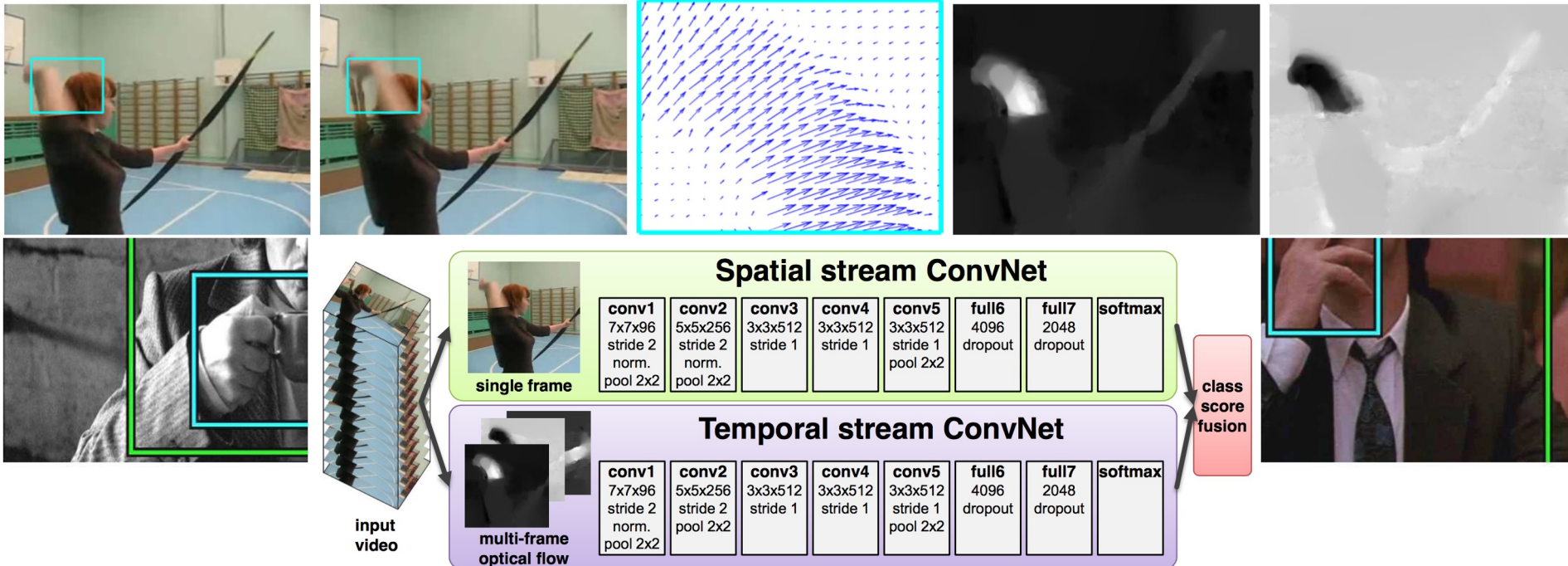
## Video and image restoration

Old video and image archive materials can be restored or revamped, addressing both analog media (e.g. b/w photos, films) or digital materials (e.g. old video formats with low res.)

Colorisation is a specialized version of restoration.

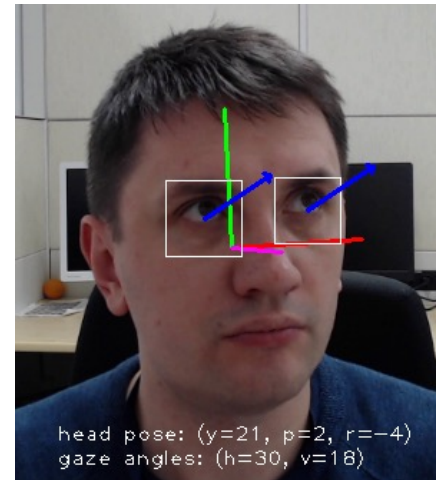
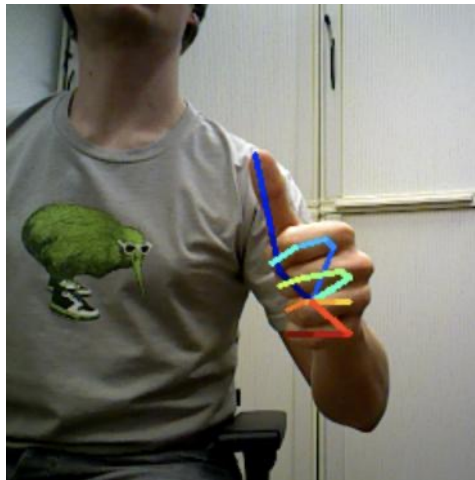
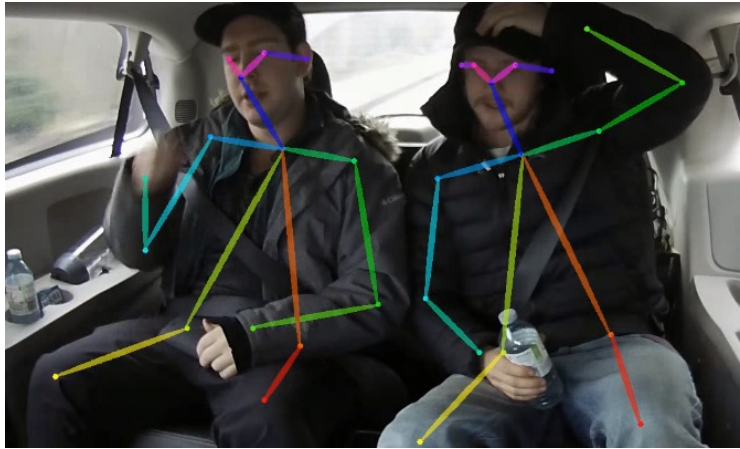
# Action recognition

- To detect and recognize events in videos there's need to use features related to motion, like trajectories or optic flow
- The CV community has developed many tasks (e.g. human action recognition, sports, event detection, etc.)



# Pose recognition

- Pose recognition can be performed on whole body or body parts, like head, hands or eyes

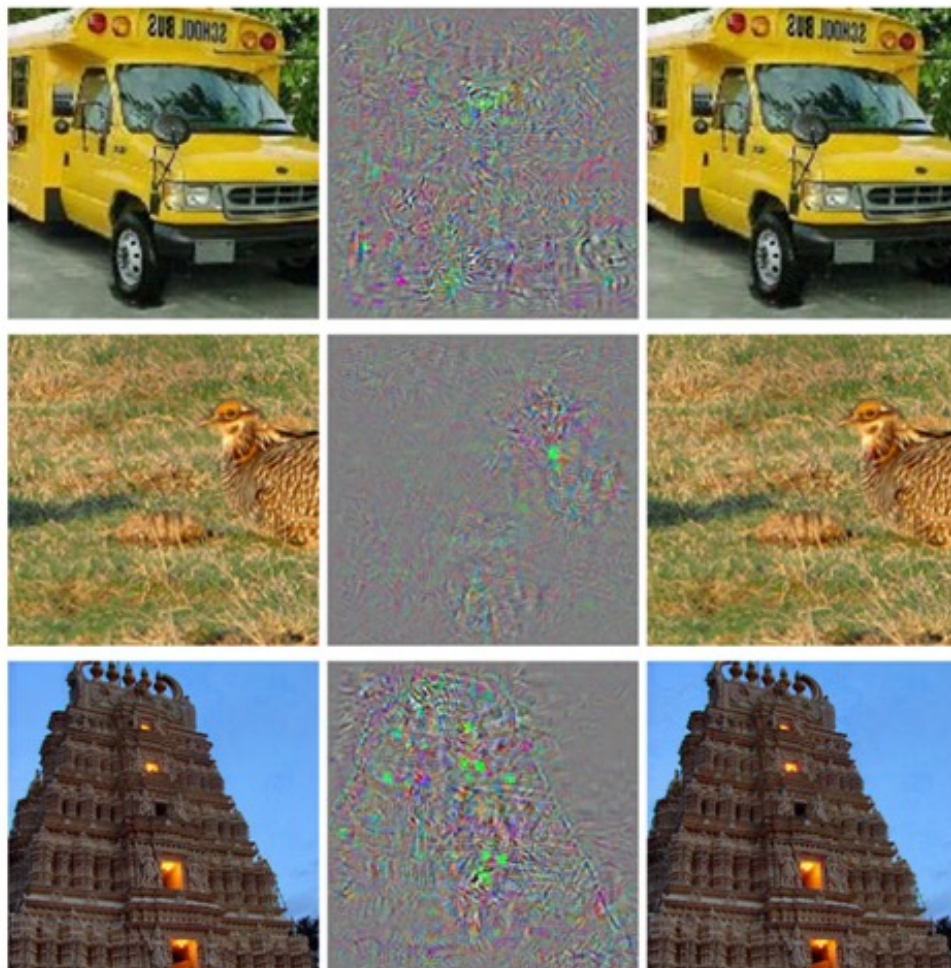


# What can go wrong ?



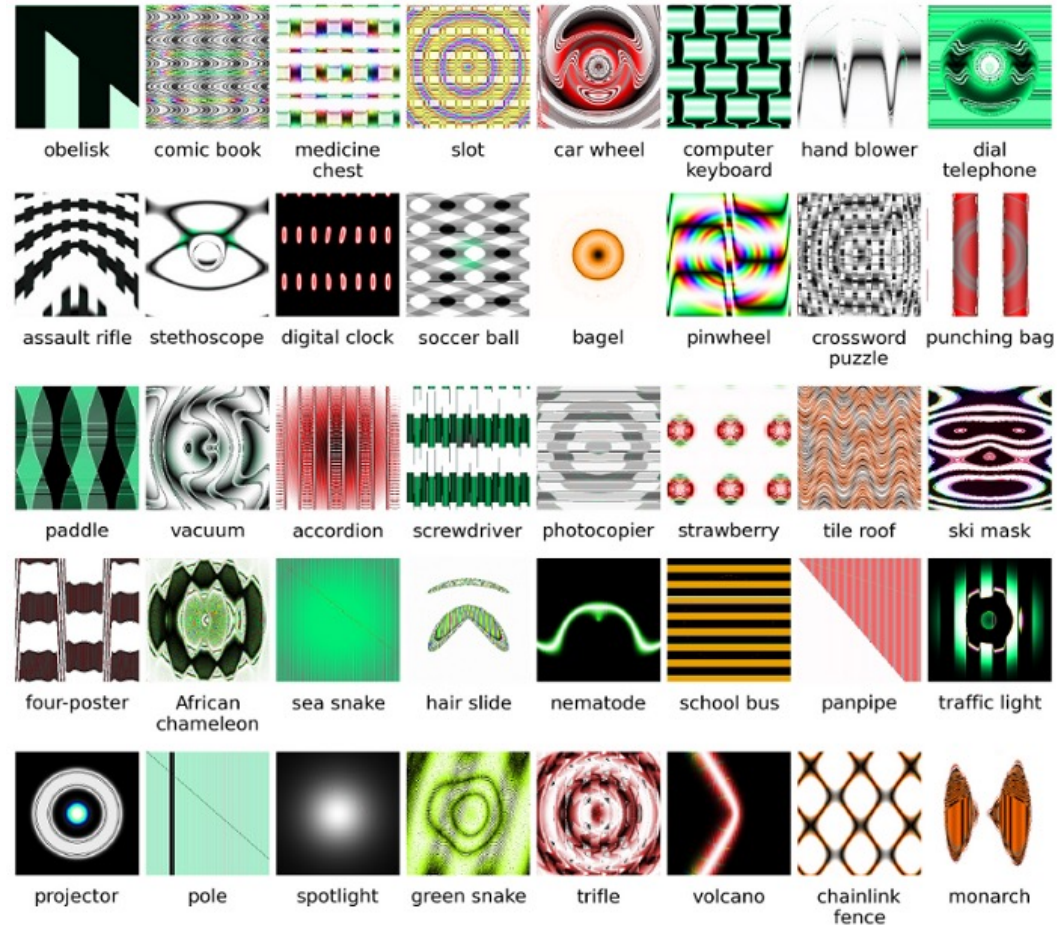
## Similar images

- (Left) is a correctly predicted sample, (center) difference between correct image, and image predicted incorrectly magnified by 10x, (right) adversarial example.
- When CNNs see the examples on the right they completely mistake their classification.



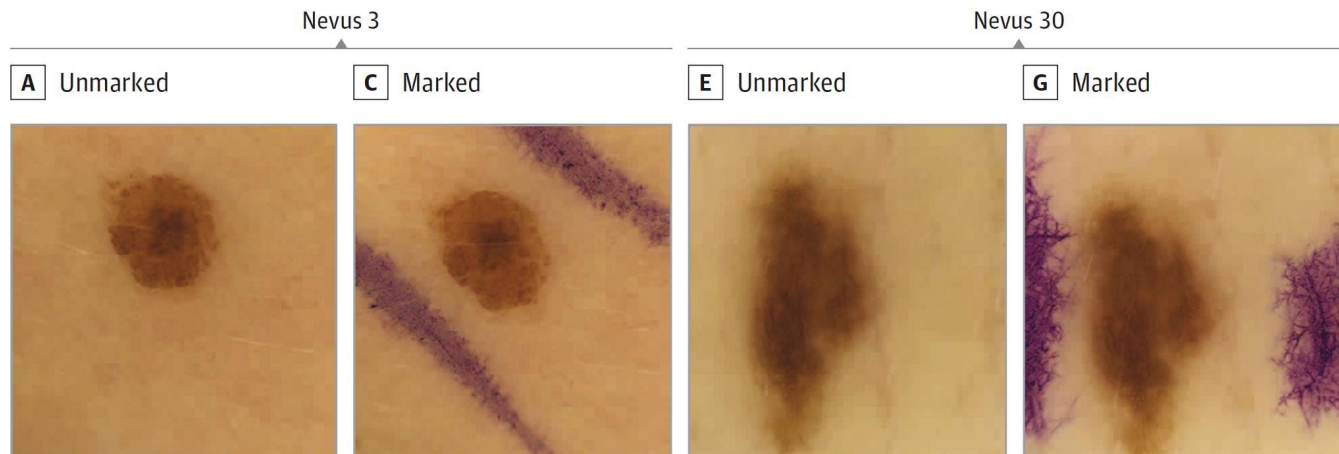
## You say what ?

- It is possible to produce images totally unrecognizable to human eyes that CNNs believe with near certainty are familiar objects.
- This shows interesting differences between human vision and current CNNs, and raise questions about the generality of CNN computer vision



## Learning wrong things

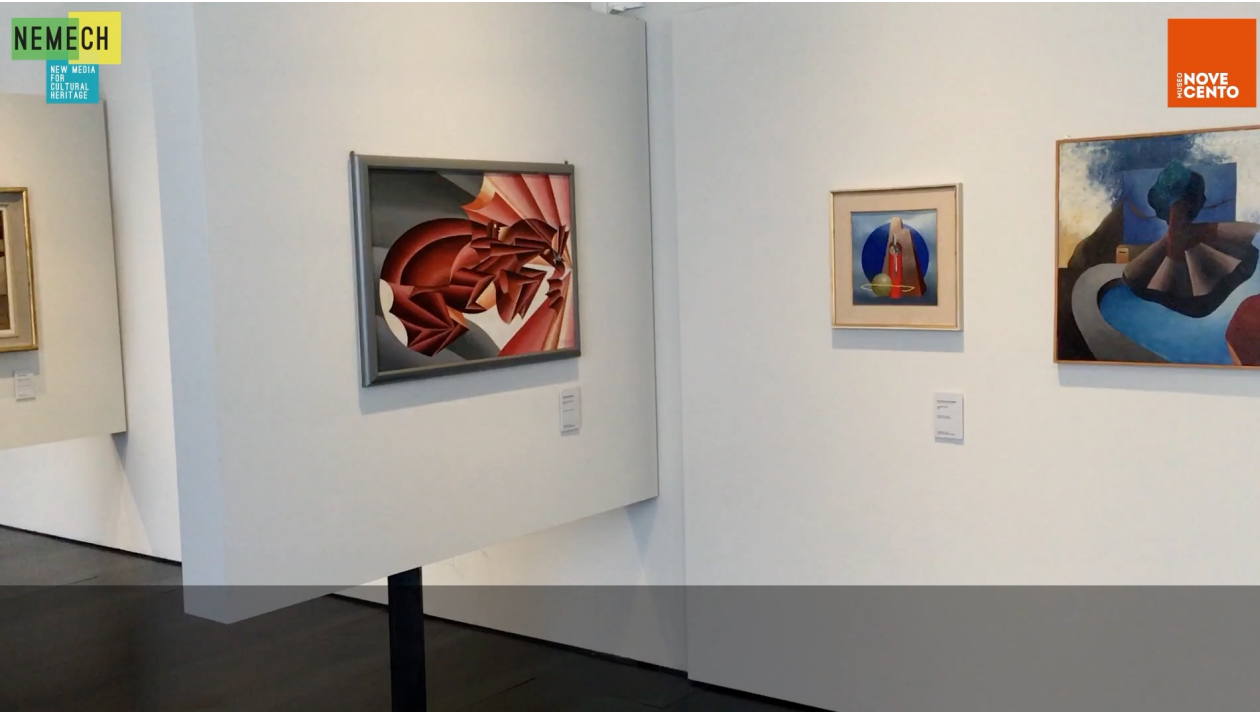
- A CNN-based system has been approved for use as a medical device in the European market to detect skin neoplasms...
- ... but further tests (published on JAMA) have shown that if a marker sign is near a nevi, then the nevi is classified as perilous.
- The fact is that dermatologists mark risky lesions with a marker, and these images were used to train the network !



# Fantastic CH applications and how to make them

Examples of use of AI & CV  
techniques for CH applications

# NEMECH experiences



## The Imaging 900 project at Museo del Novecento in Florence

Imaging 900 - Nitrito in velocità Indietro

 **Nitrito in velocità**  
Fortunato Depero - 1932



Con il suo Nitrito in velocità, appartenente a una seconda fase futurista all'inizio degli anni Trenta, l'artista risolve il soggetto, un cavallo e il suo cavaliere, in una fusione di geometrie dove predominano i con. La velocità, la macchina, la cavalcata folle verso il progresso, sono altrettanti temi cari all'immaginario futurista.

Il cavallo e il cavaliere del dipinto sono composti attraverso l'assemblaggio di geometrie coniche dai colori smaltati, simili alla carrozzeria metallica di una macchina.

Rendi una tua foto alla maniera dell'opera **Nitrito in velocità** di **Fortunato Depero**. Selezionala dalla gallery e ricevila per e-mail!

▼

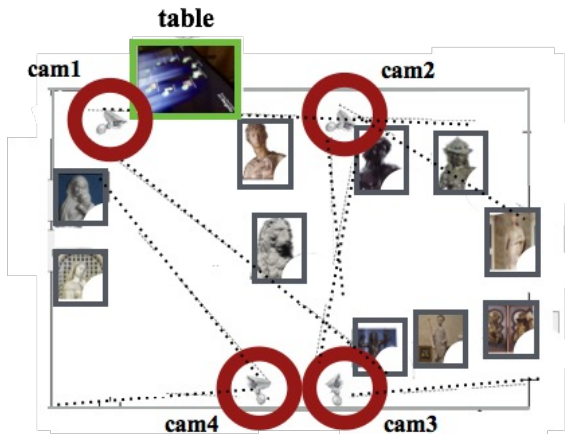


# What's in Imaging 900 ?

- Recognizing the artwork -> it is a problem of instance recognition
  - In this case we used a ML approach (based on a variation of exemplar SVM)
- Filtering the selfies -> style transfer from a specific artwork to a photo (e.g. a selfie) is akin to applying a filter. In this case the filter is a neural network
- Designing the system -> style transfer may be computationally expensive.
  - Use a client-server architecture where the server executes the style transfer

# NEMECH experiences

The MNEMOSYNE project  
at Bargello Museum in Florence



MICC | THALES  
Media Integration and Communication Center

# MNEMOSYNE

smart museums

UNIVERSITÀ DEGLI STUDI FIRENZE  
DINFO  
CENTRO DI COMPETENZA REGIONALE DELL'INFORMAZIONE

NEMECH  
Centro di competenza  
UNIFI Regione Toscana

M7BAC

MINISTERO PER I BENI E LE ATTIVITÀ CULTURALI

SPSAE-IVICE

REGIONE TOSCANA

## Computer Vision enabling technology



avg. detection rate

70%

avg. re-identification accuracy r1

99% - 61% - 54% - 18%

avg. profiling accuracy

89%

# What's in Mnemosyne ?

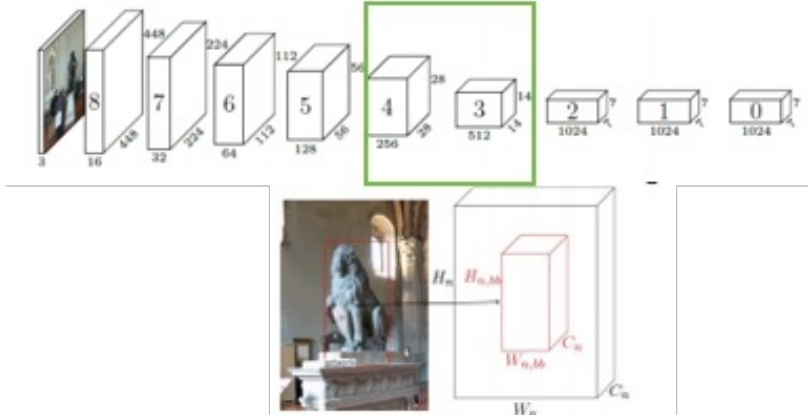
- Recognizing where is a person -> it is a problem of object detection
  - Use a neural network
- Computing the distance from an artwork -> it is a problem of camera calibration and real world distance computation
  - Use Computer Vision techniques to calibrate a camera and map pixels to real-world coordinates
- Tracking a person in the museum -> it is a problem of person re-identification
  - Compute a descriptor of the person from the dresses (privacy preserving...) and use it to track the movements. No identity is associated, no personal features (e.g. face) are used
- Providing suggestions on other artworks -> it is a problem of recommendation
  - Solve using collaborative filtering. Nowadays also deep-learning approaches can be used
- Designing the interface -> design “natural” interfaces
  - Design for a touch table, simplify interactions, mask latencies of the vision system

# NEMECH experiences

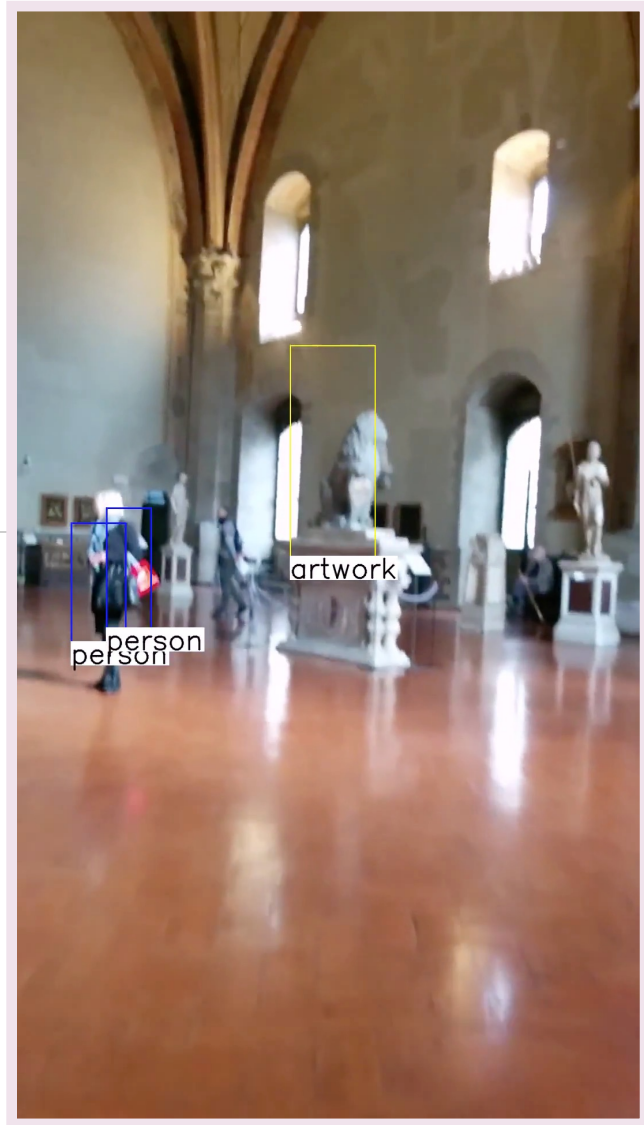
The See For Me project



Nvidia Shield K1



Computer Vision enabling technology



Wearable vision system

true	false	miss
62,8%	0,5%	37%



<https://vimeo.com/187957085>

# What's in See for me ?

- Recognizing what is a person or an artwork -> it is a problem of object detection
  - Solved using a neural network
- Recognizing what specific artwork is framed -> it is a problem of instance recognition, that can be cast as a problem of content-based image recognition
  - Solved using a neural network
- How to speed up the cascade of object detection and instance recognition -> use the same neural network, employing different layers for the different tasks
- How to understand if somebody is speaking -> it's a problem of audio classification
  - Solved using ML (no neural network). It can be solved also using NN
- How to understand if somebody is moving -> it is a problem of signal (accelerometers) classification
  - Solved using a simple ML classifier
- How to design the app -> it is a problem of dealing with a specific platform to access the NN computation
  - But multi-OS tools can be used to address both iOS and Android – must check how to deal with NN inference

# Re-Identification System for reactive personalized audio-guide

- Our system implements a reactive personalized audio-guide in the context of a museum visit.
- The system identifies the visitors in real-time but it stores the detection in the database only if he stands for a sufficient interval of time in front of the camera (and of the painting), adding a set of categorical and numerical features (i.e. age, gender, recognized emotions) used to build a profile of the user and the history of his visit and of his emotional reactions to artworks.
- Custom audio explanations are activated by micro speakers contextual to paintings for each user interested in an artwork and standing in front of it on the basis of analogies, similarities and differences with other artworks of the same artist or of the same subject, or otherwise related, viewed during the visit.

# Re-Identification System for reactive personalized audio-guide



**Giuseppe Becchi, Andrea Ferracani, Filippo Principi, Alberto Del Bimbo**

**An AI Powered Re-Identification System For Real-Time  
Contextual Multimedia Applications**

**ACM Multimedia 2022 - DEMO Session**

# What's in the reidentification system ?

- face detection
  - Specific versions of object detection networks are used for this task
  - A specialized network provides information about gaze (in terms of face orientation), to estimate the interest
- face recognition and matching
  - This is a variation of image retrieval applied to descriptors of networks trained for face recognition
- face age estimation & gender detection
  - Networks for attribute recognition are used
  - Training these networks is tricky, plenty of issues with dataset biases...
- emotion recognition
  - Training these networks is tricky, plenty of issues with dataset biases...