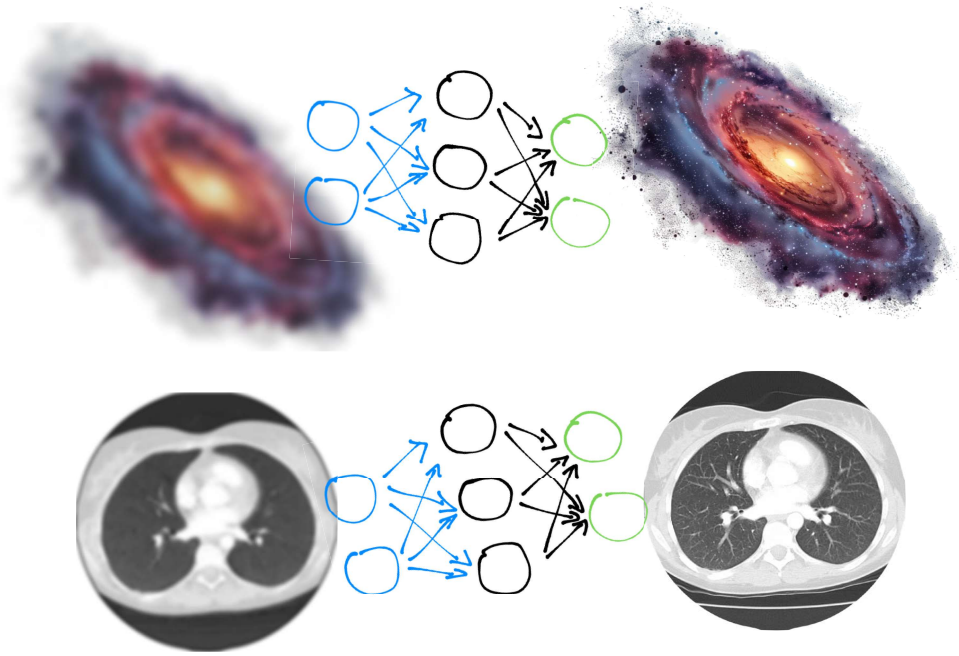
Ricostruzione di immagini con intelligenza artificiale (pillole di visione computazionale)

Michele Ginolfi — UniFi

Cover seria…



Cover anche lei seria!

UMANO chiede:

hey amico artificiale intelligente, cosa vedi qui? Un muffin o un chihuahua?

RETE NEURALE risponde:

…

Corso LM: **"ANALISI DELLE IMMAGINI E COMPUTER VISION, CON APPLICAZIONI ALLE SCIENZE FISICHE"**

Anno di immatricolazione 2023/2024 IT EN

[B032475] - DATA SCIENCE PER L'ANALISI DELLE IMMAGINI IN ASTRONOMIA E FISICA MEDICA

Informazioni generali

Corso di studi	DATA SCIENCE, CALCOLO SCIENTIFICO & INTELLIGENZA ARTIFICIALE
Tipo di corso	Laurea Magistrale
Anno di offerta	2024/2025
Anno di corso	2
Tipo Attività Formativa	Affine/Integrativa
Ambito	Attività formative affini o integrative
Crediti	6 CFU
Tipo attività didattica	Lezione
Valutazione	Voto Finale
Periodo didattico	Secondo Semestre (dal 17/02/2025 al 13/06/2025)
Responsabili	GINOLFI MICHELE
Docenti	DORIA SANDRA
Durata	48 ore (48 ore Lezione)

Dipartimento di Fisica e Astronomia UNIFI

@fisica_astro_unifi · 404 iscritti · 30 video

Canale Youtube del Dipartimento di Fisica e Astronomia dell'Università degli Studi di Firen ...altro

Iscritto

In collaborazione con Giovanna Pacini

umAnI | La scoperta della galassia più lontana mai...

Dipartimento di Fisica e Astronomi...

umAnI | Avventure cosmologiche, dall'Antartide all'Atacama -...

Dipartimento di Fisica e Astronomi...

umAnI | Scacchi: valore educativo, evoluzione del gioco,...

Dipartimento di Fisica e Astronomi...

umAnI | Illustrare la scienza - Maria Cristina Fortuna

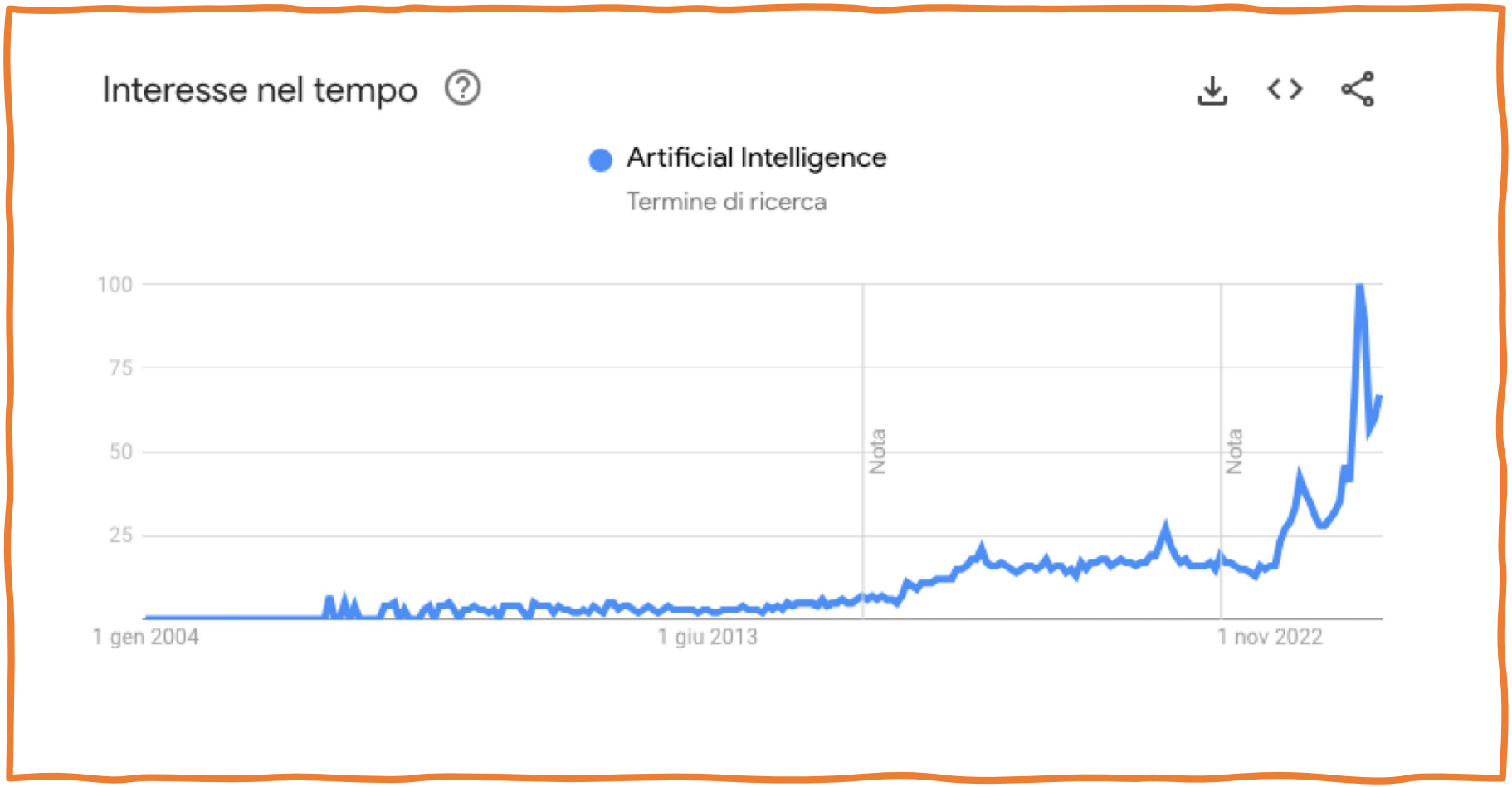
Dipartimento di Fisica e Astronomi...

umAnI | Progettare grandi telescopi: la sfida del progetto...

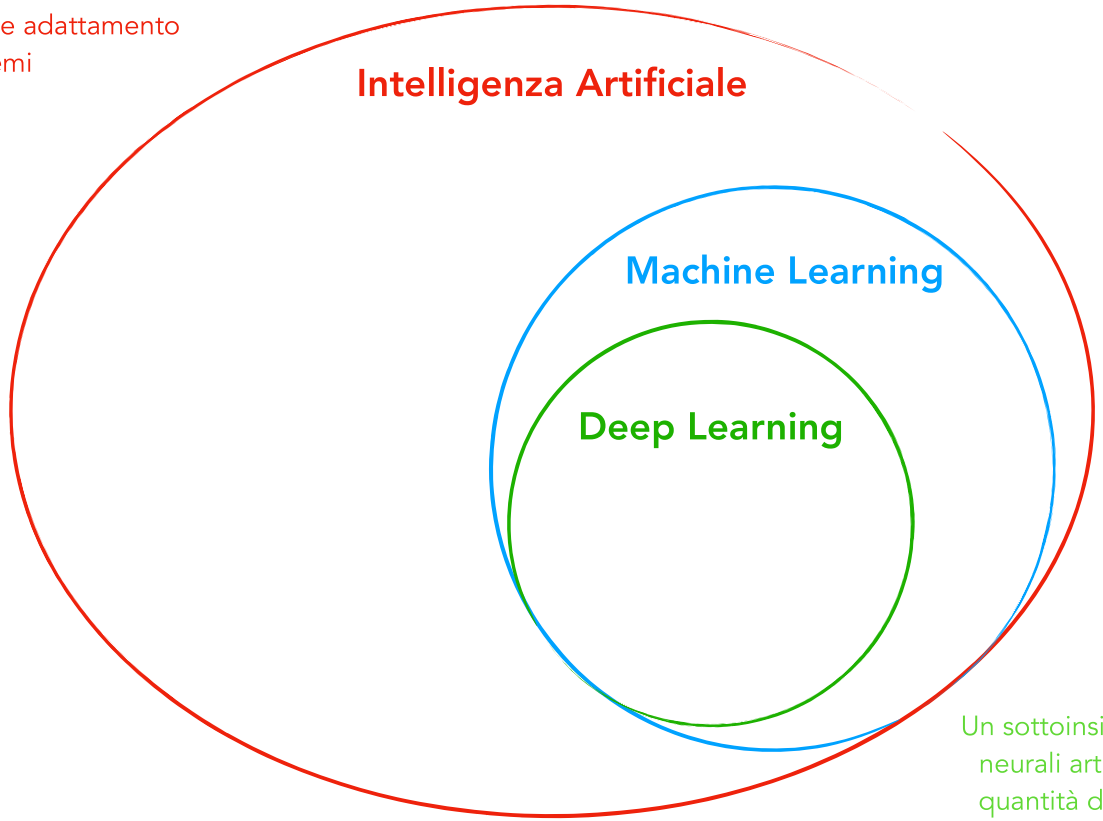
Dipartimento di Fisica e Astronomi...

umAnI | Sonificazione dei dati e multisensorialità - Anita Zanella ...

Dipartimento di Fisica e Astronomi...

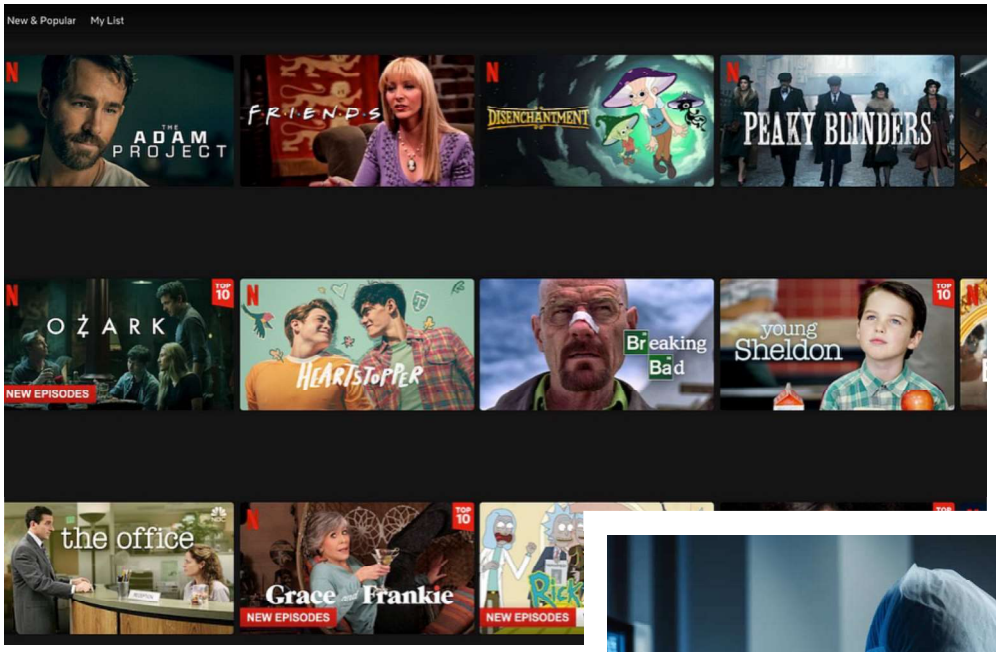

Interesse nel tempo
Artificial Intelligence
Termine di ricerca
100
75
50
25
Nota
Nota
1 gen 2004
1 giu 2013
1 nov 2022

Sistemi digitali che simulano capacità umane di apprendimento, ragionamento e adattamento per risolvere problemi



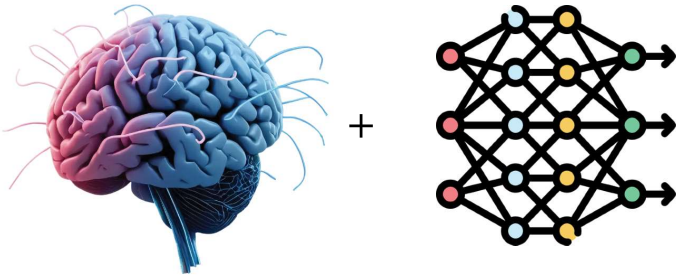
Sistemi che apprendono da dati e migliorano autonomamente le proprie performance su compiti specifici, senza istruzioni esplicite

Un sottoinsieme del machine learning che usa reti neurali artificiali profonde per analizzare grandi quantità di dati e risolvere problemi complessi.


New & Popular
My List
THE ADAM PROJECT
FRIENDS
DISENCHANTMENT
PEAKY BLINDERS
OZARK
NEW EPISODES
HEARTSTOPPER
Breaking Bad
young Sheldon
the office
Grace and Frankie
NEW EPISODES
NEW EPISODES

0218
80213648

Il deep learning come acceleratore di conoscenza scientifica

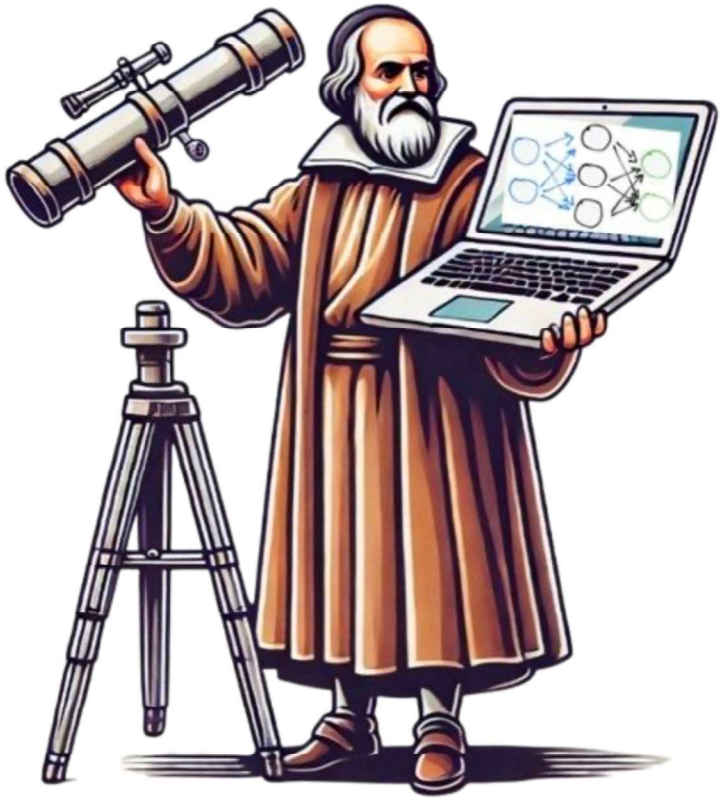


- Ottimizzazione e velocizzazione di processi *standard*
- Rivelazione di schemi e relazioni nascoste nei dati, che possono suggerire nuove teorie

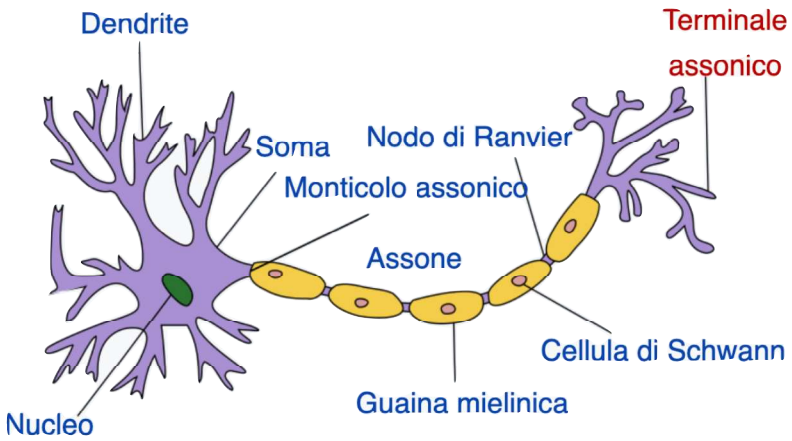
Osservazione

Ipotesi

Esperimento



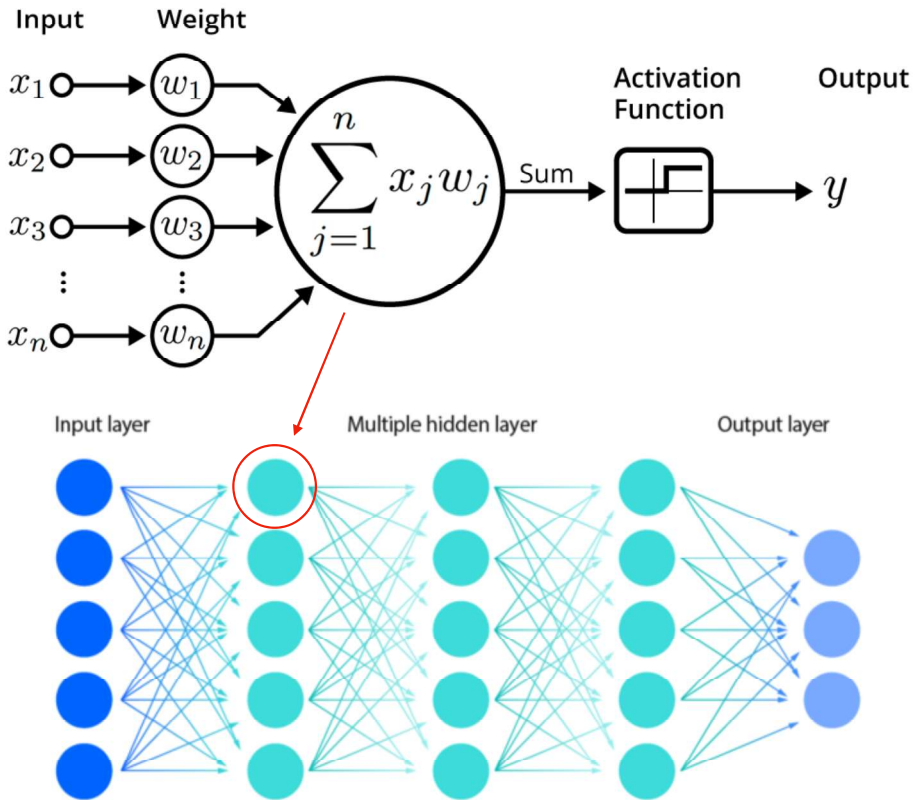
Neurone biologico



I dendriti del neurone ricevono segnali chimici da altri neuroni tramite le sinapsi. Il neurone li elabora e, se il segnale risultante supera una soglia, invia un impulso lungo l'assone verso altre cellule.

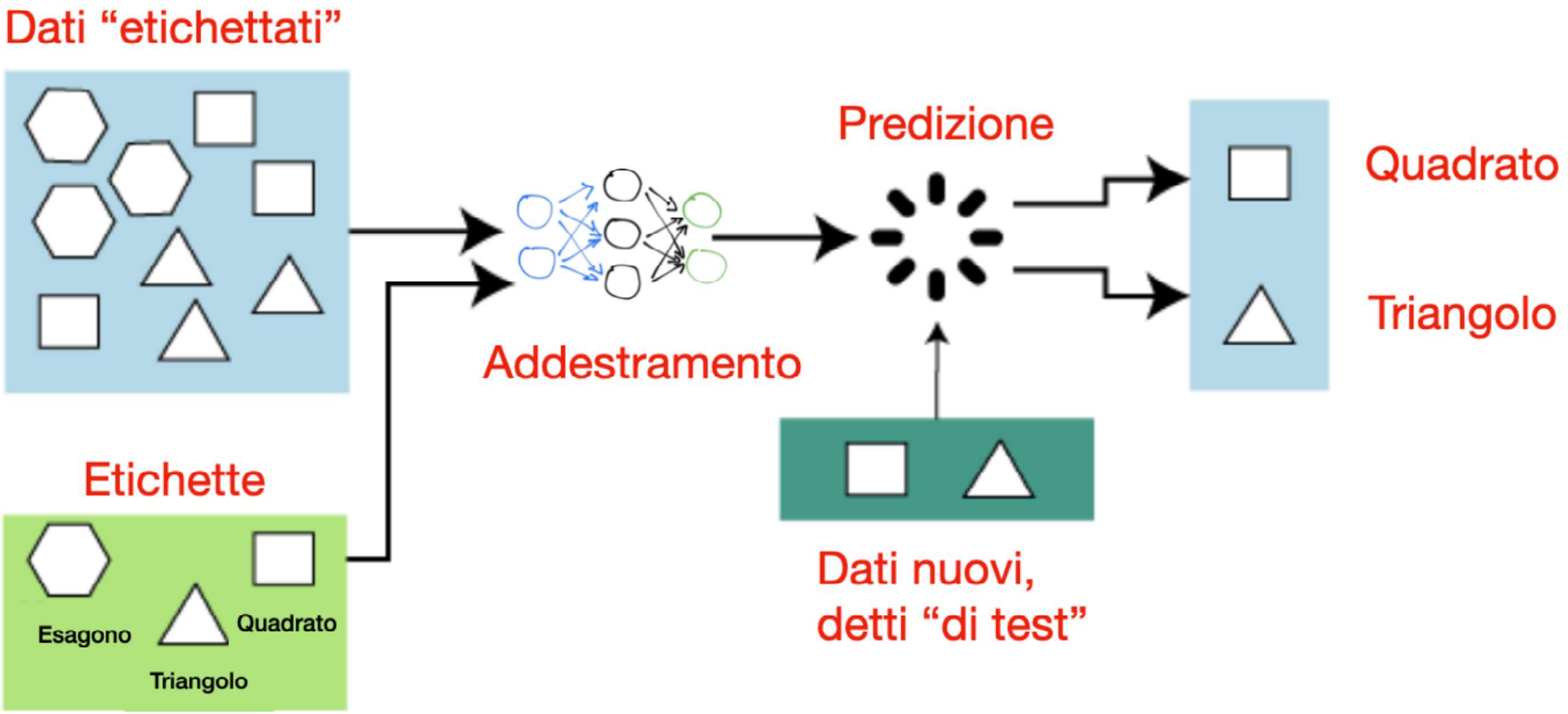
@wikipedia

Neurone artificiale



@Xiang, CONF-CIAP 2022 @IBM

Contesto supervisionato: il *cervello digitale* impara da esempi



Ma come funziona l’addestramento?

Alcuni esempi pratici… “della vita reale”

Predizione del prezzo delle case: dati come dimensione (metri quadrati), numero di stanze, anno di costruzione e posizione possono essere usati per allenare una rete neurale che stimi il prezzo di una casa.

Diagnosi di malattie: per un determinato tipo di malattia, gli input possono essere valori di test clinici (colesterolo, pressione sanguigna, glicemia etc) e caratteristiche demografiche.

Analisi del rischio di credito: usando dati come reddito, età, storico creditizio, e numero di prestiti in corso, si può prevedere la probabilità di default di un cliente. Gli diamo il mutuo oppure no?

E tanti tanti alti esempi…

NB: in questi cari, gli input, anche se correlati, possono essere trattate come caratteristiche indipendenti. Sarà chiaro il perché dopo!

Predizione del prezzo delle case

Feature 1: dimensione

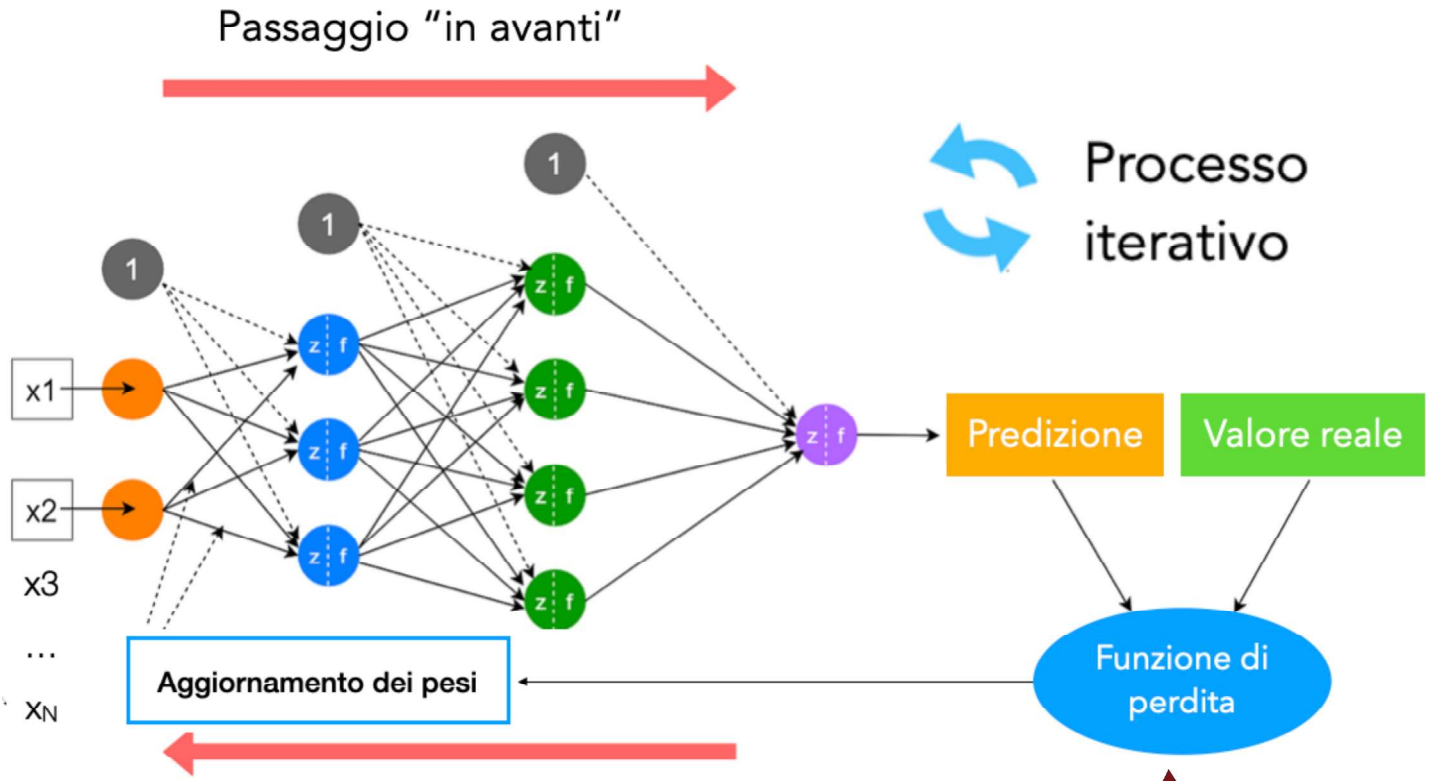
Feature 2: N stanze

Feature 3: Anno di costruzione

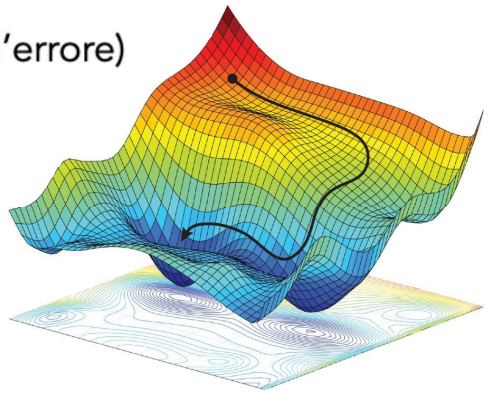
Feature 4: posizione

…

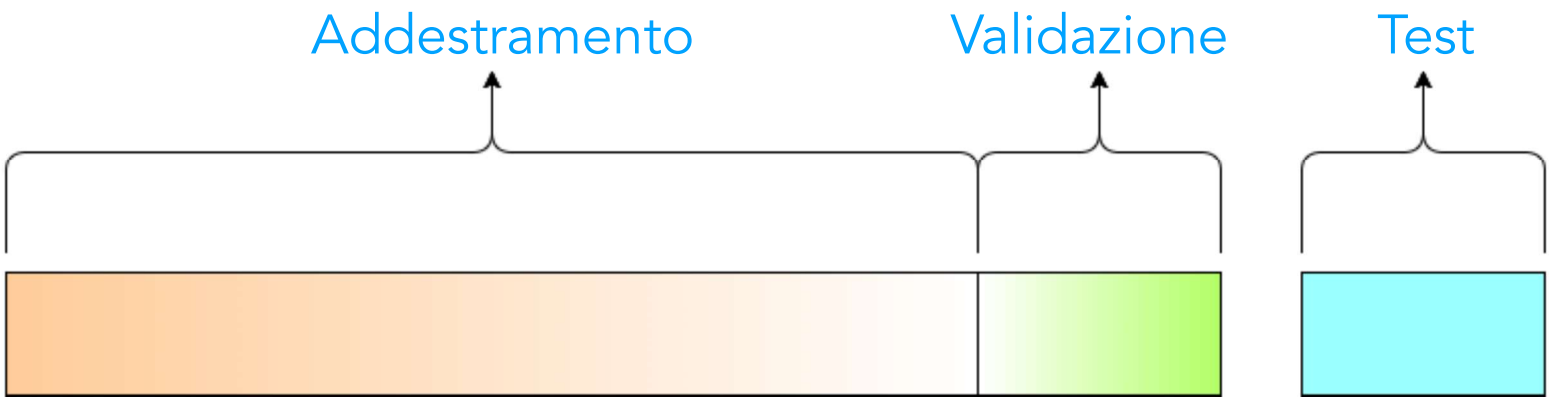
Feature N: …

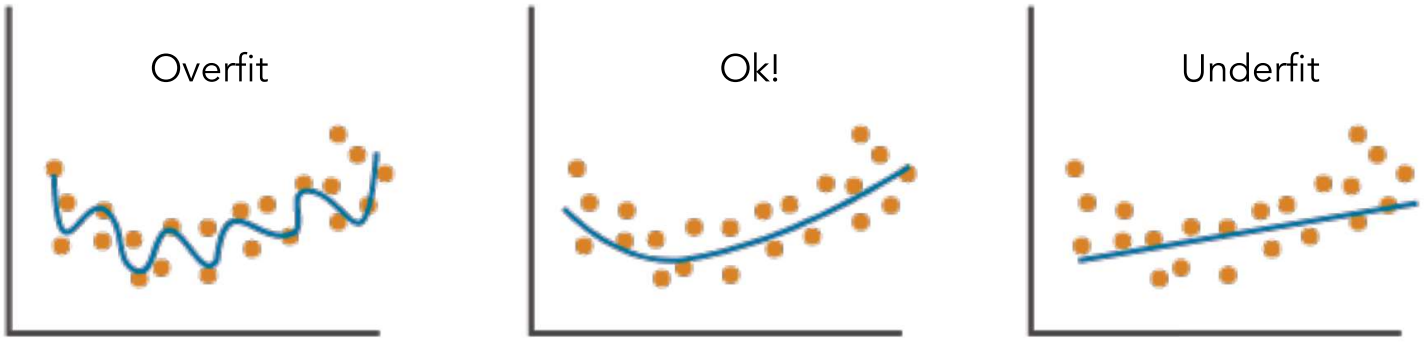


Se il modello e' addestrato bene, riuscirà a prevedere *bene* i prezzi di nuove case!



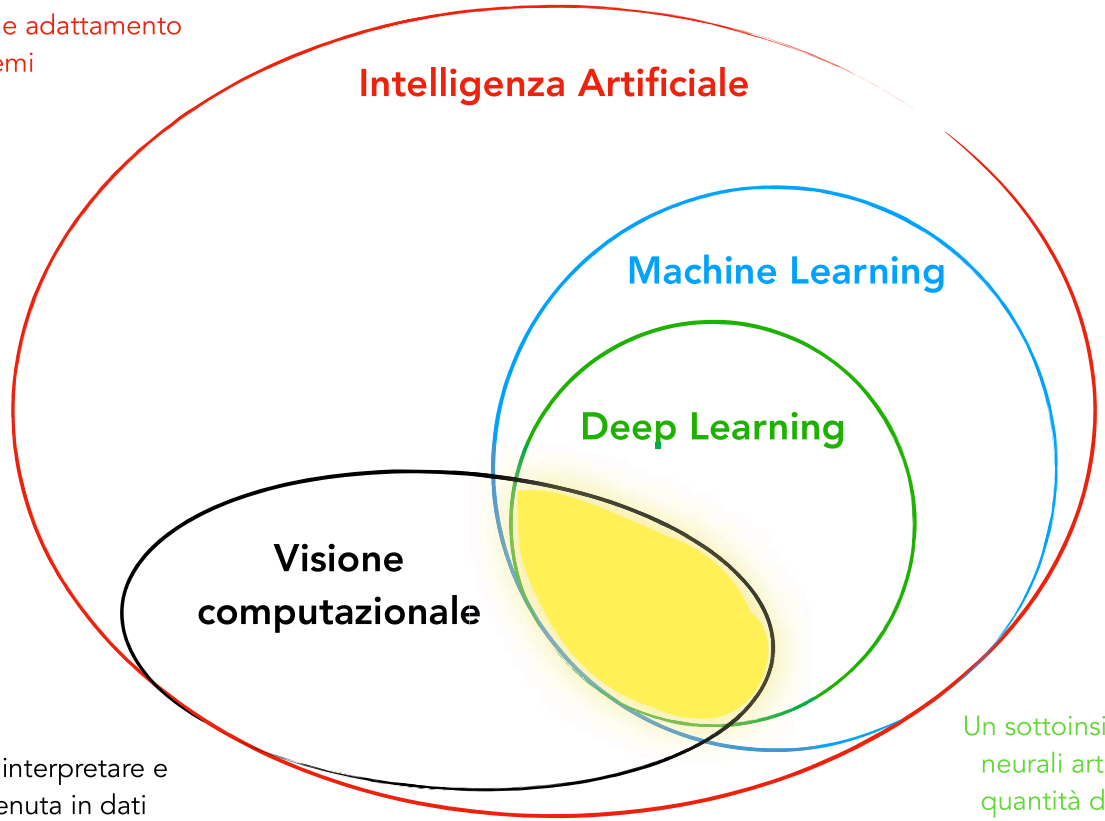
Ma come scelgo la rete neurale giusta?





Posso fare il “tuning” degli iper-parametri, aggiungere regolarizzatori, raccogliere o simulare nuovi dati etc..

Sistemi digitali che simulano capacità umane di apprendimento, ragionamento e adattamento per risolvere problemi



Sistemi che apprendono da dati e migliorano autonomamente le proprie performance su compiti specifici, senza istruzioni esplicite

Sistemi digitali capaci di elaborare, interpretare e comprendere l'informazione contenuta in dati visual, ed effettuare dei compiti di conseguenza

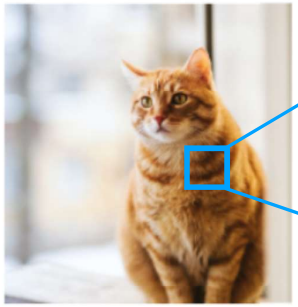
Un sottoinsieme del machine learning che usa reti neurali artificiali profonde per analizzare grandi quantità di dati e risolvere problemi complessi.

Cosa vediamo noi umani

Se il gatto si sposta un po', o se cambia la luce, o se ci avviciniamo / allontaniamo, riusciamo a riconoscerlo ancora senza problemi.
Questo perche' il nostro cervello ha imparato a collegare le proprietà dell'immagine a strutture semantiche.

Cosa vede un computer

La rappresentazione cambia drasticamente (in modo non intuibile) se l'immagine e' un po' diversa!
Dobbiamo convertire questi numeri in un significato semantico, che possa servire alla rete neurale per imparare.

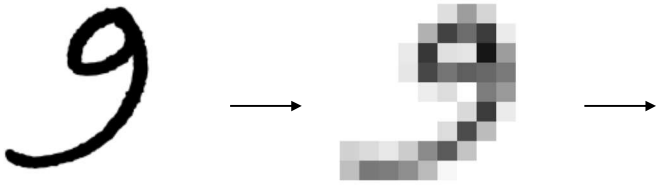


```
[[105 112 108 111 104  99 106  99  96 103 112 119 104  97  93  87]
 [ 91  98 102 106 104  79  98 103  99 105 123 136 110 105  94  85]
 [ 76  85  90 105 128 105  87  96  95  99 115 112 106 103  99  85]
 [ 99  81  81  93 120 131 127 100  95  98 102  99  96  93 101  94]
 [106  91  61  64  69  91  88  85 101 107 109  98  75  84  96  95]
 [114 108  85  55  55  69  64  54  64  87 112 129  98  74  84  91]
 [133 137 147 103  65  81  80  65  52  54  74  84 102  93  85  82]
 [128 137 144 140 109  95  86  70  62  65  63  63  60  73  86 101]
 [125 133 148 137 119 121 117  94  65  79  80  65  54  64  72  98]
 [127 125 131 147 133 127 126 131 111  96  89  75  61  64  72  84]
 [115 114 109 123 150 148 131 118 113 109 100  92  74  65  72  78]
 [ 89  93  90  97 108 147 131 118 113 114 113 109 106  95  77  80]
 [ 63  77  86  81  77  79 102 123 117 115 117 125 125 130 115  87]
 [ 62  65  82  89  78  71  80 101 124 126 119 101 107 114 131 119]
 [ 63  65  75  88  89  71  62  81 120 138 135 105  81  98 110 118]
 [ 87  65  71  87 106  95  69  45  76 130 126 107  92  94 105 112]
 [118  97  82  86 117 123 116  66  41  51  95  93  89  95 102 107]
 [164 145 112  80  82 120 124 104  76  48  45  66  88 101 102 109]
 [157 170 157 120  93  86 114 132 112  97  69  55  70  82  99  94]
 [130 128 134 161 139 100 109 118 121 134 114  87  65  53  69  86]
 [128 112  96 117 150 144 120 115 104 107 102  93  87  81  72  79]
 [123 107  96  86  83 112 153 149 122 109 104  75  80 107 112  99]
 [122 121 102  80  82  86  94 117 145 148 153 102  58  78  92 107]
 [122 164 148 103  71  56  78  83  93 103 119 139 102  61  69  84]]
```

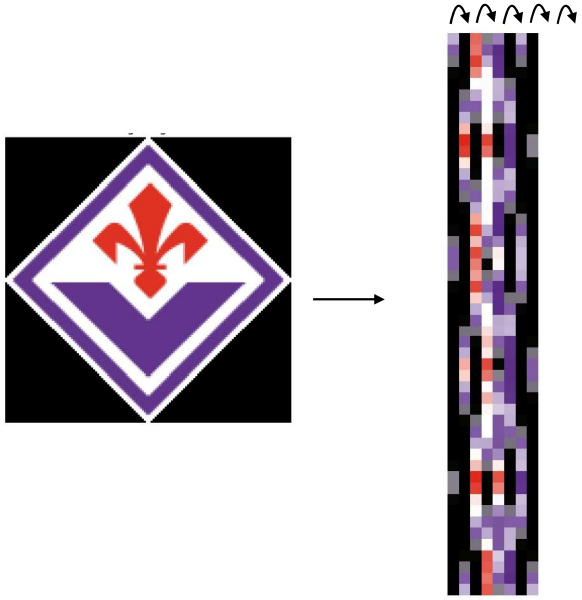
Un'immagine e' un tensore 3D di numeri (3 canali RGB)

Come insegniamo ad una rete neurale a riconoscere un'immagine?

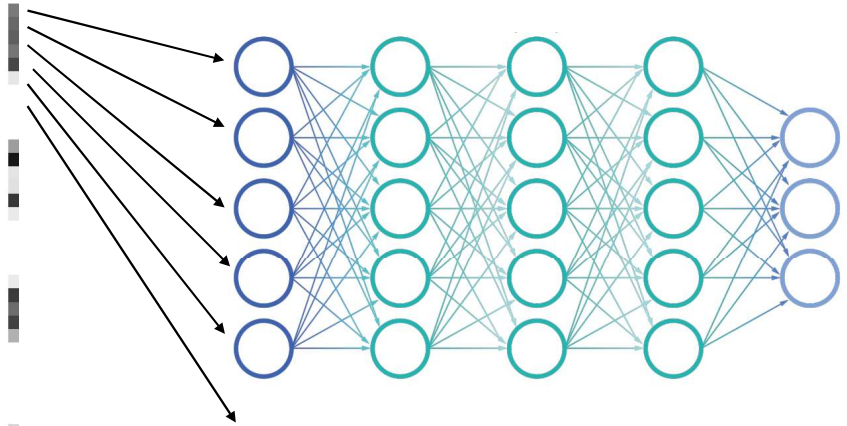
Come insegniamo ad una rete neurale a riconoscere un'immagine?



10 x 10 pixel



100 features



…

…

…

- Difficilmente gestibile per grandi immagini (1e6 features per 1000pix images)

- Sto considerando che tutti i pixel sono indipendenti. Ma la realtà non e' cosi..

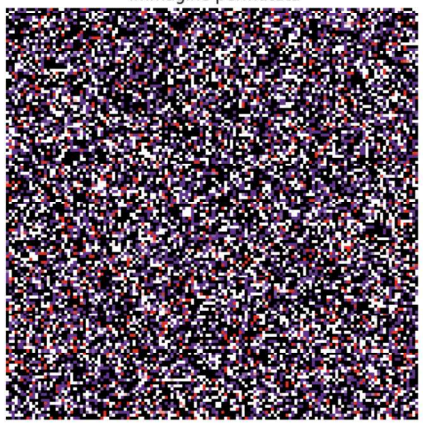
Come insegniamo ad una rete neurale a riconoscere un’immagine?

Una permutazione random fa perdere significato semantico.
Il significato visivo di un'immagine non dipende dai singoli pixel isolati, ma piuttosto dalla loro disposizione spaziale.
L’informazione da “valorizzare” e’ la relazione spaziale tra i pixel.

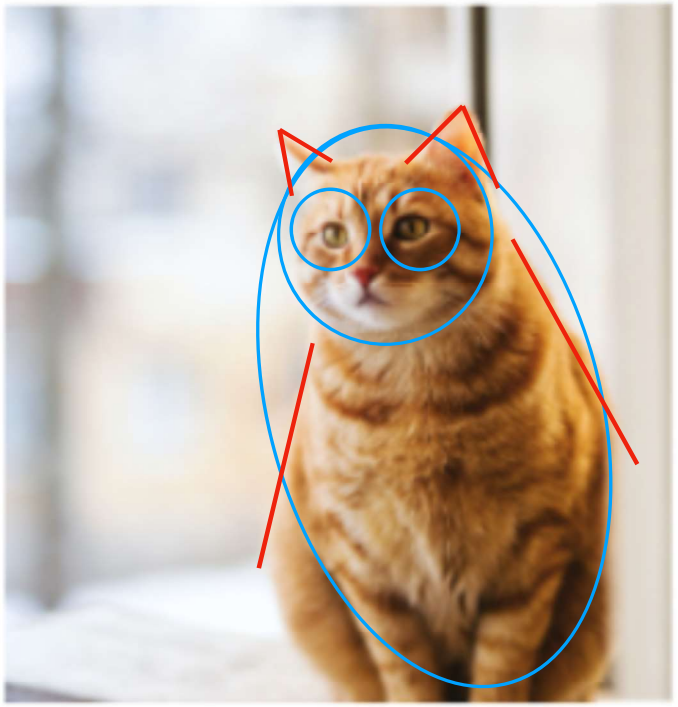
Image originale

Bassa entropia.
Facile da “spiegare”,
Facile da comprimere (~70 KB),
Facile ricorrere a strutture semantiche per interpretarla.

Immagine permutata



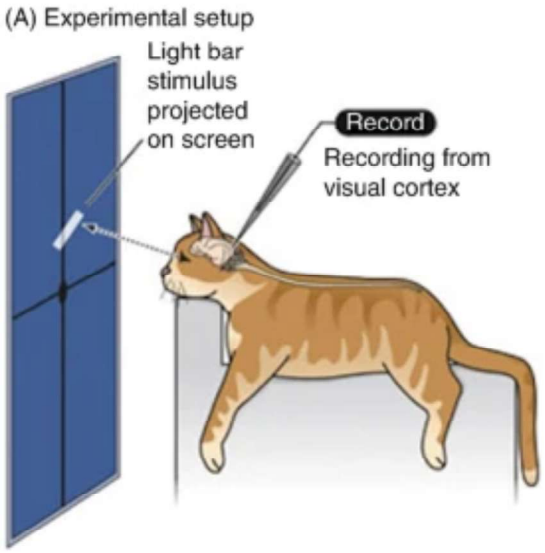
Altissima entropia.
Difficile da “spiegare”,
Difficile da comprimere (~170 KB),
Difficile ricorrere a strutture semantiche per interpretarla.

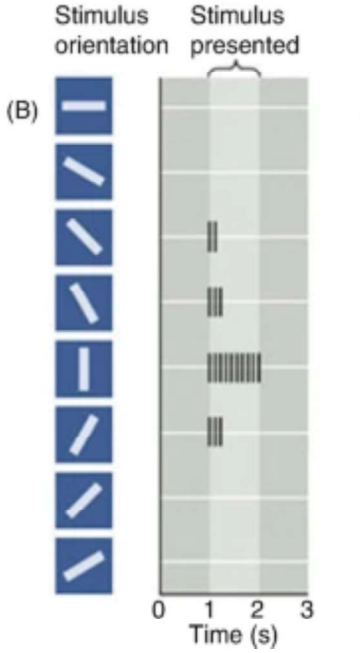


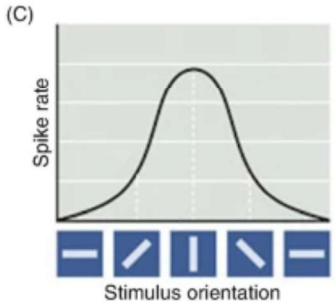
Le immagini naturali sono composizionali, possiamo farci aiutare da questa proprietà e
sfruttare l’informazione spaziale!
Dobbiamo insegnare ad una rete neurale a farlo!

Analogia con il cervello naturale: gli esperimenti di Hubel & Wiesel (’60-’70)

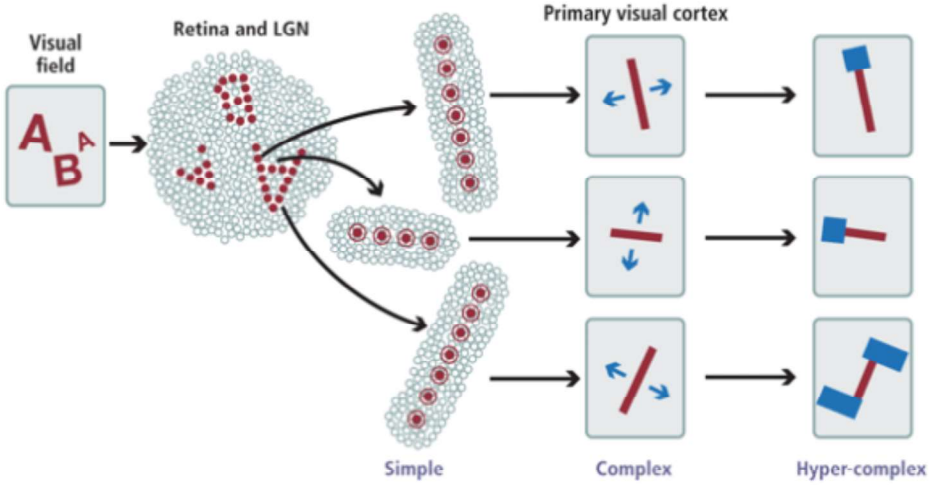
@“Feature Detection”, Springer







@uniderstandingcontext.com

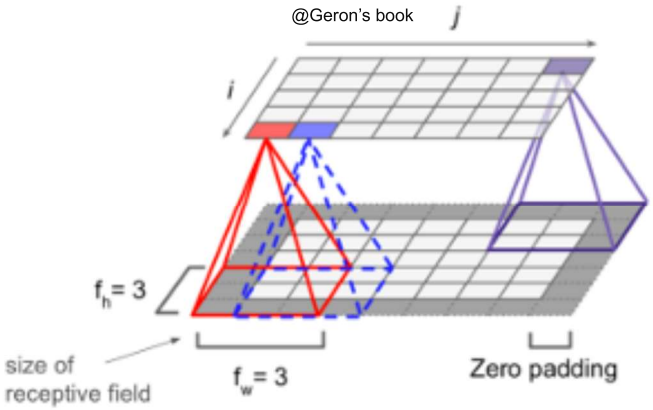


Rispondono all’orientamento di linee semplici

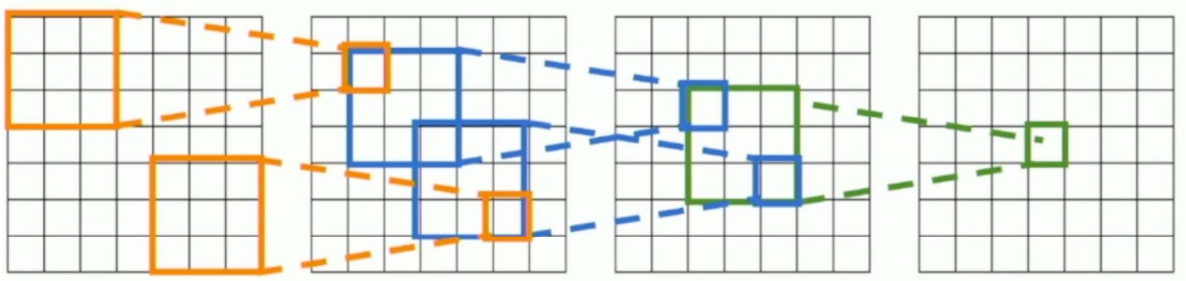
Rispondono all'orientamento e al movimento della luce

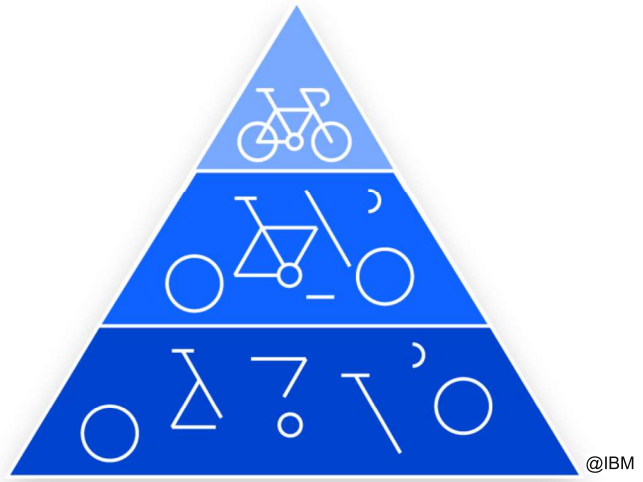
Rispondono a caratteristiche come i bordi e gli angoli, dove il movimento ha un punto di arresto o un cambiamento netto, come il limite di un oggetto

Reti neurali convoluzionali (CNN)



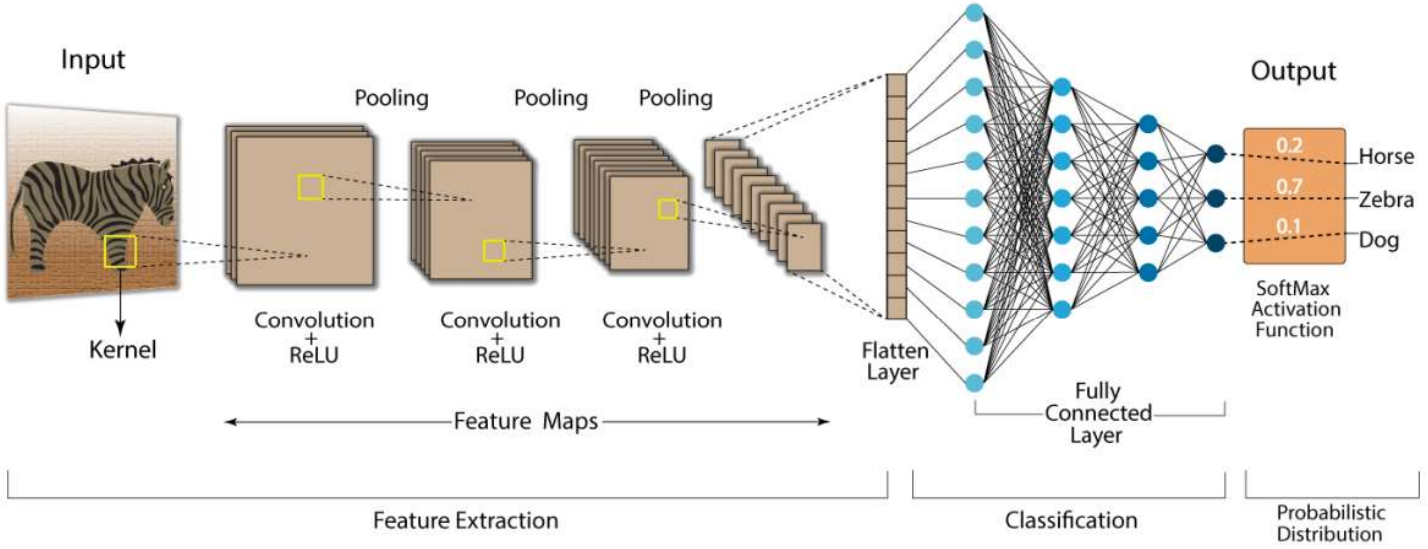
@Stanford University School of Engineering

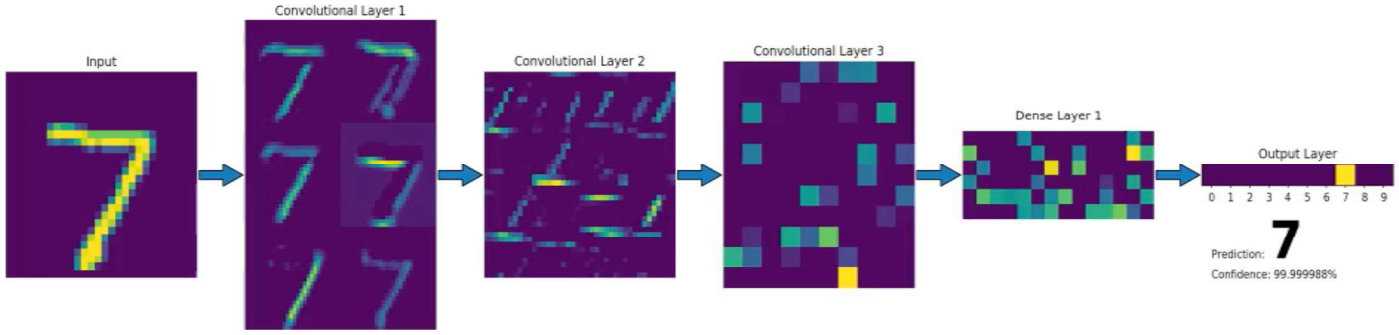




I filtri vengono ottimizzati durante l'addestramento

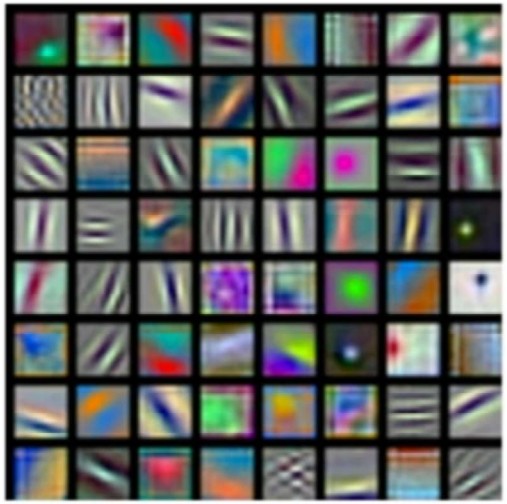
Reti neurali convoluzionali (CNN)





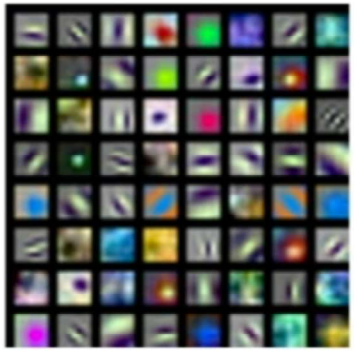
Guardiamo i filtri del primo strato convoluzionale

@Stanford University School of Engineering

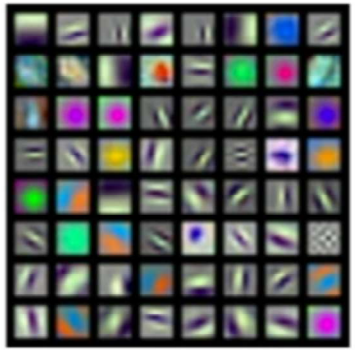


AlexNet:

ResNet-18:
64 x 3 x 7 x 7



ResNet-101:
64 x 3 x 7 x 7

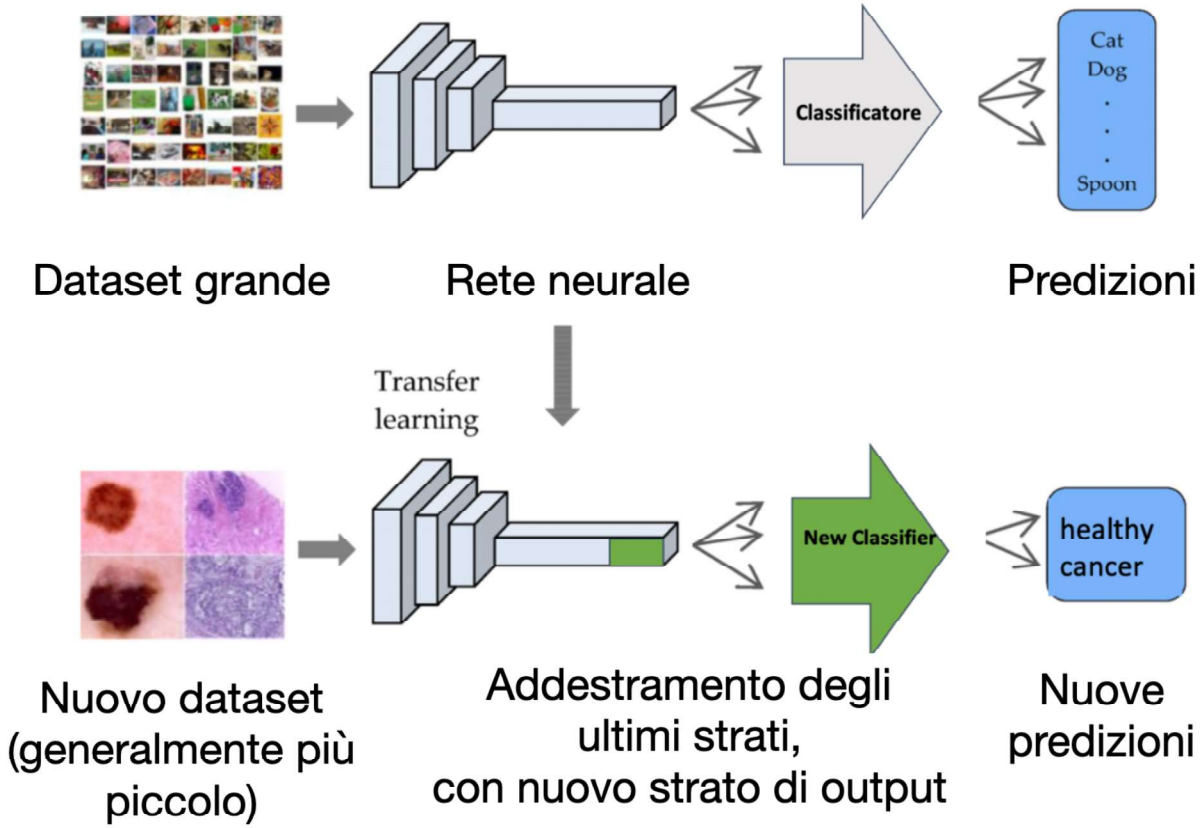


DenseNet-121:
64 x 3 x 7 x 7

Krizhevsky, "One weird trick for parallelizing convolutional neural networks", arXiv 2014
He et al, "Deep Residual Learning for Image Recognition", CVPR 2016
Huang et al, "Densely Connected Convolutional Networks", CVPR 2017

Le caratteristiche di basso livello come bordi, angoli, cerchi e pattern semplici sembrano essere abbastanza universali indipendentemente dalle specifiche architetture. Ha senso… e' conseguenza della natura composizione dei dati.
Alcune caratteristiche estratte, anche negli strati successivi, continuano ad essere generali.
Come possiamo sfruttare questo a nostro favore?

Transfer learning (riutilizzare la conoscenza acquisita)



Analogia *naturale*

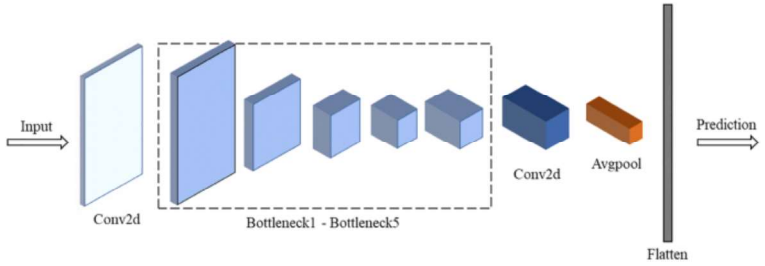
Diversi studi comportamentali e di risonanza magnetica funzionale suggeriscono che alcune cose, come la capacita di identificare facce, sia *hardwired* (non si apprende, ma la si ha dalla nascita).

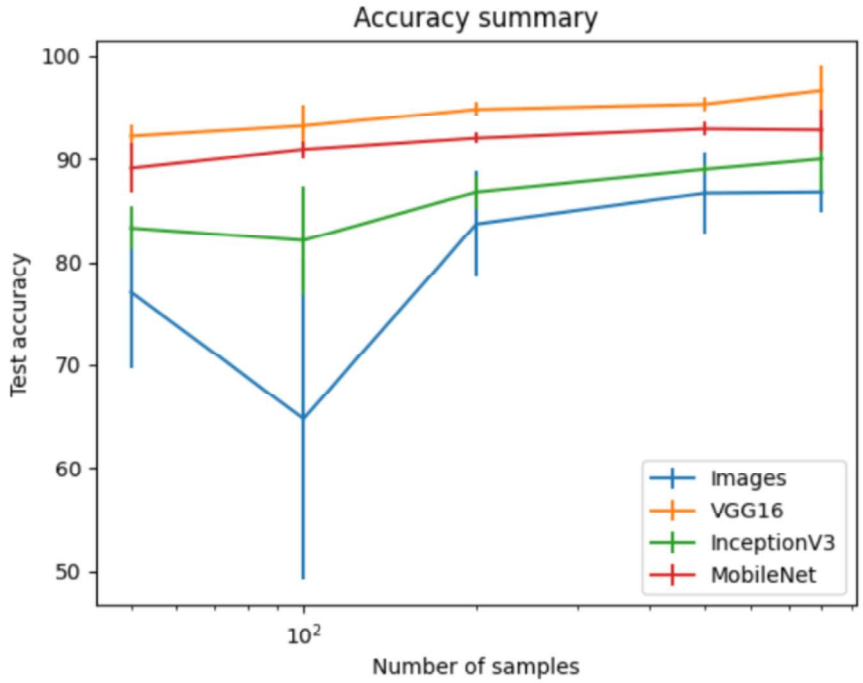
Classificare galassie partendo da immagini *quotidiane*

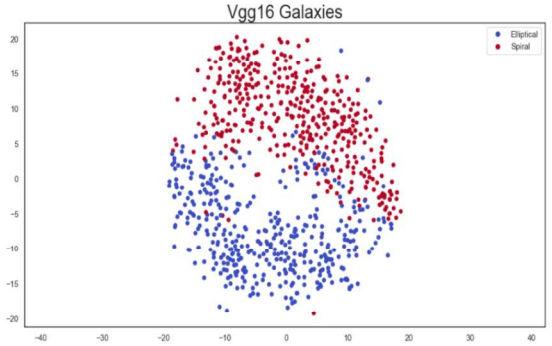
@Michael Cavicchioli, B.Sc, UniFi

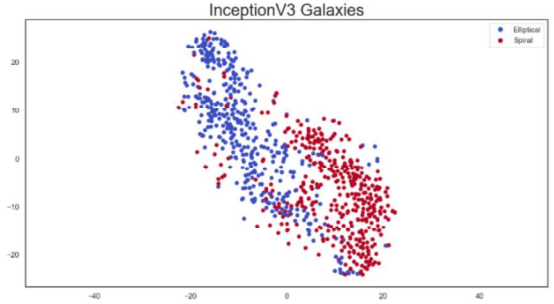
400 ellittiche

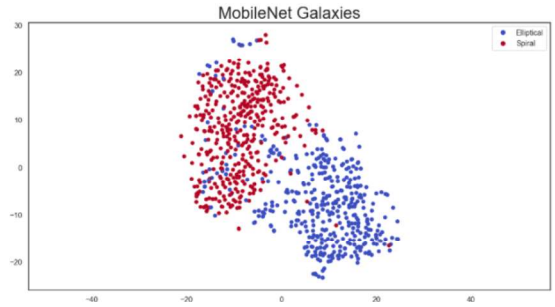
400 spirali





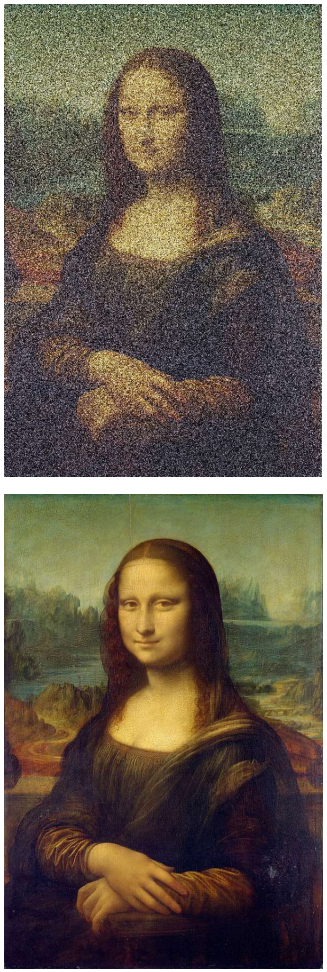




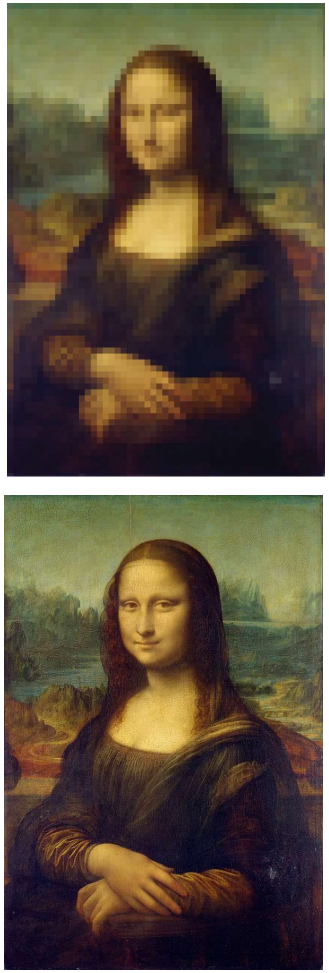


Alcune delle altre tipiche tasks di visione artificiale

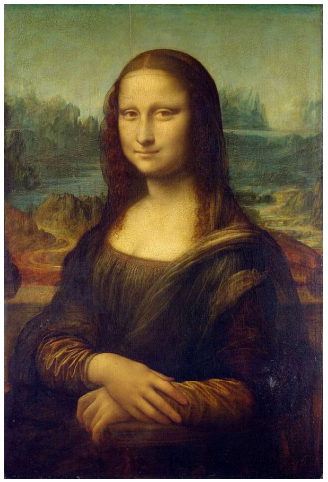
Rimozione del rumore

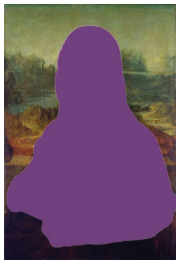


Aumento della risoluzione

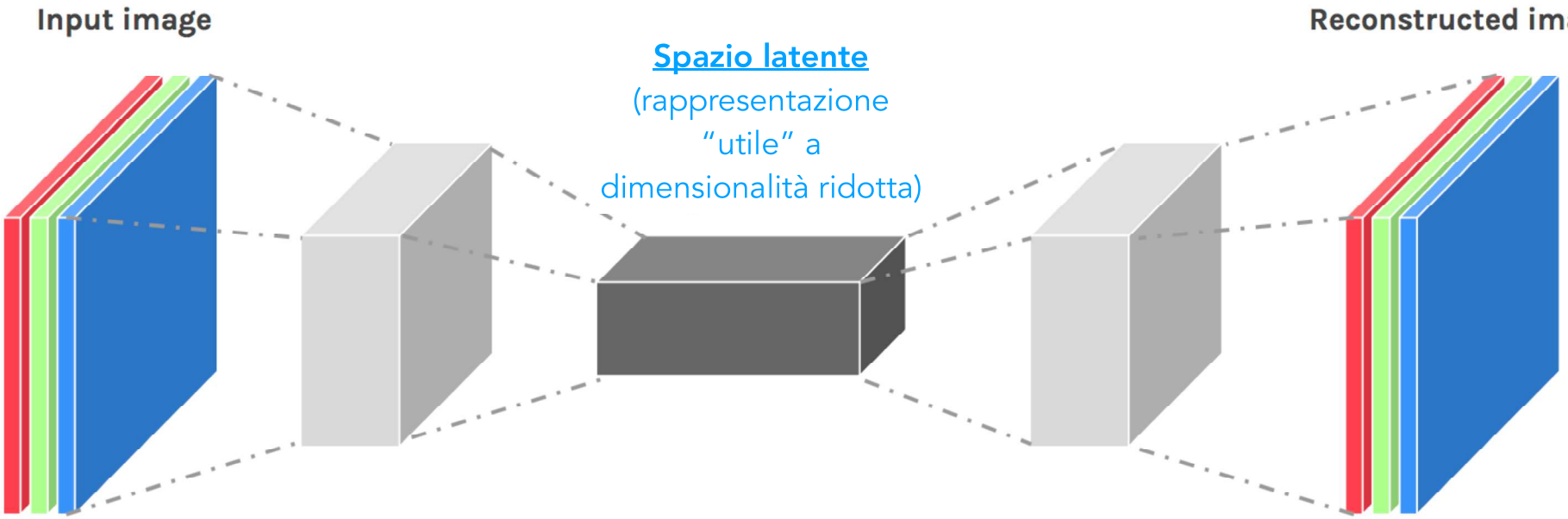


Segmentazione





Reti neurali convoluzionali di tipo encoder-decoder



Spazio latente
(rappresentazione "utile" a dimensionalità ridotta)

Encoder: codifica l'informazione rilevante per l'obiettivo, comprimendola

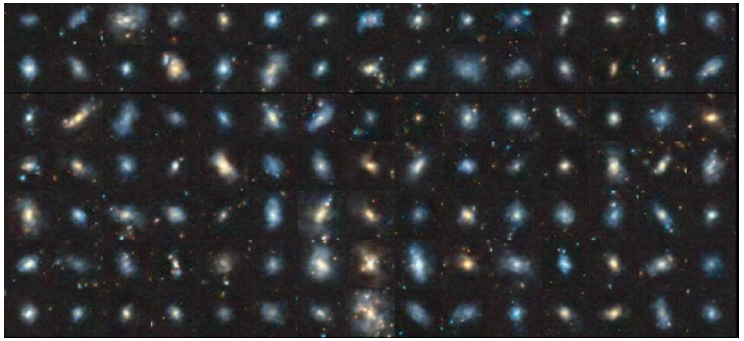
Decoder: decodifica l'informazione compressa, trasformandola nell'output desiderato. E' un processo di "generazione". Infatti queste task possono essere affrontate da modelli generativi (vedrete poi)

Modelli generativi — reti avversarie generative (GANs)

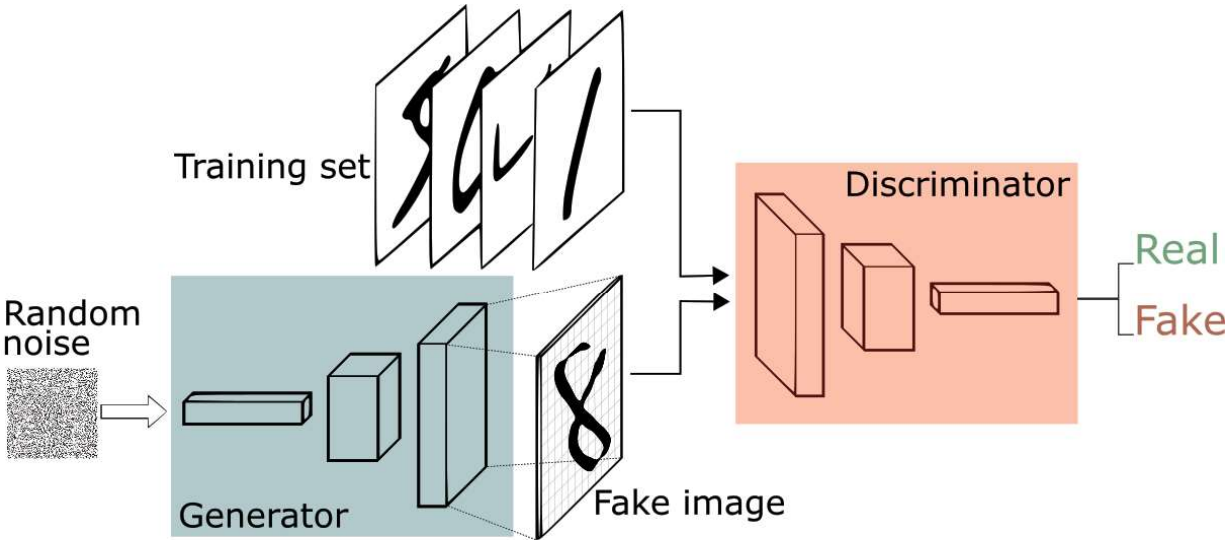
Task generativa

[Vedremo più dettagli ed applicazioni nella seconda parte]

https://this-person-does-not-exist.com/en

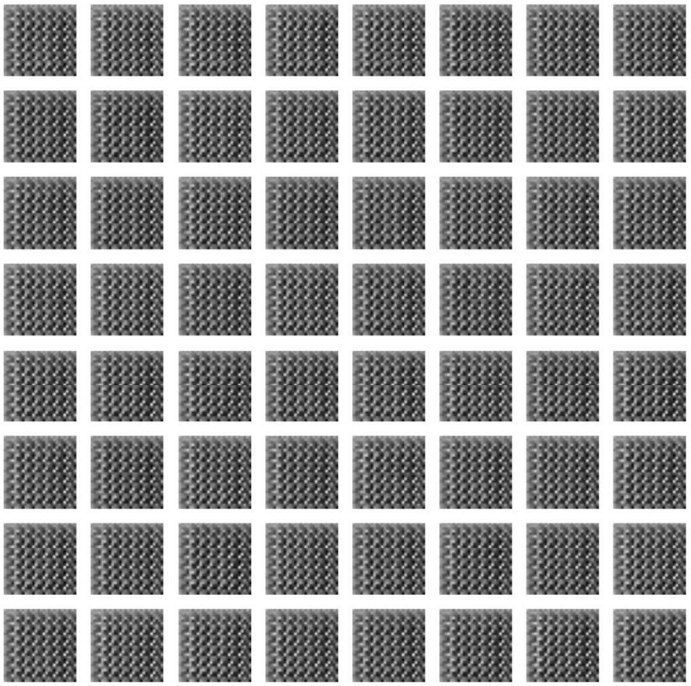


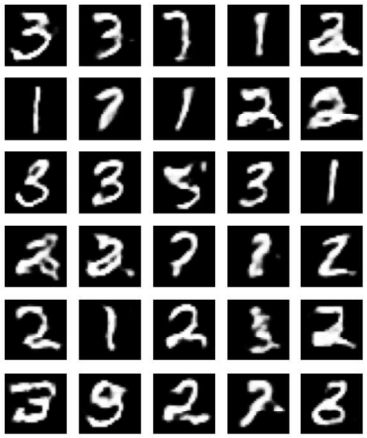
These galaxies do not exist! :-)

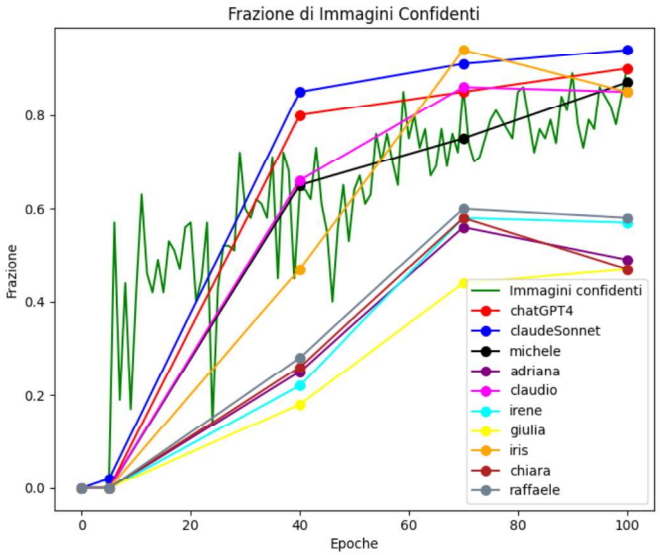


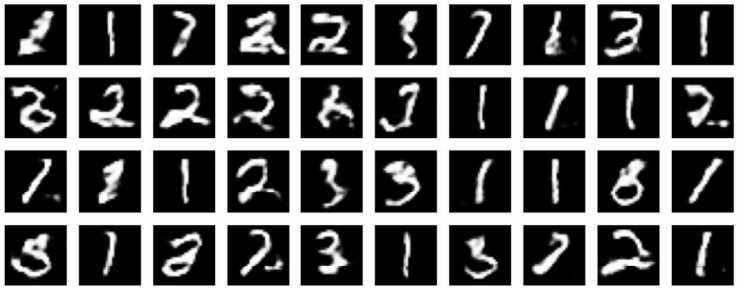
Modelli generativi — reti avversarie generative (GANs)

@Claudio Soricaro, B.Sc, UniFi









Altri modelli generativi

Lo *zoo* di modelli generativi (alcuni)

Generative Adversarial Networks (GANs):

- **Vanilla GAN**: Il modello base, in cui una rete generativa compete contro una rete discriminativa.
- **Deep Convolutional GAN (DCGAN)**: Una variante che usa reti convoluzionali profonde, migliorando la qualità delle immagini generate.
- **Conditional GAN (cGAN)**: Permette di generare immagini condizionate da etichette specifiche (es. classe dell'immagine).
- **StyleGAN** e **StyleGAN2**: Permettono un controllo più dettagliato sullo stile e l'aspetto delle immagini generate.

Variational Autoencoders (VAEs):

- **Conditional VAE (CVAE)**: Simile al cGAN, permette di generare immagini condizionate su etichette o altre informazioni.
- **β-VAE**: Una variante che aggiunge un termine di peso al termine KL, migliorando la qualità delle rappresentazioni latenti e permettendo un controllo maggiore sui fattori di variazione.

Diffusion Models:

- **Denoising Diffusion Probabilistic Models (DDPMs)**: Un modello che apprende a generare immagini denoising iterativo di un rumore gaussiano, ottenendo risultati di alta qualità.
- **Latent Diffusion Models (LDMs)**: Una versione più efficiente dei modelli di diffusione, che opera nello spazio latente per ridurre la complessità computazionale, usata da modelli come Stable Diffusion.
- **Score-Based Generative Models**: Basati sulla stima delle funzioni di score, sono simili ai modelli di diffusione ma applicano metodi stocastici per generare immagini partendo da rumore.

Autoregressive Models:

- **PixelRNN** e **PixelCNN**: Modelli autoregressivi che generano immagini pixel per pixel. PixelCNN è più efficiente grazie alla convoluzione.
- **PixelCNN++**: Un'evoluzione del PixelCNN, migliora la qualità dell'immagine e riduce l'effetto di "griglia" nei pixel.
- **Image Transformer**: Usa il meccanismo di attenzione per generare immagini pixel per pixel, mostrando risultati promettenti per immagini ad alta risoluzione.

Flow-Based Models:

- **RealNVP**: Usa trasformazioni invertibili per modellare distribuzioni di immagini complesse, permettendo un training efficiente.
- **Glow**: Un flow-based model che migliora la qualità e interpretabilità delle immagini generate, con un'architettura più semplice rispetto a RealNVP.

Transformer-Based Models:

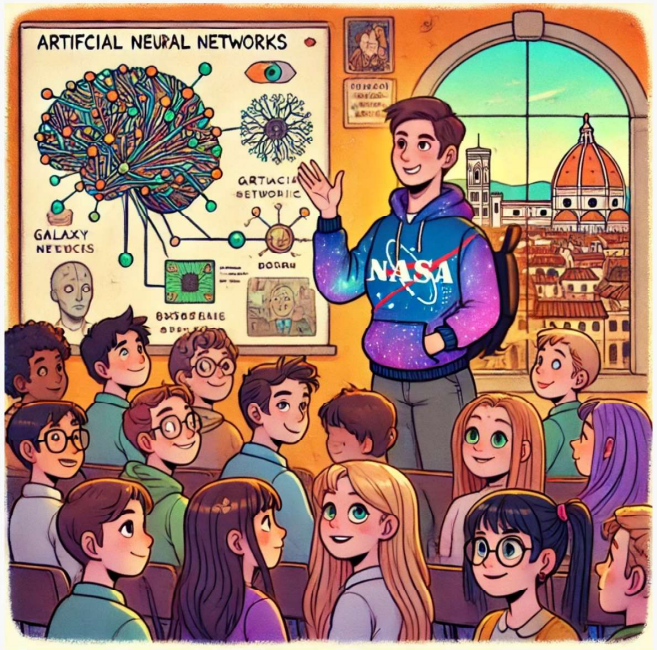
- **Vision Transformer (ViT) per generazione**: Una versione modificata del Transformer per generare immagini trattandole come sequenze.
- **DALL-E** e **DALL-E 2**: Modelli Transformer capaci di generare immagini da descrizioni testuali, aprendosi al text-to-image.
- **Imagen** di Google: Un modello basato su Transformer e Diffusion per generare immagini fotorealistiche a partire da testo.

Energy-Based Models (EBMs):

- **Deep Energy-Based Models**: Utilizzano una funzione di energia per modellare la distribuzione delle immagini e generare dati minimizzando l'energia del sistema.

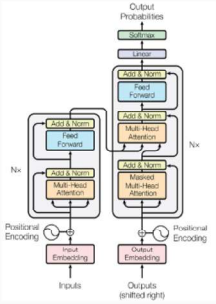
Modelli multi-modali

Puoi creare un'immagine di un giovane astrofisico che sta spiegando come funzionano le reti neurali artificiali ad pubblico di ragazzi e ragazze? Puoi fare l'immagine con lo stile di un cartone animato, ma anche con uno stile un po' anni '90. Dal momento che la lezione avviene a Firenze, puoi inserire anche un tocco di vibes di Firenze? Grazie :)

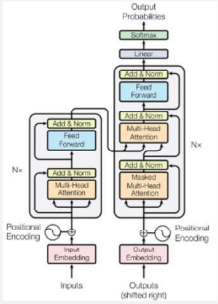


E a proposito di modelli muti-modali…

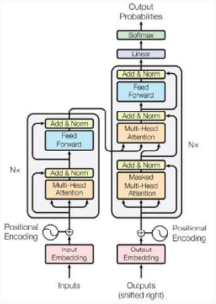
Computer Vision



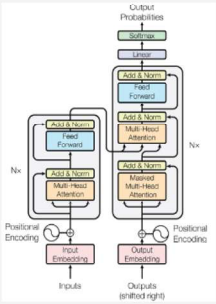
Natural Lang. Proc.



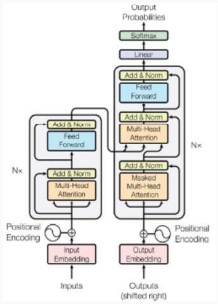
Reinf. Learning



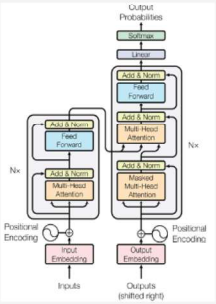
Speech



Translation

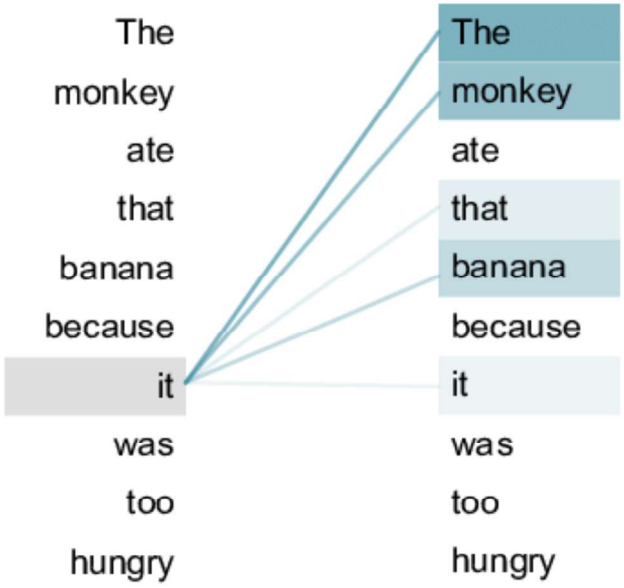


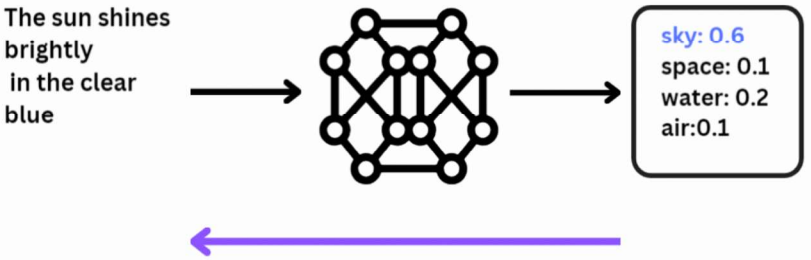
Graphs/Science



ChatGPT Claude Llama 3 From Meta MISTRAL AI_ Gemini

Transformers (meccanismo di attenzione) + Training auto-regressivo (o *mascherato*)



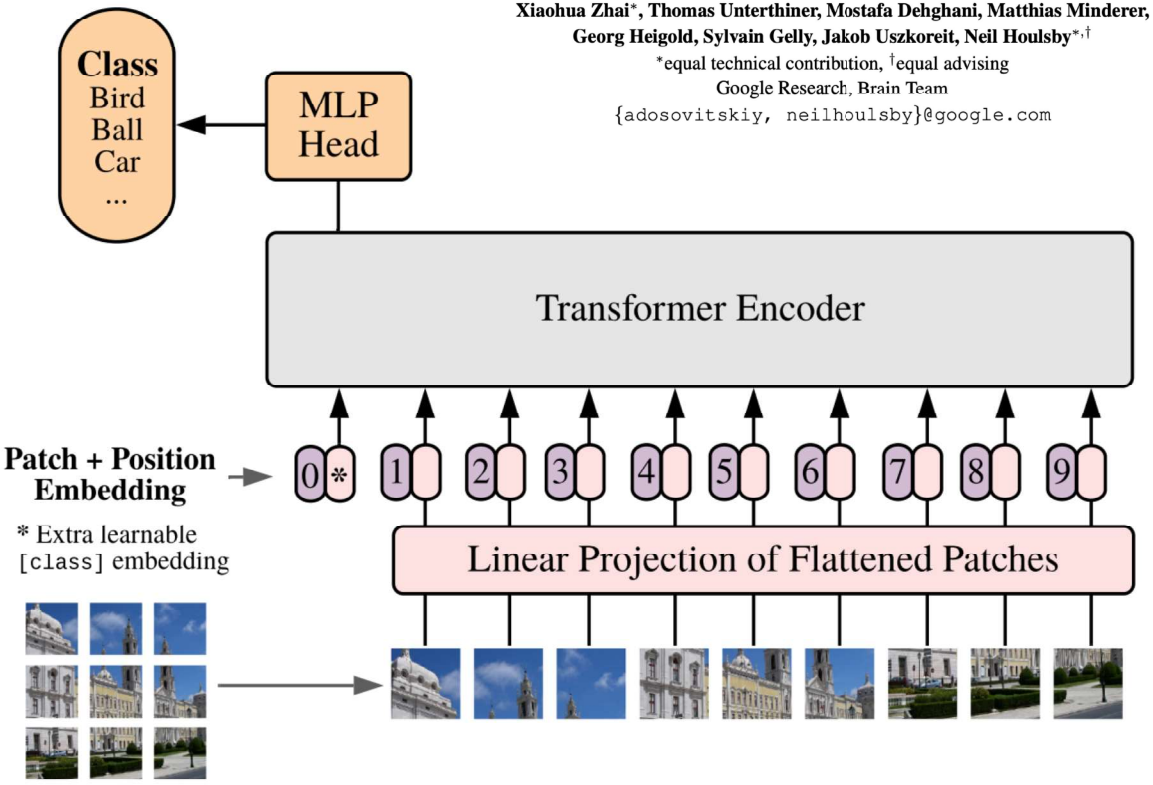


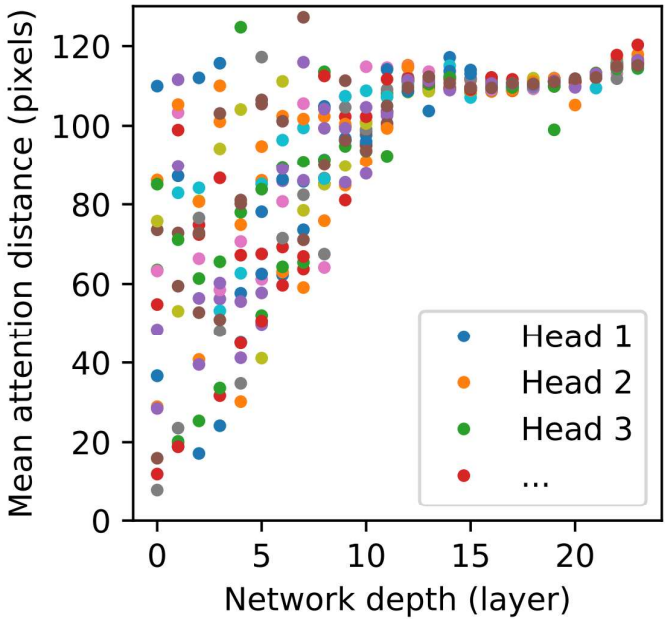
Word	Integer Token	Word Embedding
The	4	[0.455 0.345 0.234 -0.32 ..... 0.74]
sun	532	[0.342 -0.210 -0.543 0.482 ..... 0.2]
shines	323	[0.322 0.345 0.120 -0.45 ..... 0.832]
brighter	567	[0.212 -0.90 0.8 -0.321 ..... -0.12]
in	6	[0.321 0.623 0.111 -0.221 ..... 0.78]
the	4	[0.455 0.345 0.234 -0.32 ..... 0.74]
clear	65	[0.212 0.312 0.244 -0.512 ..... -0.43]
blue	823	[0.542 0.291 0.431 -0.322 ..... 0.95]

Vision Transformers (ViTs)

AN IMAGE IS WORTH 16X16 WORDS:
TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

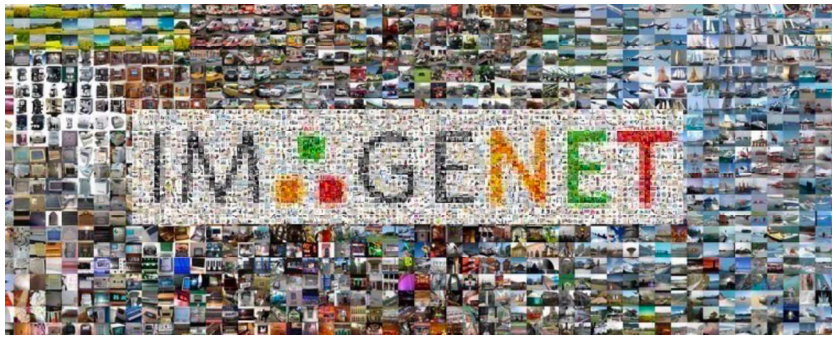
**Alexey Dosovitskiy*,†, Lucas Beyer*, Alexander Kolesnikov*, Dirk Weissenborn*,
Xiaohua Zhai*, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,
Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby*,†**
*equal technical contribution, †equal advising
Google Research, Brain Team
{adosovitskiy, neilhoulsby}@google.com

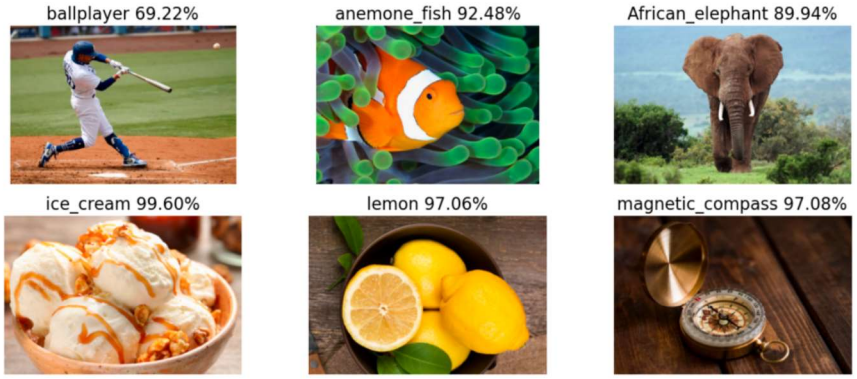




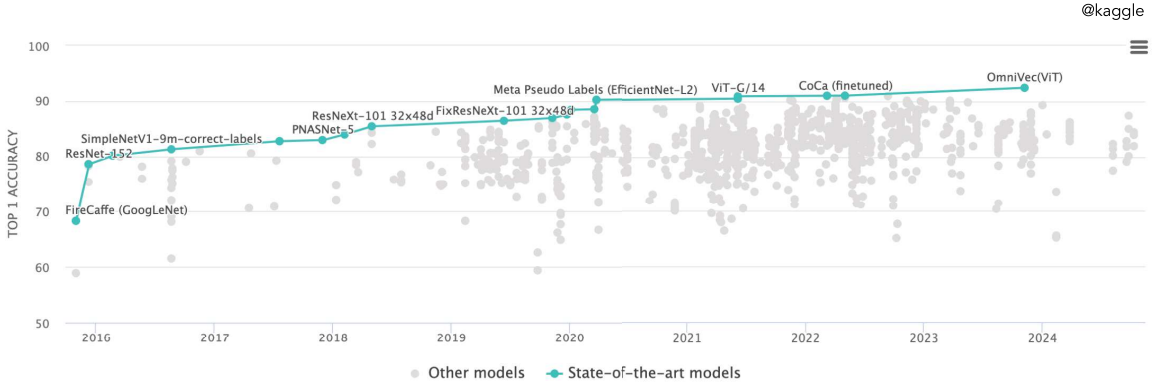
La *leaderboard* di ImageNet

Quasi 15milioni di immagini etichettate.
Circa 1000 classi.





ImageNet e' un *benchmark* per vision



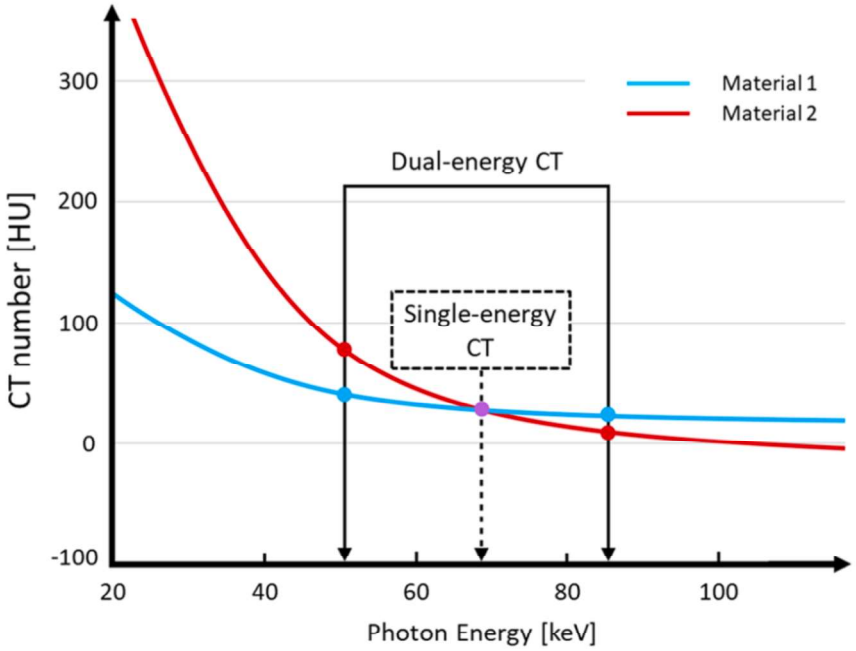
Dal ~2021 in poi i modelli più performanti sono tipicamente basati su architetture transformer.
Un buon modo di procedere per gran parti delle task in vision e' prendere un ViT (allenato su milioni di dati) e fare un fine-tuning.

Puoi fare una immagine di un gatto seduto ad un banco di scuola, tra tanti umani compagni di classe?

Progetto *in house*: Artificial Intelligence and Dual Energy Computed Tomography (AICT-DEnoise)

Project leader: Dr. Sandra Doria – CNR (w. Marco Rossetti, PhD student UniFi)

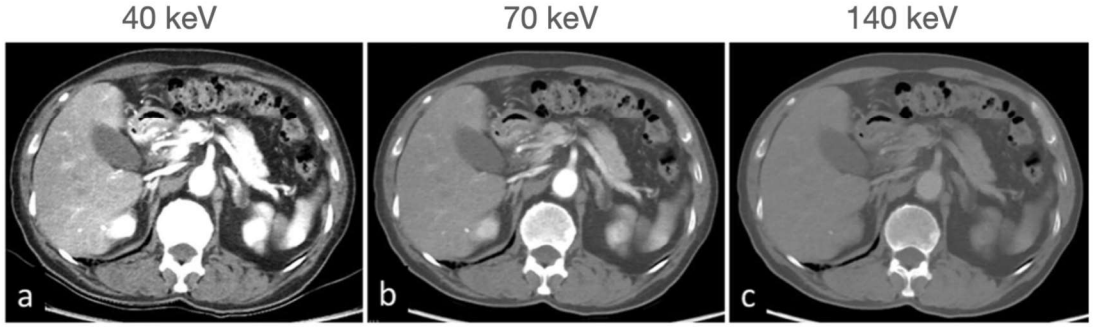
Collaborazione: CNR & UniFi (Florence), Careggi University Hospital (A.O.U. Careggi, Florence), the National center for Innovative Technologies in Public Health, National Institute of Health (TISP-ISS, Rome); the Bruno Kessler Foundation (FBK, Trento); the U.S.L. of center Tuscany; UNISER (Pistoia, Italy).



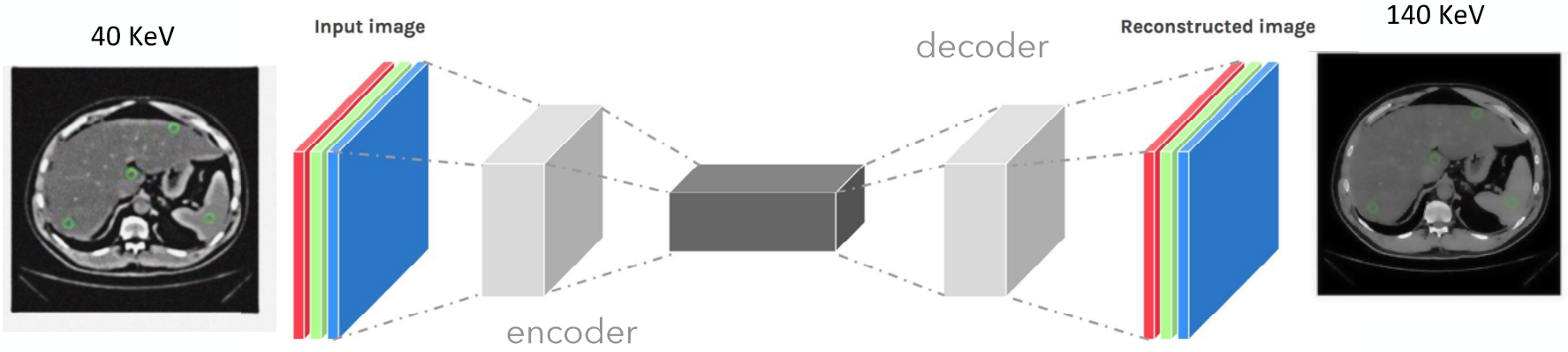
Le TC a doppia energia acquisiscono immagini a due livelli energetici (basso e alto) che interagiscono diversamente con i tessuti. Questo permette di isolare le componenti in base alle loro proprietà di assorbimento, migliorando la caratterizzazione dei tessuti e distinguendo meglio elementi come calcio, grasso o contrasto iodato, con una maggiore accuratezza diagnostica.

TC a doppia energia – denoising delle VMI a bassa energia

Dalla combinazione delle acquisizioni a due energie diverse con la TC a doppia energia si possono ricostruire immagini virtuali a energie arbitrarie (VMI; virtual mono-energetic images). A basse energie (bassi keV) il contrasto aumenta ma il rumore aumenta e così si perde in dettagli/info.



Si può addestrare una rete a imparare il mapping non lineare tra le proprietà del rumore nelle immagini VMI a bassa energia e quelle ad alta energia.
Il dataset contiene coppie di immagini VMI a bassa energia e VMI ad alta energia, dove il "segnale" è mascherato (questo potrebbe essere un problema di per sé).

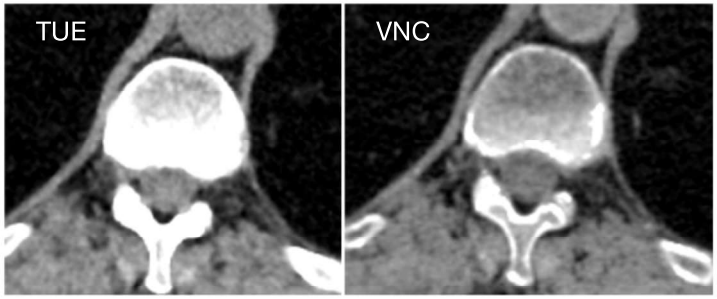


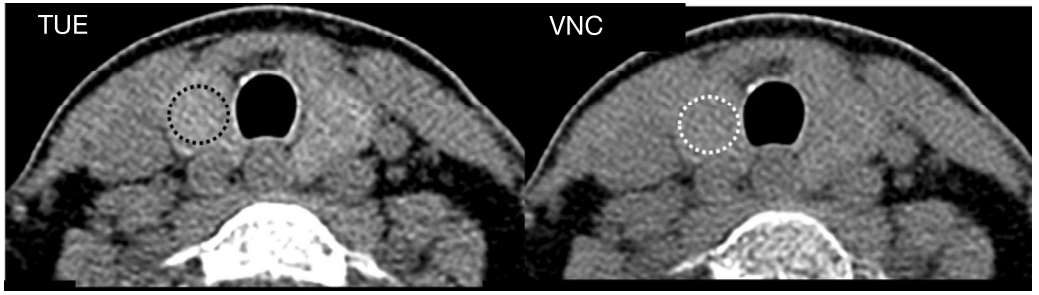
TC a doppia energia – migliorare la qualità delle VNC

Protocollo standard tipico TC singola –> due acquisizioni:

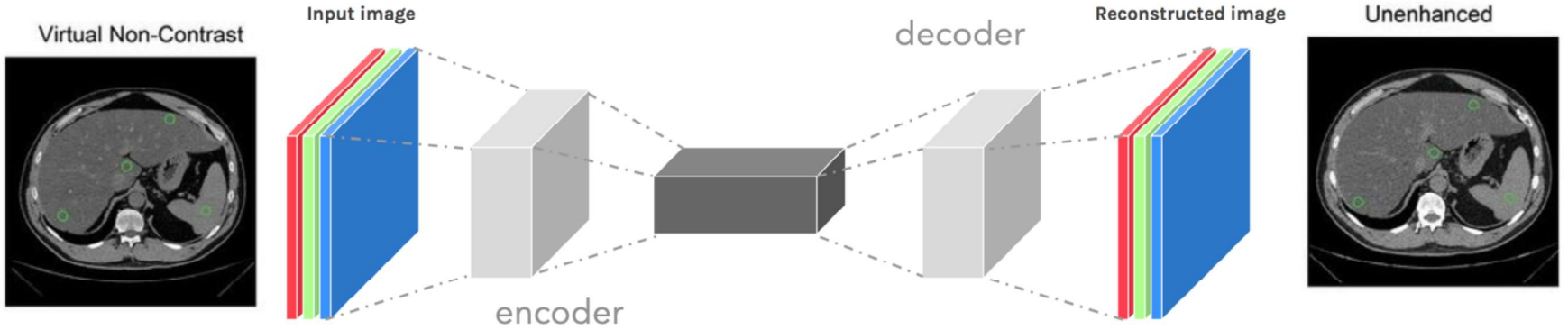
- prima (True Unhenanced, TUE)
- dopo mezzo di contrasto

Utilizzando la decomposizione dei materiali, si possono generare immagini virtuali senza contrasto (VNC) sottraendo la componente di iodio dall'immagine TC a doppia energia con contrasto. Le immagini VNC possono eliminare la necessità di esposizione alle radiazioni quando sono richieste scansioni TC senza contrasto.



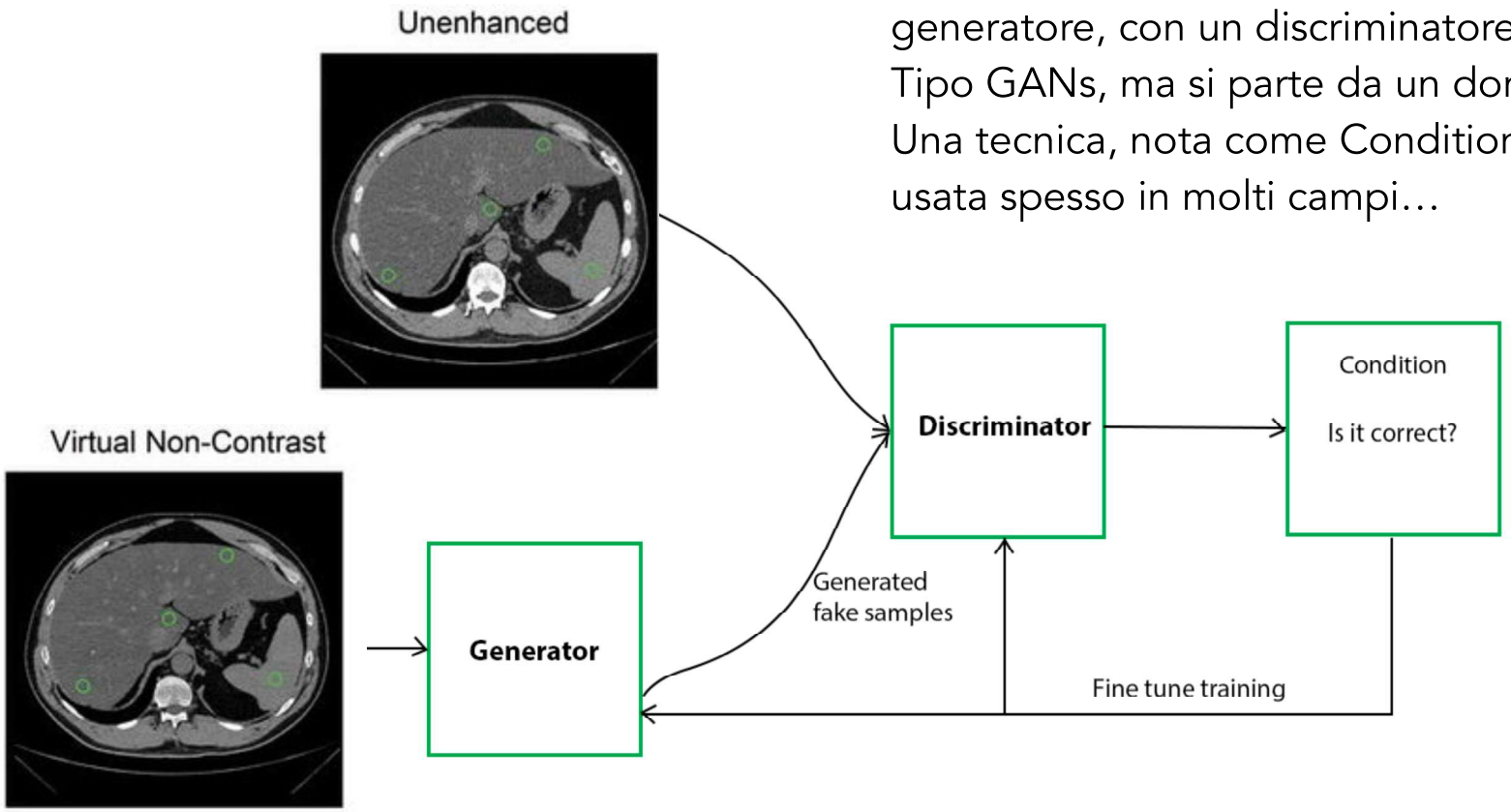


Problemi —> presenza di artefatti durante la produzione di VNC; non e' diffuso fare diagnosi solo con le immagini VNC (sembrano diverse dalle *dirette*!)



TC a doppia energia – migliorare la qualità delle VNC

E se la trasformazione la facessimo fare ad un generatore, con un discriminatore che controlla?
Tipo GANs, ma si parte da un dominio, non da rumore.
Una tecnica, nota come Conditional GAN (cGAN), usata spesso in molti campi…

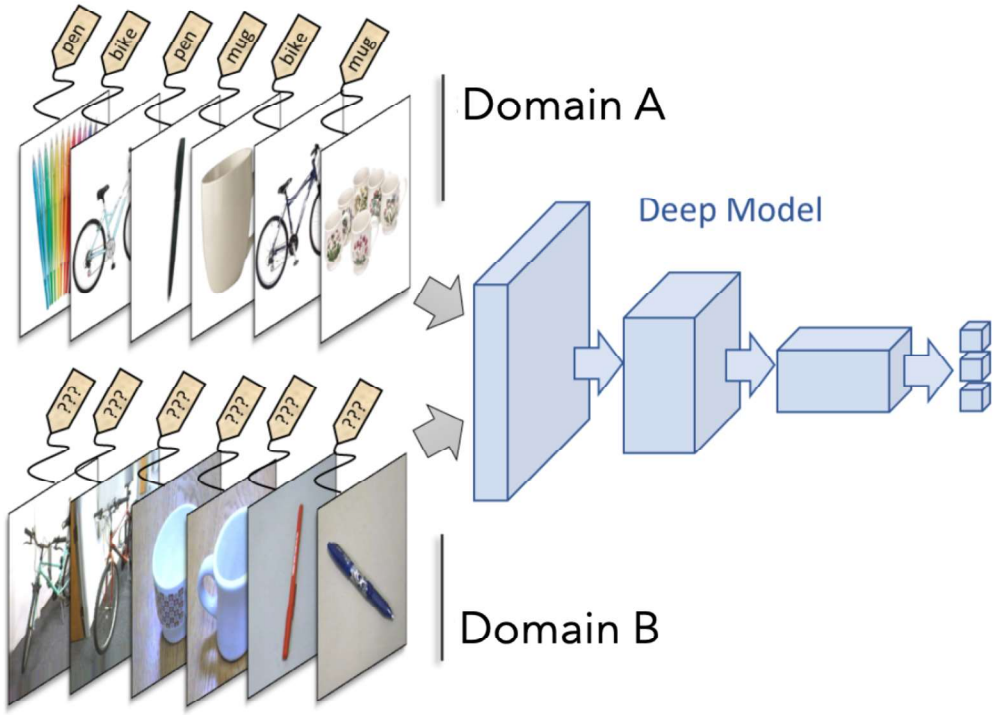


Mirza et al. 2014
Kazeminia et al., 2020
Gong et al., 2021
Skandarani et al., 2021

Problema: ma siamo sicuri che il modello funzionerà su un'altra macchina?

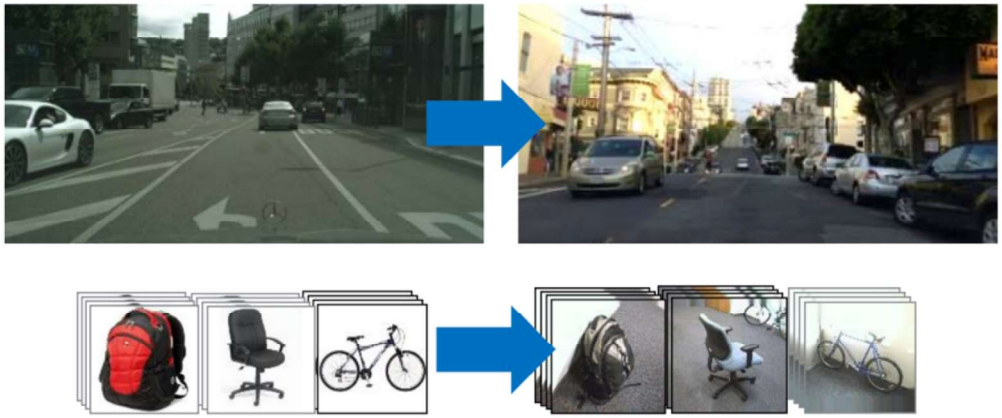
Dominio sorgente

Dominio target

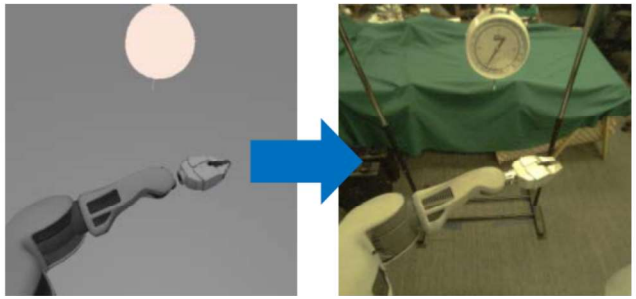


Un problema generale

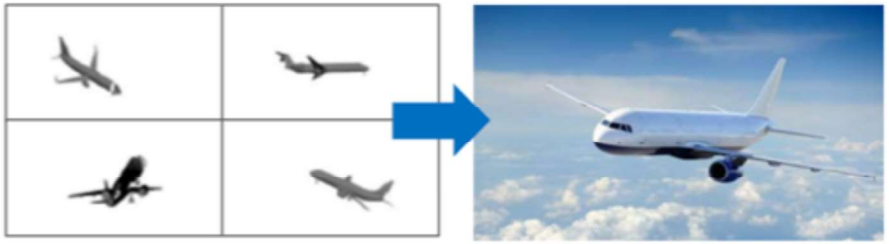
From dataset to dataset



From simulated to real control

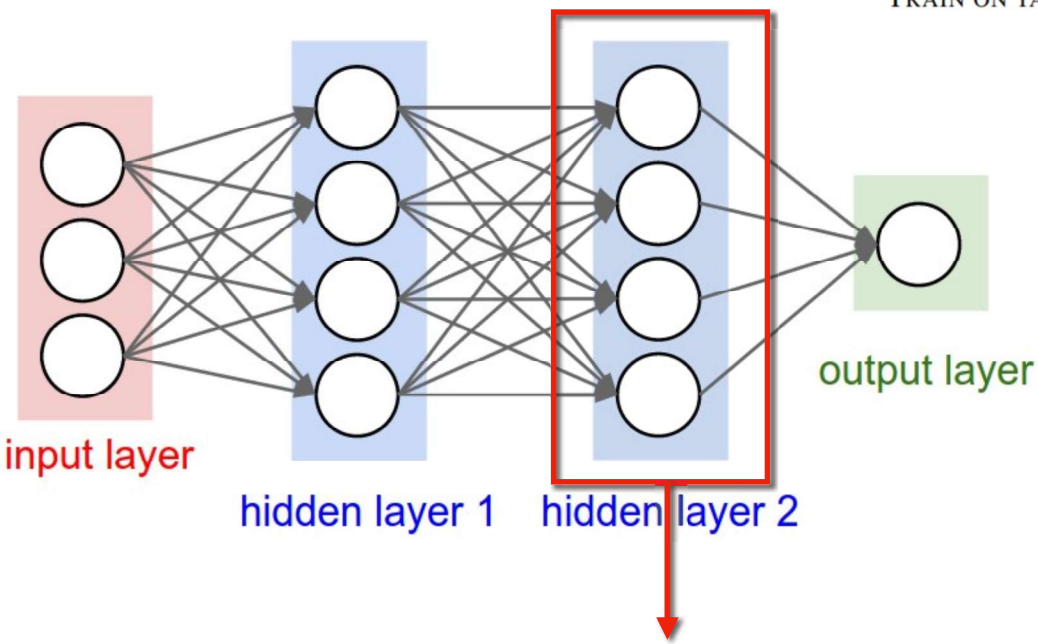


From CAD models to real images



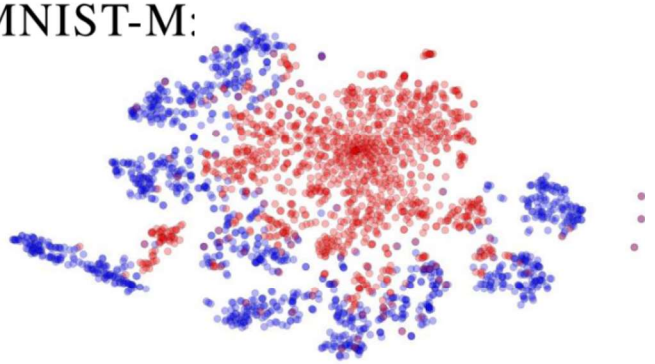
- Può esserci un netto divario tra i modelli (dominio sorgente) e i dati reali (dominio target) o tra diversi domini di dati reali.
- I modelli (dati sintetici, dominio sorgente) non catturano la complessità del dataset reale.
- I diversi domini possiedono proprietà differenti.

Encoded features don't match

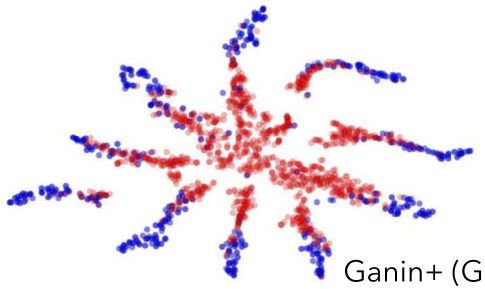


	MNIST	SYN NUMBERS	SVHN	SYN SIGNS
SOURCE				
TARGET				
	MNIST-M	SVHN	MNIST	GTSRB
SOURCE ONLY	.5749	.8665	.5919	.7400
TRAIN ON TARGET	.9891	.9244	.9951	.9987

MNIST → MNIST-M:



SYN NUMBERS → SVHN:



Ganin+ (Google DeepMind)

(a) Non-adapted

In other words, embeddings are misaligned…

VisDA: The Visual Domain Adaptation Challenge

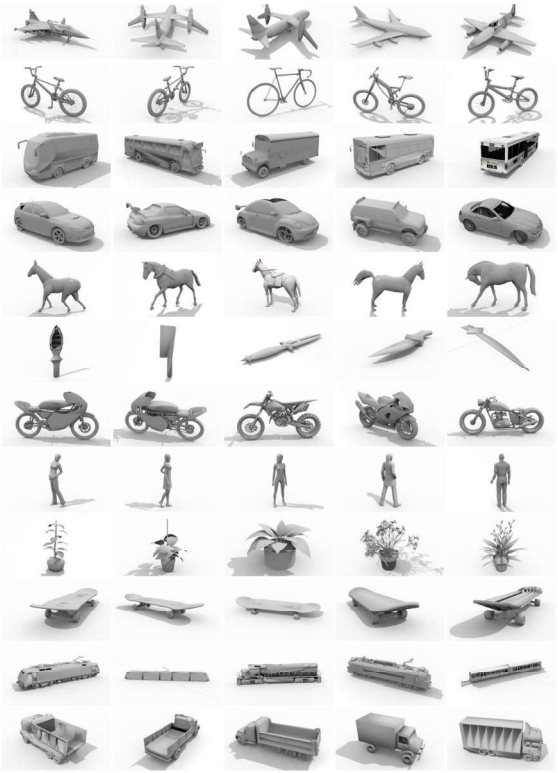
Xingchao Peng[1], Ben Usman[1], Neela Kaushik[1], Judy Hoffman[2], Dequan Wang[2] and Kate Saenko[1]

xpeng,usmn,nkaushik,saenko@bu.edu, jhoffman,dqwang@eecs.berkeley.edu

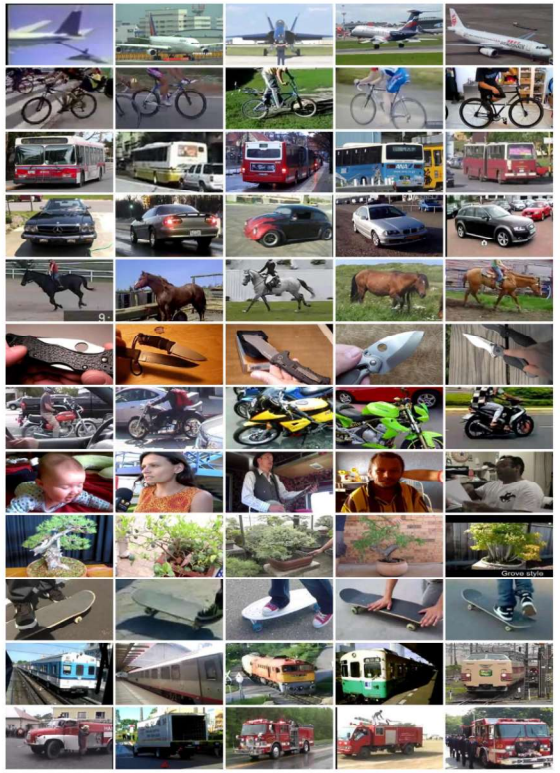
[1]Department of Computer Science, Boston University

[2]EECS, University of California Berkeley

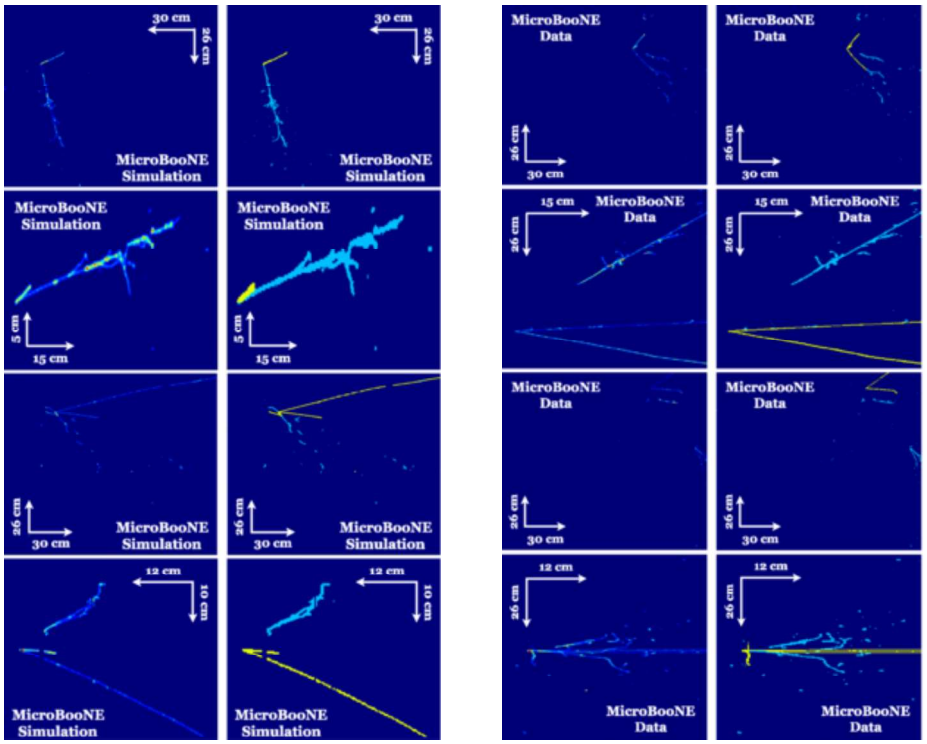
Source domain: CAD



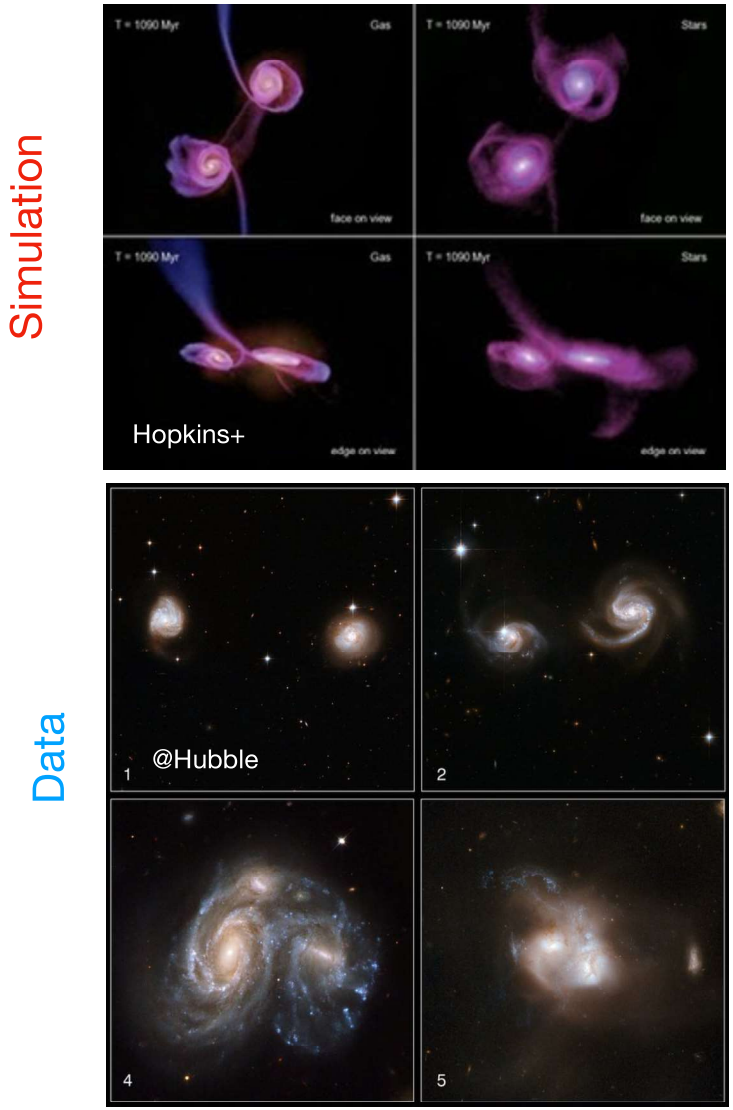
Target domain: Youtube frames



Tanti esempi in fisica



neutrinos: MicroBooNE experiment @Fermilab



Ed in imaging medicale ovviamente, dove il problema non e' solo *teorico*

Domain Adaptation for Medical Image Analysis: A Survey

Hao Guan, and Mingxia Liu, *Senior Member, IEEE*

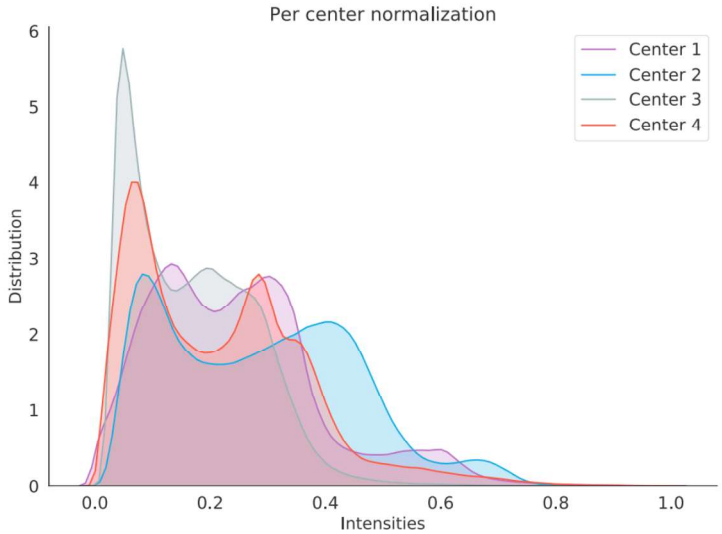


Fig. 1. Intensity distribution of MRI axial-slice pixels from four different datasets (*i.e.*, UCL, Montreal, Zurich, and Vanderbilt) that collected for gray matter segmentation. Intensity is normalized between 0 and 1 for each site. Image courtesy to C. Perone [24].

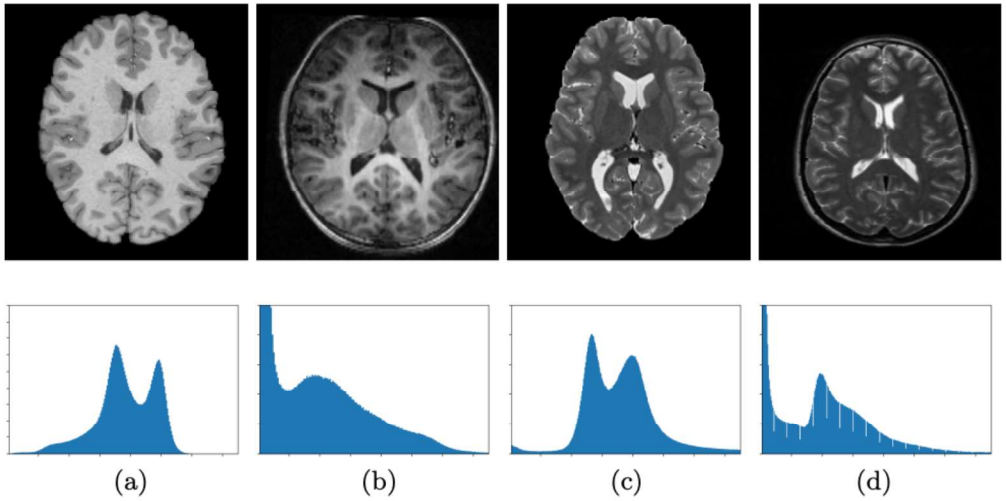
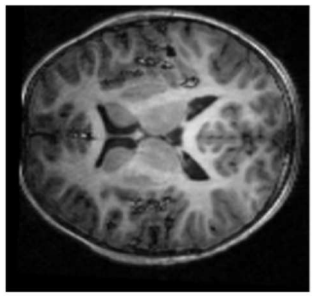


Fig. 2. Image slices (top) and corresponding intensity distribution (bottom) of normalized T1-weighted (a, b) and T2-weighted (c, d) MRIs from different scanners. Image courtesy to N. Karani [29].

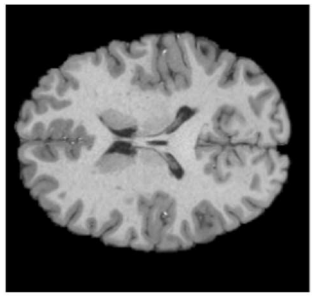
Domain shift challenge: if I have a lot of labels in the target domain…

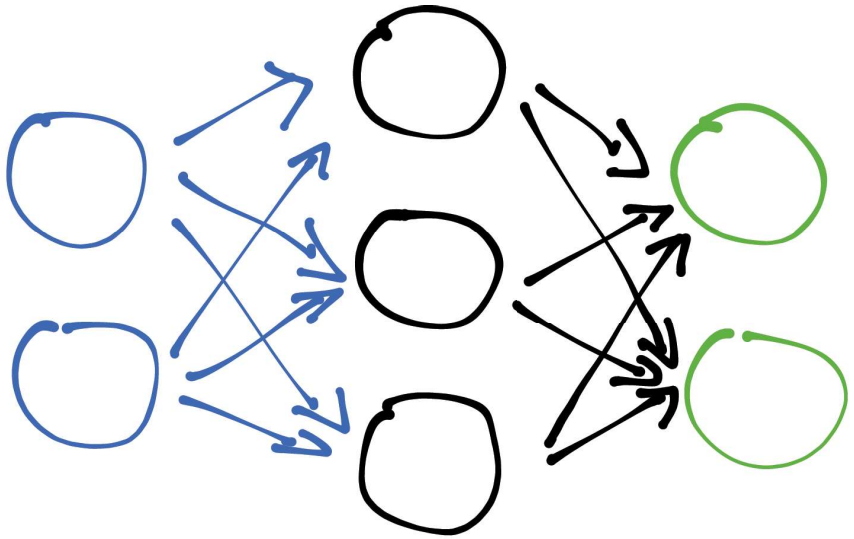
Full dataset

Domain A: large labeled dataset



Domain B: large labeled dataset

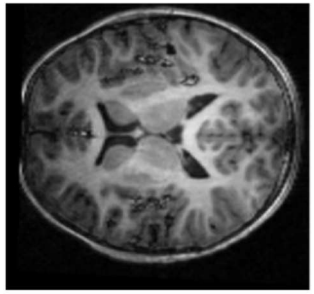




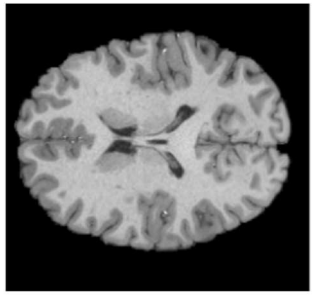
Domain shift challenge: if I have a few labels in the target domain…

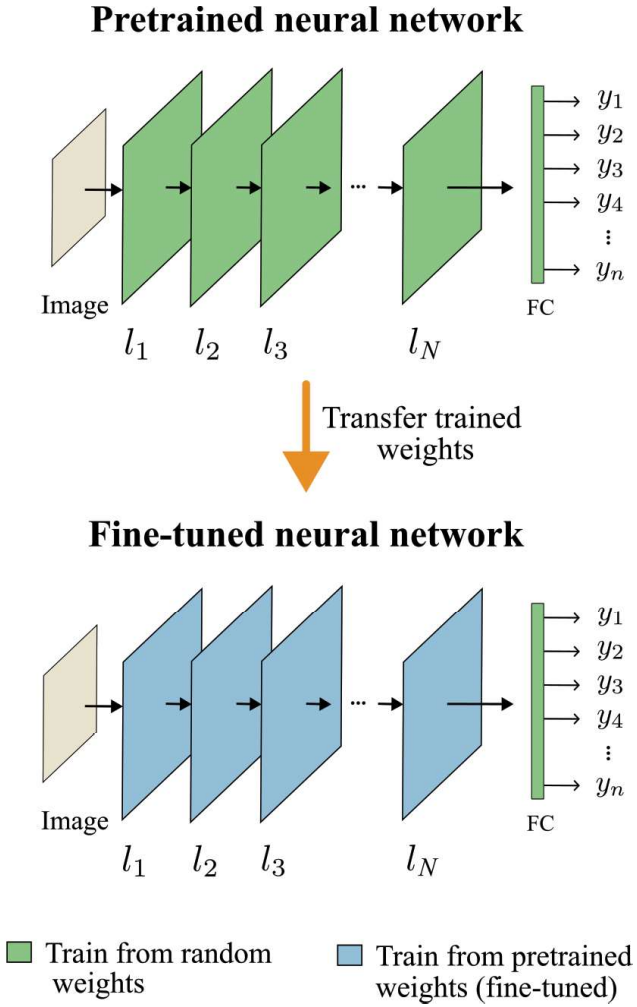
Full dataset

Domain A: large labeled dataset



Domain B: large labeled dataset



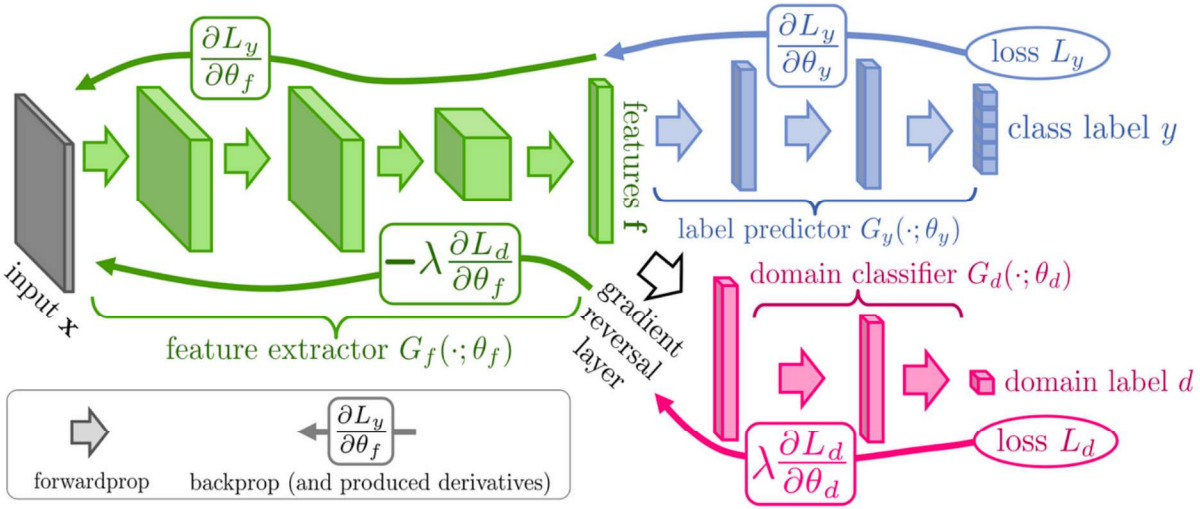


what if I don’t have labels for the target domain?

Target domain room

Feature-based domain adaptation

Domain adversarial neural nets (DANN)



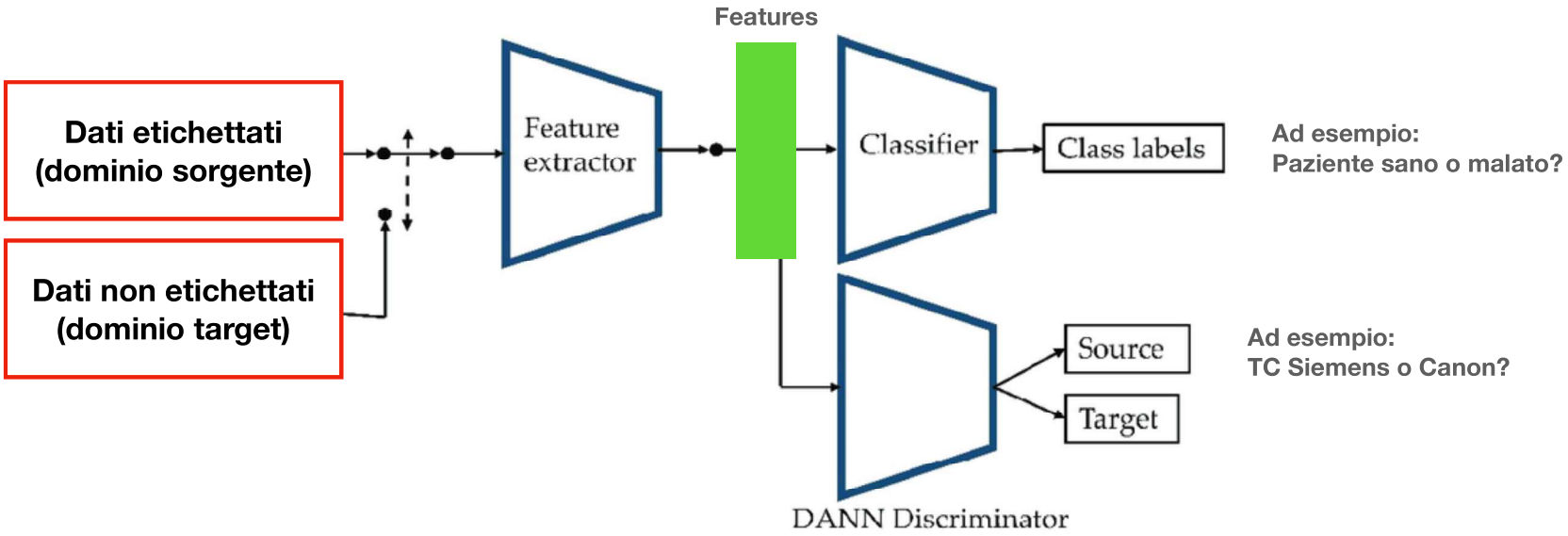
Domain classifier

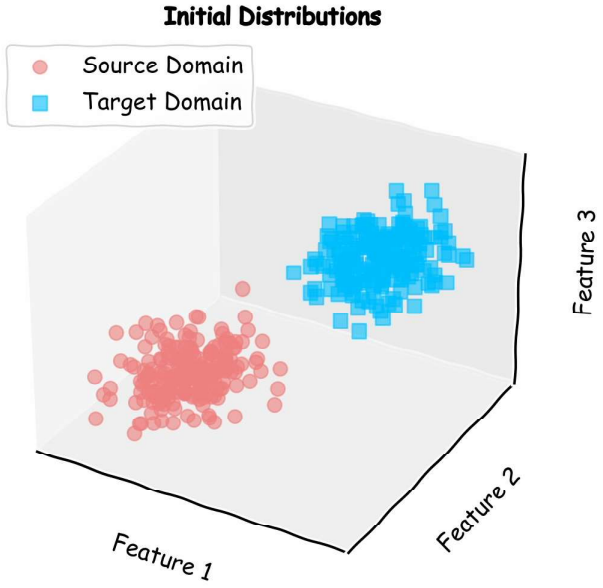
Feature extractor

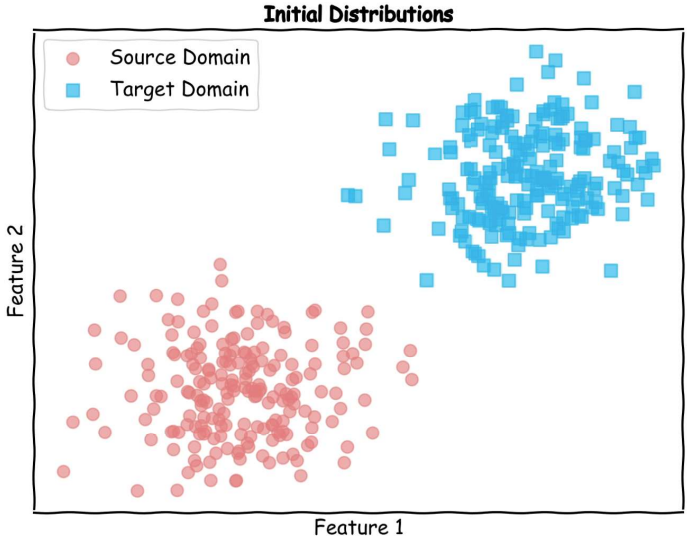
Ganin et al. 2016

(one of the best ideas in DL, IMHO)

It seems like it's often a good idea to induce competition among AIs (GAN, DANN, self-play in RF…)

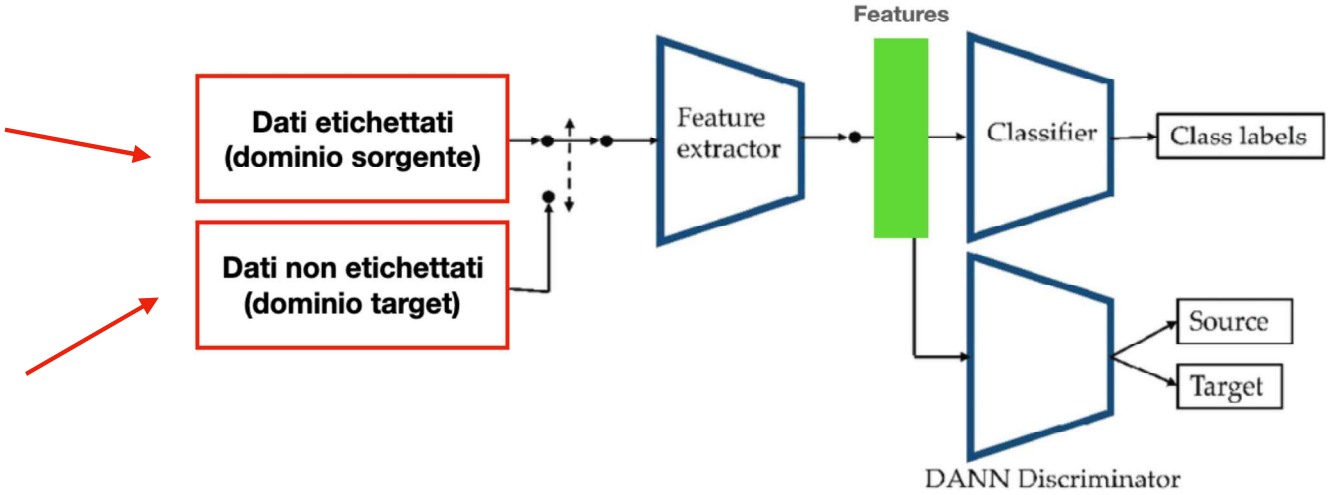

Features
Dati etichettati
(dominio sorgente)
Dati non etichettati
(dominio target)
Feature
extractor
Classifier
Class labels
Ad esempio:
Paziente sano o malato?
Source
Target
Ad esempio:
TC Siemens o Canon?
DANN Discriminator


Initial Distributions
Source Domain
Target Domain
Feature 1
Feature 2
Feature 3


Initial Distributions
Source Domain
Target Domain
Feature 2
Feature 1

Dominio sorgente
(Etichettato)

Dominio target
(Non etichettato)



Domain adversarial learning può apprendere trasformazioni invarianti rispetto alla macchina, rendendo il modello robusto rispetto agli spostamenti di dominio, senza la necessità di addestrare dataset target specificamente etichettati.

Usare modelli generativi per cambiare dominio

Un'alternativa alla creazione di una rappresentazione delle caratteristiche invariante rispetto al dominio è la mappatura da un dominio all'altro. La mappatura viene solitamente creata in modo avversariale e a livello di pixel (ossia, **adattamento avversariale di dominio a livello di pixel**). Con una cGAN per esempio!

Synthetic Vehicle Driving Image (GTA-V)

Adapted to look realistic

CyCADA: Cycle-Consistent Adversarial Domain Adaptation

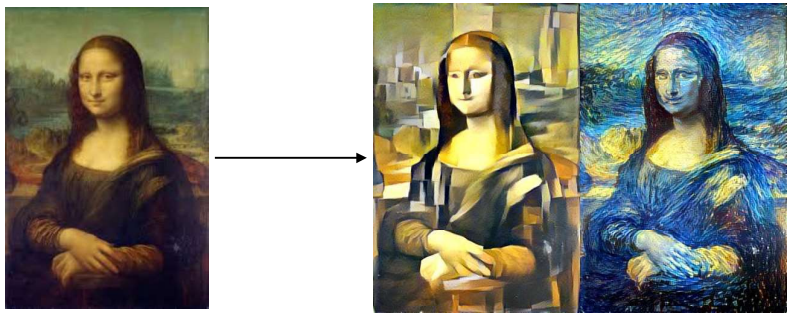
Source image (GTA5)

Adapted source image (**Ours**)

Target image (CityScapes)

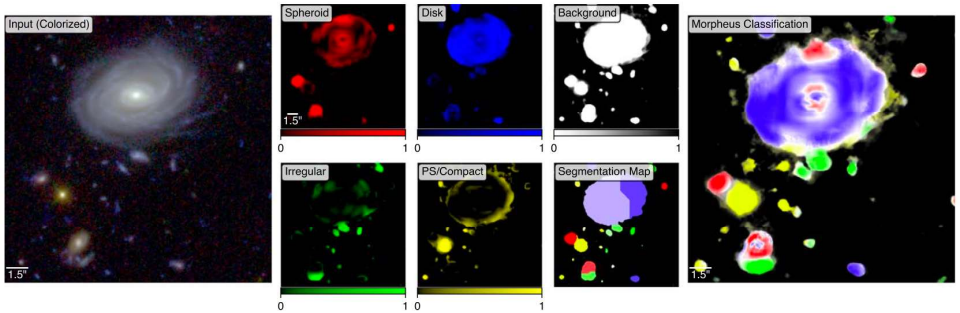
Pixel accuracy on target
Source-only: 54.0%
Adapted (**ours**): **83.6%**

Il generatore esegue l'adattamento a livello di pixel traducendo un'immagine di input del dominio sorgente in un'immagine che assomiglia da vicino alla distribuzione del dominio target (mantenendo il label!).

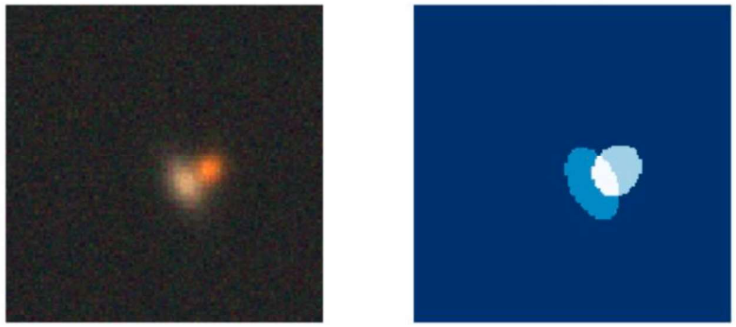


Cycle-GAN for style transfer

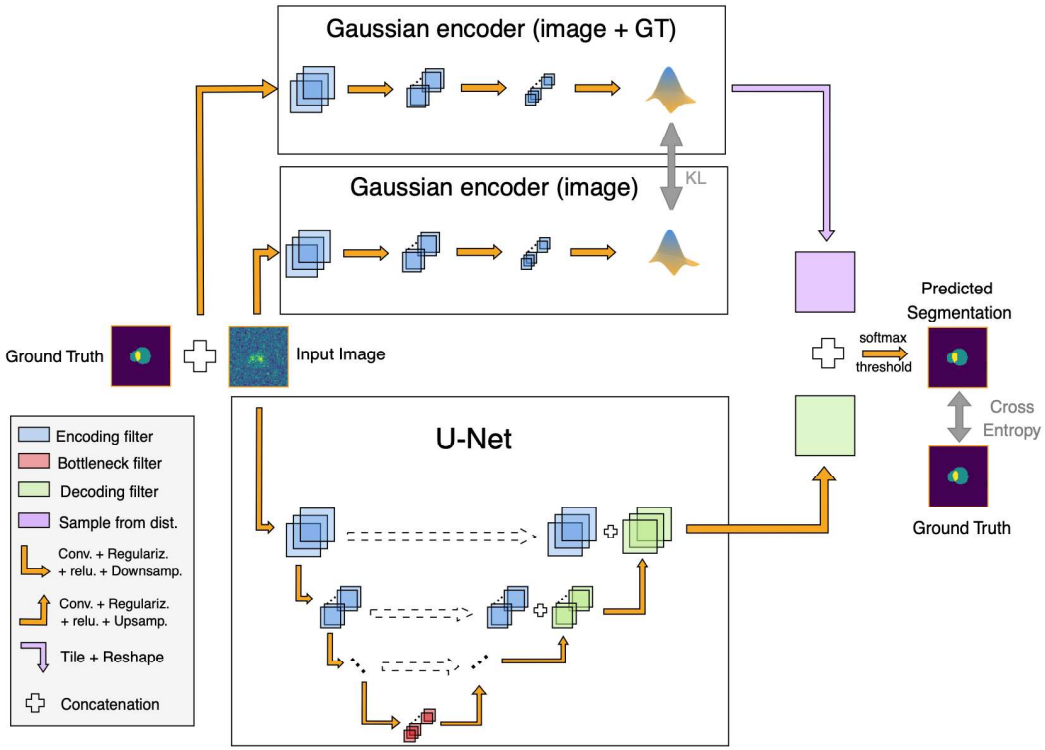
Segmentazione semantica di galassie



"Morpheus: A Deep Learning Framework for the Pixel-level Analysis of Astronomical Image Data" — Hausen & Robertson 2020

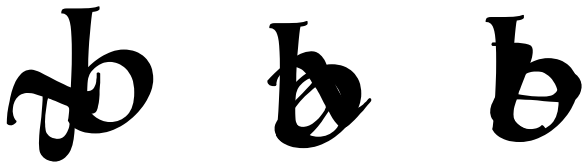


"Deblending Galaxies with Mask R-CNN" — Kamath+18



"Probabilistic segmentation of overlapping galaxies for large cosmological surveys" — Bretonnière+21

Dalle galassie alle lettere, il passo e’ breve…



Ok

Mhh

OMG

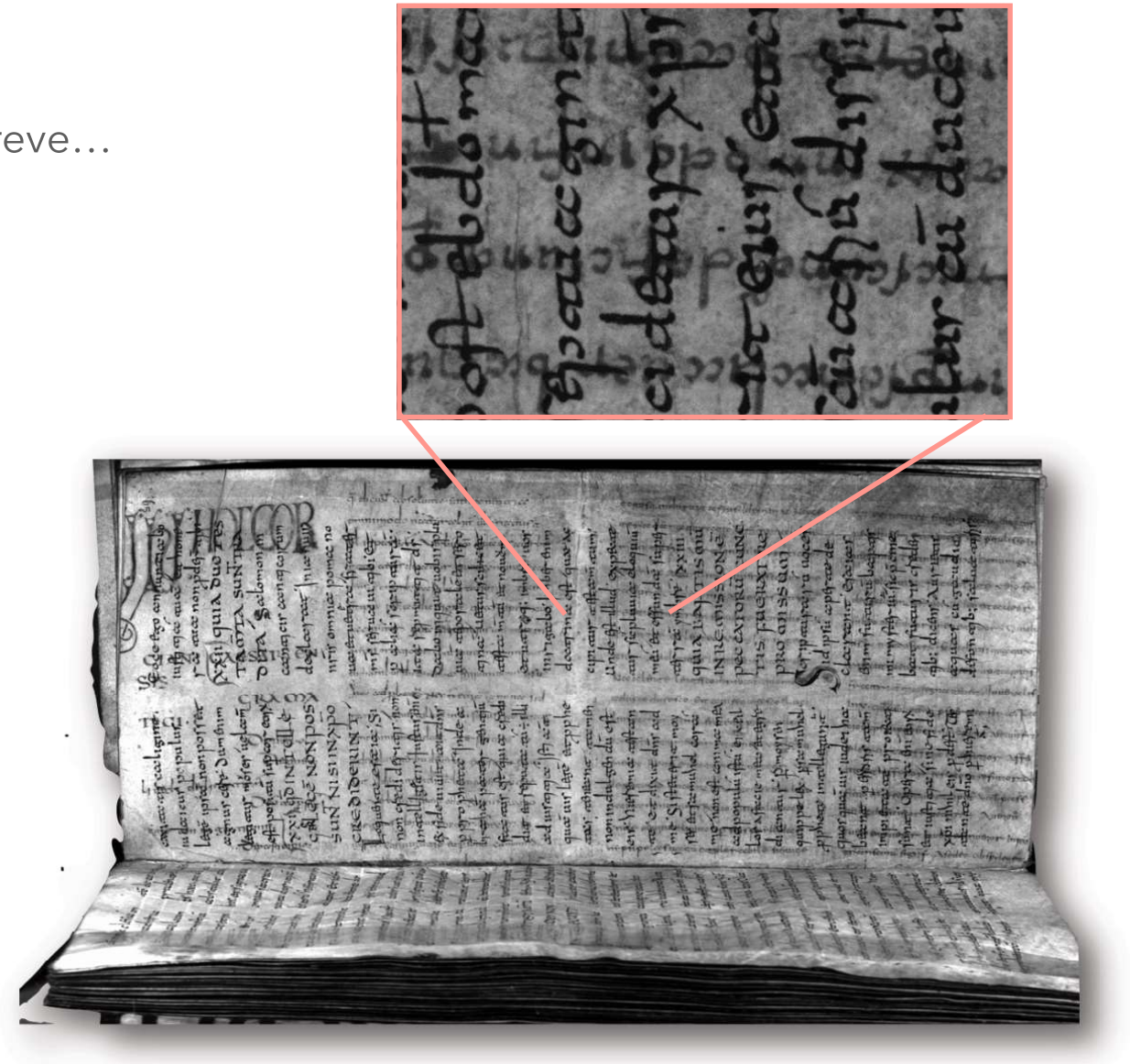
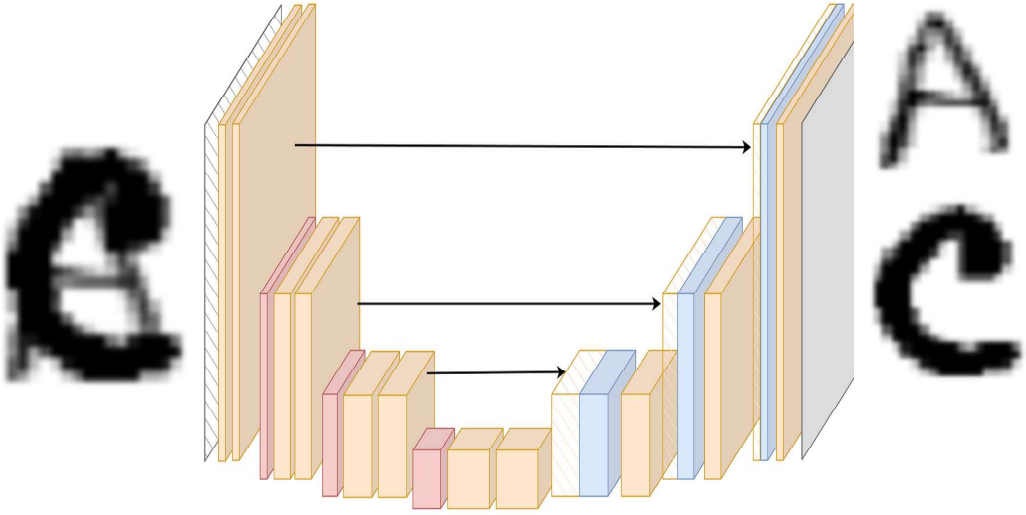
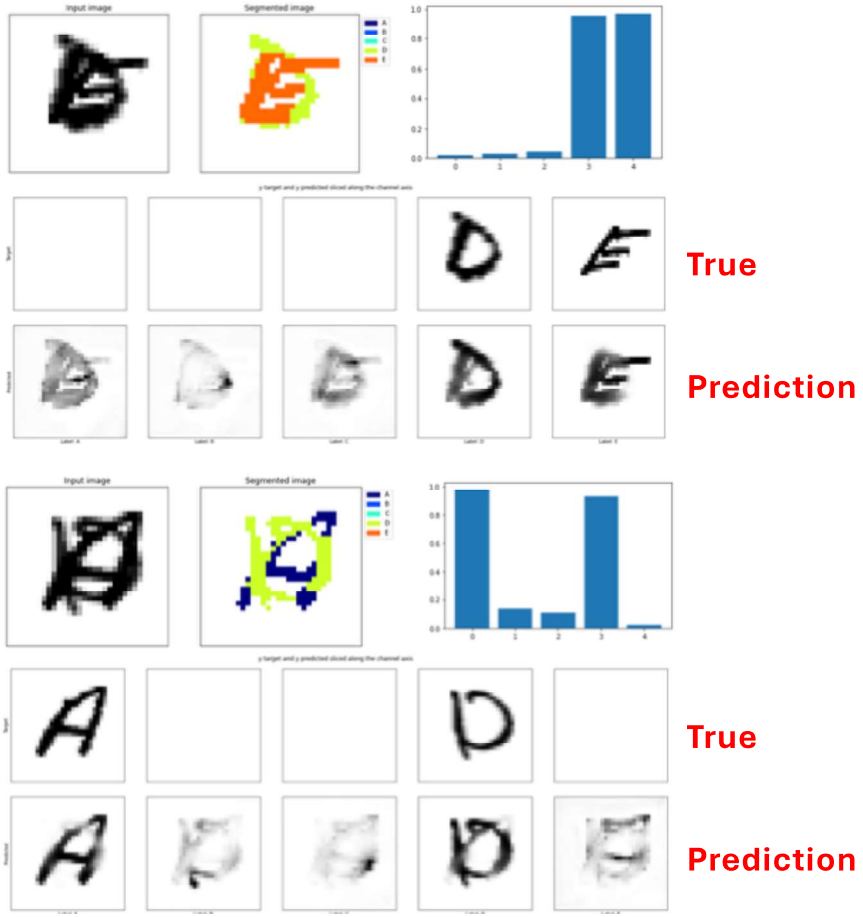


Image segmentation — overlapping letters

- x_train, x_valid, x_test : 15000, 3000, 3000
- 5 classes (letters)
- 2 overlap letters
- Random element + check for uniqueness





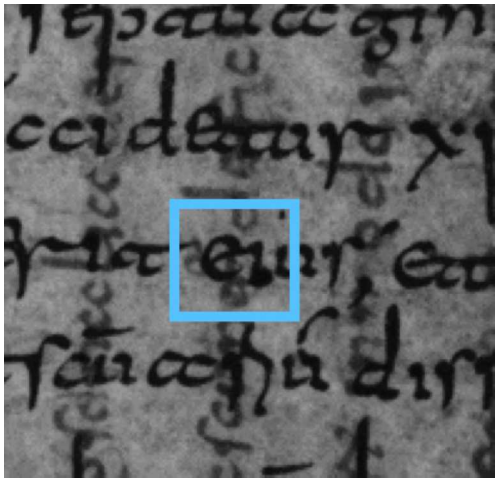
Perino & Ginolfi+23

Domain adaptation

Prossimo passo?

Source Domain (Caroline script)

Target Domain

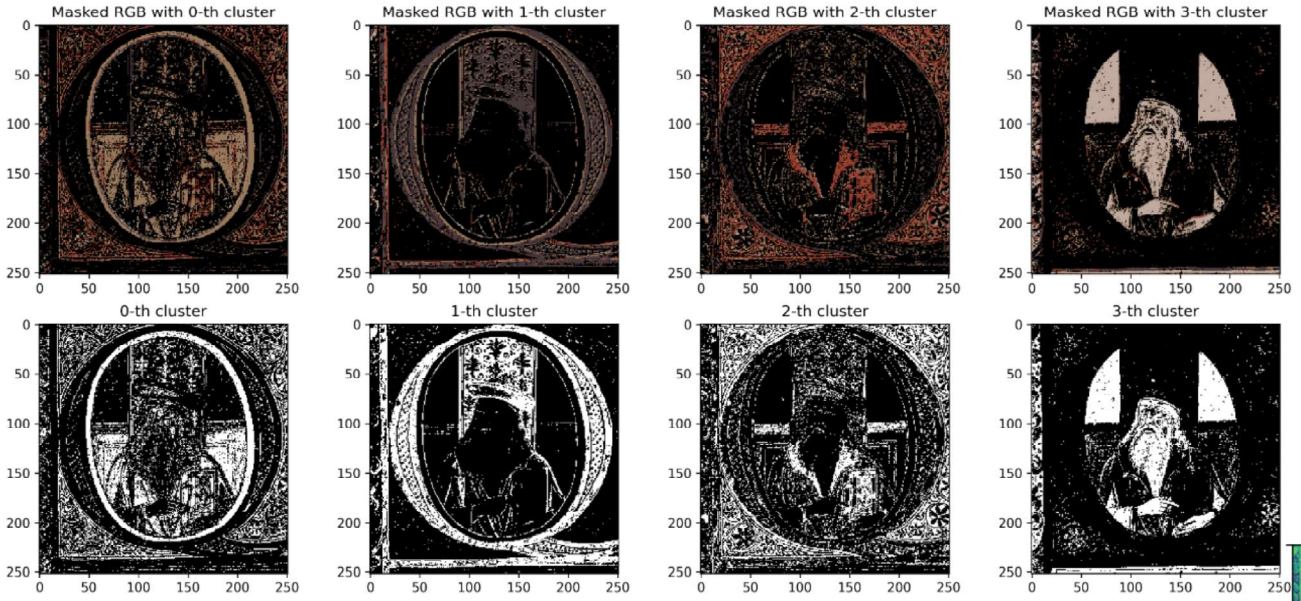


Labelled: several pages already trascribed (publicly available)

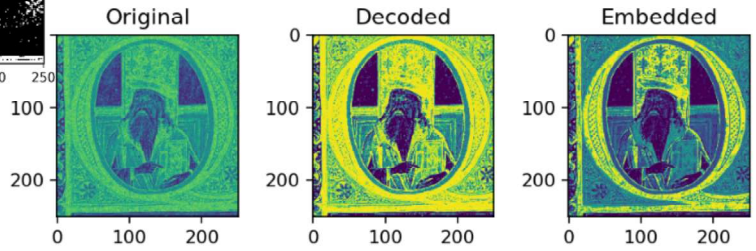
Unlabeled

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
INFN

Application: Deep *Variational* Embedding for Cultural Heritage



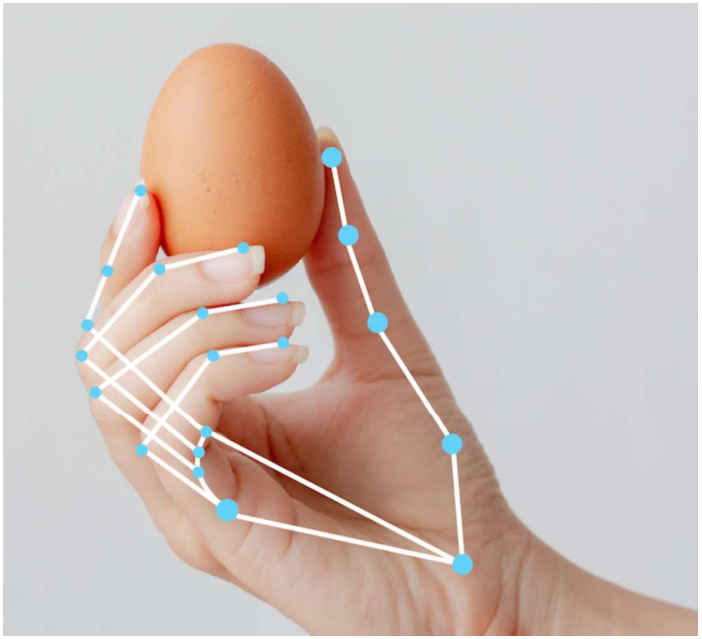
@A. Bombini (LABEC, INFN)



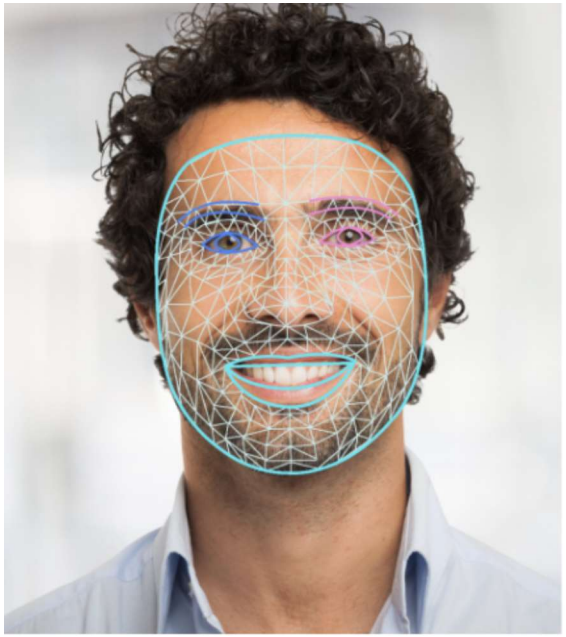
semantic segmentation obtained via deep variational spectral embedding!

MediaPipe

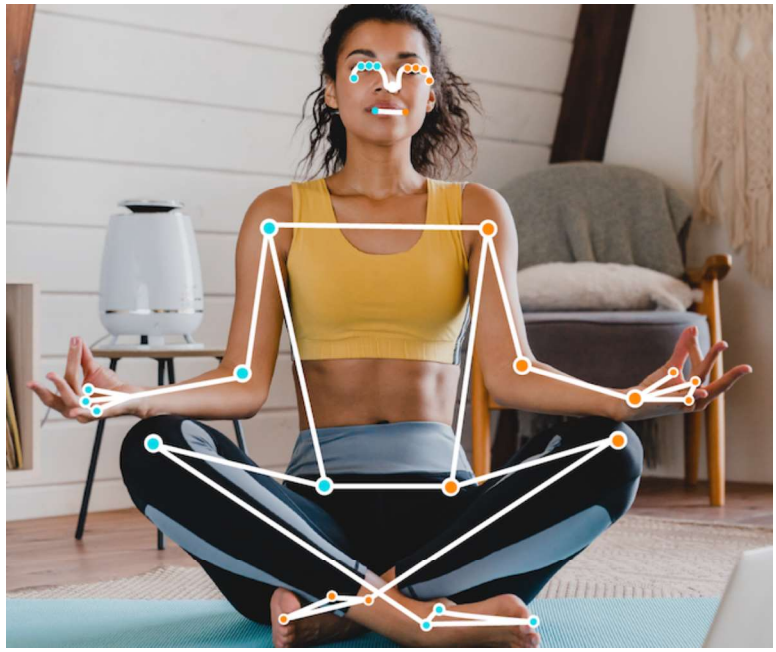
Hands Landmarker



Face Landmarker



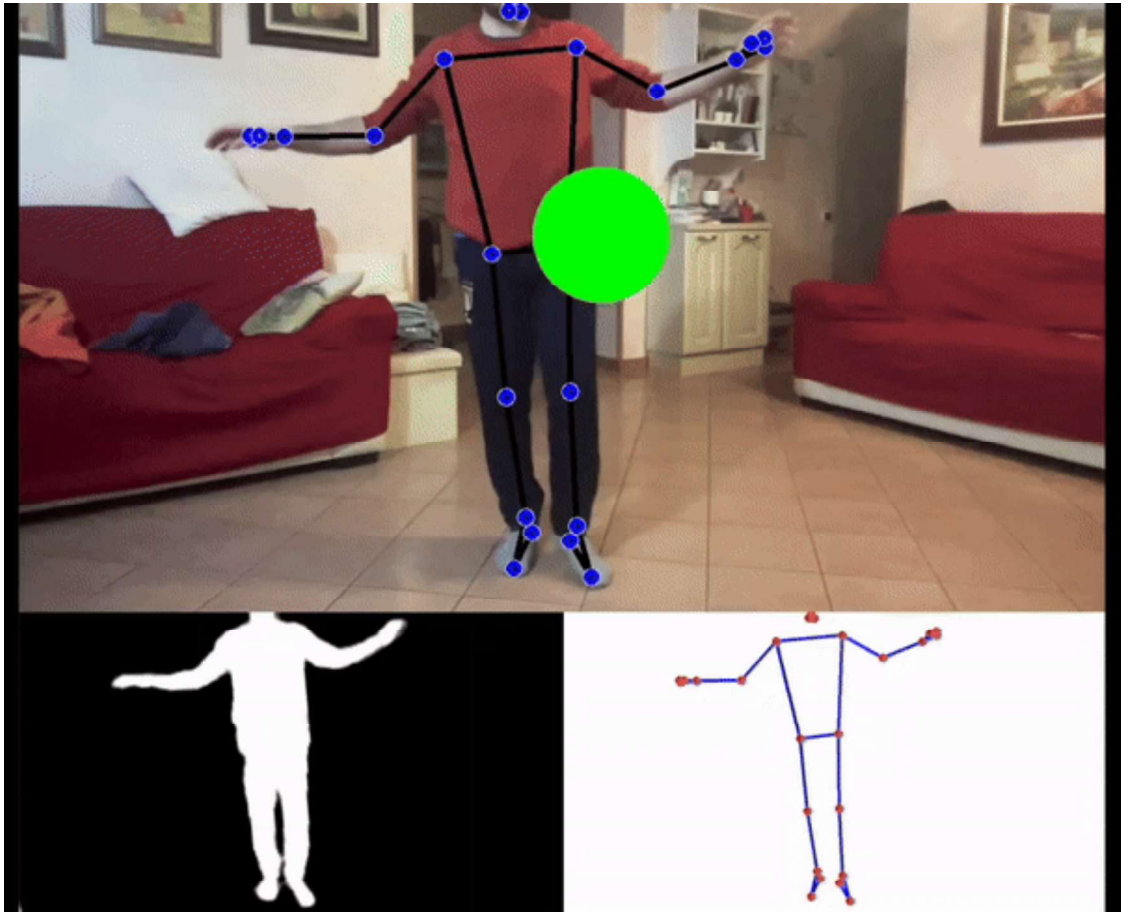
Full body pose Landmarker



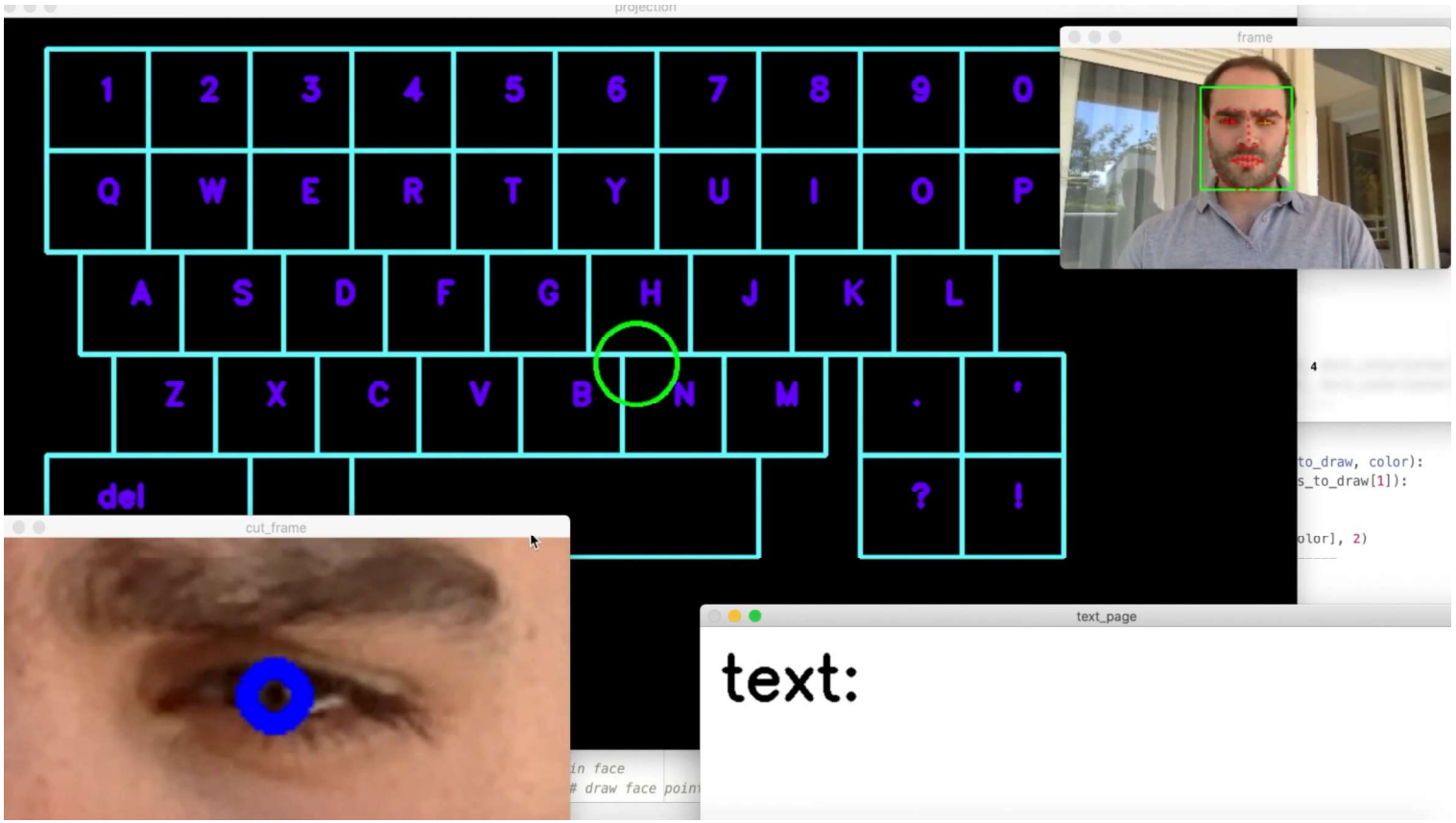
@mediapipe/google

E se la segmentazione invece di lavorare su immagini statiche, lavora sui frames di un flow video?
Possiamo interagire con il pc, associando nostre azioni a sue azioni!

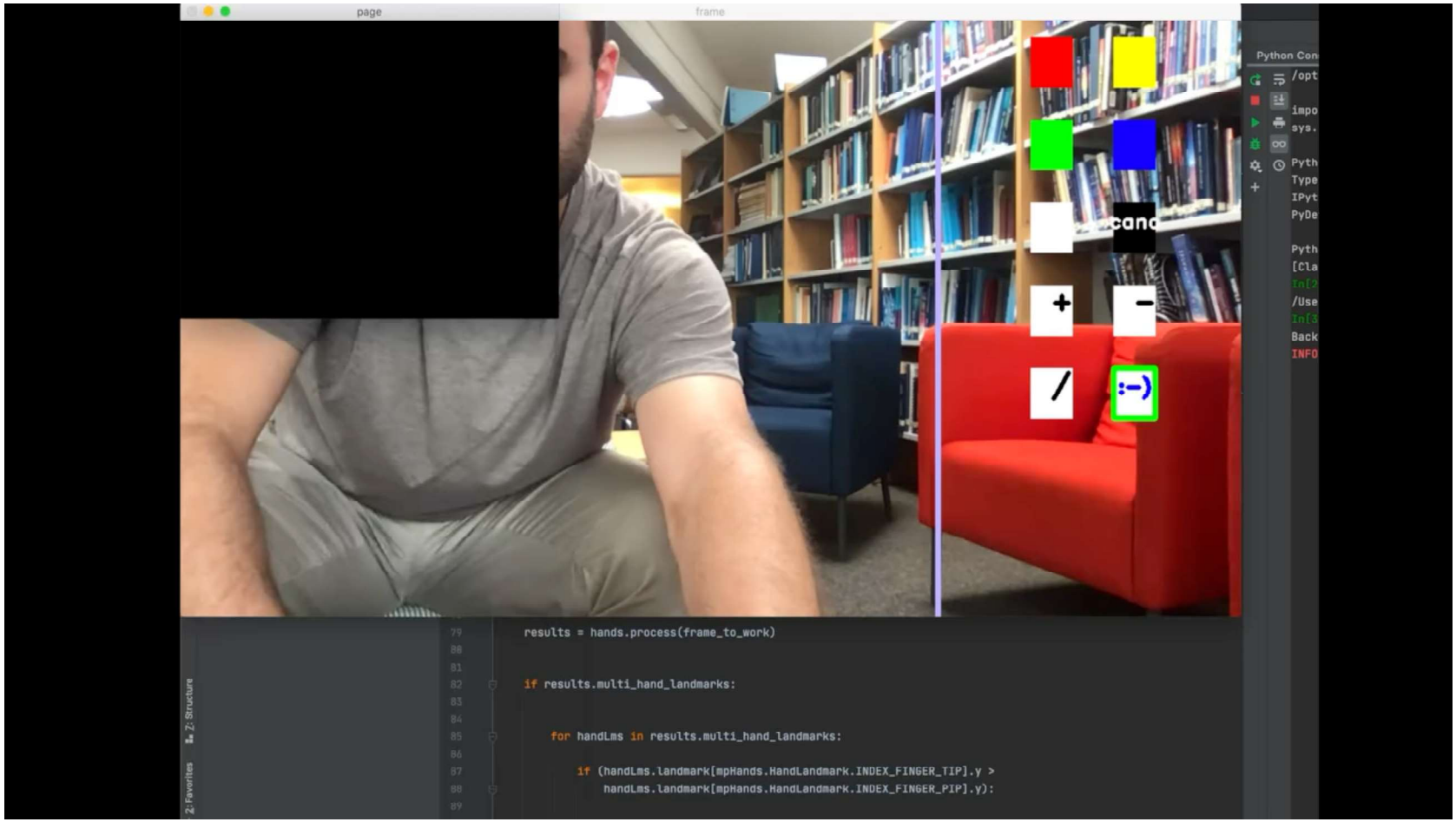
Sport in realtà aumentata — palleggiare con i pixel



Scrivere con gli occhi



Graffiti su un muro d’aria



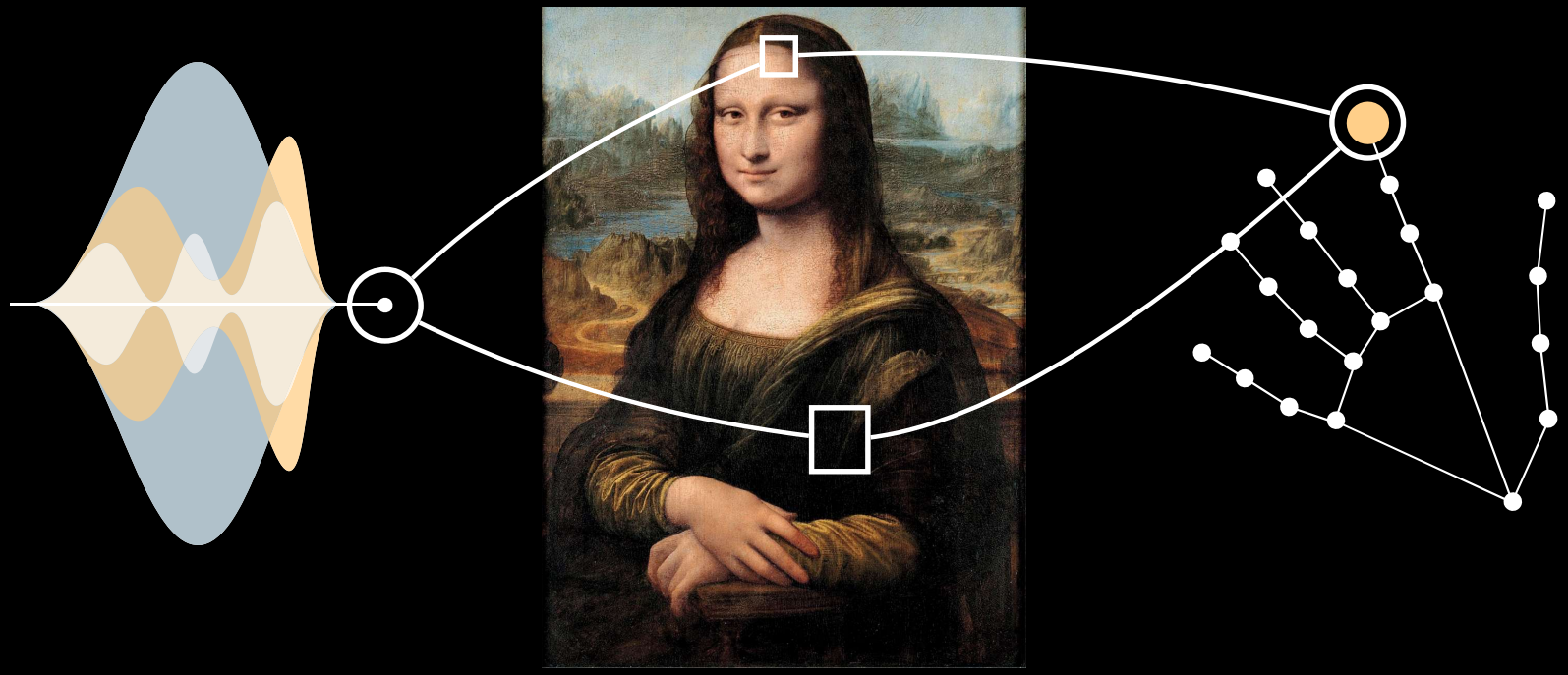
herakoi: a motion-sensing sonification experiment

github.com/herakoi/herakoi

@herakoi_

Copyright 2022 (MIT Licence)
Michele Ginolfi, Luca Di Mascolo, and contributors.

motion-aware exploration of images



Output sonoro

Mappatura dell’immagine

Tracciamento della mano in tempo reale

putting things together...

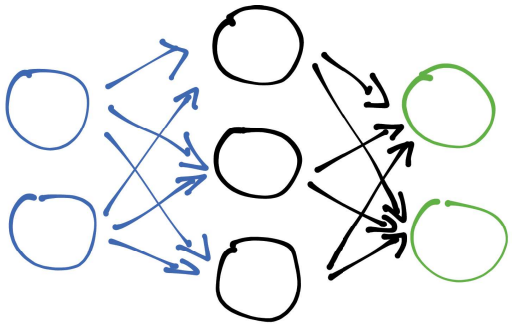
herakoi in the wild

ScienzEstate 2023 @ UniFi

Grazie
per l'attenzione

Grazie
per l'attenzione

Grazie
per l'attenzione



Grazie
per l'attenzione