UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Dipartimento di Ingegneria “Enzo Ferrari”
Corso di Laurea Magistrale in Ingegneria Informatica (D.M. 270/04)

Relatore:
Prof.ssa Sonia Bergamaschi

Correlatore:
Pietro Leo

Candidato:
Matteo Gabrielli

ACTION RECOGNITION PER STIMARE LE ACTIVITIES OF DAILY LIVING (ADL) DI PERSONE ANZIANE

Anno Accademico
2017 - 2018

IBM

OUTLINE

1. Analisi use-case fornito da IBM Italia e «IRCCS Casa Sollievo della Sofferenza»
2. Scelta del dataset, estrazione e pre-processing dei dati
3. Creazione del modello usando un approccio di Auto ML
4. Creazione del modello tramite Transfer Learning
5. Presentazione dei risultati e sviluppi futuri

MULTIDIMENSIONAL PROGNOSTIC INDEX

«Il Multidimensional Prognostic Index (MPI) è un indice prognostico di mortalità ad un anno basato su informazioni ottenute da una Valutazione Multidimensionale (VMD) del soggetto anziano»

MPI

- 8 domini, 63 elementi
- La conoscenza prognostica dell'outcome del paziente anziano può indirizzare o influenzare le scelte diagnostiche e terapeutiche.

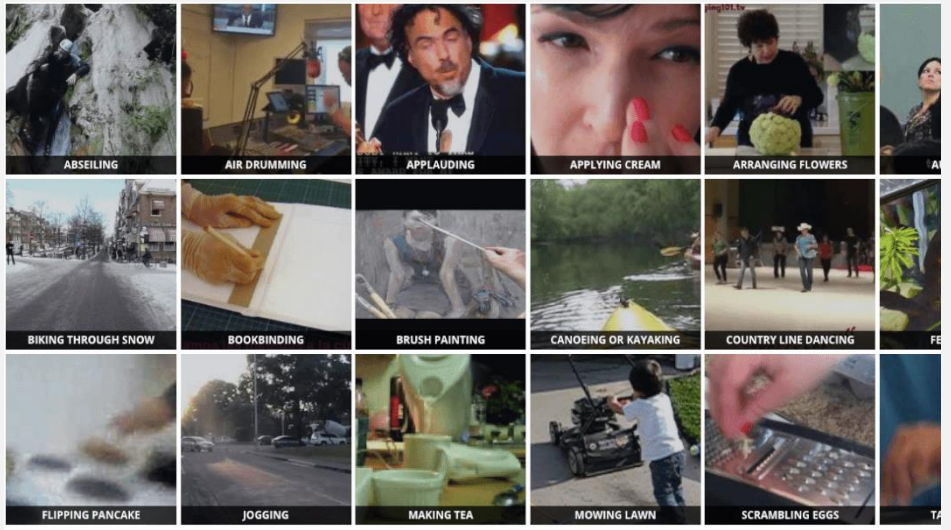
Activities of Daily Living

Le ADL sono attività di base eseguite durante la vita di un essere umano: lavarsi, mangiare, vestirsi, …

OBIETTIVI DEL PROGETTO

Progettare e sviluppare un modello di Deep Learning che sia in grado di riconoscere azioni svolte in un video contenente scene di vita quotidiana, svolte da un soggetto anziano all'interno dell'abitazione.

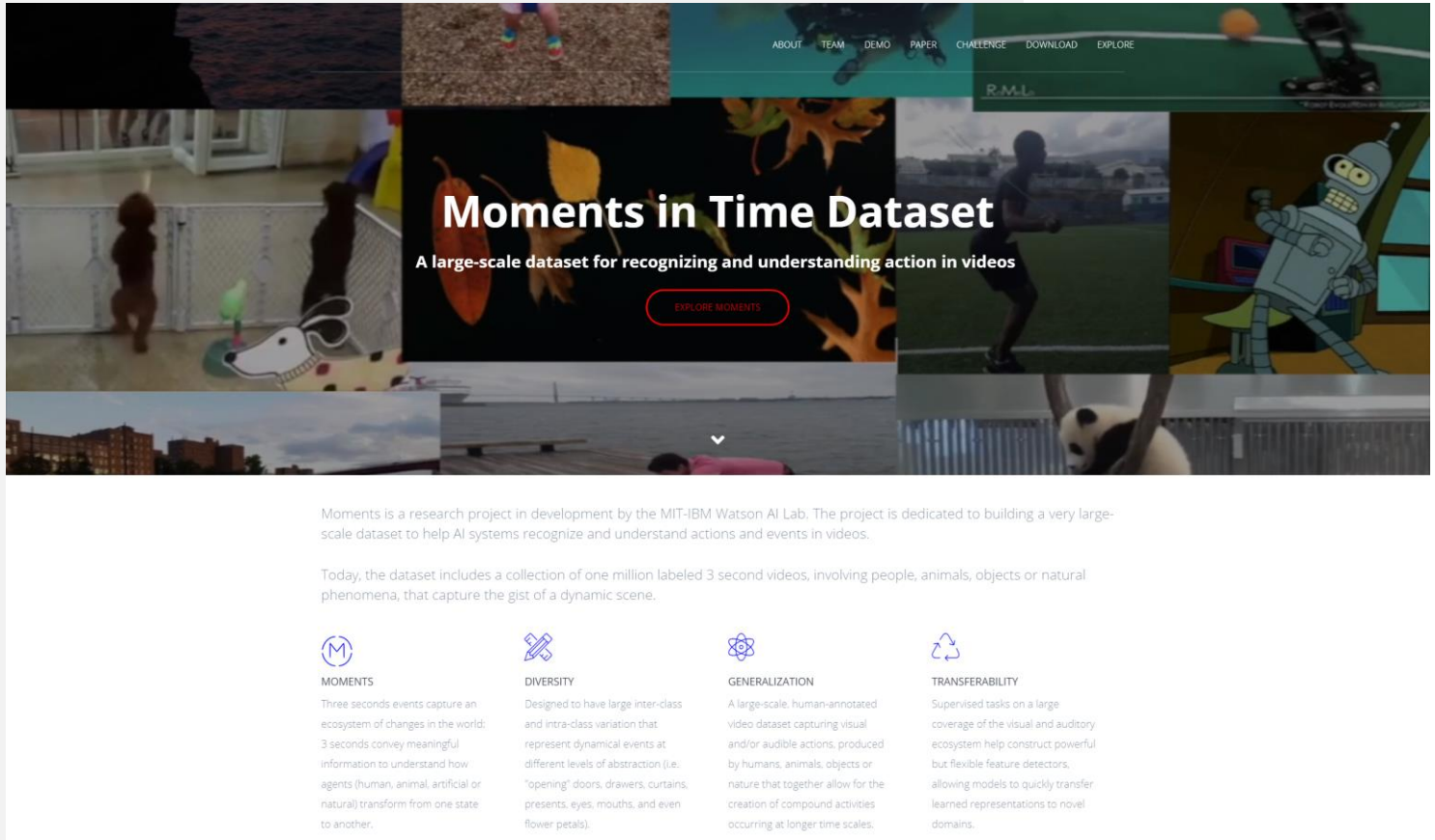
I risultati del modello sono a supporto della compilazione dell'ADL e IADL.



SCELTA DEI DATI: «MOMENTS IN TIME»

La generazione dei dati è molto onerosa in termini di tempo e costi

È stata eseguita un'analisi dei vari dataset disponibili per la Action Recognition per la ricerca di una fonte di dati pronta e validata.



Il dataset «Moments in Time» ha:

- 339 classi
- 1 milione di video di 3 secondi con label
- ampia variazione intra-classe dei video

DATA EXTRACTION & CLEANING - PIPELINE

	A	B	C	D	E	F	G
1	**CATEGORIES**					**CATEGORIES MERGED**	**CATEGORIES (final name)**
2	calling					calling/telephoning	telephoning
3	cooking					cooking	cooking
4	dining					dining/eating	eating
5	dressing					dressing	dressing
6	drinking					drinking	drinking
7	dusting					dusting/mopping/vacuuming	housekeeping
8	eating					reading/studying/writing	reading-writing
9	mopping					sewing	sewing
10	reading					smoking	smoking
11	sewing					socializing	socializing
12	smoking					standing	standing
13	socializing						
14	standing						
15	studying						
16	telephoning						
17	vacuuming						
18	writing						

Scelta delle classi (11)

Estrazione dei dati

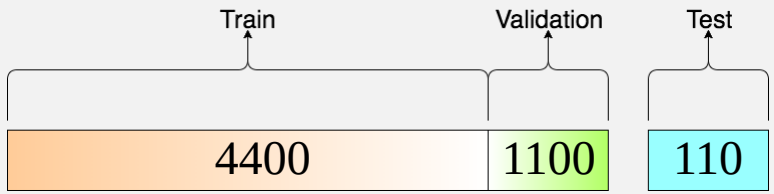
Pre-processing dei video

Estrazione frames dai video

Data cleaning manuale sulle immagini

Creazione del dataset,
500 immagini per classe

500 x 11 = 5500 immagini totali



Linguaggio utilizzato:

python

Interazione con i video:

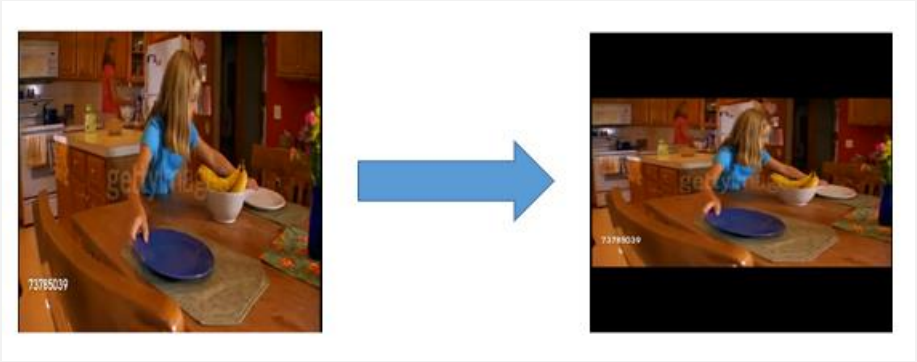
MoviePy

FFmpeg

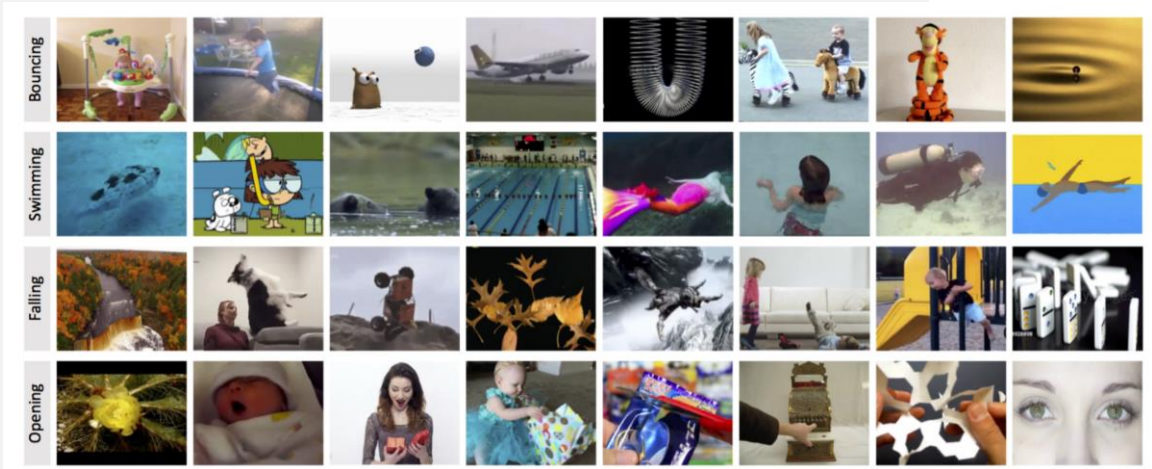
OpenCV

DATA PRE-PROCESSING

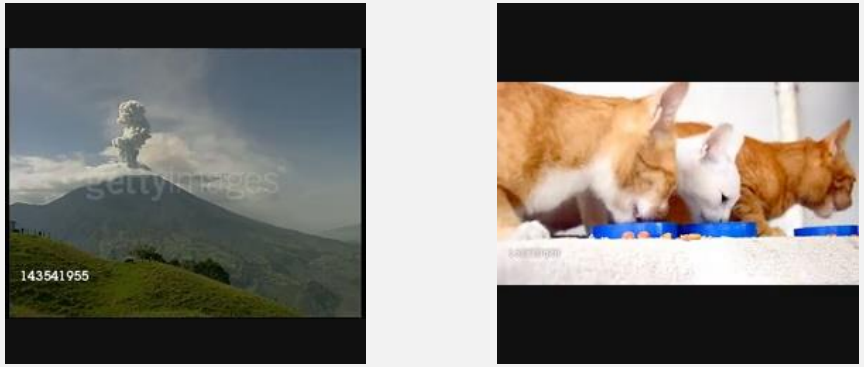
«La scarsa qualità dei dati è il nemico numero uno per l'uso diffuso e redditizio dell'apprendimento automatico.»



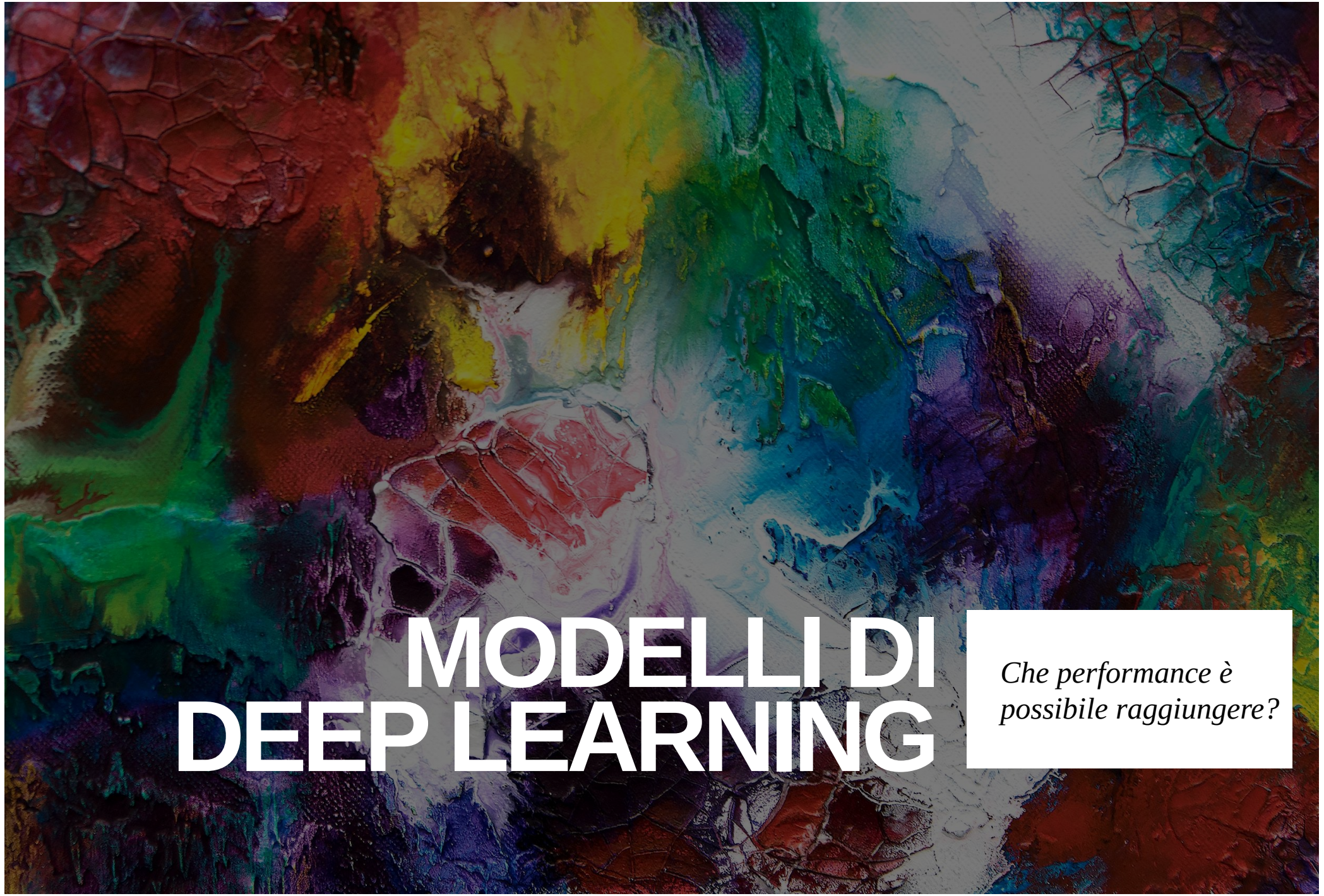
Garantire la correttezza dell'immagine



Interpretazione del nome dell'azione



Immagini che non rispecchiano il nostro caso



MODELLI DI DEEP LEARNING

Che performance è possibile raggiungere?

APPROCCIO AUTO ML - NEUNETS

Sintesi automatica di modelli di Reti Neurali

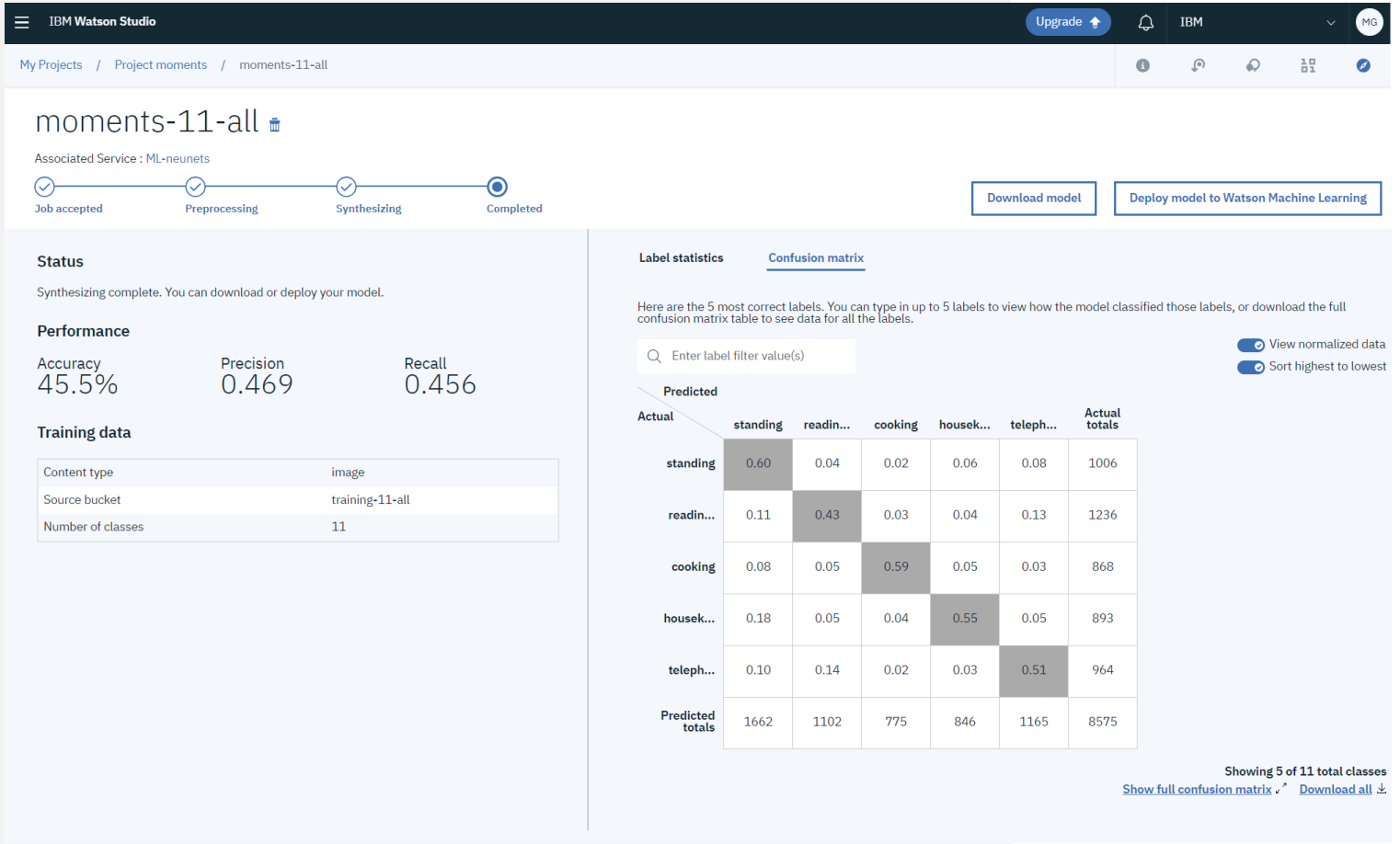
1. Forniti in input i dati

2. NeuNetS crea automaticamente la rete neurale più adatta per i dati forniti ed effettua il training

3. Risultati forniti in ~2 ore

Punti di forza:

- Accessibile gratuitamente in IBM Cloud
- Nessuna riga di codice per avere un primo risultato



MODELLO ALLENATO SU «MOMENTS IN TIME»

Modello di partenza per effettuare il Transfer Learning

Top-5 Actions:

31.5% -> dining

4.0% -> feeding

3.9% -> eating

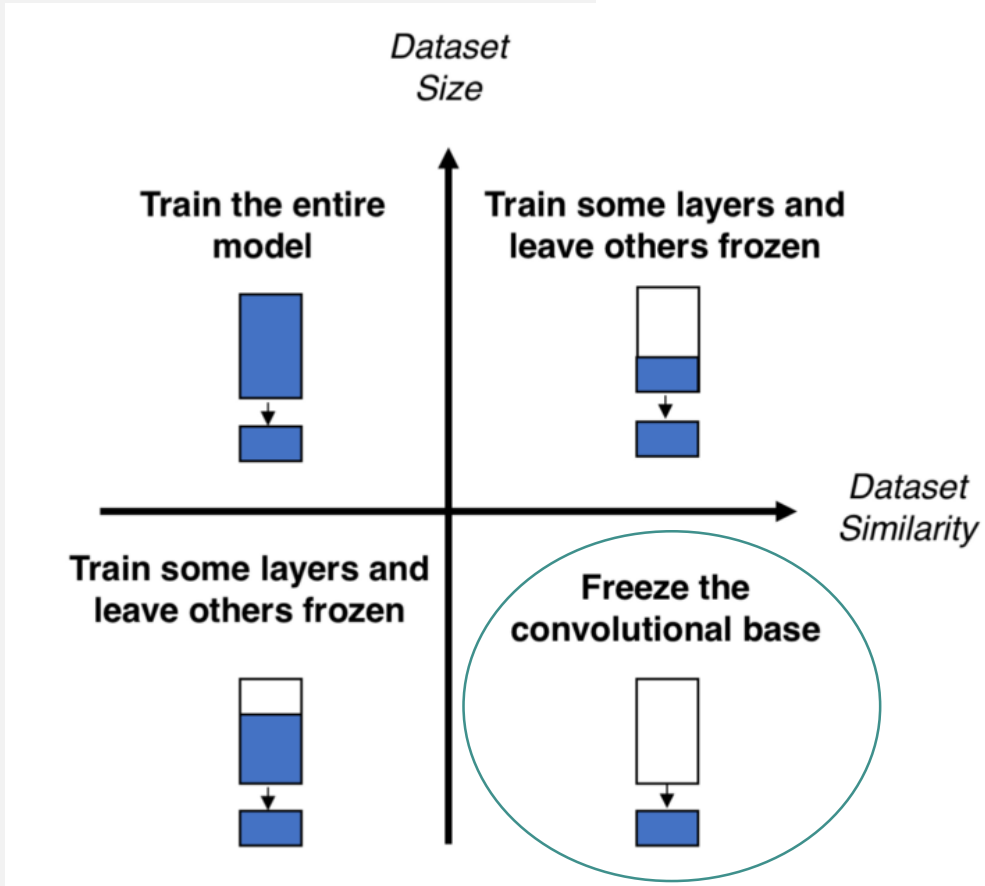
2.8% -> serving

2.6% -> socializing

Class Activation Mapping

Model	Modality	Top-1 (%)	Top-5 (%)
Chance	-	0.29	1.47
ResNet50-scratch	Spatial	23.65	46.73
ResNet50-Places	Spatial	26.44	50.56
ResNet50-ImageNet	Spatial	27.16	51.68
TSN-Spatial	Spatial	24.11	49.10
BNInception-Flow	Temporal	11.60	27.40
TSN-Flow	Temporal	15.71	34.65
SoundNet	Auditory	7.60	18.00
TSN-2stream	Spatial+Temporal	25.32	50.10
TRN-Multiscale	Spatial+Temporal	28.27	53.87
I3D	Spatial+Temporal	29.51	56.06
Ensemble (SVM)	S+T+A	31.16	57.67

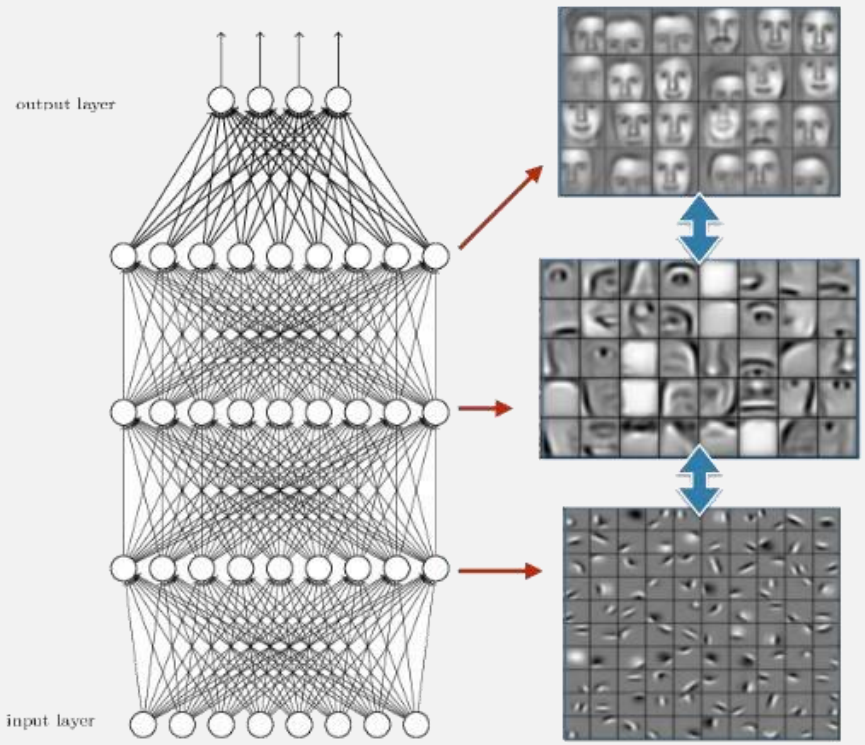
TABLE 1: **Classification Accuracy:** We show Top-1 and Top-5 accuracy of the baseline models on the validation set.

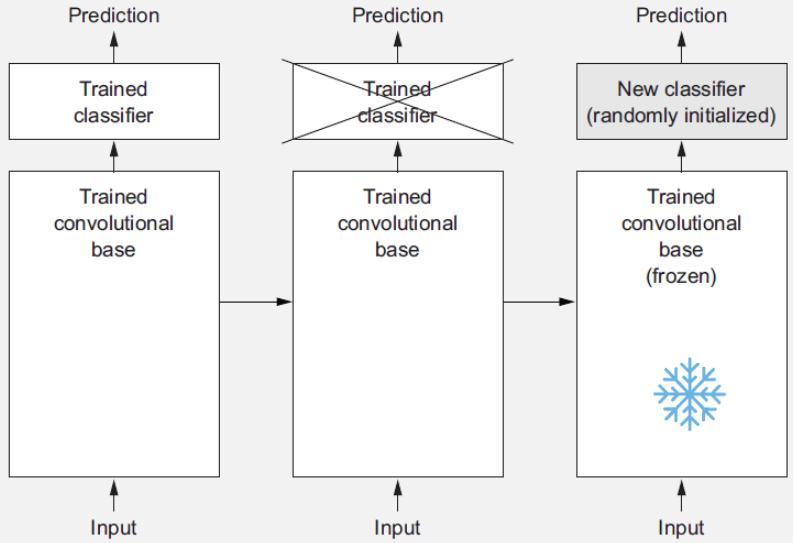


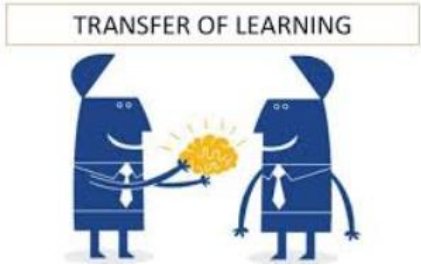
TRANSFER LEARNING

"Transfer learning is a popular method in computer vision because it allows us to build accurate models in a timesaving way" (Rawat & Wang, 2017).

"The applications of skills, knowledge and/or attitudes that were learned in one situation to another learning situation (Perkins, 1992)"







Training su Kaggle Kernel:

- GPU NVIDIA Tesla K80
- ~ 1 minuto ogni epoca
  (un iterazione su tutto il dataset)

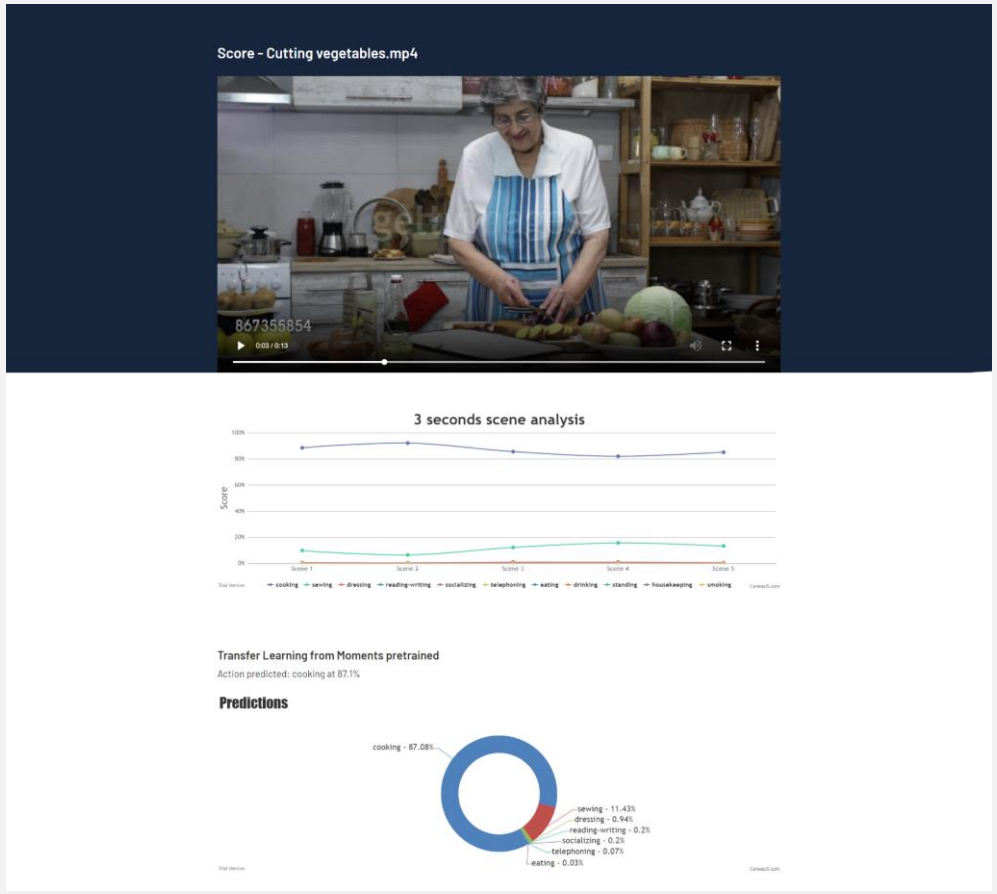
PYTORCH

kaggle

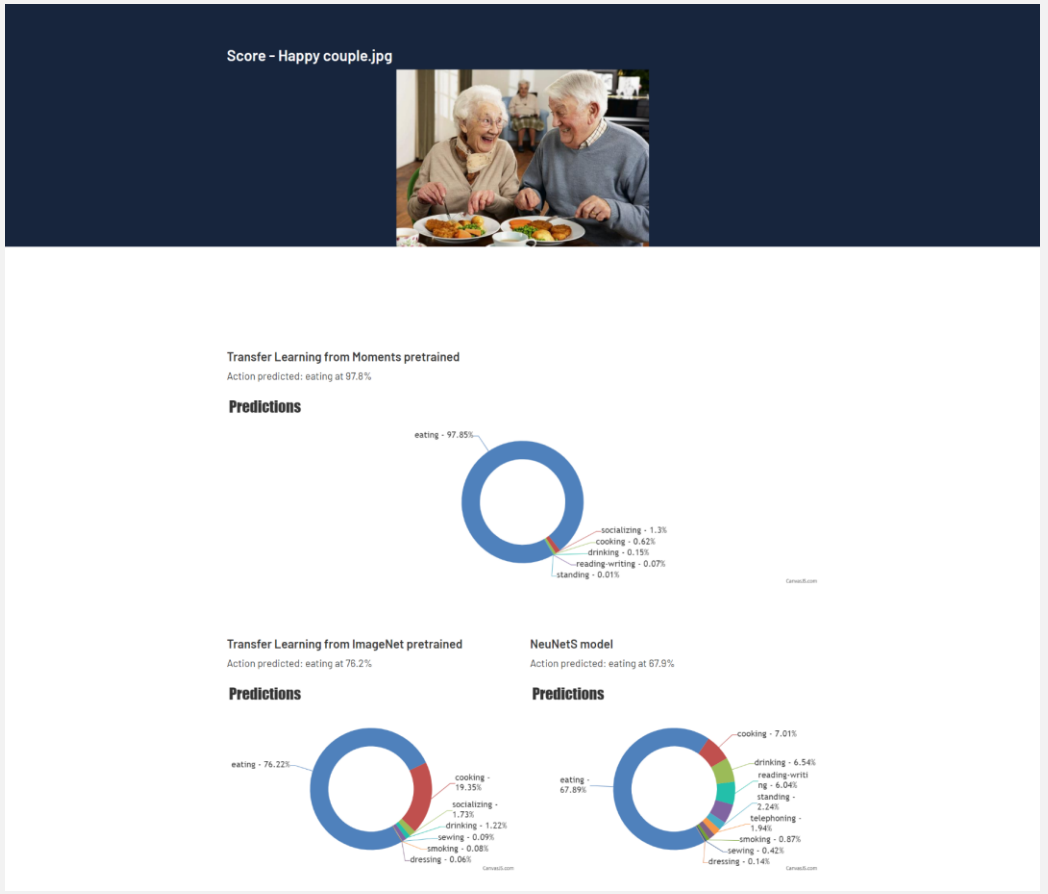
RISULTATI SUL DATASET DI TEST (110 ELEMENTI)

	NeuNetS su 500 immagini per classe	NeuNetS su tutte le immagini di una classe	Transfer learning sul modello Moments 	Transfer learning sul modello ImageNet
Accuracy	18.18 %	36.36%	64.55%	48.18%
Cooking	70.0 %	70.0 %	90.0 %	90.0 %
Dressing	0.0 %	0.0 %	60.0 %	50.0 %
Drinking	20.0 %	70.0 %	100.0 %	70.0 %
Eating	10.0 %	50.0 %	80.0 %	50.0 %
Housekeeping	50.0 %	80.0 %	90.0 %	100.0 %
Reading-writing	10.0 %	40.0 %	60.0 %	50.0 %
Sewing	0.0 %	0.0 %	50.0 %	20.0 %
Smoking	30.0 %	20.0 %	30.0 %	10.0 %
Socializing	0.0 %	30.0 %	50.0 %	20.0 %
Standing	0.0 %	0.0 %	0.0 %	0.0 %
Telephoning	10.0 %	40.0 %	100.0 %	70.0 %

WEB-APP: DEMO FOR MODEL INTERACTION



Test modello su video
(evoluzione temporale delle scene)



Test modelli su immagine

IBM Cloud

SVILUPPI FUTURI

- Da Multi-Class a Multi-Label (concorrenza azioni)
- Integrare la componente audio e quella temporale al modello (solo componente spaziale)
- Migliorare l’accuracy del modello collezionando più dati e lavorando sui parametri che influiscono sul training del modello
- Integrare soluzioni di Object Detection per identificare gli oggetti nella scena
- Estendere il numero di azioni riconoscibili
- Cominciare una fase sperimentale per poter raccogliere feedback e dati sul campo

GRAZIE

Matteo Gabrielli 🔗 *https://github.com/mattegab13*