Prof. Michele Nappi

BABELE

Riconoscimento lingua parlata tramite video senza audio

Corso di Fondamenti Di
Visione Artificiale e Biometria

Gruppo 24

Il Team

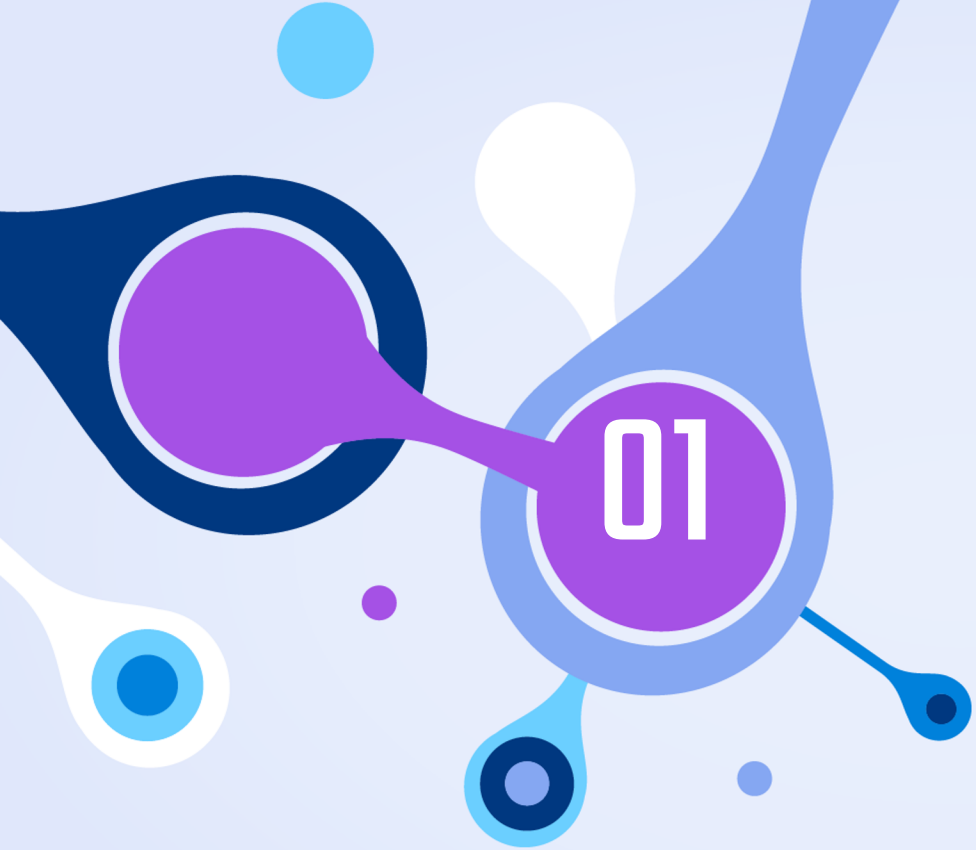
Lucia Cascone

Francesco Paciello

Luca Boffa

Vincenzo Di Leo

Indice della presentazione



IL PROBLEMA

Cos’è Babele?

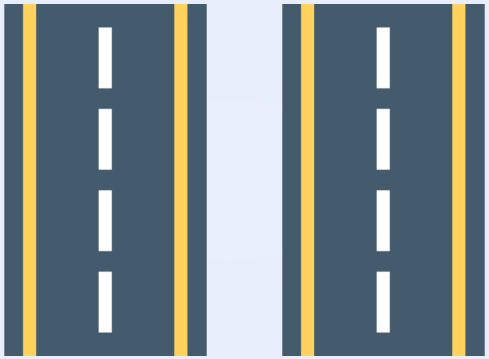
OBIETTIVO

Il progetto **BABELE** consiste nel creare un modello che sfrutta le **reti neurali artificiali** per riconoscere la **lingua parlata** da un soggetto, attraverso un video senza audio

L'approccio

Qual è stato il nostro approccio al Progetto Babele?

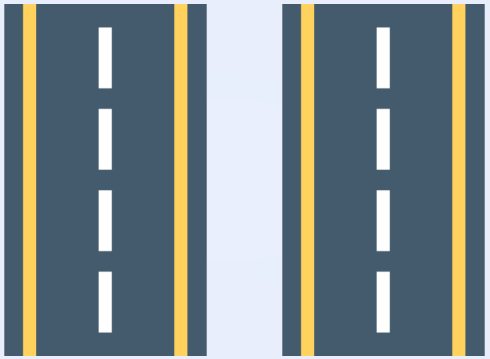
Due strade parallele



2D CNN 3D CNN

Due strade parallele

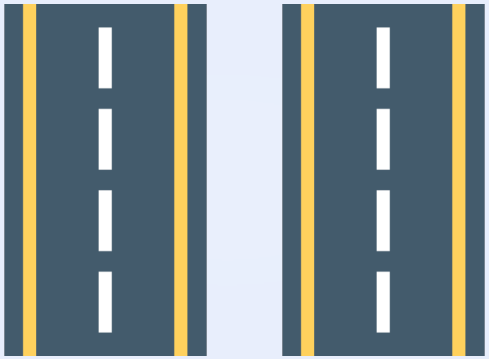
Non tengono traccia della linea temporale dei frames



2D CNN 3D CNN

Due strade parallele

Non tengono traccia della linea temporale dei frames



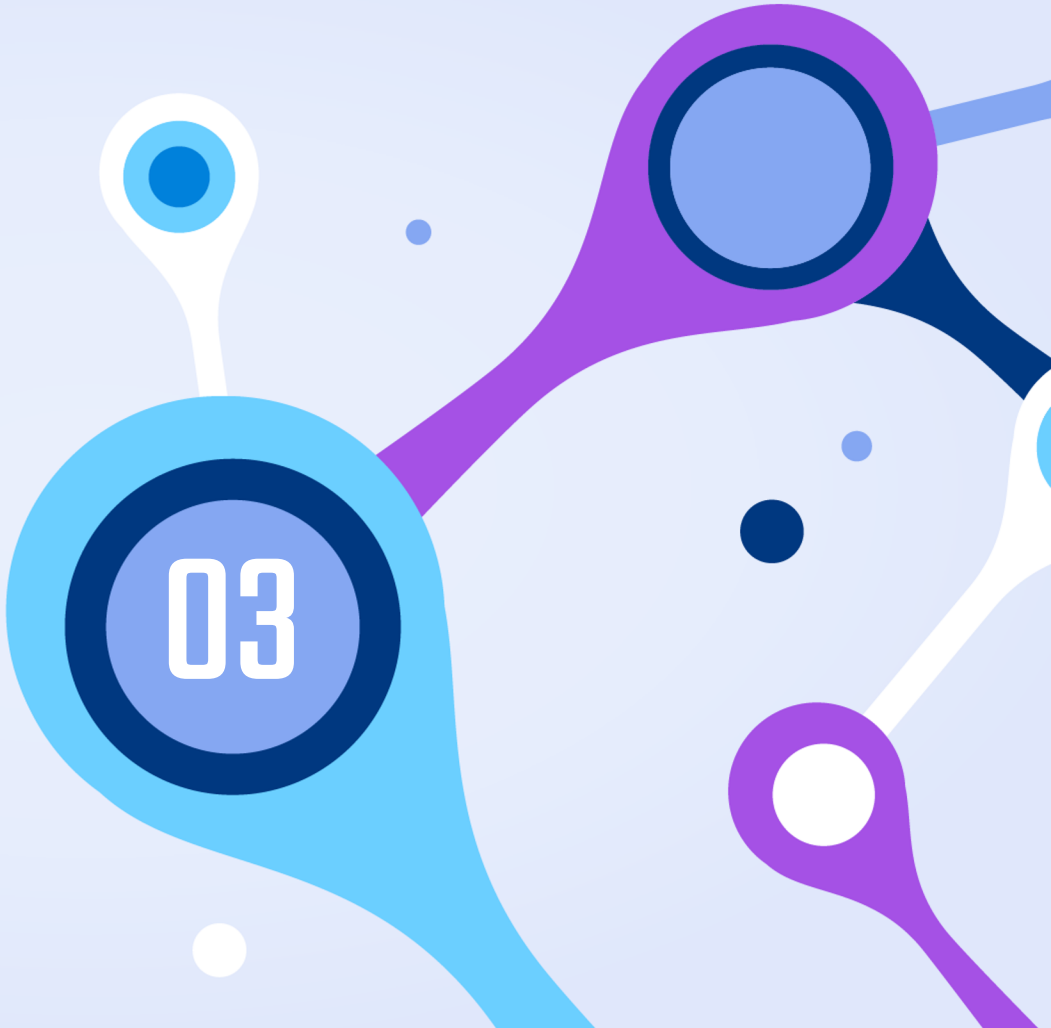
2D CNN

3D CNN

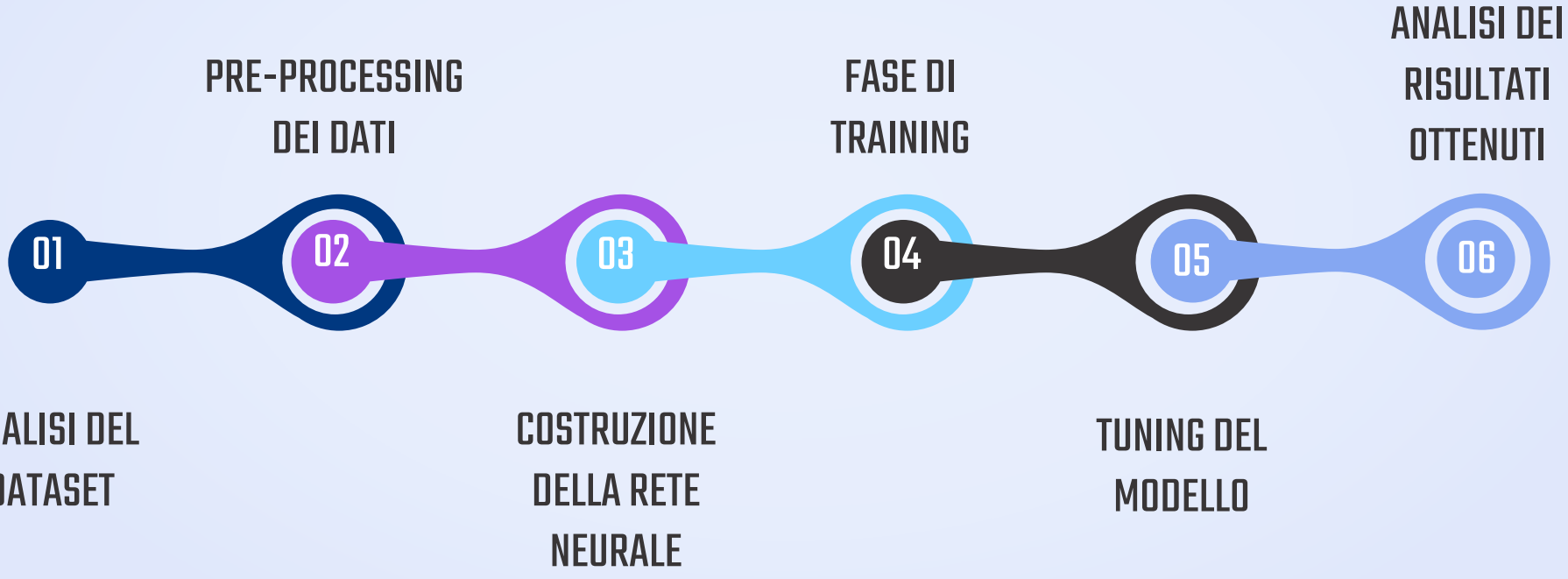
Tengono traccia della linea temporale dei frames

La pipeline

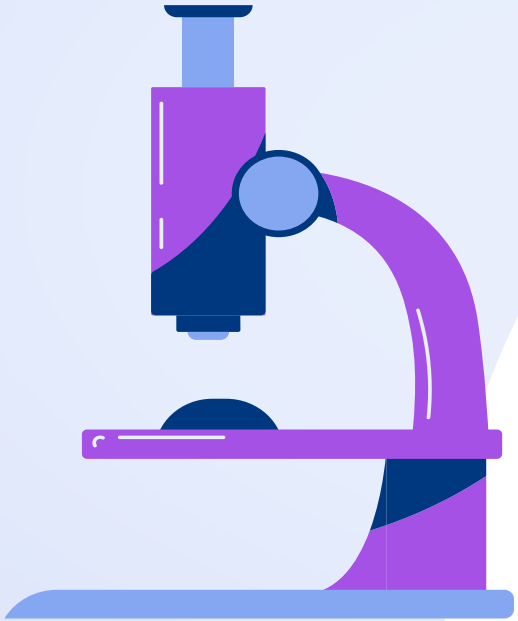
Insieme delle fasi del Progetto



La pipeline


01
ANALISI DEL DATASET
02
PRE-PROCESSING DEI DATI
03
COSTRUZIONE DELLA RETE NEURALE
04
FASE DI TRAINING
05
TUNING DEL MODELLO
06
ANALISI DEI RISULTATI OTTENUTI

ANALISI DEL DATASET



Prima scelta:

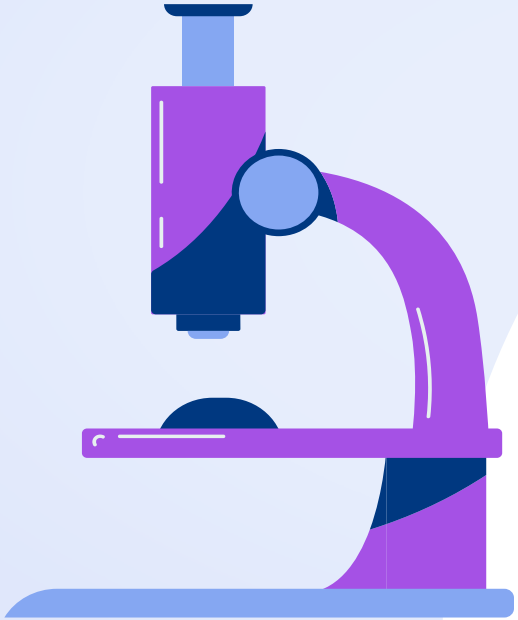
DEPENDENT

Stesse persone tra test e training set

INDEPENDENT

Persone **diverse** tra test e training set

ANALISI DEL DATASET



Scelta effettuata:

Independent

IDEA 💡

Utilizzare quello dependent avrebbe portato la rete ad associare la lingua parlata ad una persona, **non generalizzando!!**

ANALISI DEL DATASET

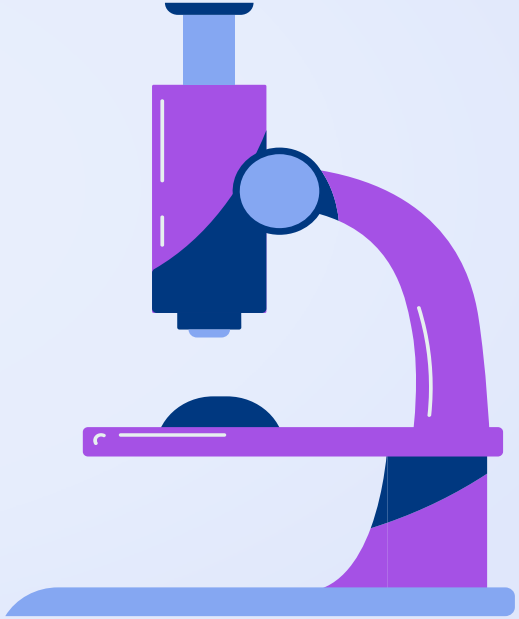
Seconda scelta:

Full videos

Video che mostrano **tutto il viso** della persona

Only lips

Video **ritagliati** che mostrano **solo la parte delle labbra**



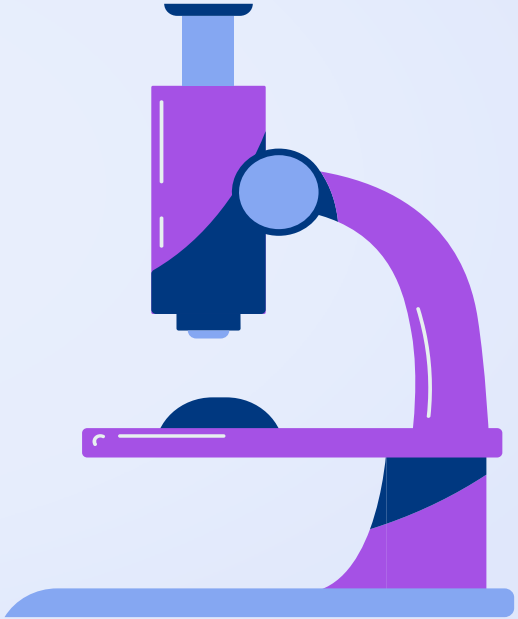
ANALISI DEL DATASET

Scelta effettuata:

Only lips

IDEA 💡

Utilizzare video di sole labbra avrebbe ridotto drasticamente l'area d'interesse, velocizzando anche la rete

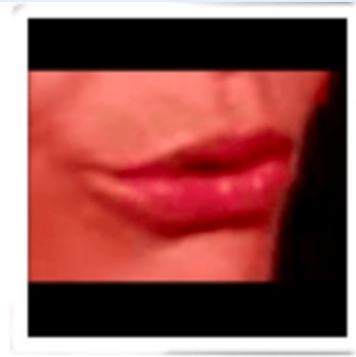


PRE-PROCESSING DEI DATI

- Abbiamo utilizzato la nomenclatura dei file per **dividerceli in cartelle in base alla loro label** (lingua)

- La divisione in train, test e validation set **era già stata effettuata**, quindi abbiamo saltato questo step

- Abbiamo utilizzato una classe chiamata **FrameGenerator** per **estrarre i frames dai video**



1_1_1_4_30_1_m

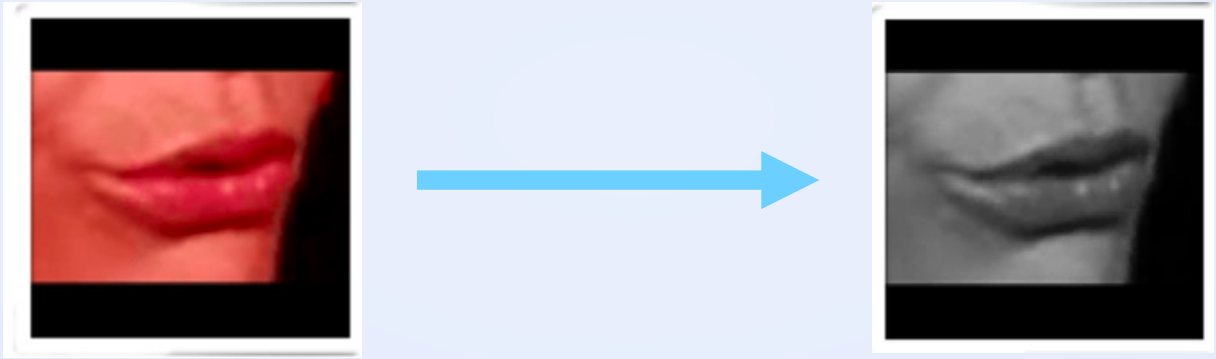
ESTRAZIONE DEI FRAMES

Con la classe **FrameGenerator** impostiamo due parametri:

- **n_frames**: indica il numero di frames da prendere per un singolo video
- **frame_step**: indica ogni quanti frames devo prenderne uno

PRE-PROCESSING DEI DATI

Abbiamo effettuato degli esperimenti
convertendo le immagini in **scala di grigi**



IDEA

Rendere la rete più leggera

PRE-PROCESSING DEI DATI

Risultati **non convincenti**, siamo tornati con i video a **colori** 👎🏼

Test accuracy (B&W)

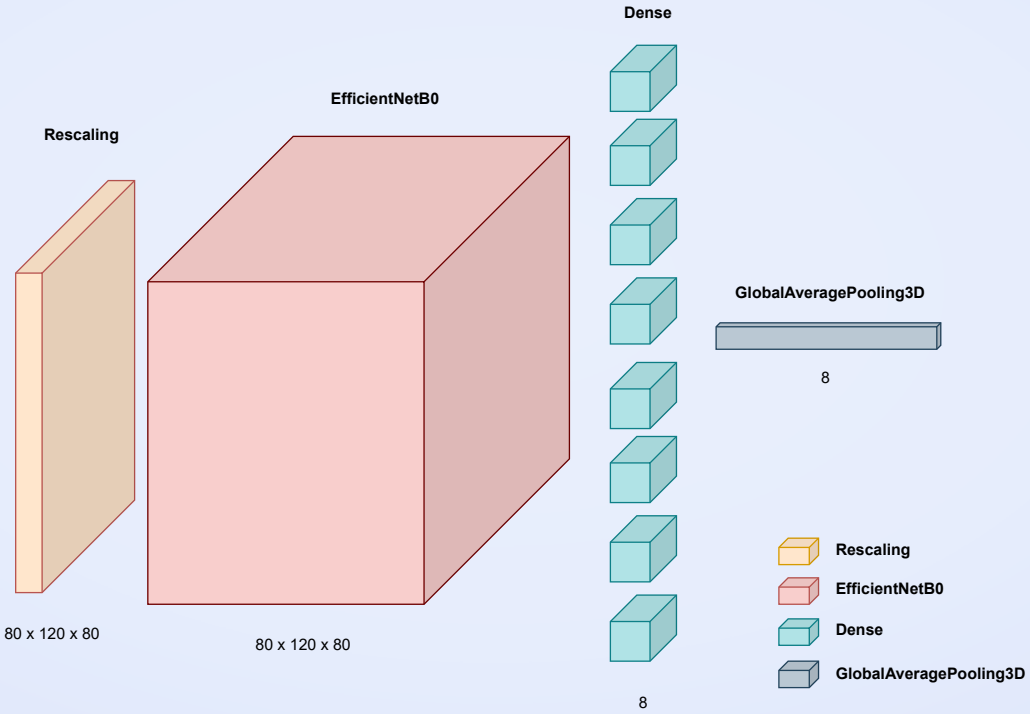
2D	3D
0.30	0.28

COSTRUZIONE DELLA RETE NEURALE

Struttura della rete 2D

COSTRUZIONE DELLA RETE NEURALE

Struttura della rete 2D

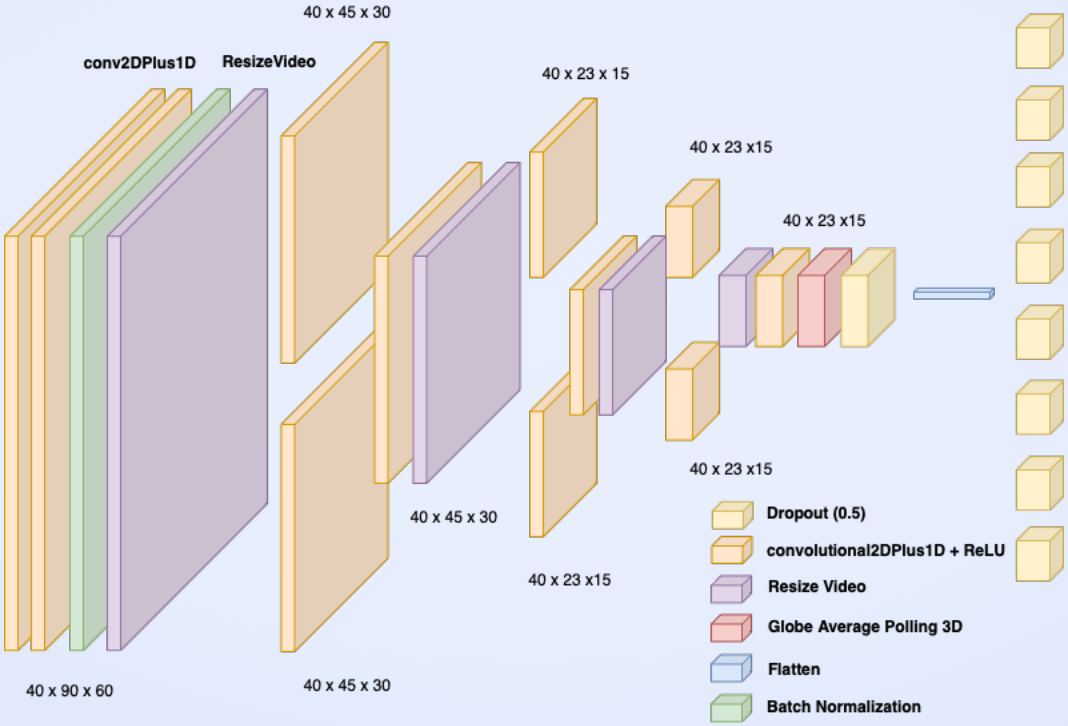


COSTRUZIONE DELLA RETE NEURALE

Struttura della rete 3D

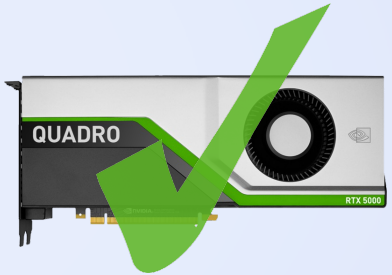
COSTRUZIONE DELLA RETE NEURALE

Struttura della rete 3D



FASE DI TRAINING

Inizialmente abbiamo effettuato la fase di training sulla piattaforma in cloud **Google Colab** ma abbiamo avuto **forti limitazioni**.

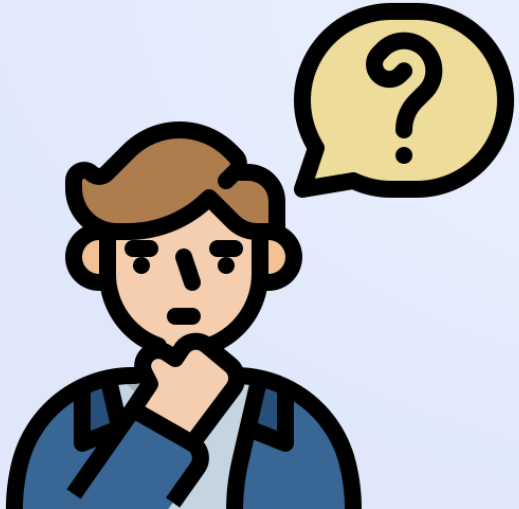


TUNING DEL MODELLO

Durante questa fase abbiamo effettuato varie prove modificando sia la rete che i suoi parametri:

- Image dimension
- frame_step
- batch_size
- Shuffle

IMAGE DIMENSION INFLUENCE



Domanda

Cosa succede se faccio variare la **dimensione dell'immagine** in input?

IMAGE DIMENSION INFLUENCE

2D CNN

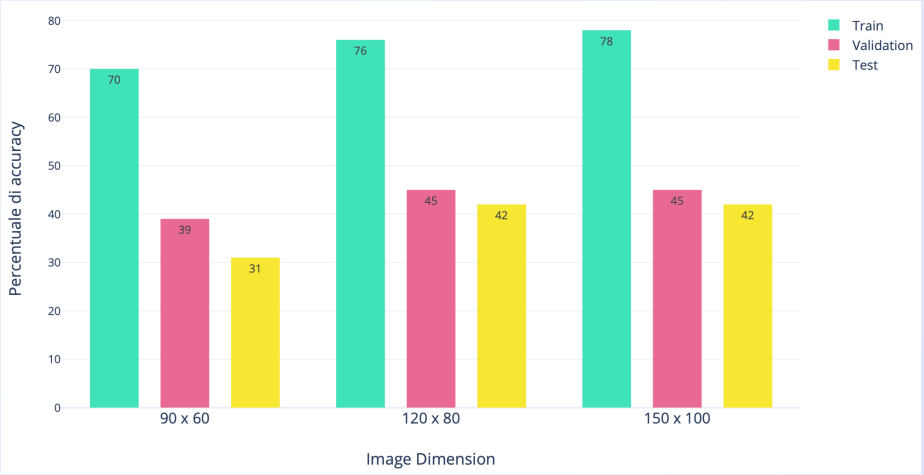


IMAGE DIMENSION INFLUENCE

2D CNN

Miglioramenti passando da **90 x 60** a **120 x 80**, ma aumentando ancora la dimensione non ce ne erano di significativi, quindi abbiamo optato per **120 x 6080**

Nota: non abbiamo scelto 150 x 100 perché volevamo ridurre il carico computazionale

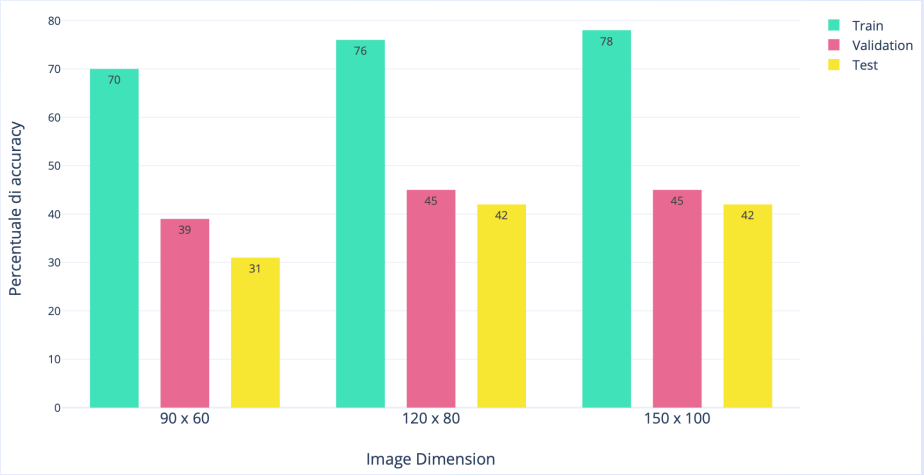


IMAGE DIMENSION INFLUENCE

3D CNN

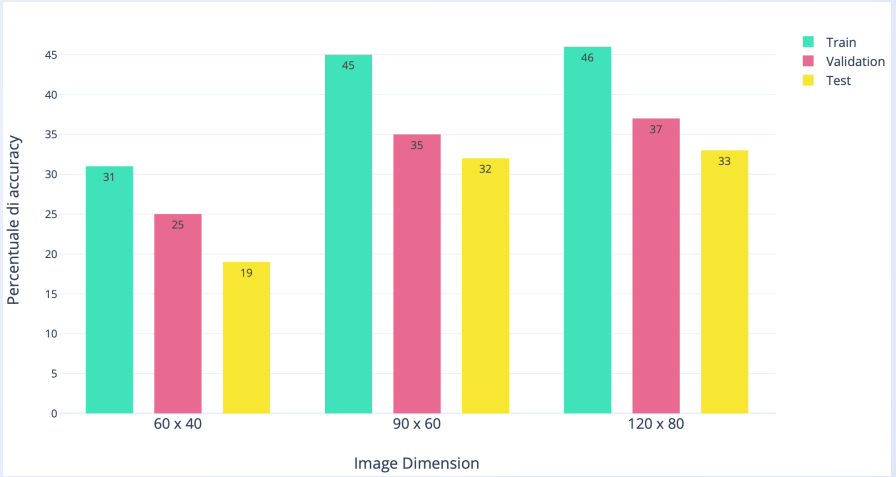
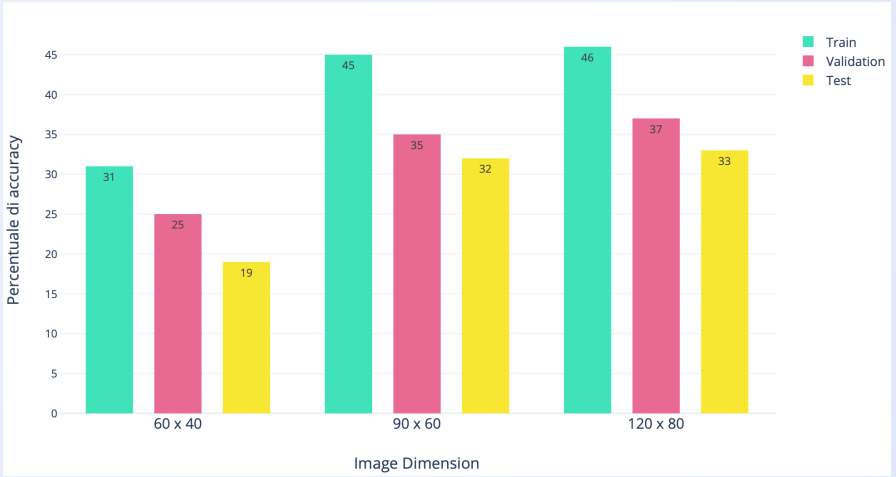


IMAGE DIMENSION INFLUENCE

3D CNN

Stesso discorso, miglioramenti da **60 x 40** a **90 x 60**, ma attenuazioni quando salivamo a **120 x 80**, in questo caso abbiamo scelto **90x60**

Nota: non abbiamo scelto 120 x 80 perché volevamo ridurre il carico computazionale



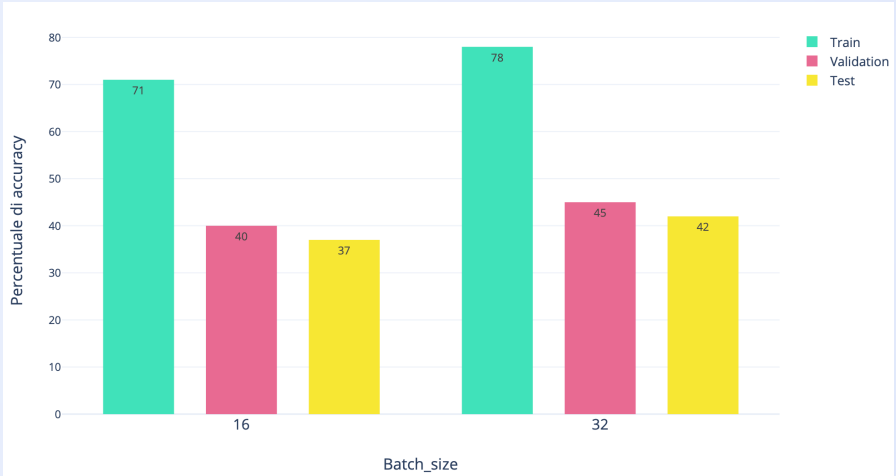
BATCH SIZE INFLUENCE

Domanda

Cosa succede se faccio variare la **batch size?**

BATCH SIZE INFLUENCE

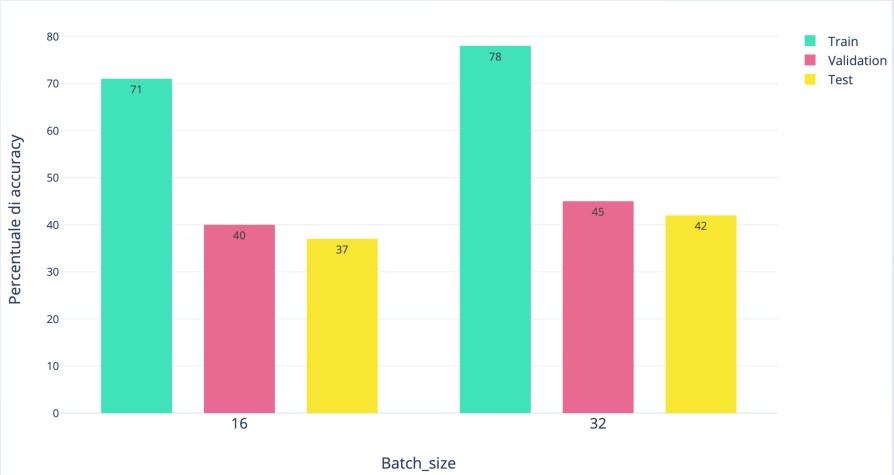
2D CNN



BATCH SIZE INFLUENCE

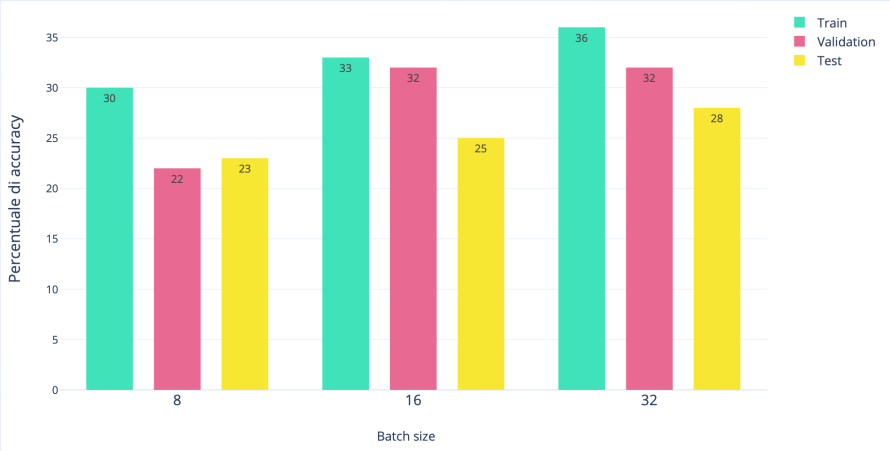
2D CNN

La **batch_size impostata a 32** ha portato a risultati migliori, quindi abbiamo sperimentato con questo valore



BATCH SIZE INFLUENCE

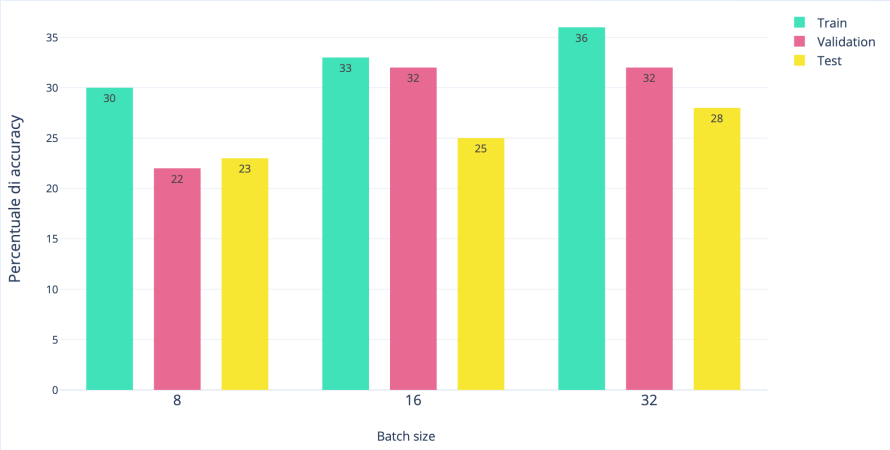
3D CNN



BATCH SIZE INFLUENCE

3D CNN

Anche per le 3D CNN si è rivelato vantaggioso impostare il parametro **batch_size a 32**



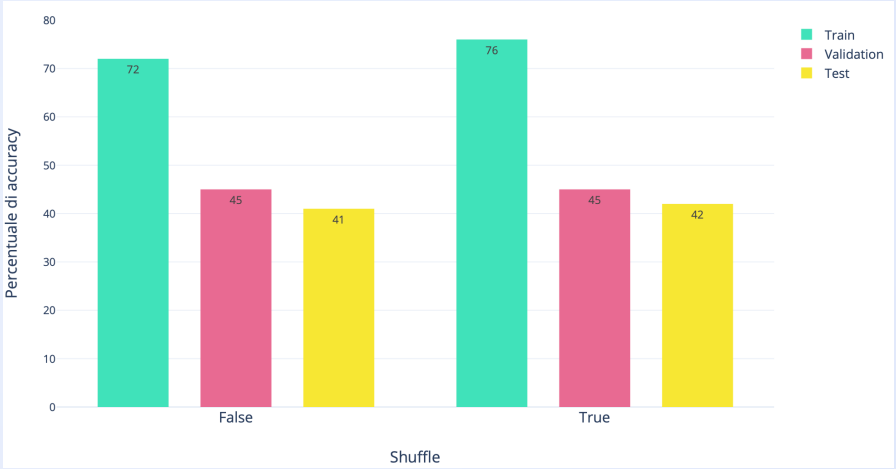
SHUFFLE INFLUENCE

Domanda

Cosa succede se **randomizzo** il dataset durante l'addestramento?

SHUFFLE INFLUENCE

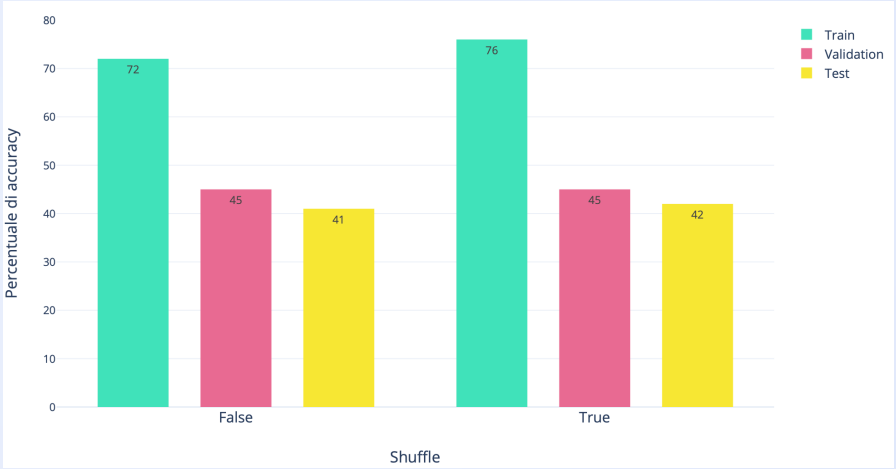
2D CNN



SHUFFLE INFLUENCE

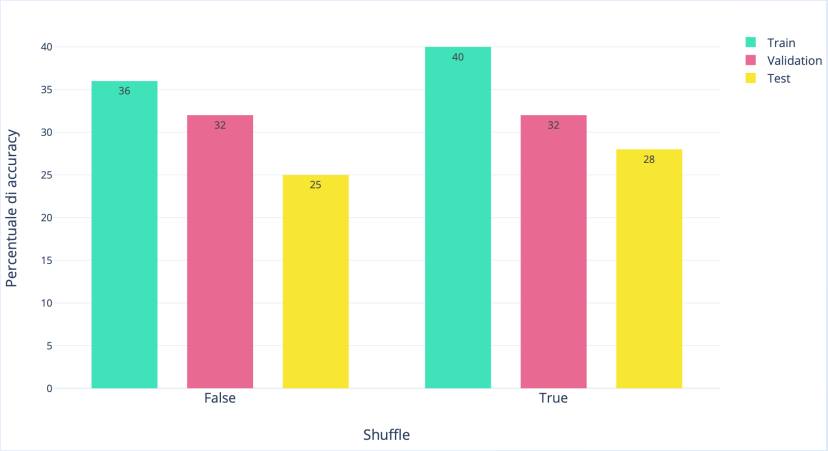
2D CNN

Alternando lo shuffle, tra **true** e **false,** si è potuto notare come quest'ultimo **non porti un notevole impatto** sui risultati della rete



SHUFFLE INFLUENCE

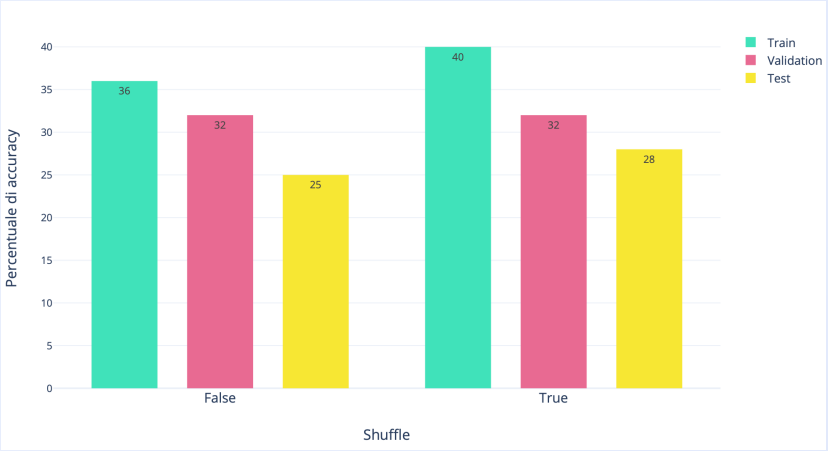
3D CNN



SHUFFLE INFLUENCE

3D CNN

In questo caso, il parametro **shuffle = true,** portava a **risultati leggermente migliori** rispetto ad averlo impostato a **false**



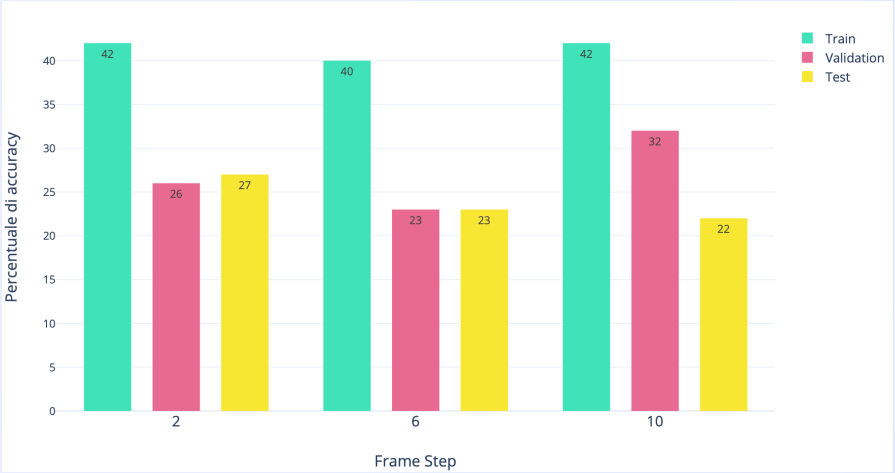
FRAME STEP INFLUENCE

Domanda

Cosa succede se faccio variare il parametro **frame_step**?

FRAME STEP INFLUENCE

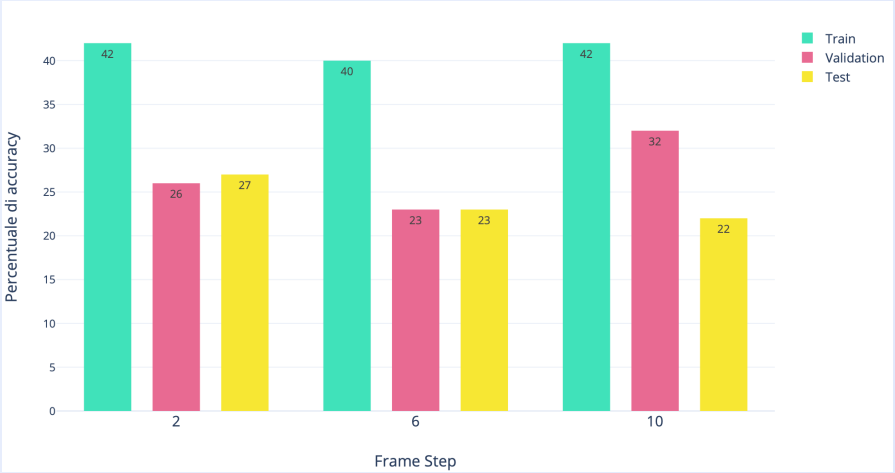
2D CNN



FRAME STEP INFLUENCE

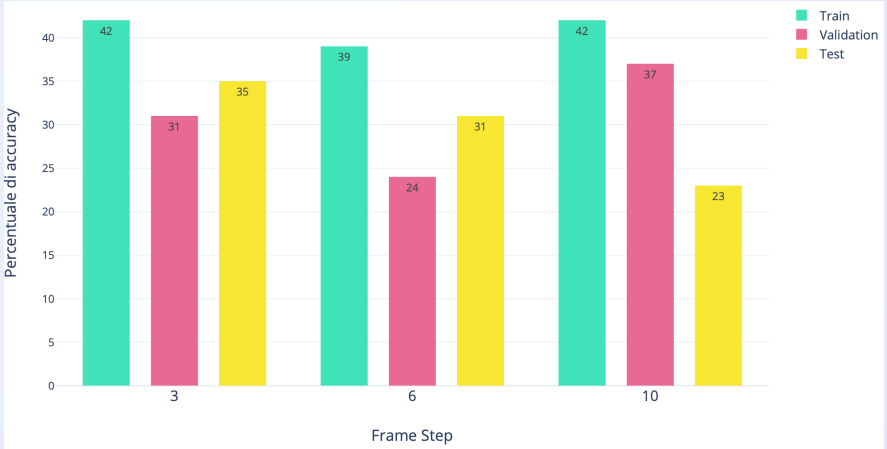
2D CNN

Impostando un frame_step relativamente **basso**, si ottengono **risultati migliori**, quindi abbiamo continuato con **frame_step = 2**



FRAME STEP INFLUENCE

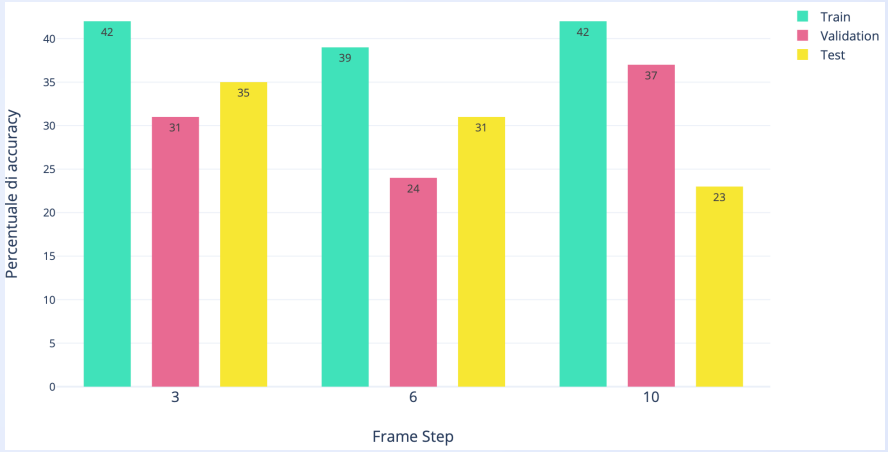
3D CNN



FRAME STEP INFLUENCE

3D CNN

Anche in questo caso, un frame_step basso portava a risultati migliori, per le 3D CNN è stato mantenuto un **frame_step = 3**



ANALISI DEI RISULTATI

RISULTATI 2D CNN

Migliore configurazione

Parametro	Valore
Image res.	80 x 120
Numero frames	60
Frame step	2
Batch size	32
Shuffle	True
Learning rate	Adam (0.0001)

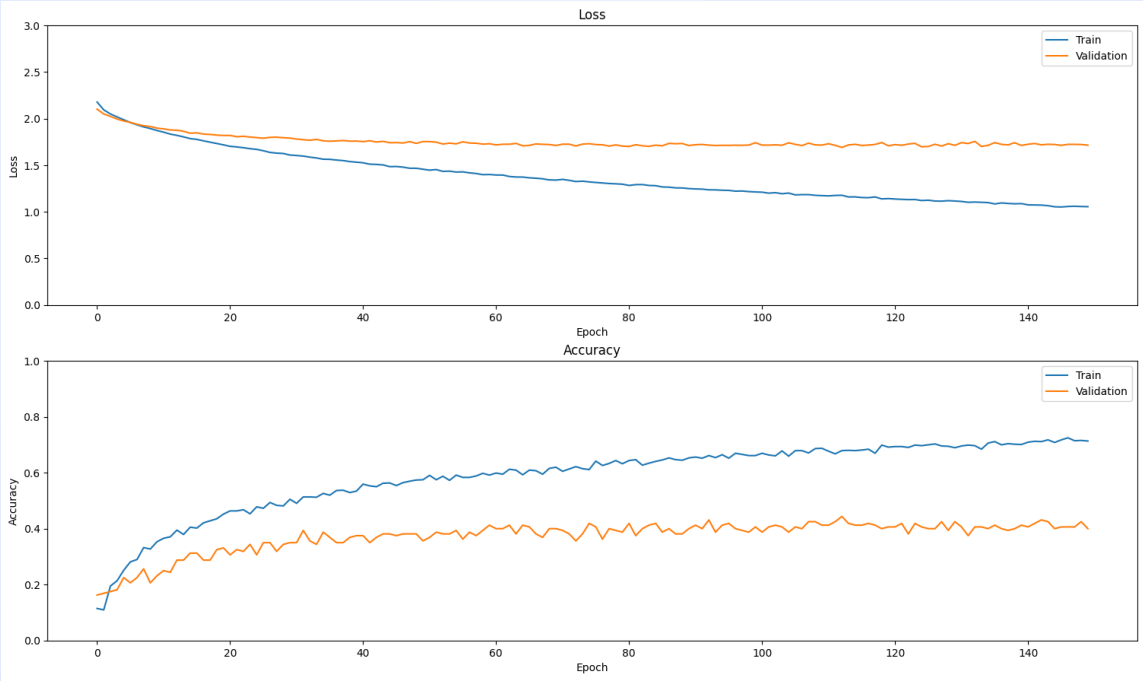
Classe	Precision	Recall
Italiano	0.45	0.45
Inglese	0.11	0.10
Tedesco	0.78	0.75
Spagnolo	0.30	0.20
Olandese	0.53	0.85
Russo	0.13	0.05
Giapponese	0.23	0.25
Francese	0.38	0.60

Test Accuracy

0.42

RISULTATI 2D CNN

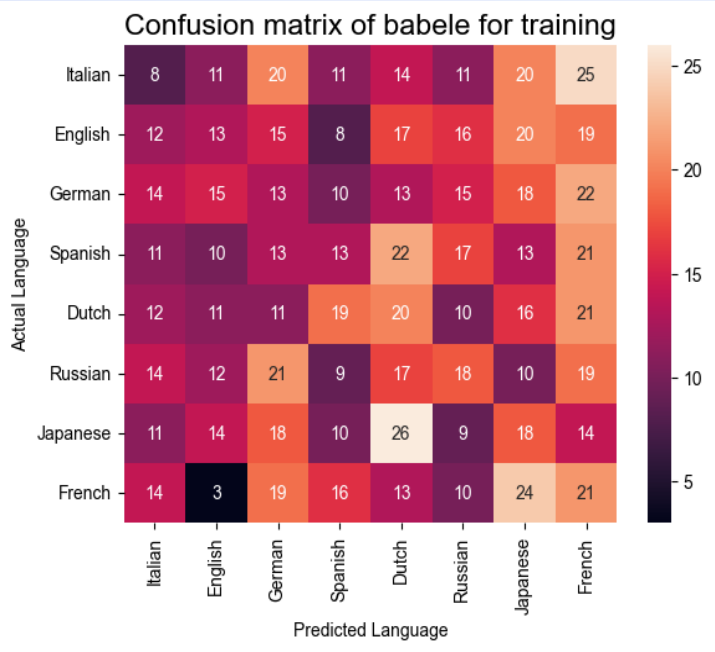
Accuracy e loss



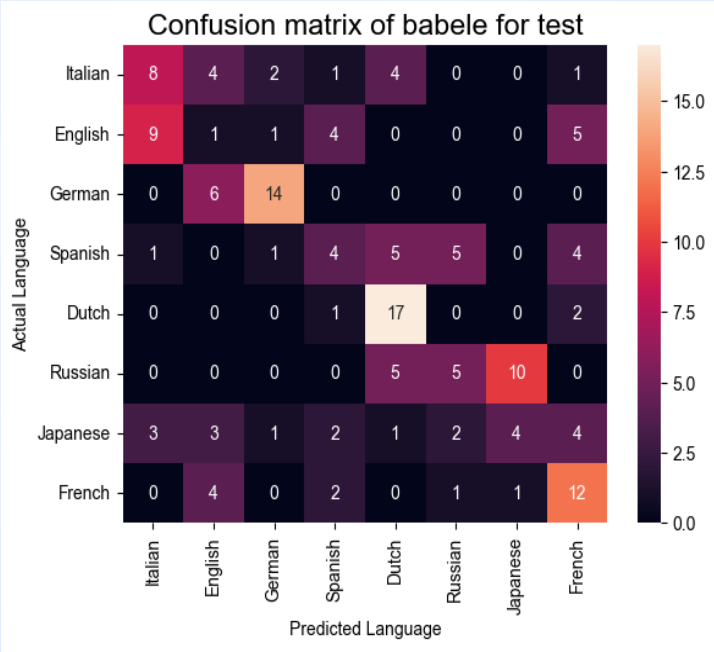
RISULTATI 2D CNN

Matrice di confusione

Training



Test



RISULTATI 3D CNN

Migliore configurazione

Parametro	Valore
Image res.	90 x 60
Numero frames	40
Frame step	3
Batch size	32
Shuffle	True
Learning rate	Adam (0.001)
Space conv.	7x7
Time conv.	3

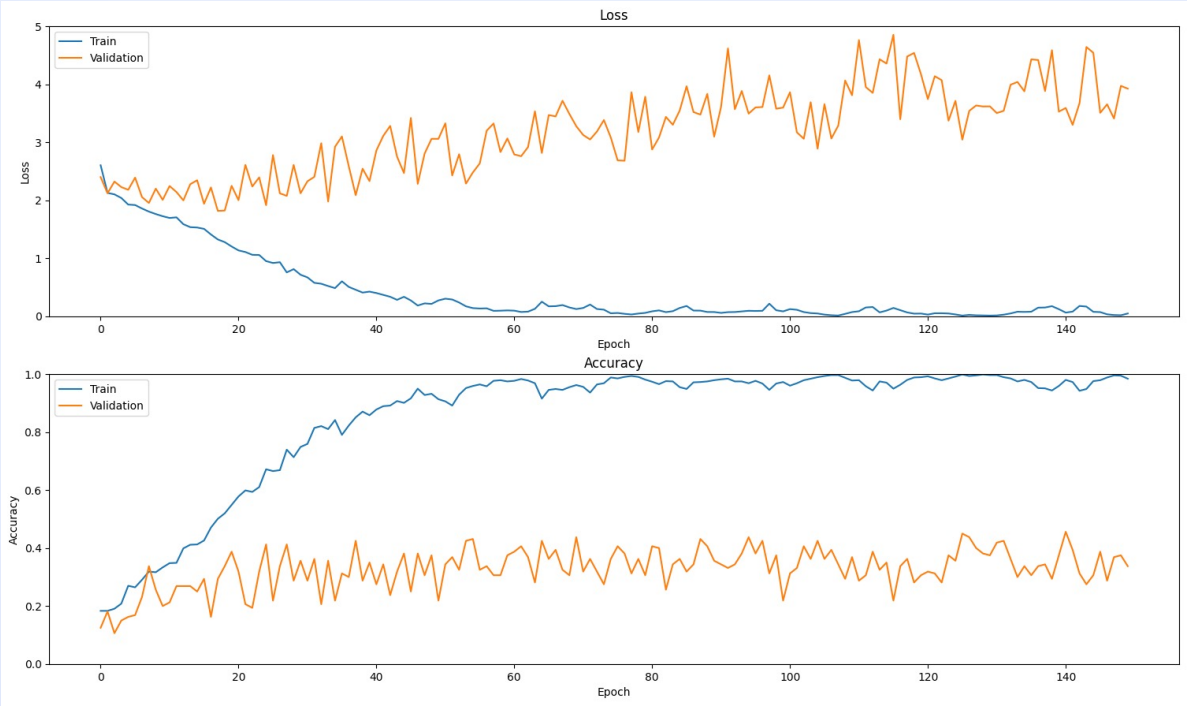
Classe	Precision	Recall
Italiano	0.38	0.25
Inglese	0.00	0.00
Tedesco	0.38	0.80
Spagnolo	0.13	0.25
Olandese	0.73	0.95
Russo	0.75	0.30
Giapponese	0.00	0.20
Francese	0.17	0.25

Test Accuracy

0.36

RISULTATI 3D CNN

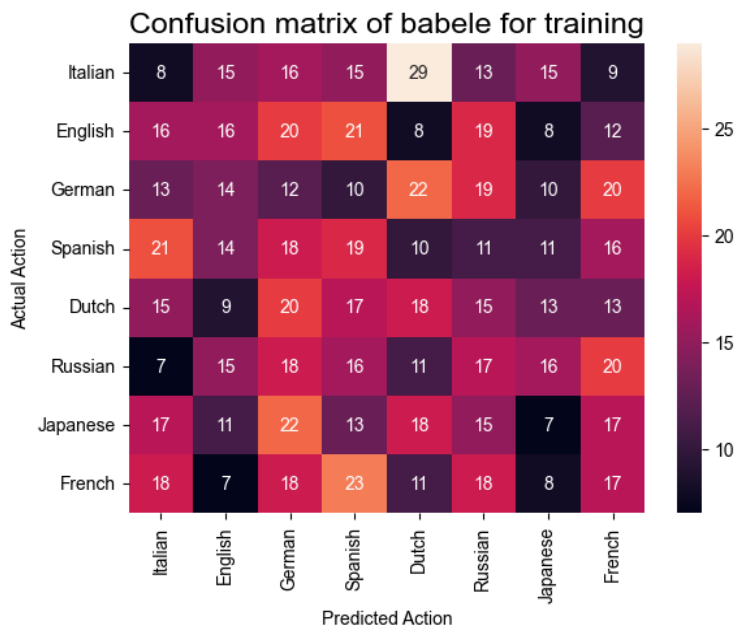
Accuracy e loss



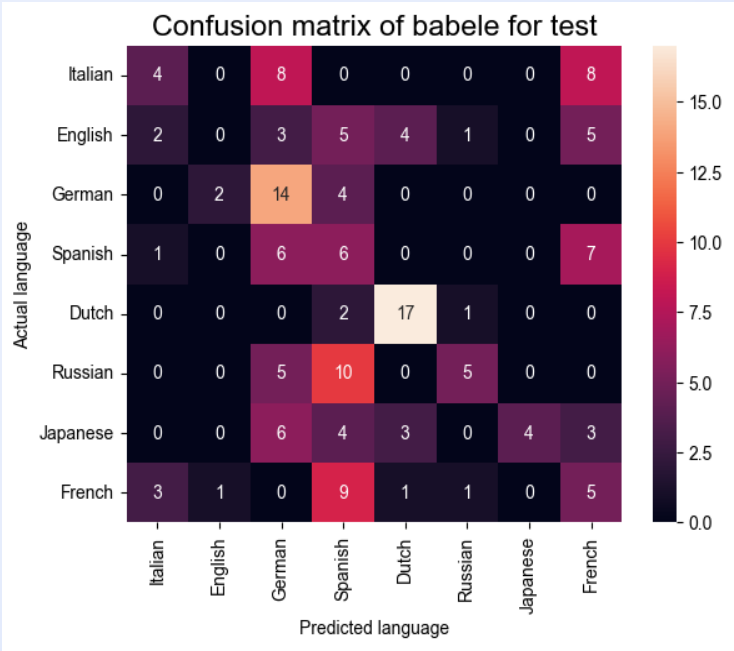
RISULTATI 3D CNN

Matrice di confusione

Training



Test



CONCLUSIONI

Conclusioni e possibili sviluppi futuri

CONCLUSIONI E VALUTAZIONI

Valutazione delle reti

La **rete 2D** ha avuto risultati migliori rispetto alla 3D

CONCLUSIONI E VALUTAZIONI

Valutazione delle reti

La **rete 2D** ha avuto risultati migliori rispetto alla 3D

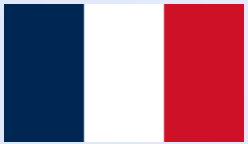
Analizzare i frame in modo **indipendente**, invece che effettuare delle **correlazione temporali** come fa la 3D CNN, potrebbe portare a risultati migliori poiche' il modello riesce ad **identificare dei pattern di immagini** all'interno dei singoli frames, estraendo le informazioni utili per identificare la lingua

CONCLUSIONI E VALUTAZIONI

Lingue riconosciute

Le **lingue** che sono state riconosciute **meglio** sono:

2D CNN



3D CNN


SVILUPPI FUTURI

In futuro si potrebbe ambire a **migliorare i risultati** di questo progetto, seguendo le seguenti **strategie**:

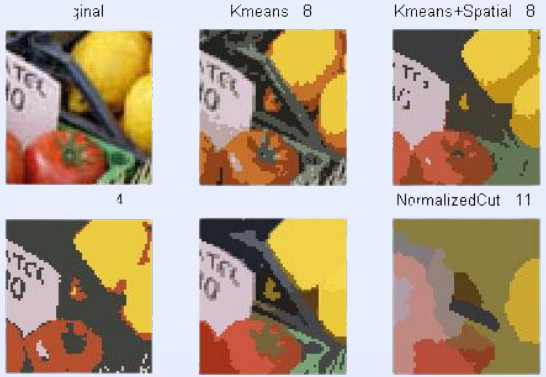
SVILUPPI FUTURI

In futuro si potrebbe ambire a **migliorare i risultati** di questo progetto, seguendo le seguenti **strategie**:

DATASET PIU' GRANDE

SVILUPPI FUTURI

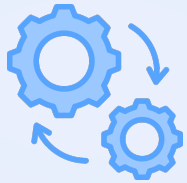
In futuro si potrebbe ambire a **migliorare i risultati** di questo progetto, seguendo le seguenti **strategie**:



MEAN SHIFT ALGORITHM

SVILUPPI FUTURI

In futuro si potrebbe ambire a **migliorare i risultati** di questo progetto, seguendo le seguenti **strategie**:



MIGLIORARE FASE DI PRE-PROCESSING

Grazie!

Se avete domande, chiedete pure

Francesco
Luca
Vincenzo