

# FIND THAT LOOK

## FASHION RECOMMENDATION

Digital Signal and Image Management - A.A. 2020/2021

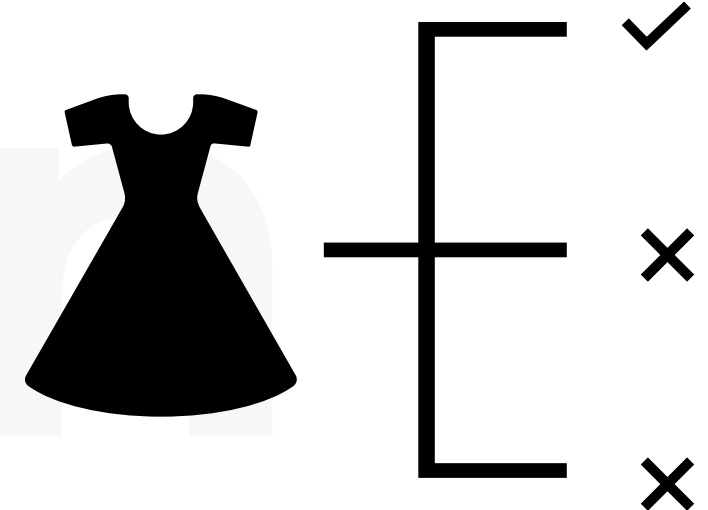
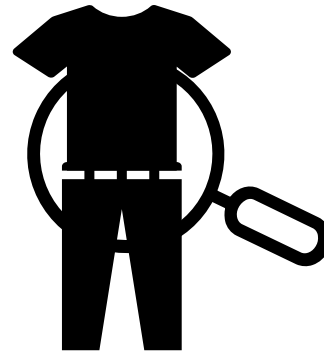
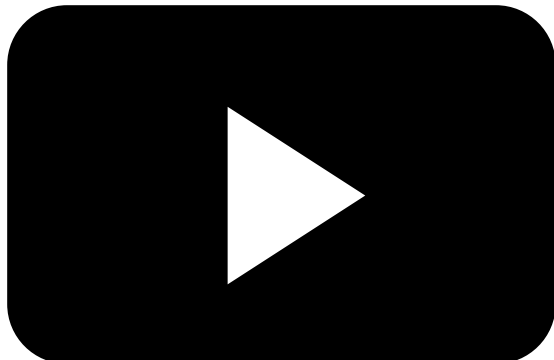


**MATTEO CAMPIRONI**

matricola 801850

**SERENA DI MAGGIO**

matricola 821063

- ① **Image Recognition** 
- ② **Image Retrieval** 
- ③ **Demo** 

# Image Recognition

# Obiettivo

L'obiettivo del task di Image Recognition è quello di sfruttare reti neurali convoluzionali per determinare a quale categoria appartengono i capi di abbigliamento che vengono forniti in input.



# Dataset

Il dataset considerato è stato costruito a partire da collezioni di immagini messe a disposizione su Kaggle:

- [Dataset 1](#)
- [Dataset 2](#)

Il risultato ottenuto è un dataset costituito da 10390 immagini, suddivise in 10 categorie in modo bilanciato.



DRESSES



JACKETS



SANDALS



SHOES



PANTS



SKIRTS



SHIRTS



SHORTS



SWEATERS



T-SHIRTS

# Data Augmentation

Sfruttando ImageDataGenerator di Keras, è stata applicata la tecnica di Data Augmentation durante l'addestramento dei modelli.

Le trasformazioni utilizzate sono le seguenti:

- brightness\_range,
- channel\_shift\_range
- vertical\_flip
- horizontal\_flip
- shear\_range
- zoom\_range
- rotation\_range
- width\_shift\_range
- height\_shift\_range

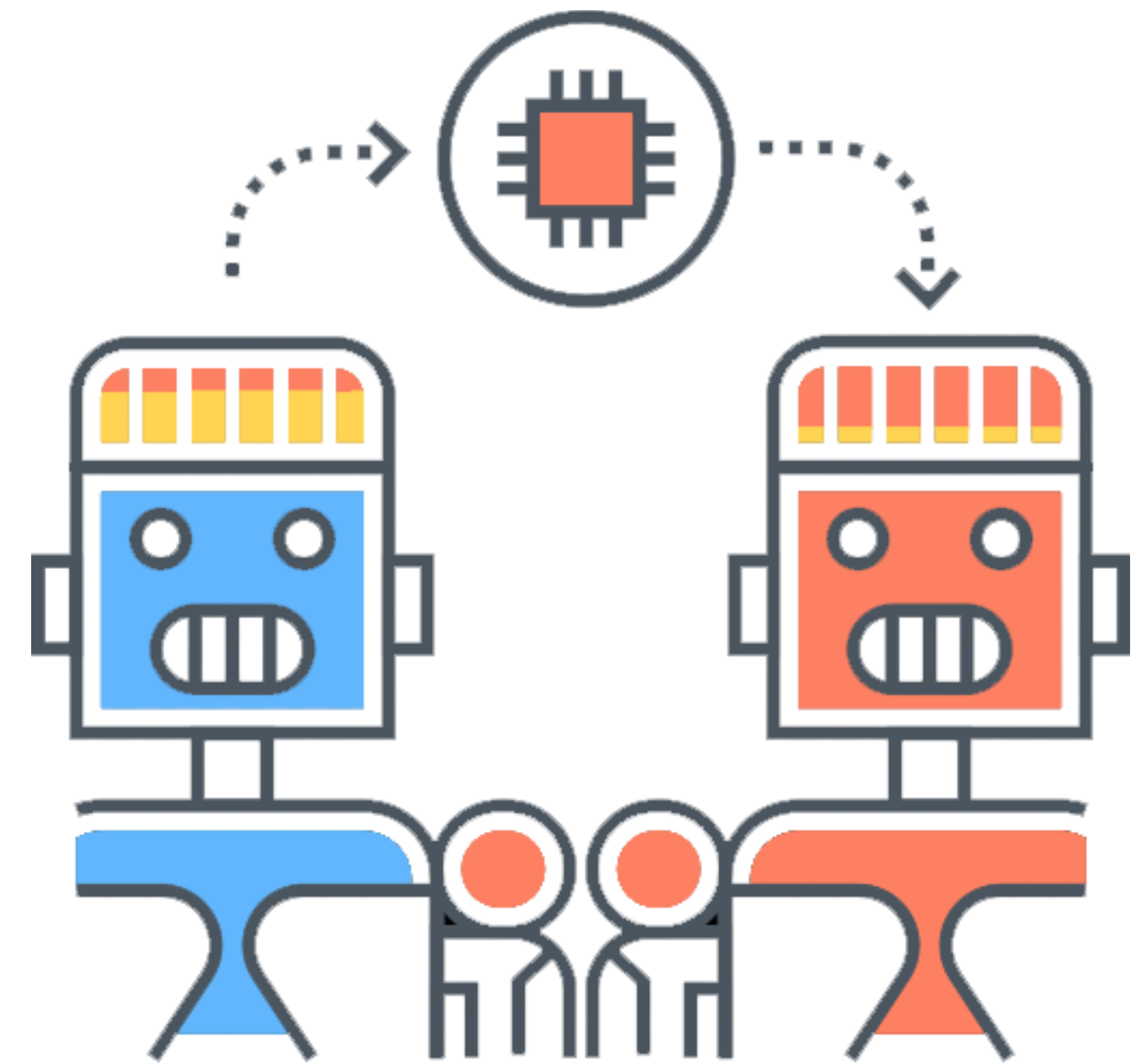




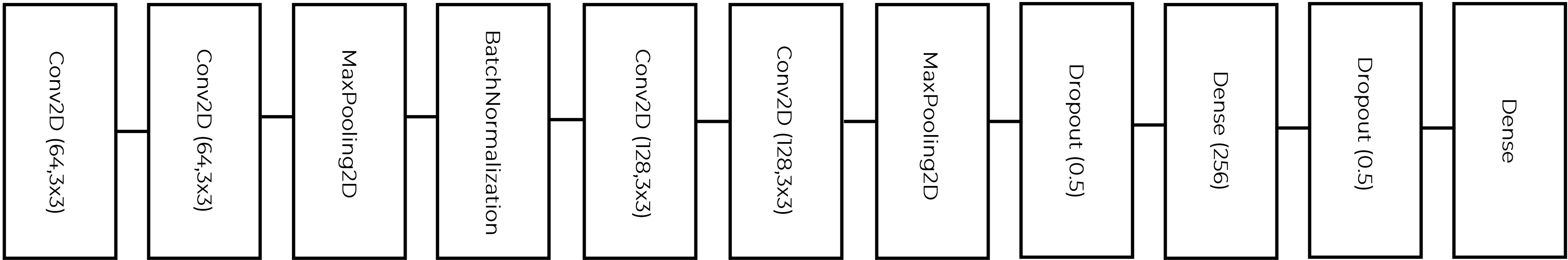
# Modelli

Per raggiungere lo scopo prefissato, si è deciso di procedere in due modi:

- realizzare un modello di rete neurale dall'inizio;
- utilizzare la tecnica di transfer learning, che consente di riutilizzare gran parte dei parametri di una rete neurale già addestrata in precedenza.

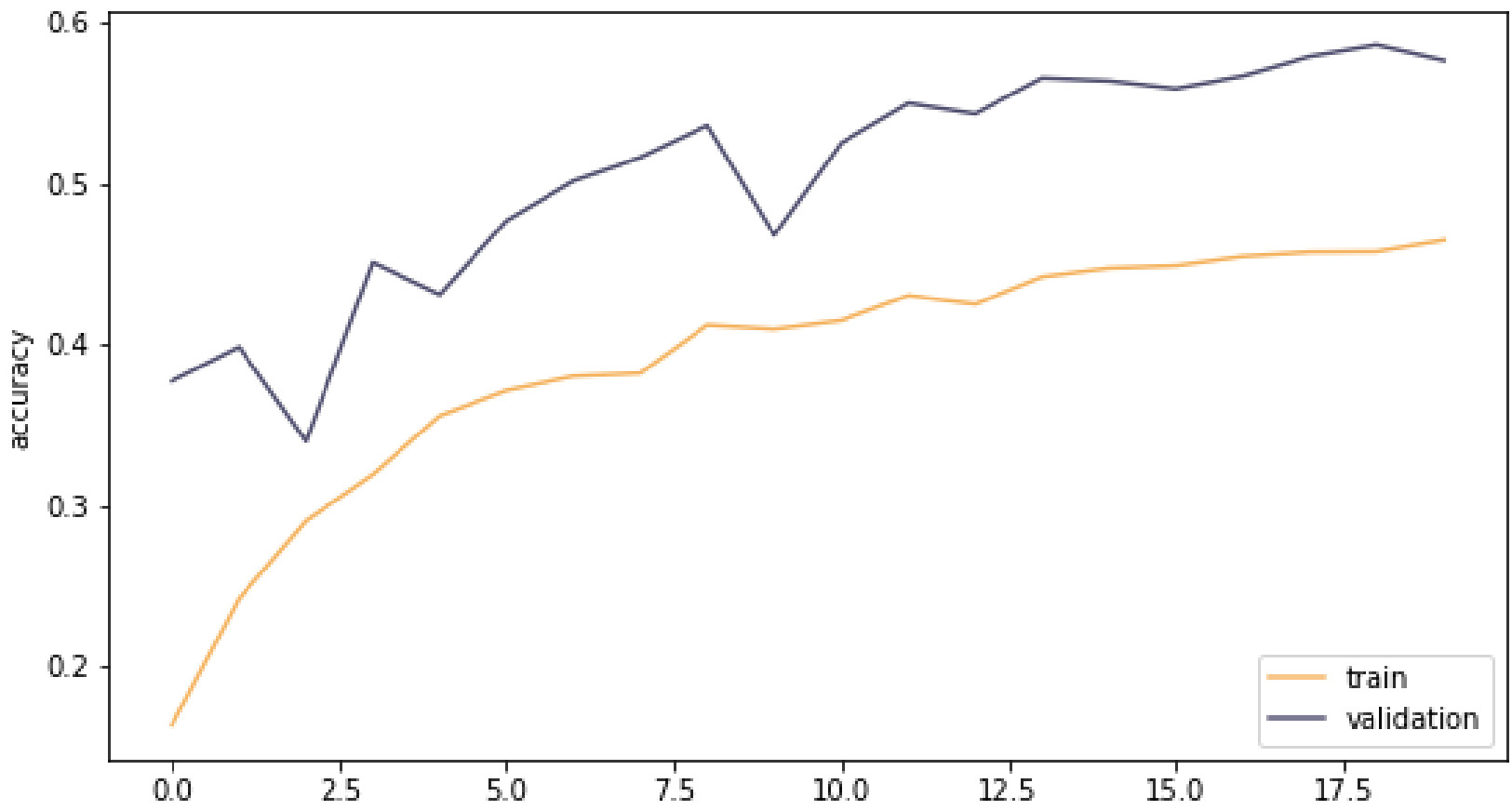


# Custom Model



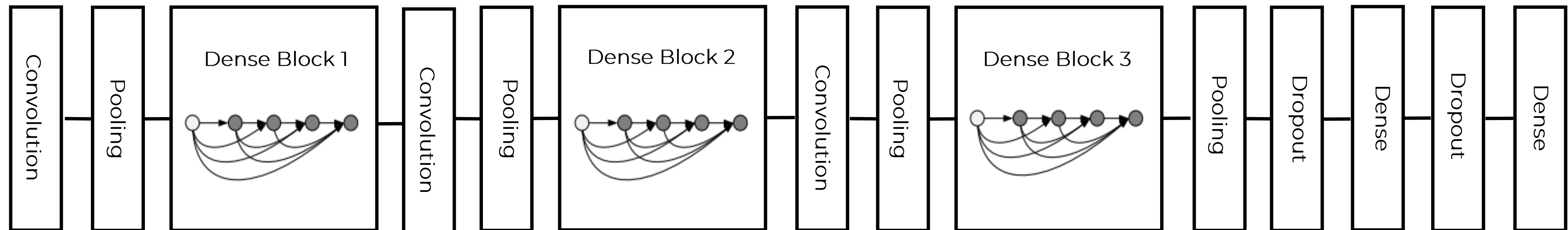
train accuracy
46%

val. accuracy
58%





# DenseNet201

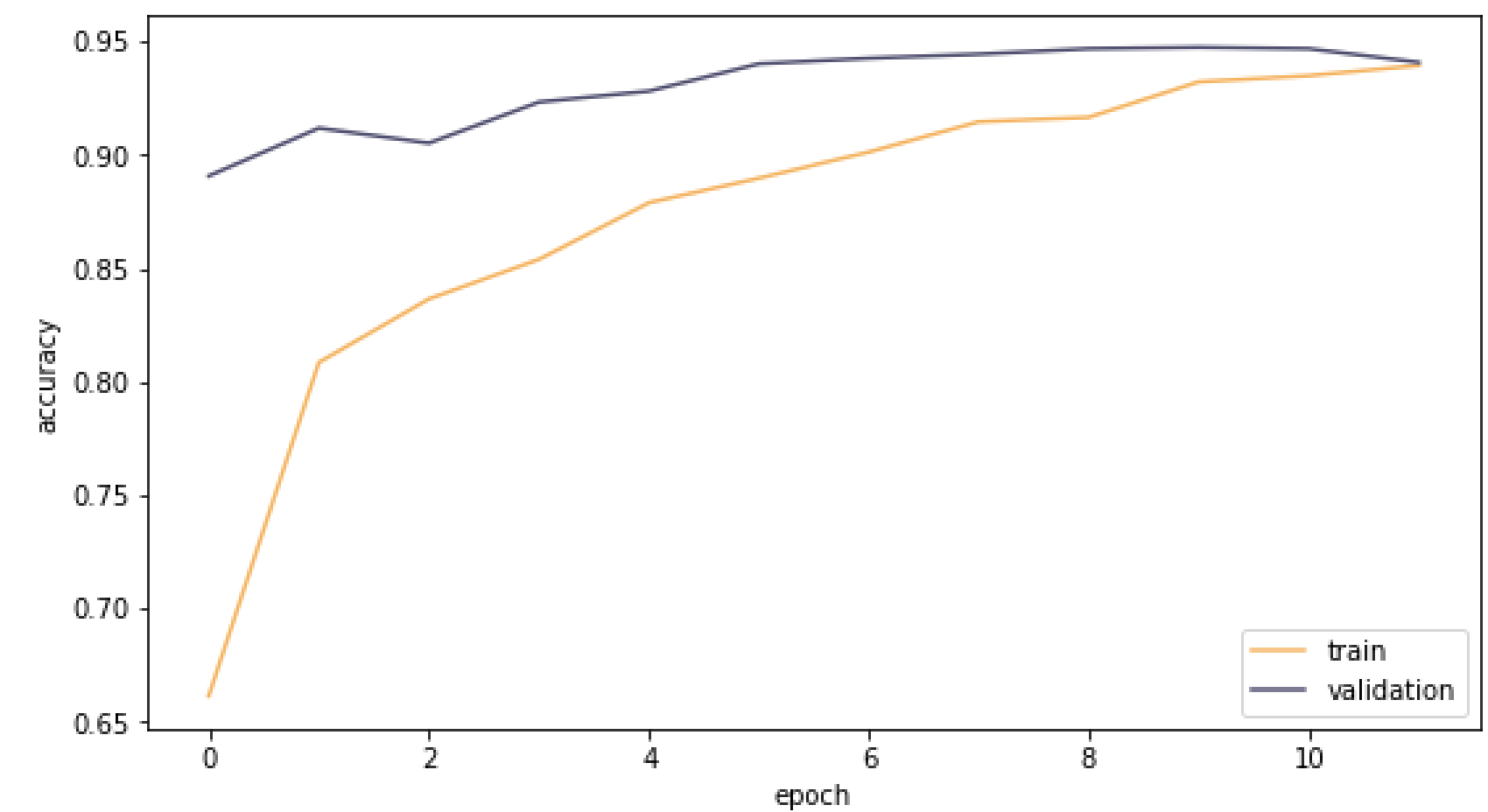


**train accuracy**

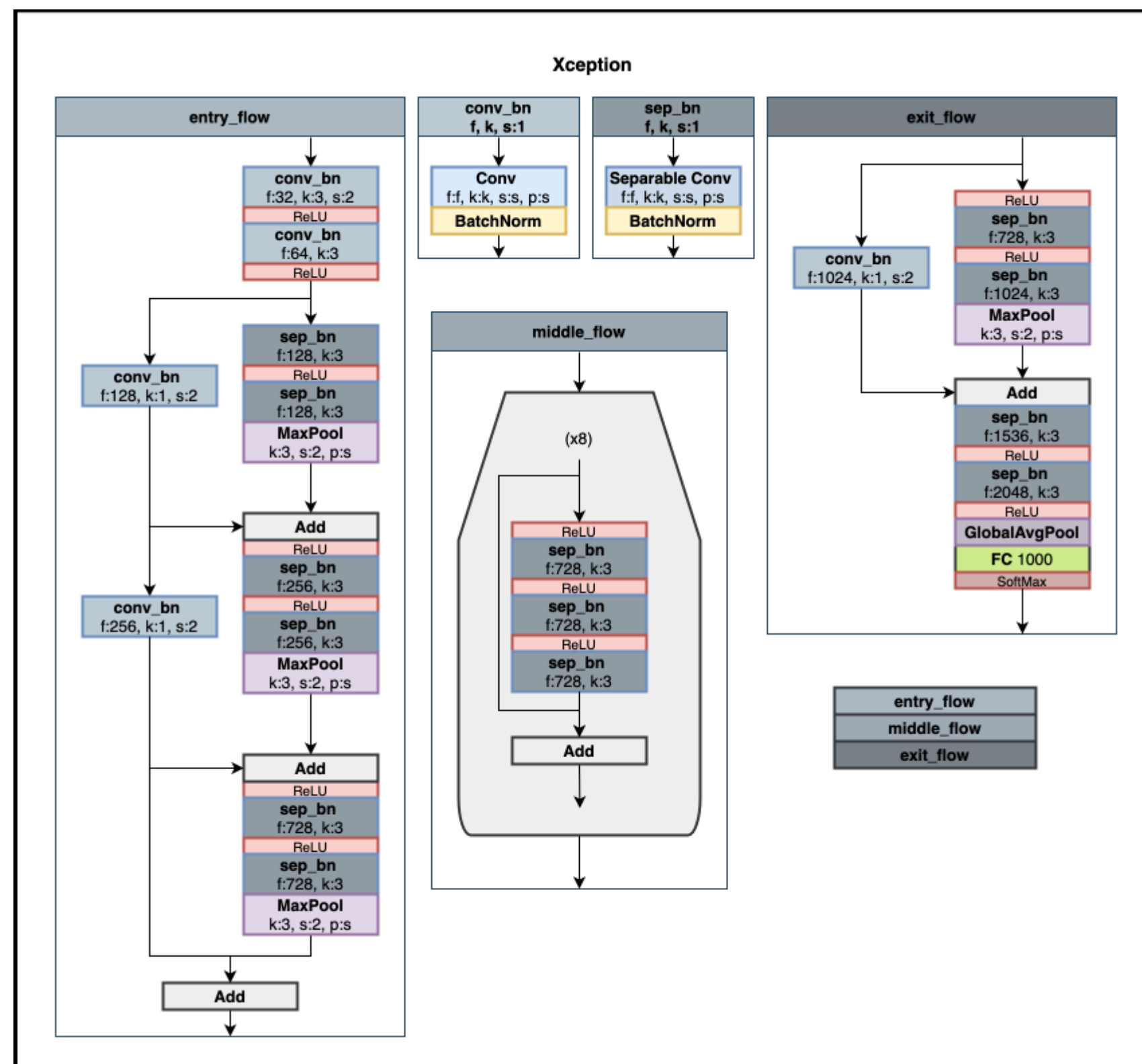
93%

**val. accuracy**

94%



# Xception

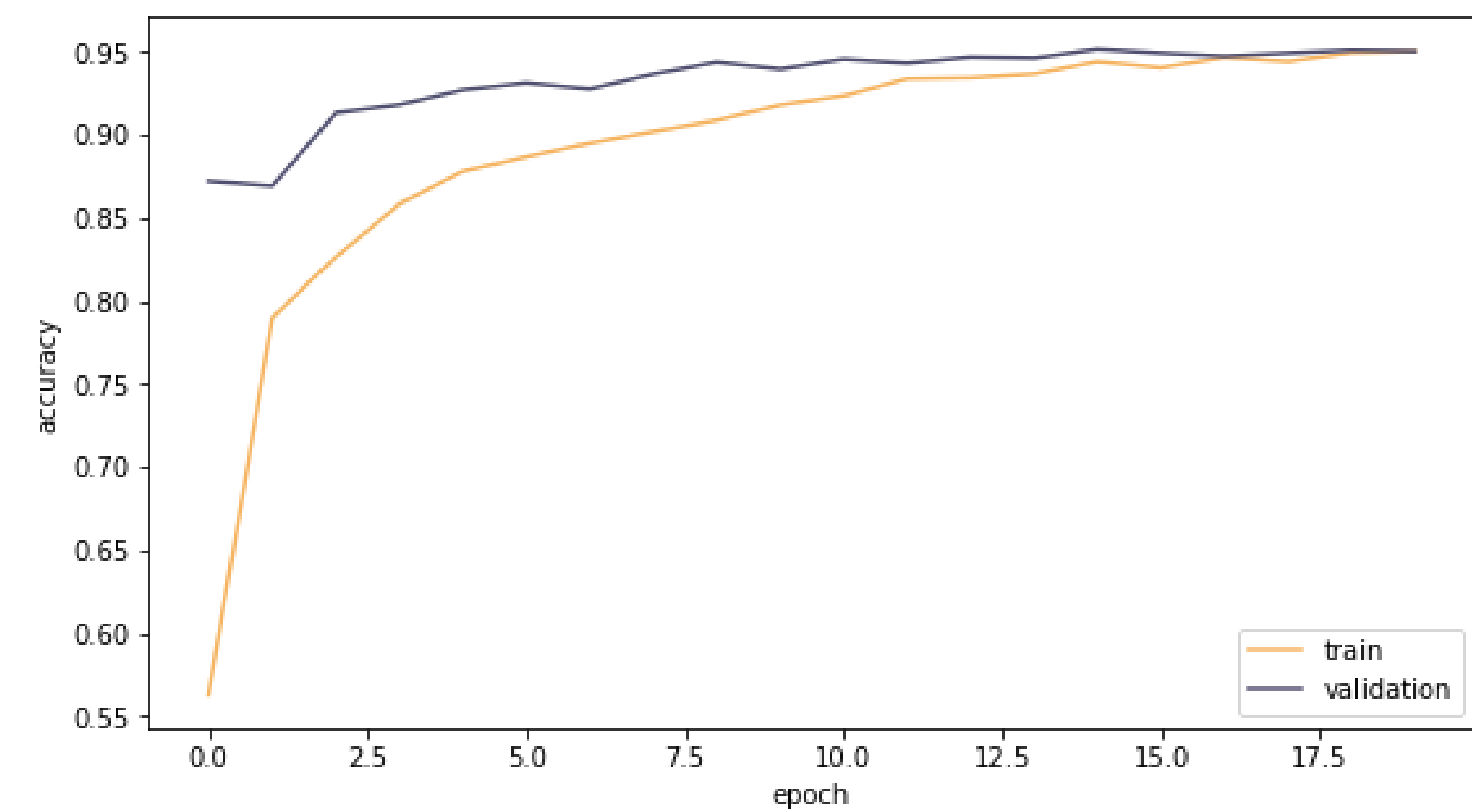


**train accuracy**

95%

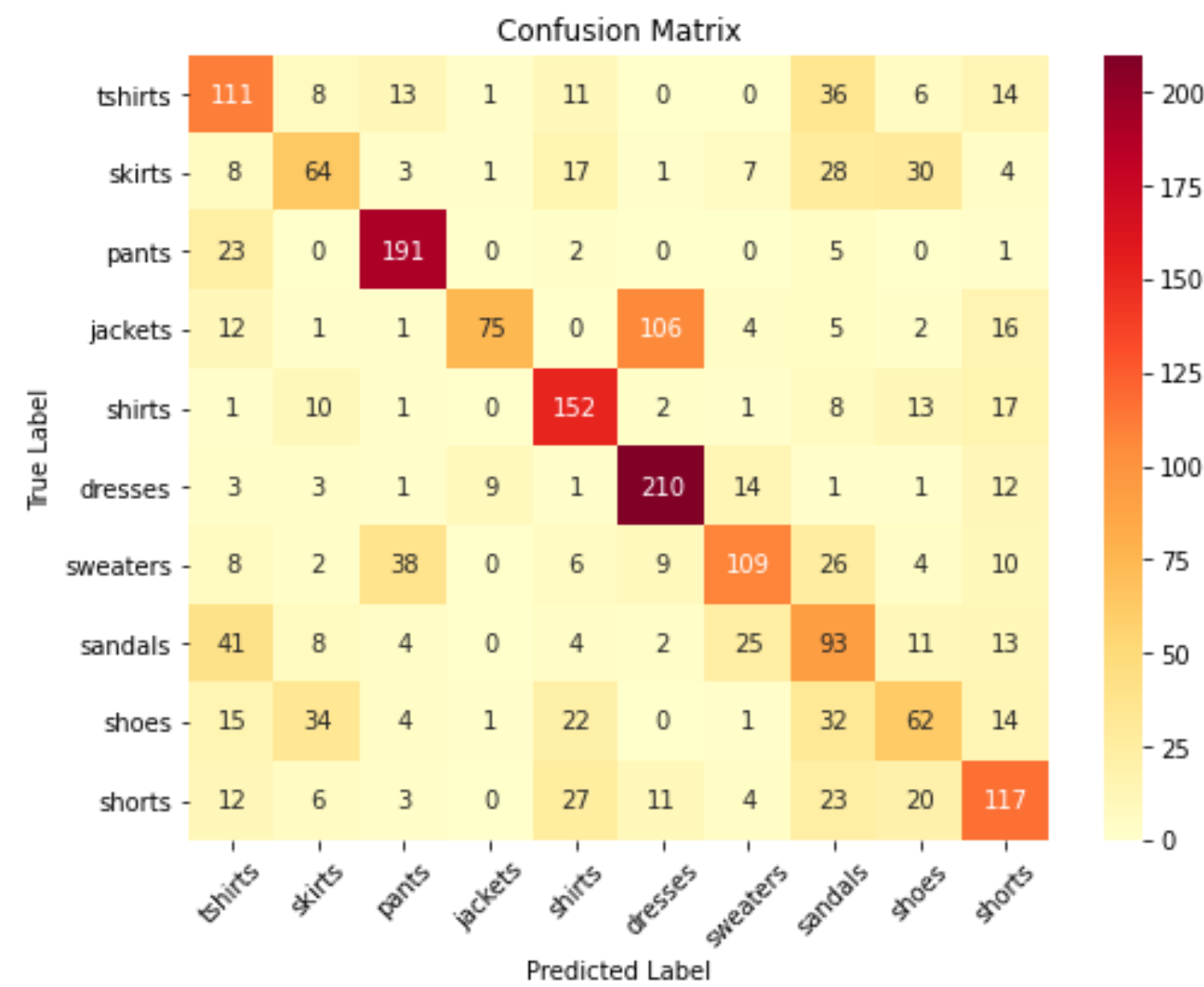
**val. accuracy**

95%



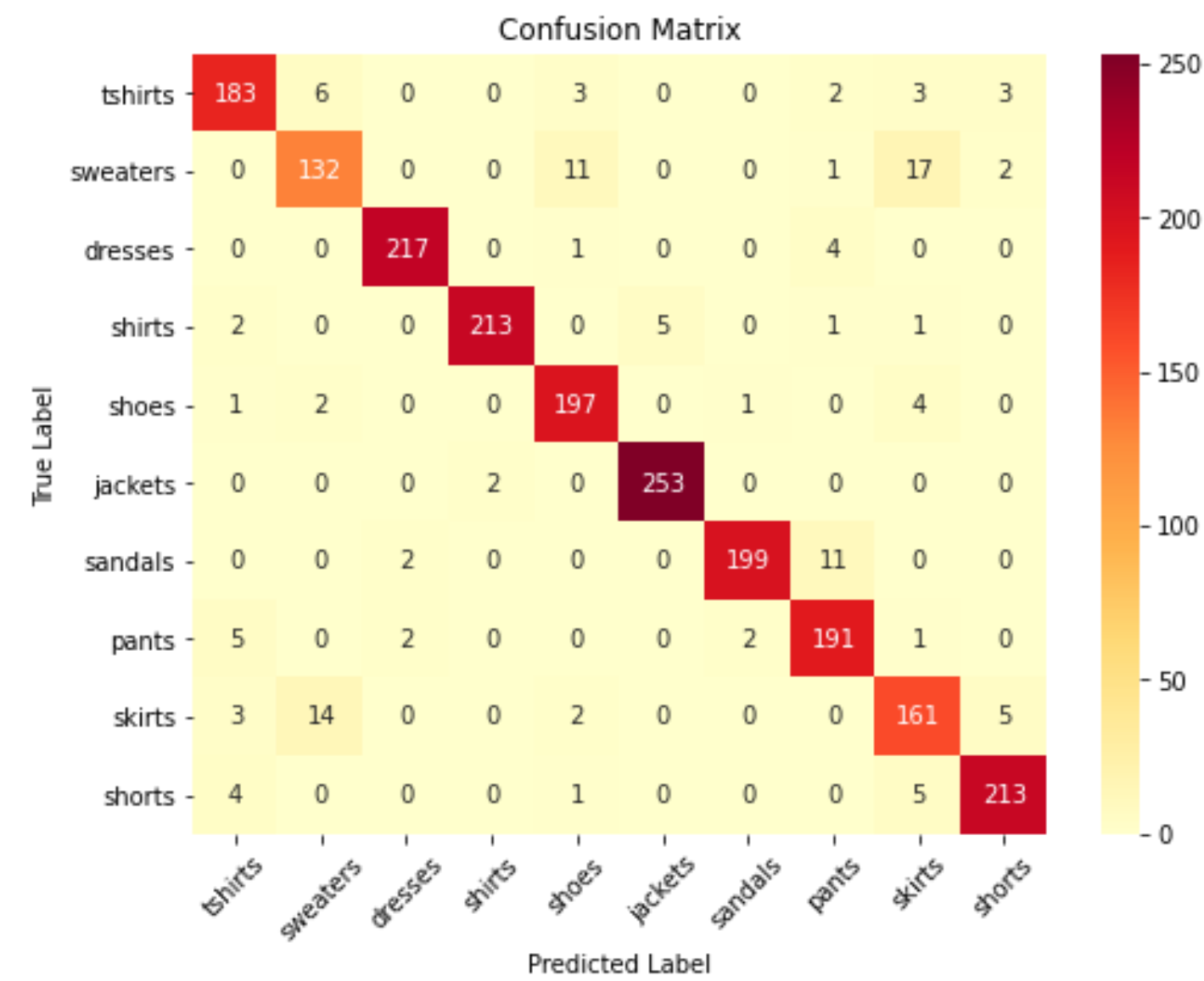
# Risultati

CUSTOM MODEL



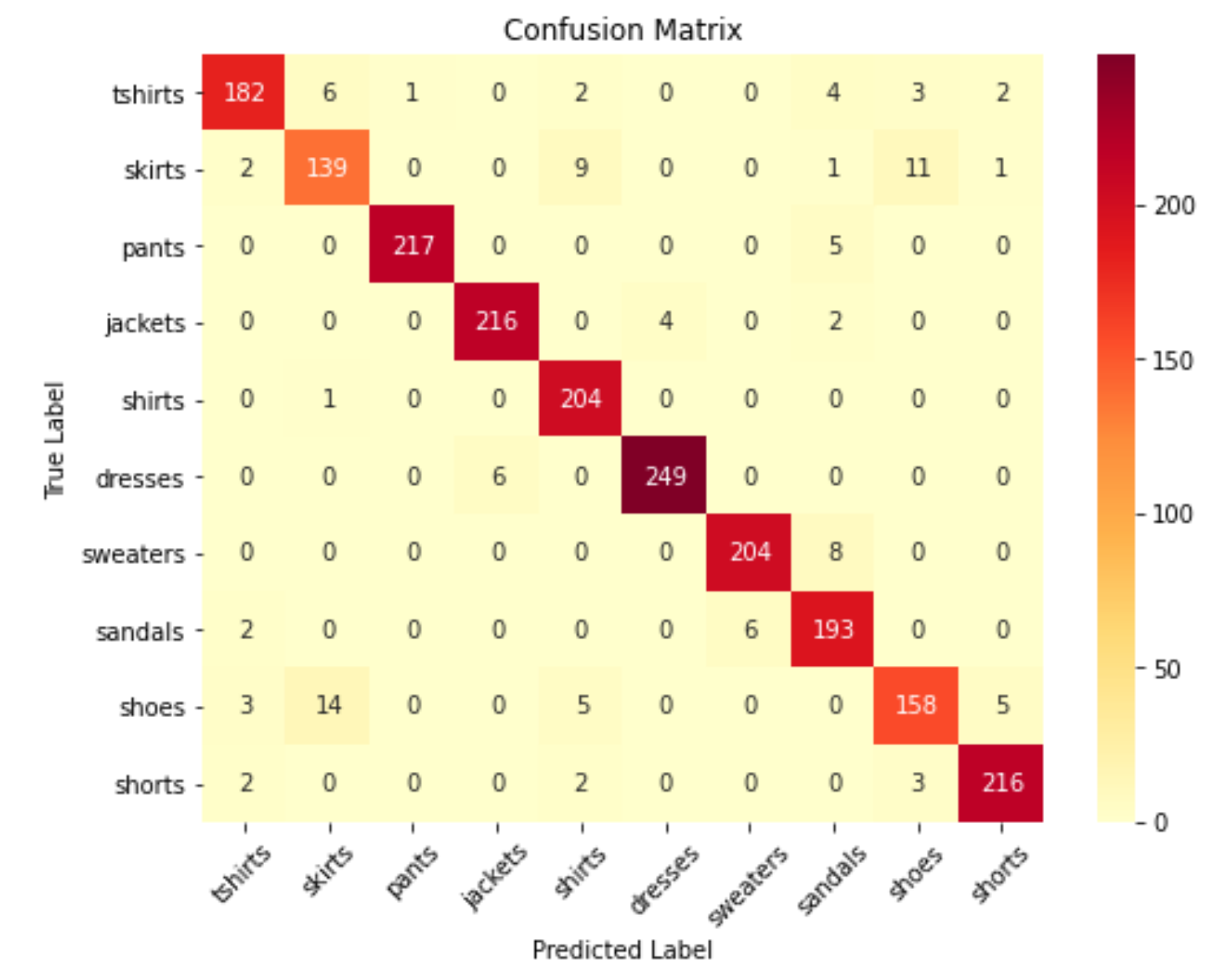
Test Accuracy : **0.567**

XCEPTION



**0.938**

DENSENET201



**0.930**

# **Image Retrieval**

# Obiettivo

L'obiettivo del task di Image Retrieval è quello di restituire i capi d'abbigliamento più simili a quelli indossati dal soggetto dell'immagine di query. In particolare verranno restituiti 10 vestiti, 5 per la parte superiore e 5 per la parte inferiore.



# Dataset

Il dataset selezionato è un sottoinsieme di quello considerato per la parte di classificazione. In particolare sono state rimosse le classi riguardanti le calzature e sono state considerate le seguenti:

- Dresses
- Jackets
- Pants
- Shirts
- Shorts
- Skirts
- Sweaters
- T-Shirts



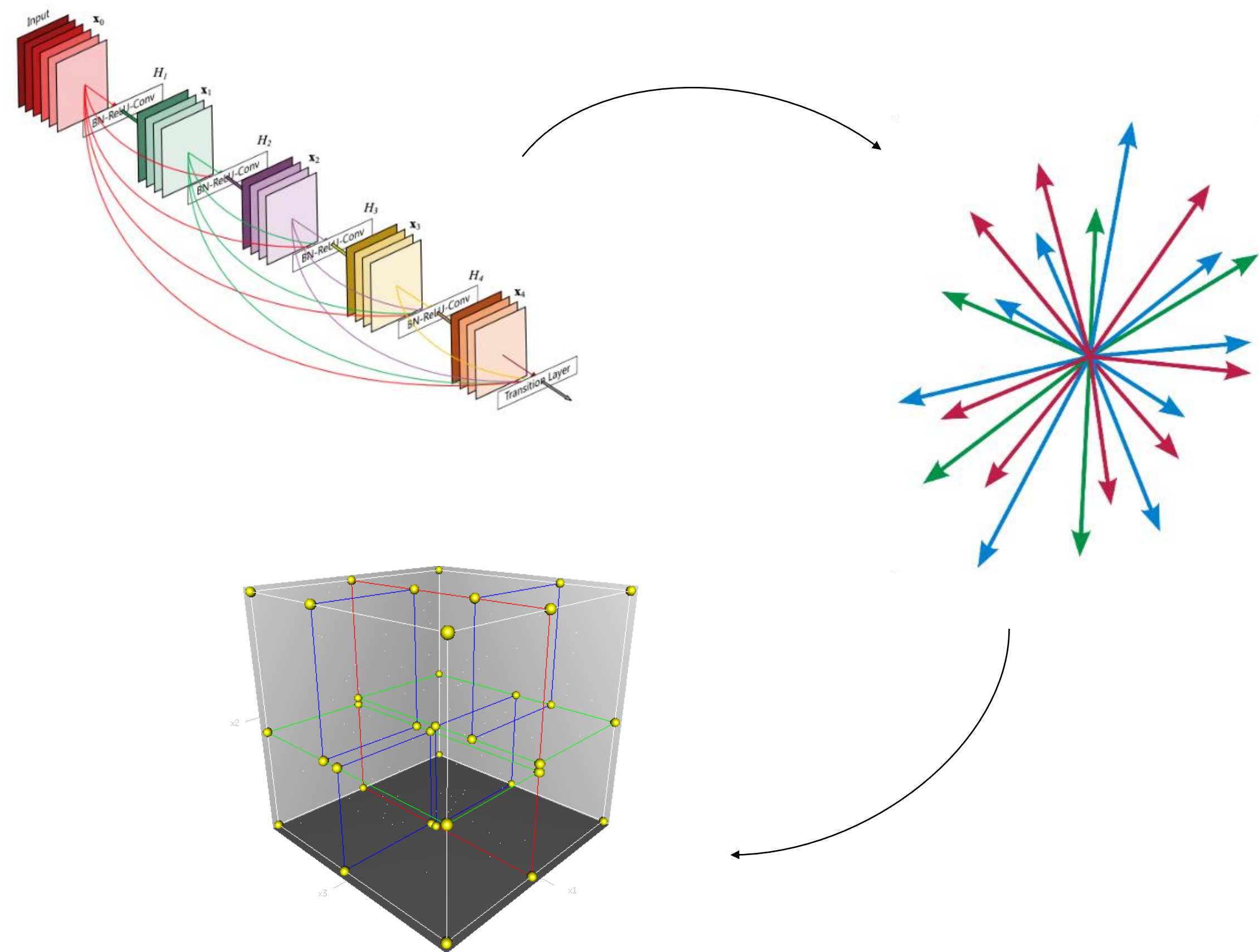


# Feature space

Per generare lo spazio delle feature è stato utilizzato il modello basato su DenseNet201 proposto per il task di classificazione.

Ogni immagine viene quindi rappresentata da un vettore di 1920 dimensioni.

Viene infine utilizzato un KDTree per il retrieval delle immagini.

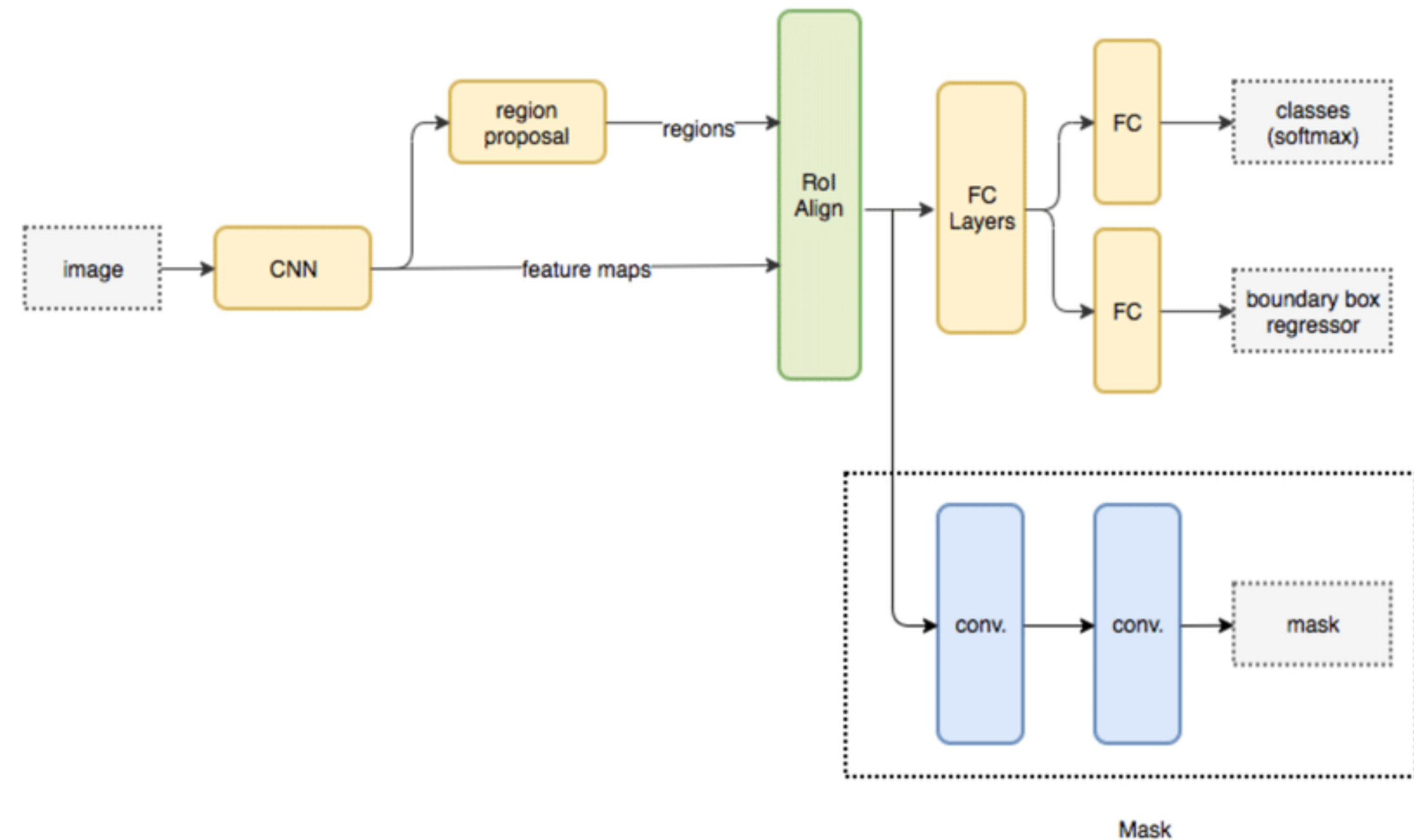




# Mask R-CNN

Come illustrato precedentemente, lo scopo è quello di restituire i vestiti più simili a partire da quelli indossati dal soggetto nell'immagine di query. Per individuare i diversi abiti nella foto è stata utilizzata una Mask R-CNN.

Il modello utilizzato è stato pre-addestrato sul dataset DeepFashion, utilizzando come backbone ResNet101.



# Esempio





# Primo risultato



Come è possibile osservare, in questo caso vengono restituiti vestiti della classe corretta. Non viene tenuto però conto del colore, elemento fondamentale per un capo d'abbigliamento.

OUTPUT:



# Color feature

Per estrarre le informazioni riguardante il colore si è scelto di costruire manualmente le feature. In particolare è stato utilizzato l'istogramma per ogni canale nello spazio RGB.

Dal momento che le immagini del dataset (e in generale quelle di vestiti) sono composte dal soggetto poste su uno sfondo bianco, quest'ultimo dominerebbe l'istogramma rendendo difficile utilizzarlo come feature.



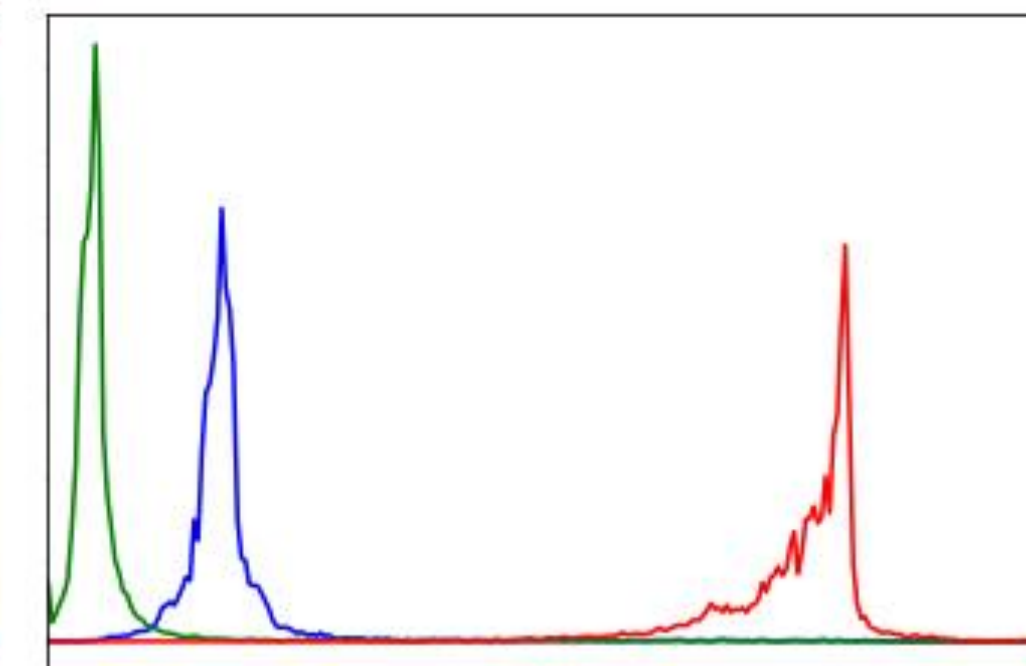
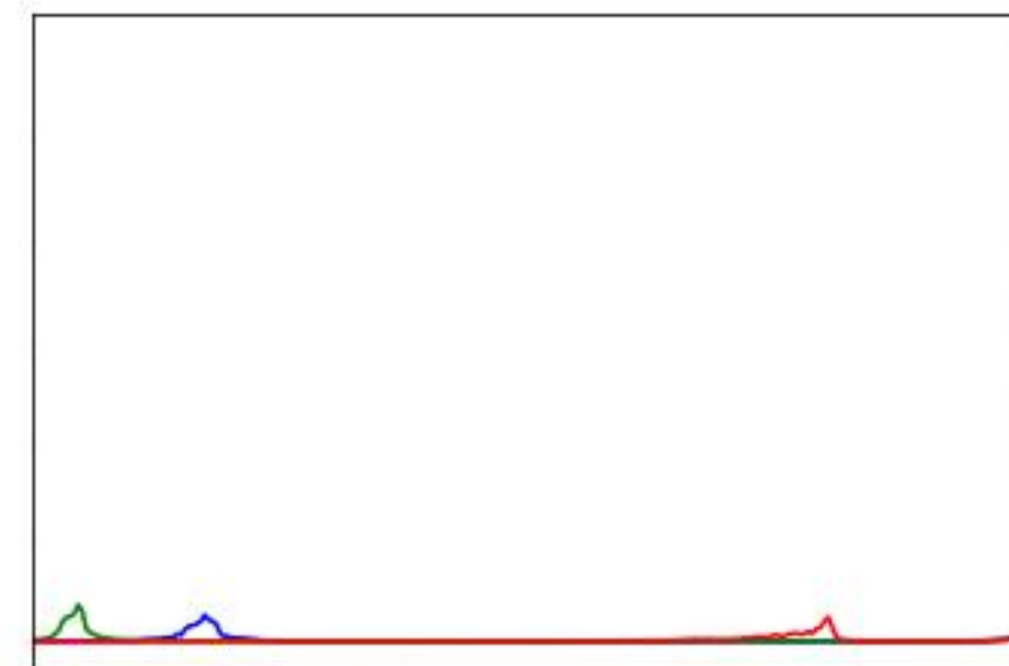


# Color feature

Per questo motivo viene utilizzato l'algoritmo GrabCut offerto da OpenCV per generare una maschera del soggetto, in modo tale da estrarre solamente l'istogramma del vestito.

Le immagini di query invece sono generalmente più difficili da scontornare, in quanto scattate in condizioni meno controllate, con sfondi che possono variare. Di conseguenza viene utilizzata la maschera generata dalla Mask-RCNN e successivamente viene calcolato l'istogramma.

Infine viene utilizzata la funzione flatten per ottenere un vettore di 343 dimensioni, che verrà concatenato con quello generato dalla CNN.



# Query image

Come già sottolineato, le immagini di vestiti sono principalmente scattate su uno sfondo neutro (generalmente bianco). Essendo il nostro modello allenato su immagini di questo tipo, abbiamo utilizzato la maschera generata dalla Mask R-CNN per rendere l'immagine di query più simile a quanto visto durante la fase di training.





# Risultato finale



Ora le immagini restituite sono più coerenti tra loro dal punto di vista del colore. Osserviamo che ci sono degli svantaggi nell'utilizzo dell'istogramma:

- Sensibilità ai cambiamenti di illuminazione dell'immagine
- Non vengono considerate le informazioni spaziali

OUTPUT:





**Demo**