

Visual Information Processing and Management

Foodx-251 classification

Davide Mazzitelli 851657

Gianmarco Corti 851864

Descrizione problema

Il progetto consiste nella progettazione e valutazione di un sistema che usi il DB di immagini **Foodx-251** composto da diverse componenti:

- Pulizia del training set da immagini non appartenenti alla categoria cibo corretta
- Classificazione del validation set nelle 251 classi previste dal dataset
- Classificazione del validation set degraded nelle 251 classi
- Valutazione oggettiva dei risultati di classificazione
- Analisi visuale di casi considerati significativi

Descrizione Dataset

Il dataset a nostra disposizione è il Foodx-251 [\[paper\]](#) per il task di food classification.

Caratteristiche del dataset:

- 158k immagini, 251 classi
- Suddivisione del dataset:
 - Training set: 118475 immagini (da 100 a 600 immagini per classe)
 - Validation set: 11994 immagini, utilizzato come primo test set
 - Validation set degraded: 11994 immagini, utilizzato come test set rappresentante delle condizioni di acquisizione non ottimali, il set consiste infatti delle stesse immagini di Validation set a cui sono aggiunte diverse tipologie di rumore
- Errori:
 - Il training set fornitoci contiene circa il 20% di labels scorrette



Osservazioni iniziali sul training set

Osservando il training set e le classi abbiamo potuto fare due osservazioni:

- Domain errors:




- Presenza di classi sbilanciate: es. classe 162
 - 117 elementi
 - presenza di immagini errate + presenza di molte tipologie diverse di cibi



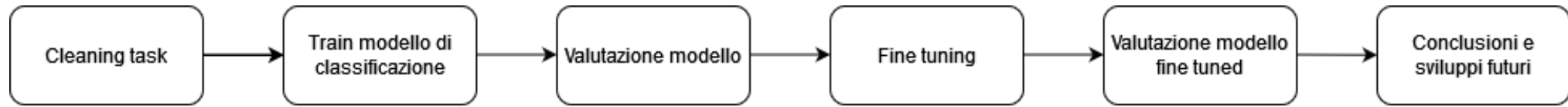
Gyro error
If you are on a range bearing 089°T and the centerline pelorus reads 090°pgc, what is the gyro error?

- Gyro = 090° and True = 089°
- Gyro = Best → Error = West
- $090^\circ - 089^\circ = 1^\circ$
- **Answer** 1° Westerly Error

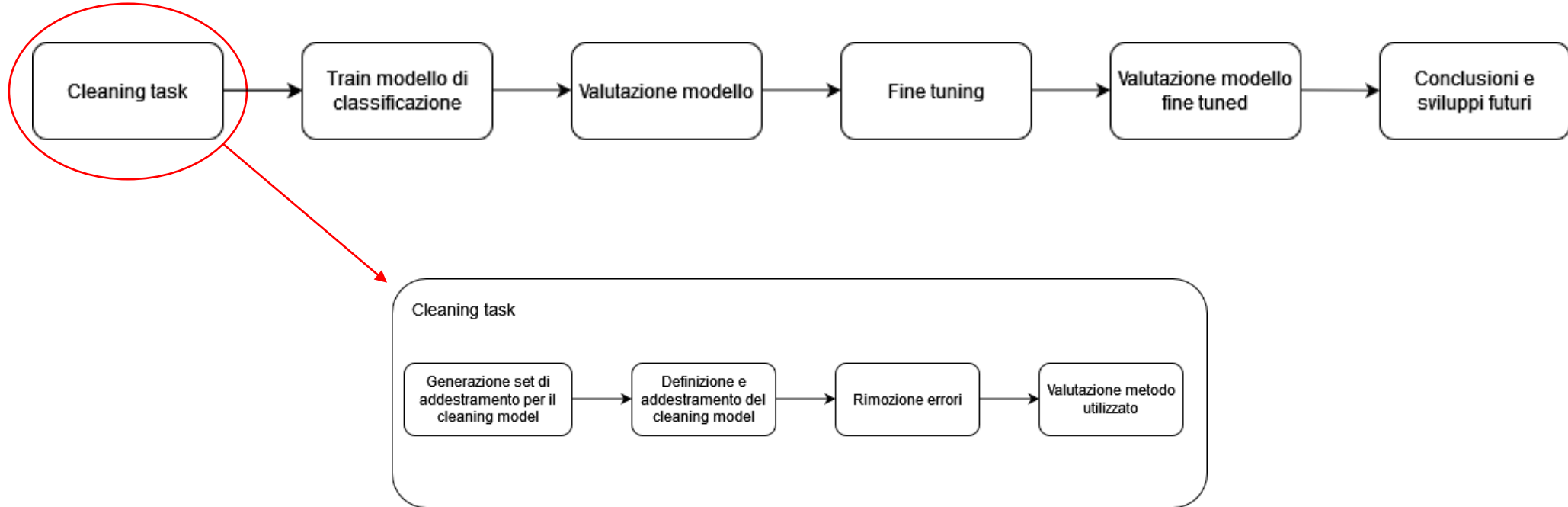


- Aspettative: minori performance del modello nella classificazione di queste classi

Descrizione pipeline



Descrizione pipeline



Cleaning Task

Approccio basato sullo studio presente nel paper *Identifying Mislabeled Instances in Classification Datasets* [1]

Tre step:

- Addestramento modello di cleaning
 - Creazione train set per il modello
 - Definizione architettura e addestramento
- Utilizzo del modello per la rimozione degli errori dal train set
- Valutazione del metodo utilizzato

Creazione train set per il cleaning model

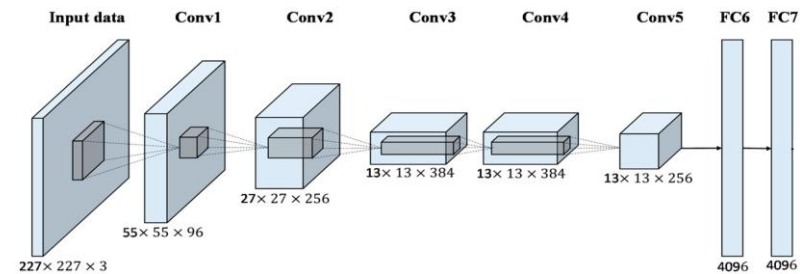
Obiettivo: rimozione per ogni classe di immagini considerate 'outlier'

→ addestramento con un set più pulito

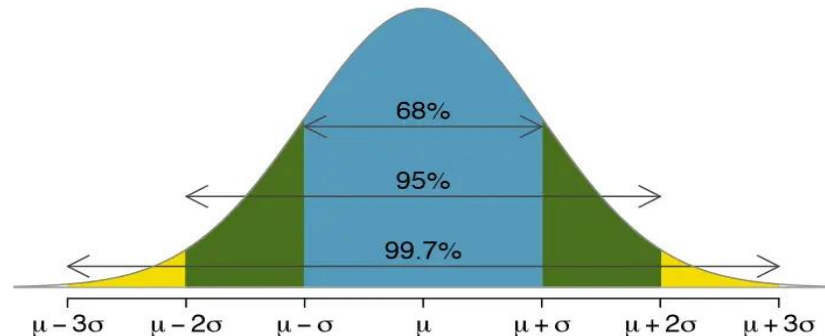
Approccio intrapreso:

- Estrazione feature delle immagini tramite transfer learning

- Rete utilizzata: Alexnet
- Cutpoint: layer 'fc7'



- Calcolo della feature 'media' per ogni classe
- Calcolo della distanza coseno tra le feature di ogni elemento del set con la feature 'media' della classe con cui è etichettato
- Filtraggio delle immagini errate sulla base dello z-score calcolato su queste distanze

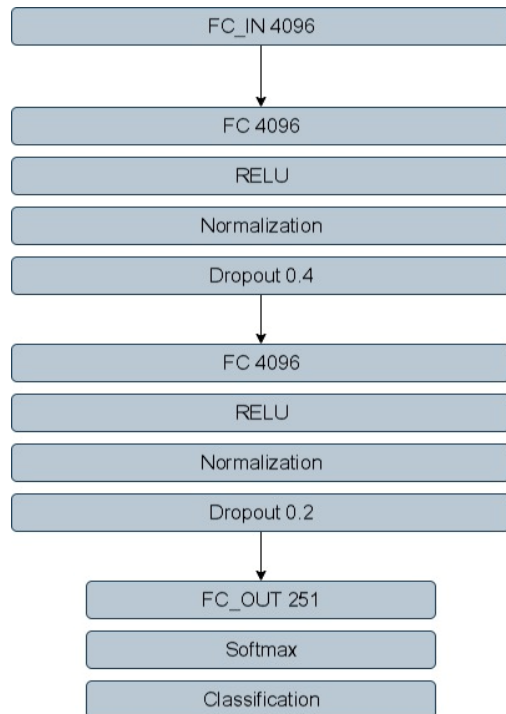


Soglia scelta per la rimozione: 1.5

Numero di immagini rimosse: 16178

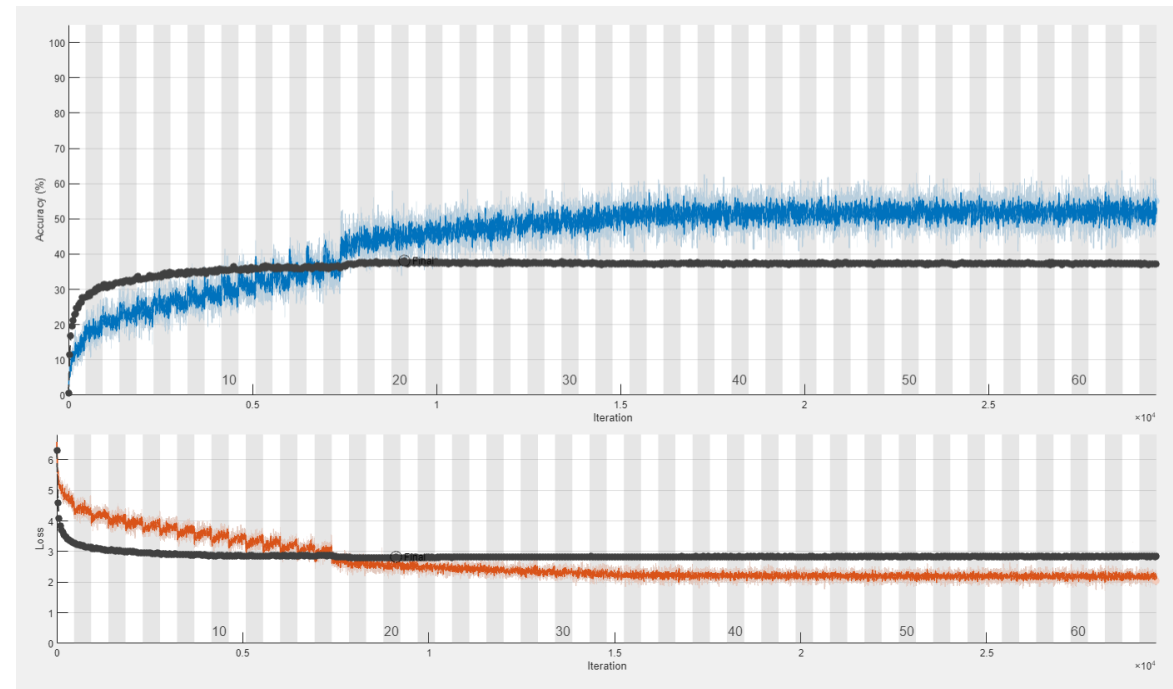
Definizione architettura ed addestramento

Architettura del modello



Training

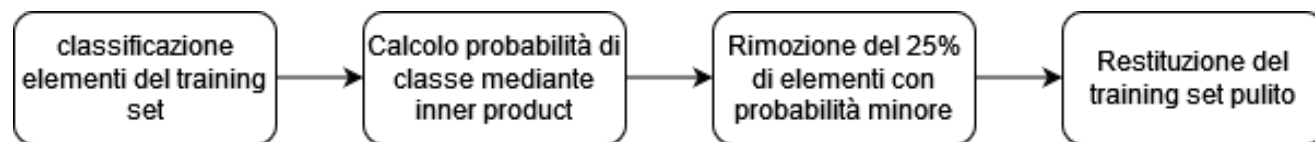
- Optimizer: Adam
- Mini batch size: 256
- lr : $1e-4$
- Output Network: best validation loss
- N. epochs : 64



Validation accuracy: 38.19%

Rimozione degli errori

Metodo di cleaning:



Risultati ottenuti:

Immagini rimosse: 29619

Dataset risultante: 88856 immagini

Esempio immagini errate rimosse dalla classe 0 'macaron':



Altre osservazioni:

La classe 162 risulta essere quella maggiormente ridotta, elementi rimasti: 51

Valutazione del metodo utilizzato

- Modifica casuale delle label del 20% di immagini del validation set
- Classificazione immagini mediante cleaning model
- Eliminazione di una percentuale α di immagini classificate con minor confidenza

Metriche di valutazione:

Dato il dataset D:

- $I \subseteq D$: insieme di istanze con errori di label
- I_α : istanze errate trovate con il nostro metodo

$$\alpha \text{ precision} = \frac{|I_\alpha \cap I|}{|I_\alpha|}$$
$$\alpha \text{ recall} = \frac{|I_\alpha \cap I|}{|I|}$$

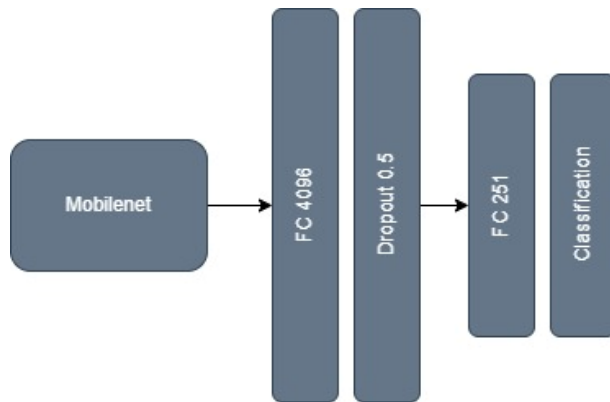
Risultati ottenuti:

	α precision	α recall
$\alpha = 0.2$	0.8	0.8
$\alpha = 0.25$	0.69	0.87
$\alpha = 0.3$	0.6	0.9

Configurazione selezionata

Train modello classificazione

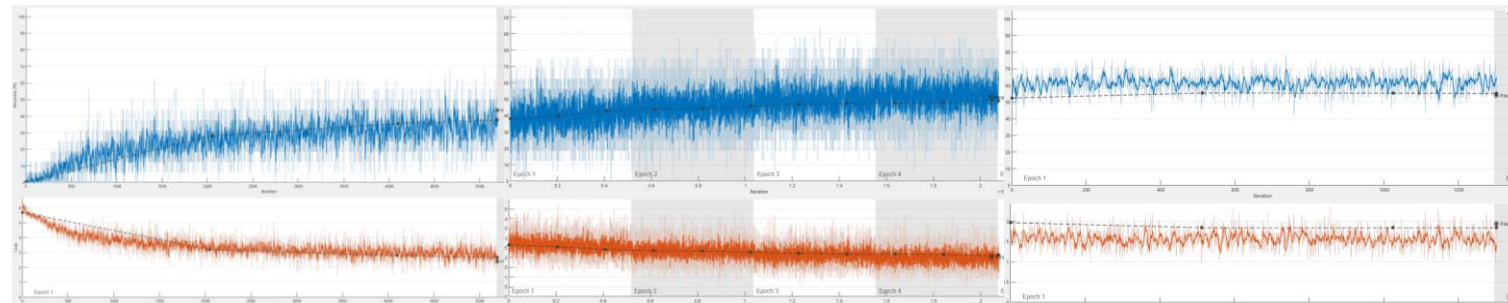
Modello: MobilenetV2



- Aggiunti layer FC e Dropout
- Sostituzione layer classificazione

Training:

- Optimizer: Adam
- Mini batch size: 16
- lr : 1e-4
- Output Network: best validation loss
- N. epochs : 6



- Training set: dataset pulito con metodo di cleaning
- Data augmentation:
 - Rotazione
 - Traslazione

Valutazione modello base

Metriche di performance:

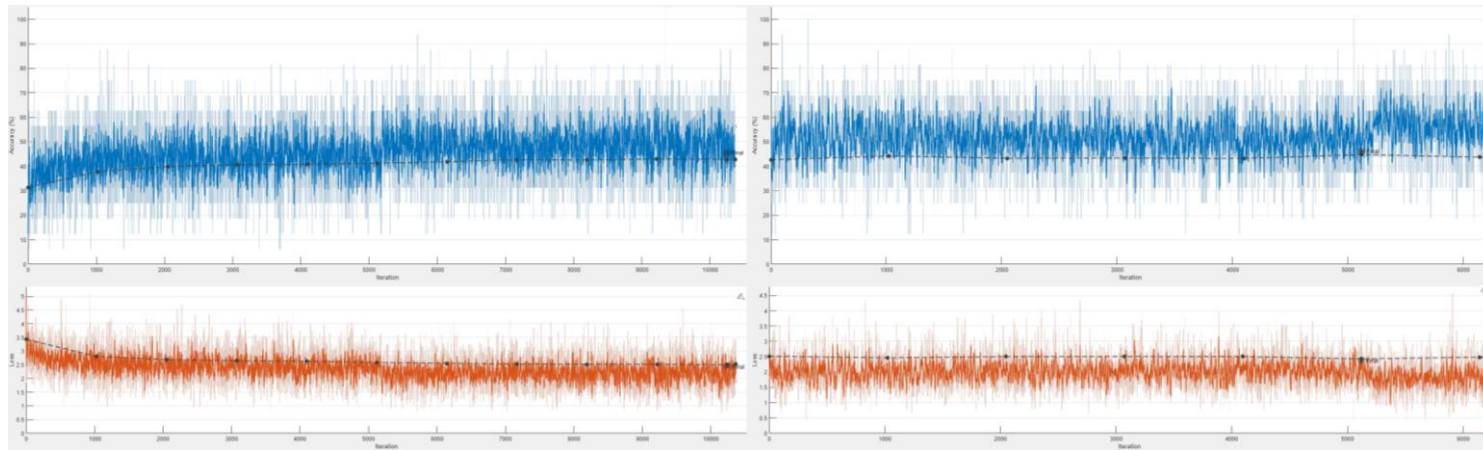
- Accuracy (1-Shot, 3-Shot)
- Precision media
- Recall media

	Accuracy - 1 Shot	Accuracy - 3 Shot
Validation Set	0.5394	0.7354
Validation Set Degraded	0.3369	0.4805

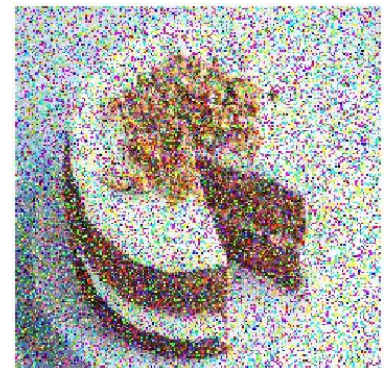
	Precision media	Recall media
Validation Set	0.5261	0.5675
Validation Set Degraded	0.3300	0.4593

Fine Tuning

- Osservazione validation set degradato:
 - 3 tipologie principali di rumore:
 1. Blur
 2. Salt & pepper
 3. pixelation
- Data augmentation → aggiunta randomica di rumore al training set



1.



2.



3.

Performance fine tuning

Metriche di performance:

- Accuracy (1-Shot, 3-Shot)
- Precision media
- Recall media

Modello base:

	Accuracy - 1 Shot	Accuracy – 3 shot
Validation Set	0.5394	0.7354
Validation Set Degraded	0.3369	0.4805
	Precision media	Recall media
Validation Set	0.5261	0.5675
Validation Set Degraded	0.3300	0.4593

Modello fine tuned:

	Accuracy - 1 Shot	Accuracy – 3 shot
Validation Set	0.5732	0.7637
Validation Set Degraded	0.4613	0.6507
	Precision media	Recall media
Validation Set	0.5596	0.5772
Validation Set Degraded	0.4507	0.4854

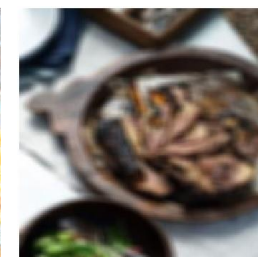
Approcci alternativi

Identificazione e rimozione delle immagini a bassa qualità

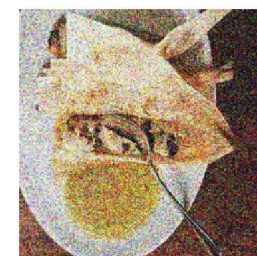
- Utilizzo di NIQE per il calcolo di quality score del validation set degradato
- Rimozione immagini > soglia scelta (10)
- Valutazione performance del modello fine tuned sul validation set filtrato



NIQE score: 4.3260



NIQE score: 7.4888

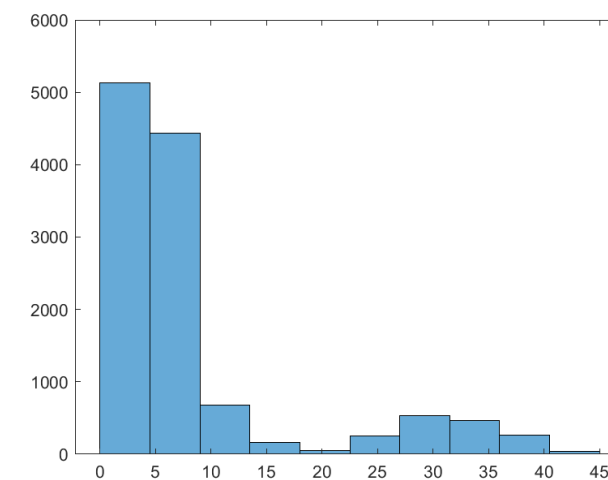


NIQE score: 29.3487



NIQE score: 11.8733

Modello Fine Tuned	Accuracy - 1 Shot	Accuracy - 3 Shot
Validation Set Degraded	0.5019	0.6970
	Precision media	Recall media
Validation Set Degraded	0.4920	0.5153



Istogramma dei NIQE scores

Conclusioni

Il modello di classificazione prodotto presenta buoni risultati solamente per quanto riguarda la classificazione di tipo 3 shot, in particolare abbiamo ottenuto un notevole aumento delle performance di accuracy tramite fine tuning del modello di base (oltre il 10% su validation set degraded).

I valori ottenuti riguardo le metriche di precision e recall ci suggeriscono performance del modello variabili a seconda delle singole classi

Top 5 precision scores (fine tunde model – validation set)

Precision	Class 27 : 0.9259	Class 76 : 0.9245	Class 173 : 0.9245	Class 49 : 0.9137	Class 97 : 0.8852
Recall	Class 221 : 0.9310	Class 101 : 0.9285	Class 63 : 0.9250	Class 100 : 0.9024	Class 245 : 0.8979

Worst 5 recall scores (fine tunde model – validation set)

Precision	Class 162 : 0	Class 116 : 0	Class 58 : 0.0909	Class 209 : 0.0975	Class 75 : 0.0983
Recall	Class 162 : 0	Class 116 : 0	Class 84 : 0.1842	Class 147 : 0.2105	Class 58 : 0.2222

Il modello è in grado sia di classificare molto bene determinate classi, ma anche di non riuscire affatto nella classificazione di altre. Questo presupponiamo sia dovuto al fatto che il nostro metodo di pulizia del dataset sfavorisca molto determinate classi che inizialmente presentavano un numero maggiore di errori (es. classe 162) .

Conclusioni

Osservazione :

- Classe 116: osservando le immagini etichettate come appartenenti a questa classe pre e post cleaning abbiamo notato che il nostro metodo ha eliminato per la maggior parte immagini di cibi

Tipologia di immagini mantenute:



Limite del metodo di cleaning utilizzato: presenza di un gran numero di immagini errate della stessa tipologia

Sviluppi futuri

- Utilizzo di un' architettura più complessa per il cleaning model, in modo da ottenere un training set migliore per il modello di classificazione
- Rimozione strato aggiunto a mobilenet e fine tune per ottenere un modello più semplice con meno parametri
- Utilizzo di mezzi con risorse computazionali avanzate per addestrare modelli più sofisticati ed accurati

Extra: NIQE

- Opinion-Unaware (OU) Distortion-Unaware (DU) NR-IQA method
- Valutazione basata su fattori come nitidezza, rumore e altri dettagli visivi
 - Simile alla percezione umana
 - Coefficienti MSCN :
 - Normalizzazione dell'immagine pesata tramite gaussian weighting function
 - Catturano le proprietà statistiche che vengono modificate dalla presenza di distorsione
 - Coefficienti NSS : Generalized Gaussian Distribution delle MSCN

