

# Classificazione di materiali e riconoscimento loghi di borse

Tesina per il progetto di esame del corso di Computer  
Vision e Deep Learning

Andrea Civitarese<sup>1\*</sup> and Gioele Pasquini<sup>1†</sup>

<sup>1\*</sup>Dipartimento di Ingegneria dell'Informazione, Università Politecnica  
delle Marche, Via Brecce Bianche, 12, Ancona, 60131, Italy.

\*Corresponding author(s). E-mail(s): [s1112974@studenti.univpm.it](mailto:s1112974@studenti.univpm.it);

Contributing authors: [s1113072@studenti.univpm.it](mailto:s1113072@studenti.univpm.it);

<sup>†</sup>These authors contributed equally to this work.

## Sommario

In questa tesina descriveremo i diversi esperimenti svolti al fine di realizzare i due task a noi assegnati: la classificazione di materiali di borse e la detection di loghi di brand di abbigliamento.

Nella prima parte della relazione, verranno riassunti i lavori correlati al nostro progetto ed elencati i diversi paper da noi utilizzati come punto di riferimento. In seconda istanza, verranno discussi i materiali e i metodi di cui abbiamo usufruito per poter svolgere i due task, soffermandoci particolarmente sulla descrizione dei dataset cercati, utilizzati, scartati e creati.

Verranno poi descritte in dettaglio le reti neurali utilizzate nel progetto e i vari esperimenti fatti al fine di raggiungere il risultato migliore. In generale, il lavoro svolto può essere riassunto come l'utilizzo di diverse architetture di reti neurali, sulle quali è stato effettuato il fine tuning con due dataset contenenti immagini di borse e loghi specifici per i task.

Nella parte finale verranno riportati dati sul consumo energetico e sulle emissioni dei diversi addestramenti da noi utilizzati.

Infine, verranno discussi i risultati da noi ottenuti nei vari esperimenti rapportandoli anche, appunto, con i consumi energetici delle reti e verranno proposti degli sviluppi futuri dei lavori da noi svolti.

**Keywords:** Classification, Detection, Brand, Fashion, Materials, CNN, YOLO

# 1 Introduzione

La moda, al giorno d'oggi, è in continuo divenire, anche grazie all'avvento dei social network che assumono un ruolo molto importante nella nostra quotidianità. Coloro che nel marketing definiamo "influencer" hanno acquisito un ruolo predominante nella scelta dei capi d'abbigliamento, influenzando le donne ma anche gli uomini di tutto il mondo. Se la moda, prima, era vista come qualcosa di privilegiato, ora è certamente alla portata di tutti. Bisogna anche considerare che, in questi tempi, tutte le grandi aziende cercano di stare al passo con il progredire della tecnologia per poter ottimizzare produzione, vendite e profitti. Grazie a ciò, è nata l'esigenza di utilizzare branche dell'informatica, come machine learning e deep learning, anche in campi che sembrerebbero diametralmente opposti all'informatica, come la moda.

Il lavoro svolto per questo progetto ha come obiettivo quello dell'identificazione dei brand a partire dai loghi presenti sulle borse e dei materiali presenti in queste a partire dalle loro immagini.

Alla luce di ciò, il tutto è stato diviso in due task principali: uno di classificazione e uno di detection.

In particolare, il task di classificazione prevedeva di addestrare una rete neurale per classificare i materiali con cui sono state realizzate delle borse partendo dalle loro immagini. Per fare ciò sono state testate diverse reti neurali convoluzionali di classificazione (CNN) su un dataset utilizzato per training, validation e test che comprende circa 1400 borse sia da uomo che da donna raccolto dal sito MyTheresa.

Il task di detection, invece, riguarda il rilevamento, sempre su borse, di loghi di brand fashion; per questo task è stata usata la rete neurale preaddestrata YOLO v8, sulla quale è stato effettuato un fine tuning tramite un dataset creato appositamente da noi reperendo immagini da internet.

[Link](#) al GitHub del progetto.

# 2 Lavori correlati

Riguardo alla classificazione dei materiali di borse non è stato trovato alcun paper che avesse le stesse caratteristiche dei nostri task, esiste già tuttavia un lavoro svolto da ricercatori dell'Università Cattolica di Rio de Janeiro [1] che si occupa di riconoscere il tipo di borsa (zaino, marsupio, pochette, borsone, tote bag, etc.). Il workflow di questo task è, però, abbastanza differente dal nostro dato che, oltre ad usare una rete di classificazione, nel lavoro dell'università brasiliana vengono testate anche diverse reti di detection per effettuare il crop delle immagini in modo da individuare solo la parte contenente la borsa. Questa parte dell'immagine sarà quella che poi verrà data in input alla rete classificatrice.

Per quanto riguarda il task di detection, non esiste alcun lavoro svolto riguardante specificatamente la detection di loghi dei brand sulle borse.

Esistono, tuttavia, una serie di dataset che propongono delle sfide di detection riguardanti i loghi dei più famosi brand sul mercato. Nel survey di Sujuan Hou et al. [2], sono riportati i principali dataset, le diverse sfide e le migliori reti neurali in grado di affrontarle.

Inoltre, sono stati molto utili e interessanti diversi paper trovati in rete, come

”Brand-Aware Fashion Clothing Search using CNN Feature Encoding and Re-ranking” [3], ”Brand Logo: Visual Analysis of Fashion Brands” [4] e ”Deep Cross-Domain Fashion Recommendation” [5]

## 3 Materiali e metodi

In questo capitolo andremo a parlare dei dataset utilizzati e degli esperimenti effettuati nei due task.

### 3.1 Classificazione di materiali

Il task di classificazione dei materiali consiste nella previsione del materiale con cui è stata realizzata la borsa presente nell’immagine data in input alla rete neurale.

#### 3.1.1 Dataset utilizzato e relative sfide

Il dataset utilizzato è composto da immagini prese dal sito MyTheresa<sup>1</sup> suddivise tra borse da donna e da uomo estratte dalle pagine dell’e-commerce riferite al catalogo di diversi paesi quali:

- Italia;
- Cina;
- Korea;
- Stati Uniti;
- Giappone.

Nel dataset sono presenti borse di 4 grandi brand del fashion:

- Gucci;
- Balenciaga;
- Yves Saint Laurent;
- Tod’s.

Le borse presenti sono, inoltre, annotate con i loro materiali di produzione e altre informazioni non utili al task che stiamo perseguendo, come il costo, il colore dominante, la tipologie di borsa e altro.

Una volta unite le immagini provenienti dai cataloghi dei vari paesi ed aver eliminato i duplicati, abbiamo dovuto gestire le borse realizzate con più di un materiale; a tal proposito, abbiamo deciso di non considerare le etichette di materiali presenti in piccola percentuale nella borse, mentre per i materiali presenti in percentuali simili abbiamo duplicato l’immagine cambiandone l’etichetta.

Successivamente abbiamo eliminato le immagini di borse il cui materiale aveva un supporto totale nel dataset inferiore a 5.

Al termine di tutti questi step, il dataset risultante contiene 1402 immagini, di cui 220 di borse da uomo e 1182 di borse da donna.

I supporti per i materiali presenti nel dataset sono mostrati in tabella 1.

---

<sup>1</sup><https://www.mytheresa.com/>

**Tabella 1** Supporti del dataset per la classificazione

Materiale	Supporto
Calf leather	409
Fabric	352
Leather	329
Polyurethane	34
Canvas	18
Polyester	24
Cow leather	87
Lamb leather	116
Raffia	20
Goat Lether	13

Come si può notare, sono presenti 10 materiali e da ciò deriva la dimensione dell'ultimo layer della head di classificazione utilizzata nelle reti testate.

Un altro elemento importante da notare è il forte sbilanciamento del dataset nei confronti di alcuni materiali come "Calf leather", "Fabric" e "Leather". Per cercare di bilanciarlo abbiamo effettuato della data augmentation sulle classi con supporto inferiore a 100 applicando alle immagini rotazioni, zoom, variazioni di luminosità e flip orizzontali e verticali. Naturalmente, le immagini augmentate sono state utilizzate solo ed esclusivamente per il training, lasciando la validation e il testing invariati.

### 3.1.2 Esperimenti realizzati

Per il task di classificazione sono state realizzate e testate quattro reti neurali convoluzionali. In particolare, abbiamo utilizzato le reti Mobilenet\_v3, Resnet50, VGG16 e Inception. In tutte queste reti, la head di classificazione è stata realizzata con 3 layer fully connected rispettivamente da 1024, 512 e 10 neuroni; i primi due layer sono attivati dalla ReLu, mentre il terzo dalla Softmax. Inoltre, i pesi dei layer convoluzionali di tutte le reti sono state inizializzati con quelli derivanti dall'addestramento su imagenet.

Per ogni tipologia di rete sono stati effettuati diversi test andando ad addestrare con differenti iperparametri. In particolare, sono stati testati diversi valori di learning rate, batch size, numero di layer della rete congelati e tipologia di pooling effettuato all'uscita dell'ultimo layer convoluzionale.

## 3.2 Detection dei loghi

Il task di detection dei loghi consiste nel rilevamento, tramite bounding box, dei loghi presenti nelle immagini di borse date in input alla rete neurale e la loro classificazione in base al brand.

### 3.2.1 Dataset utilizzato e relative sfide

Inizialmente, si era pensato di utilizzare l'unione di due dataset per poter proseguire con l'addestramento della rete: LogoDet-3K<sup>2</sup> e QMUL-OpenLogo<sup>3</sup>. In un secondo momento, però, ci siamo accorti che il dataset non era adatto alla realizzazione del nostro obiettivo: di lì, abbiamo creato e annotato (utilizzando Roboflow<sup>4</sup>) l'intero dataset.

Nello specifico, il dataset è costituito da 24 classi, ciascuna con 80 supporti (55 utilizzati per il training, 15 per la validation e 10 per il testing), corrispondenti ad alcuni dei più famosi brand di alta moda presenti nel mercato, quali:

1. Balenciaga;
2. Burberry;
3. Celine;
4. Dior;
5. DKNY;
6. Elisabetta Franchi;
7. Fendi;
8. Ferragamo;
9. Givenchy;
10. Goyard;
11. Gucci;
12. Guess;
13. Liu-Jo;
14. Loewe;
15. Louis Vuitton;
16. Marc Jacobs;
17. MCM;
18. Michael Kors;
19. Moncler;
20. Prada;
21. Tory Burch;
22. Valentino;
23. Versace;
24. Yves Saint Laurent.

La traccia iniziale prevedeva la detection di 26 brand: quelli sovraelencati, ai quali si aggiungevano Hermes e Max Mara (che non sono stati integrati nel dataset poichè, per quanto riguarda Hermes, le borse risultavano essere molto anonime non avendo nessun tratto distintivo da utilizzare per la detection, mentre per Max Mara risultavano esserci pochi supporti).

Il dataset è stato poi aumentato variando le immagini andando a effettuare rotazioni e cambi di luminosità.

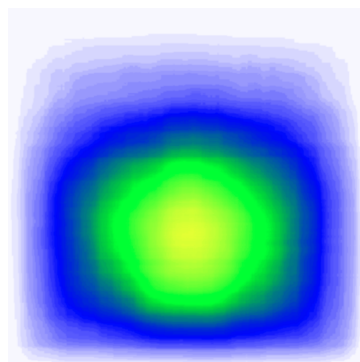
Nell'immagine 1 è riportata la heatmap delle label del dataset: è possibile vedere

---

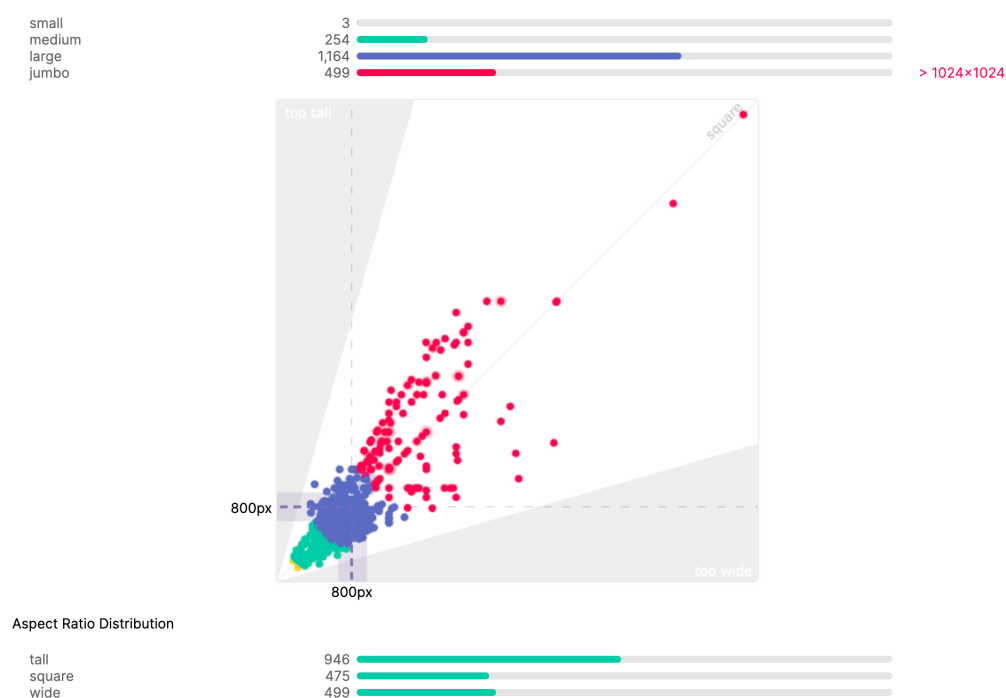
<sup>2</sup><https://www.kaggle.com/datasets/lyly99/logodet3k>

<sup>3</sup><https://hangsu0730.github.io/qmul-openlogo/>

<sup>4</sup><https://roboflow.com/>



**Figura 1** Heatmap dei loghi

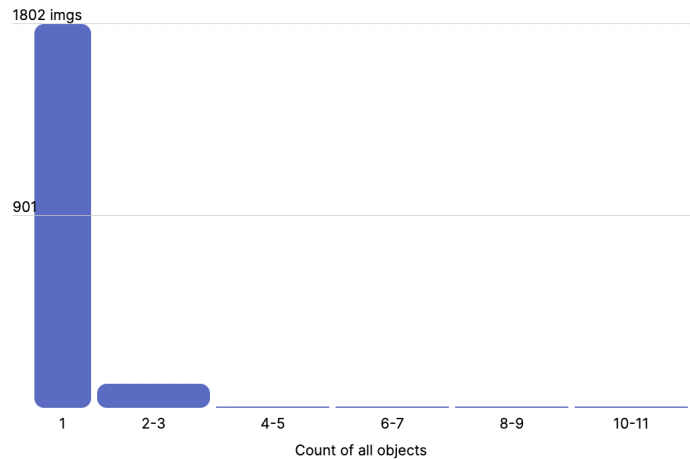


**Figura 2** Distribuzione in base alla grandezza delle immagini

la distribuzione nelle foto del dataset dei loghi dei diversi brand, in particolare, in corrispondenza dei colori caldi vi è una maggiore frequenza, mentre in corrispondenza dei colori freddi vi è una minor frequenza.

Nella figura 2 è riportata la distribuzione delle immagini del dataset in base alla loro dimensione.

Nella figura 3 è possibile valutare quanti label per immagine ci sono nel dataset.



**Figura 3** Numero di label per immagine

### 3.2.2 Esperimenti realizzati

Per il task di detection, è stata utilizzata una sola rete: la YOLO v8, pubblicata nel gennaio del 2023 da Ultralytics<sup>5</sup>. In particolare, YOLOv8<sup>6</sup> è un modello all'avanguardia che si basa sul successo delle versioni precedenti di YOLO e introduce nuove funzionalità e miglioramenti per aumentare prestazioni e flessibilità: una nuova backbone, una nuova head di detection senza ancore e una nuova funzione di loss. YOLOv8 è anche altamente efficiente rispetto alle altre reti di detection e può essere eseguito su una varietà di piattaforme hardware, dalle CPU alle GPU.

Nello specifico, la head è stata aggiornata separando le head di classificazione e detection e passando da un sistema Anchor-Based a uno Anchor-Free (prevede direttamente il centro di un oggetto invece dell'offset da una casella di ancoraggio nota). Inoltre, YOLOv8 aumenta le immagini durante il training: ad ogni epoca, il modello vede variazioni leggermente diverse delle immagini fornite inizialmente (come mosaici a 4 di immagini già presenti nel dataset, tecnica di data augmentation che non viene utilizzata nelle ultime 10 epoche).

Tuttavia, non è ancora stato pubblicato nessun paper a riguardo.

## 4 Risultati

In questa sezione andremo a presentare i risultati ottenuti dalle prove effettuate per i task di classificazione e di detection.

### 4.1 Classificazione

Abbiamo effettuato molteplici addestramenti delle reti, tanti quanti sono le diverse combinazioni di parametri da noi scelte.

<sup>5</sup><https://ultralytics.com/yolov8>

<sup>6</sup>[https://blog.csdn.net/qq\\_29788741/article/details/128626422](https://blog.csdn.net/qq_29788741/article/details/128626422)

**Tabella 2** Migliori risultati Inception, con 0 layer non congelati, learning rate 0.001, batch size 32 e average pooling.

	Precision	Recall	f1-Score	Support
leather	0.44	0.52	0.47	33
fabric	0.61	0.71	0.66	35
canvas	0.00	0.00	0.00	2
calf-leather	0.59	0.73	0.65	41
polyester	0.00	0.00	0.00	2
lamb-leather	0.86	0.50	0.63	12
raffia	0.00	0.00	0.00	2
goat-leather	1.00	1.00	1.00	1
polyurethane	0.00	0.00	0.00	4
cow-leather	1.00	0.11	0.20	9
micro avg	0.57	0.57	0.57	141
macro avg	0.45	0.36	0.36	141
weighted avg	0.57	0.57	0.54	141
samples avg	0.57	0.57	0.57	141
accuracy	0.5673758865248227			

In particolare, le reti addestrate sono state quattro: Mobilenet v3, Resnet50, Inception e VGG16.

I diversi parametri testati sono stati:

- Layer della CNN non congelati: 0,3,6 (VGG16), 0,15,30 (Resnet50), 0,20,40 (Mobilenet) e 0,40,80 (Inception);
- Learning rate: 0.0001, 0.001 e 0.005 (per tutte le reti);
- Batch size: 16 e 32 (per tutte le reti);
- Pooling alla fine dell'ultimo layer convoluzionale: avg e max (per tutte le reti);
- Dropout: 0.5 (per tutte le reti tra i layer fully connected);
- Ottimizzatore: SGD con momento 0.9 (per tutte le reti).

È possibile consultare i risultati degli addestramenti delle diversi reti cliccando sui seguenti link:

- [Resnet50](#);
- [Inception](#);
- [Mobilenet v3](#);
- [VGG16](#).

I pesi risultanti dai migliori addestramenti di ogni rete possono essere reperiti al seguente [link OneDrive](#).

È interessante, però, notare le differenze tra i risultati migliori delle reti addestrate, riportati nelle tabelle 2, 3, 4, 5, mentre nella tabella 6 è possibile vedere il risultato del training della rete che ha dato il risultato migliore (Resnet), addestrato questa volta per 100 epoche.

Le figure 4 e 5 mostrano due esempi di risultato della classificazione effettuata con ResNet50 addestrata per 100 epoche.



**Tabella 3** Migliori risultati Resnet50, con 15 layer non congelati, learning rate 0.001, batch size 32 e max pooling.

	Precision	Recall	f1-Score	Support
lamb-leather	0.78	0.58	0.67	12
calf-leather	0.64	0.88	0.74	41
canvas	1.00	0.50	0.67	2
goat-leather	0.50	1.00	0.67	1
cow-leather	0.60	0.33	0.43	9
fabric	0.80	0.91	0.85	35
raffia	1.00	0.50	0.67	2
polyester	0.00	0.00	0.00	2
leather	0.65	0.45	0.54	33
polyurethane	0.33	0.25	0.29	4
micro avg	0.69	0.69	0.69	141
macro avg	0.63	0.54	0.55	141
weighted avg	0.68	0.69	0.67	141
samples avg	0.69	0.69	0.69	141
accuracy	0.6879432624113475			

**Tabella 4** Migliori risultati MobileNet, con 0 layer non congelati, learning rate 0.001, batch size 32 e average pooling.

	Precision	Recall	f1-Score	Support
lamb-leather	0.71	0.83	0.77	12
polyester	0.00	0.00	0.00	2
leather	0.62	0.73	0.67	33
fabric	0.63	0.83	0.72	35
calf-leather	0.62	0.56	0.59	41
canvas	0.00	0.00	0.00	2
raffia	1.00	0.50	0.67	2
cow-leather	0.00	0.00	0.00	9
goat-leather	1.00	1.00	1.00	1
polyurethane	0.00	0.00	0.00	4
micro avg	0.62	0.62	0.62	141
macro avg	0.46	0.45	0.44	141
weighted avg	0.56	0.62	0.59	141
samples avg	0.62	0.62	0.62	141
accuracy	0.624113475177305			

## 4.2 Detection

Il modello utilizzato, come detto precedentemente, è YOLO v8.

L'addestramento è stato effettuato per 50 epoche con le impostazioni di default consigliate nella documentazione di Ultralytics<sup>7</sup> tranne che per la dimensione delle immagini che è stata impostata su 800x800. Gli andamenti di loss, precision, recall e mean average precision durante il training di YOLO sono riportati in figura 6.

<sup>7</sup><https://docs.ultralytics.com/usage/cfg/#train>

**Tabella 5** Migliori risultati VGG16, con 6 layer non congelati, learning rate 0.0001, batch size 16 e average pooling.

	Precision	Recall	f1-Score	Support
leather	0.56	0.73	0.63	33
polyurethane	0.00	0.00	0.00	4
lamb-leather	0.60	0.25	0.35	12
goat-leather	0.00	0.00	0.00	1
fabric	0.81	0.83	0.82	35
raffia	0.40	1.00	0.57	2
calf-leather	0.68	0.73	0.71	41
polyester	0.00	0.00	0.00	2
cow-leather	0.75	0.33	0.46	9
canvas	1.00	0.50	0.67	2
micro avg	0.65	0.65	0.65	141
macro avg	0.48	0.44	0.42	141
weighted avg	0.65	0.65	0.63	141
samples avg	0.65	0.65	0.65	141
accuracy	0.6524822695035462			

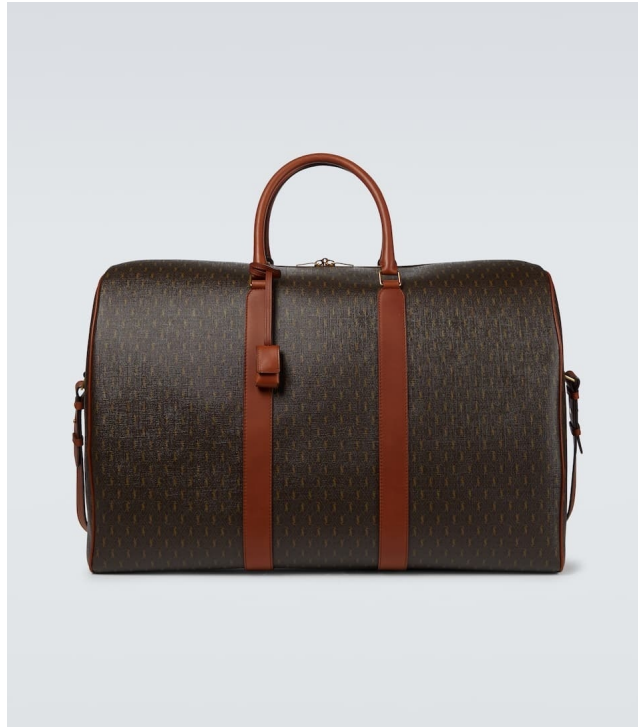
**Tabella 6** Migliori risultati Resnet per 100 epoche, con 15 layer non congelati, learning rate 0.001, batch size 32 e max pooling.

	Precision	Recall	f1-Score	Support
goat-leather	1.00	1.00	1.00	1
raffia	1.00	1.00	1.00	2
polyurethane	0.00	0.00	0.00	4
calf-leather	0.79	0.76	0.77	41
leather	0.68	0.70	0.69	33
lamb-leather	0.67	0.67	0.67	12
fabric	0.67	0.86	0.75	35
cow-leather	0.75	0.33	0.46	9
canvas	0.00	0.00	0.00	2
polyester	0.33	0.50	0.40	2
micro avg	0.70	0.70	0.70	141
macro avg	0.59	0.58	0.57	141
weighted avg	0.69	0.70	0.69	141
samples avg	0.70	0.70	0.70	141
accuracy	0.7021276595744681			

Una volta terminato l'addestramento, il modello è stato verificato con il dataset di test. Le performance sono riportate in tabella 7.

Nella figura 7 è possibile visualizzare la matrice di confusione del nostro modello.

Alcuni esempi di detection sono riportati in figura 8, tutte le immagini con le relative predizioni possono essere trovate al seguente [link GitHub](#).



**Figura 4** Esempio di classificazione  
Label: 4.SAINT-LAURENT\_fabric  
Ground truth: [0. 0. 0. 0. 0. 0. **1.** 0. 0. 0.]  
Prediction: [0.0025 0.0015 0.0052 0.0101 0.0415 0.0020 **0.9138** 0.0097 0.0043 0.0088]

### 4.3 Consumi energetici

Durante gli esperimenti abbiamo utilizzato il tracker di emissioni fornito dalla libreria CodeCarbon<sup>8</sup>. Gli esperimenti sono stati tutti effettuati sul server del VRAI equipaggiato con Intel(R) Xeon(R) Silver 4214 CPU @ 2.20GHz (CPU), GeForce RTX 2080 Ti (GPU) e 220 GiB di RAM. I risultati del tracking dei consumi sono riportati in tabella 8 e sono riferiti agli addestramenti delle reti con i migliori risultati.

Come si può notare, nonostante sia la rete più performante per il task di classificazione, Resnet50 non è la più dispendiosa in termini energetici, è tuttavia al secondo posto in questa classifica.

Notevoli sono le performance di Mobilenet v3 che è la meno dispendiosa in termini energetici ed offre comunque delle buone performance, ciò è dovuto al fatto che questa rete è stata progettata appositamente per essere eseguita su dispositivi con limitate capacità di calcolo.

---

<sup>8</sup><https://github.com/mlco2/codecarbon>



**Figura 5** Esempio di classificazione

Label: 249\_TOD'S\_leather

Ground truth: [0. 0. 0. 0. **1.** 0. 0. 0. 0. 0.]

Prediction: [0.0060 0.0024 0.0074 0.3222 **0.4934** 0.0178 0.0115 0.1346 0.0024 0.0019]

## 5 Conclusioni

In questo capitolo verranno trattate le conclusioni riguardanti gli esperimenti da noi effettuati per il lavoro assegnatoci.

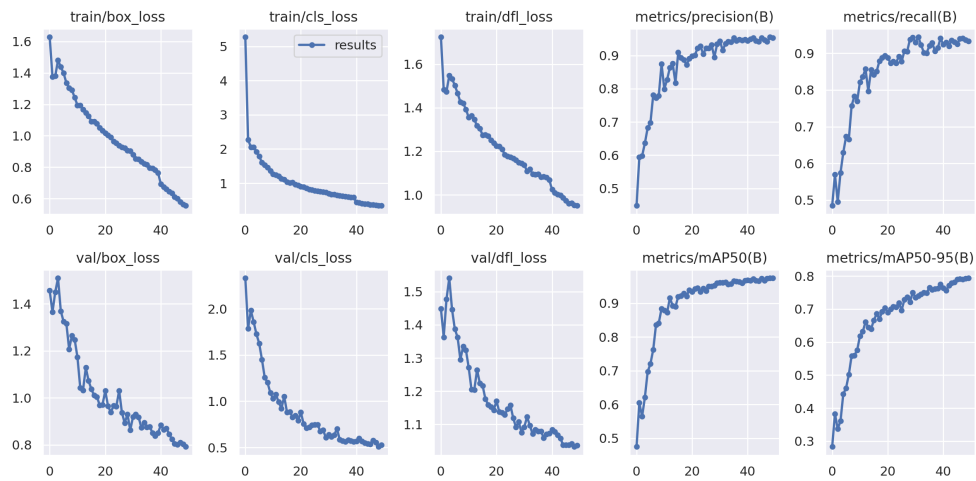
### 5.1 Classificazione

Dai risultati degli esperimenti effettuati, possiamo notare come la rete neurale Resnet50 sia risultata la migliore in termini di performance per il task di classificazione, mentre la peggiore risulta essere stata la rete neurale Inception.

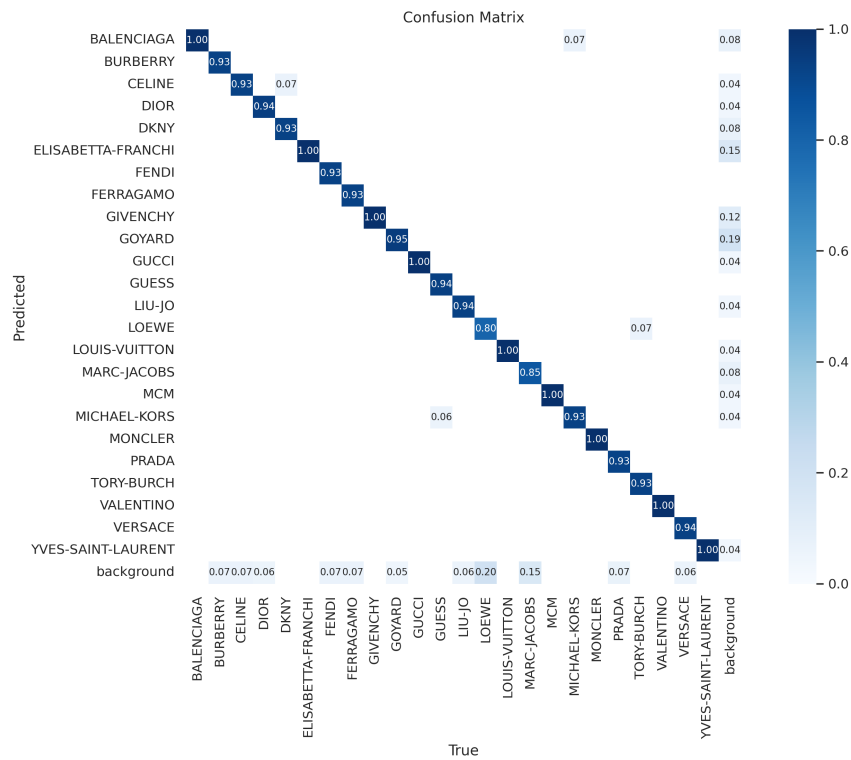
Un risultato molto interessante è però quello della MobileNet v3, che nonostante non abbia i risultati migliori, risulta essere il miglior compromesso tra consumo energetico e performance.

Dai risultati di tutte le reti si può comunque vedere come lo sbilanciamento del dataset abbia influenzato i risultati. Infatti, per i materiali più presenti, in tutte le reti la precision e la recall superano la soglia di 0.6, mentre si osservano performance decisamente scadenti per quanto riguarda i materiali meno presenti nel dataset.

Particolare attenzione merita l'importante margine di miglioramento tra i due addestramenti fatti con la rete Resnet50 (il primo [3] fatto per 50 epoche, il secondo



**Figura 6** Andamento del training di YOLO v8



**Figura 7** Matrice di confusione del modello

**Tabella 7** Performance YOLO v8 sul dataset di test

Classe	Images	Instances	C Precision	C Recall	mAP50	mAP50-95
Tutte	354	377	0.954	0.933	0.974	0.796
BALENCIAGA	354	18	0.946	0.971	0.987	0.813
BURBERRY	354	15	1	0.891	0.991	0.813
CELINE	354	15	0.9	0.933	0.942	0.821
DIOR	354	18	0.959	0.889	0.948	0.551
DKNY	354	15	0.942	0.8	0.971	0.788
ELISABETTA-FRANCHI	354	15	0.906	1	0.995	0.723
FENDI	354	15	1	0.921	0.995	0.898
FERRAGAMO	354	15	1	0.86	0.953	0.697
GIVENCHY	354	16	0.937	0.938	0.988	0.606
GOYARD	354	20	0.882	0.95	0.954	0.849
GUCCI	354	15	0.93	1	0.995	0.951
GUESS	354	17	0.987	0.941	0.964	0.803
LIU-JO	354	16	0.973	0.938	0.983	0.853
LOEWE	354	15	0.925	0.826	0.979	0.77
LOUIS-VUITTON	354	15	0.92	1	0.962	0.751
MARC-JACOBS	354	20	0.939	0.765	0.886	0.639
MCM	354	15	0.973	1	0.995	0.799
MICHAEL-KORS	354	15	0.933	0.926	0.983	0.806
MONCLER	354	14	0.971	1	0.995	0.901
PRADA	354	14	1	0.961	0.995	0.85
TORY-BURCH	354	15	0.975	0.933	0.976	0.823
VALENTINO	354	13	0.976	1	0.995	0.942
VERSACE	354	16	0.987	0.938	0.948	0.794
YVES-SAINT-LAURENT	354	15	0.93	1	0.995	0.854

**Tabella 8** Consumi energetici addestramento reti

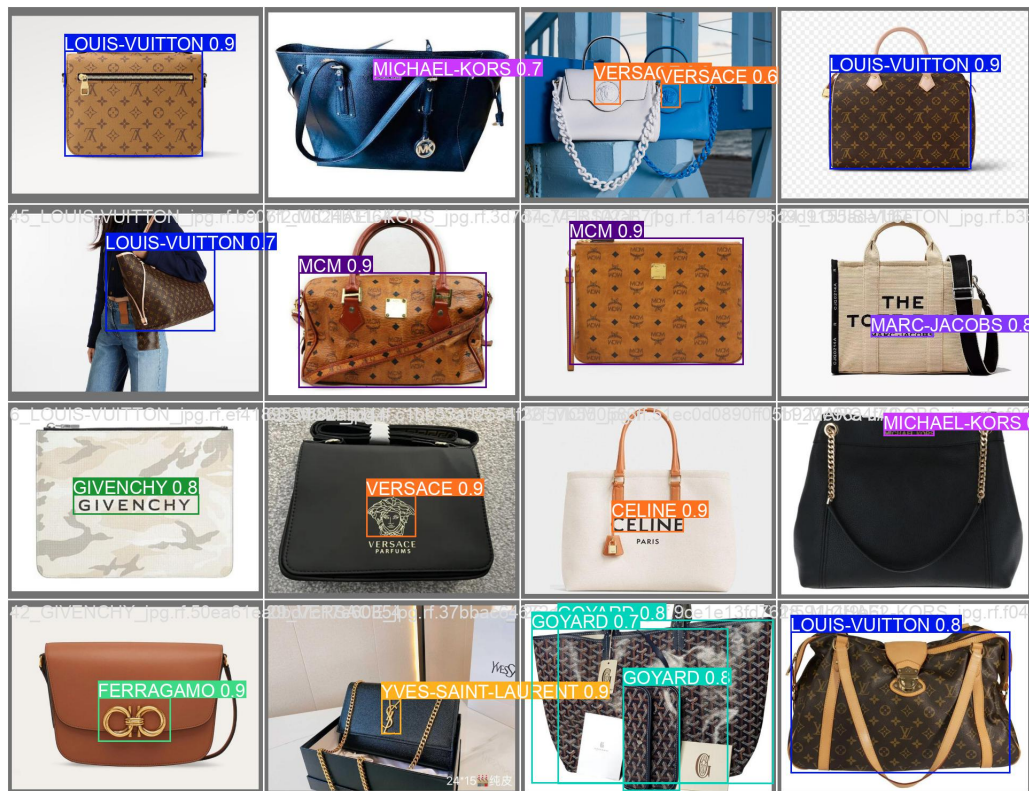
	INCEPTION	MOBILENET	RESNET	RESNET 100 ep	VGG16	YOLO
Consumo [kwh]	2,95E-06	1,72E-06	2,99E-06	6,87E-06	3,49E-06	1,23E-04
Emissioni [kg CO2]	1,95E-06	1,14E-06	1,98E-06	4,54E-06	2,30E-06	8,15E-05
Durata [s]	154	105	165	327	196	4.867

[6] per 100).

## 5.2 Detection

Per quanto riguarda il task di detection, i risultati mostrano delle ottime performance del modello YOLO v8. Dalle prove effettuate, è possibile notare quanto influisca la complessità della rete sui consumi e sulle emissioni: non c'è confronto considerando che nella detection è stata utilizzata la versione small di YOLO v8, anche paragonandola alla rete di classificazione più complessa e dispendiosa, la VGG16.

Dalla matrice di confusione, invece, è possibile notare che è difficile che la rete sbagli a classificare un logo, ma è più comune che non venga rilevato alcun logo nella



**Figura 8** Esempio di detection

foto oppure che parti dello sfondo vengano rilevate come logo. Da ciò, deriva che le metriche di precision e recall per la classificazione sono molto alte, mentre la mAP è più bassa.

### 5.3 Sviluppi futuri

In lavori futuri sarebbe sicuramente utile effettuare degli esperimenti riguardanti le reti di classificazione utilizzando un dataset completamente bilanciato, magari espandendolo con altre immagini raccolte dal catalogo MyTheresa di altri paesi. In questo modo sarebbe possibile vedere le effettive prestazioni delle reti con tutti i 10 materiali senza alcun bias.

Riguardo il task di detection sarebbe interessante confrontare i risultati ottenuti con quelli di altre reti di detection sia basate su convoluzioni (SSD, RCNN) che con la rete di detection basata su transformer DETR[5], quest'ultima mostra risultati molto promettenti sul dataset COCO nonostante sia meno complessa in termini di impiego di risorse delle reti di detection convoluzionali tradizionali.

## 6 Ringraziamenti

Nell'ultima sezione della relazione, ci teniamo a ringraziare la Professoressa Migliorelli per gli insegnamenti, l'impegno, la dedizione, la sopportazione e la pazienza nei nostri confronti.

Un particolare ringraziamento va al Professor Balloni, per i consigli e la disponibilità e la cortesia durante tutta la realizzazione del progetto.

Ringraziamo, infine, tutto il VRAI per averci dato la disponibilità del loro server per poter effettuare i diversi esperimenti.

## Riferimenti bibliografici

- [1] Braga, D., Braz, L., Souza, C., Silva, T.: Handbags classification model via deep learning. (2019)
- [2] Hou, S., Li, J., Min, W., Hou, Q., Zhao, Y., Zheng, Y., Jiang, S.: Deep Learning for Logo Detection: A Survey (2022)
- [3] Manandhar, D., Yap, K.H., Bastan, M., Heng, Z.: Brand-Aware Fashion Clothing Search using CNN Feature Encoding and Re-ranking (2018)
- [4] Kiapour, M.H., Piramuthu, R.: Brand & Logo: Visual Analysis of Fashion Brands (2018)
- [5] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-End Object Detection with Transformers (2020)