

Report_{Esercizio}₂

Leonardo Del Bene

April 2025

1 Report Esercizio 2

L'obiettivo di questo esercizio è implementare una **Class Activation Map (CAM)** per visualizzare su quali regioni dell'immagine la rete neurale si concentra durante la classificazione.

A tal fine, ho confrontato la CAM ottenuta dalla mia implementazione di una CNN (realizzata per l'Esercizio 1.3) con quella generata da una ResNet-18.

Gli esperimenti sono stati condotti utilizzando il dataset ImageNette-160, una versione ridotta del dataset ImageNet.

Per seguire l'implementazione del CAM suggerita dal docente, è stato necessario modificare leggermente la mia CNN, aggiungendo un layer di *Global Average Pooling* prima dell'ultimo layer lineare, che non era presente nella versione originale.

Per quanto riguarda la ResNet-18, ho utilizzato una versione pre-addestrata su ImageNet, modificando la testa del classificatore per adattarla a un numero di classi pari a 10 (anziché 1000). Su questo modello ho eseguito un *fine-tuning*, congelando tutti i pesi tranne quelli dell'ultimo layer. Il fine-tuning è durato 5 epoche e il modello ha raggiunto un'accuratezza di circa il 96–97% sul validation set.

La CNN, invece, è stata addestrata da zero su ImageNette per un totale di 100 epoche, salvando il modello con le migliori performance. Alla fine dell'addestramento, ha raggiunto un'accuratezza del 72% sul validation set. La performance ridotta è dovuta al fatto che, per poter gestire un batch di dimensioni adeguate in memoria GPU, è stato necessario eseguire un *RandomCrop* delle immagini a dimensioni 64×64 .

In Figura 1 è mostrata la Class Activation Map ottenuta. Nella prima riga sono riportate le immagini originali (una per ciascuna delle 10 classi del dataset); nella seconda riga è rappresentata la CAM generata dalla ResNet, mentre nella terza riga è mostrata la CAM relativa alla CNN.

Come si può osservare, la ResNet tende a focalizzarsi con precisione sull'oggetto principale dell'immagine, mentre la CNN mostra una distribuzione dell'attenzione più ampia. Tuttavia, nonostante la CNN abbia raggiunto solo il 72% di accuratezza, riesce comunque a porre l'attenzione sugli oggetti corretti in diversi casi — ad esempio nel caso del lettore di cassette, delle palline da golf o della pompa di benzina. In altri casi, invece, questa focalizzazione è meno evidente.

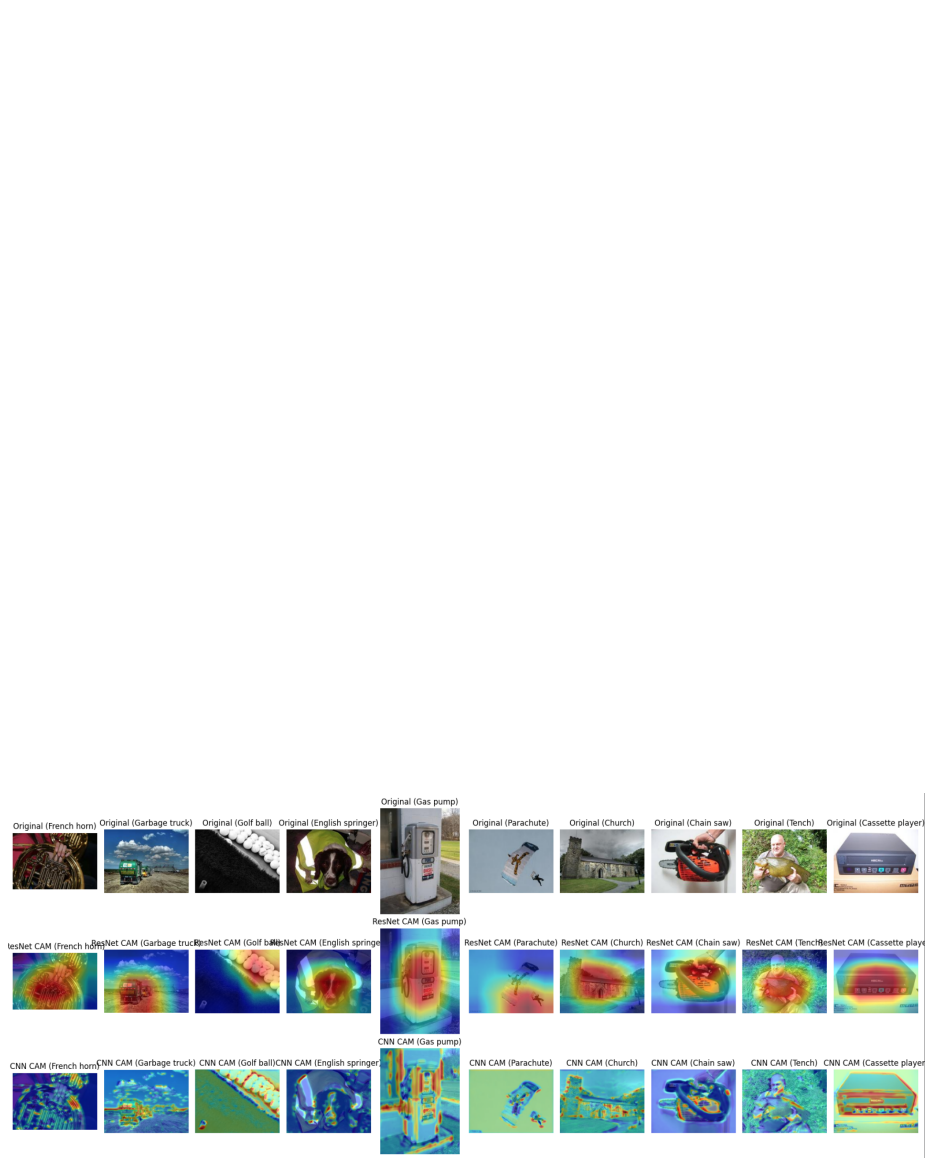


Figure 1: Class Activation Map