

# Tecniche di Deep Learning e Object Detection per la foto-identificazione dei cetacei

---

Gianvito Losapio

Laurea Triennale in Ingegneria Informatica e dell'Automazione

Tesi di Laurea in Calcolo Numerico

Anno Accademico 2018 - 2019

Relatore: Prof. Tiziano Politi

Correlatore: Dott. Vito Renò

# Outline

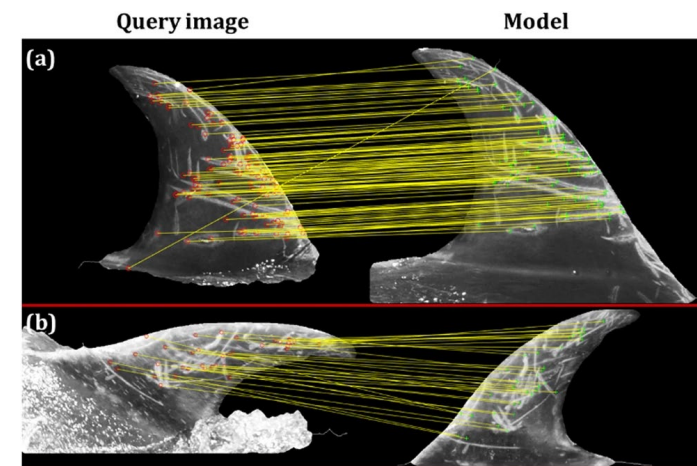
---



- Problema affrontato e suo scopo
- Panoramica delle soluzioni proposte
- Presentazione di due differenti metodologie utilizzate
- Conclusioni

# Scopo del lavoro e interessi scientifici

- Studiare i cetacei oggi significa provare a comprendere le modalità e gli effetti dell'**alterazione** in atto degli **ecosistemi marini** (Taranto, plastica negli oceani, Artico)
- Il monitoraggio delle specie attraverso raccolta di parametri (es. abbondanza, *site fidelity*) può essere largamente facilitata mediante l'utilizzo di una tecnica non-invasiva basata su algoritmi innovativi e su una grande disponibilità di dati: la **foto-identificazione automatica** degli esemplari
- La specie ***Grampus griseus*** è adatta per l'applicazione di questa tecnica di studio perché le pinne dorsali presentano un *pattern* di segni unico che contraddistingue ciascun individuo all'interno della popolazione (paragonabili alle impronte digitali dell'uomo). Per essi è stato proposto *SPiR* (Maglietta R., Renò V. et. al., *DolFin: an innovative digital platform for studying Risso's dolphins in the Northern Ionian Sea*, Scientific Reports 8, 2018)

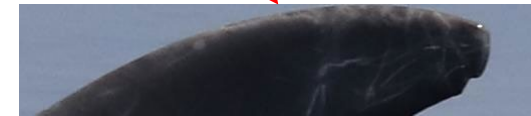
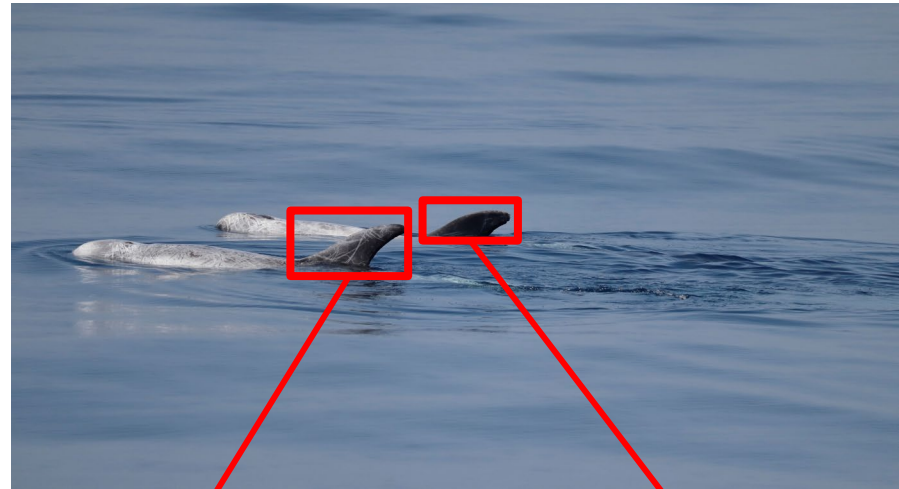
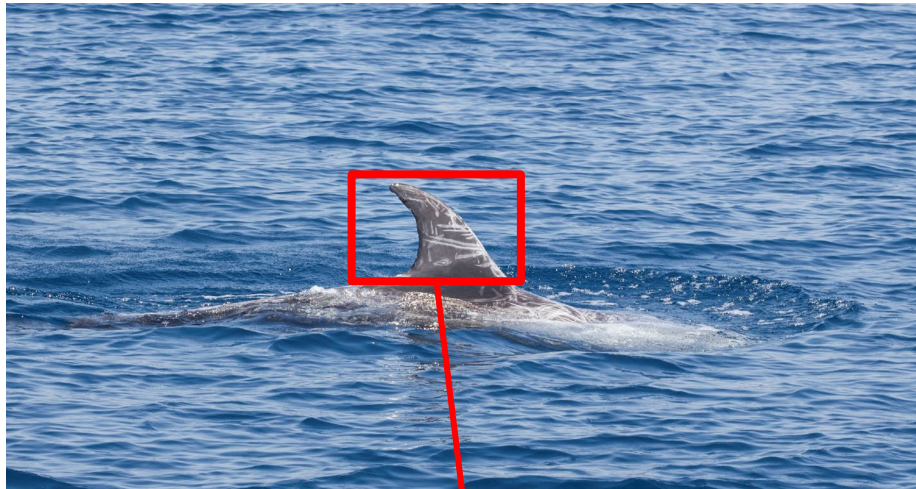


# Problema



La completa automatizzazione della foto-identificazione resta vincolata alla necessità di ritagliare manualmente le pinne dorsali a partire dalle immagini iniziali (**collo di bottiglia**)

- Introduzione di una tecnica di **riconoscimento e ritaglio automatico** di pinne dorsali di delfini a partire da una collezione di immagini scattate durante campagne di avvistamento in mare



# Soluzioni proposte



- Sono stati sviluppati due differenti programmi Matlab che consentono di identificare e ritagliare pinne dorsali a partire da una collezione di immagini con i seguenti passaggi salienti:
  1. **Segmentazione** dell'immagine basata sui colori
  2. **Classificazione** delle immagini ritagliate
- Il lavoro è stato concentrato principalmente sull'introduzione di:
  1. Una tecnica innovativa di segmentazione automatica
  2. Un criterio di classificazione basato su *deep learning*
- Dati utilizzati per gli esperimenti:
  - ~10.000 foto scattate tra il 2013 ed il 2018 nel Golfo di Taranto
  - ~14.000 foto scattate nel 2018 a largo delle isole Azzorre





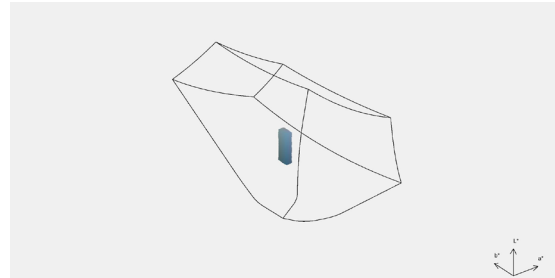
# Segmentazione: poliedri nello spazio Lab

- **Metodo innovativo:** definizione a priori di un insieme di tonalità di colore che permette di distinguere in maniera univoca le pinne dorsali all'interno delle immagini
- Creazione di **5 modelli risolutivi** nello spazio colori CIE 1976  $L^*a^*b^*$  (Lab):
  1. Stima della tonalità predominante del mare per l'individuazione del modello risolutivo di appartenenza
  2. Applicazione del filtro corrispondente al modello individuato per la localizzazione delle pinne

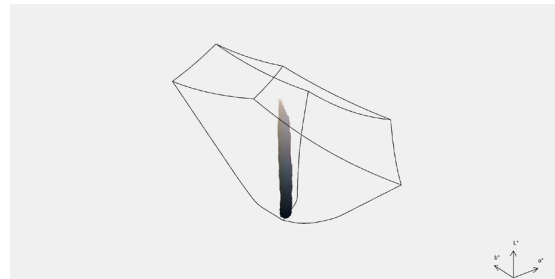
Foto di esempio



Tonalità del mare stimata



Filtro corrispondente applicato

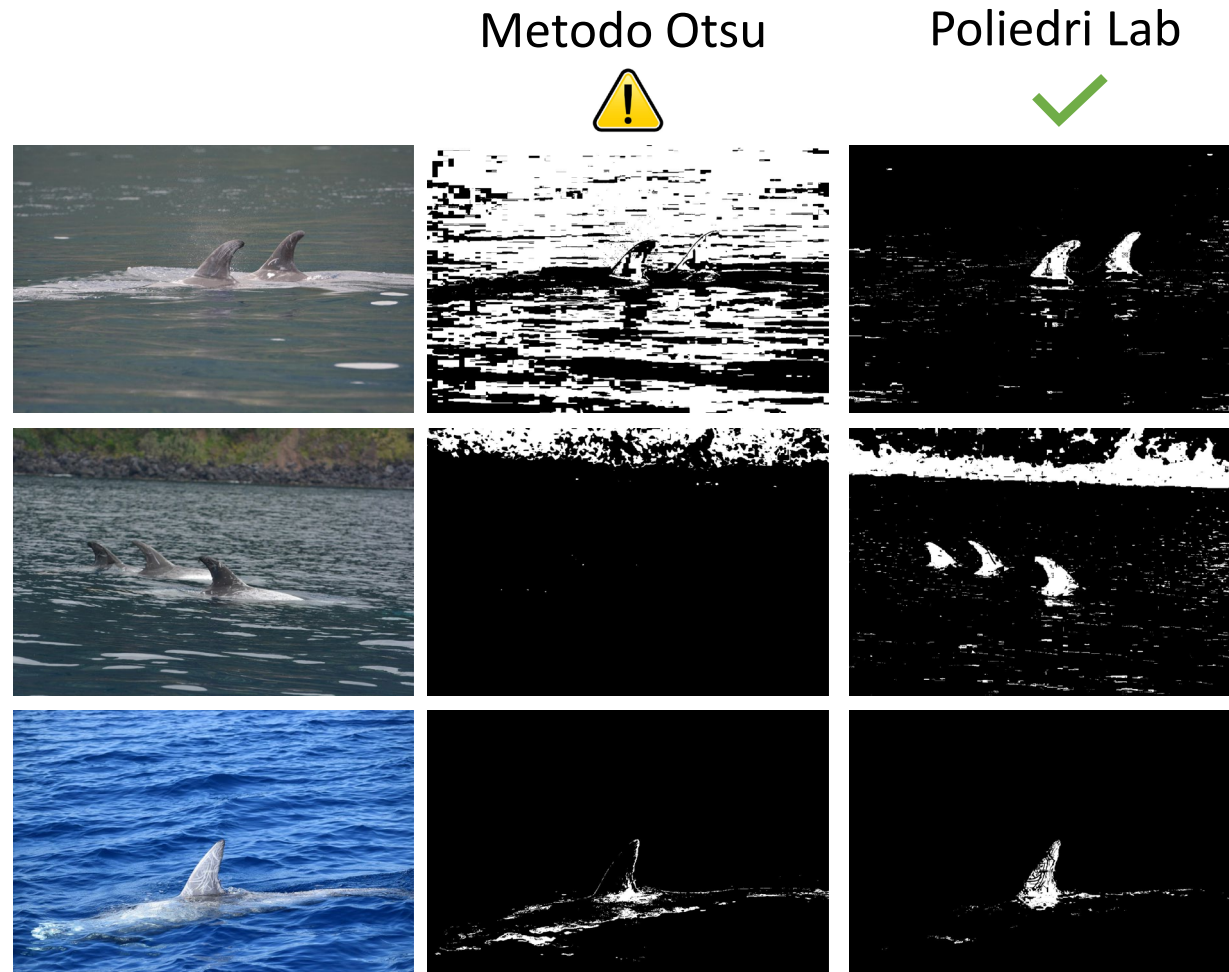


Binarizzazione



# Segmentazione: analisi qualitativa

- Confronto con un metodo di sogliatura allo stato dell'arte (metodo Otsu con soglie multilivello)



# Segmentazione: analisi quantitativa

- **Efficienza** di risoluzione del problema di ***object detection***

$$\eta_{od} = \frac{\text{pinne correttamente ritagliate in modo automatico}}{\text{pinne ritagliabili da un operatore umano}} \times 100$$

- Misura su campione casuale di 500 foto scattate durante diverse campagne di avvistamento

Metodo Otsu	Metodo dei poliedri nello spazio Lab
$\eta_{od} = 58 \%$	$\eta_{od} = 84 \%$

- Guadagno complessivo di **26 punti percentuali** con il metodo innovativo di segmentazione automatica grazie ad un approccio basato sull'intero dominio dei dati a disposizione
- $\eta_{od} = 84 \%$  significa riuscire ragionevolmente a ritagliare, in media, le pinne di tutti gli esemplari avvistati in mare

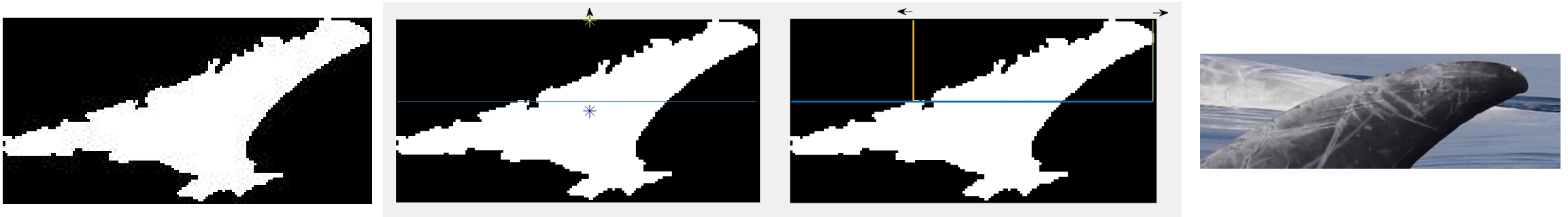


# Filtraggio e ritaglio adattivo di regioni connesse

- Miglioramento del risultato della segmentazione basato su:
  - Applicazione del filtro mediano per la riduzione del rumore «sale e pepe»
  - Filtraggio di regioni connesse sulla base di statistiche verosimili per una pinna dorsale

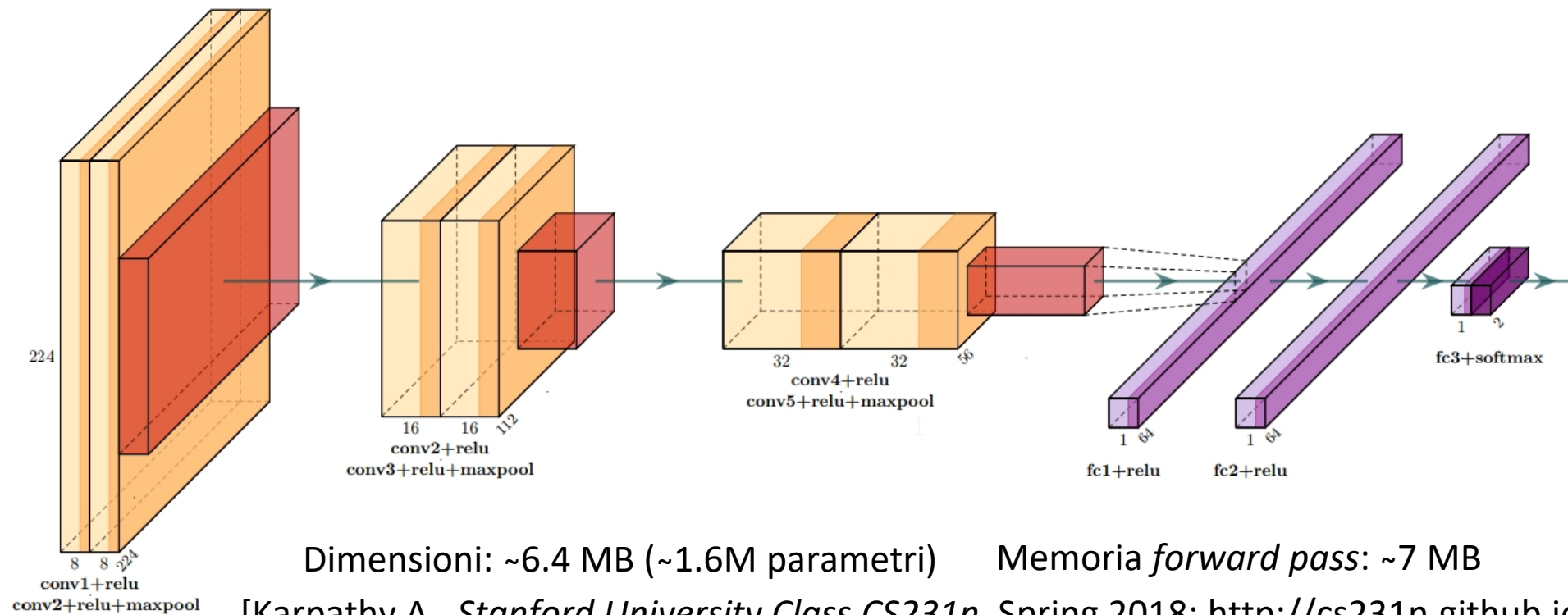


- Creazione di un algoritmo che consente di ottenere un ritaglio preciso in corrispondenza delle pinne a partire da precisi punti geometrici calcolati per ciascuna regione binaria («adattivo»)



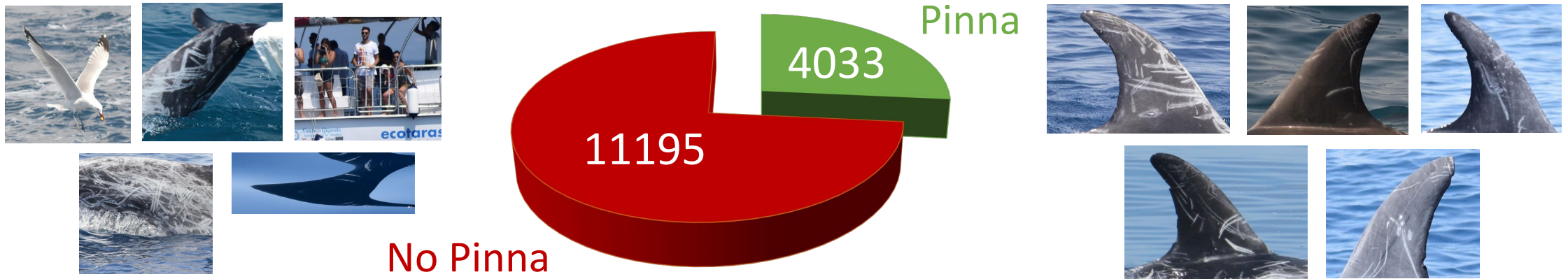
# Classificazione: rete neurale convoluzionale

- **Necessità di una fase di classificazione:** pur migliorando le fasi iniziali, il risultato complessivo è costituito mediamente da un numero di ritagli che non contengono pinne almeno pari a quelli che ne contengono.
- INPUT--> [CONV -> RELU -> CONV -> RELU -> MAXPOOL] \* 3 --> [FC -> RELU] \* 2 --> FC con filtri di dimensione 3x3 e dimezzamento delle dimensioni mediante max pooling



# Classificazione: rete neurale convoluzionale

- Dimensione del dataset: 15.228 ritagli manualmente etichettati (scatti nel Golfo di Taranto)



- Partizionamento secondo la tecnica **k-Fold Cross Validation** stratificata con  $k = 5$  :
  - Training set: 12182 (80%)
  - Test set: 3046 (20%)
- **Addestramento** eseguito con il metodo *stochastic gradient descent with momentum*, con dimensione del *minibatch* pari a 20, numero di epoche pari a 30 e *learning rate* iniziale pari a 0.0003.
- **Image Augmentation** per un bilanciamento «virtuale» delle classi in fase di addestramento (riflessione, rotazione, traslazione)

# Classificazione: rete neurale convoluzionale

- Il **tempo di addestramento** è stato di circa 3 ore e 20 minuti per ciascuna rete, utilizzando la modalità a singolo processore grafico su una macchina equipaggiata di Intel Core i5-6400T @ 2.20 GHz con 8 GB di RAM e Nvidia GeForce 930M con 2 GB dedicati in modalità DDR3.
- Risultato della *5-Fold Cross Validation* come **media** delle prestazioni sui ***fold* di test**

Accuracy	Sensitivity	Specificity
99 %	99 %	99 %

- Le performance sono state, inoltre, misurate su un campione di 20.888 ritagli ottenuti dall'elaborazione di un dataset con caratteristiche completamente differenti (scatti alle isole Azzorre). Come atteso, si è osservato un calo «fisiologico» delle prestazioni ma allo stesso una discreta **capacità di generalizzazione** da parte dei classificatori.

Accuracy	Sensitivity	Specificity
92 %	85 %	95 %

# Conclusioni

---



- La fase del ritaglio manuale rimane un collo di bottiglia in numerose tecniche di foto-identificazione automatica per lo studio di diverse specie animali proposte in letteratura (Buehler et. al, *An automated program to find animals and crop photographs for individual recognition*, Ecological Informatics 50 (2019), 191-196).
- Sebbene la tendenza sia una risoluzione del problema di *object detection* mediante tecniche di machine/deep learning, il **dominio limitato** di questo problema ha consentito di sperimentare un approccio differente mediante tecniche di segmentazione automatica.
- Entrambe le metodologie esaminate presentano i propri *pro/contro*.
  - L'utilizzo dei poliedri nello spazio di colore consente di codificare una soluzione generale a partire da una quantità limitata di immagini, purché rappresentative di numerose condizioni di scatto.
  - L'utilizzo della CNN permette di ottenere una soluzione generale senza una fase manuale di **feature engineering** ma a partire da un insieme di immagini di numerosità molto più elevata. La soluzione così ottenuta è una *black-box*.
- Si è constatata una **forte interdisciplinarità** tra le tecniche del calcolo numerico, dell'*image processing* e del *machine learning*. Esse possono facilitare enormemente l'analisi delle grandi quantità di dati prodotta nell'ambito di studi di carattere ecologico («*ecological informatics*»).