# TWO ALTERNATIVE FORCED CHOICE CLASSIFIERS

Symbolic and Evolutionary Artificial Intelligence's Project

Stefano Petrocchi

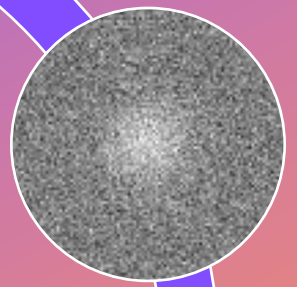
# Introduction

## Objectives

Verify that the ***two alternative forced choice (TAFC)*** problem is simpler than the ***single image classification*** problem

Verify that a ***Siamese network*** is better suited to the TAFC than a ***single image classifier***

Experiments

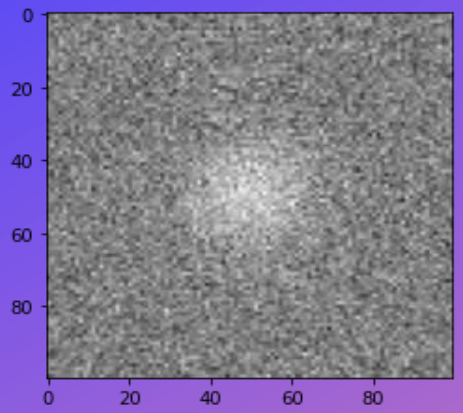


Classification of artificially created signals

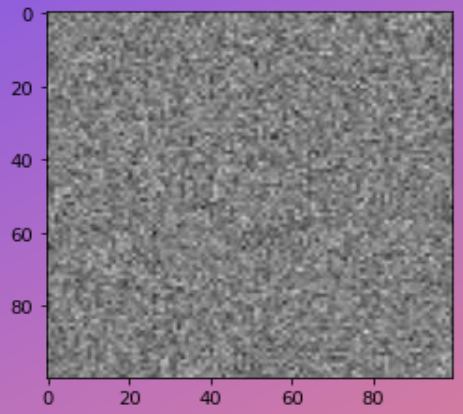
Distinction of leopards from jaguars in a real-world dataset

# Two Alternative Forced Choice

***Single Image Classifier:***




Classifier → $\lambda^+$ Single Image Score



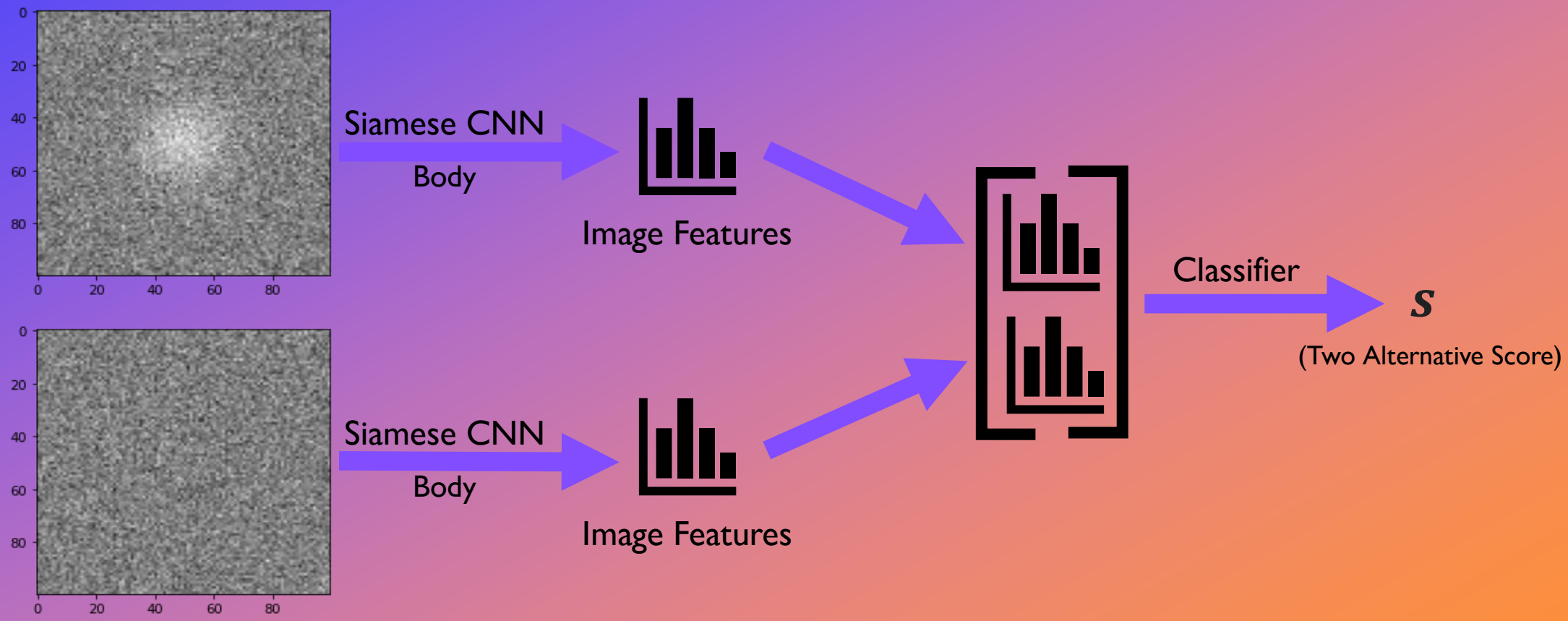

Classifier → $\lambda^-$ Single Image Score

$$s = \begin{cases} 1 & if \ \lambda^+ > \lambda^- \\ 0 & if \ \lambda^+ < \lambda^- \end{cases}$$

Two Alternative Score

# Two Alternative Forced Choice

***Siamese Classifier:***



# Artificial Signals Experiment

Objectives

*Single Image Classification:*

Classify single images by the presence or absence of the signal

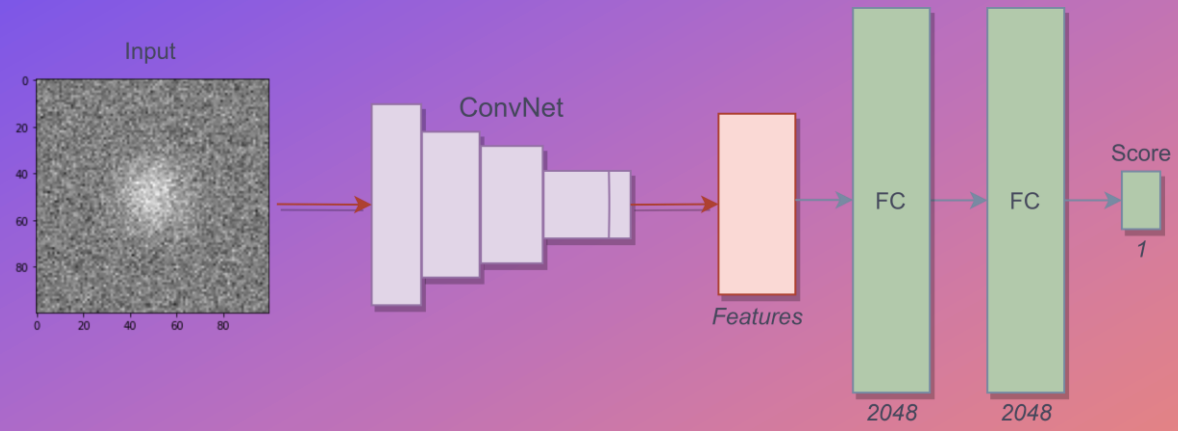
*Two Alternative Images Classification:*

Two images are presented to the network that must distinguish which of the two input contains the signal

# Artificial Signals Experiment

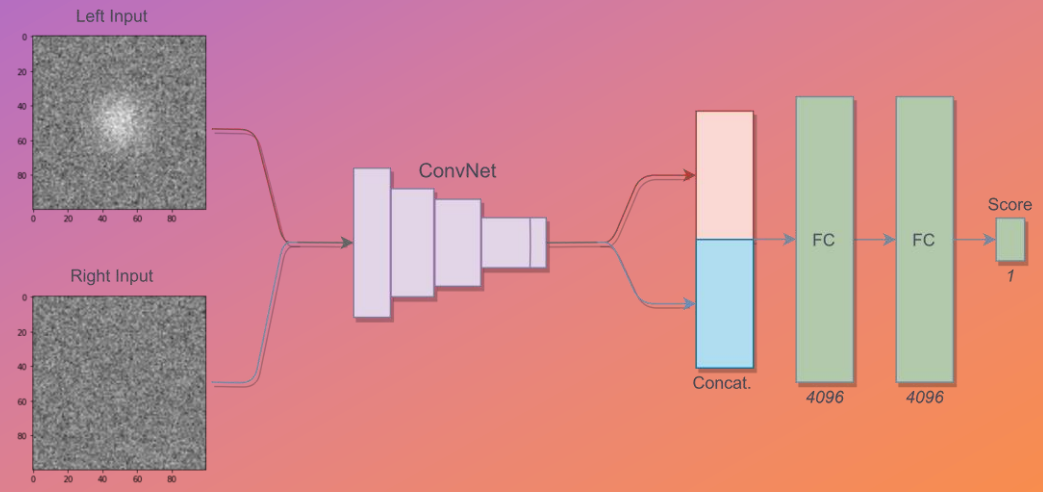
## Models

*Single Image Classifier*



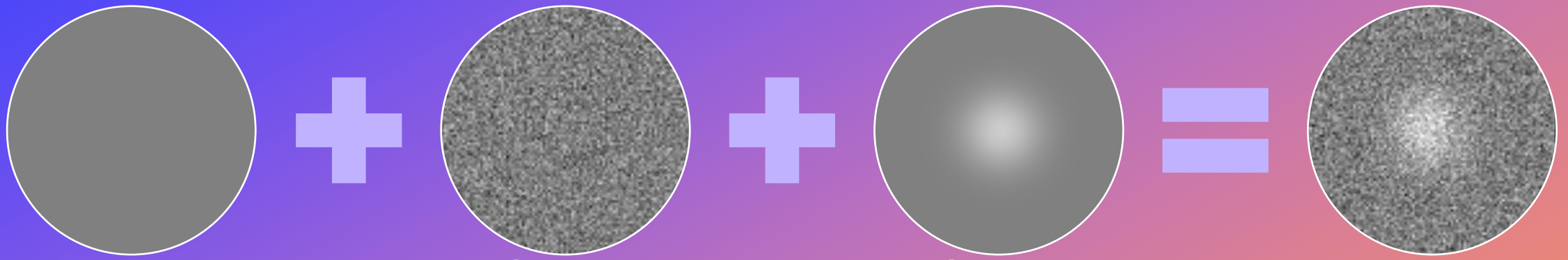

*Siamese Network*

*Koch et al. 2015*



# Artificial Signals Experiment

Signals composition as defined in *Abbey et al. 2002*



Background + Gaussian Noise + Gaussian Signal = Signal Image

**Parameters**

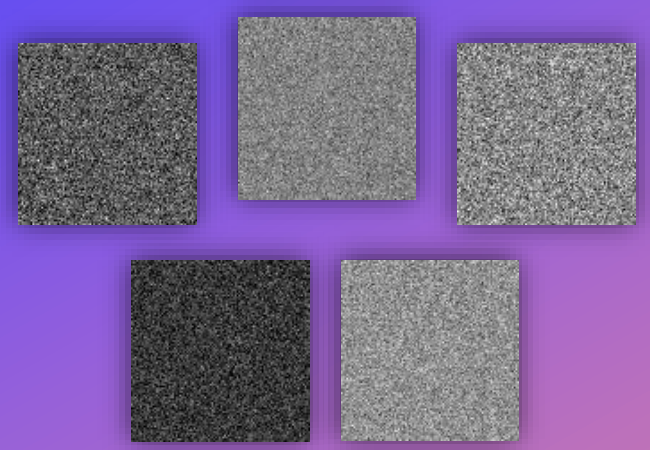
(Decided Heuristically)

| *Background* | *Noise* | | *Signal* | | |
|---|---|---|---|---|---|
| $value \in [0.2, 0.8]$ | $\mu = 0$ | $\sigma \in [0.01, 0.2]$ | $\mu = 0$ | $\sigma \in [5, 25]$ | $\alpha \in [0.005, 0.05]$ |

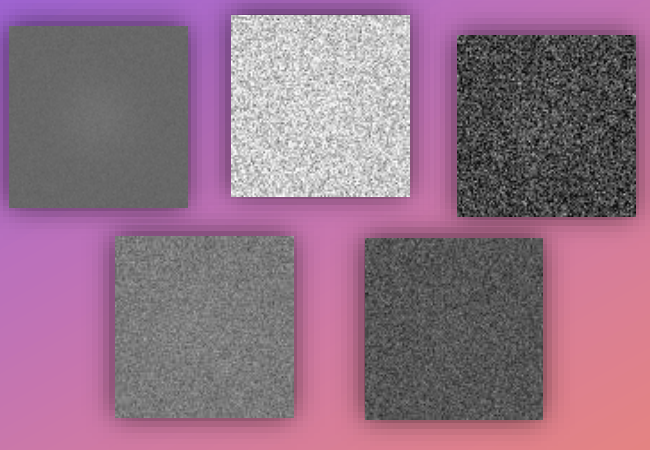
# Artificial Signals Experiment

***Single Image Classification:***

Signal Absent

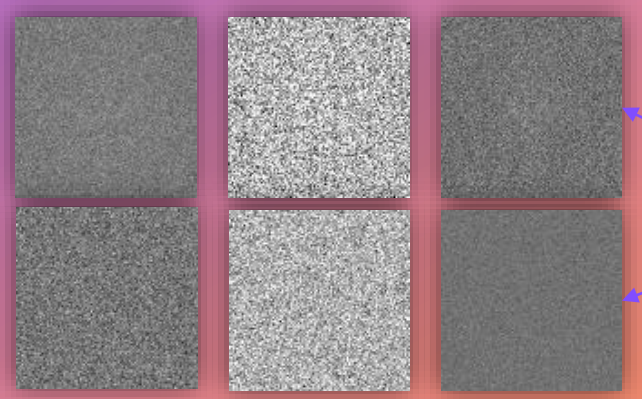


Signal Present

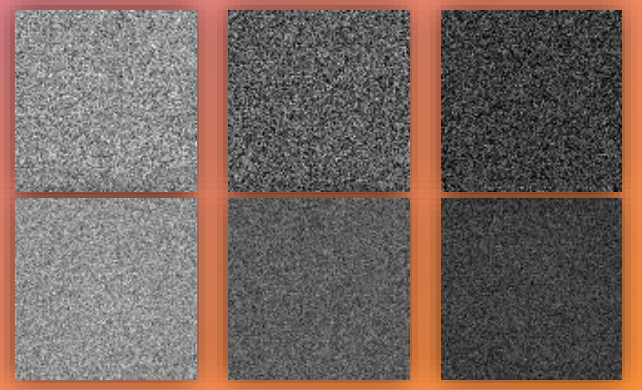


***Two Alternative Images Classification:***

Signal on Top

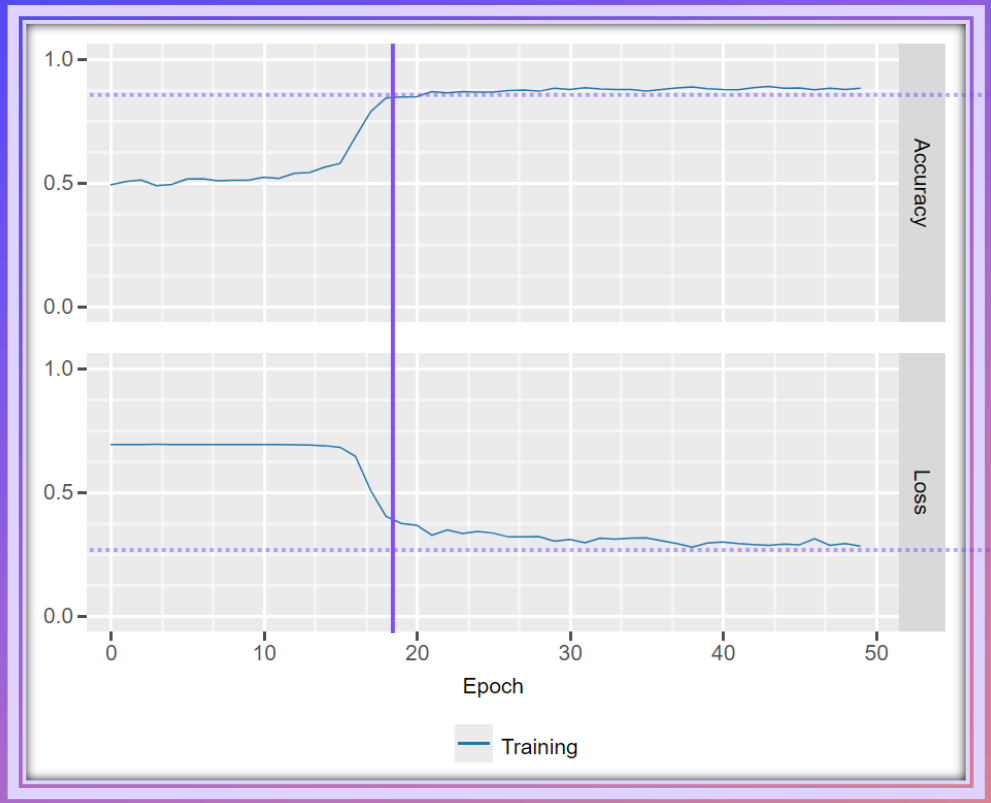



Signal at Bottom

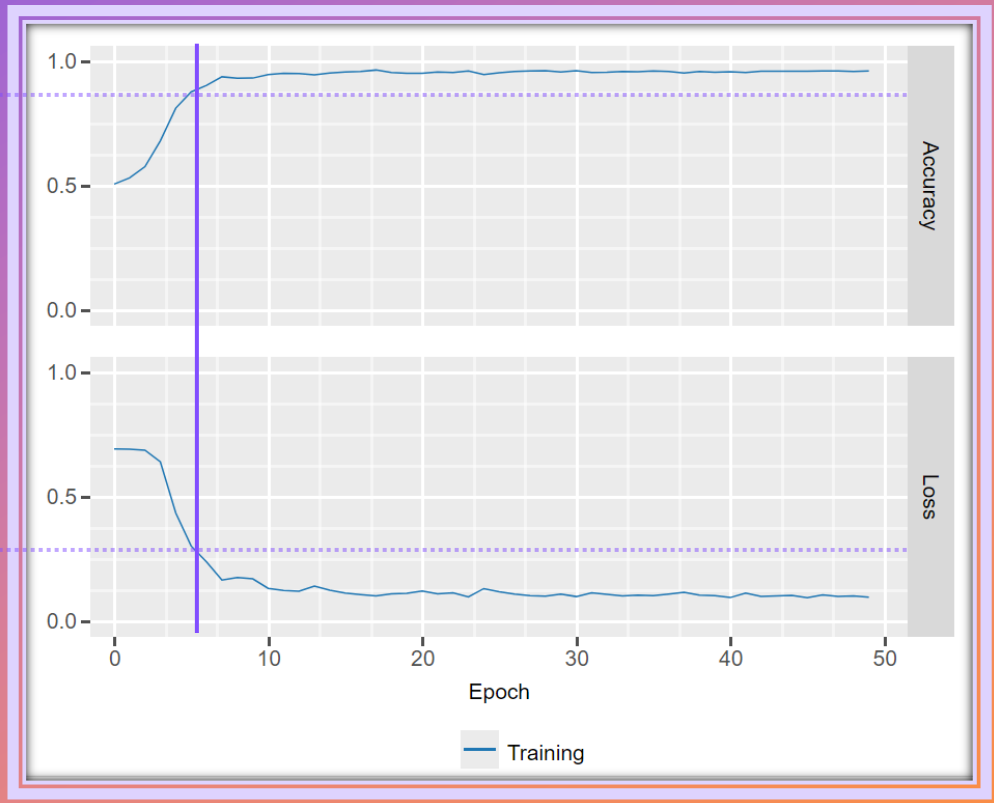


# Artificial Signals Experiment

***Single Image Model:***




***Siamese Model:***

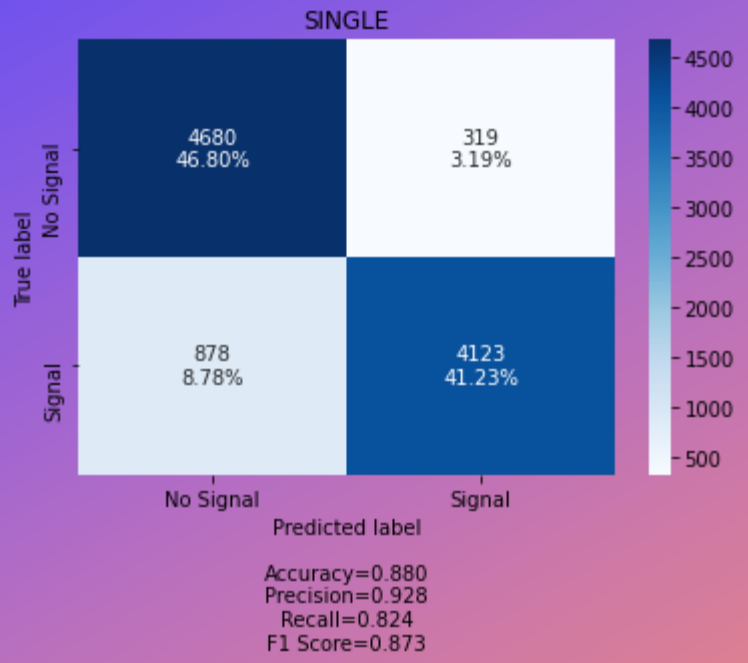


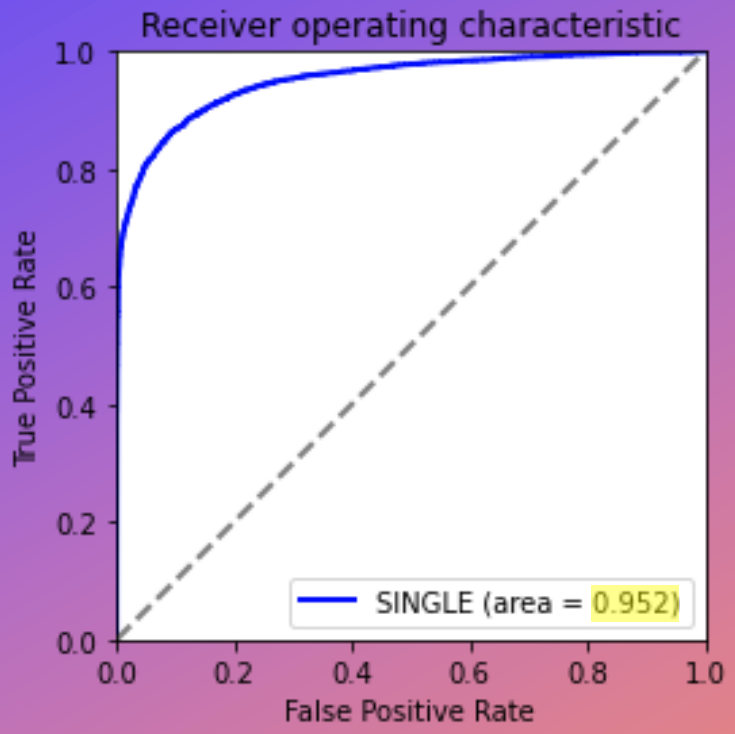

Training performed over 50 epochs (composed by 128 batches of 32 samples) to allow comparison. The data is artificially generated so it is not necessary to use a validation set as overfitting cannot occur.

# Artificial Signals Experiment

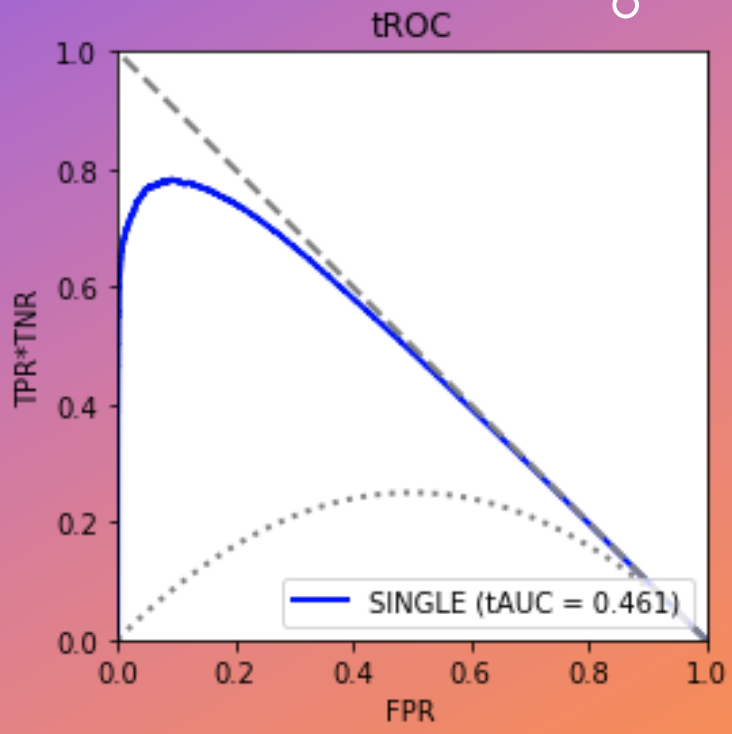
Single Input Results

To evaluate the results obtained from the single input model, a test set consisting of *10 000 single images* was generated




It should be noted that in the TAFC problem, as expected for large test sets, the value of the single input model's *accuracy* is the same as the *AUC* for the single images classification



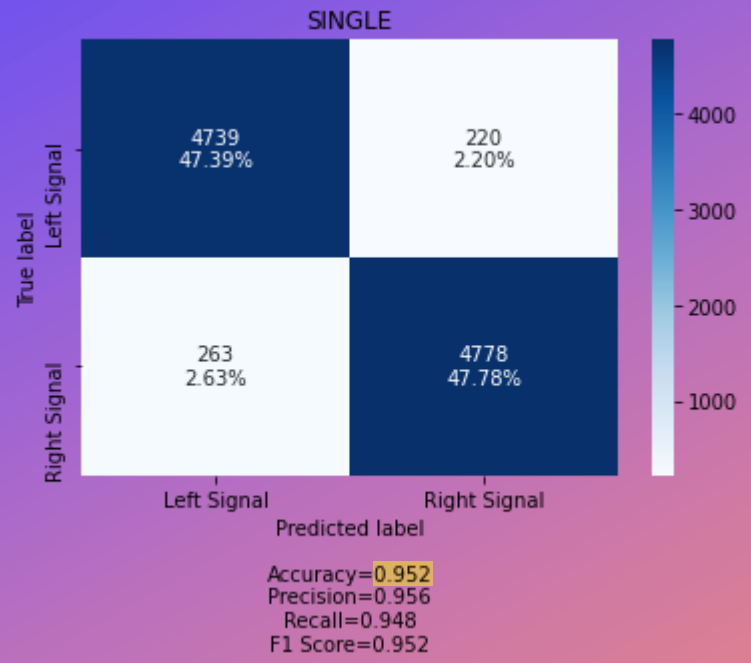

As described in Marafini's thesis, it is possible to obtain the *tROC* curve by replacing the TPR in the ordinates with TPR * TNR

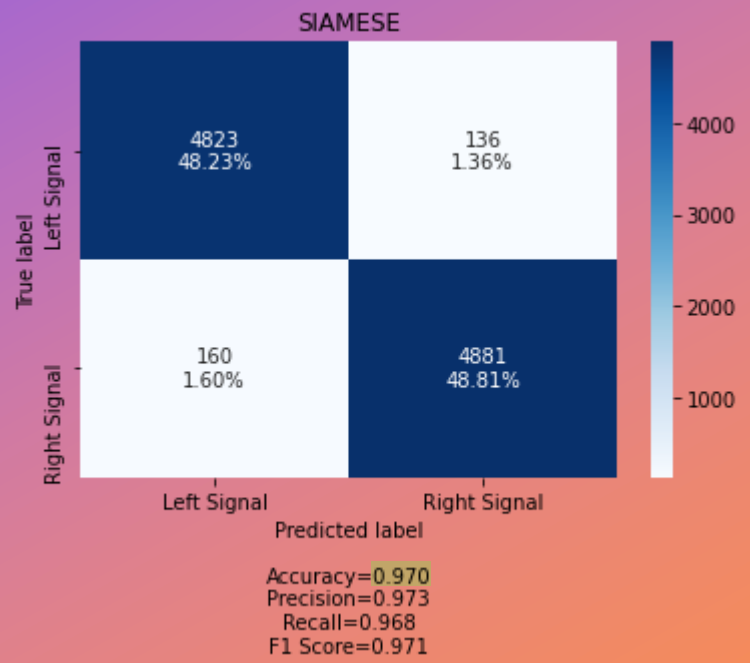
# Artificial Signals Experiment

TAFC Results

Also for the TAFC problem, *10 000 pairs* of images were generated to evaluate the results obtained from the *Siamese model* compared with the results obtained from the *single input* model

To do this, the images in pairs are supplied one at a time to the single input classifier and then the two scores obtained are compared, the image with the highest score is the one classified as containing the signal



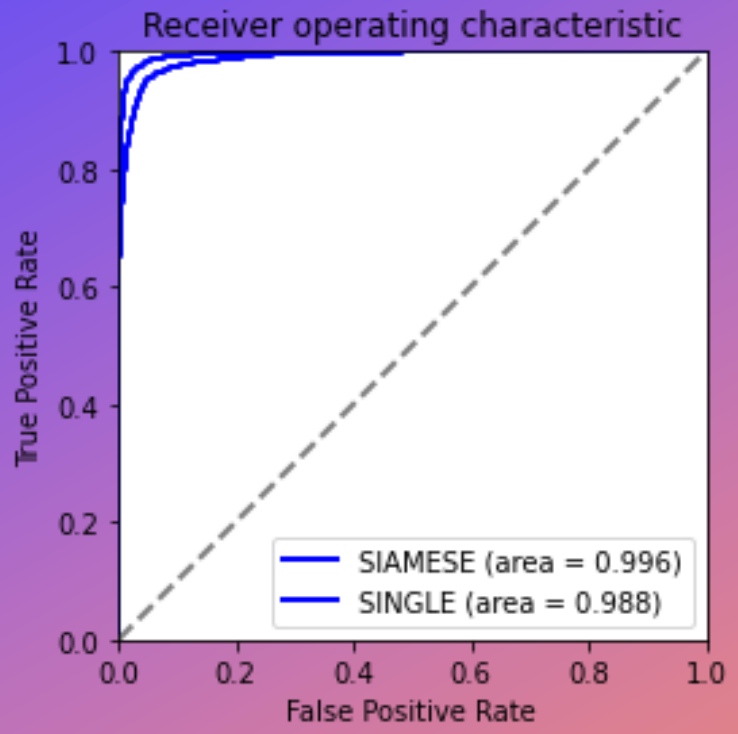

By comparing the accuracy value of the *Siamese model* with that of the *single input model*, it can be seen that this has a better accuracy. This is probably due to the fact that the Siamese model is able to directly compare images as opposed to the single model that sees them individually.

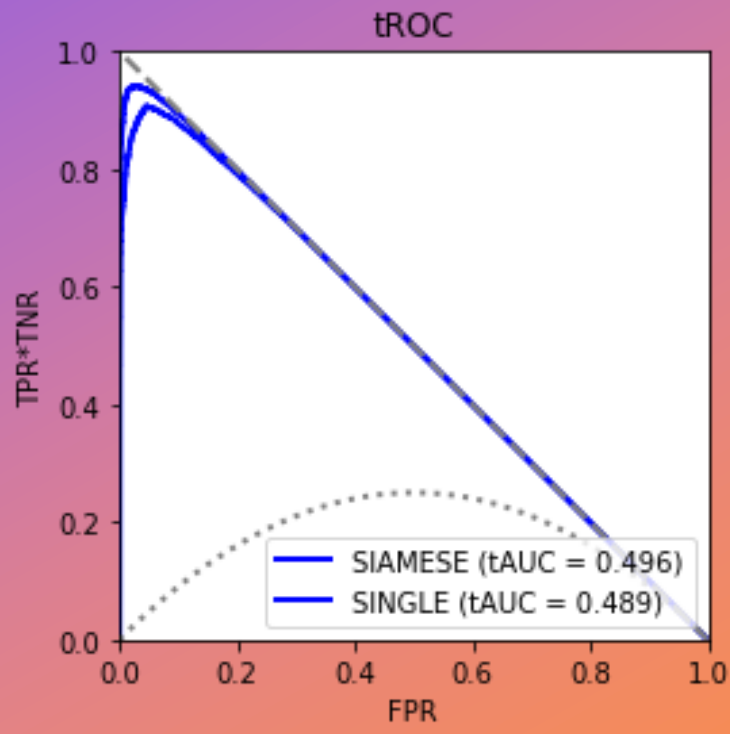
# Artificial Signals Experiment

To draw the ROC and tROC curves also for the single input model, the *confidence* in the classification of a pair of images was interpreted with the following formula:

$$c = \begin{cases} \frac{1 + (r - l)}{2} & if\ r \geq l \\ \frac{1 - (l - r)}{2} & if\ r < l \end{cases}$$

Given that the optimal confidence is 1 if $r \geq l$ and 0 otherwise



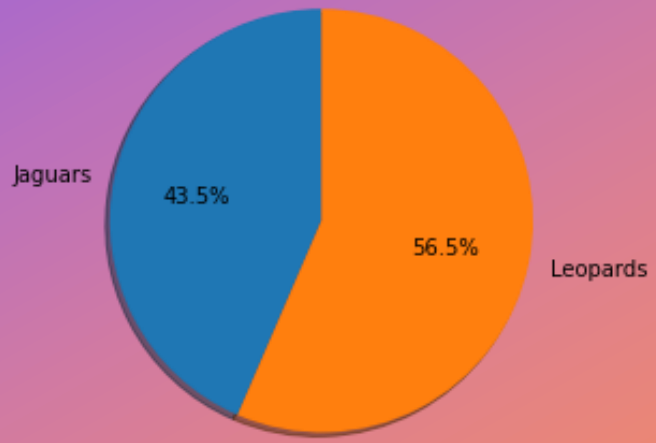




It can be seen that both for the ROC curve and for the tROC curve the value of AUC and tAUC is better for the *Siamese model* than for the *single input* model, confirming the hypotheses.

# Real-World Dataset Experiment

## Dataset

The results obtained from the previous experiment with artificial data were validated using a *real-world dataset* consisting of photos of *leopards* and *jaguars*.




The dataset is composed of 1418 images divided for 60% in training set, for 20% in validation set and for the remaining 20% in test set, using a stratified split.

## Models

In this experiment 4 different models were compared: two *single input networks* and two *Siamese networks*, in which one is defined *from scratch* and the other uses a pretrained *ResNet-50*

# Real-World Dataset Experiment

Training

Two different trainings were carried out, one for the *single image classification* and one for the *TAFC* one

Strong *augmentation* was applied to the training batches to compensate for the small size of the dataset

Batches for the TAFC are made by random permutations of training set images, while validation set is decided initially and remains constant

*Early stopping* with 5 epochs patience was used for all models

Leopards

Jaguars

# Real-World Dataset Experiment

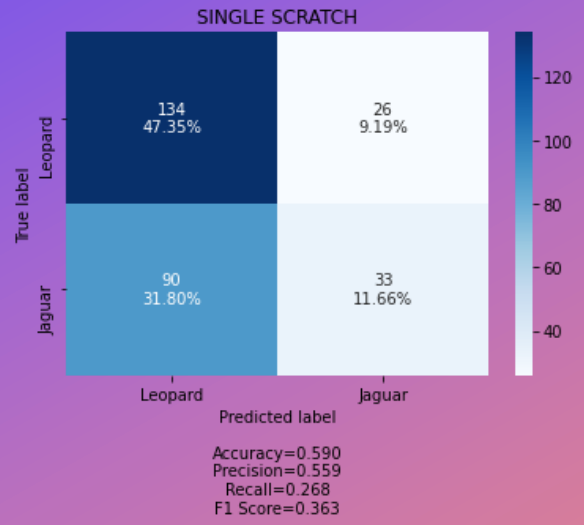
## Single Input Results

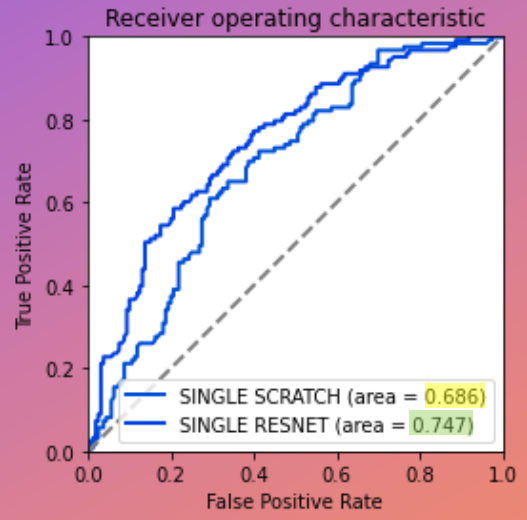
To the side are reported the results obtained from the *single input models* obtained using the test set

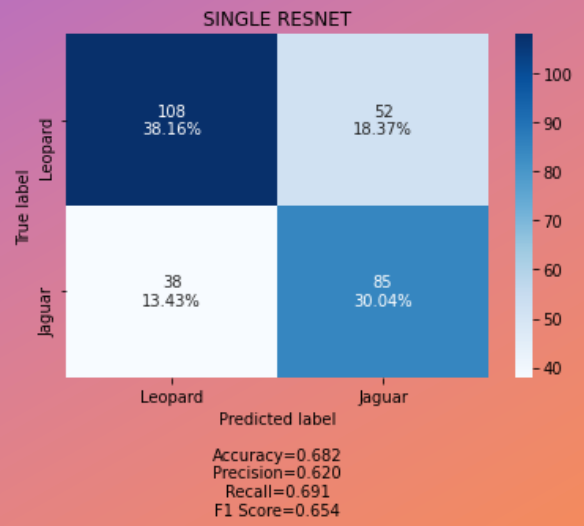
It can be seen that the *from scratch model* is not able to achieve optimal accuracy, since the problem is difficult and the data is small

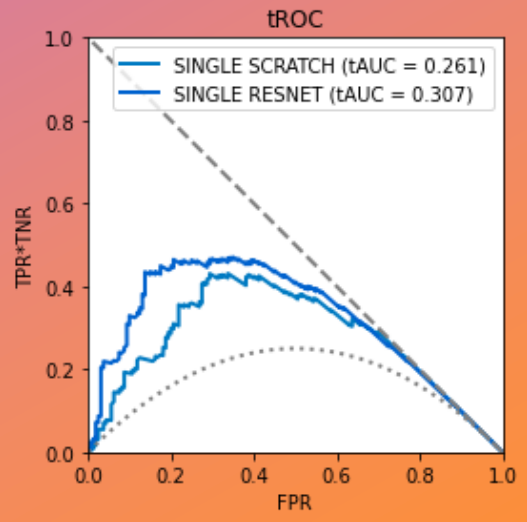
The *pretrained model*, on the other hand, manages to achieve better results, even if not optimal

It is interesting to note that the *AUC* values of the two models are similar to their *accuracy* in the TAFC problem in the next slide



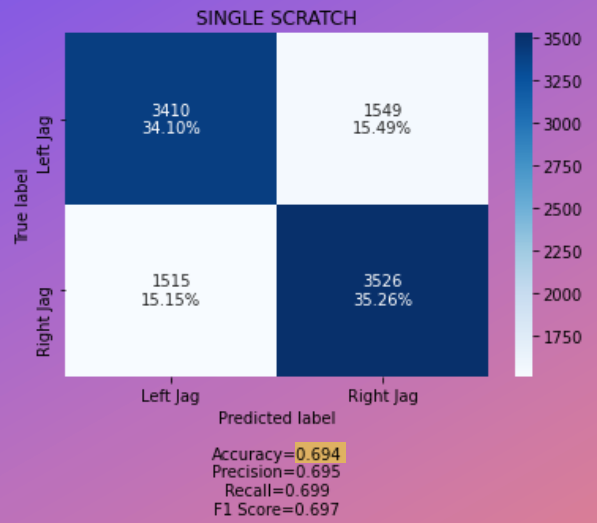
# Real-World Dataset Experiment

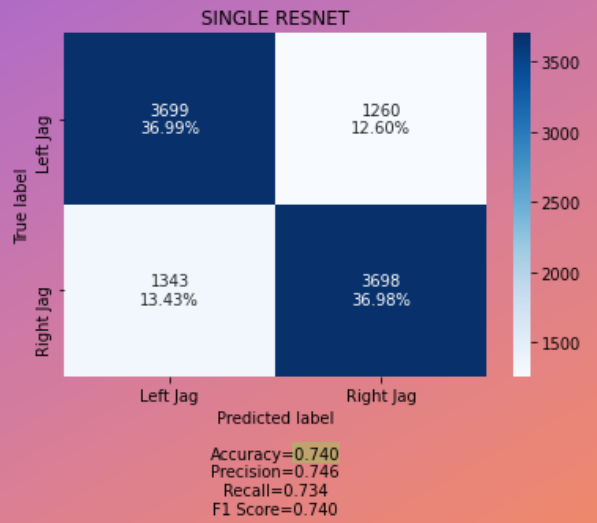
## TAFC Results

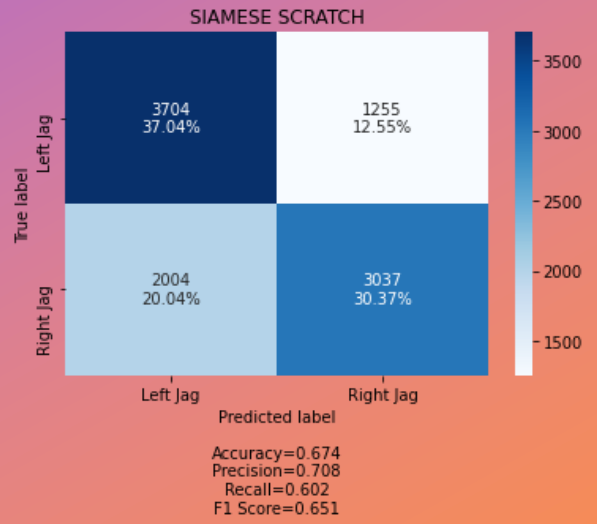
The test set for the TAFC problem consists of *10 000 pairs* of images (note that about a 40 000 different test set images permutations are possible)

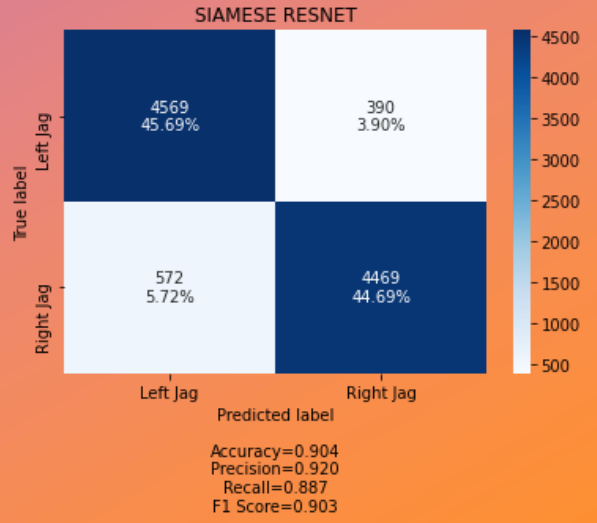
It is interesting to note that the *Siamese scratch model* performs worse than its single input counterpart

As expected, the *pretrained Siamese* version has the best results, confirming, together with the experiment with synthetic data, that the Siamese model is better than a single model for the TAFC problem

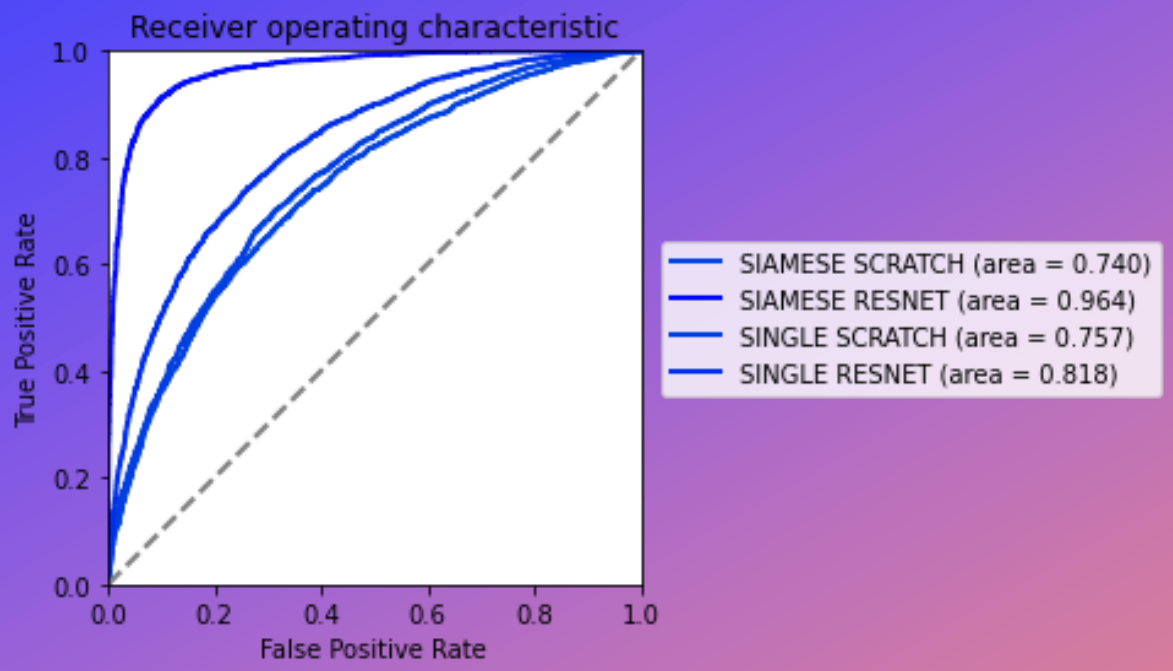


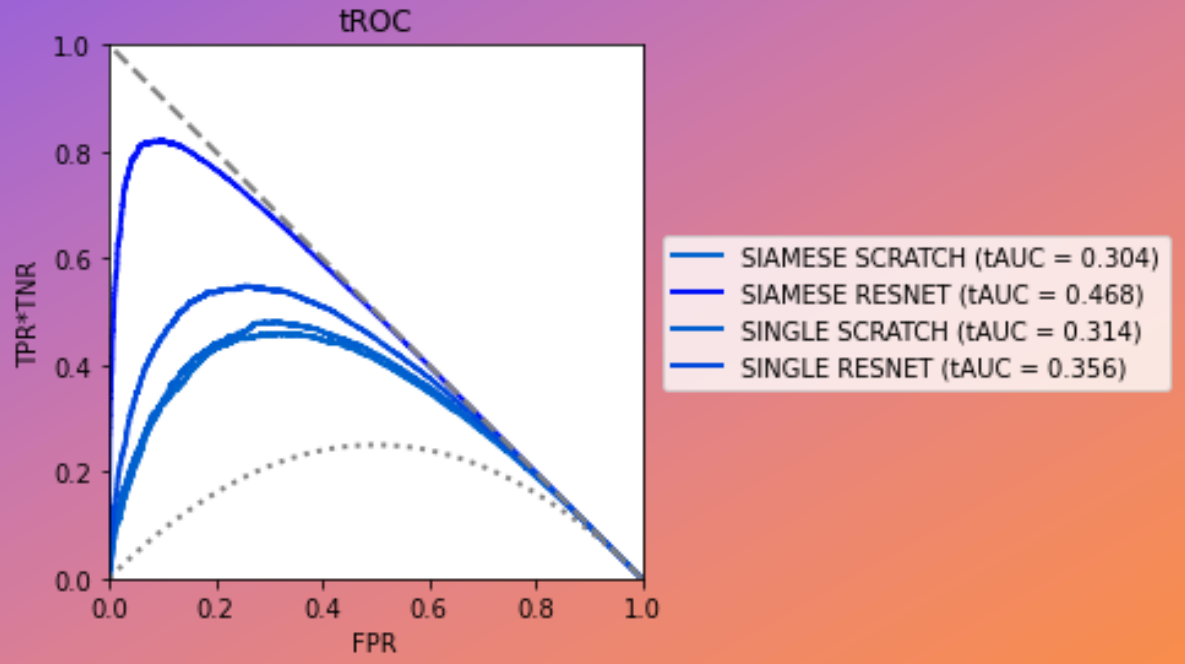



# Real-World Dataset Experiment




The ROC and tROC curves confirm the considerations made previously. It is possible to notice how the *pretrained Siamese model* stands out in terms of effectiveness, as it can consider both inputs at the same time and reduce overfitting thanks to the many possible permutations of the images and to the weights already initialized.

# QUESTIONS?