# Duplicated Ads Solution

### The Business Problem

Detect duplicate classified ads by comparing the *images* from the classified ads.

### Exploring the Data Samples

I explored the data samples provided, i found the following observations.

* 110 training samples.
* All images with different dimensions. Some with 480 \* 270, others with more than 900 \* 600 dimensions. Which is not normalized.
* There are no different training images for each ad.

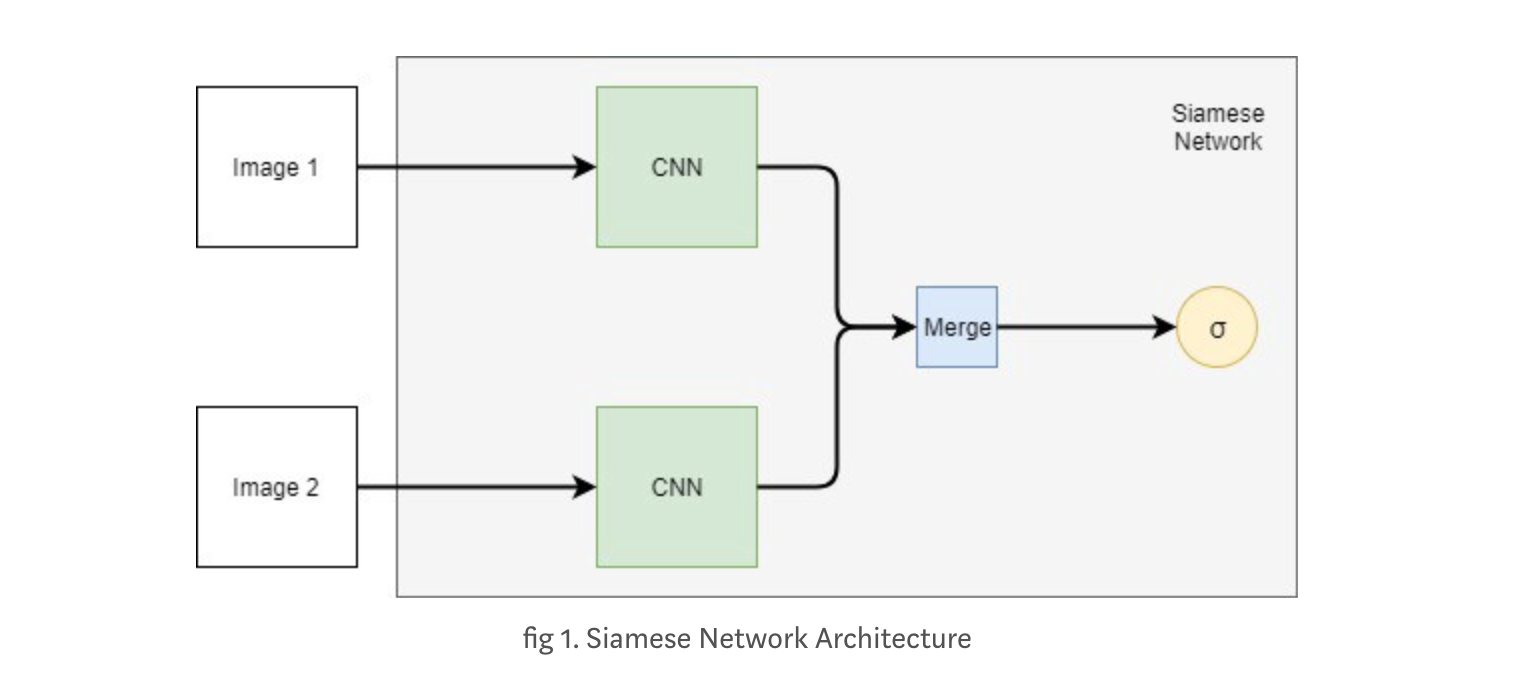
### The Approach

**Deep Learning Classification**

First approach comes to my mind is a classification approach, but for sure it needs many samples for each ad to be able to train a neural network and predict the class of the add for each image and compare the two classes together to predict the duplication.

**Deep Learning Siamese Network (One Shot learning)**

After researching, i found better approach to measure the similarity between an image and other images. Which is Siamese network.



Siamese Network is a Neural Network Architecture which compares two input images, and decides, whether those two images are same or not. Which the input is two images, vectorized into CNN, to extract the features with the same base CNN network. Using same weights. Then find the similariy distance using the Euclidean distance function.

A full description at the Siamese network paper

https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

### The Code Implementation

All code exists in the following jira url

https://github.com/samehamin/schibsted-challenge