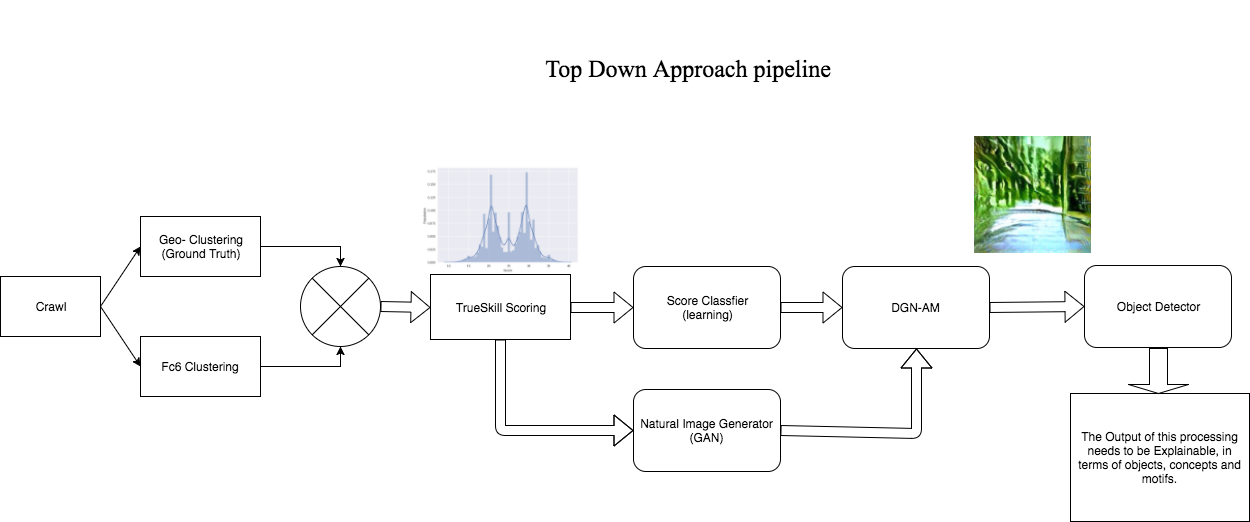
# Explainability Pipeline: Top Down

This is the pipeline for the top down approach.



# Block-wise Explaination:

## Crawl:

* The data needs to be crawled from google streetview. Currently the data is annotated from the Placepulse project, but in the future, this step can include a crowd sourced annotation process or a game that feeds in annotated data.

## Geo - Clustering:

* From discussions, and the output of exploratory experiments, we have found out that the concept of beauty can vary for images coming from diverse locations. This implies that a beautiful rural image is not made of the same visual motifs as a beautiful urban image. For this reason, there needs to be some data pruning put in place using GIS data which enables us to filter out images using the lattitudes and longitudes into images from the city and from outside the city. This makes our ground truth for seggregation of images.

## Fc7 clustering:

* There are neural nets in the art, which can discriminate between scenes in an image. PlacesCNN is one such neural net, which can out an image into one of the 205 places categories.
* Here we use PlaceCNN to extract higher level features viz. the Fc7 features, which can be then used to cluster images accordingly.
* These clusters belong to a 200 dimensional abstract space, which theoretically should separate images into clusters of visual similarity.

## Compounding Block:

* At this point, we use the ground truth labelling of rural and urban images from the Geo-clustering block and augment it using the visual similarity features found in the Fc7 layer.
* This is done, because in the due course of geo-clustering, a lot of data is going to be lost because of sparsity of samples across the cities.

## TrueSkill:

* Using the trueskill algorithm we score each image in terms of a particular aesthetic metric, using either pairwise comparisons as proposed in the Placepulse paper, or by some other ranking method.

## Scoring Classifier:

* We train a Deeplearning classifier, that can classifiy images from each cluster, into a predetermined number of bins representing their trueskill scores. The current system is classifying images into two bins for high and low trueskill score.
* The idea is to train multiple such classifiers, each for a particular class of images e.g. rural, urban or commertial, residential.

## Natural Image generator:

* For each class, we need a Generator network (GAN) which can generate an approximate natural looking image drawn from distribution of a particular class of images. This is important for the next step of the pipeline.

## DGN-AM:

* We plug in the newly trained generator and classifier networks, into the Activation maximization network so as to generate an ideal image that maximizes the activation for a particular neuron.
* In our case this can be Beauty for rural class of images, or depression for urban class of images etc.
* The output of this step would be an image, which is very close to a natural streetview image, but has all the right motifs that maximize the sense of beauty, depression, safety, happiness or any other abstract affective dimension.
* The important part is this image needs to be very close to a natural image.

## Explainability of generated images:

* The last step deals with analysis of these synthetic naturalish images, which would allow us to detect objects or concepts that can be explained in terms of perception of aesthetics or the abstract affective dimensions.