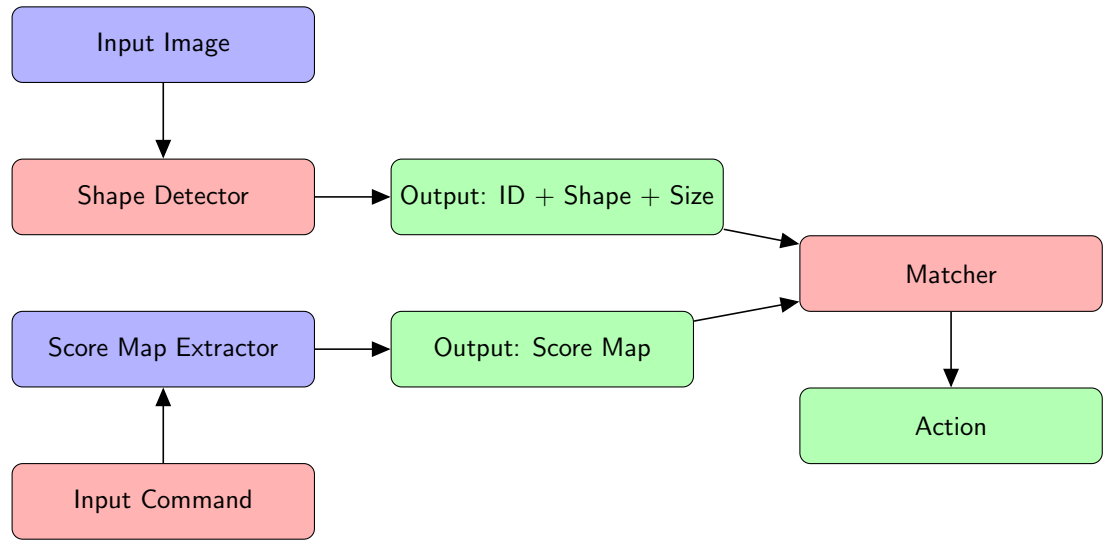


# Pipeline

Irdi Balla

09/02/2018



## 1 Architecture

The Architecture is composed of several parts that pass together to achieve the final Result or as i call it 'Action'.

### 1.1 Environemnt

The environment is a 3D world with different objects that is generated randomly on every test. The environment randomly selects 20 times from the set  $\{Cube, Sphere, Pyramid, Cylinder\}$  and for each objects it assigns a random size and color. These Objects are placed in the scene in such a way that they dont overlap with each other. The camera represents the robot which looks at this scene from a 45° angle to mimic a situation where the robot is sitting on a table. These objects have to be rearranged to follow the commands.

## 1.2 Shape Detector

The Shape Detector is a model trained to recognize the basic shapes from the set  $\{, Sphere, Pyramid, Cylinder\}$  in an image taken from the environment. This module outputs an ID, a color and a size for the all the shapes that it detects. A bounding region is considered to be a shape if it surpasses a threshold.

## 1.3 Score Map Extractor

The Score Map Extractor is a trained model that takes a command as an input. From the input it extracts the action, the subject on which the action has to happen and the description of the subject. Its output is a score map of the geometric cues and semantic segmentation of the words as described by [CITE].

## 1.4 Matcher

The matcher has to figure out the best combination of the objects that matches the command. It takes the output of the Shape Detector and Score Map Extractor as its input. The Matcher tries to find the best combination of basic operations [translate, rotate, scale] for every object from the Shape Detector that matches best the Result from the Score Map Extractor. The output from this Module is the 'Action' that is sent to the environment which then performs this action.