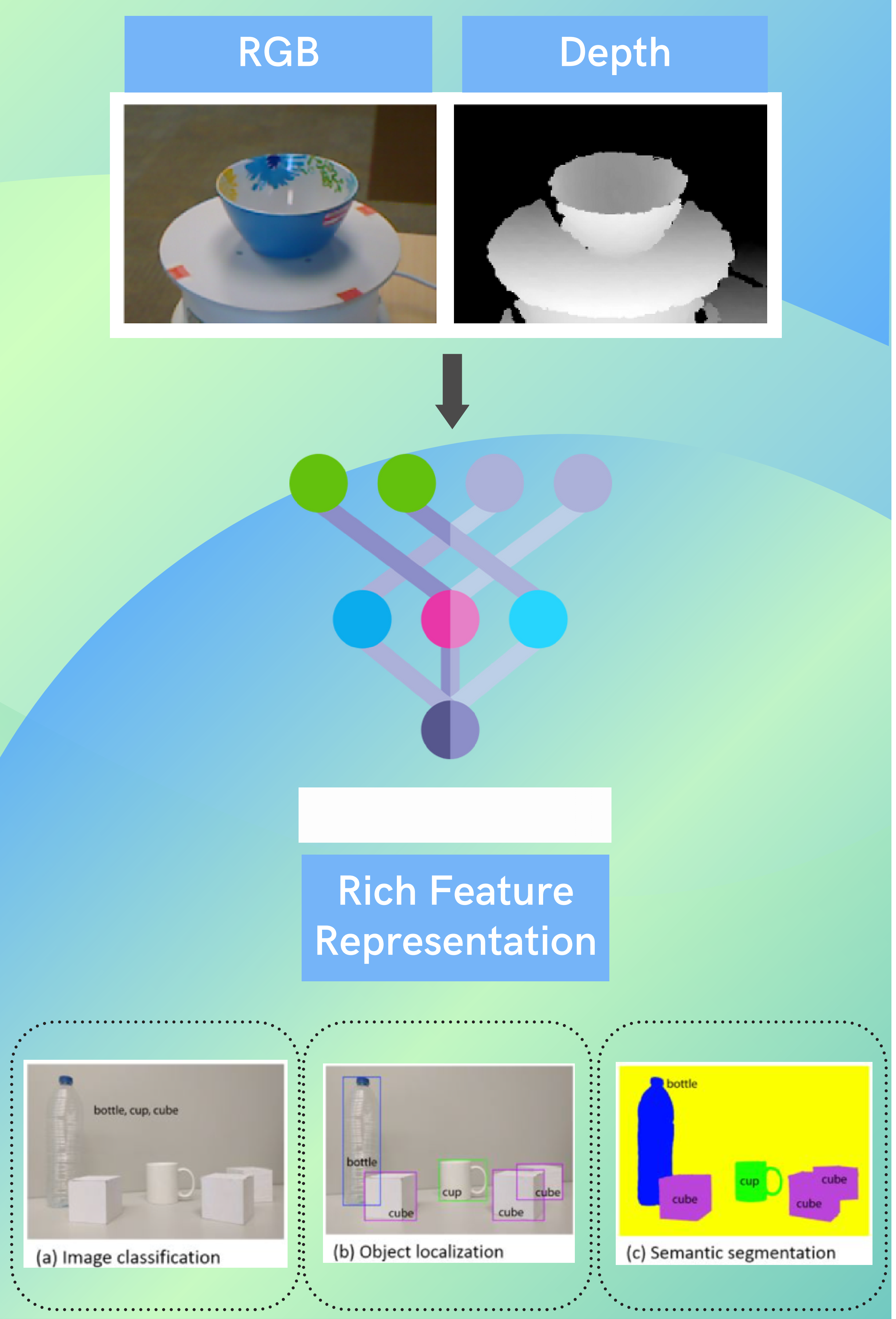



# Learning Rich Features from RGB-D Images for Object Detection and Segmentation

**Team number:** 4
**Team name:** Team number 105

# Agenda →

01

INTRODUCTION

Intro to the Research paper we will be presenting.

# Agenda →

01

INTRODUCTION

Intro to the Research paper we will be presenting.

02

OVERVIEW OF THE PROBLEM

What problem is this research paper attempting to solve?

Why this problem is interesting and what challenges it poses?

# Agenda →

01

INTRODUCTION

Intro to the Research paper we will be presenting.

02

OVERVIEW OF THE PROBLEM

What problem is this research paper attempting to solve?

Why this problem is interesting and what challenges it poses?

03

SCOPE OF THE PROJECT

Defining the exact scope of what we are attempting as a part of the project.

What are the specific goals and objectives of our work?

# Agenda →

01

## INTRODUCTION

Intro to the Research paper we will be presenting.

02

## OVERVIEW OF THE PROBLEM

What problem is this research paper attempting to solve?

Why this problem is interesting and what challenges it poses?

03

## SCOPE OF THE PROJECT

Defining the exact scope of what we are attempting as a part of the project.

What are the specific goals and objectives of our work?

04

## METHODOLOGY

How do we plan to accomplish our goals?

The dataset we will be using and how to access it.
The baseline model we will be using.
Experiments, ablations, or comparisons we plan to conduct.
Computing resources we will be using
and other details ...

# Agenda →

05

TIMELINE

Outlining a timeline for how we plan to complete the above tasks

Identification of any major milestones or deadlines

# Agenda →

05

TIMELINE

Outlining a timeline for how we plan to complete the above tasks

Identification of any major milestones or deadlines

06

INITIAL PROGRESS

Initial progress we have made on the code level

Explain any challenges we have faced and how we have overcome them

# Agenda →

05

TIMELINE

Outlining a timeline for how we plan to complete the above tasks

Identification of any major milestones or deadlines

06

INITIAL PROGRESS

Initial progress we have made on the code level

Explain any challenges we have faced and how we have overcome them

07

CONCLUSION

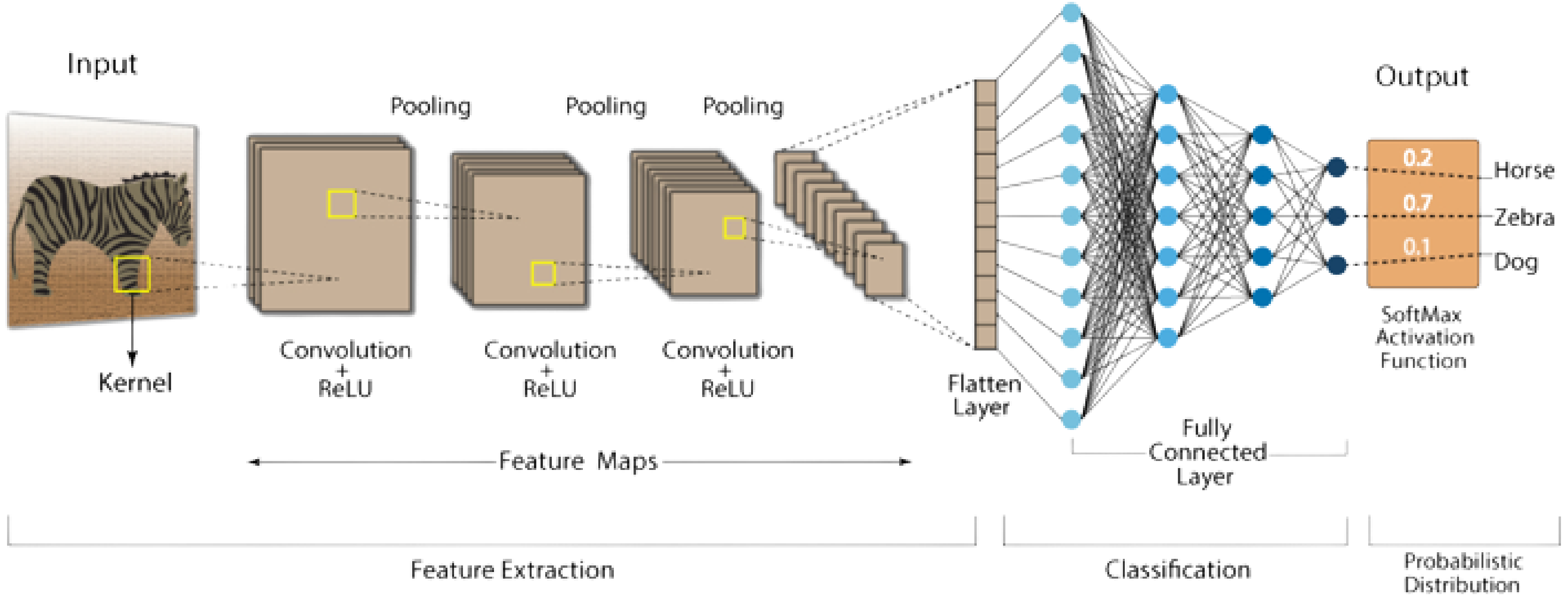
Summarize the key points of the presentation

Explain the potential impact of our work

Invite questions and discussion from the audience

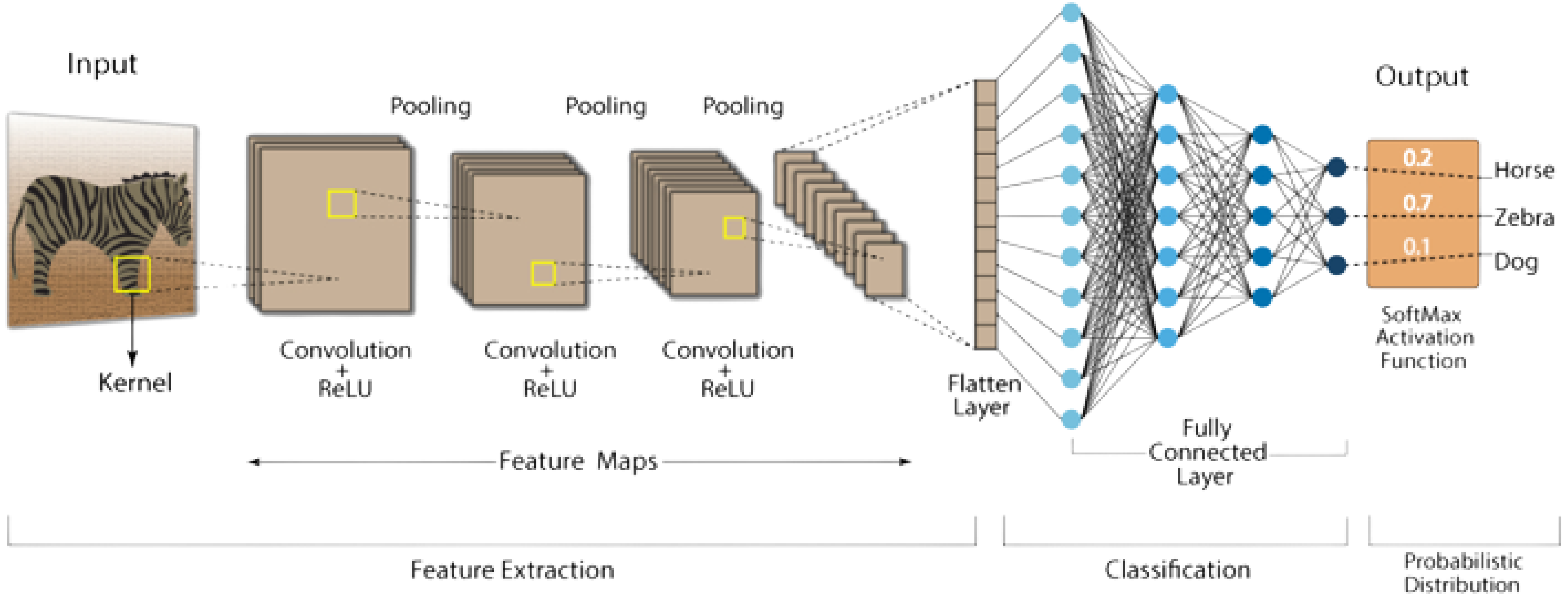
# Introduction to Problem Statement

## Strengthening Vision Perception via Convolutional Neural Networks

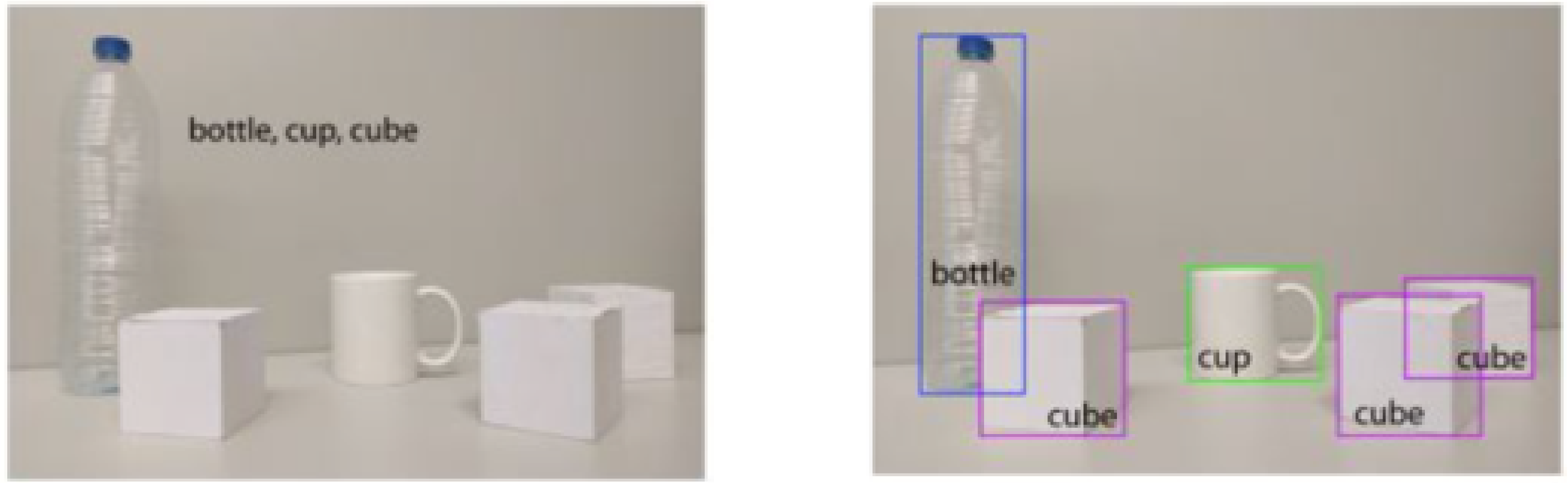


# Introduction to Problem Statement →

## Strengthening Vision Perception via Convolutional Neural Networks




## Learning Rich Feature Representations!

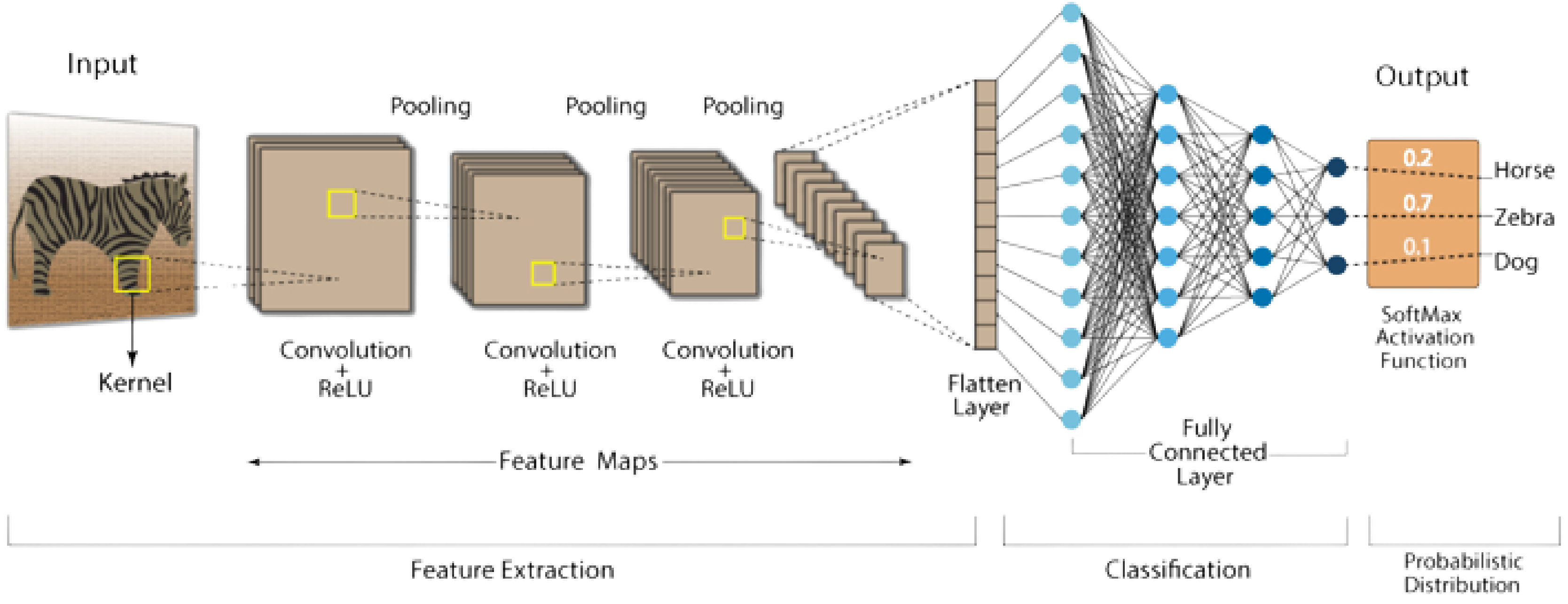



(a) Image classification

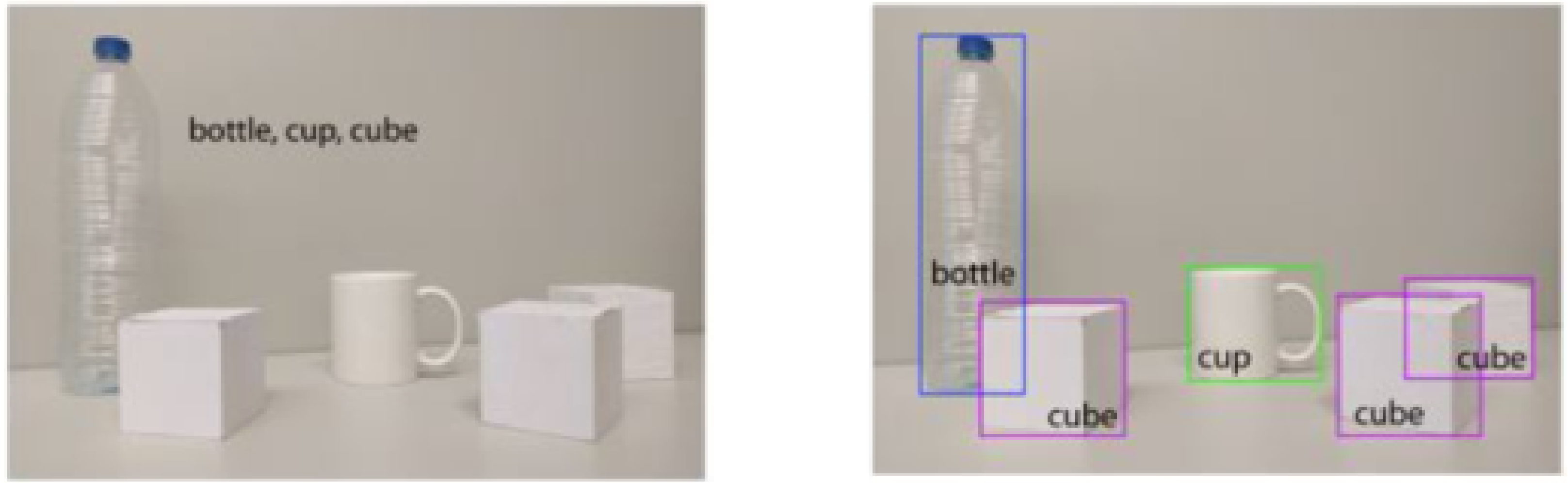
(b) Object localization

# Introduction to Problem Statement →

## Strengthening Vision Perception via Convolutional Neural Networks



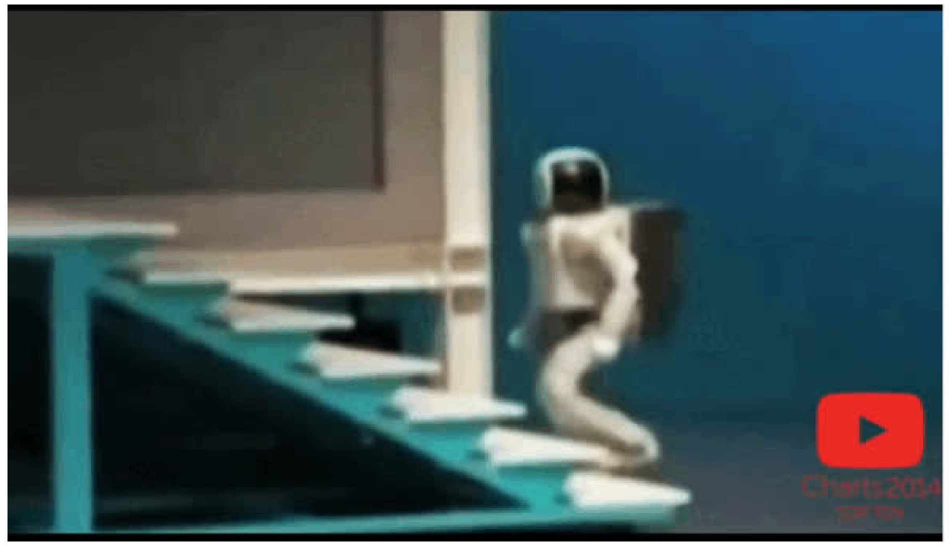

## Learning Rich Feature Representations!

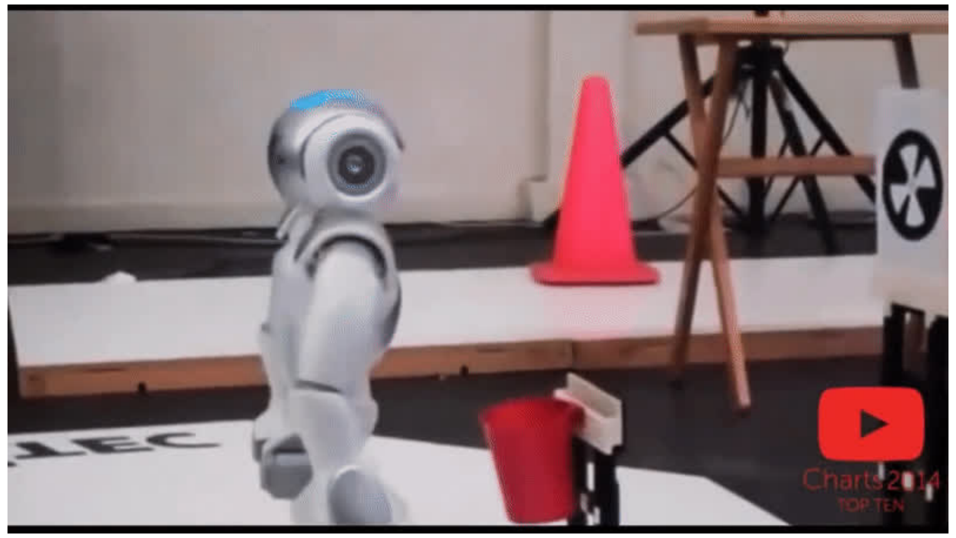



(a) Image classification

(b) Object localization

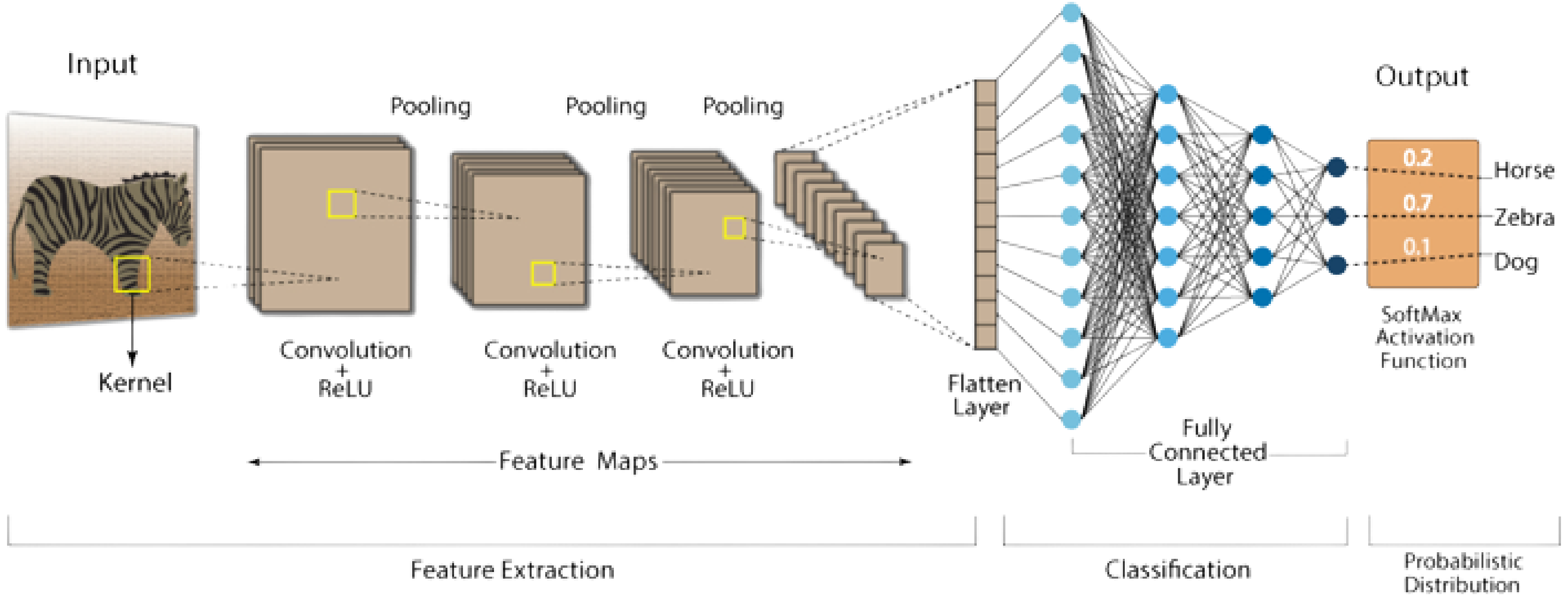
## Easy for human cognition - difficult for machines



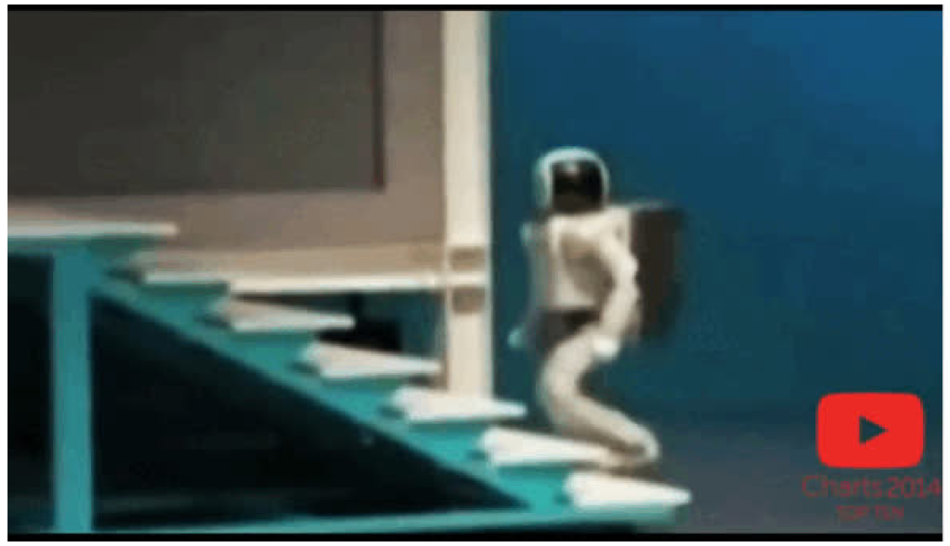


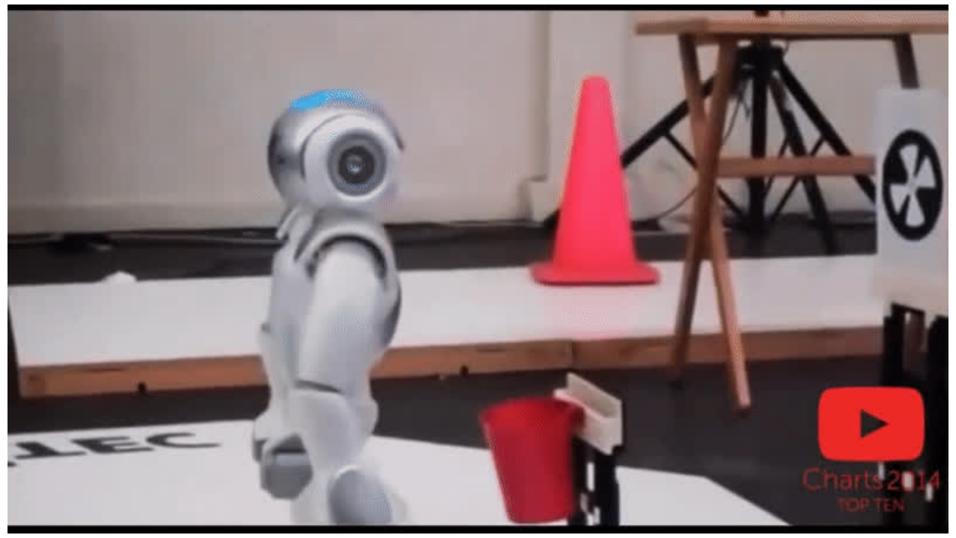
# Introduction to Problem Statement →

## Strengthening Vision Perception via Convolutional Neural Networks



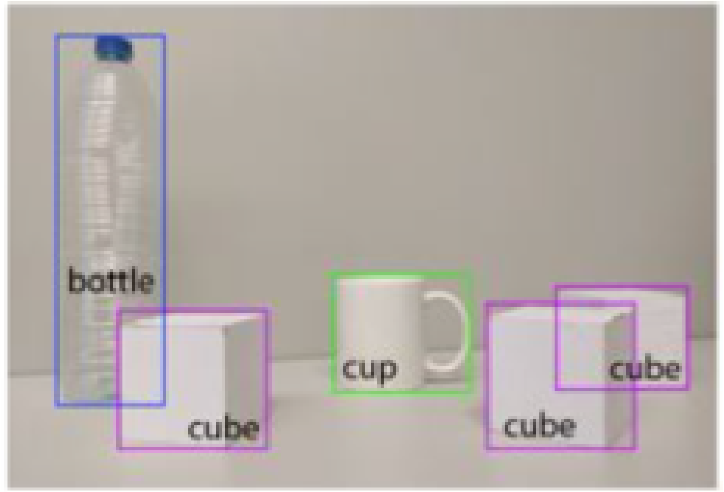

## Easy for human cognition - difficult for machines



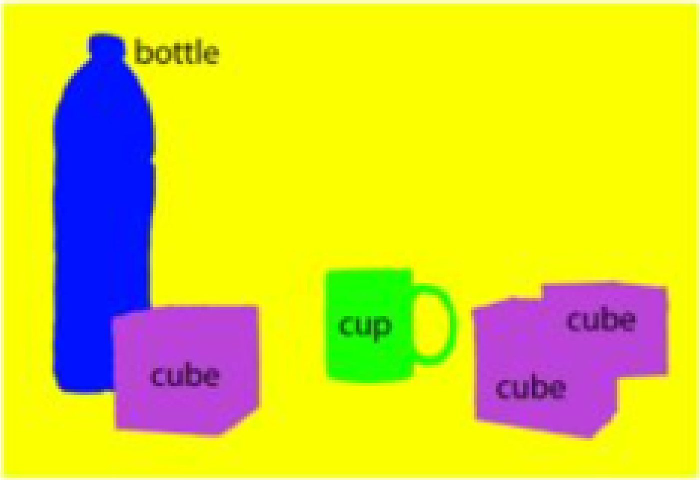


## Learning Rich Feature Representations!

(a) Image classification



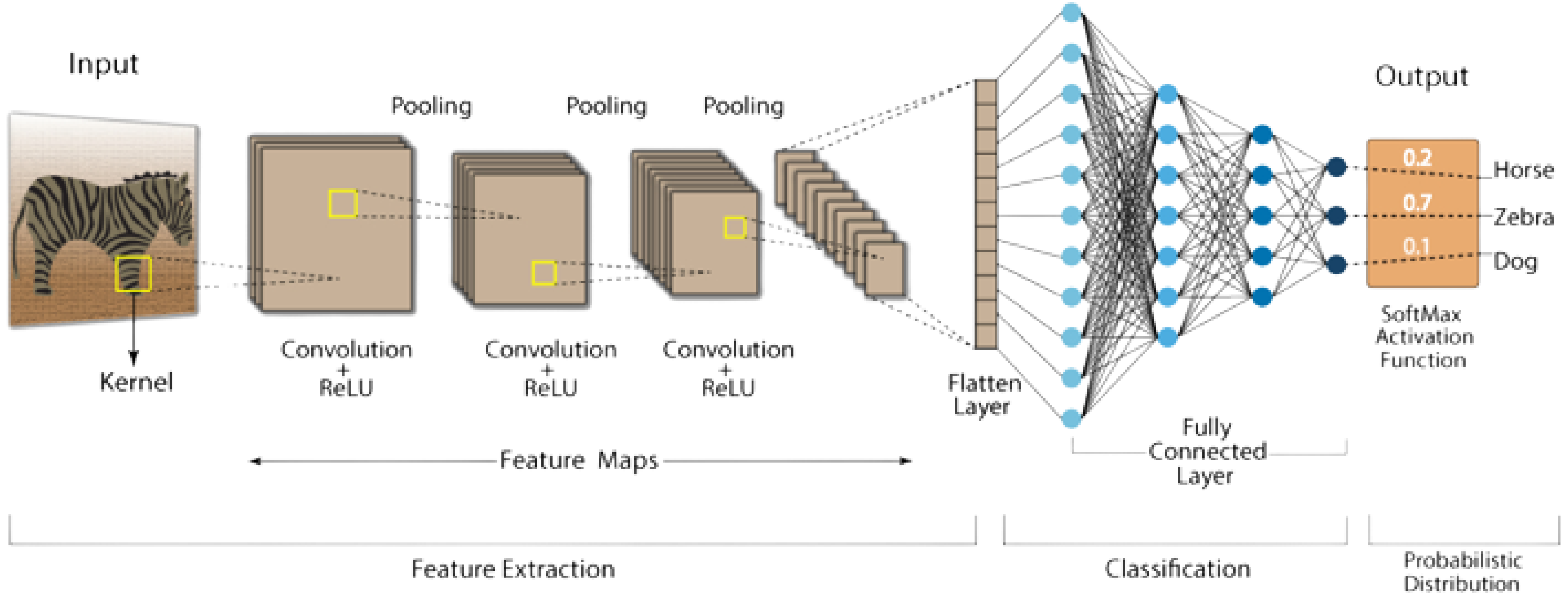
(b) Object localization



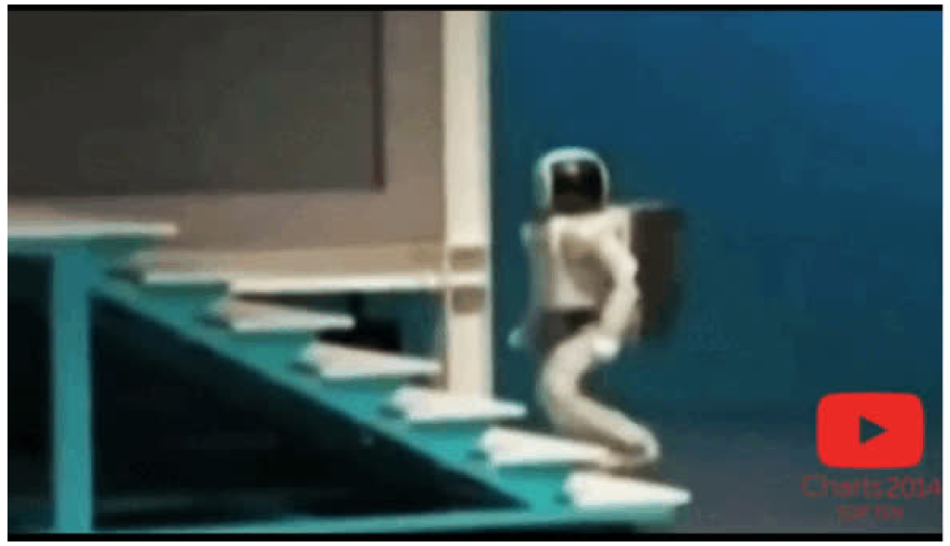
(c) Semantic segmentation

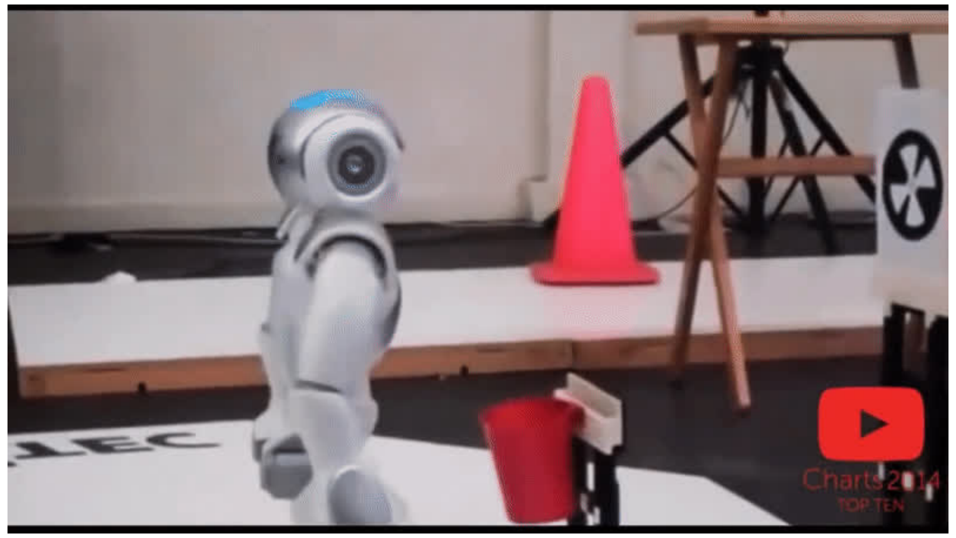
# Introduction to Problem Statement →

## Strengthening Vision Perception via Convolutional Neural Networks



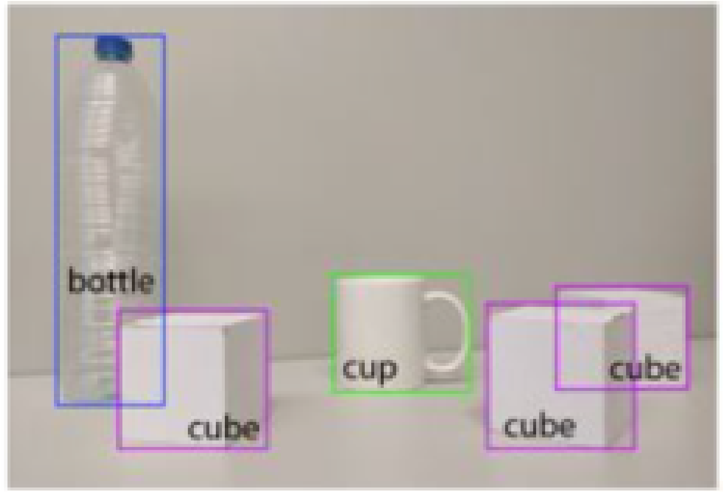

## Easy for human cognition - difficult for machines



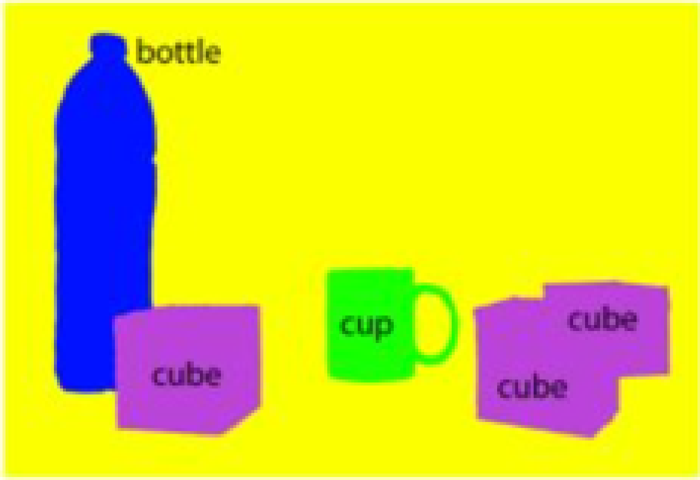


## Learning Rich Feature Representations!

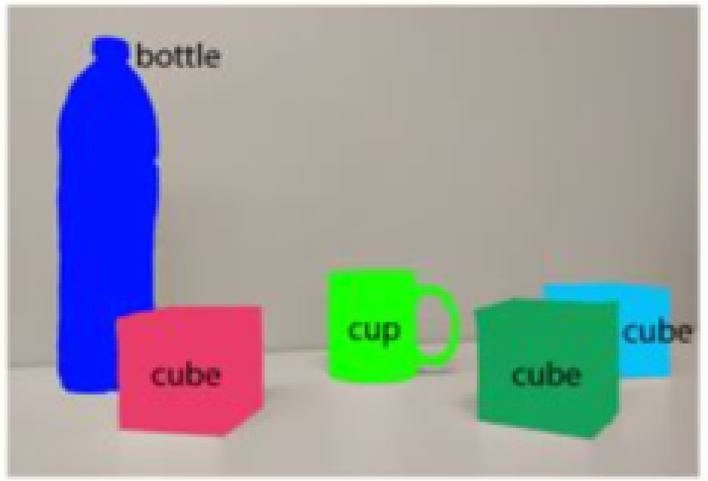

(a) Image classification




(b) Object localization




(c) Semantic segmentation



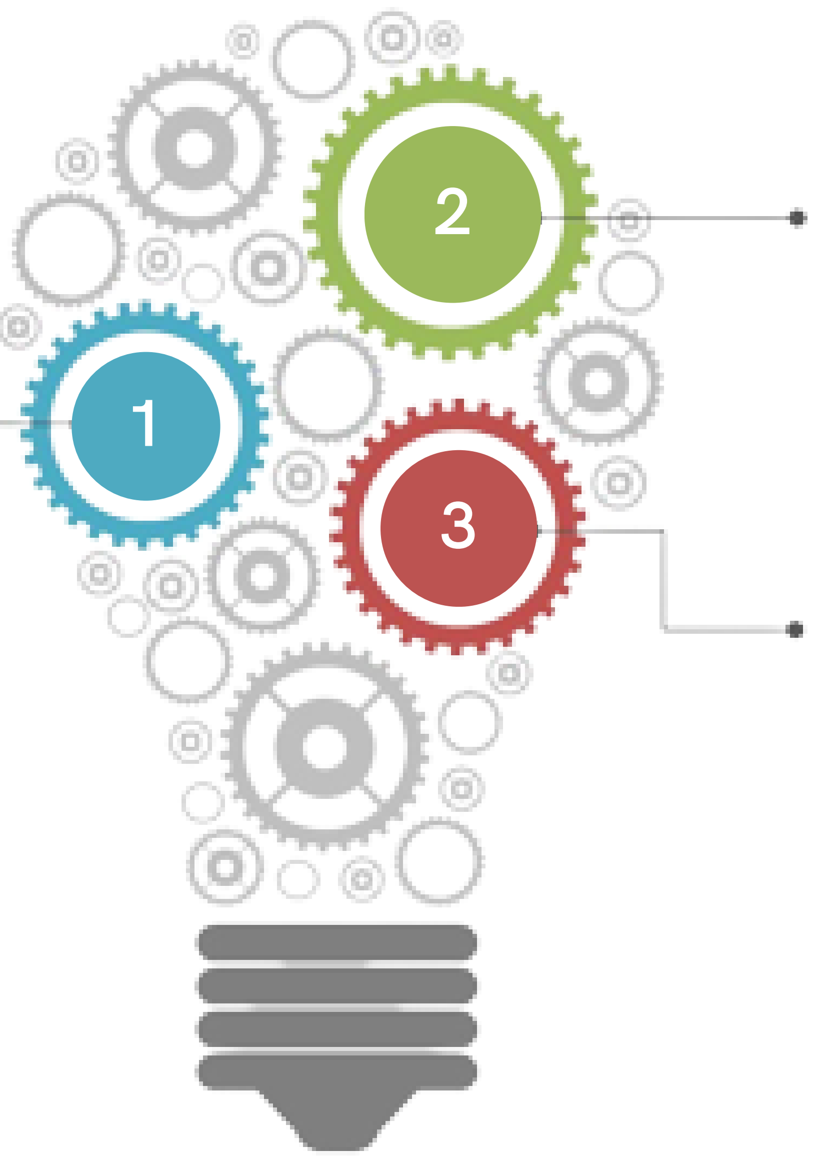

(d) Instance segmentation

# Overview: Research Idea →

- Develop an Integrated system for scene understanding from RGB-D images

# Overview: Research Idea →

- Develop an Integrated system for scene understanding from RGB-D images




### Object Detection

- <u>Geocentric embedding</u> instead of raw depth images to CNN = better feature representation (56% relative improvement).

# Overview: Research Idea →

- Develop an Integrated system for scene understanding from RGB-D images

- Decision forest approach that labels pixels of object instances as foreground or background by using the geocentric pose features.

Instance Segmentation


Object Detection

- Geocentric embedding instead of raw depth images to CNN = better feature representation (56% relative improvement).

# Overview: Research Idea →

- Develop an Integrated system for scene understanding from RGB-D images

- <u>Decision forest</u> approach that labels pixels of object instances as foreground or background by using the geocentric pose features.

Instance Segmentation


Object Detection

- <u>Geocentric embedding</u> instead of raw depth images to CNN = better feature representation (56% relative improvement).

Semantic Segmentation

- Output of object detectors in a <u>superpixel classification framework</u> for semantic scene segmentation (24% relative improvement).

# Overview: Proposed Pipeline →

Architecture of the integrated system

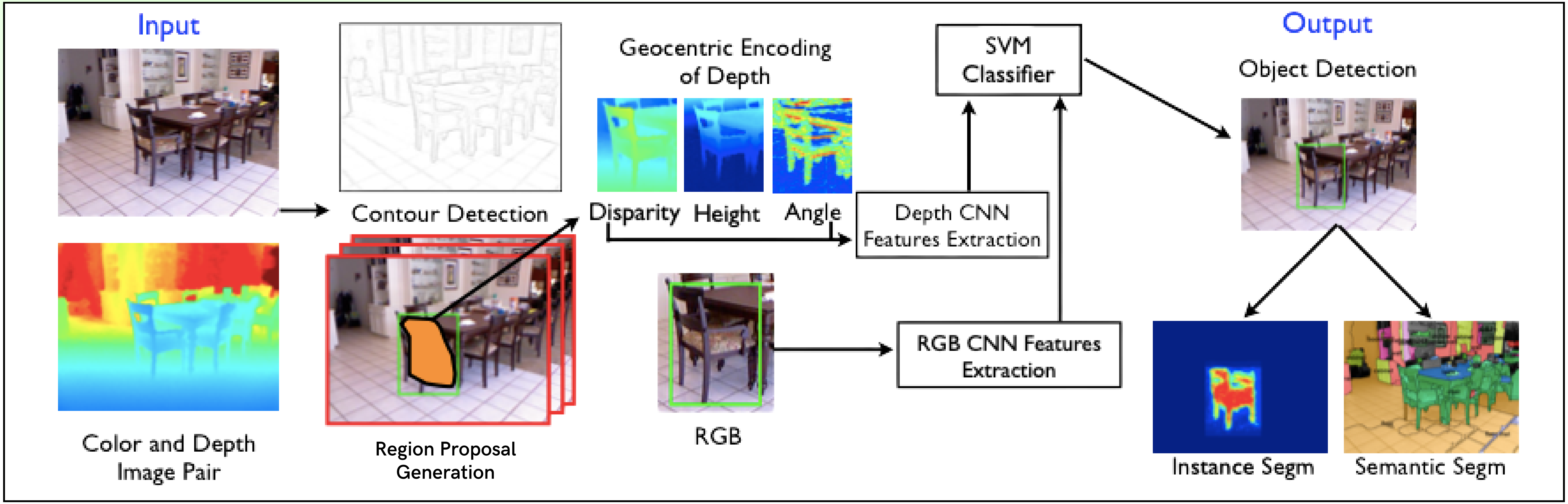
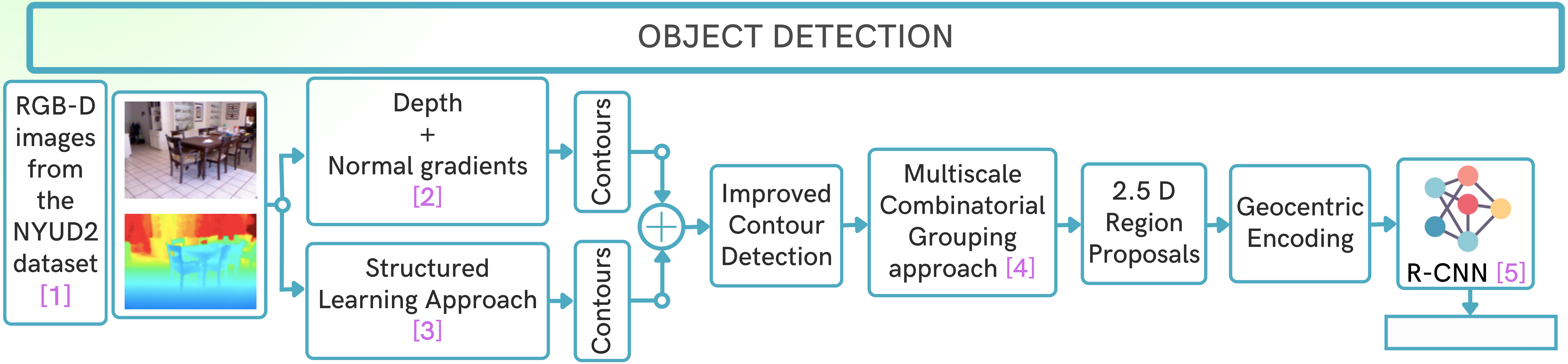
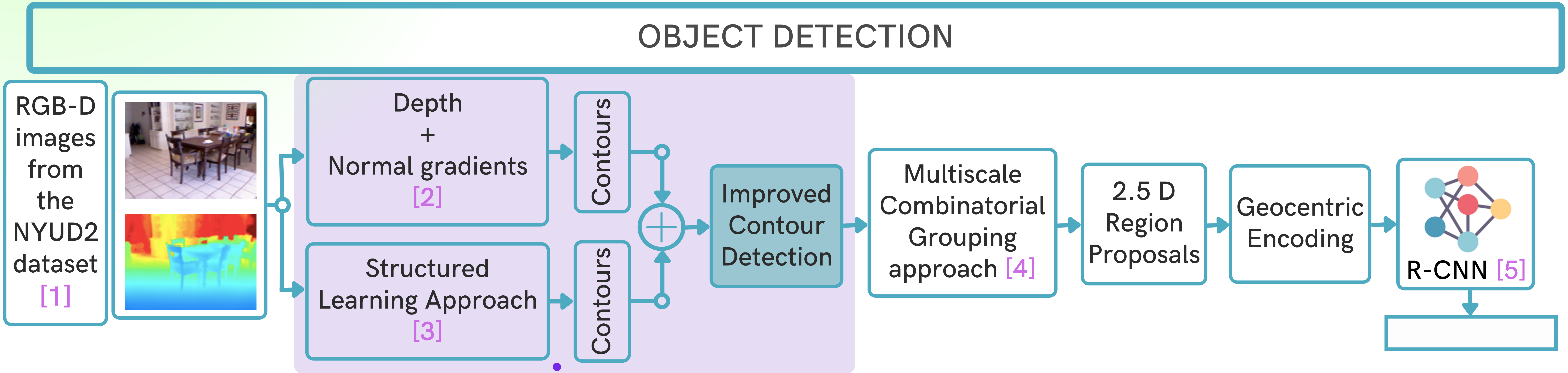



Fig 1. Overview: from an RGB and depth image pair, the system detects contours, generates 2.5D region proposals, classifies them into object categories, and then infers segmentation masks for instances of "thing"-like objects, as well as labels for pixels belonging to "stuff"-like categories.

# Architecture of the integrated system



# Architecture of the integrated system

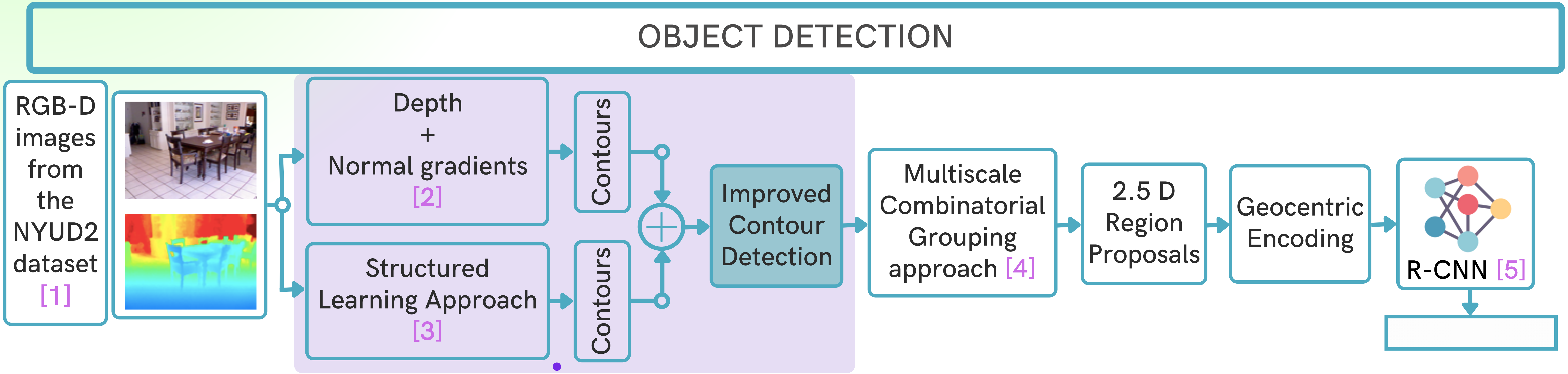



[2] uses gPb_ucm [6] and proposed local geometric gradients dubbed NG-, NG+ and DG to capture convex, concave normal gradients and depth gradients

[3] proposed an approach based on structured random forests to directly classify a pixel as being a contour pixel or not

[3] produces better localized contours that capture fine details, but tend to miss normal discontinuities that [2] easily finds

# Architecture of the integrated system

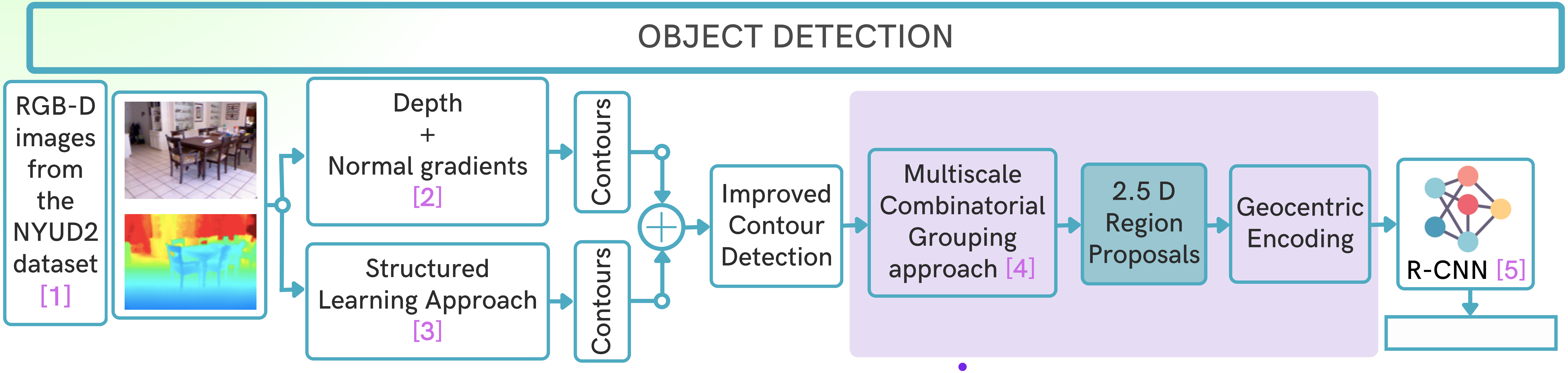



[2] uses gPb_ucm [6] and proposed local geometric gradients dubbed NG-, NG+ and DG to capture convex, concave normal gradients and depth gradients

[3] proposed an approach based on structured random forests to directly classify a pixel as being a contour pixel or not

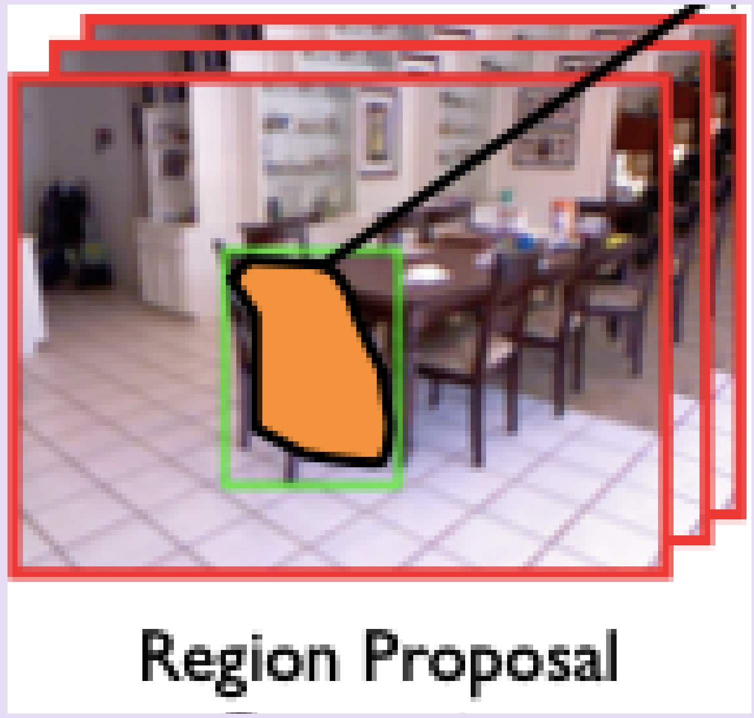
[3] produces better localized contours that capture fine details, but tend to miss normal discontinuities that [2] easily finds

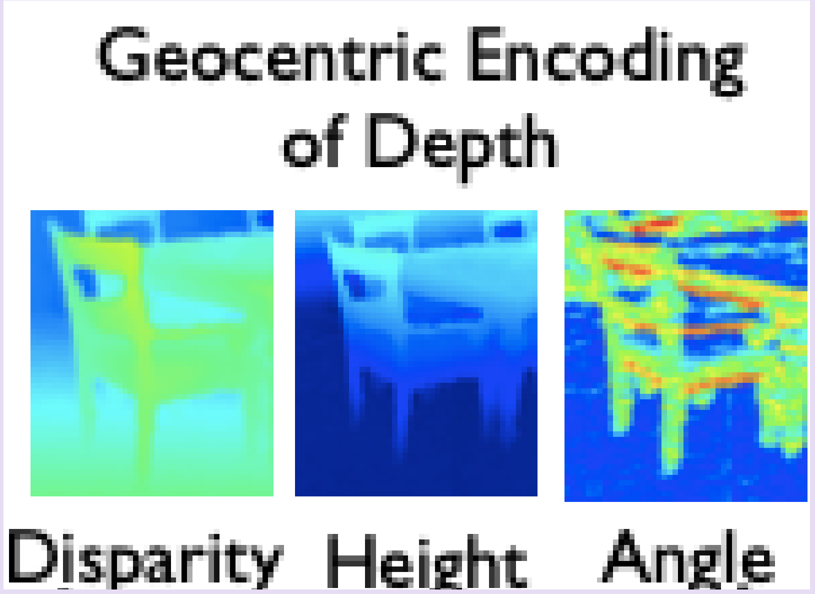
## Architecture of the integrated system




MCG is an approach used for object proposal generation.

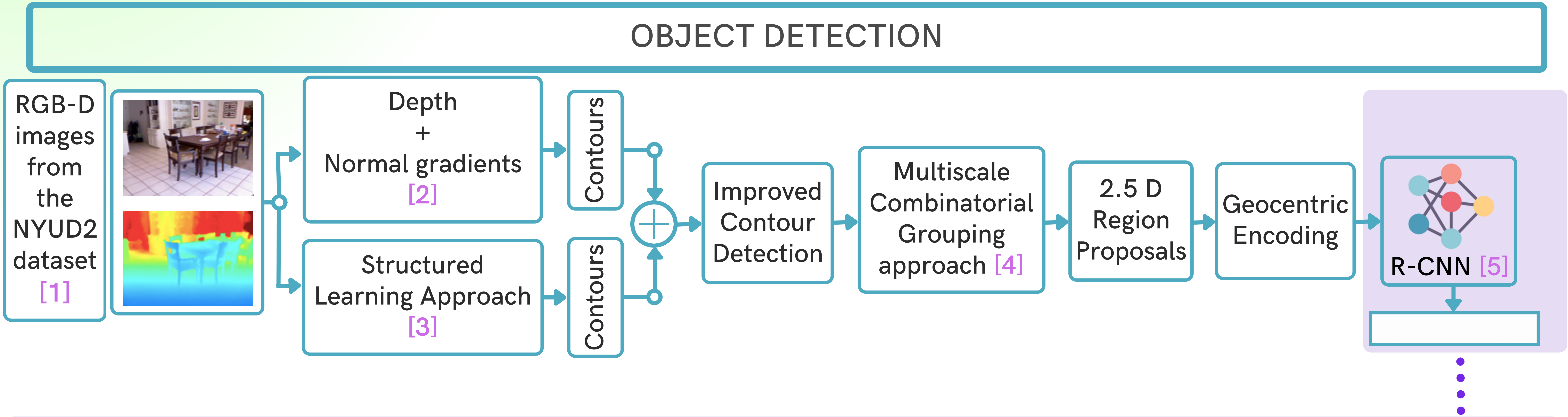
From improved contour, we obtain object proposals by generalizing MCG to RGB-D images



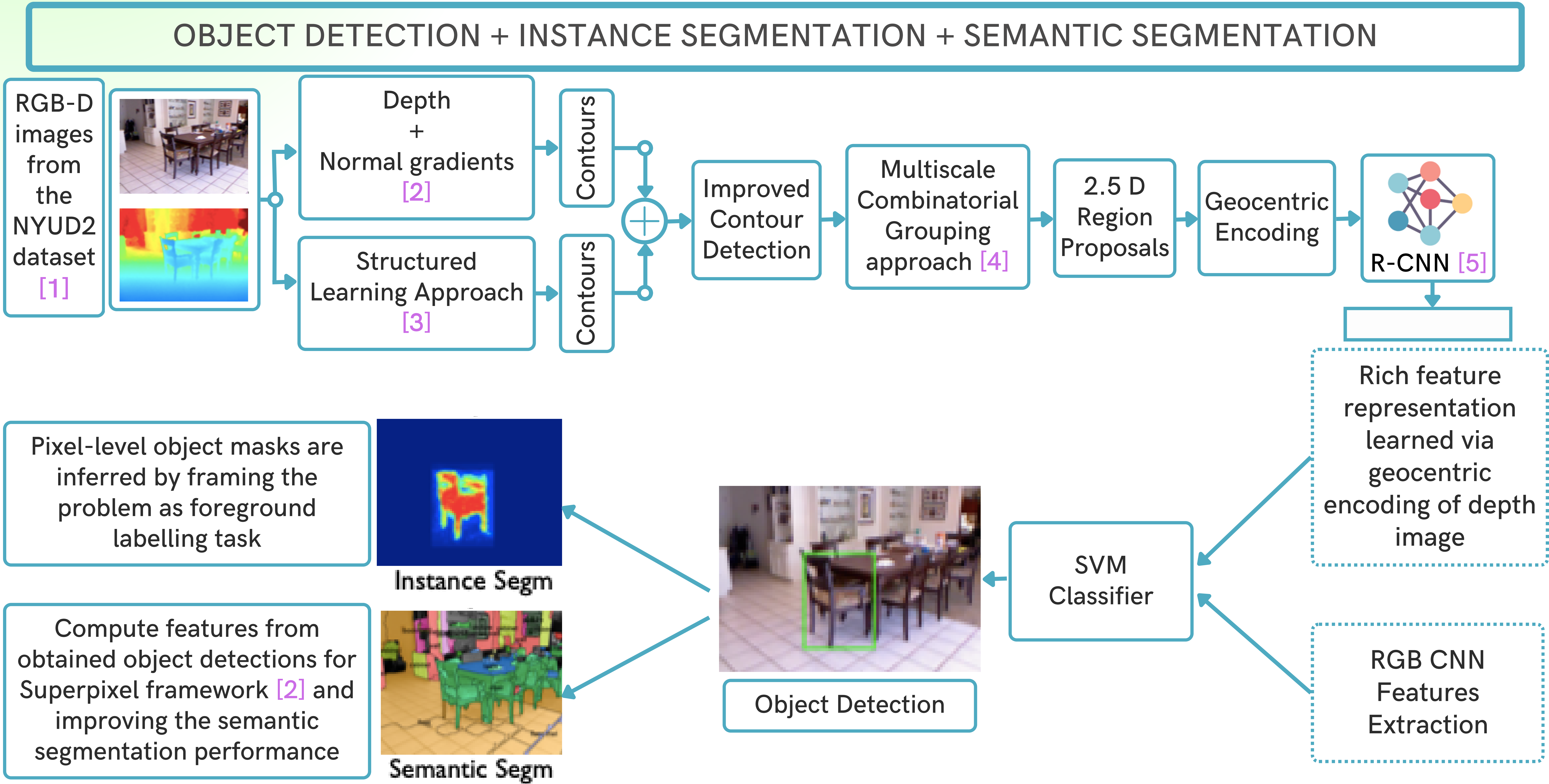

Geocentric encoding allows CNN to learn strong features than by using disparity alone

# Architecture of the integrated system




A large R-CNN pre-trained on RGB images, can be adapted to generate rich features from this representation of depth images.

# Architecture of the integrated system



## Architecture of the integrated system

# RESULTS

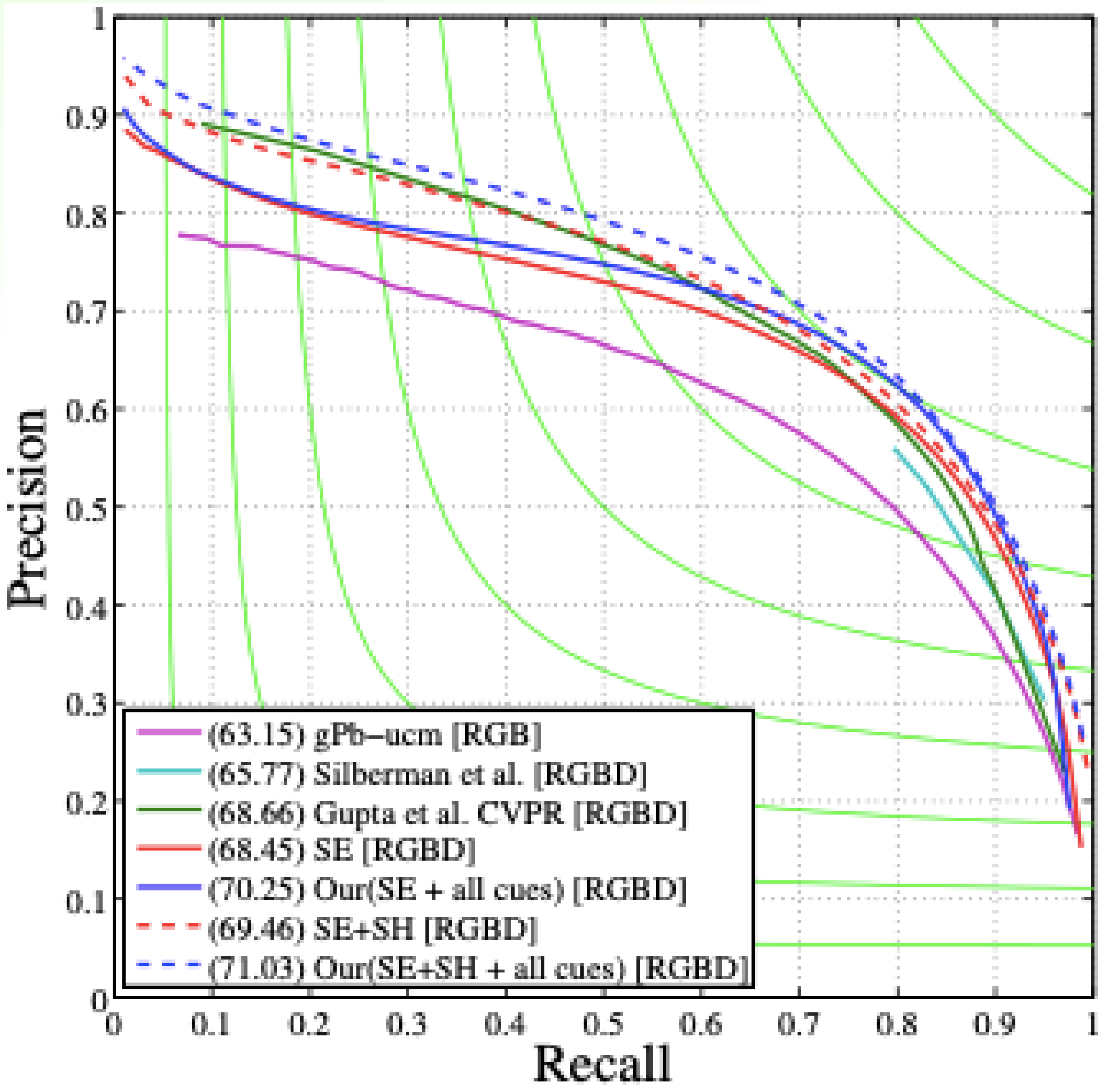



Fig 2: Precision-Recall curve on boundaries on the NYUD2 dataset

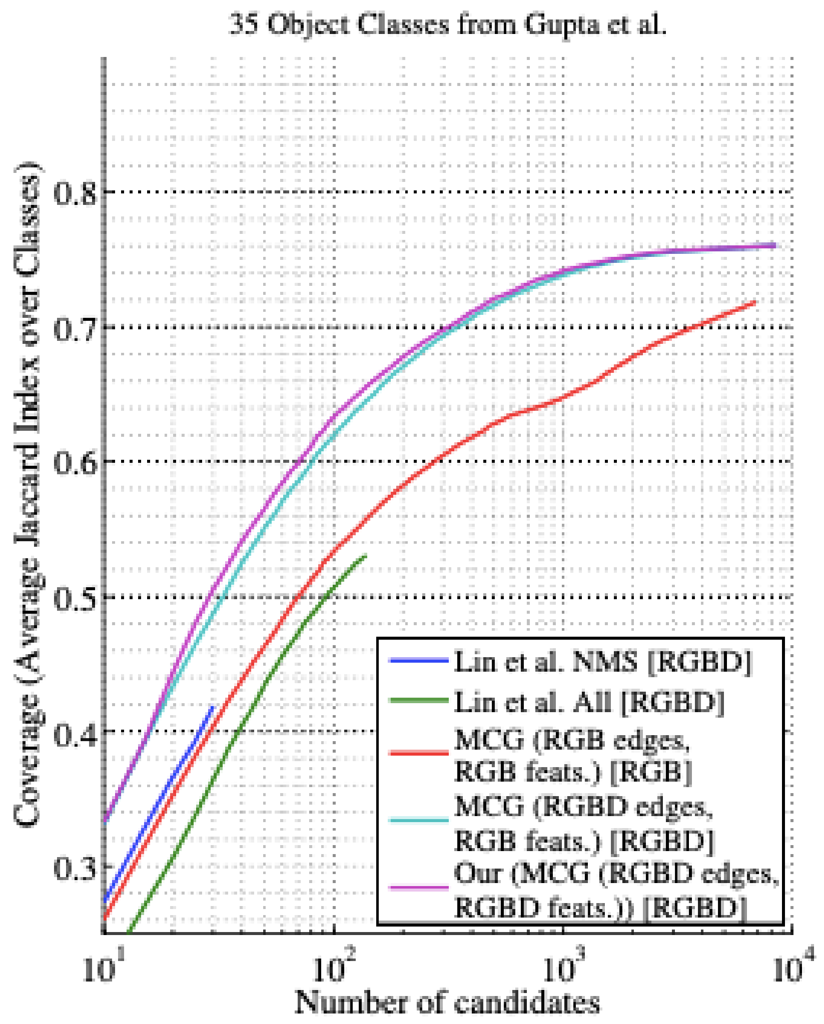



Fig. 4. Region Proposal Quality: Coverage as a function of the number of region proposals per image: Depth-based region proposals using the improved RGB-D contours work better than all past approaches, while at the same time being more general. Note that the X-axis is on a log scale.

## OBJECT DETECTION

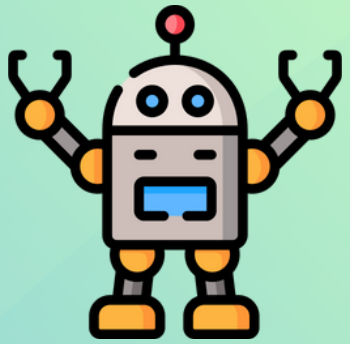
Final object detection system achieves an average precision of 37.3%, which is a 56% relative improvement over existing methods

# Why is this Problem Interesting

FACILITATE THE USE OF PERCEPTION IN FIELDS LIKE SMART TRAFFIC MONITORING, SUPPLY-CHAIN MONITORING, MEDICAL ASSISTANCE

- Advanced human-machine interfaces (gesture recognition, motion tracking, etc.)
- Crowd monitoring
- Self-driving cars
- Navigation systems for unmanned robotic vehicles, and much more


Virtual Reality



Robot Navigation

Augmented Reality

# Challenges →

Depth images may contain noise due to sensor limitations or environmental factors such as reflections or interference from other devices. This can lead to inaccurate depth measurements, which can affect the accuracy of object detection and segmentation

Processing both RGB and depth information requires more computational resources than traditional 2D image-based methods. This can lead to slower detection and segmentation times, which may be a problem in real-time applications

Objects can vary in shape, size, color, and texture, which can make it difficult to create accurate models for object detection and segmentation. This can lead to false positives or false negatives

Training deep learning models for RGB-D object detection and instance segmentation requires large amounts of annotated data. However, there are currently only a limited number of publicly available annotated datasets for this task

- Data Quality
- Computational Complexity
- Variability in object appearance
- Limited availability of annotated dataset

# Scope of the Project

What are we attempting as a part of the project

Components we are going to implement

OBJECT DETECTION

INSTANCE SEGMENTATION

Exact details regarding the implementation of all the sub-components - Timeline Slide

# Methodology →

## 1 Datasets
What and how to access?

**Dataset name**

NYU Depth Dataset v2

**Access**

Open Source

## 2 Computing Resources
What and how much?

**Specifications**

**GPU**: RTX 2080Ti
**CPU**: Intel Xeon E5-2640 v4

**Estimated Time**

Finetuning: 90 hrs (100 epochs)

## 3 Baseline models
The requirements

2.5D region proposal

Geocentric encoding of depth information (HHA)

Random forest for **Instance Segmentation**

Finetune R-CNN for HHA

Implement **Semantic Segmentation**

## 4 Improvements
The extras

Replace R-CNN with faster R-CNN

Using v2 of NYUD dataset

Replace SVM classifier with a Deep Learning base method

# Timeline →

**4 Apr**

Write code for 2.5 D region proposal.

**8 Apr**

Implement geocentric encoding of depth information

**10 Apr**

Fine tune R-CNN for geocentric encoding.

**12 Apr**

Random forest for instance segmentation.

# Timeline →

**14 Apr**

Implement Semantic Segmentation.

**16 Apr**

Replace SVM classifier for object detection.

**18 Apr**

Replace R-CNN with faster R-CNN.

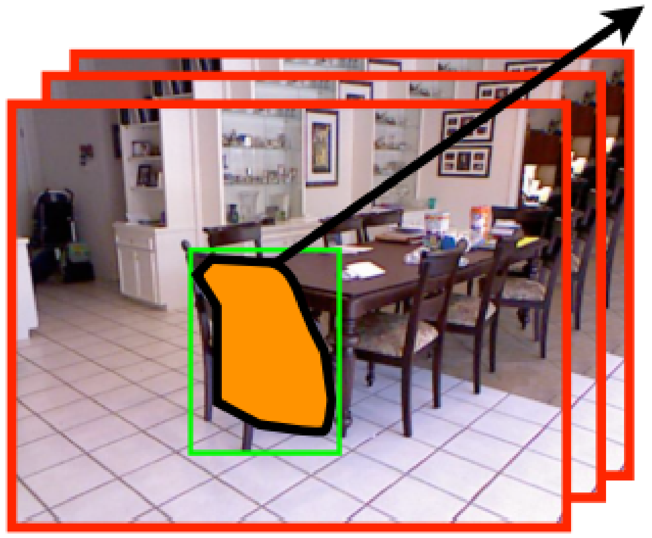
**21 Apr**

Final presentation.

# Initial Progress →

## Implementation of 2.5D region proposal

[4] Arbela´ez, P., Pont-Tuset, J., Barron, J., Marques, F., Malik, J.: Multiscale combinatorial grouping. In: CVPR (2014)

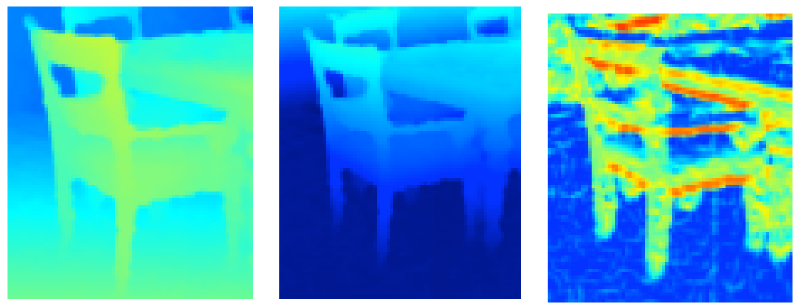


Region Proposal
Generation

## Encoding Depth Images for Feature Learning

[18] Gupta, S., Arbel´ aez, P., Malik, J.: Perceptual organization and recognition of indoor scenes from RGB-D images. In: CVPR (2013)

Geocentric Encoding
of Depth



Disparity Height Angle

# Team →

**Deepti Rawat**
deepti.rawat@research.iiit.ac.in
Roll No: 2022801016

**Rishabh Khanna**
rishabh.khanna@research.iiit.ac.in
Roll No: 2019113025

**Mihir Rawat**
mihir.r@research.iiit.ac.in
Roll No: 2020113005

**Kshitijaa Jaglan**
kshitijaa.jaglan@research.iiit.ac.in
Roll No: 2019115005

# Work Distribution →

## Feature Representation

- ✓ **2.5D Region Proposal**
  Kshitijaa, Mihir
- ✓ **Geocentric Encoding**
  Deepti, Rishabh
- **Fine Tuning R-CNN**
  Deepti, Rishabh
- **RGB R-CNN**
  Mihir, Kshitijaa

## Segmentation Pipeline

- **SVM Classifier**
  Deepti, Mihir
- **Instance Segmentation**
  Rishabh, Kshitijaa
- **Semantic Segmentation**
  Deepti, Mihir

## Proposed Changes

- **Faster R-CNN**
  Mihir, Kshitijaa
- **DL Classifier**
  Kshitijaa
- **Instance Segmentation**
  Deepti, Rishabh

# References


- [1] https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html
- [2] Gupta, S., Arbela´ez, P., Malik, J.: Perceptual organization and recognition of indoor scenes from RGB-D images. In: CVPR (2013)
- [3] Dolla´r, P., Zitnick, C.L.: Structured forests for fast edge detection. In: ICCV (2013)
- [4] Arbela´ez, P., Pont-Tuset, J., Barron, J., Marques, F., Malik, J.: Multiscale combinatorial grouping. In: CVPR (2014)
- [5] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR (2014)
- [6] Arbela´ez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. TPAMI (2011)