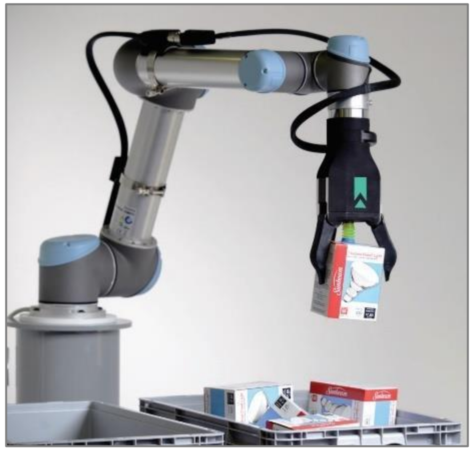
Ensemble Learning Methods:



for Robot Grasp Quality Estimation

Omri Green – Miles Gregg – Justin Smith – Fadi Alladkani

Group 4

OUTLINE

PLANAR GRASPING PROBLEM

Robots often used in factories for pick-and place

Revolutionized several industries

Why not use grasping in other places? (**Generalized Grasping**)

HOUSEHOLDS **RECYCLING PLANTS** **WAREHOUSES**

PLANAR GRASPING PROBLEM

Grasping general objects is difficult

Excel in controlled environments

- **Known** object shapes
- **Known** rough object locations
- **Known** assumptions
- **Facilitates** widespread adoption

Lose Performance in unknown environments

- **Unknown** object shapes
- **Unknown** object types and quantity
- **No** a-priori assumptions on environment
- **Prevents** widespread adoption

MODERN GRASPING APPROACHES

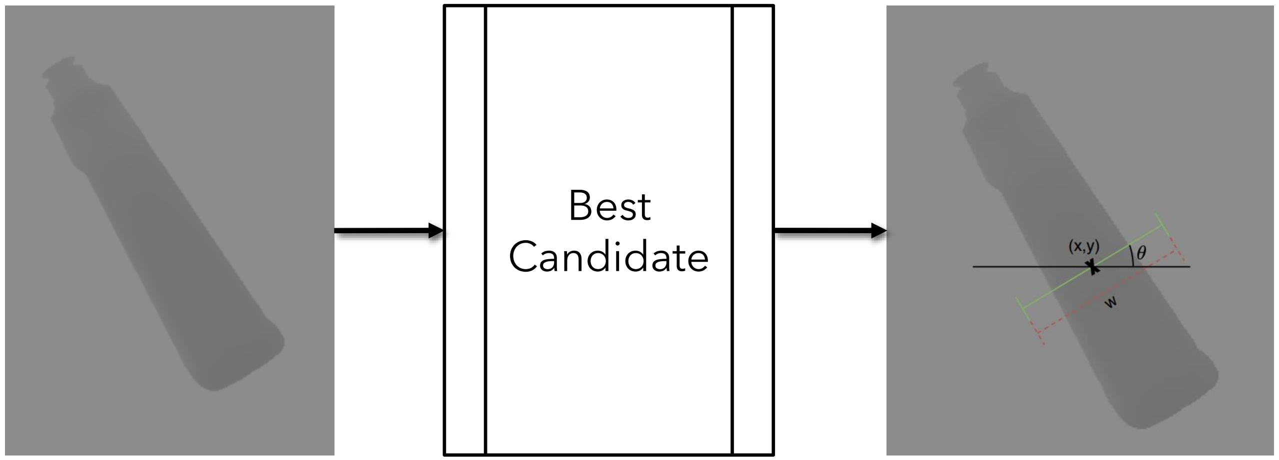
- Best Candidate
- Quality Estimation
- Generative

Modern solutions to generalized planar grasping

- **Data-Driven**
- **Attempt to generalize**
- **Three major approaches**

MODERN GRASPING APPROACHES

- Best Candidate
- Quality Estimation
- Generative



MODERN GRASPING APPROACHES

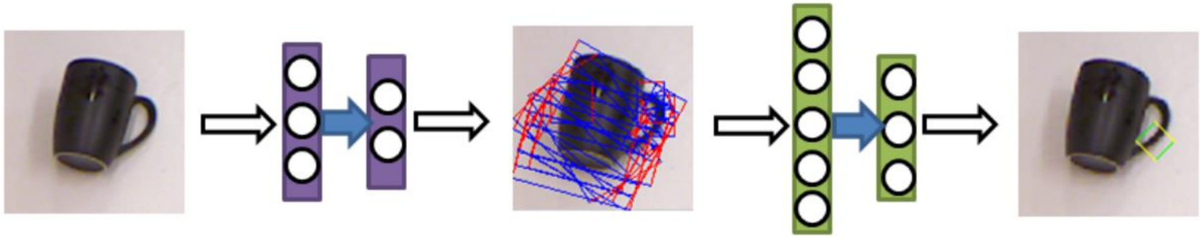
- Best Candidate
- Quality Estimation
- Generative

Representative Algorithm

Fast-Search [1]

Two-Stage Process

1. Neural Network estimates top candidate grasps
2. Second Neural Network selects best grasp from candidates chosen in **step 1**



[1] LENZ ET AL. "**DEEP LEARNING FOR DETECTING ROBOT GRASPS**", THE INTERNATIONAL JOURNAL OF ROBOTICS RESEARCH, **2015**

MODERN GRASPING APPROACHES

Best Candidate

Quality Estimation

Generative

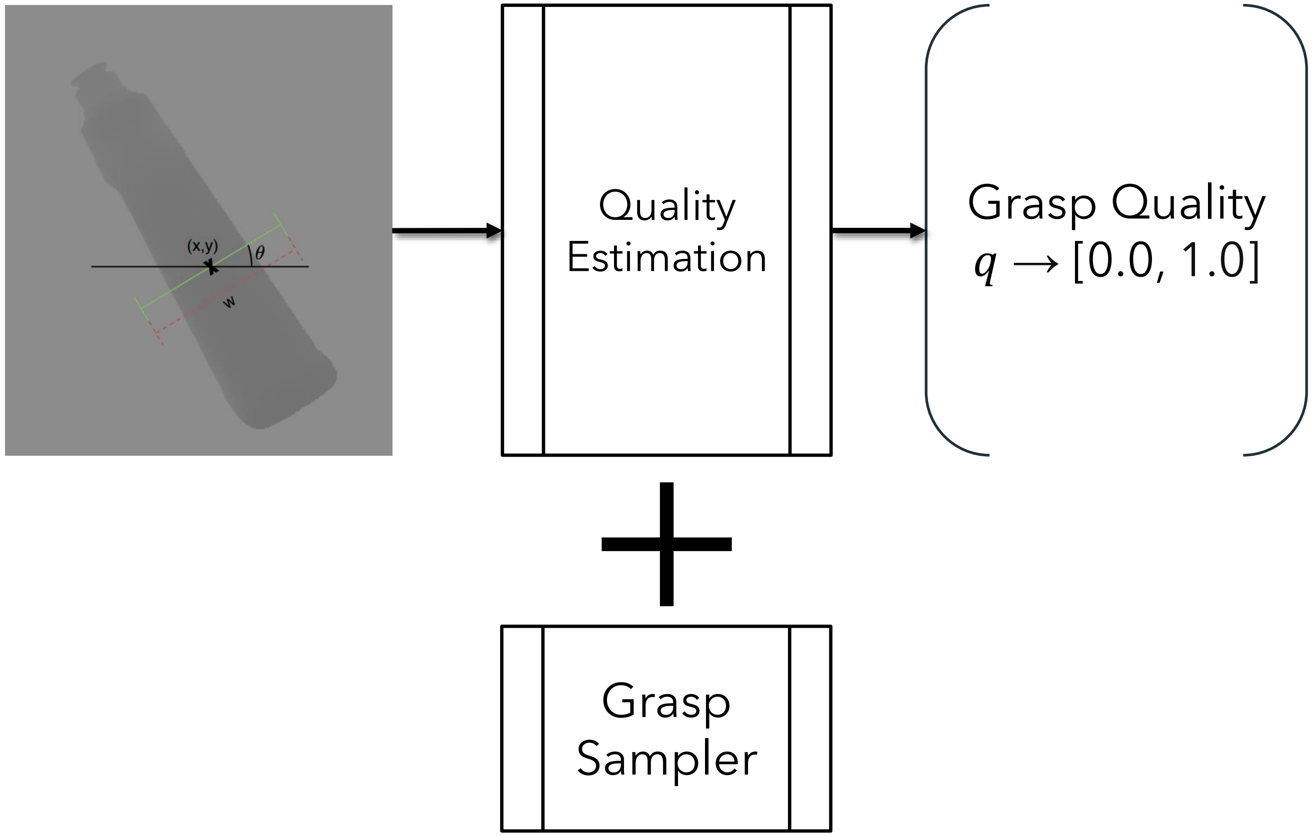
Question How stable is this grasp on this object?

Grasp Quality used as a metric for grasp stability ($[0.0, 1.0]$)

Quality Estimation uses Convolutional Neural Networks to estimate **Grasp Quality**

MODERN GRASPING APPROACHES

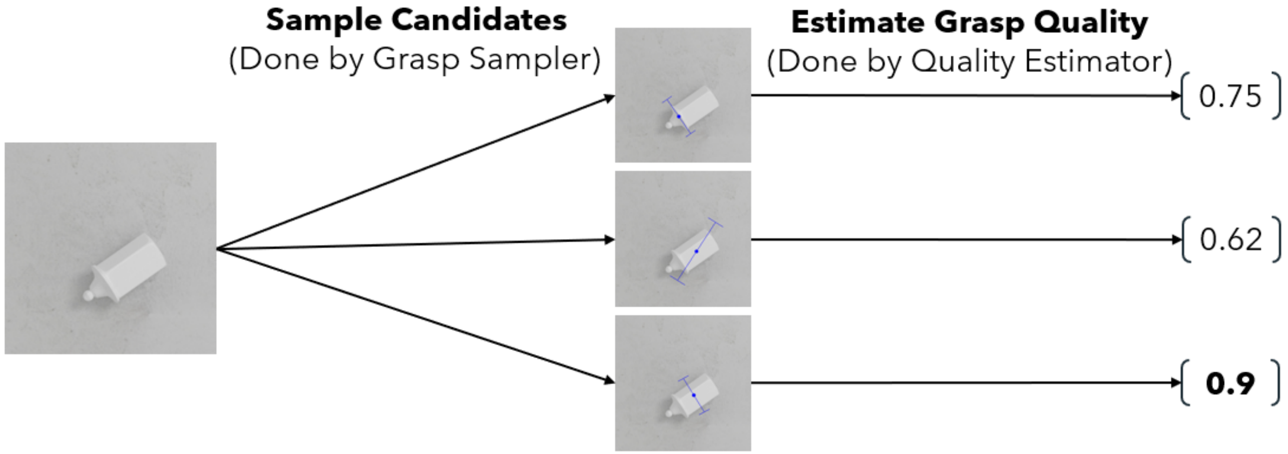
- Best Candidate
- Quality Estimation
- Generative



MODERN GRASPING APPROACHES

- Best Candidate
- Quality Estimation
- Generative

Sample multiple grasps and rank
Select best one (highest *grasp quality*)



MODERN GRASPING APPROACHES

Best Candidate

Quality Estimation

Generative

Representative Algorithm

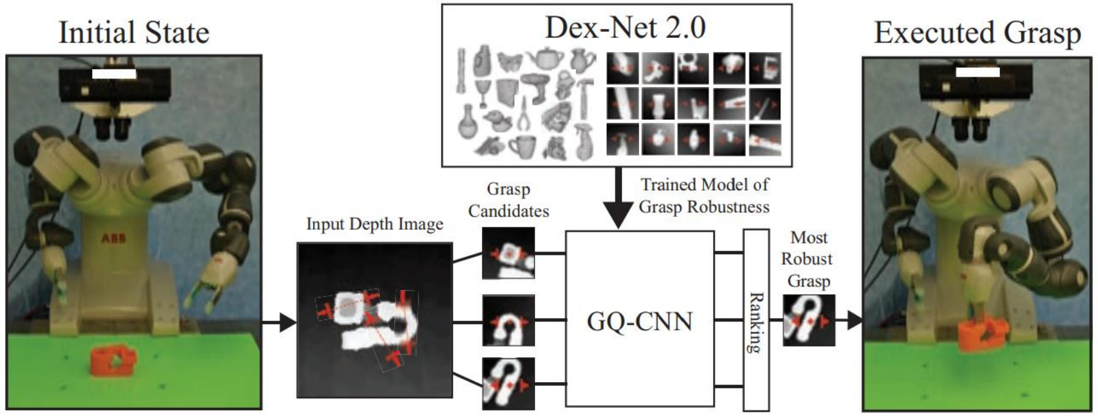
Dexnet-4.0

Grasp Quality Convolutional Neural Network [1] (GQCNN)

Trained on large synthetic dataset

Fully convolutional neural network

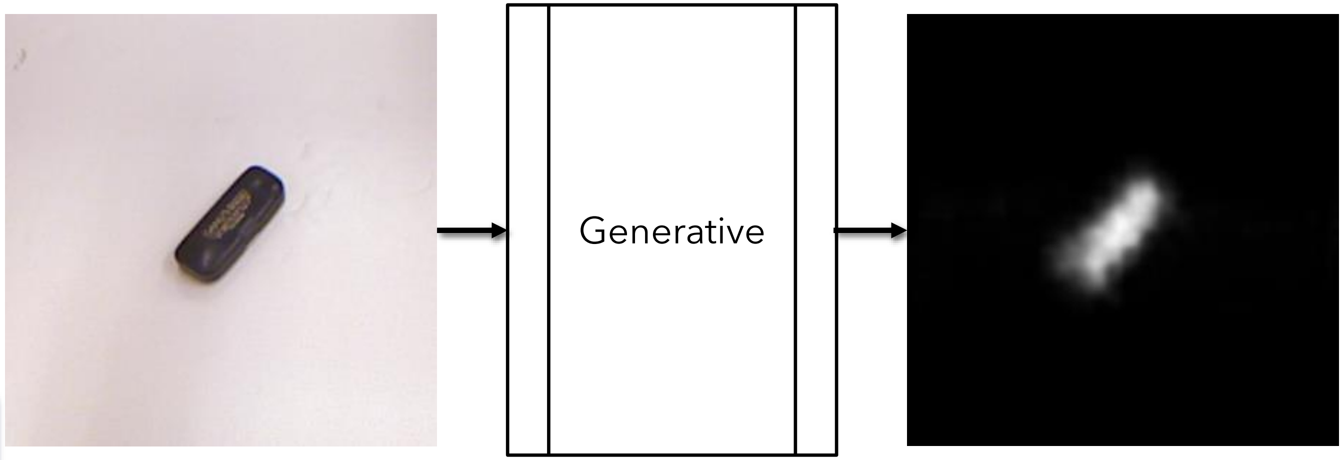
Uses depth images



[1] MAHLER ET AL. "**LEARNING AMBIDEXTROUS ROBOT GRASPING POLICIES**", SCIENCE ROBOTICS **2019**

MODERN GRASPING APPROACHES

- Best Candidate
- Quality Estimation
- Generative



MODERN GRASPING APPROACHES

Best Candidate

Quality Estimation

Generative

Avoid sampling multiple grasps

Select pixel where grasp quality is highest

MODERN GRASPING APPROACHES

- Best Candidate
- Quality Estimation
- Generative

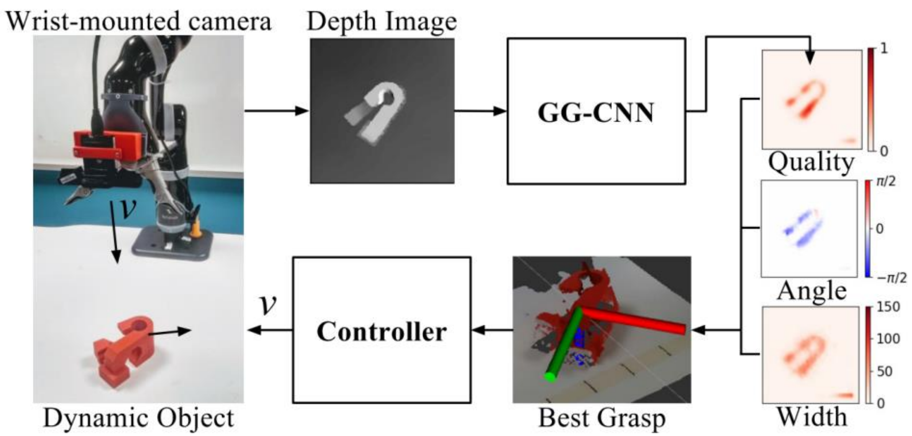
Representative Algorithm

Generative Grasping Convolutional Neural Network [1] (GGCNN)

Trained on sets of real-life images and grasping rectangles

Can run in real time

Uses depth images



[1] MORRISON ET AL. " **CLOSING THE LOOP FOR ROBOTIC GRASPING: A REAL-TIME, GENERATIVE GRASP SYNTHESIS APPROACH**", RSS **2014**

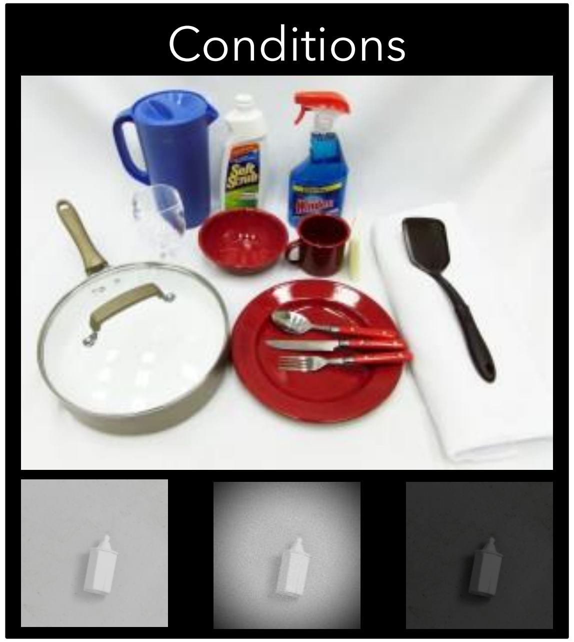
PROBLEM WITH EXISTING METHODS

- Above grasping algorithms attempt to generalize
- Still show difficulty grasping
  - Sensitivity to object shapes
  - Sensitivity to environmental conditions
  - Sensitivity to camera/lighting

PROBLEM WITH EXISTING METHODS

LARGE INPUT SPACE

Large **variety** of objects
Different environmental **conditions**
Different gripper **parameters**



PROBLEM WITH EXISTING METHODS

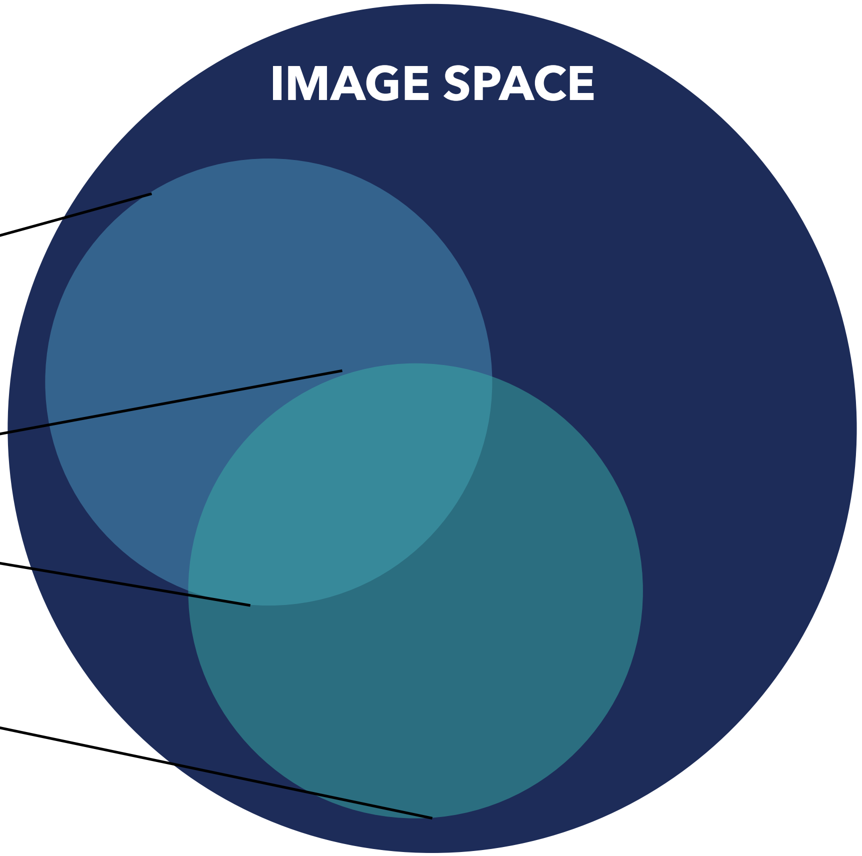
Inability to generalize

Sensitivity to environmental conditions

Grasping difficult / unknown objects

Grasping Algorithm A

Grasping Algorithm B



Input region where Algorithm A’s estimates is closest to ground truth

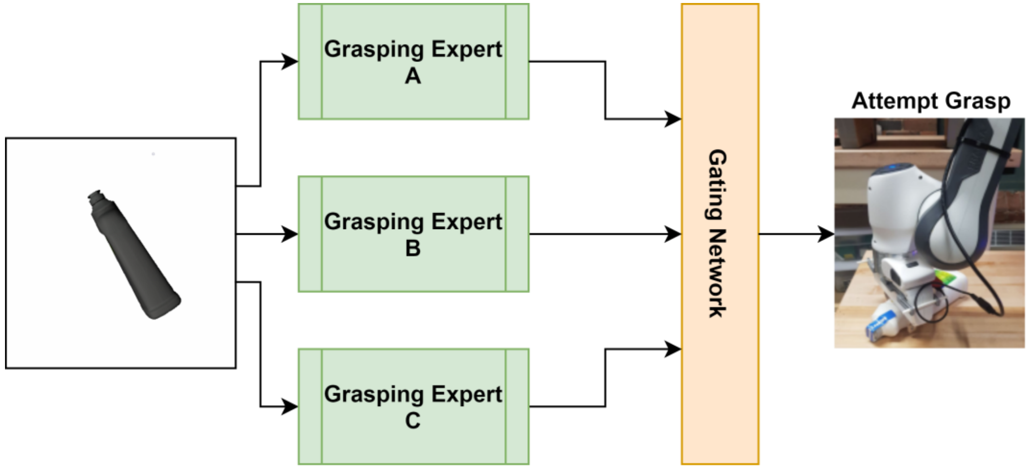
Input region where Algorithm B’s estimates is closest to ground truth

ENSEMBLE-BASED SOLUTION

ECNN: Ensemble Convolutional Neural Network

Combine multiple grasping algorithms

- Combination done by Gating Network
- Take advantage of strengths of each expert
- Overcome weaknesses of each



Emphasis on performance and flexibility

ENSEMBLING TECHNIQUES

- Best Candidate
- Quality Estimation
- Generative

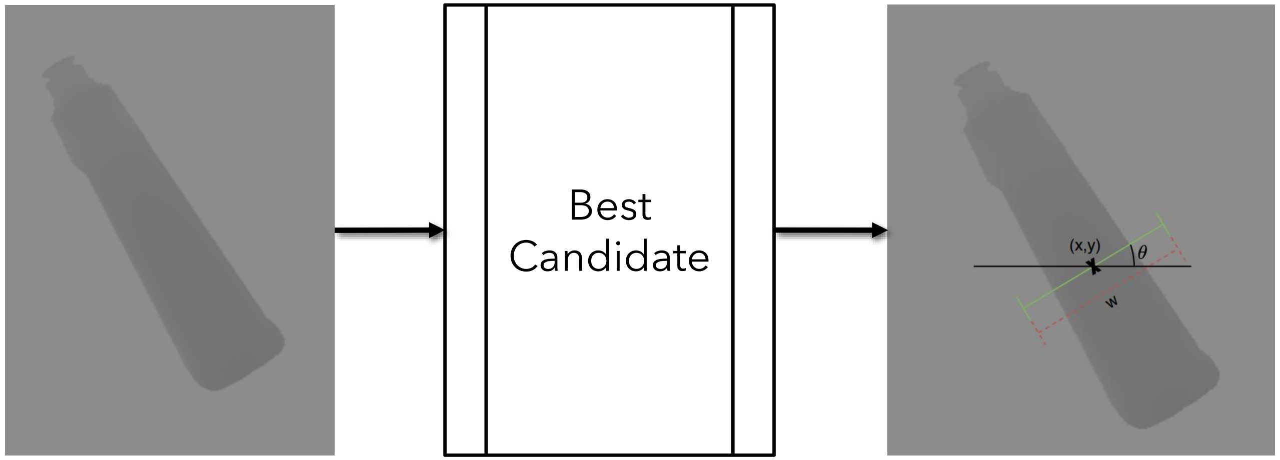
Ensemble expert candidacy

Different experts which can be used

Impacts ensemble network architecture

ENSEMBLING TECHNIQUES

- **Best Candidate**
- Quality Estimation
- Generative

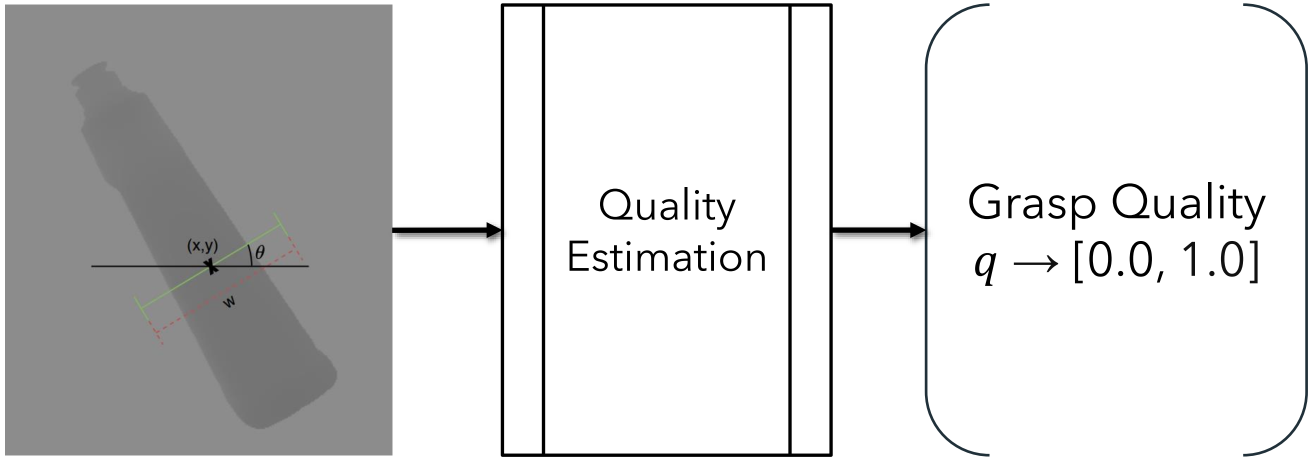


Combination through selection

Gating Network selects which grasp to execute

ENSEMBLING TECHNIQUES

- Best Candidate
- Quality Estimation
- Generative

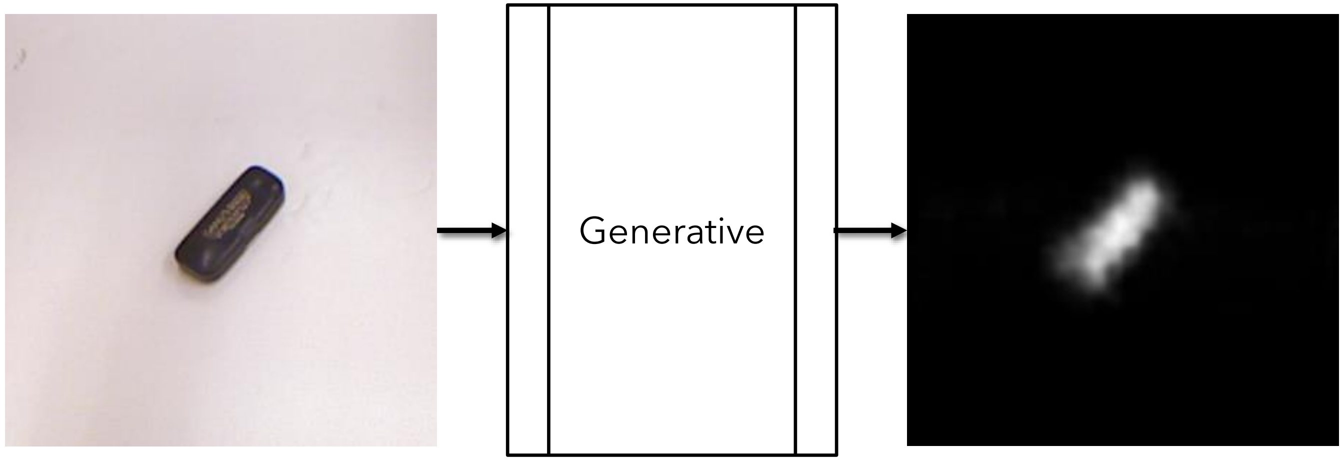


Combination through grasp quality

Gating Network calculates weighted average quality from each expert

ENSEMBLING TECHNIQUES

- Best Candidate
- Quality Estimation
- Generative



Combination through grasp quality

Gating Network calculates weighted average quality from each expert

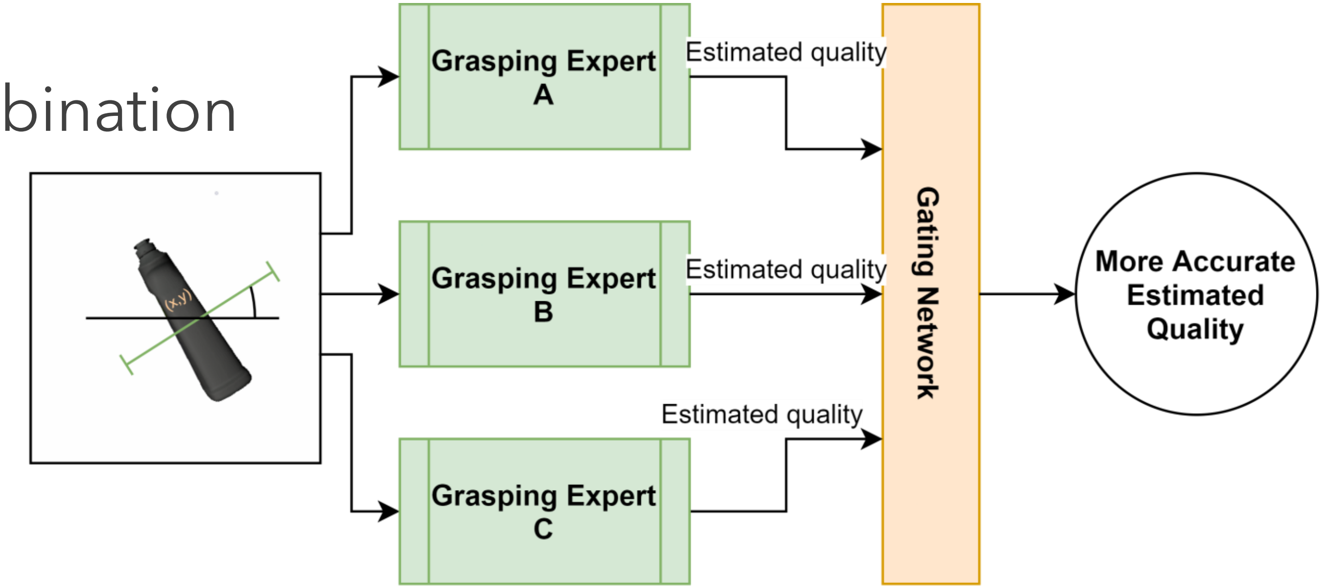
ENSEMBLE-BASED SOLUTION

ECNN: Ensemble Convolutional Neural Network

Choose **Quality Estimation** combination

- Avoid discarding expert opinions

  (Weighted sum ensures all experts contribute)
- Pair with Grasp Sampler
- Use **Mixture Of Experts** model



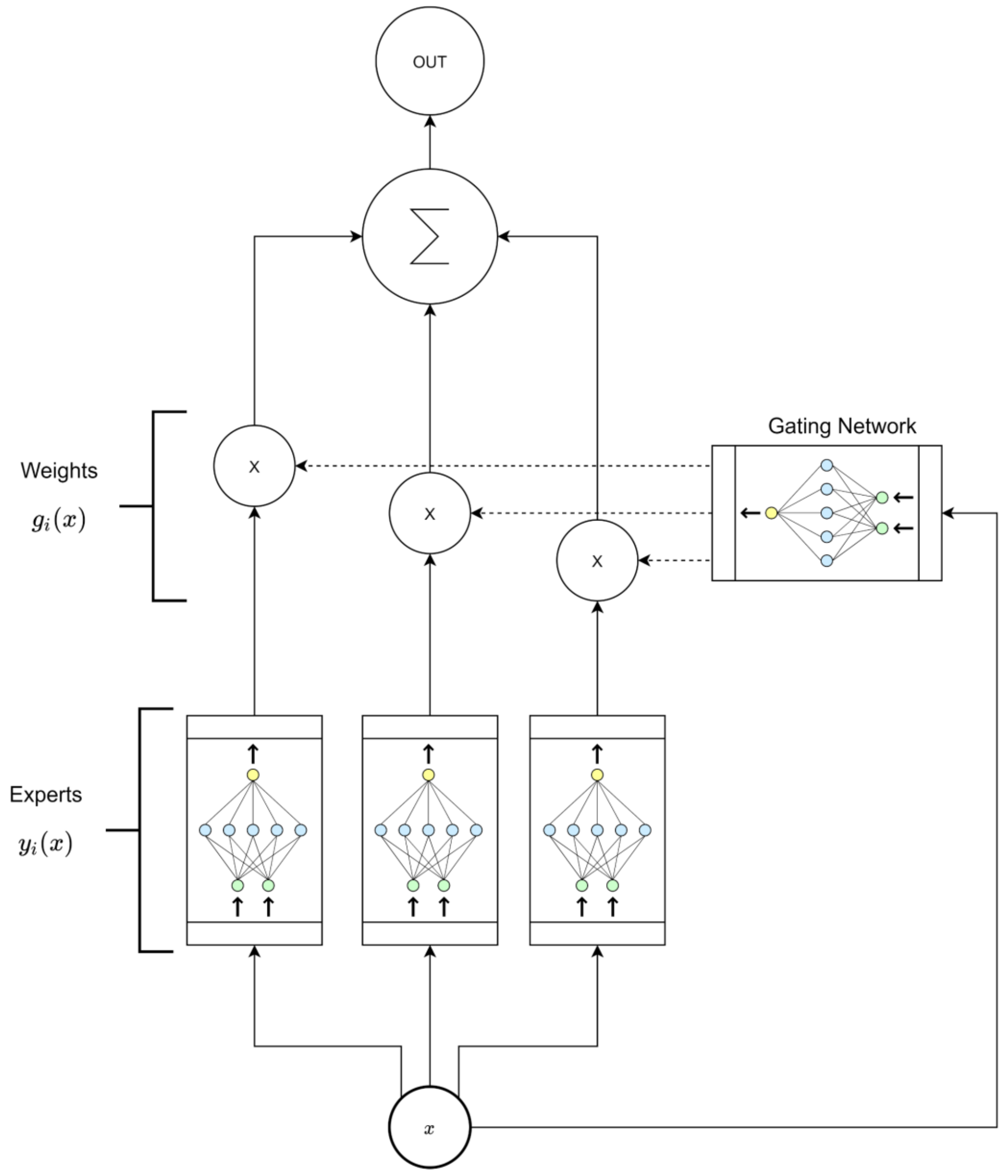
MIXTURE OF EXPERTS

Statistical Ensemble

Combine multiple classifier outputs

Improve overall performance

- Elimination of generalization errors
- Improve estimation accuracy



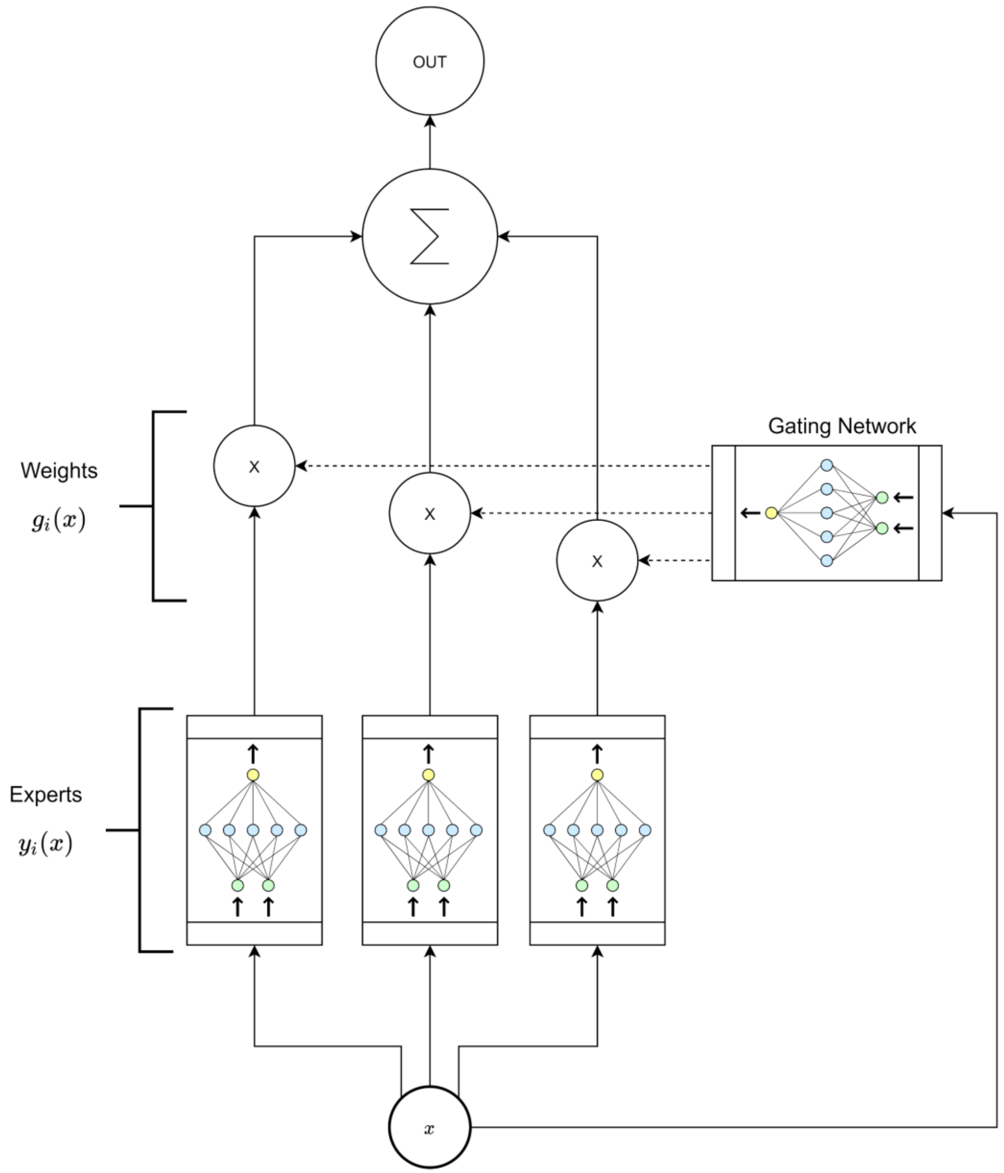
MIXTURE OF EXPERTS

Input-dependent weighted combination

- Weights as a function of {image, grasp}
  - **Assign weights to expert opinion based on the input**
  - **Learn which experts provide grasp quality closest to ground truth for which input**
- Gating Network

Benefits from expert diversity

$$y(x) = \sum_{i=1}^{n} y_i(x) g_i(x)$$

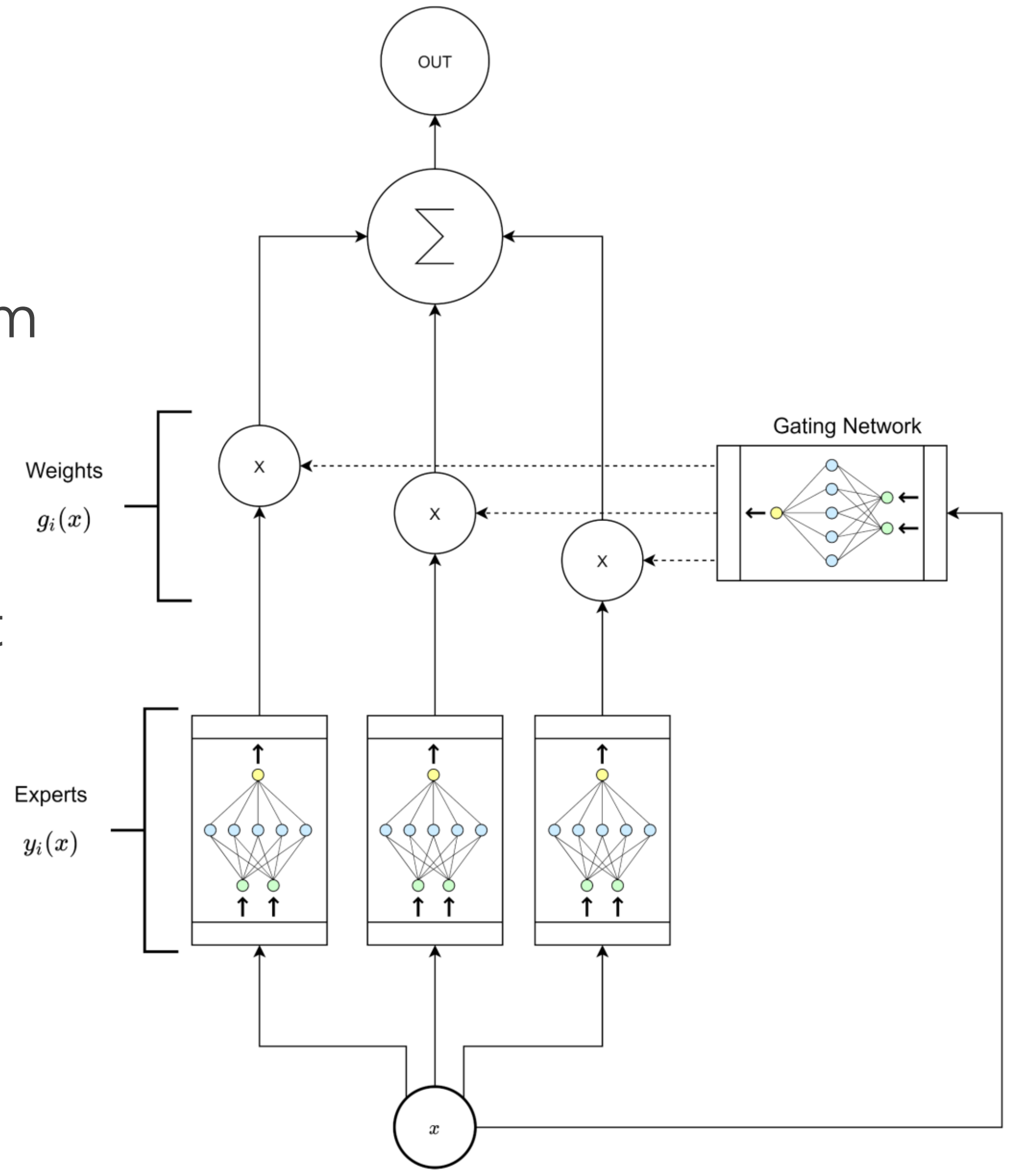


MIXTURE OF EXPERTS

Training Phase Learn which experts perform best on which inputs

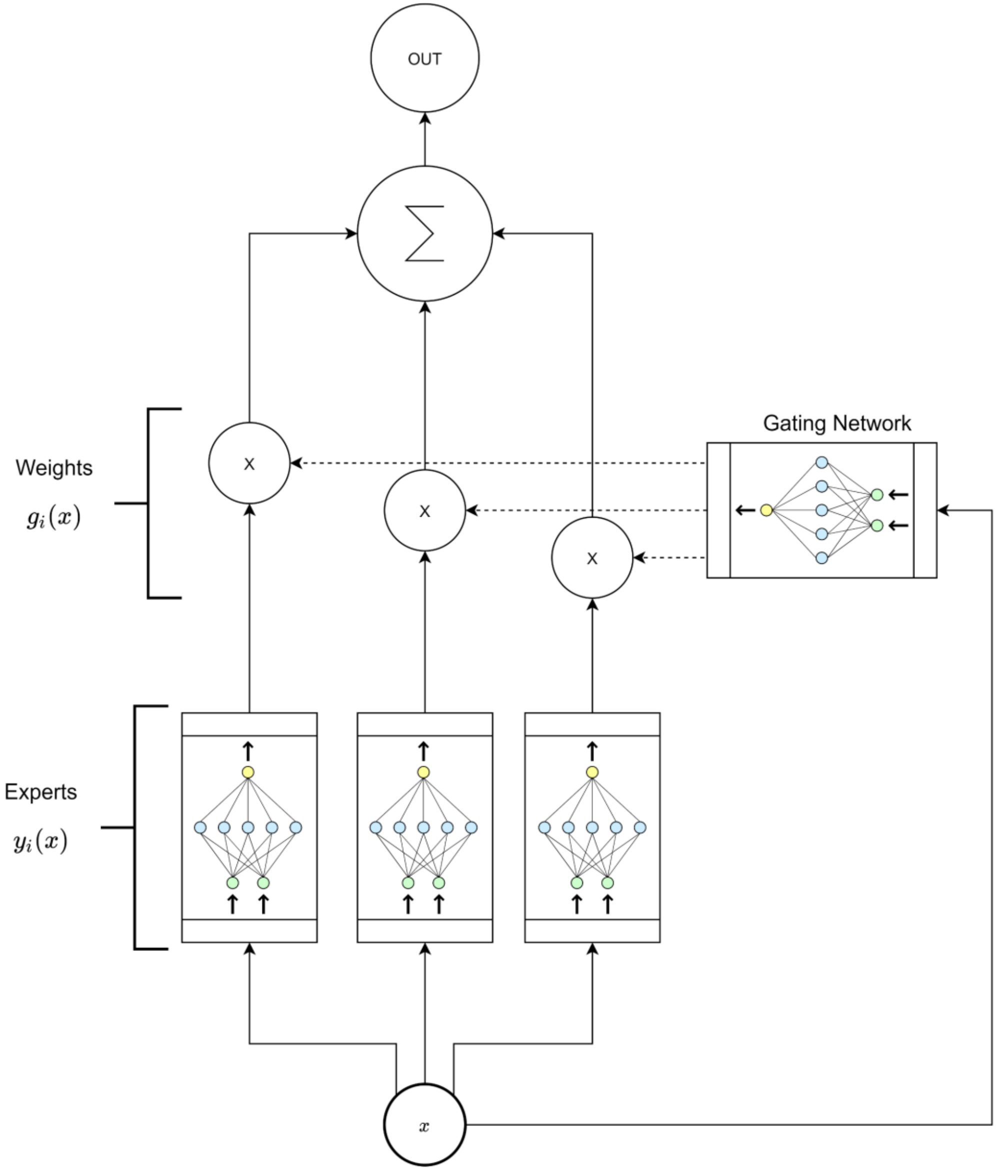
Evaluation Phase Use learned information to assign weights to experts based on input

$$y(x) = \sum_{i=1}^{n} y_i(x) g_i(x)$$



MIXTURE OF EXPERTS

Input	Grasping Expert Qualities	Gating Network Weights	Output
I	q_0	g_0	q
G	q_1	g_1	
	q_2	g_2	



MIXTURE OF EXPERTS

Advantages

- Existing open-source solutions and methods
- Less training
- Increased generalization

ECNNs: Ensemble CNNs

Goal: Design Gating Network

- Constant Weights (reference)
- Image (ImECNN)
- Grasp-Image (GrImECNN)

Goal: Expert Selection

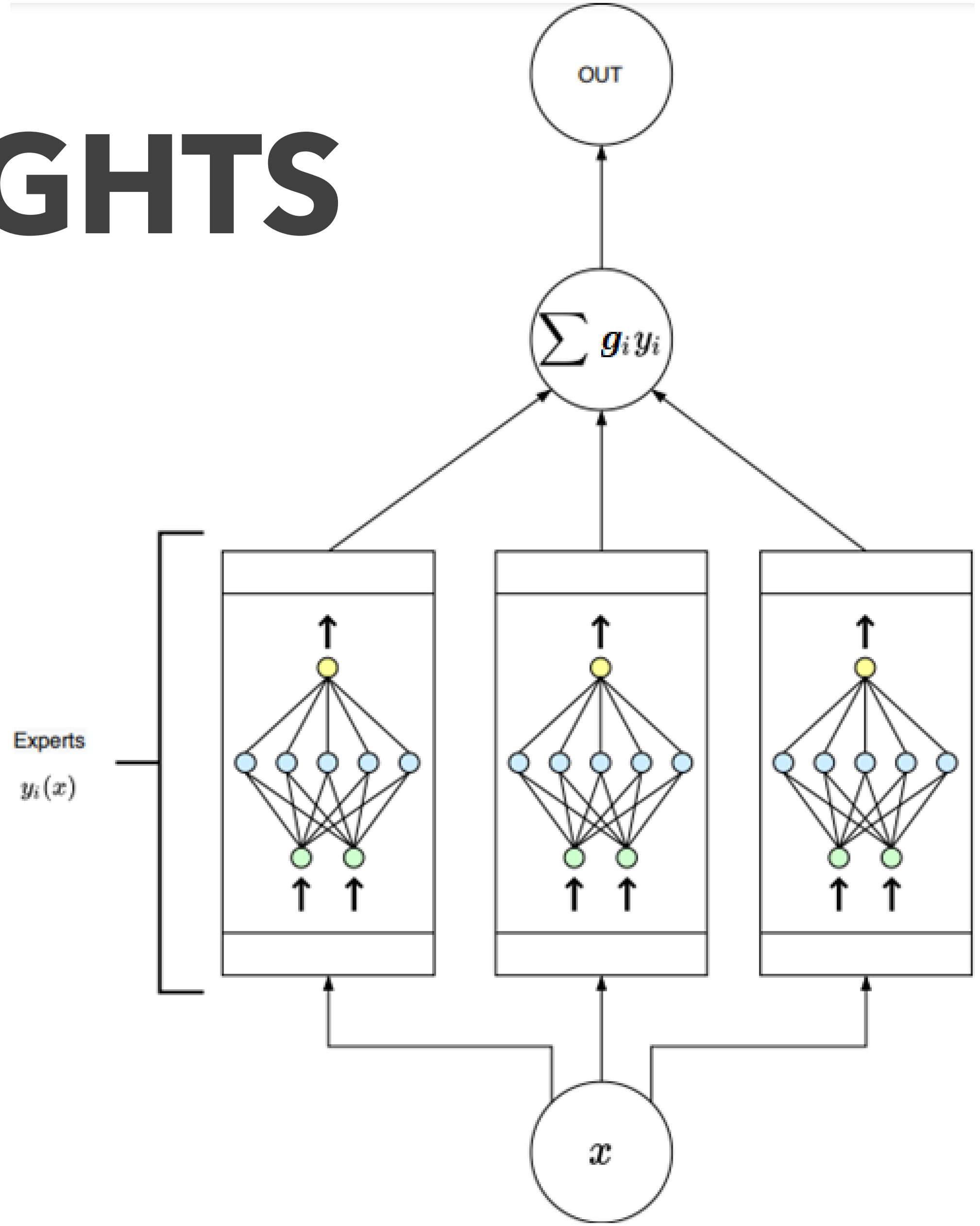
- Diversity
- Availability
- Accuracy

CONSTANT WEIGHTS

For Comparison

- Weights learned offline
- Weights independent of input

$$y = \sum_{i=1}^{n} g_i y_i(x)$$

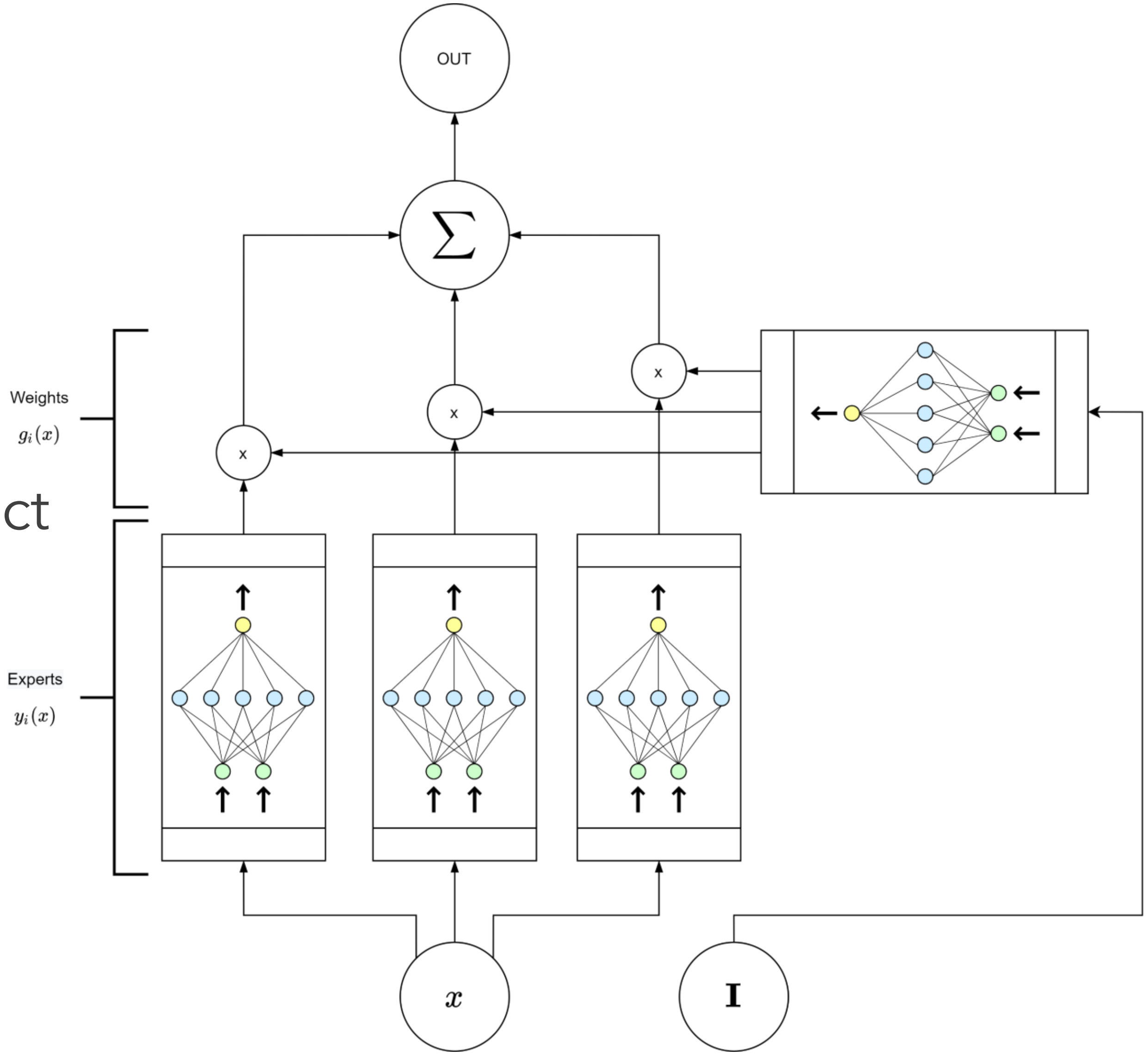


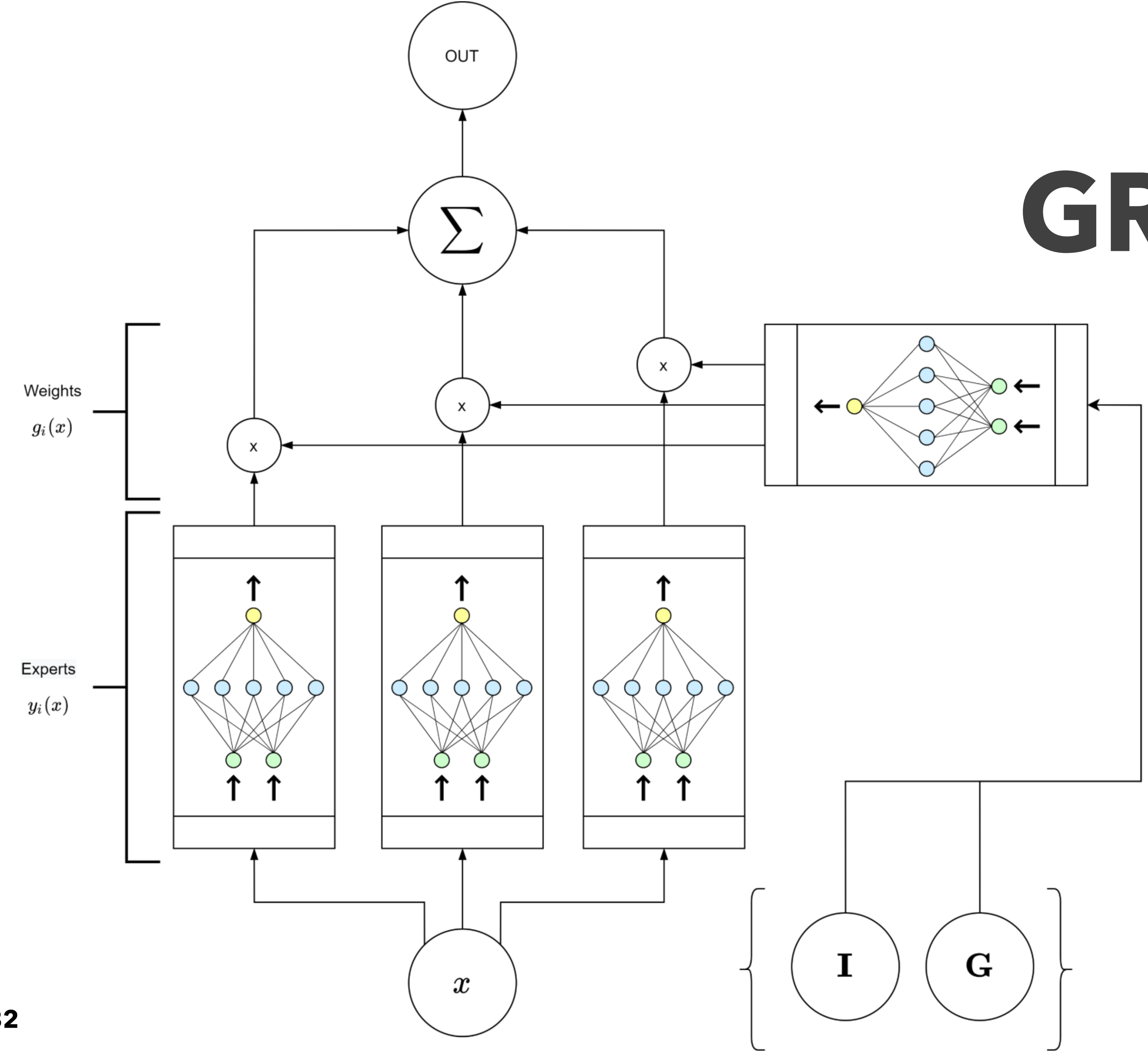
IMAGE

Expert Classification: per-object

- Convolutional Gating Network
- Weights dependent on image of object

$$y = \sum_{i=1}^{n} g_i(\mathbf{I}) y_i(x)$$





GRASP-IMAGE

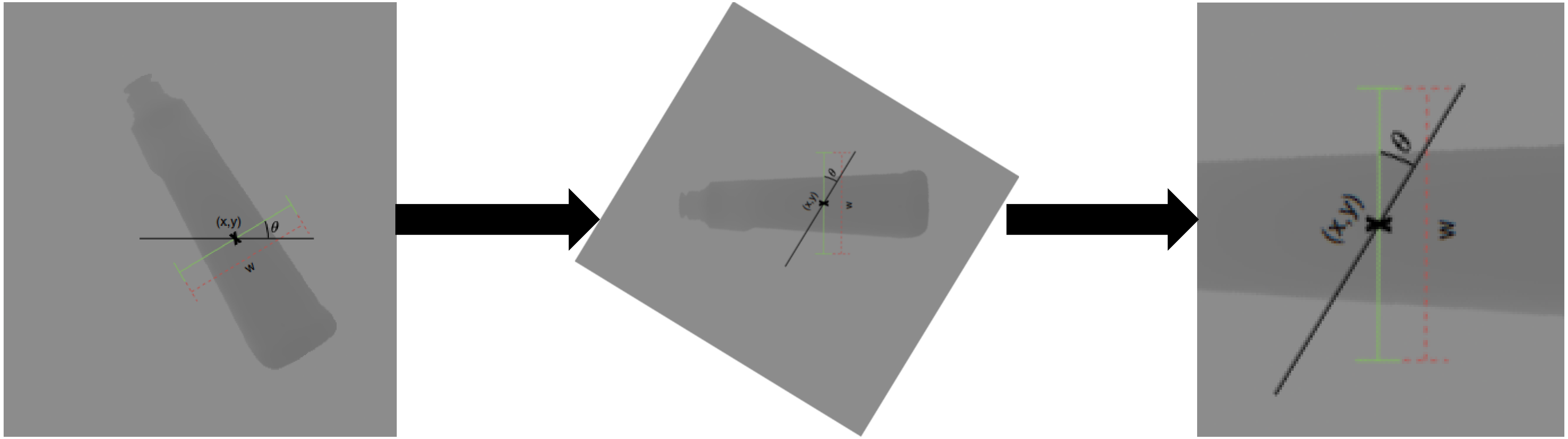
Expert Classification:
per-object, per-grasp

Grasp Integration:
Crop + Rotate Image

GRASP-IMAGE

Expert Classification: per-object, per-grasp

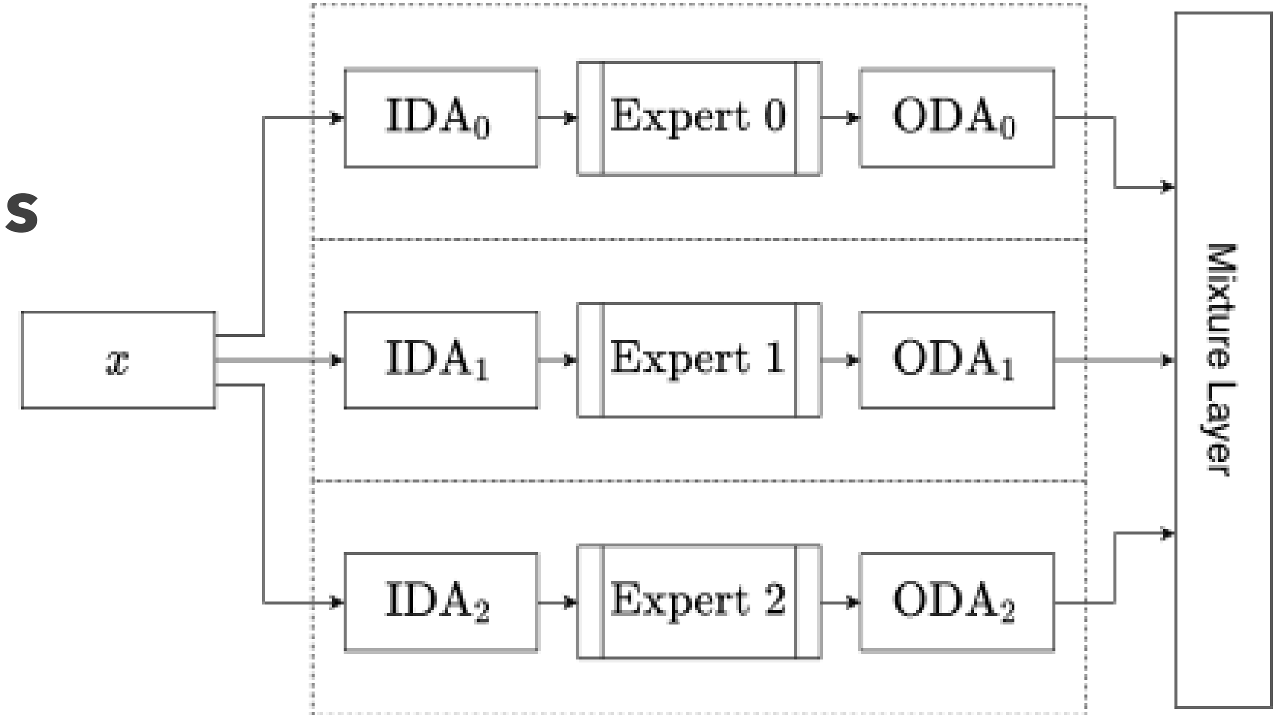
Grasp Integration: Crop + Rotate Image



COMPATIBILITY – DATA ADAPTERS

Varied grasping algorithms

High flexibility

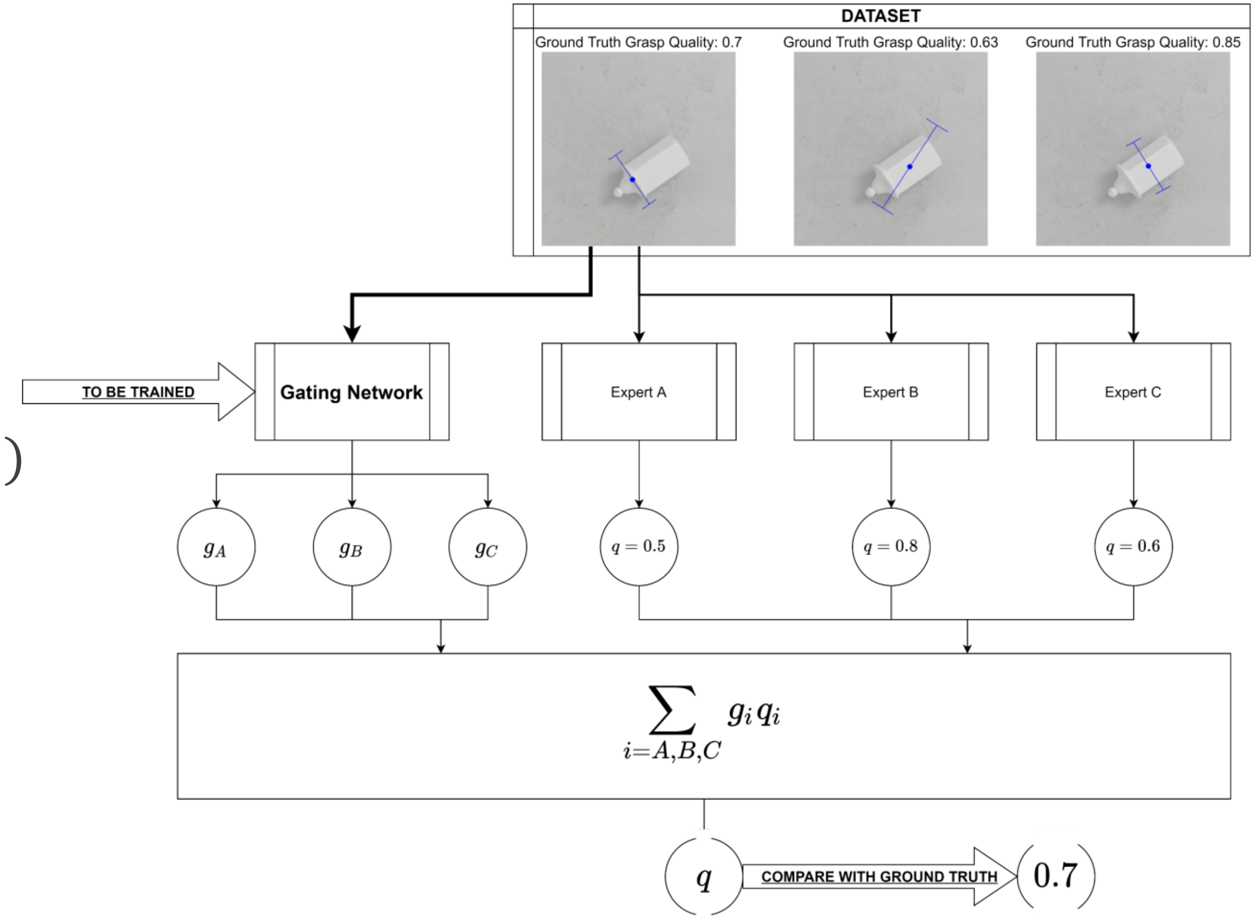


PERFORMANCE & TRAINING

Training

- Training Gating Network
- Network should learn mapping

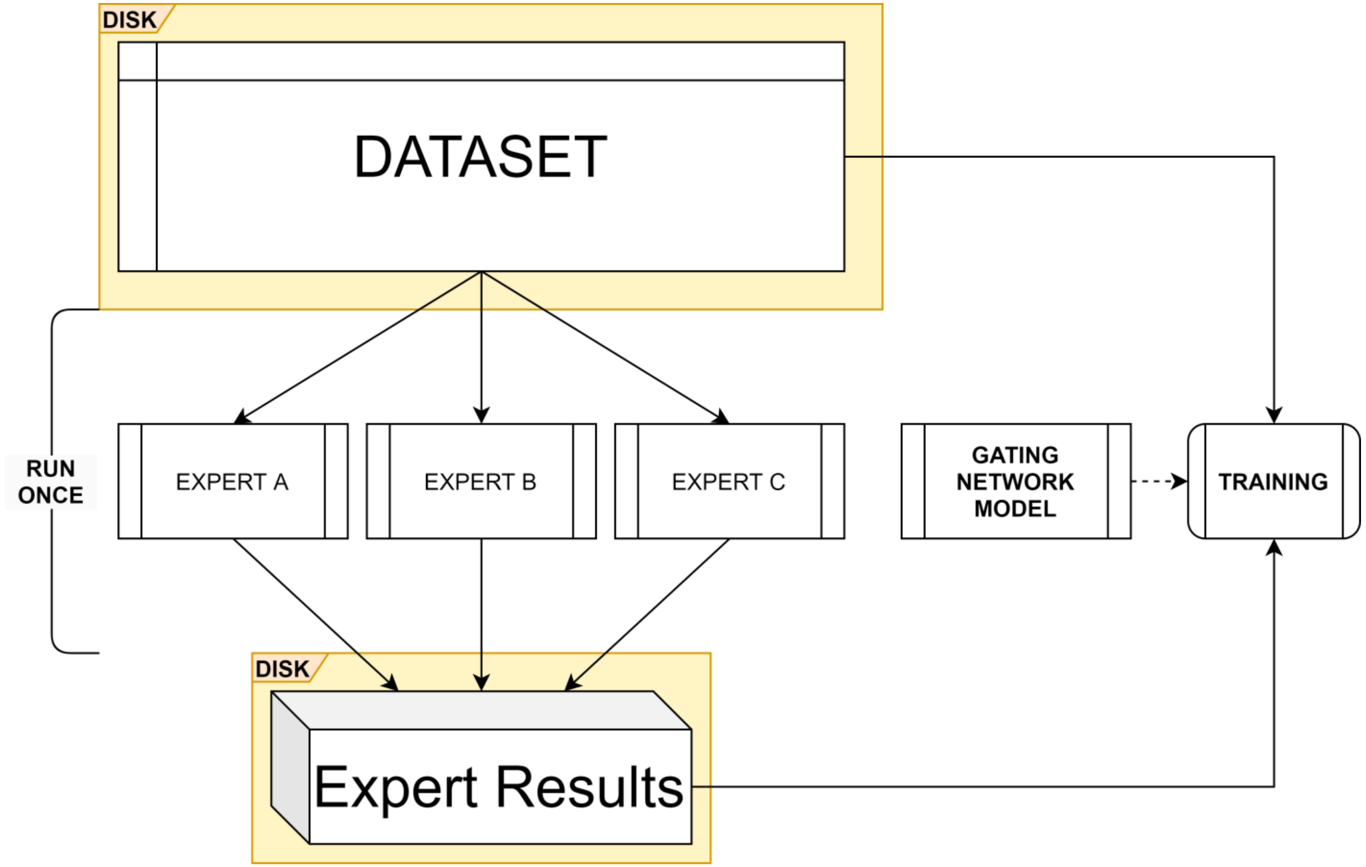
$(\text{Image}, \text{Grasp}) \rightarrow \text{Expert Weights } (g_i)$



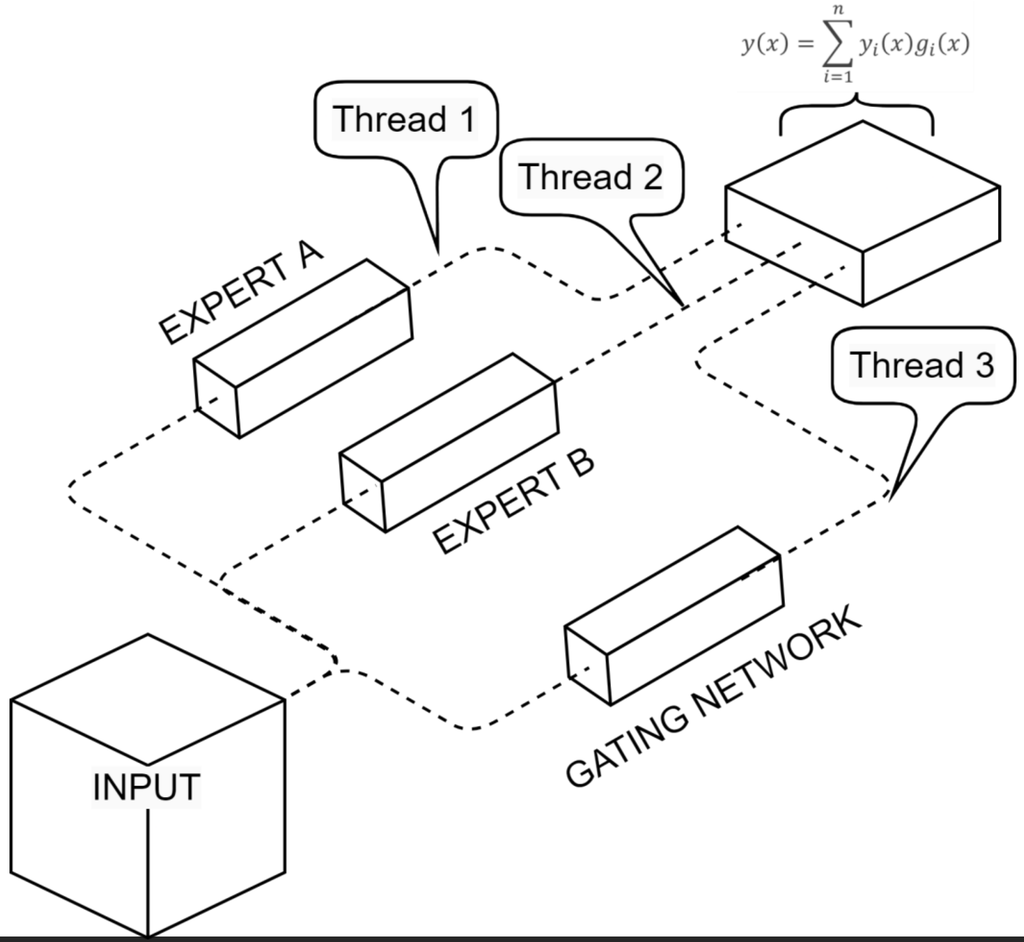
PERFORMANCE & TRAINING

Efficient Training

- Frozen expert models
- Run experts once, cache results



PERFORMANCE & TRAINING



Low Performance Overhead

- If possible, parallelize networks
- Small Gating Network

VERIFICATION

Sample ECNNs

Three Experts

- Finetuned Dexnet 4.0 (GQCNN-4.0)
- Generative Grasping CNN (GGCNN-D)
- Custom Generative Grasping CNN (GGCNN-RGB)

Training
Cornell Dataset[1]

→

Experiment
YCB Dataset[2]
Franka Emika + RealSense

[1] LENZ ET AL. "**DEEP LEARNING FOR DETECTING ROBOTIC GRASPS**", IJRR **2014**
[2] CALLI ET AL. "**YALE-CMU-BERKELEY DATASET FOR ROBOTIC MANIPULATION RESEARCH**", IJRR **2017**

EXPERT SELECTION

Emphasis on

DIVERSITY	AVAILABILITY	ACCURACY
GQ-CNN	**GGCNN-D**	**GGCNN-RGB**
Training: Synthetic Data	Training: Real Data	Training: Real Data
Depth Input	Depth Input	Color Input
Good on adversarial objects	Good in clutter	Good in clutter

TRAINING

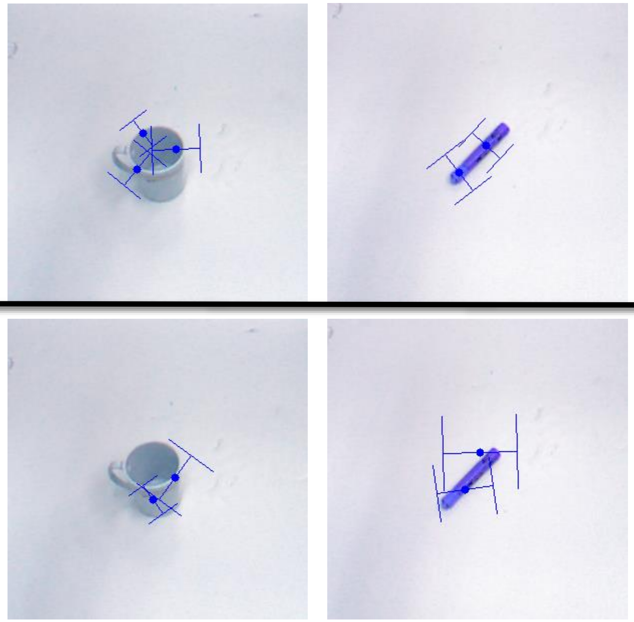
Three Ensembles

- Constant Weights
- ImECNN
- GrImECNN

Dataset

- Cornell Grasping Dataset
- Handlabelled good/bad grasps
- Images from real camera

	Mug	Pen	**Ground Truth Grasp Quality** (Each Grasp)
Good Grasps			1.0
Bad Grasps			0.0



GATING NETWORK	Constant Weights	ImECNN	GrImECNN
WEIGHT FUNCTION $g_i(\cdot)$	Input-Independent	Image-Dependent	(Grasp + Image)-Dependent

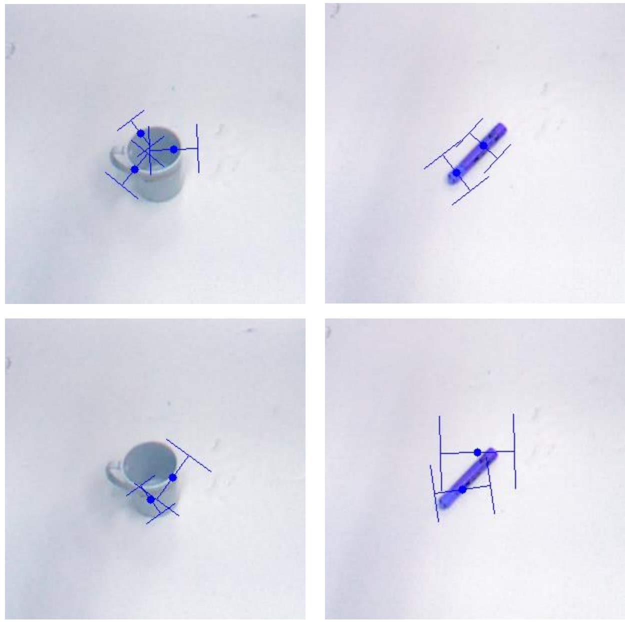
TRAINING

Metric

- Testing classification accuracy
- Grasp successfully classified if
  - Estimated $q < 0.5$ for **Bad Grasp**
  - Estimated $q > 0.5$ for **Good Grasp**

TRAINING

Mug | Pen



	Expert A	Expert B	
Good Grasps	3/5	2/5	Successfully classified **good grasps** (Estimated $q > 0.5$)
Bad Grasps	2/4	4/4	Successfully classified **bad grasps** (Estimated $q < 0.5$)
Grasp Classification Accuracy:	5/9	6/9	

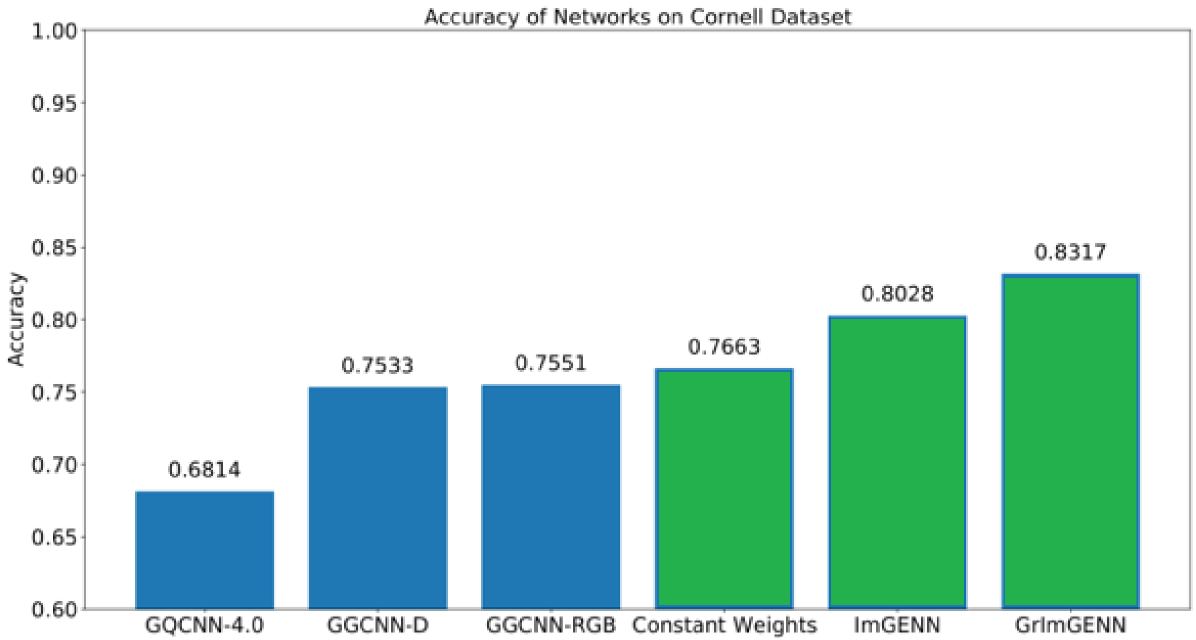
TRAINING

Three Ensembles

- Constant Weights
- ImECNN
- GrImECNN

Cornell Dataset

Up to 6% classification accuracy increase vs strongest expert



GATING NETWORK	Constant Weights	ImECNN	GrImECNN
WEIGHT FUNCTION $g_i(\cdot)$	Input-Independent	Image-Dependent	(Grasp + Image)-Dependent

EXPERIMENT

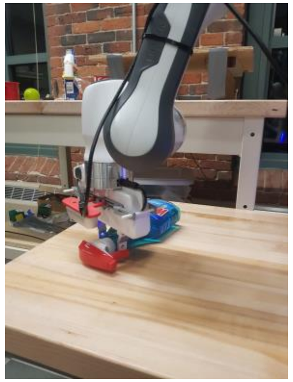
Verify performance of GrImECNN

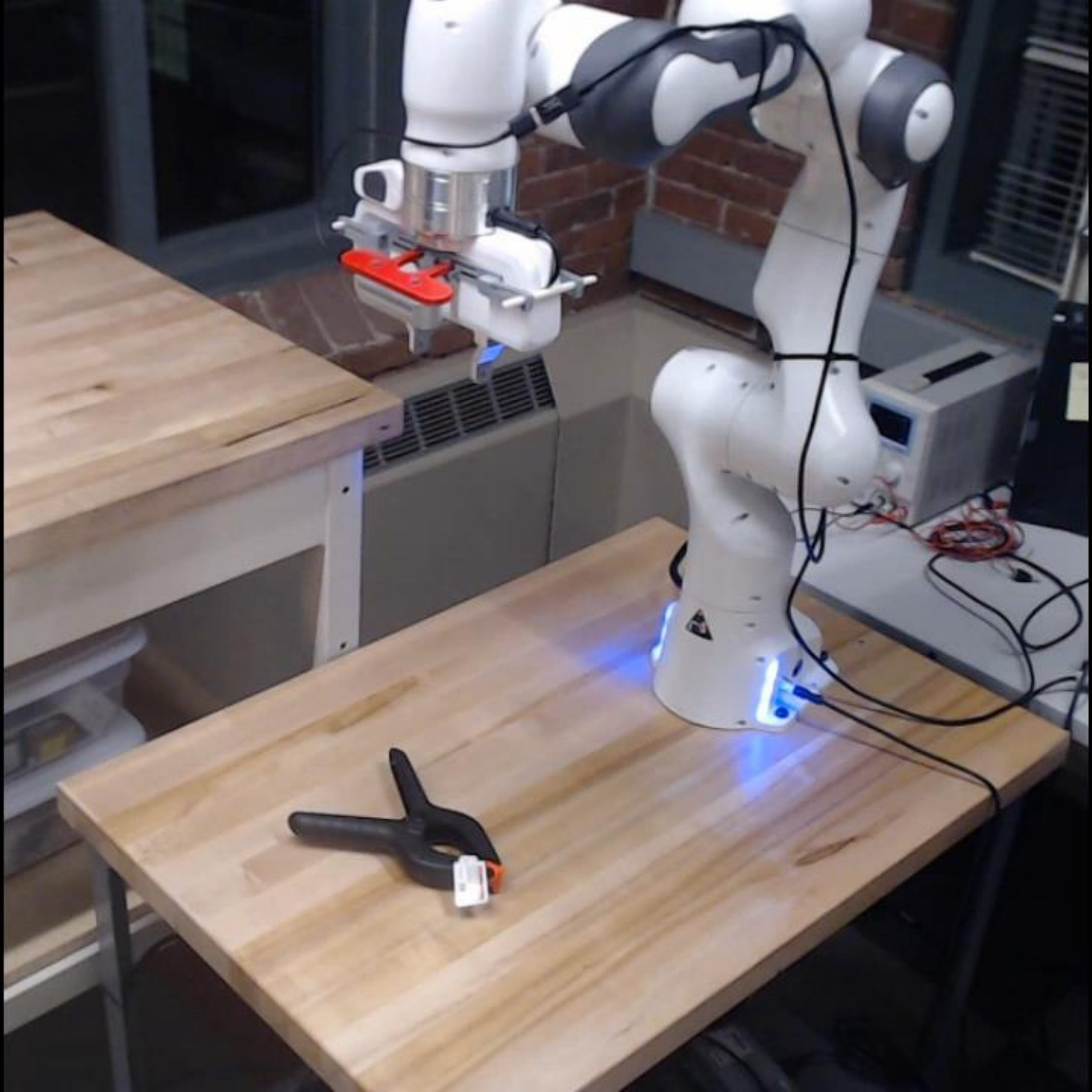
- 10 Challenging YCB Objects
- Three poses each

Most cases

accuracy(GrImECNN) ≥ **accuracy**(Best Expert)

Objects	Success			
	GQCNN	Gen-RGB	Gen-D	GrImECNN
Screwdriver	**3/3**	2/3	**3/3**	**3/3**
Windex	2/3	2/3	**3/3**	2/3
Mustard	1/3	1/3	**3/3**	**3/3**
Bleach	2/3	2/3	2/3	**3/3**
Pear	**2/3**	0/3	1/3	**2/3**
Banana	**3/3**	**3/3**	**3/3**	**3/3**
Mug	1/3	1/3	**2/3**	**2/3**
Spatula	**3/3**	**3/3**	2/3	**3/3**
Spring Clamp	**3/3**	**3/3**	2/3	**3/3**
Wine Glass	0/3	0/3	**1/3**	**1/3**
Total	20/30	17/30	22/30	**25/30**





CONCLUSION

Improved estimation of grasp quality
Low overhead in performance/training
Takes advantage of existing algorithms

Future Work:

- Additional experts & training data
- Impact of expert selection
- Different ensembling techniques