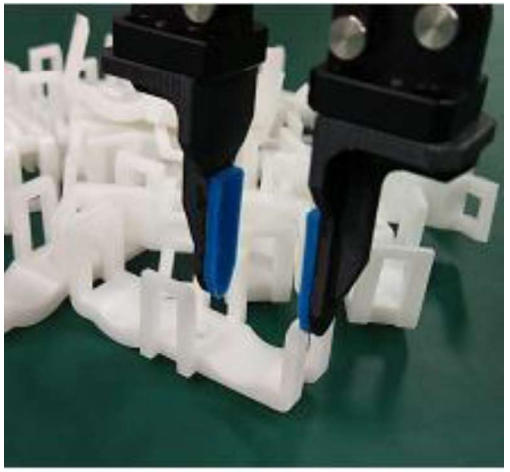
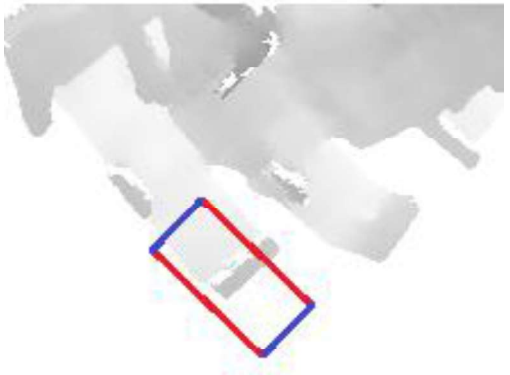
OUELINES

1. Research background
2. Research objectives
3. Related works
4. Grasp detection
5. Pose estimation
6. Conclusion and future work

4.1 Rectangle representation of grasp

- Use rectangle to represent the grasp of a parallel-jaw gripper [10].
  - Blue lines indicate the layout of fingers
  - Length of red line indicates the open width

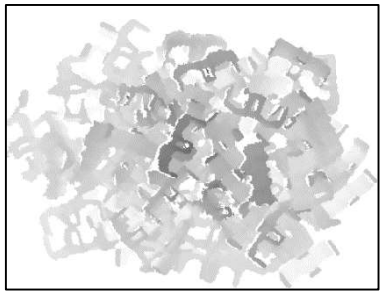




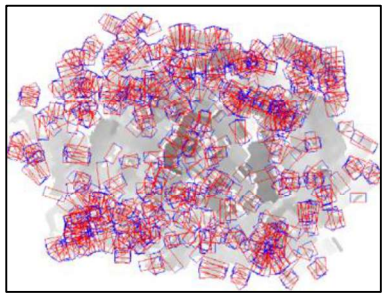
- Required parameters for grasp of a parallel-jaw gripper:
  - Intrinsic gripper finger size ***d***.
  - Position of gripper center ***x, y, z***.
  - Open width of the gripper ***w***.
  - In-plane rotation angle $\boldsymbol{\theta}$.

4.2 Algorithm pipeline

Real scene

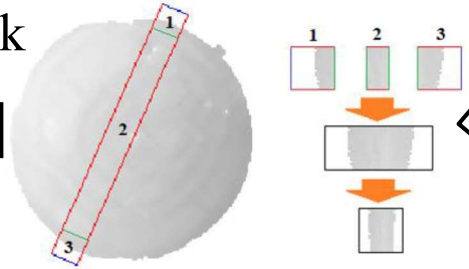


Depth image



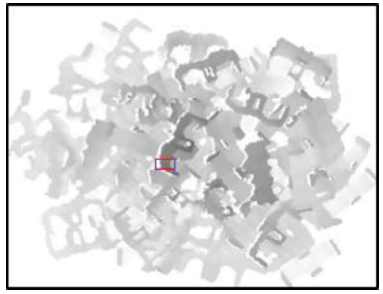
Grasp candidates

Pre-process



Processed grasps

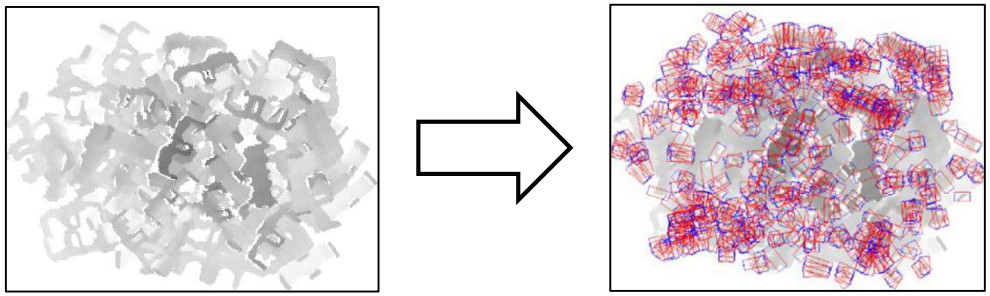
Network



Best grasp

Compute parameters for grasp execution

4.3 Generation of grasp candidates

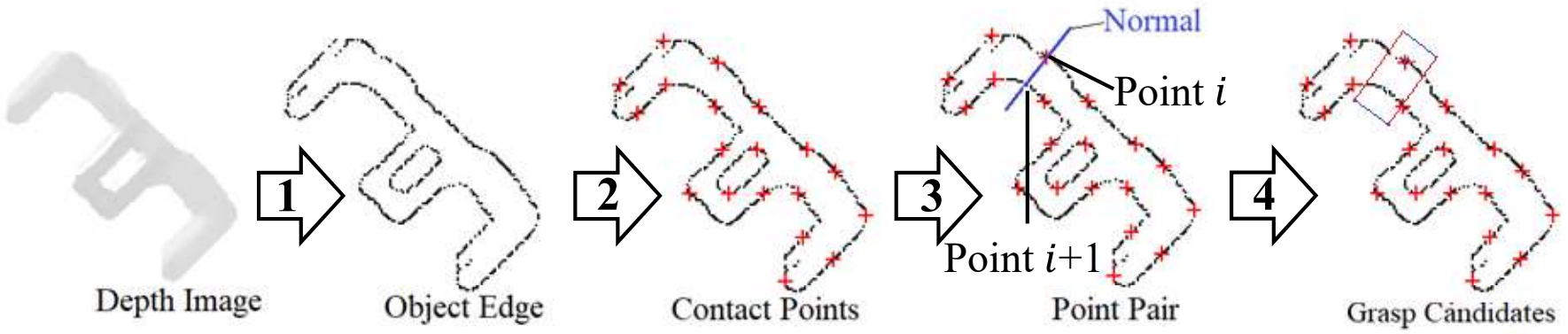


Depth image

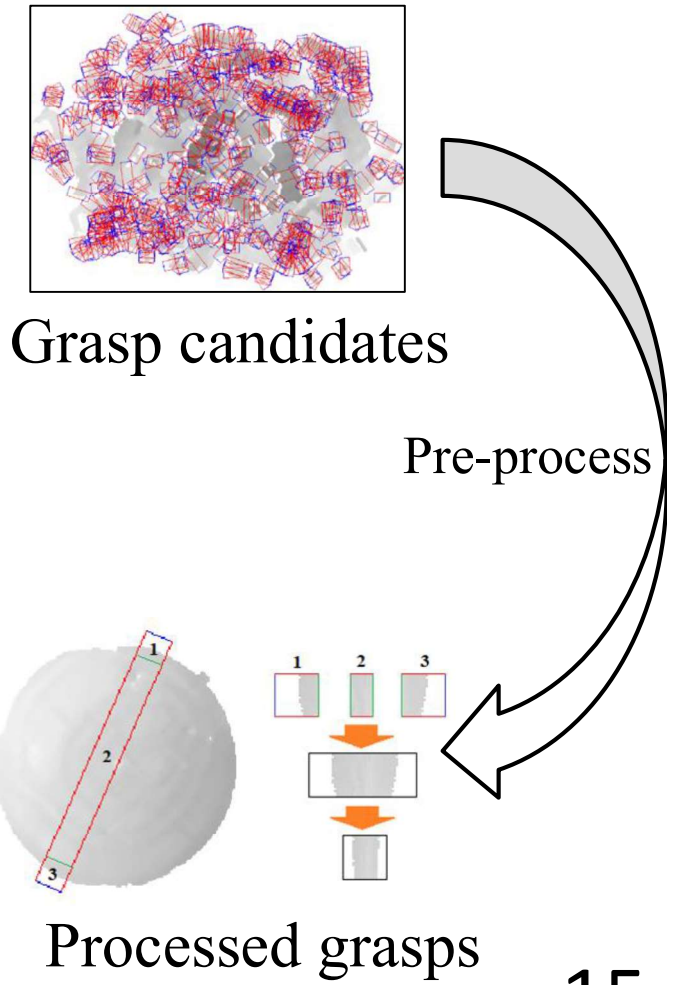
Grasp candidates

4.3 Generation of grasp candidates

- Generate grasps by contact points
  1. Obtain object edge from depth image.
  2. Find the contact points on object edges.
  3. For every contact point $\boldsymbol{i}$, find the normal and the intersection of normal and edge as points $\boldsymbol{i+1}$.
  4. Generate a grasp candidate using two contact points.

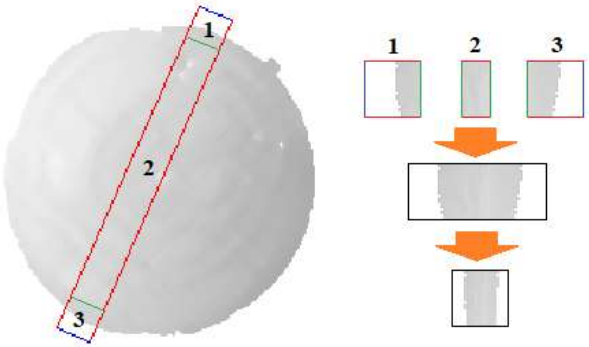


4.4 Pre-processing of grasp rectangle

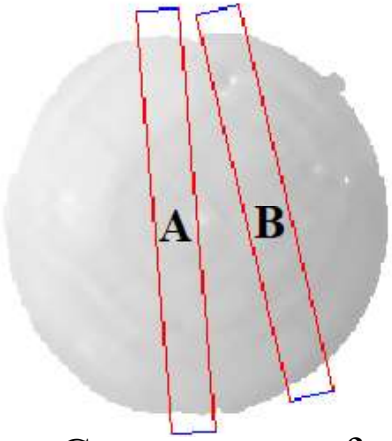


4.4 Pre-processing of grasp rectangle

- Process the input image to the same size.
  1. Separate the rectangle into 3 parts.
  2. Keep the size of part 1&3, and compress part 2
  3. Combine 3 parts and resize to the target size.



- Compared with simply resizing, the proposed method provides appropriate grasp rectangle and scores.

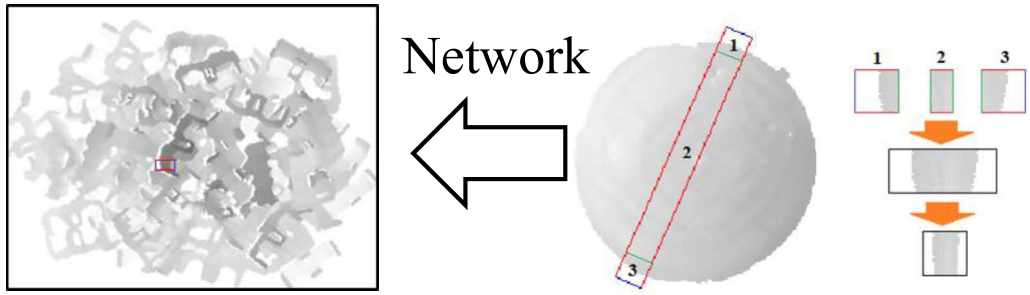


Grasp Scores Using Different Data Processing Methods

Grasp No.	A	B
Simply Resizing	7.975	8.709
Our Method	8.079	5.066

Grasp scores of simply resizing and proposed method

4.5 Network structure

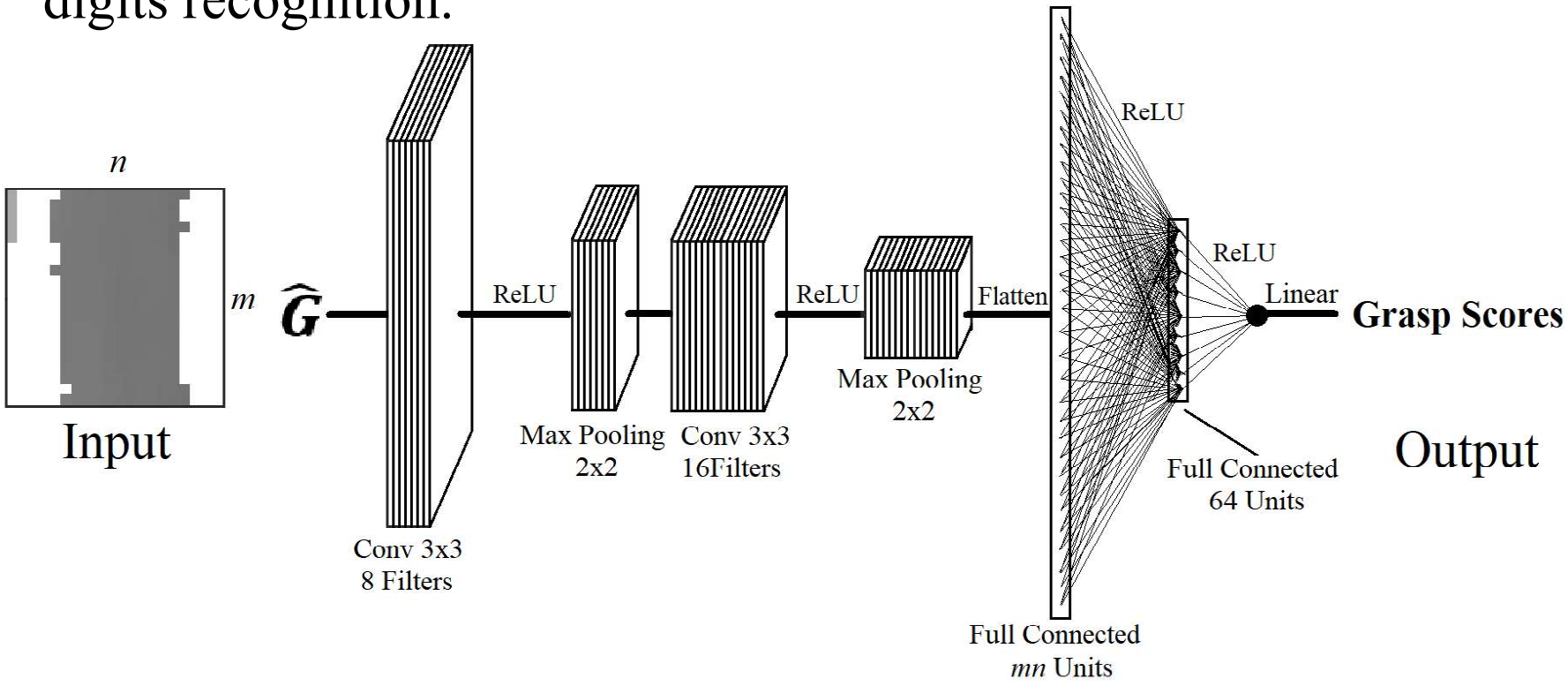


Best grasp

Processed grasps

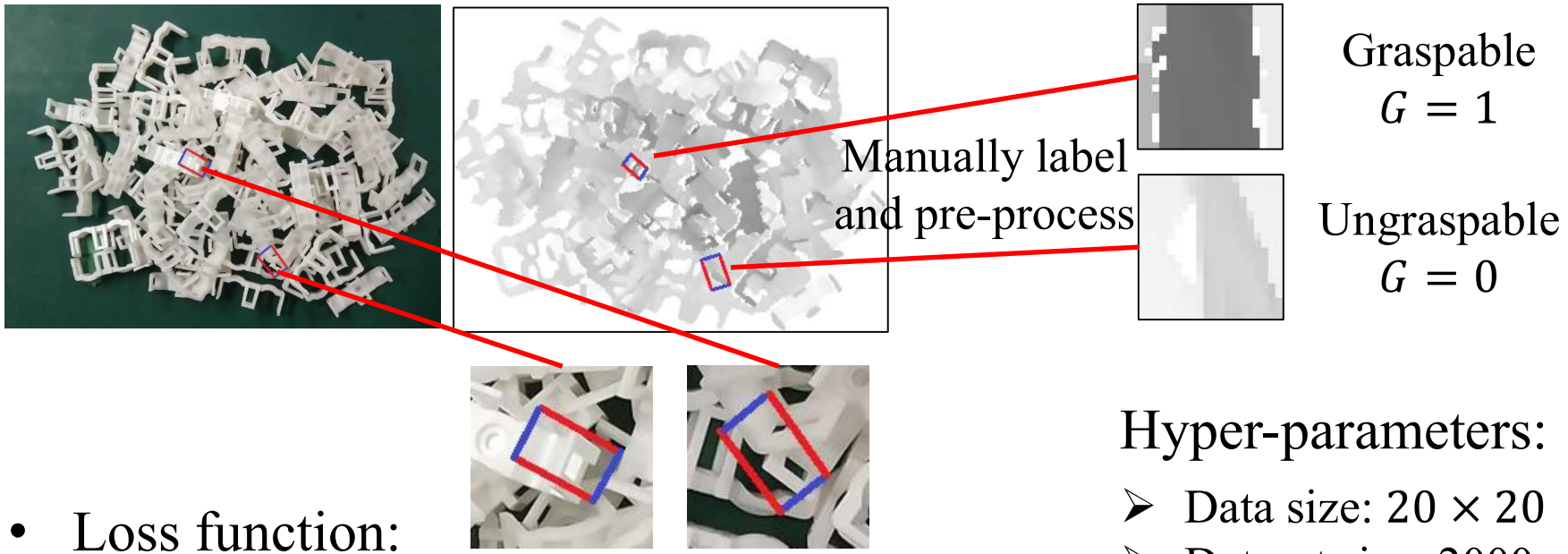
4.5 Network structure

- The structure is similar with LeNet-5 [8], we change the number of filters and layers, and use only 400 perceptrons. LeNet-5 is a very classical convolutional neural network used for handwritten digits recognition.



4.6 Training process

- For every type of object, prepare 2000 training data including 1000 graspable rectangles and 1000 ungraspable rectangles.



- Loss function:

$$L = -\frac{1}{N}\sum_{1}^{N} Gln\sigma(\hat{y}) + (1-G)ln\big(1-\sigma(\hat{y})\big)$$

$\hat{y}$ is predicted scores by network, $\sigma(\hat{y}) = \frac{1}{1+e^{-\hat{y}}}$ [12]

Hyper-parameters:

- Data size: 20×20
- Dataset size: 2000
- Batch size: 32
- Dropout: 0.5
- Optimizer: RMSProp
- Learning rate: 0.001