siléane

CV ML for ROBOTICS

Instance Segmentation for Robotic Bin Picking

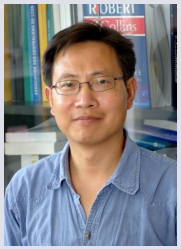
ÉCOLE CENTRALE LYON

LIRIS

Dr.Matthieu Grard,
Dr.Emmanuel Dellandréa,
Prof. Liming Chen

Imagine ECL

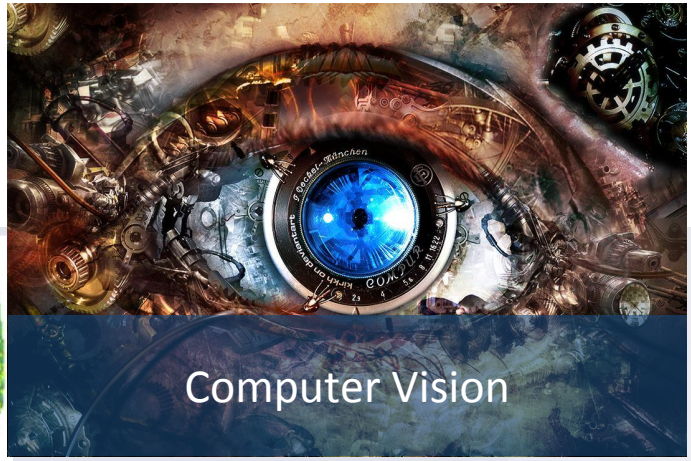
Laboratoire d'InfoRmatique en Image et Systèmes d'information



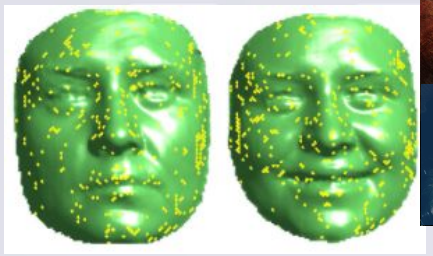
Outline

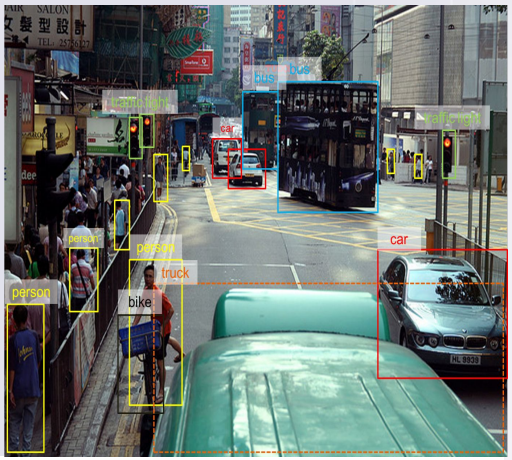
- **Human manipulation skills and robotic grasping**
- **State of the art**
- **Spatial Layout Aware Object instance segmentation**
- **Futur Work**

CV&ML@my group

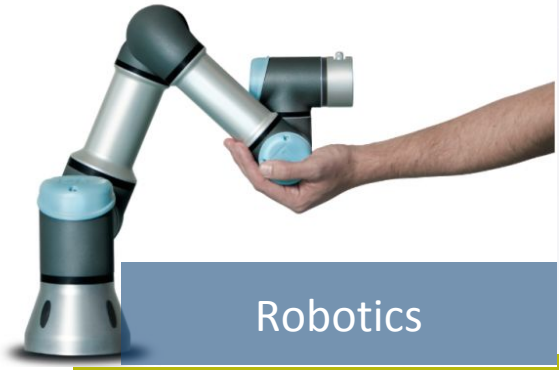


- **Computer vision**
  - Object recognition and detection
  - Face Biometrics in 2D and 3D
  - Affective Computing





- **Machine learning**
  - Deep Structured learning
  - transfer learning
  - Domain adaptation
- **Robotics**
  - Grasping
  - Human robot interaction and collaboration



Human manipulation skills and Robotic Bin Picking...

Human manipulation skills...

a dexterity requiring intelligence and vision

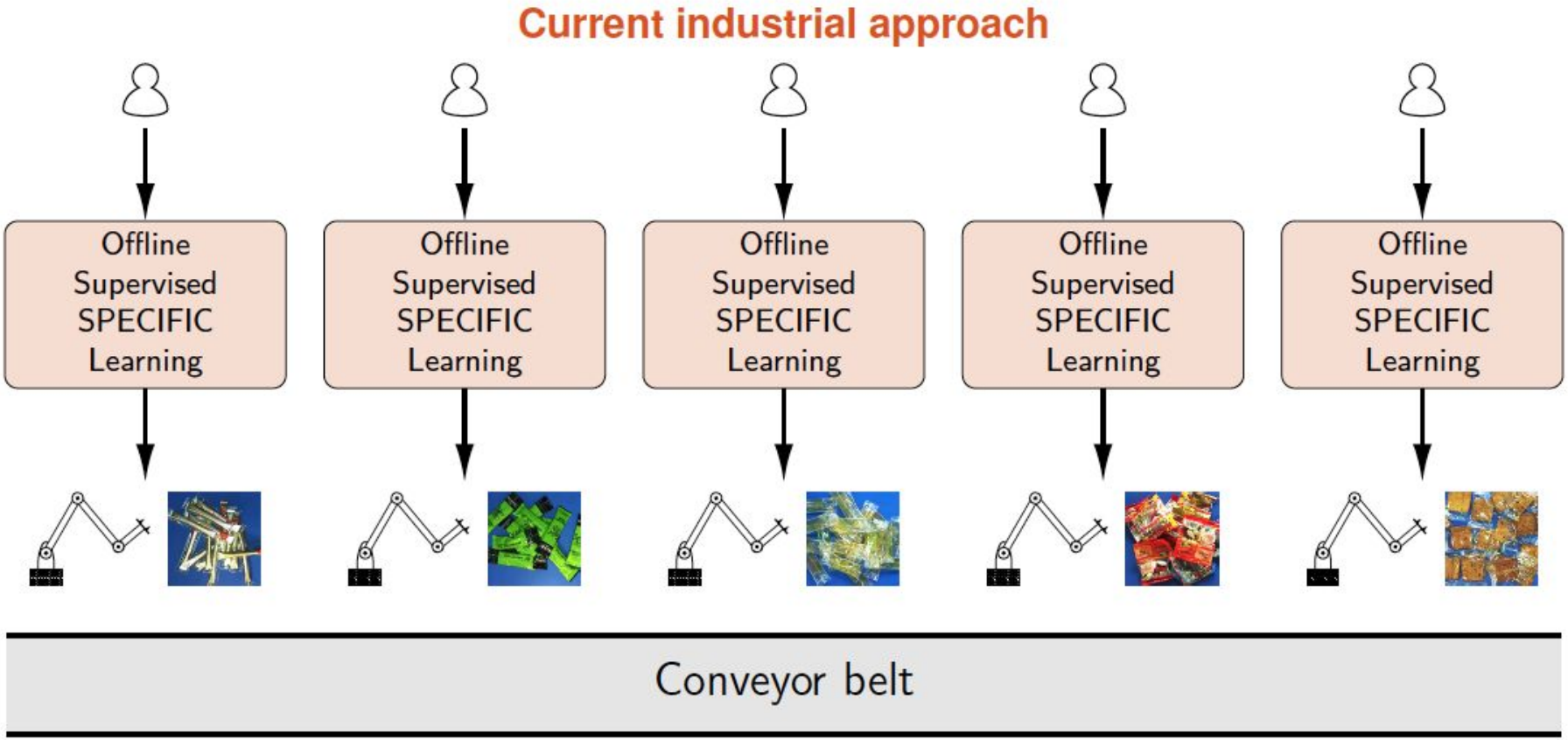
Robotic Bin Picking...

- Kamido by Siléane

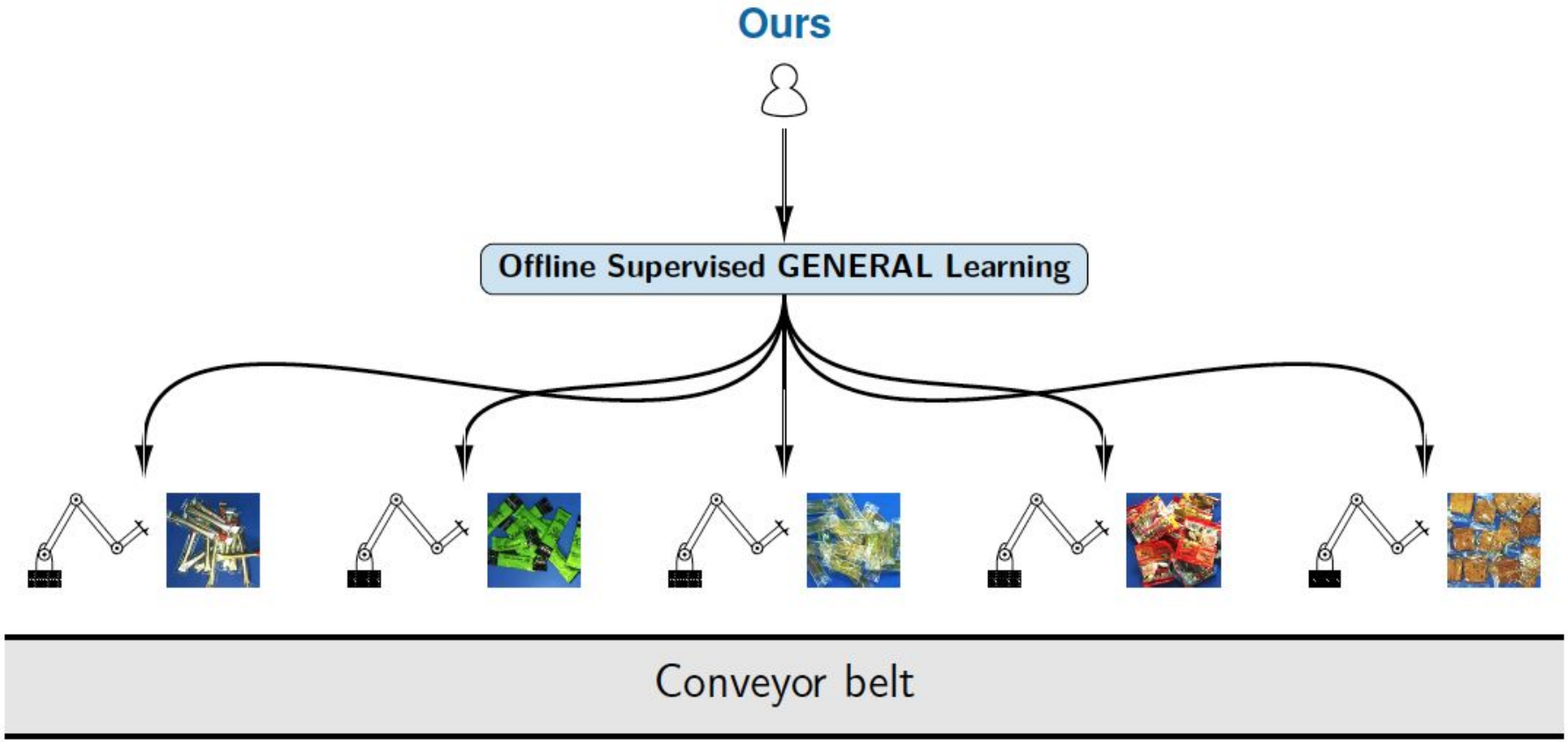
Picking and kitting...

Logistics

Current Industrial Approach



Aim



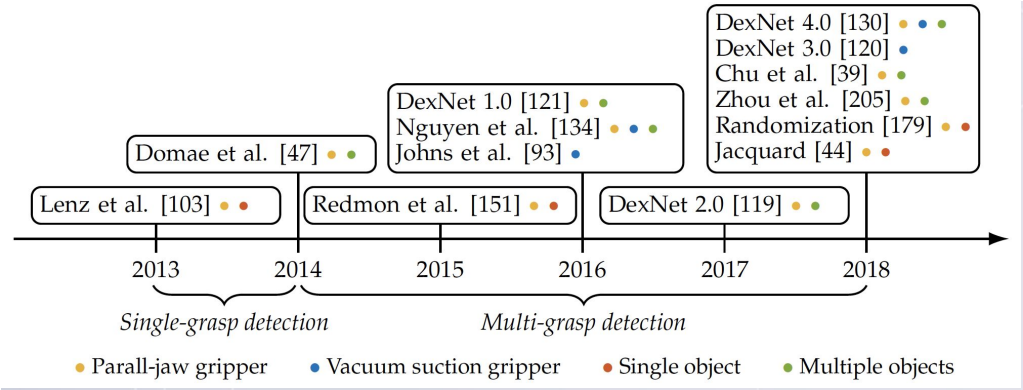
Challenges

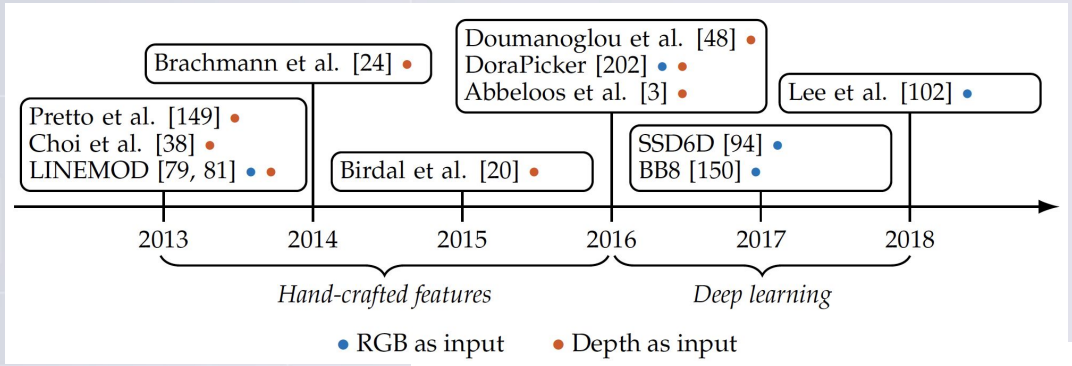
Diversity of objects

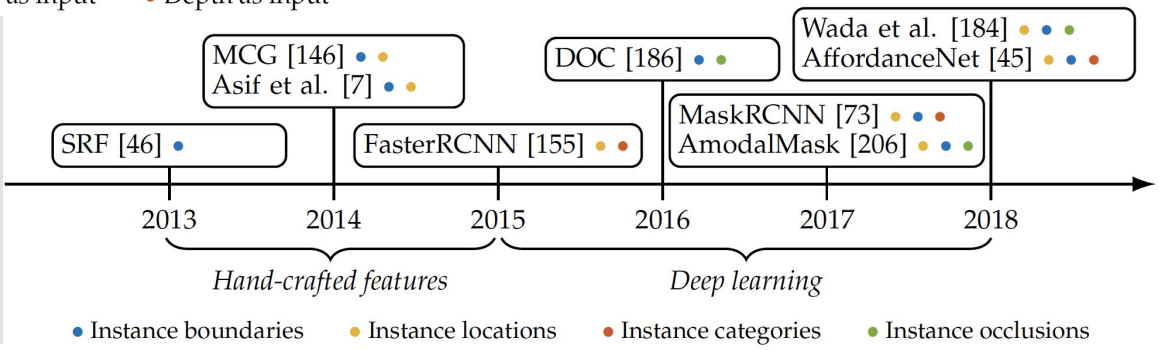
Intra-class variations

State of the art

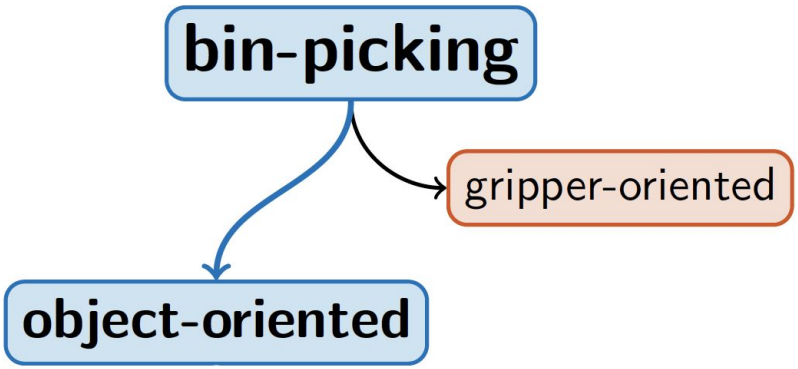
Many works...


DexNet 4.0 [130]
DexNet 3.0 [120]
Chu et al. [39]
Zhou et al. [205]
Randomization [179]
Jacquard [44]
DexNet 1.0 [121]
Nguyen et al. [134]
Johns et al. [93]
Domae et al. [47]
Lenz et al. [103]
Redmon et al. [151]
DexNet 2.0 [119]
2013
2014
2015
2016
2017
2018
Single-grasp detection
Multi-grasp detection
Parall-jaw gripper
Vacuum suction gripper
Single object
Multiple objects

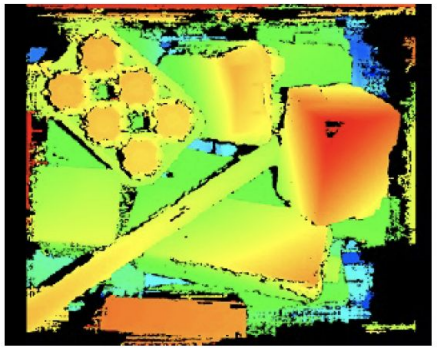

Doumanoglou et al. [48]
DoraPicker [202]
Abbeloos et al. [3]
Brachmann et al. [24]
Lee et al. [102]
Pretto et al. [149]
Choi et al. [38]
LINEMOD [79, 81]
SSD6D [94]
BB8 [150]
Birdal et al. [20]
2013
2014
2015
2016
2017
2018
Hand-crafted features
Deep learning
RGB as input
Depth as input


Wada et al. [184]
AffordanceNet [45]
MCG [146]
Asif et al. [7]
DOC [186]
MaskRCNN [73]
AmodalMask [206]
SRF [46]
FasterRCNN [155]
2013
2014
2015
2016
2017
2018
Hand-crafted features
Deep learning
Instance boundaries
Instance locations
Instance categories
Instance occlusions

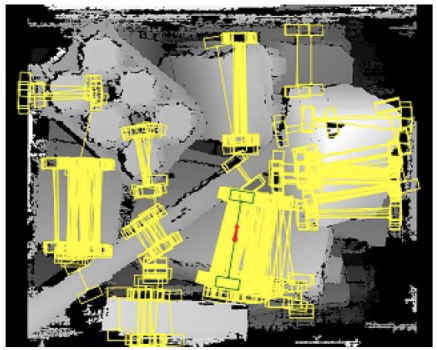
Gripper-Oriented Bin-Picking



Image



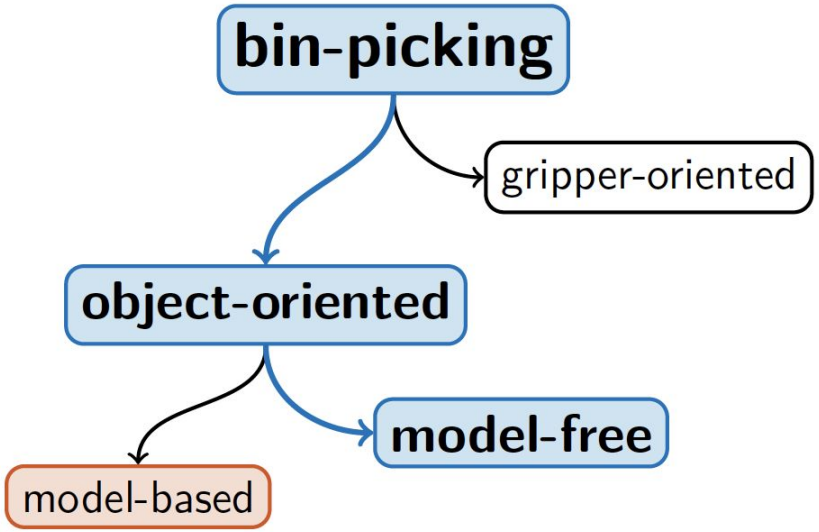
Depth (input)



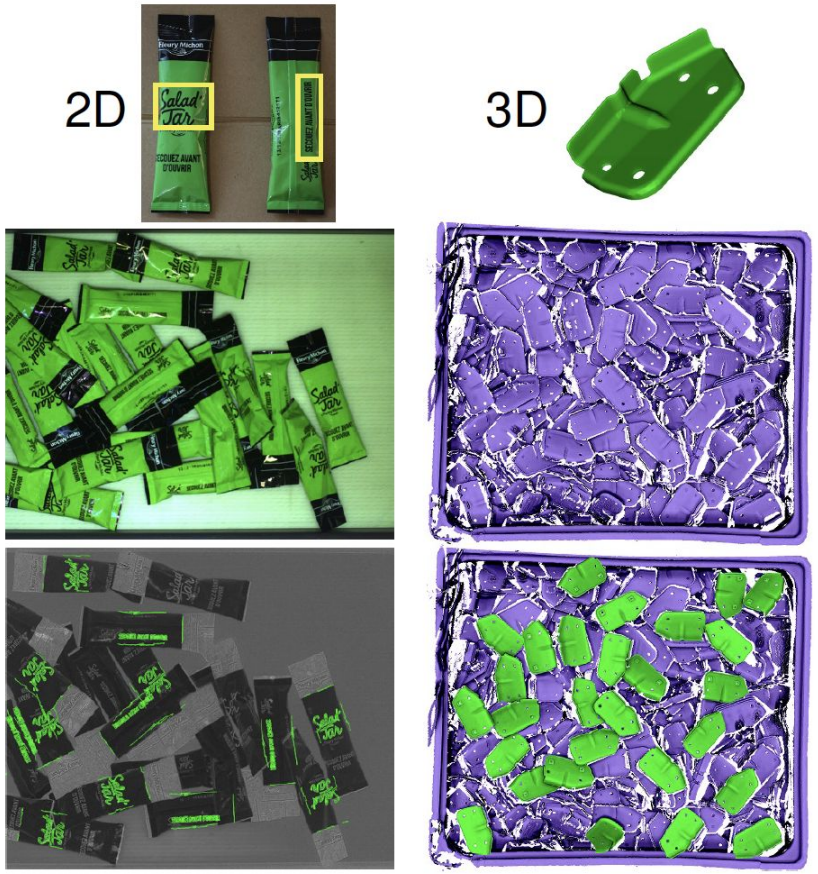
Gripper-oriented

😞no explicit notion of object instances

Model-Based Object-Oriented Bin-Picking

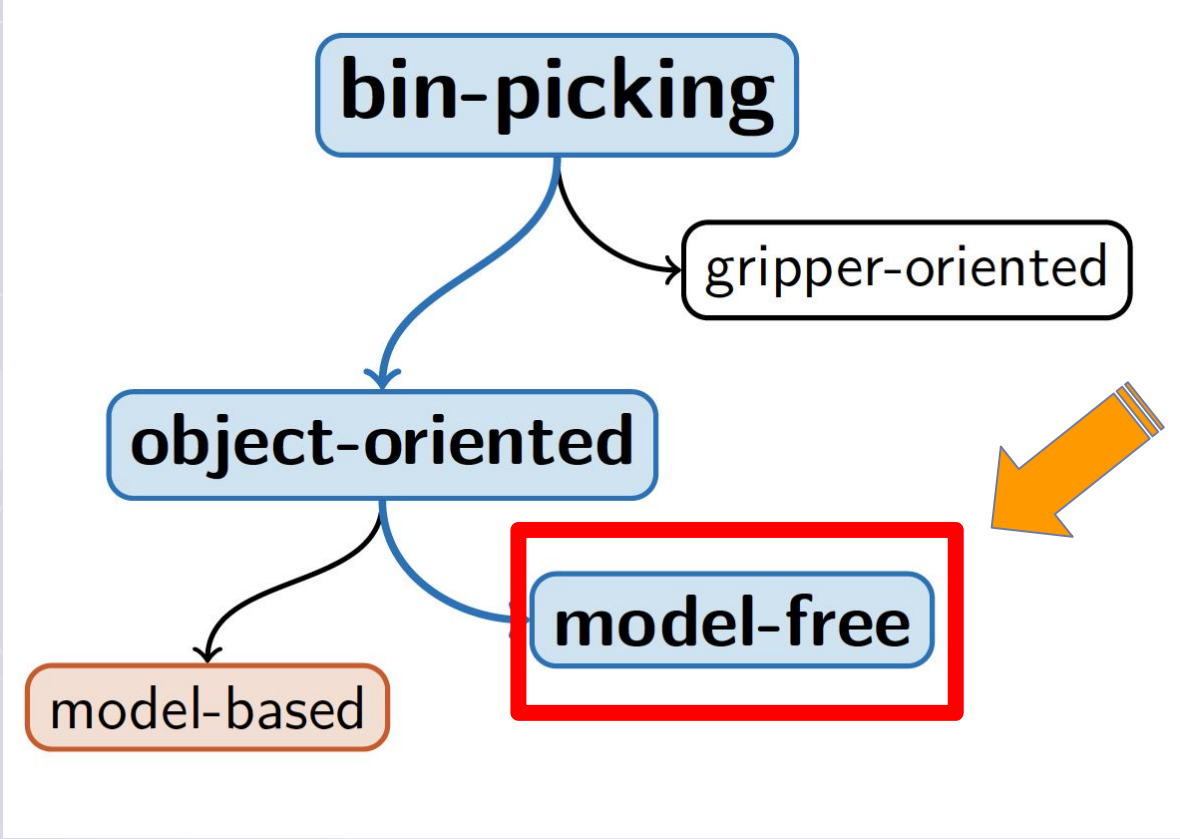


😞 explicit object model required



Our Approach

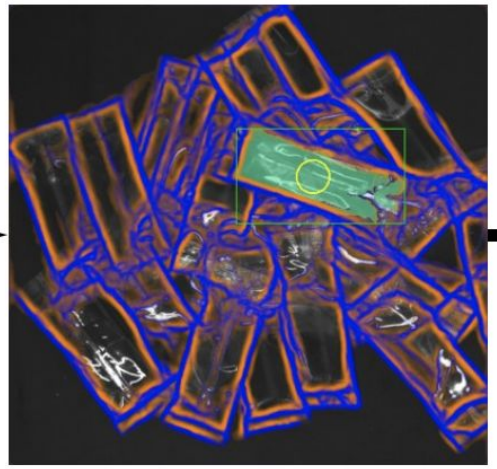
Model-Free Object Instance-Oriented Bin-Picking



Our Approach

Model-Free Object Instance-Oriented Bin-Picking

Acquisition



Detection

Grasping

Object-Oriented

⇒ Segmenting Object Instances and Spatial Layout

Object model-Free

⇒ learning only from examples

Depth or RGB as INPUT ?

MotionCam 3D, Photoneo

RGB for homogeneous piles

Scene Reconstruction

- **Unfeasible** due to shiny or transparent products

Spatial Layout Aware Object Instance Segmentation

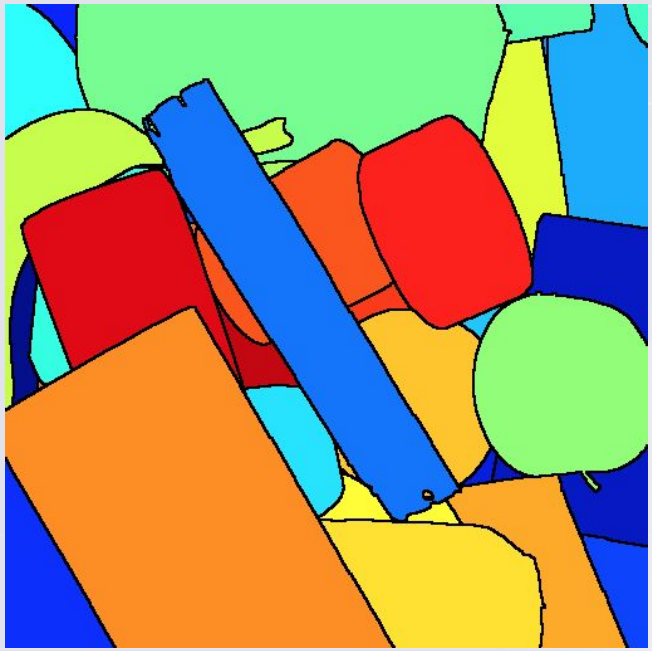
ECL 2014
PhD ECL 2019

Layout Aware Object Instance Segmentation

INPUT

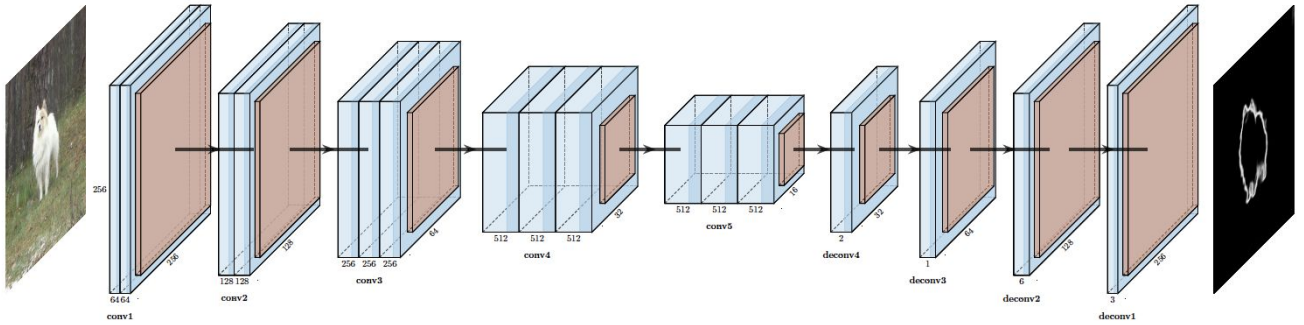
OUTPUT

f () =



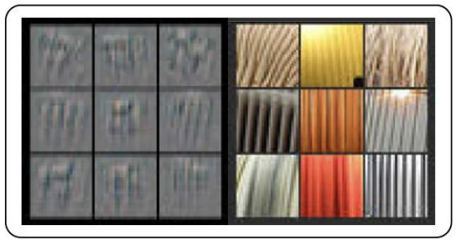
Deeply learning f...

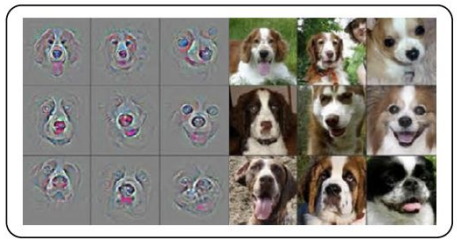
Which learnable function ?



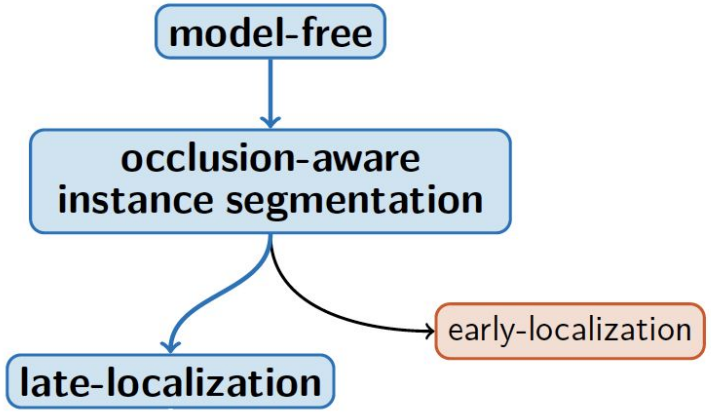
FCNs learn hierarchical representations (Zeiler et al., ECCV 2014).

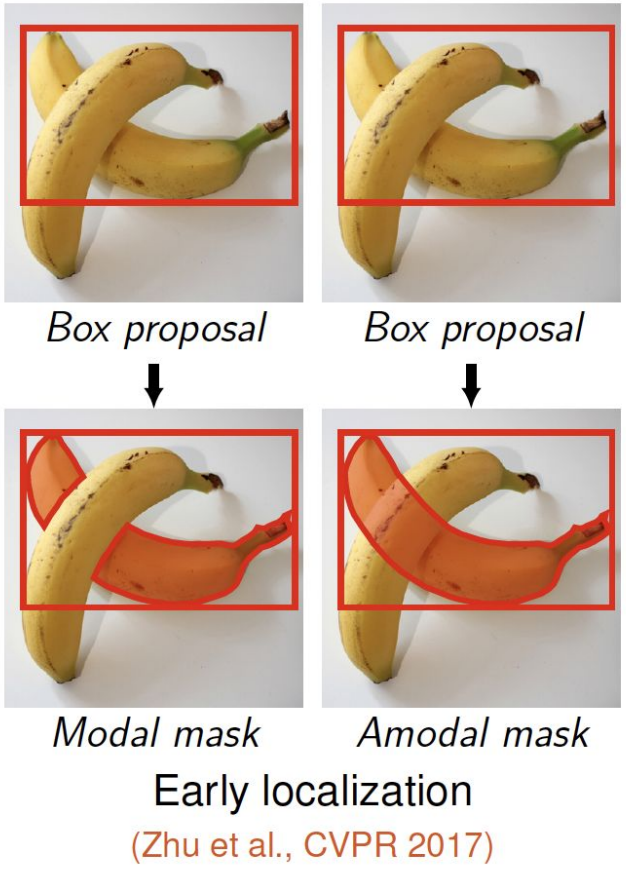
Low-level general features → High-level task-specific features





Early-Localization Instance Segmentation





✖ invalid rectangle approximation

Early-Localization Instance Segmentation

A bounding box may contain several instances.

Urban scenes

Bin-picking scenes

⇒ Similar patterns may be classified differently.

Input

Which binary segmentation?

Early localization

Early-Localization Instance Segmentation

A bounding box may contain several instances.

Urban scenes | Bin-picking scenes

⇒ **Similar patterns may be classified differently.**

Input

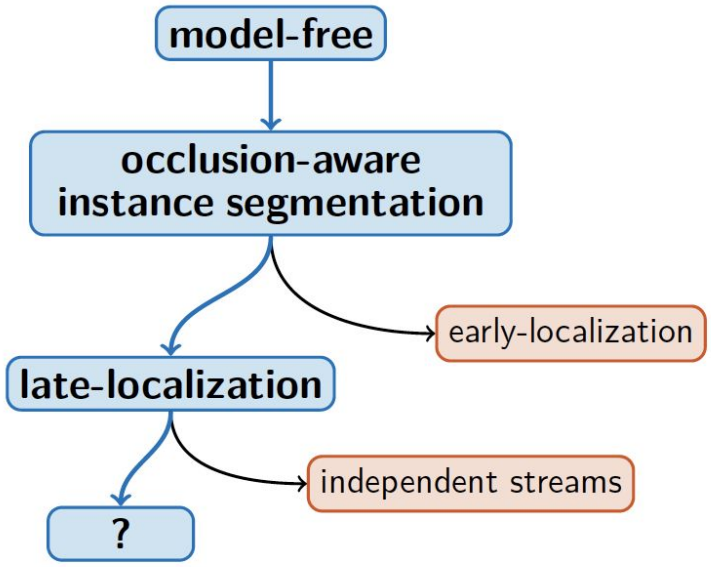
Which binary segmentation?

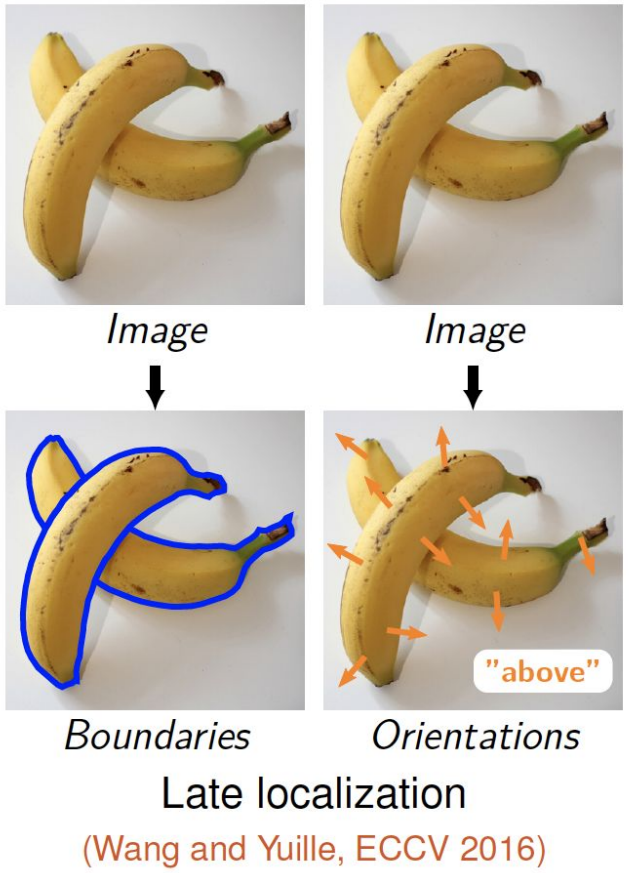
Early localization

Boundaries

Late localization

Late-Localization Instance Segmentation

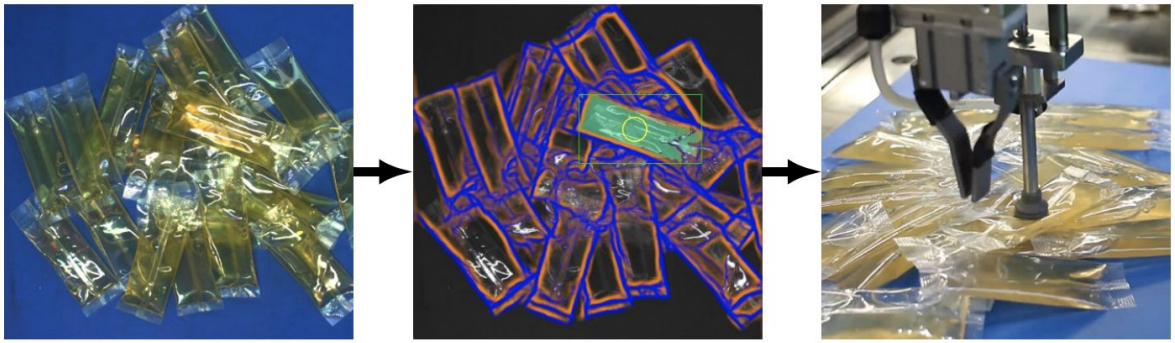




✖ Instance boundaries and occlusions are learned separately.

The Proposed Approach

Generic occlusion-aware instance segmentation
using a synthetically trained *multicameral* network



Acquisition *Detection* *Grasping*

Simulation

trains

Synthetic training images

Thee Proposed Architecture Design

Encoder/Decorder with Multi-Cameral Decoding Process

We propose to **decompose the decoding process** into 4 consecutive sub-tasks:

1. generic instance boundary detection;

img

bds

Thee Proposed Architecture Design

Encoder/Decorder with Multi-Cameral Decoding Process

We propose to **decompose the decoding process** into 4 consecutive sub-tasks:

1. generic instance boundary detection;
2. occluding boundary side detection;

img bds occ

Thee Proposed Architecture Design

Encoder/Decorder with Multi-Cameral Decoding Process

We propose to **decompose the decoding process** into 4 consecutive sub-tasks:

1. generic instance boundary detection;
2. occluding boundary side detection;
3. unoccluded instance segmentation;

img bds occ seg

Thee Proposed Architecture Design

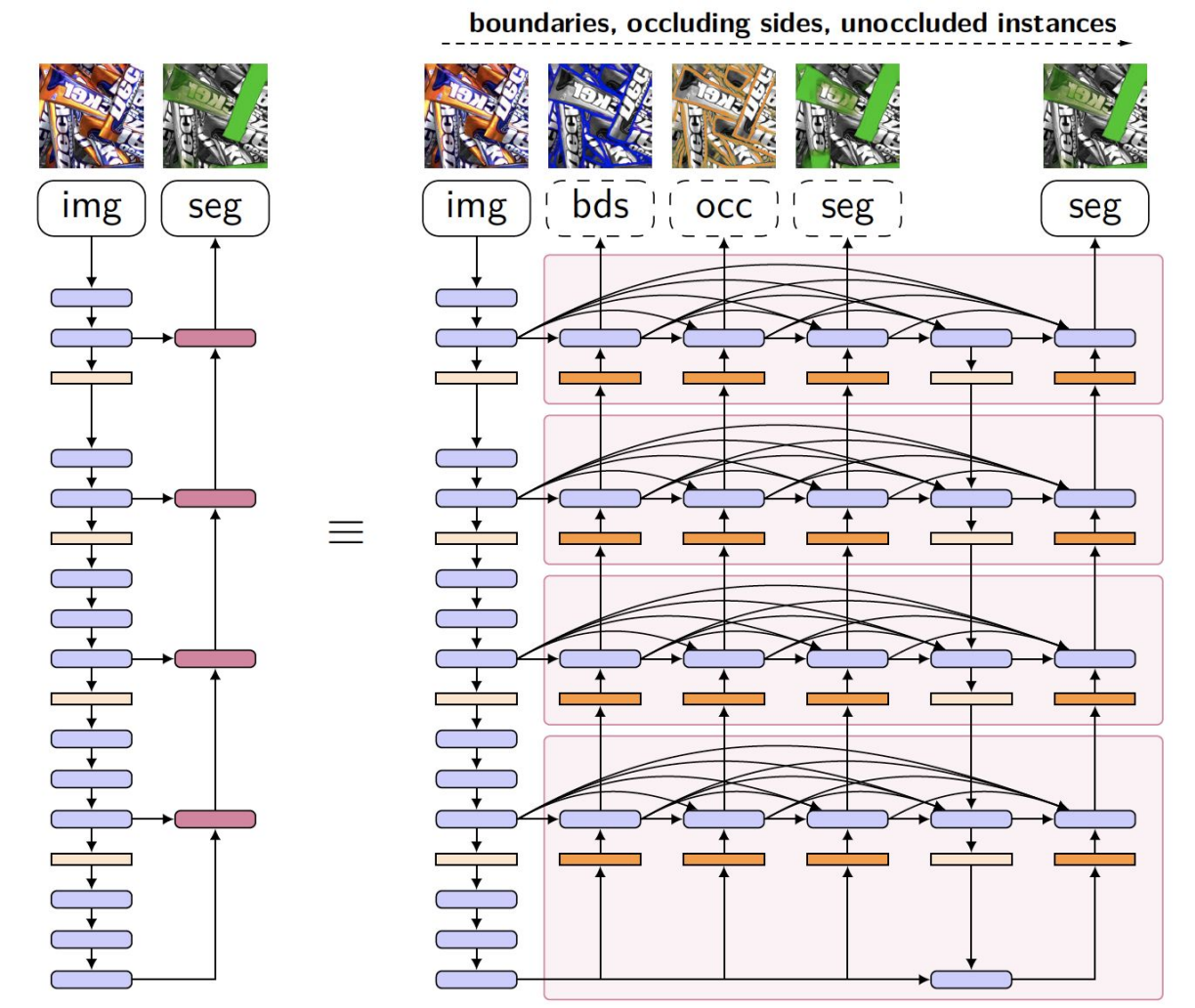
Encoder/Decorder with Multi-Cameral Decoding Process

We propose to **decompose the decoding process** into 4 consecutive sub-tasks:

1. generic instance boundary detection;
2. occluding boundary side detection;
3. unoccluded instance segmentation;
4. segmentation refinement.

img bds occ seg seg

The Proposed Architecture Design



Collecting training data...

Stereoscopic system

blender

Simulation & GPU rendering

100 images / ~8h

...leading to the Mikado Dataset

Mikado dataset

Object contour and occlusion jointly annotated !

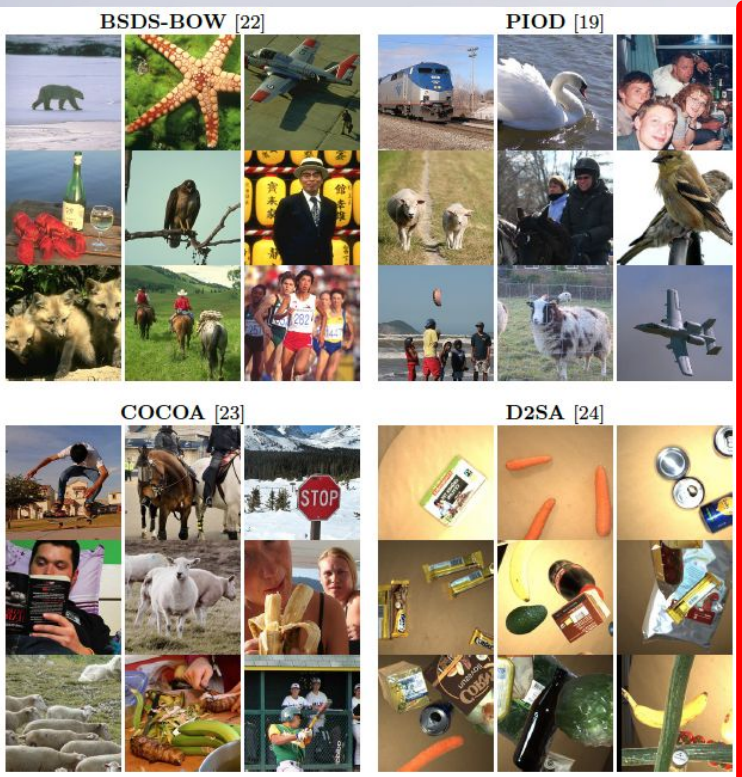
INPUT

OUTPUT

Mikado dataset vs state-of-the-art

Dataset	Average image size	Number of images	Number of instances	Instances per image	Inter-instance occlusions per image	Background pixels per image	Ground-truth annotations
BSDS-BOW[1] [22]	432×369	200	–	–	–	–	Human-made
PIOD [19]	469×386	10,100	24,797	2.5	1.3	69%	
COCOA[2] [23]	578×483	3,823	34,884	9.1	1.3	33%	
D2SA[2] [24]	1962×1569	5,600	28,703	5.1	2.8	79%	
Mikado (Ours)	640×512	2,400	48,184	**20.1**	**52.9**	**24%**	**Computer-generated**
Mikado+[3] (Ours)	640×512	14,560	459,002	**31.5**	**60.5**	**24%**	

Mikado dataset vs state-of-the-art



Dataset	Average image size	Number of images	Number of instances	Instances per image	Inter-instance occlusions per image	Background pixels per image	Ground-truth annotations
BSDS-BOW[1] [22]	432×369	200	–	–	–	–	Human-made
PIOD [19]	469×386	10,100	24,797	2.5	1.3	69%	
COCOA[2] [23]	578×483	3,823	34,884	9.1	1.3	33%	
D2SA[2] [24]	1962×1569	5,600	28,703	5.1	2.8	79%	
Mikado (Ours)	640×512	2,400	48,184	**20.1**	**52.9**	**24%**	**Computer-generated**
Mikado+[3] (Ours)	640×512	14,560	459,002	**31.5**	**60.5**	**24%**	

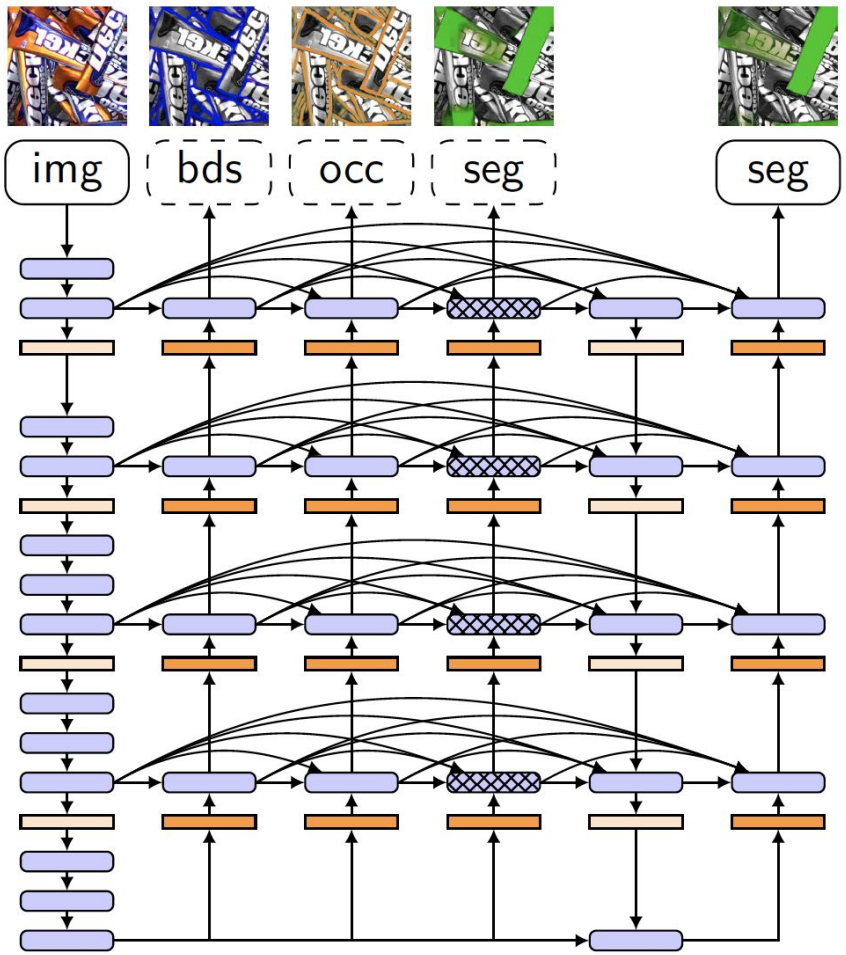
Mikado dataset vs state-of-the-art

Dataset	Average image size	Number of images	Number of instances	Instances per image	Inter… per image	Background pixels per image	Ground-truth annotations
BSDS-BOW[1] [22]	432×369	200	–	–	–	–	Human-made
PIOD [19]	469×386	10,100	24,797	2.3	1.3	69%	
COCOA[2] [23]	578×483	3,823	34,884	9.1	1.3	33%	
D2SA[2] [24]	1962×1569	5,600	28,572	5.1	2.8	79%	
Mikado (Ours)	640×512	2,400	48,184	**20.1**	**52.9**	**24%**	**Computer-generated**
Mikado+[3] (Ours)	640×512	14,560	459,002	**31.5**	**60.5**	**24%**	

14 500 images

459 002 instances

A performance-enhancing network design

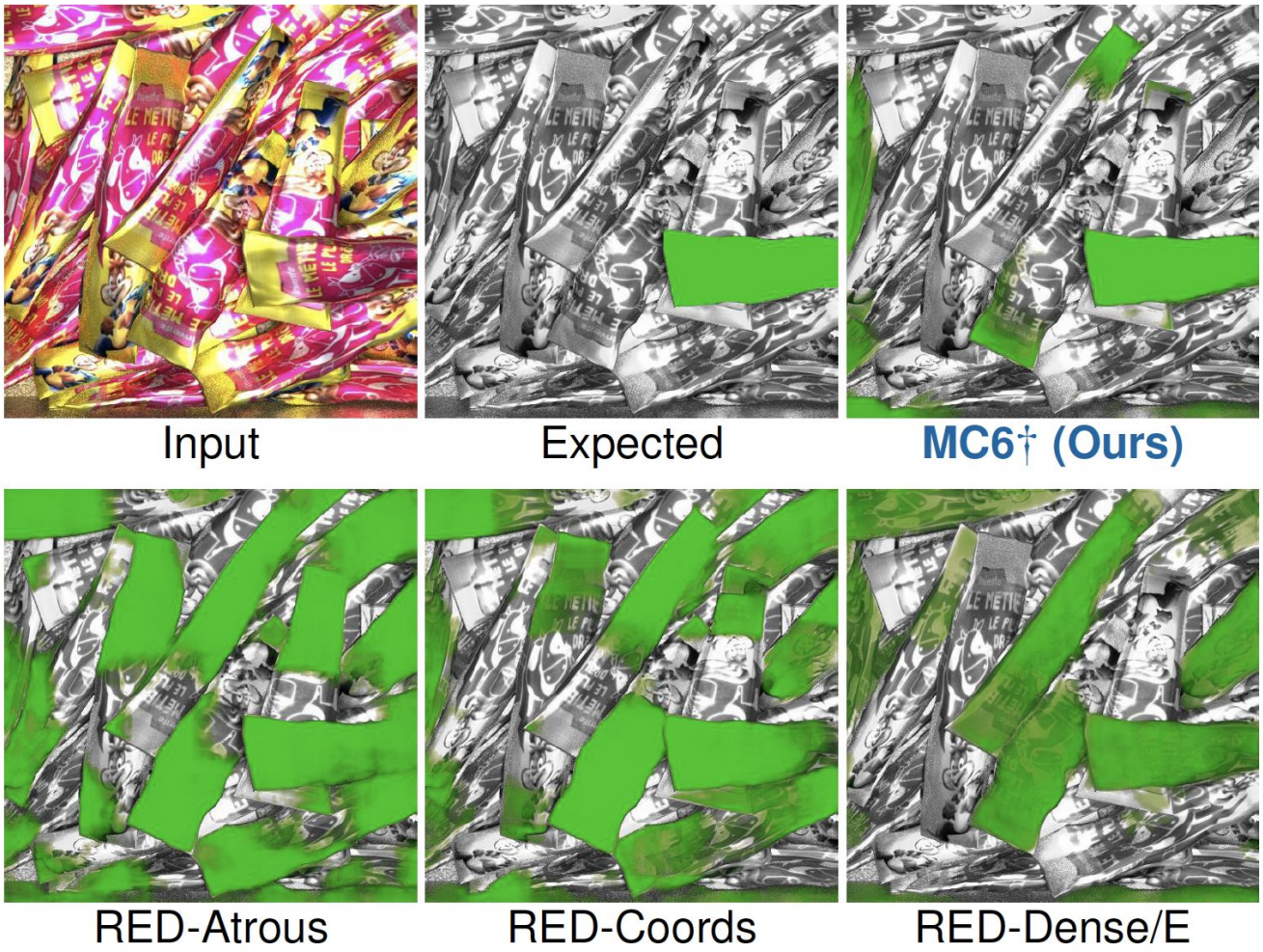


Mikado

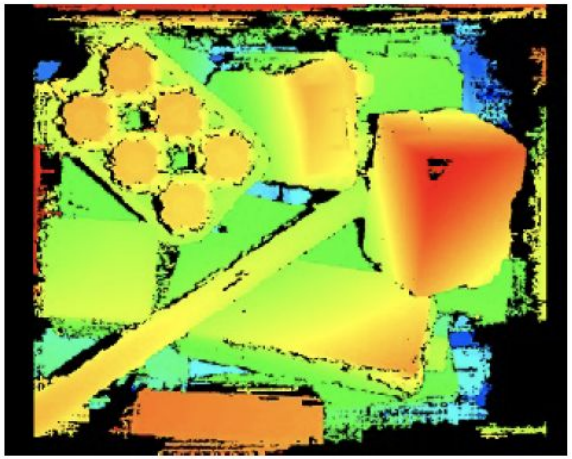
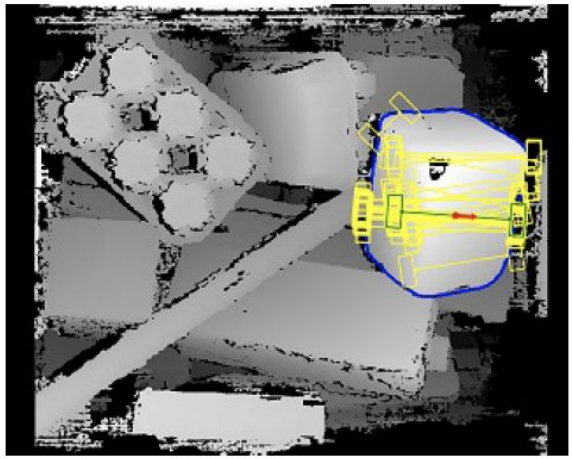
Average Precision	seg
MC6†-X/D4	.837
MC6†	.825
MC4†	.802
MC4	.762
MC3	.750
RED(=MC2)	.732

Comparative Results

Comparative Resultts



Future Work

- **Joint Object Instance Segmentation and Grasp Detection**
- **Sim-2-Real Domain Adaptation**

Future work….

Mikado Dataset

A Synthetic Dataset of Dense Homogeneous Object Layouts
for Occlusion-aware Instance Segmentation

Home Dataset Download Contact

Collaborators

- Matthieu Grard, Siléane & Ecole Centrale de Lyon, LIRIS, France
- Emmanuel Dellandréa, Ecole Centrale de Lyon, LIRIS, France
- Liming Chen, Ecole Centrale de Lyon, LIRIS, France

Citation

If you use the Mikado dataset in your research, please cite the following paper:

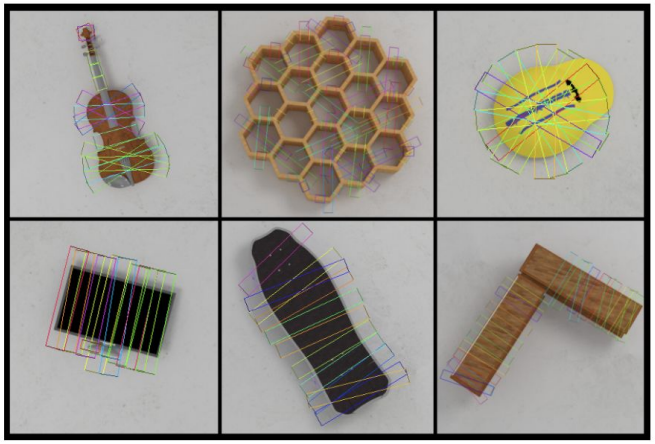
- M. Grard, E. Dellandrea, and L. Chen, "Deep Multicameral Decoding for Localizing Unoccluded Object Instances from a Single RGB Image" in *International Journal of Computer Vision (IJCV)*, 2020. DOI: https://doi.org/10.1007/s11263-020-01323-0

https://mikado.liris.cnrs.fr/

JACQUARD DATASET

A Large-Scale Dataset for Robotic Grasp Detection

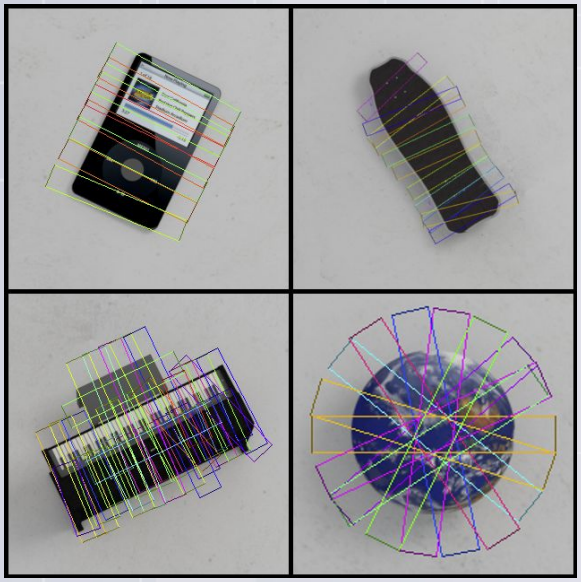
Home Database Contact Download Testing



ECL 2017

https://jacquard.liris.cnrs.fr/

Jacquard vs Cornell

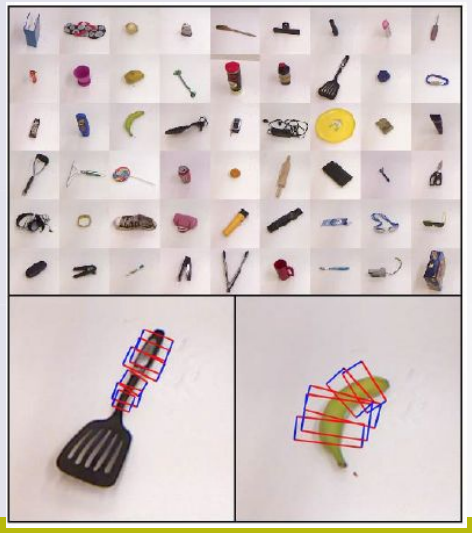


- 11k objects from ShapeNet
- 50k images
- >4 million grasp locations

Human manually labelled dataset

1035 images

from 280 objects



For further details

Matthieu Grard, Emmanuel Dellandréa, **Liming Chen**, "Deep Multicameral Decoding for Localizing Unoccluded Object Instances from a Single RGB Image", International Journal of Computer Vision (IJCV); Online 27 March 2020.

Read the preprint here: https://arxiv.org/abs/1906.07480, or here: https://hal.archives-ouvertes.fr/hal-02151828/

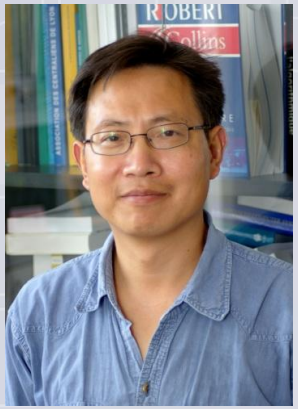
Matthieu Grard, Generic Instance Segmentation for Object-Oriented Bin-Picking, PhD thesis, Ecole Centrale de Lyon, May 2019

https://www.theses.fr/236948415

Follow our research on robotics here

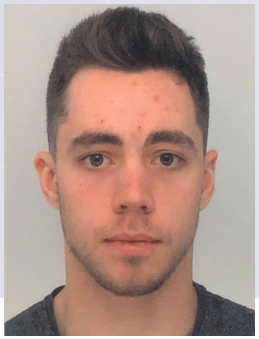
- https://ai4robot.liris.cnrs.fr/
- Tweeter: AI4Robotics ECL_Liris(@EclLiris)

People...







Thanks

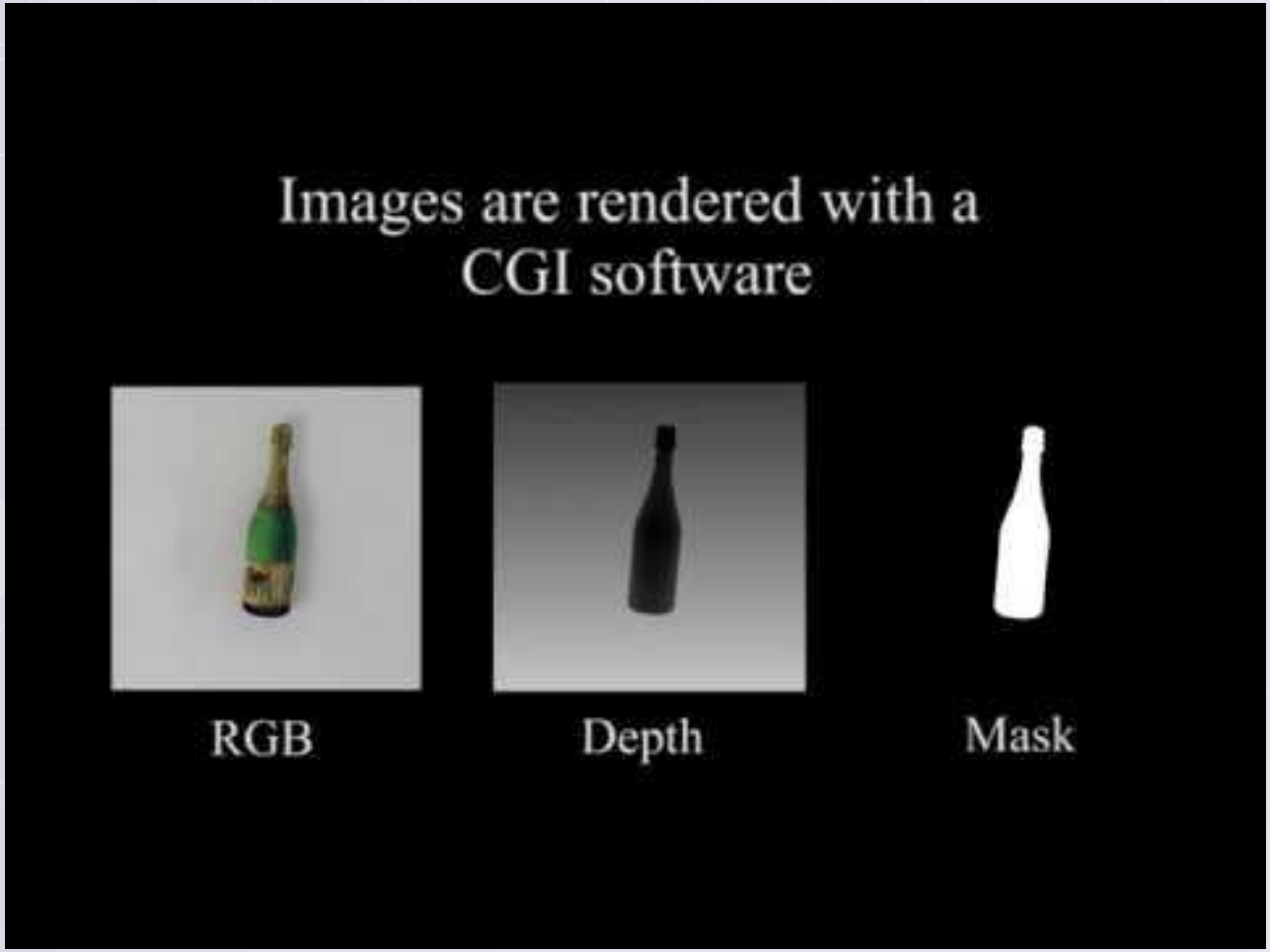
AGENCE NATIONALE DE LA RECHERCHE

L'Europe s'engage en Auvergne – Rhône-Alpes
http://www.europe-en-auvergnerhonealpes.eu/

FUI 21 PIKAFLEX

Labcom Arès

The Jacquard grasp dataset

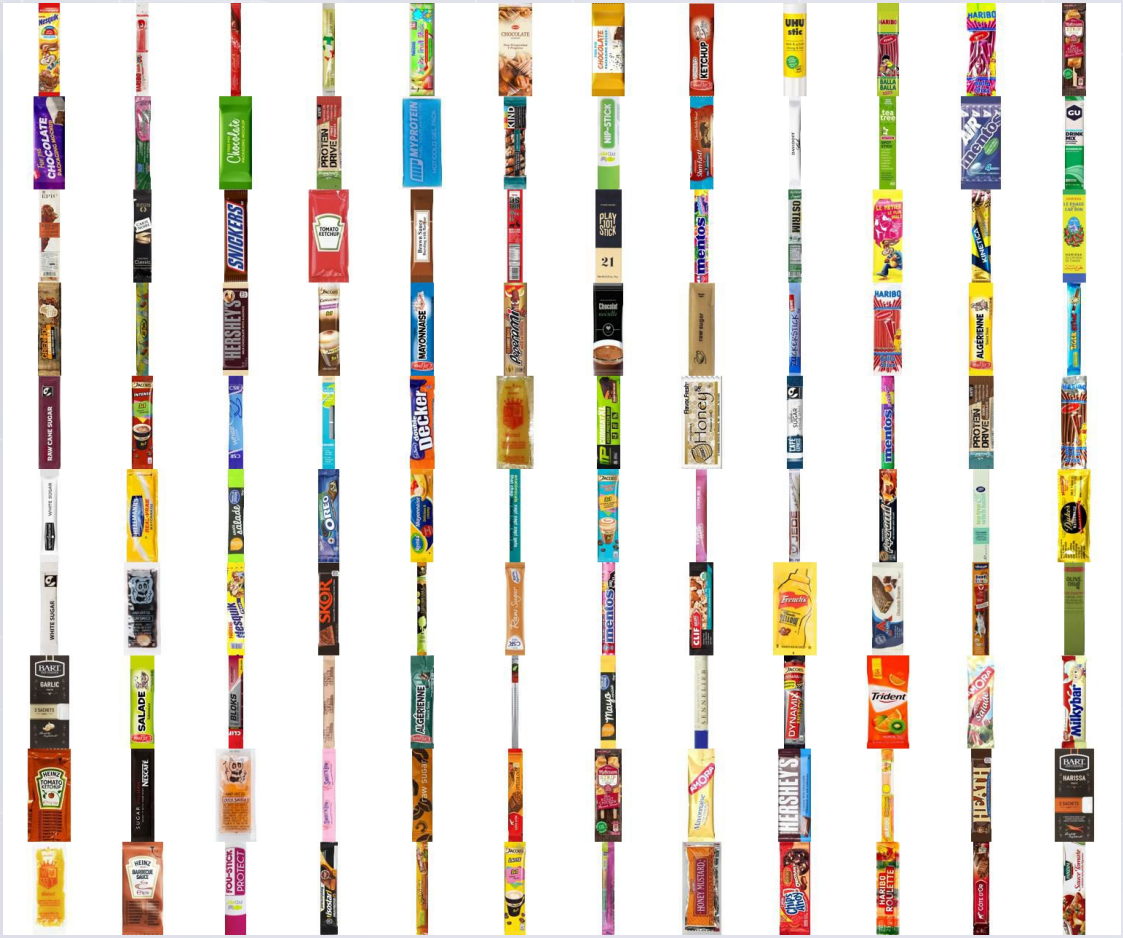


Thank you for your attention

Merci

谢
谢

Randomization (texture, background, shape)



Randomization (texture, background, shape)

Spatial Layout Object Instance Segmentation

INPUT

OUTPUT

f () =