

Synthetic Data for Robotics: Opportunities and Challenges

Maksims Ivanovs, researcher
Institute of Electronics and Computer Science (edi.lv)

16 November 2022

Intro: Robotics, Deep Learning, and Data

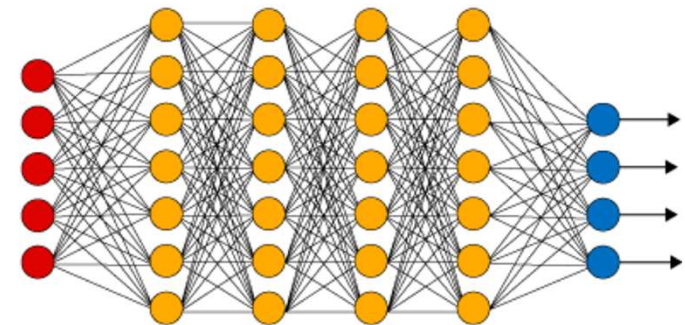
- Perception of the environment (e.g., with cameras) is an essential aspect of robotics
- State-of-the-art results on many perceptual tasks: with deep neural networks (DNNs)

- **Advantages of DNN:**

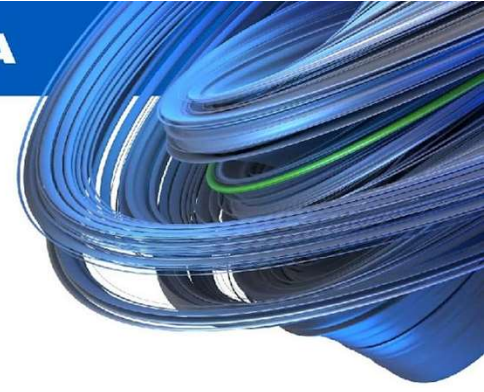
- high performance
- no need for feature engineering: learning from the data
- no need to retrain DNN from scratch: transfer learning

- **Shortcomings of DNN:**

- 'black boxes'
- need a lot of computational power (esp. GPU) for training
- need a lot of data for training



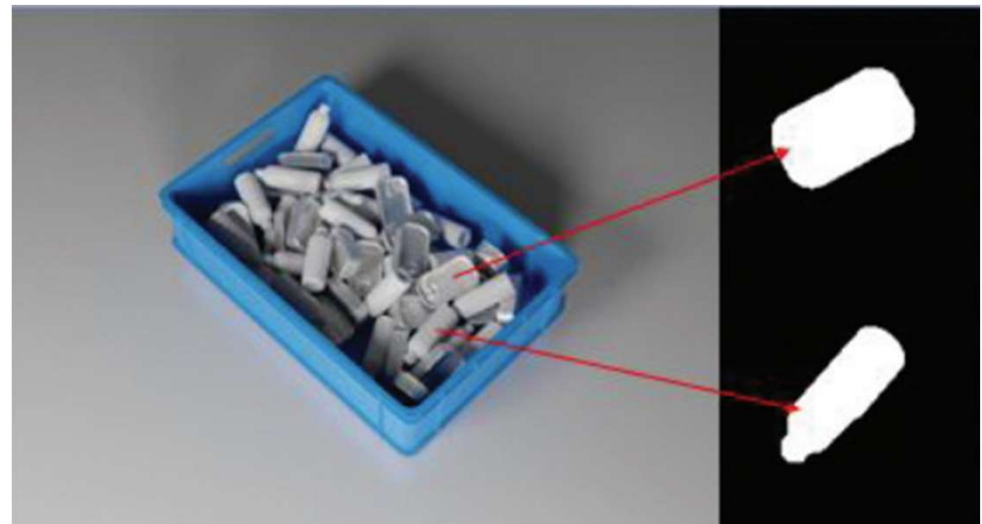
Sample robotics task: bin picking (1)



Sample robotics task: bin picking (2)

Computer vision tasks and labelling needed for them:

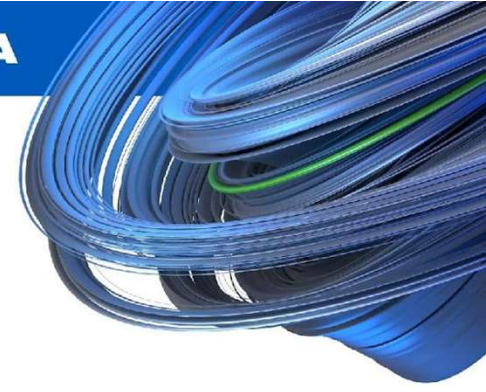
- object detection: bounding boxes
- semantic segmentation: segmentation masks



Getting data for training: some challenges

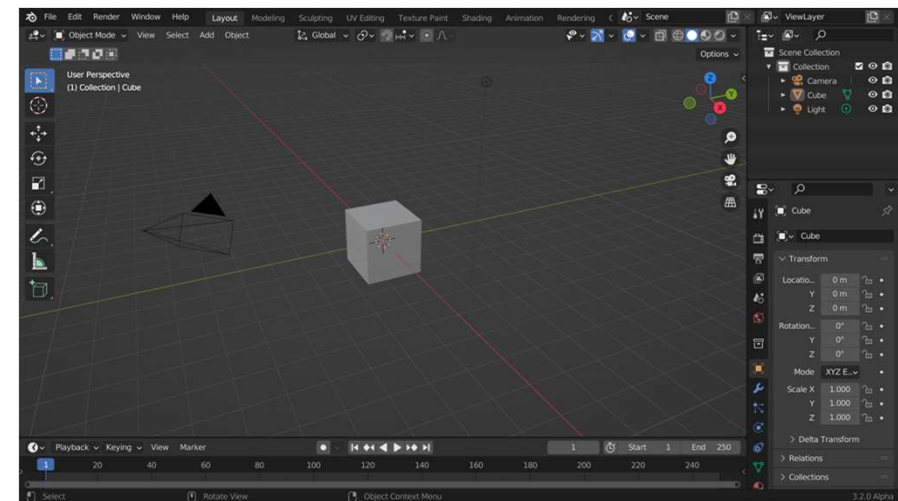
- Surveys of ML practitioners:
 - a lot time is taken by data processing rather than model improvement
 - ... and no one appears to be too happy about that
- Data acquisition, processing and labelling:
 - costly
 - time-consuming
 - tedious
- Typical computer vision workarounds: web scraping and crowdsourcing (e.g., Eyewire)
 - ... yet they may not work for robotics
- Outsourcing (e.g., Mechanical Turk) has its issues as well





Possible solution: synthetic data

- Outline of the approach:
 - generate images with 3D modelling software such as Blender
 - optional step: make them more similar to real-world images
 - train the model:
 - either on synthetic images alone
 - or on the augmented (synth + real) dataset
- **Advantages:**
 - possible to generate a lot of the data
 - possible to modify the data
 - labelling is available by default
- **Main challenge: synthetic $\neq$ real**



Synthetic data for bin-picking task (1)

Task: object detection

real-world images



synthetic images



Synthetic data for bin picking task (2)

Real-world images:

- image size: 528x342 pixels
- 2200 samples
- manually labelled
- constant angle & brightness



Synthetic images generated with Blender:

- image size: 1024 x 768 pixels;
- 8800 samples
- 16 different angles
- different brightness levels
- realistic textures & reflections
- automated labelling: object rotation angle, coordinates & visibility percentage



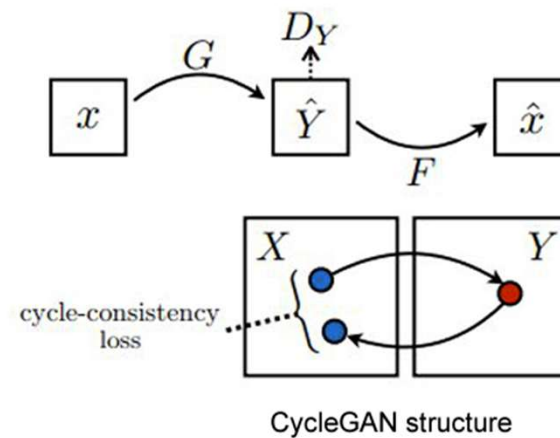
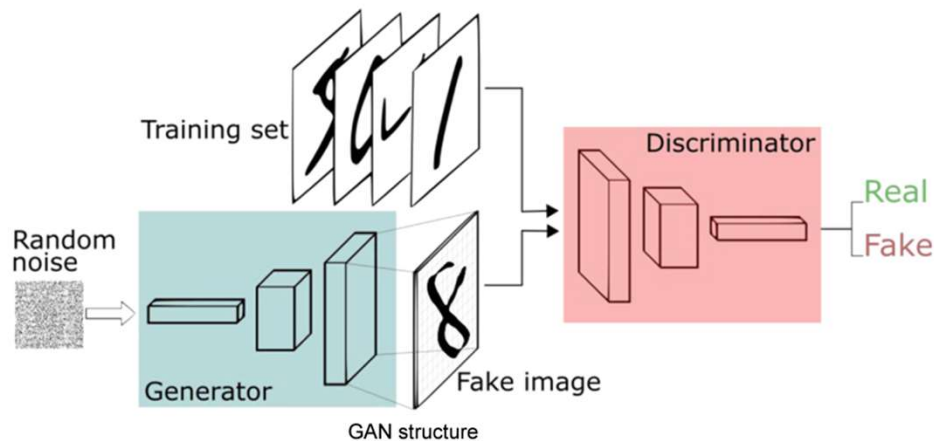
Synthetic data for bin picking task (3)

Object detection experiments (in: Arents et al (2022) *Synthetic Data of Randomly Piled, Similar Objects for Deep Learning-Based Object Detection*):

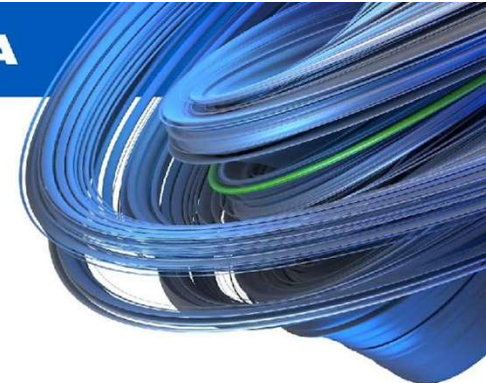
Data distribution			Test 1			Test 2		
Real	Synthetic	Real / Synthetic Ratio %	Step	AP @0.5:0.95	OD %	Step	AP @0.5:0.95	OD %
8800	0	100 / 0	9100	88.61	100.00	9100	69.22	96.95
7920	880	90 / 10	7900	88.61	100.00	9200	71.04	98.47
7040	1760	80 / 20	6000	88.36	100.00	8000	73.23	100.00
6160	2640	70 / 30	6300	88.65	100.00	7600	72.83	100.00
5280	3520	40 / 60	6900	88.22	100.00	7400	72.50	100.00
4400	4400	50 / 50	4400	85.82	100.00	5000	73.84	100.00
3520	5280	40 / 60	8000	85.59	100.00	8700	70.27	100.00
2640	6160	30 / 70	7200	84.39	100.00	5900	64.57	100.00
1760	7040	20 / 80	7200	84.23	100.00	4500	63.62	100.00
880	7920	10 / 90	8200	82.59	100.00	7000	63.25	100.00
0	8800	0 / 100	7700	70.01	100.00	5000	38.66	100.00

Synthetic data for bin picking task (4)

To bridge Sim2Real gap, we use GAN (Generative Adversarial Network) DNN, in particular CycleGAN



- Doesn't need a lot of data (~ 1000 images is enough)
- Translates image domain
- Preserves semantics of image



Synthetic data for bin picking task (5)

Examples of the experiments with CycleGAN



input image



synthetic: CycleGan original implementation in Tensorflow

- strong pixel grid
- objects in shadows erased
- draws nonexistent objects
- cropped bottles on the left-hand side



synthetic: added augmentation:

- strong pixel grid
- some shadow areas are erased



synthetic: added augmentation & Conv2DTranspose replaced with resized Conv

- small pixel grid
- shadow areas translated correctly
- less blurred



synthetic: added augmentation & noise layer:

- decreased pixel grid
- shadow area translated correctly
- best FID score

Synthetic data for bin picking task (5)

Examples of the experiments with CycleGAN



input image



synthetic: CycleGan original implementation in Tensorflow

- strong pixel grid
- objects in shadows erased
- draws nonexistent objects
- cropped bottles on the left-hand side



synthetic: added augmentation:

- strong pixel grid
- some shadow areas are erased



synthetic: added augmentation & Conv2DTranspose replaced with resized Conv

- small pixel grid
- shadow areas translated correctly
- less blurred



synthetic: added augmentation & noise layer:

- decreased pixel grid
- shadow area translated correctly
- best FID score

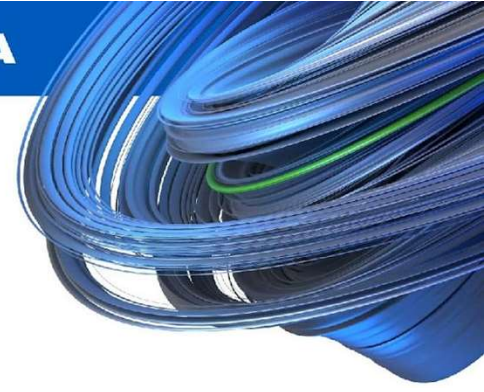
Resized transpose Augmented noise

Synthetic data for bin picking task (6)

Object detection experiments with YOLOv5 architectures

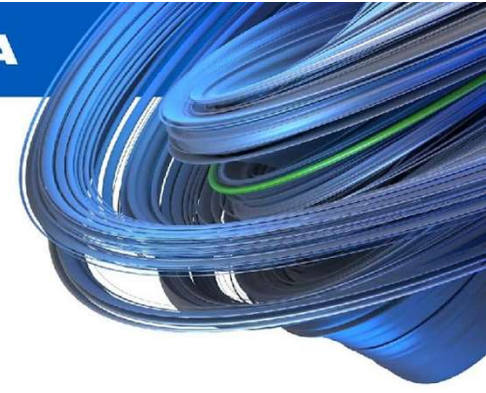
Model	Dataset	Precision	Recall	mAP (threshold 0.5)	mAP (avg for IoU ∈ [0.5 : 0.95])
YOLOv5 Small	Original	0.689	0.894	0.742	0.447
	synthetic				
	Resized	0.615	0.838	0.632	0.245
	transpose Augmented noise	0.731	0.905	0.777	0.48
YOLOv5 Medium	Original	0.693	0.802	0.72	0.42
	synthetic				
	Resized	0.581	0.782	0.589	0.204
	transpose Augmented noise	0.712	0.91	0.753	0.455
YOLOv5 XLarge	Original	0.718	0.785	0.75	0.413
	synthetic				
	Resized	0.683	0.867	0.715	0.339
	transpose Augmented noise	0.721	0.872	0.761	0.461

In: Duplevska et al. (2022)
*Sim2Real image
translation to improve a
synthetic dataset for a bin
picking task*



Opportunities & challenges: what's next?

- Our research has demonstrated the advantages of using synthetic data and Sim2Real transfer
- Future work in the iMOCO4.E project:
 - further improve image synthesis with CycleGAN
 - integrate CycleGAN in the Reinforcement Learning environment
 - deploy trained model-based robot in a factory environment
- Broad perspectives:
 - look for alternative image generation approaches, e.g., diffusion models
 - is making images more similar the only way to go?
 - are currently available Sim vs. Real metrics sufficient?



Funding acknowledgement

Intelligent Motion Control under Industry 4.E

iMOCO4.E

Budget: 31M EUR

Partners: 46

Countries: Netherlands, Czech Republic, Spain,
Greece, Ireland, Italy, Belgium, Latvia, Poland,
Denmark, Romania, Finland, Switzerland

IMOCO project has been co-funded by the Horizon 2020
Framework Programme of the European Union
Under grant agreement No. 101007311



Thank you!