Deep Learning and Computer Vision for Surface Anomaly Detection

Danijel Skočaj

Visual Cognitive Vision Laboratory
Faculty of Computer and Information Science
University of Ljubljana

ACAI 2024, Athens, 15. 7. 2024

ViCoS
Visual Cognitive Systems lab

FRI
UNIVERSITY OF LJUBLJANA
Faculty of Computer and Information Science

Danijel Skočaj

FRI
UNIVERSITY OF LJUBLJANA
Faculty of Computer and Information Science

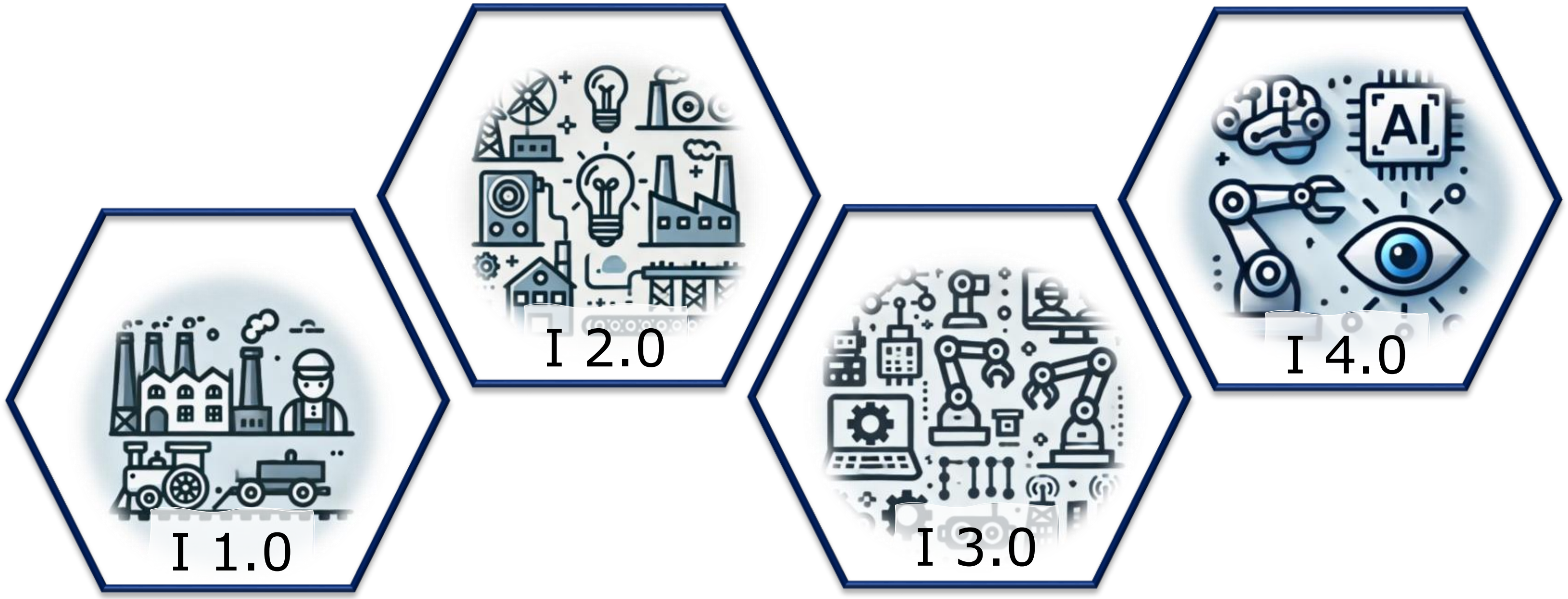
ViCoS
Visual Cognitive Systems lab

danijel.skocaj@fri.uni-lj.si

Industry 4.0

Electrical power mass production

Intelligent automation



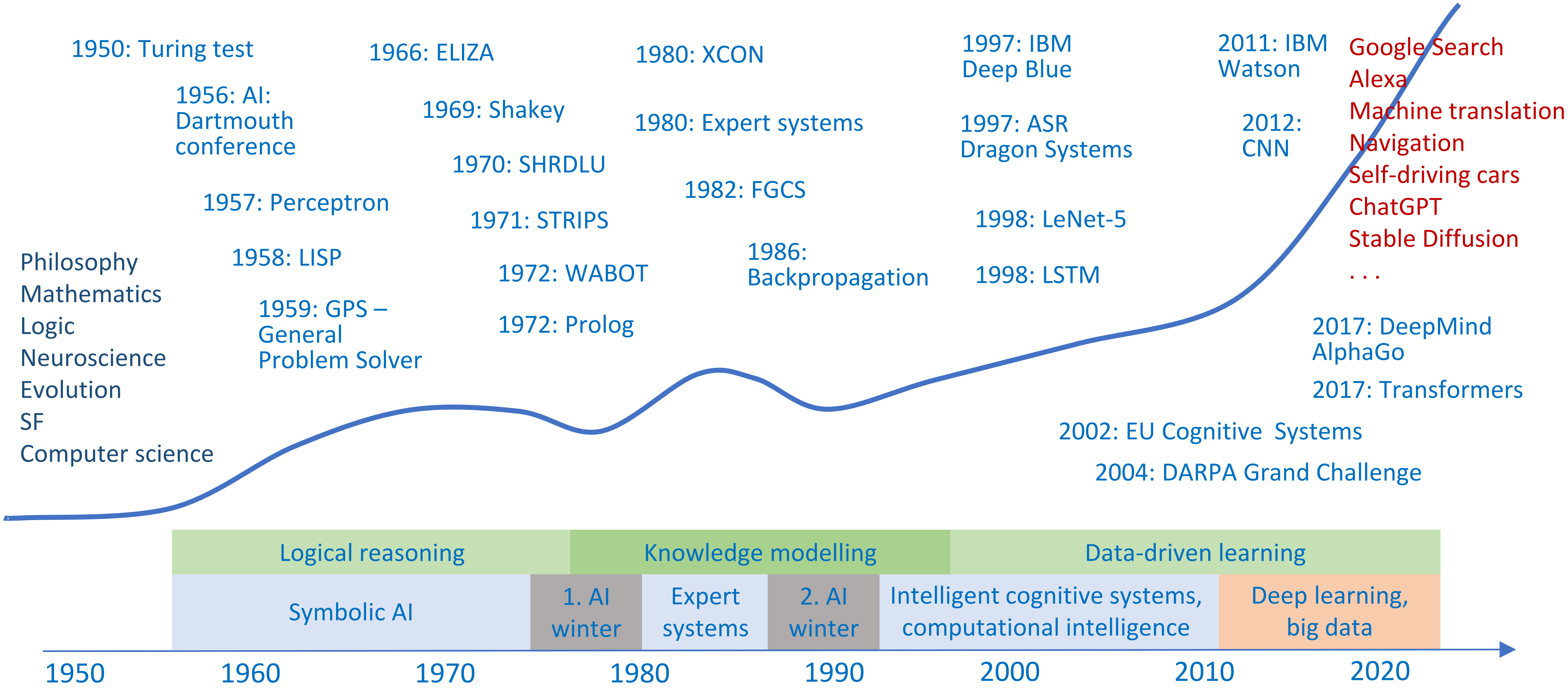
Steam power mechanisation

Automation of routine processes

Smart *
- Connectivity
- Integration
- IoT
- Robotics
- Big Data
- AI

Data-driven learning-based solutions

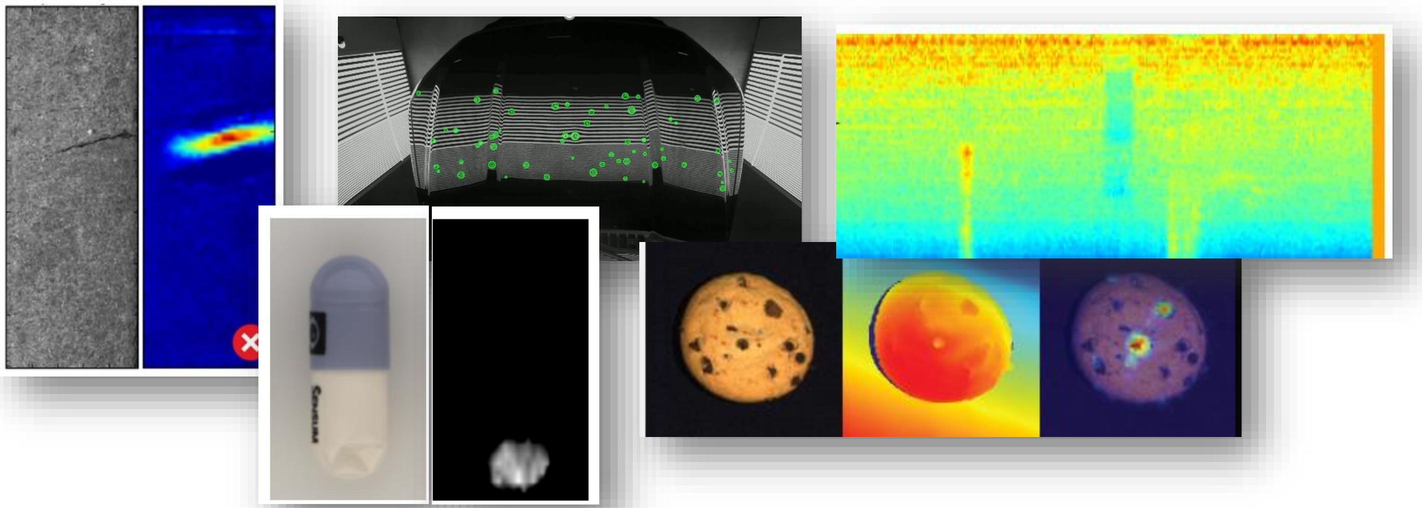
Development of artificial intellignce



Outline

Data → Machine learning → Solutions

Images → Machine vision → Solutions



Machine vision

Surface defect detection

Different learning regimes

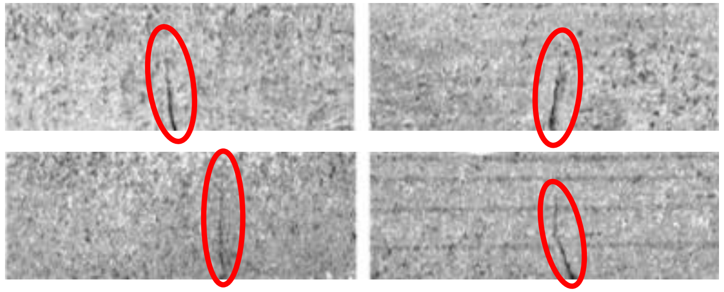
Practical considera-tions

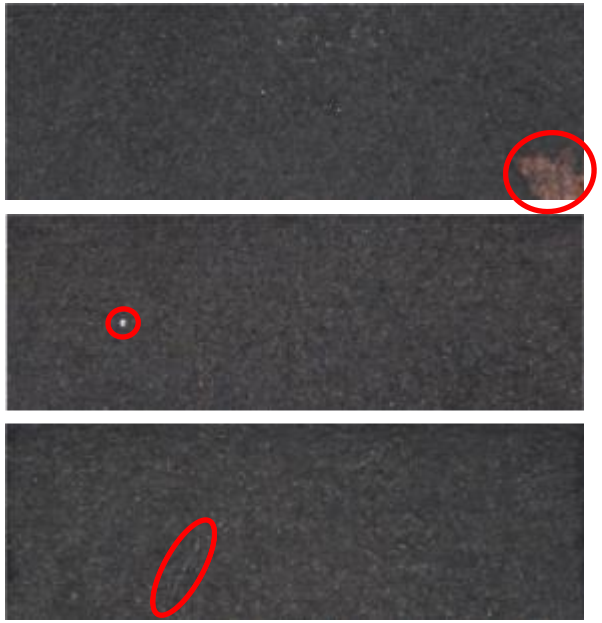
Machine vision

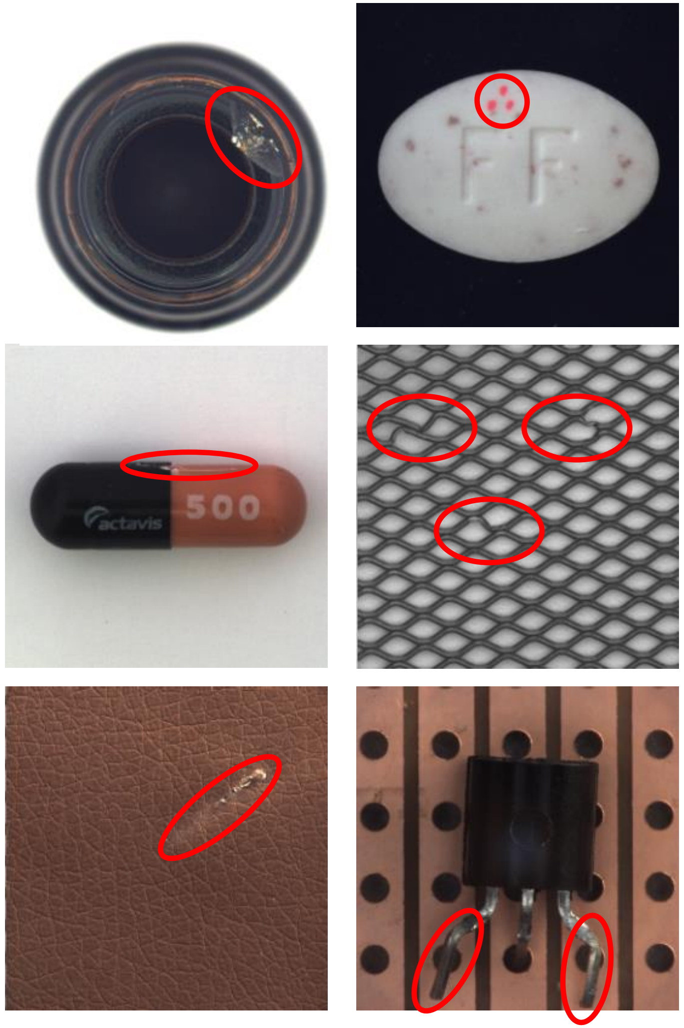
- Machine vision system
  - Environment (background)
  - Illumination
  - Camera and lenses
  - Computational power
  - (Manipulation)
  - Software

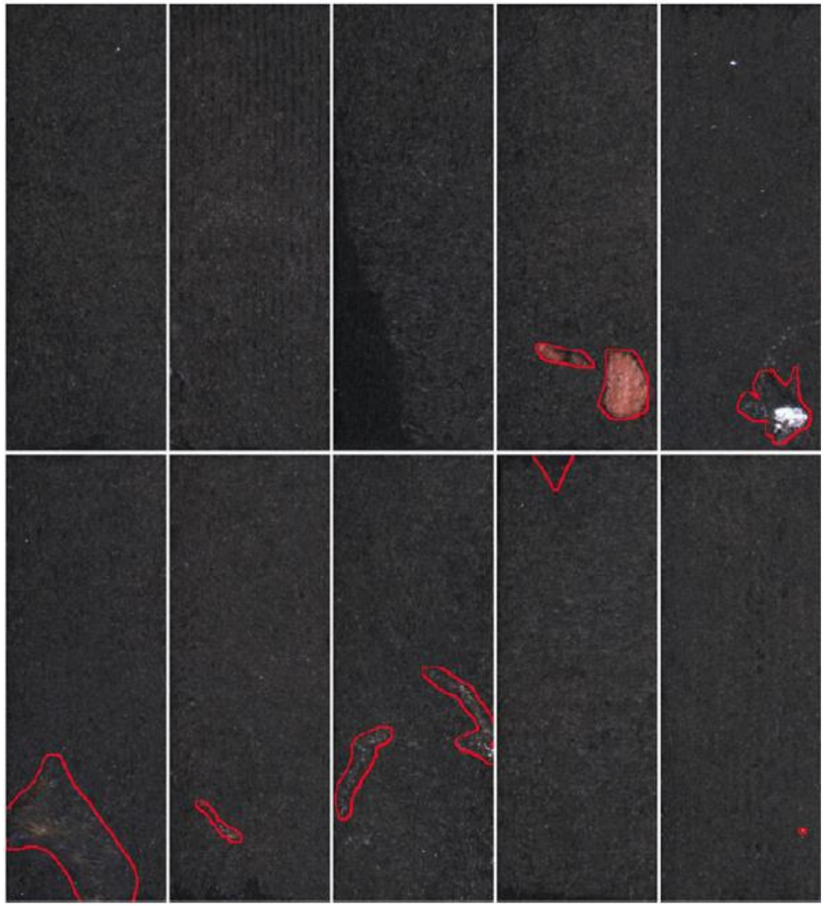
an engineering discipline that uses computer vision algorithms to develop systems for solving practical problems, especially in industrial production

Surface defect detection problem

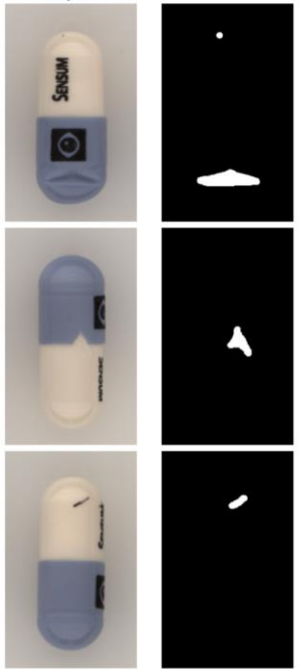






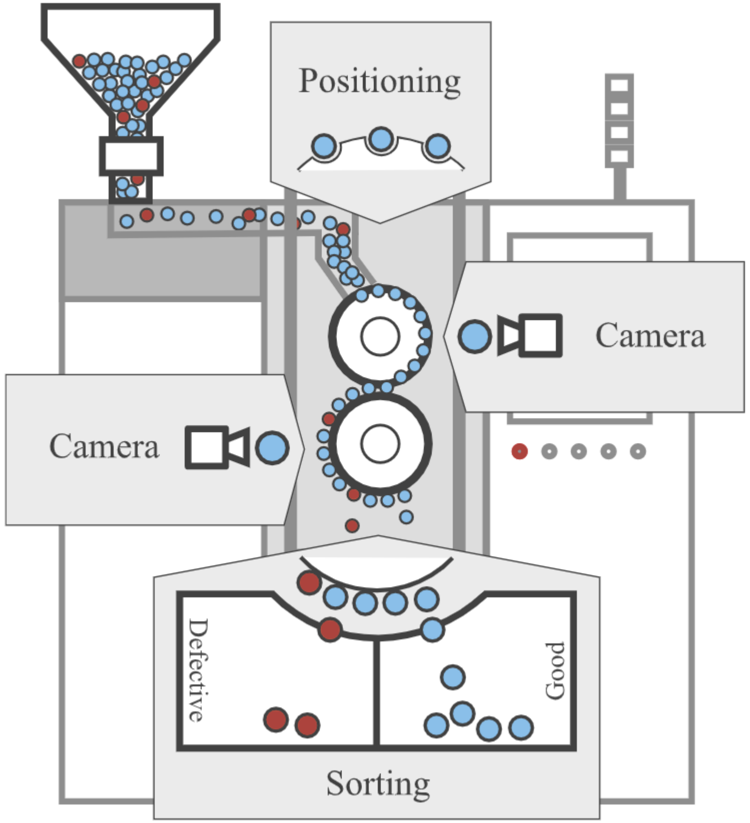


Example: Visual inspection of pharmaceutical products



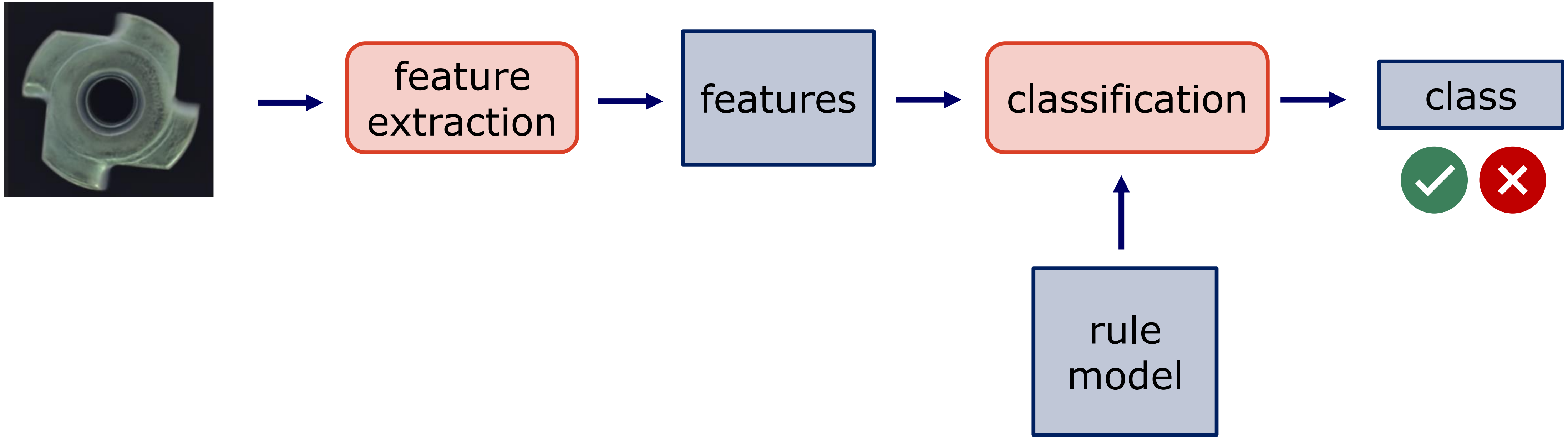
NCAA 2021

sensum.eu



Rule-based machine vision

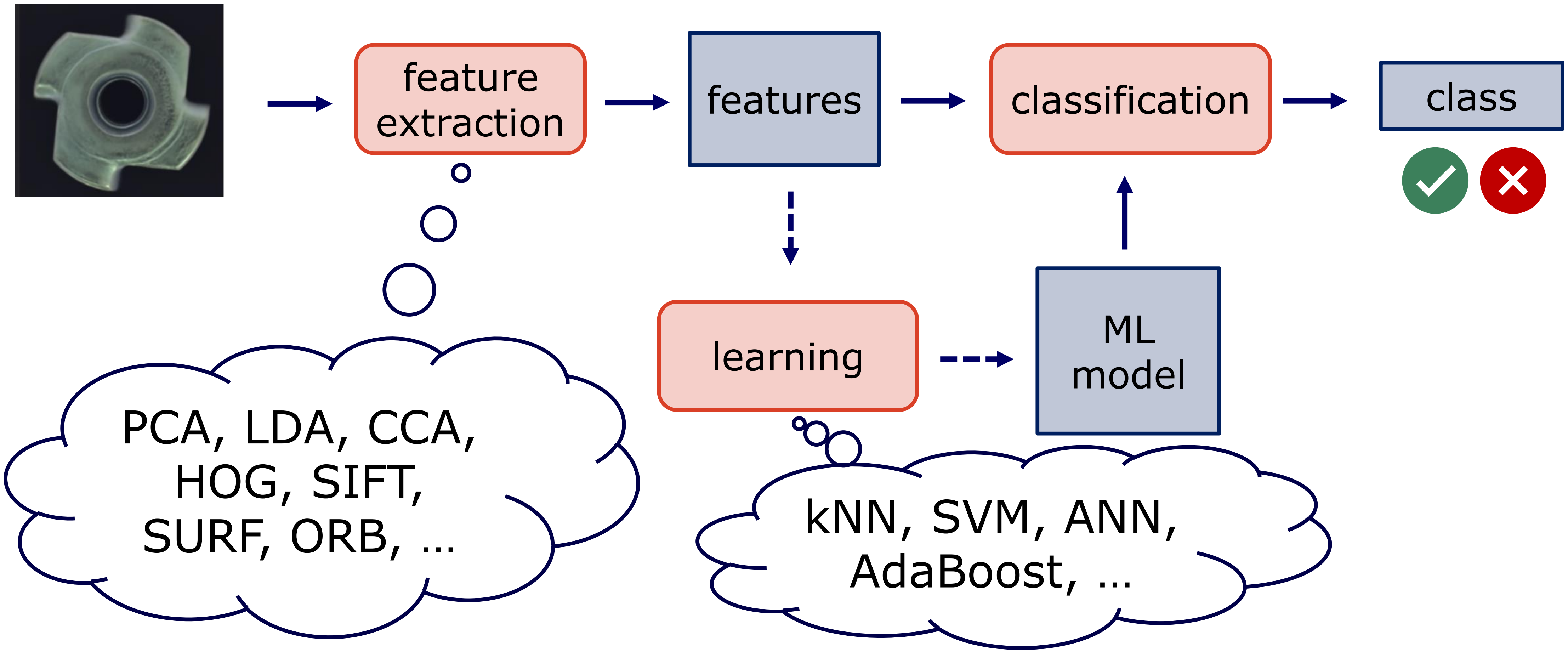
- Rule-based approach



Machine learning in computer vision

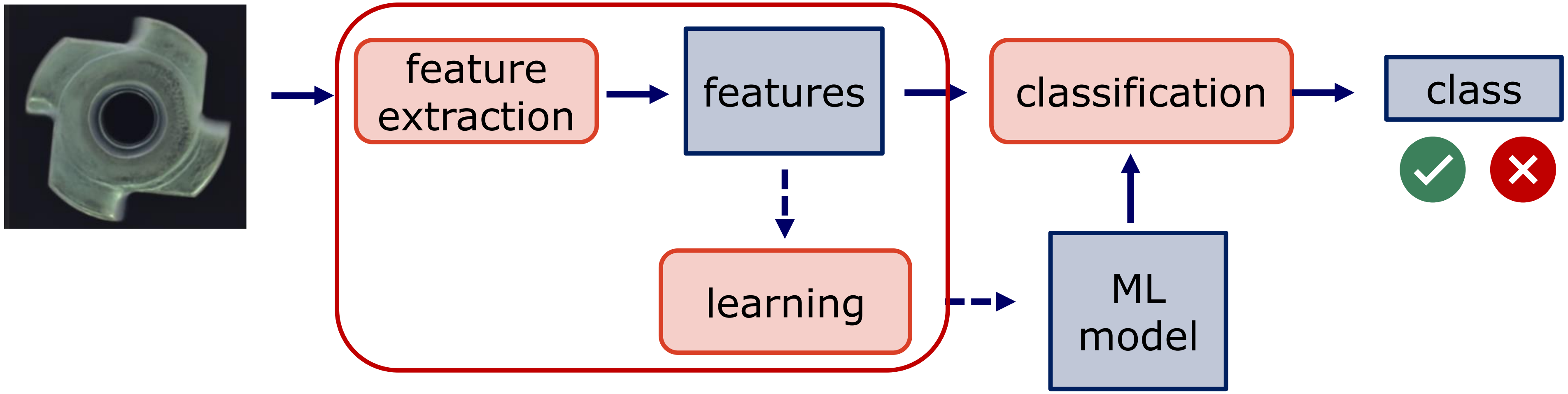
UNIVERSITY OF LJUBLJANA
FRI Faculty of Computer and Information Science

- Conventional ML approach

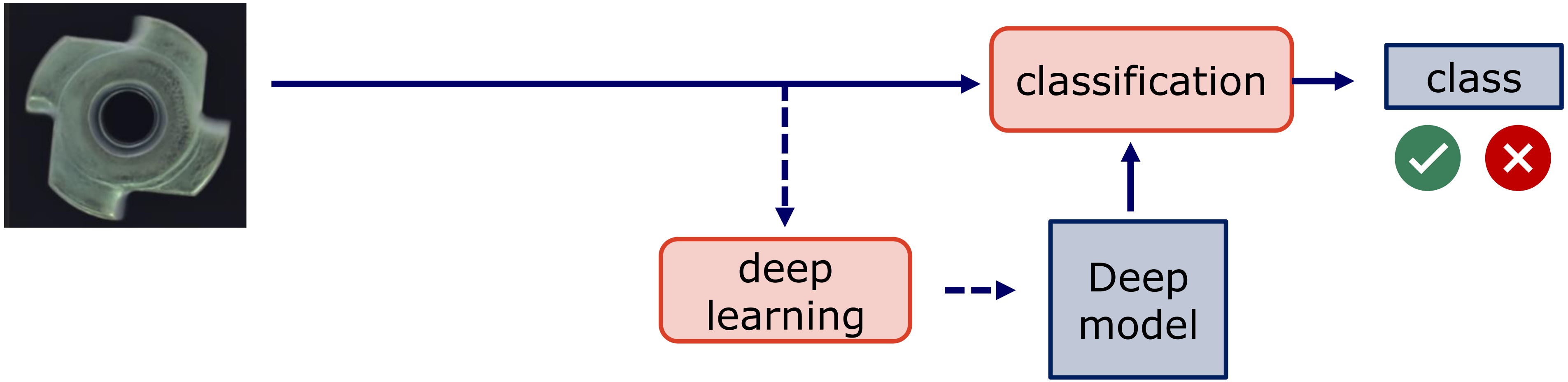


Deep learning in computer vision

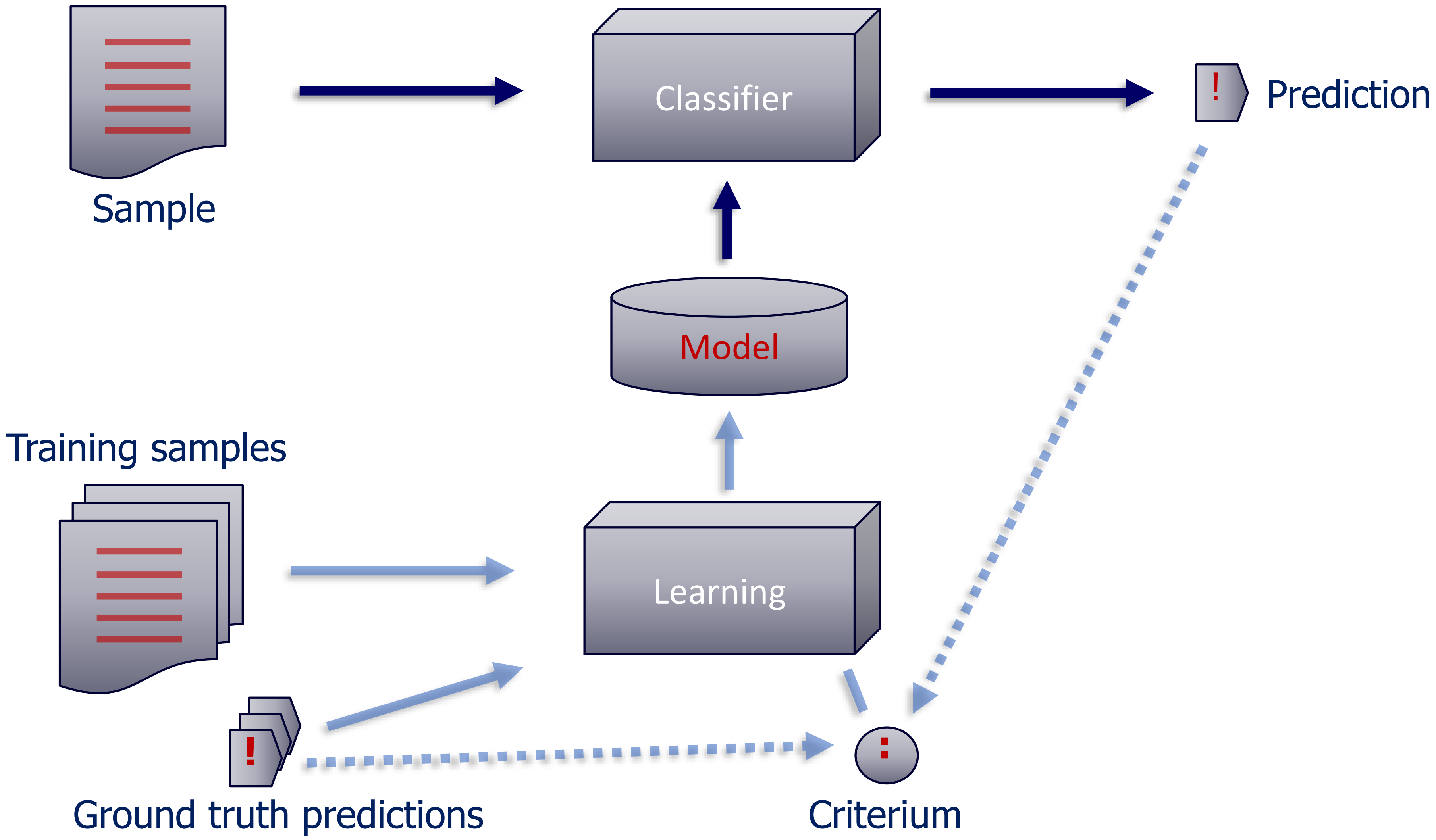
- Conventional machine learning approach in computer vision



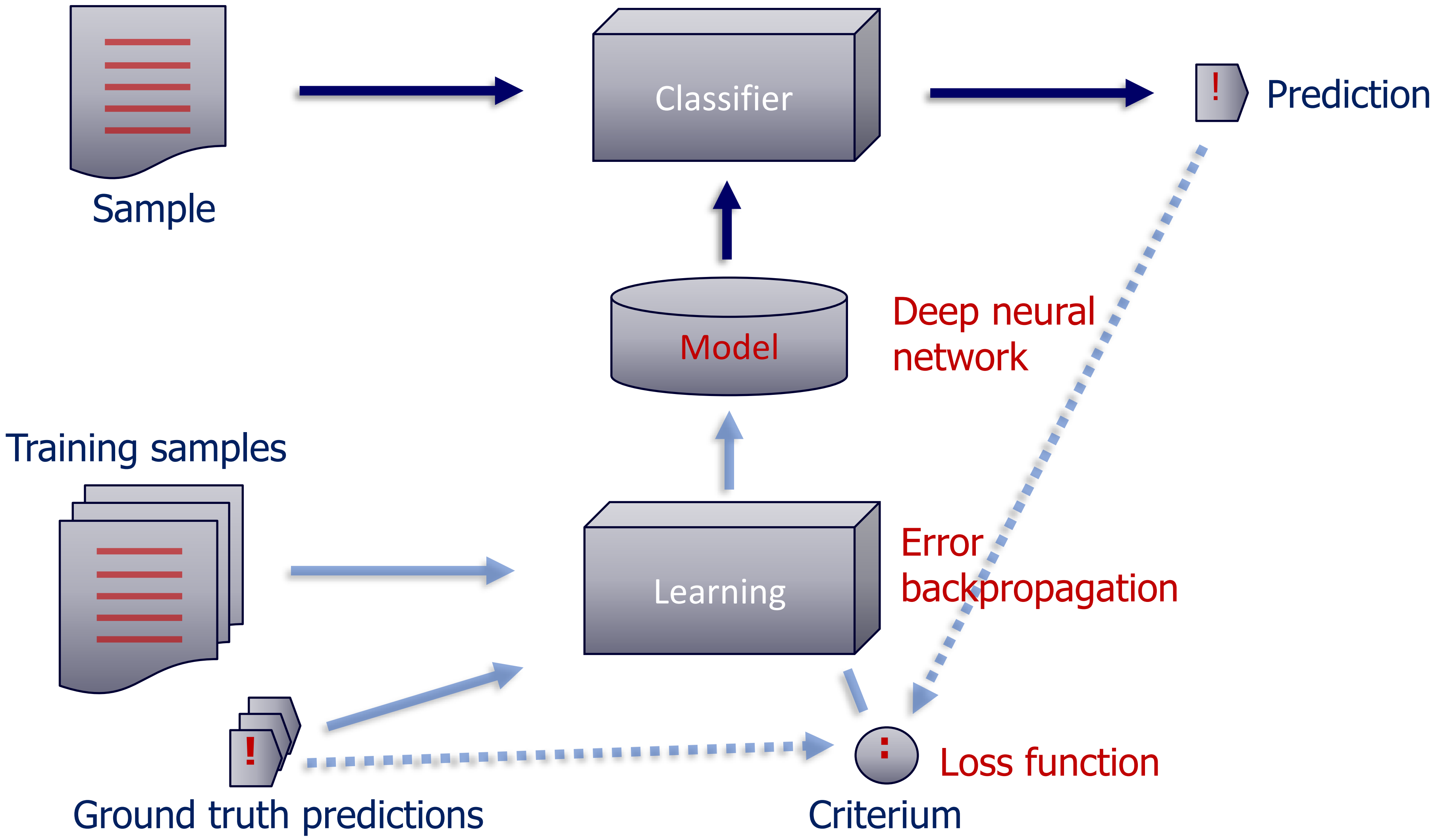
- Deep learing approach



Machine learning

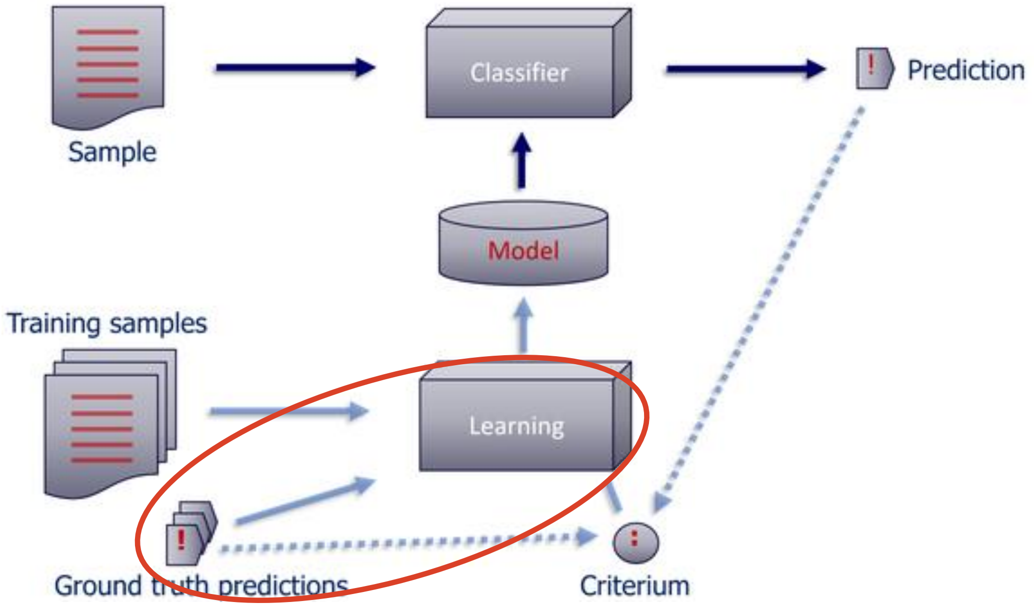


Deep learning

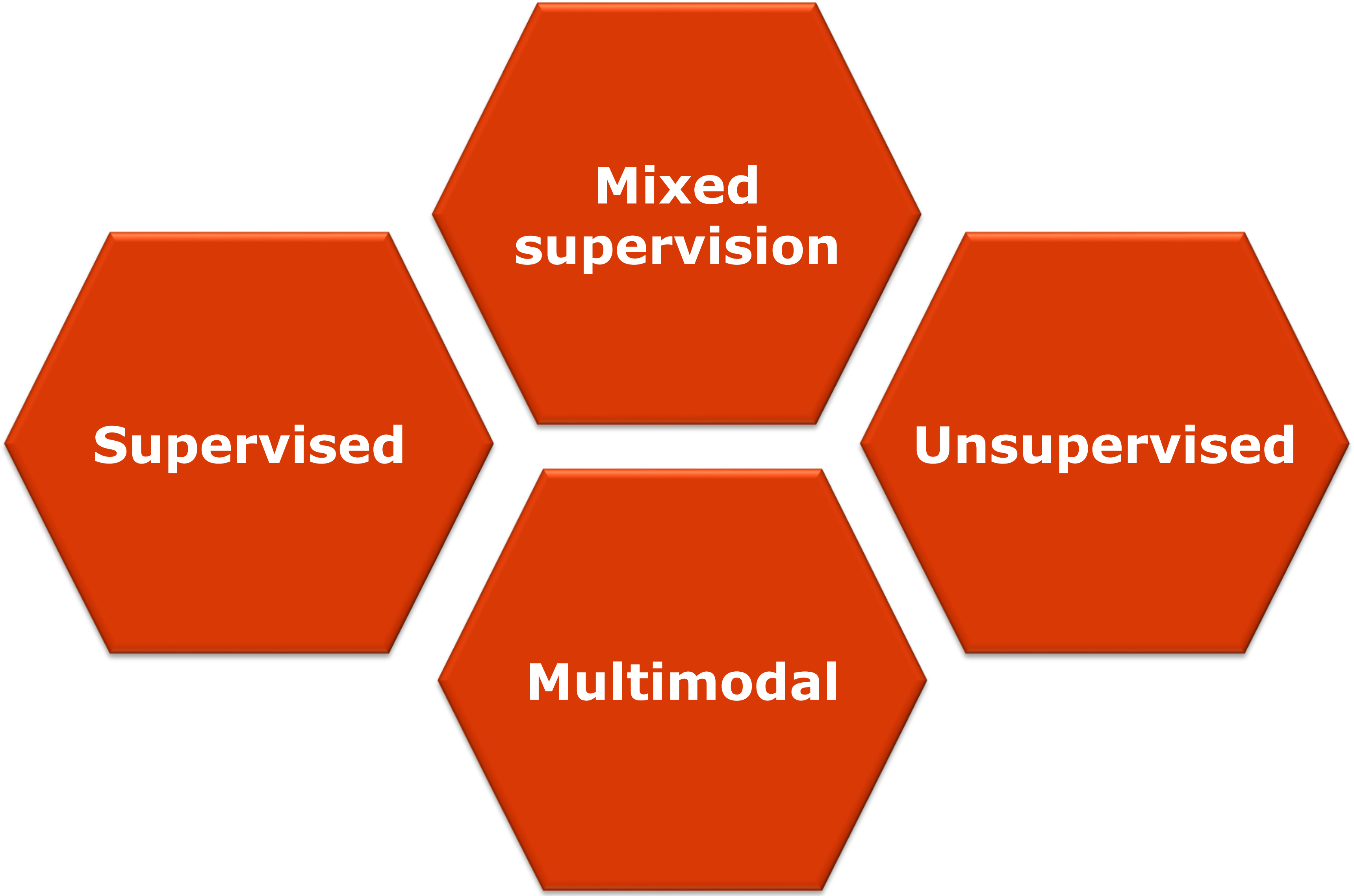


Learning regimes

- Supervised learning
- Weakly supervised learning
- Semisupervised learning
- Unsupervised learning
- Self-supervised learning
- Reinforcement learning

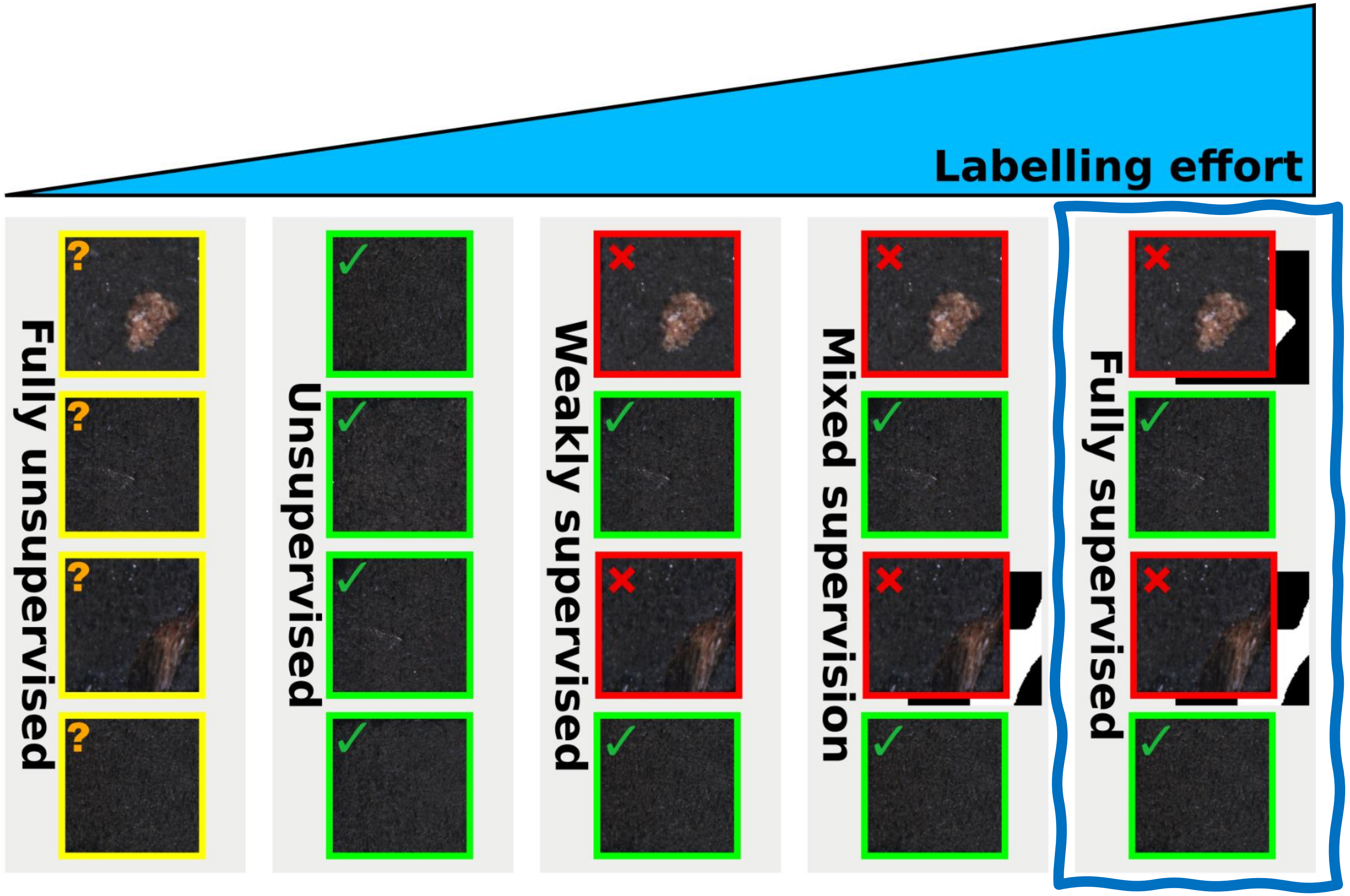


Surface defect detection



Learning regimes

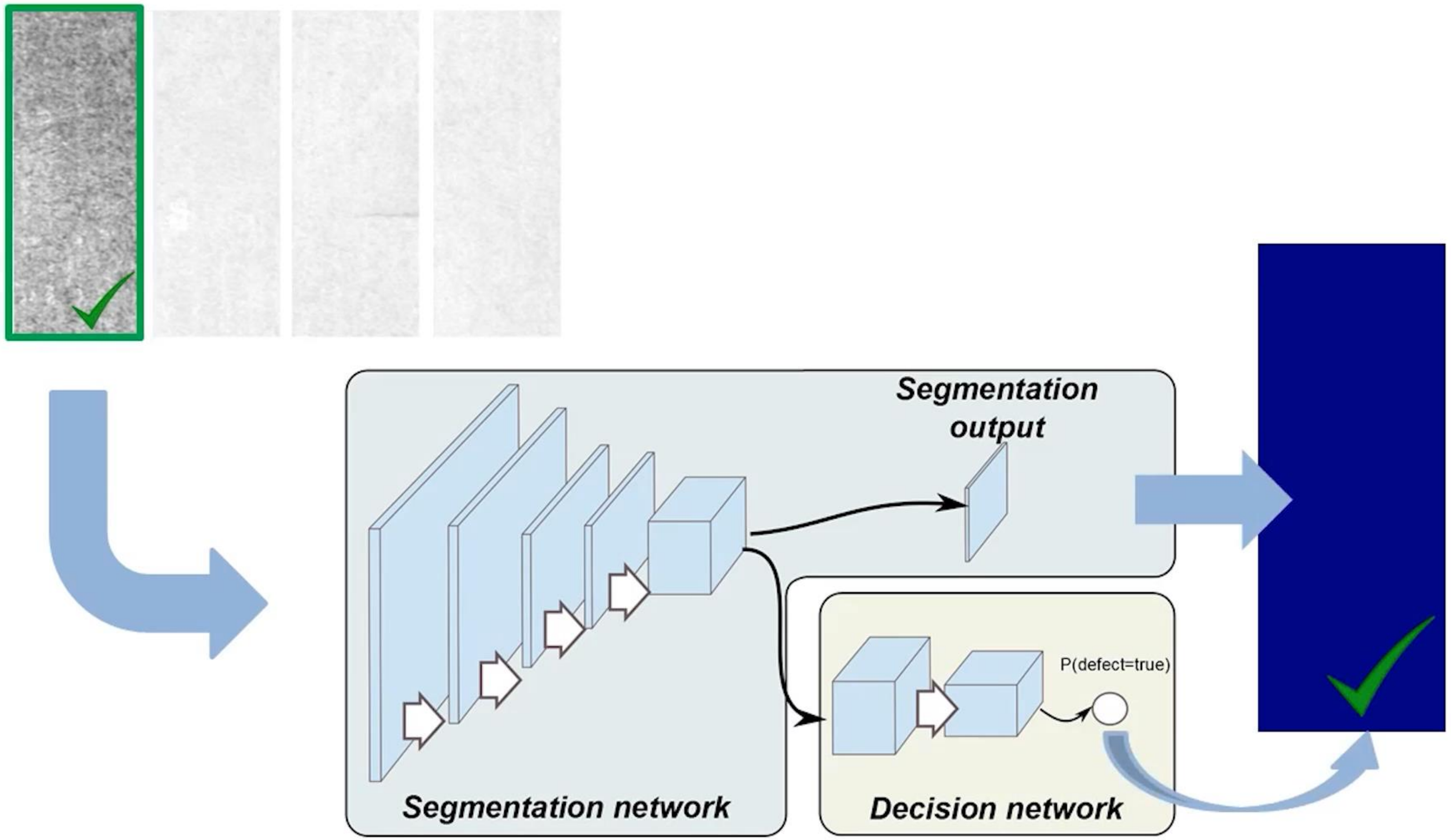
UNIVERSITY OF LJUBLJANA
FRI
Faculty of Computer and Information Science



New paradigm

UNIVERSITY OF LJUBLJANA | FRI Faculty of Computer and Information Science

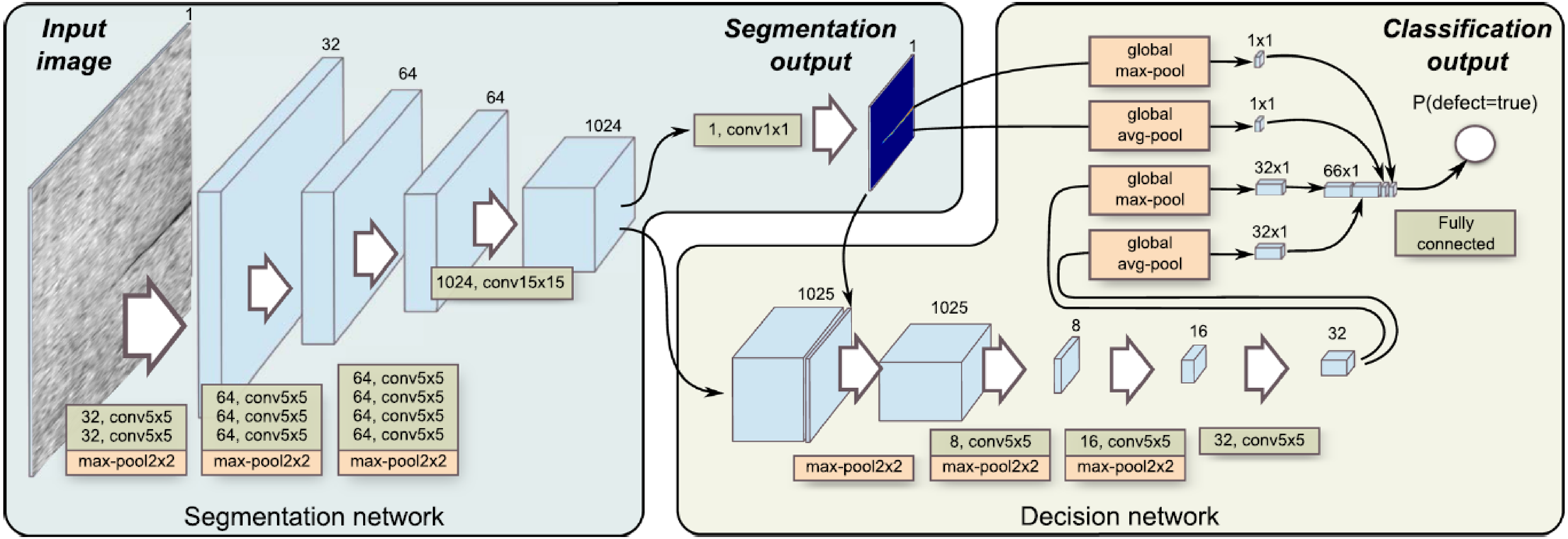
- Conventional approach: programming specific solutions
- New paradigm: data-driven learning-based sloutions



GOSTOP
2016-2020

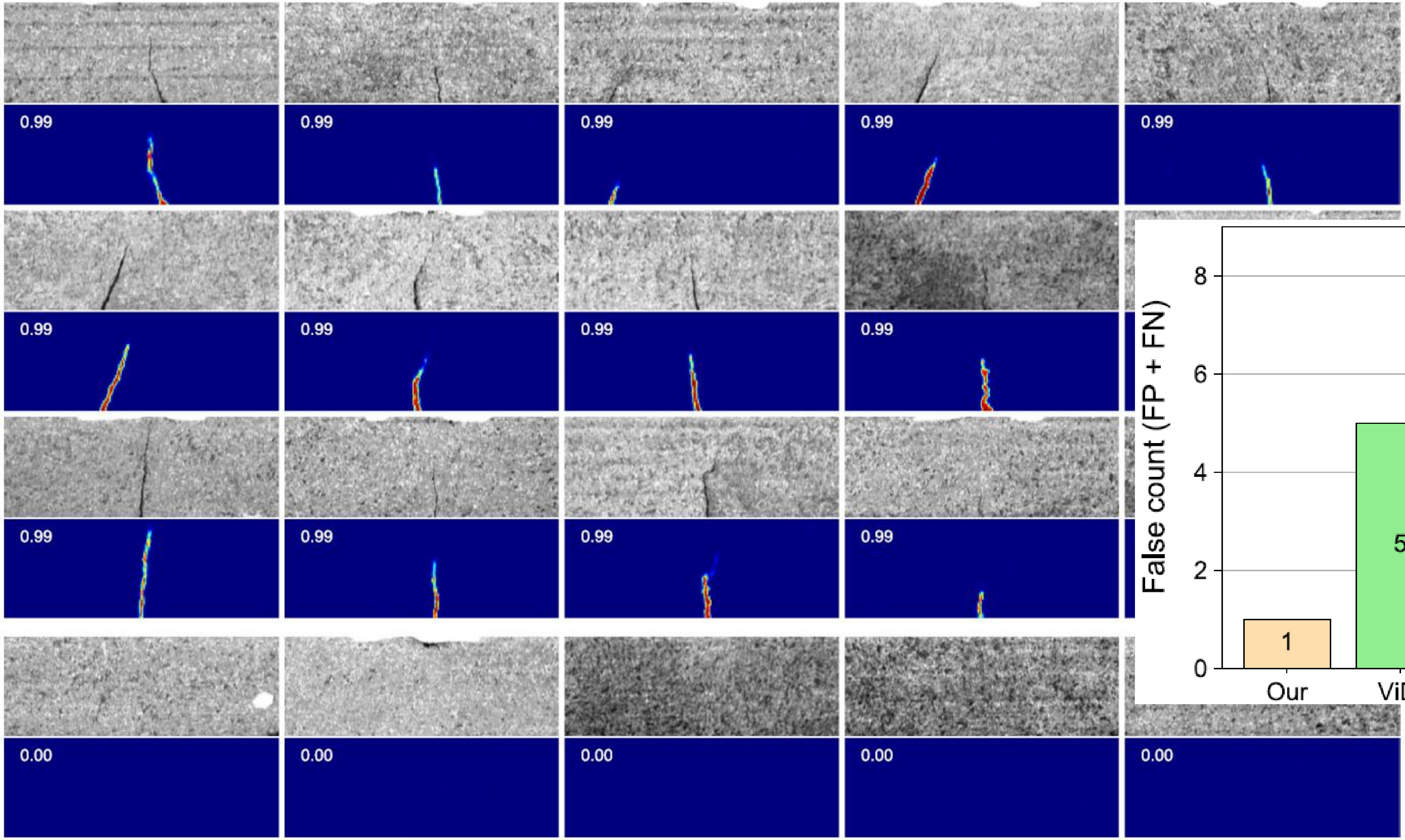
DIVID
2018-2021

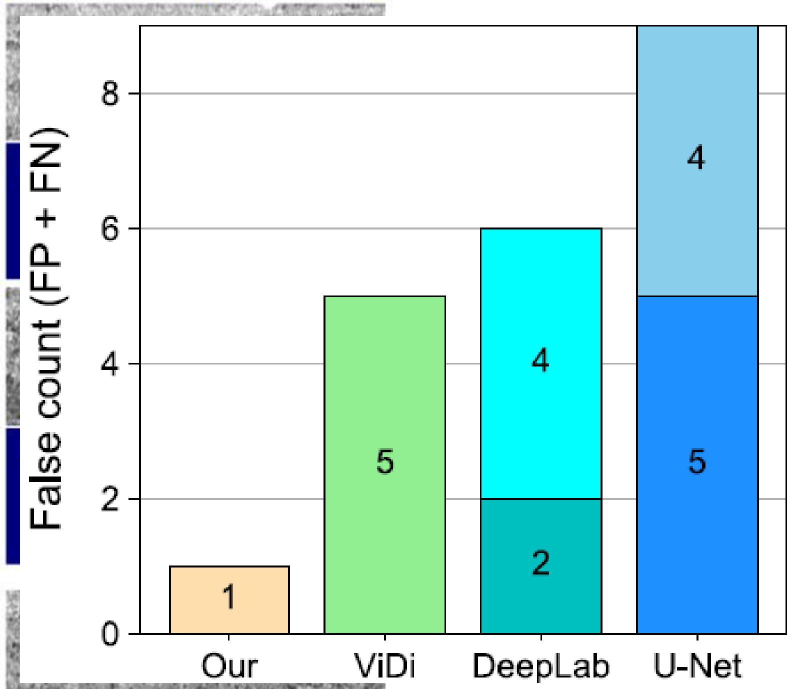
Supervised learning



ICVS 2019 JIM 2020

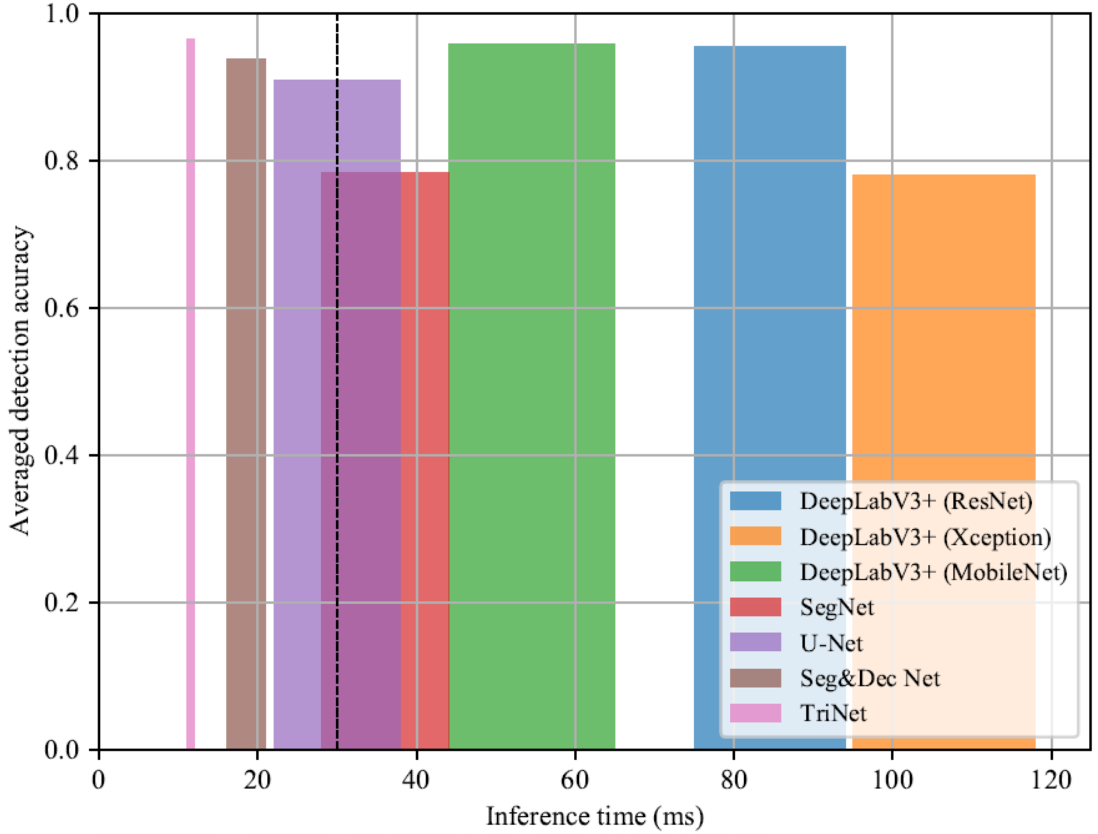
Supervised learning

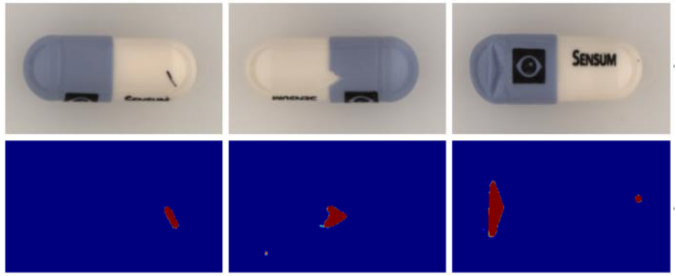




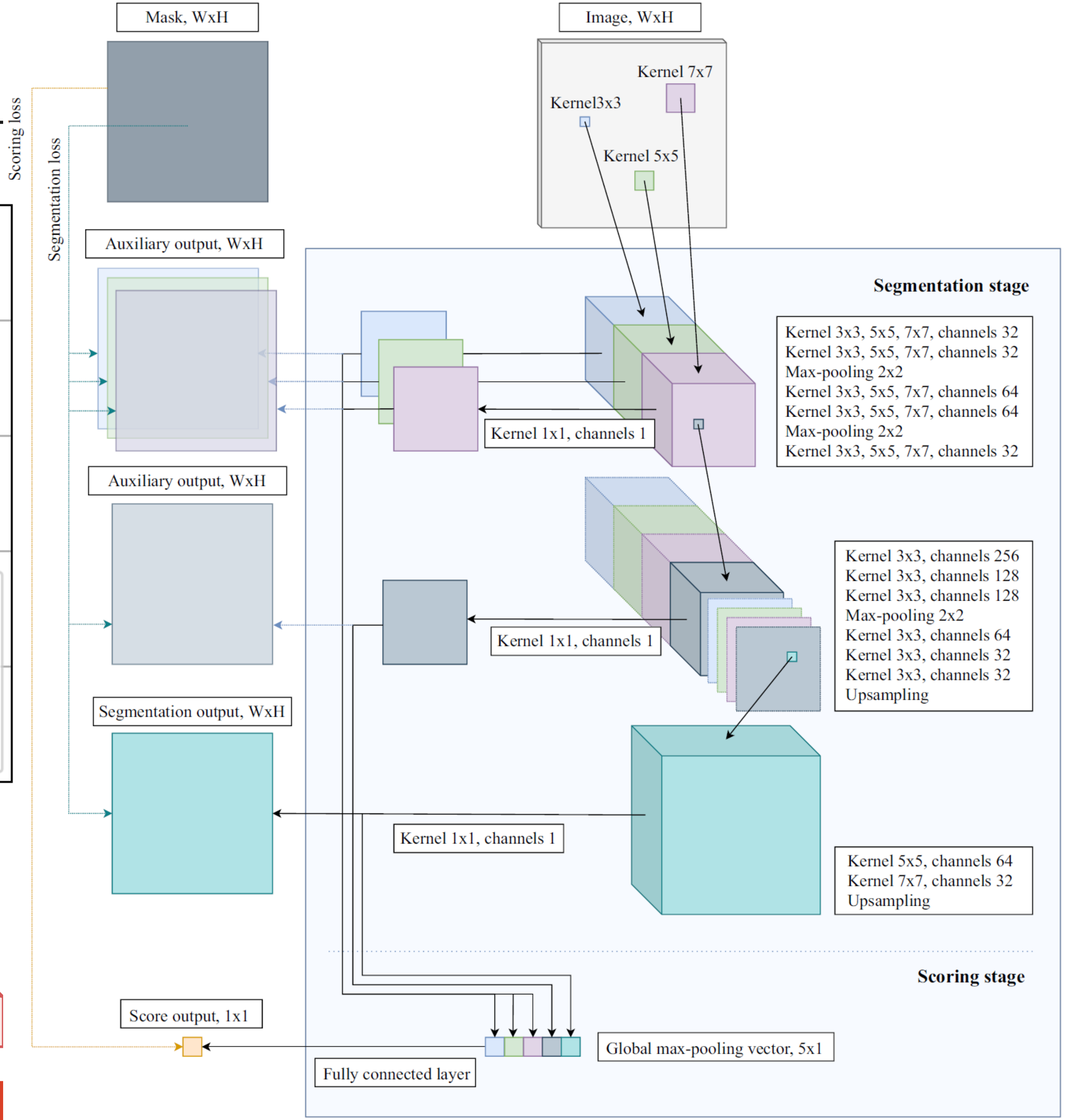
JIM 2020

Supervised learning



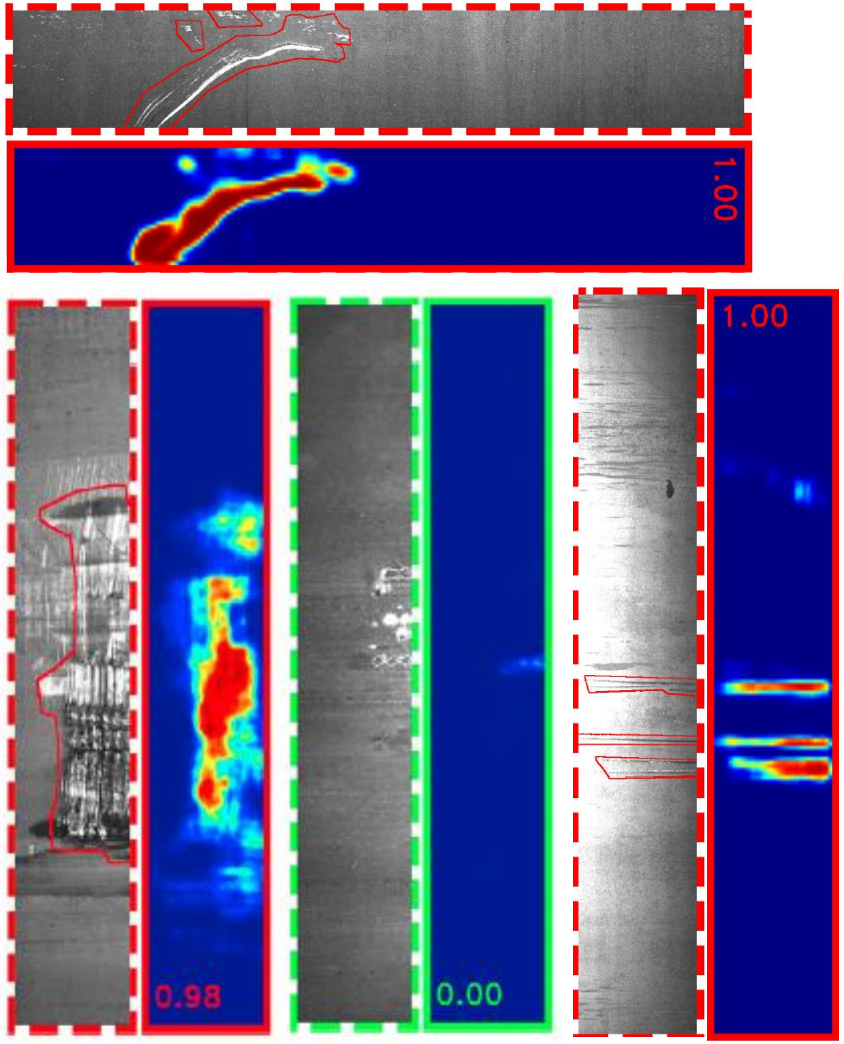


NCAA 2021



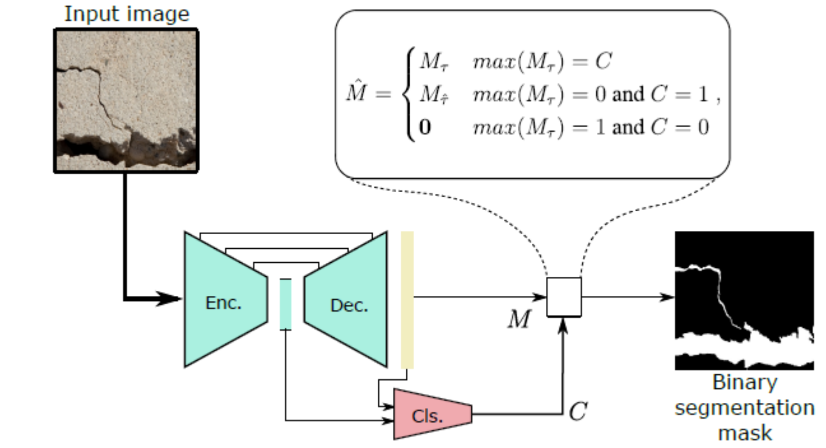
Supervised learning applications

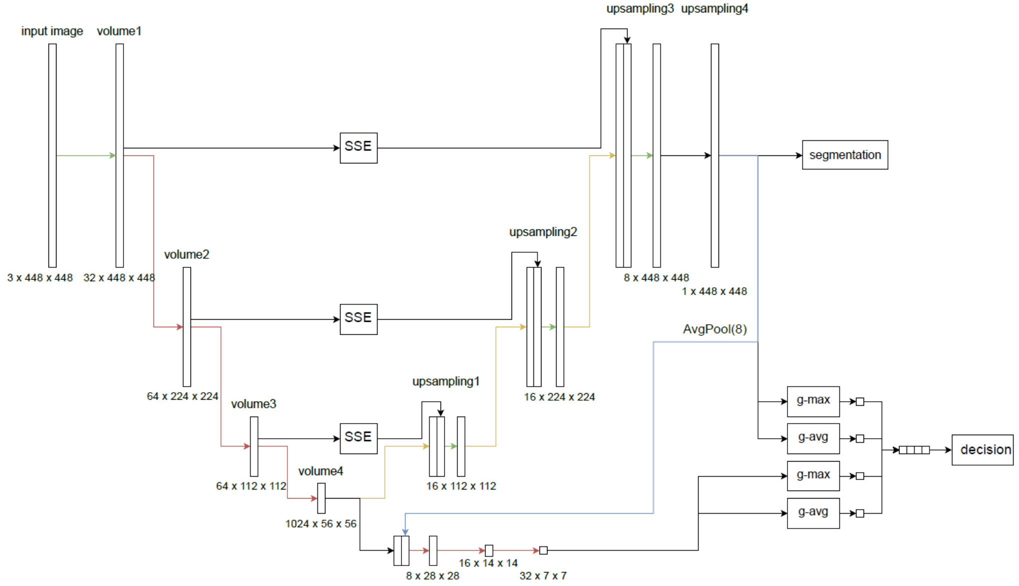
UNIVERSITY OF LJUBLJANA
FRI
Faculty of Computer and Information Science

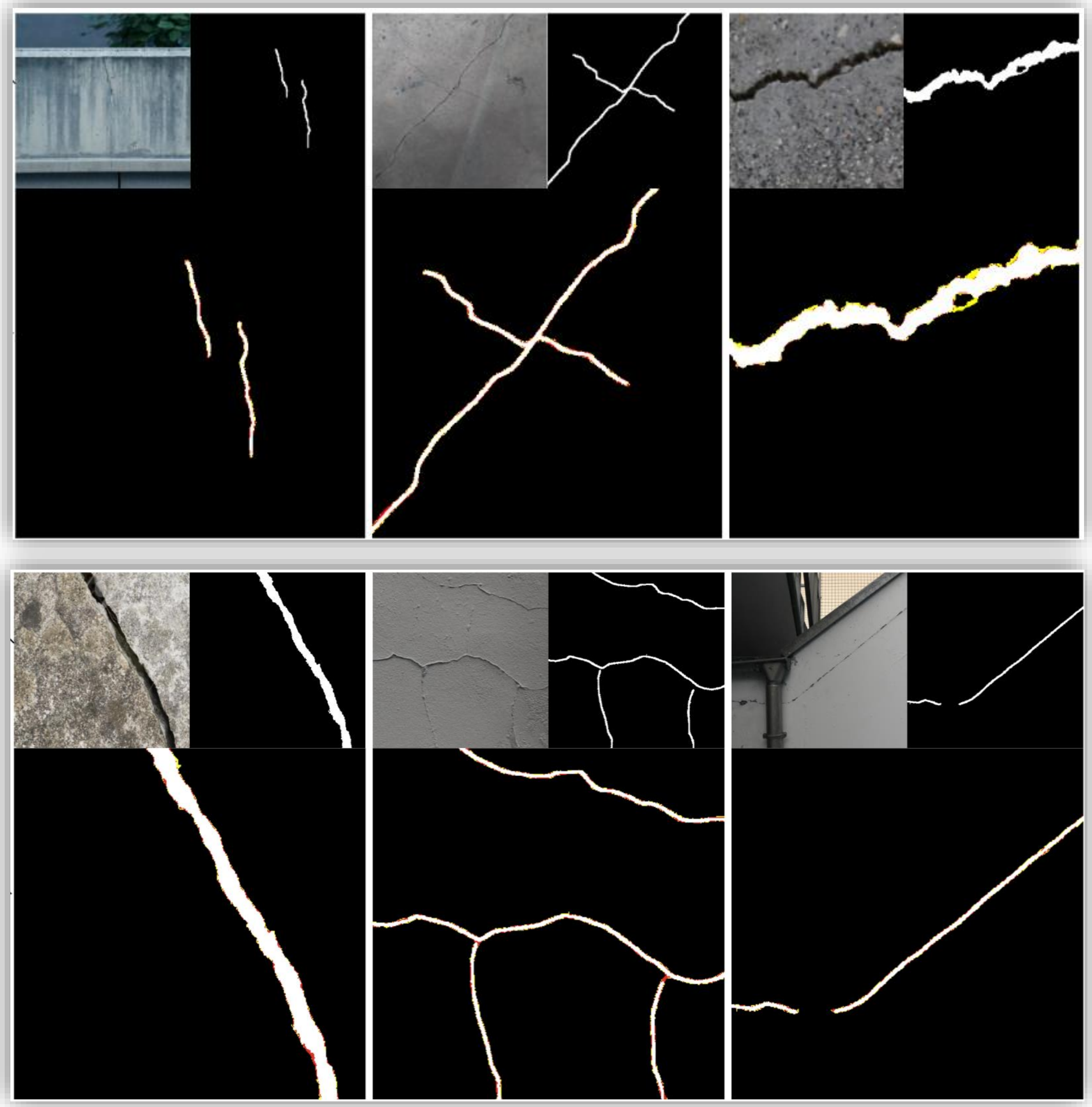


Supervised learning

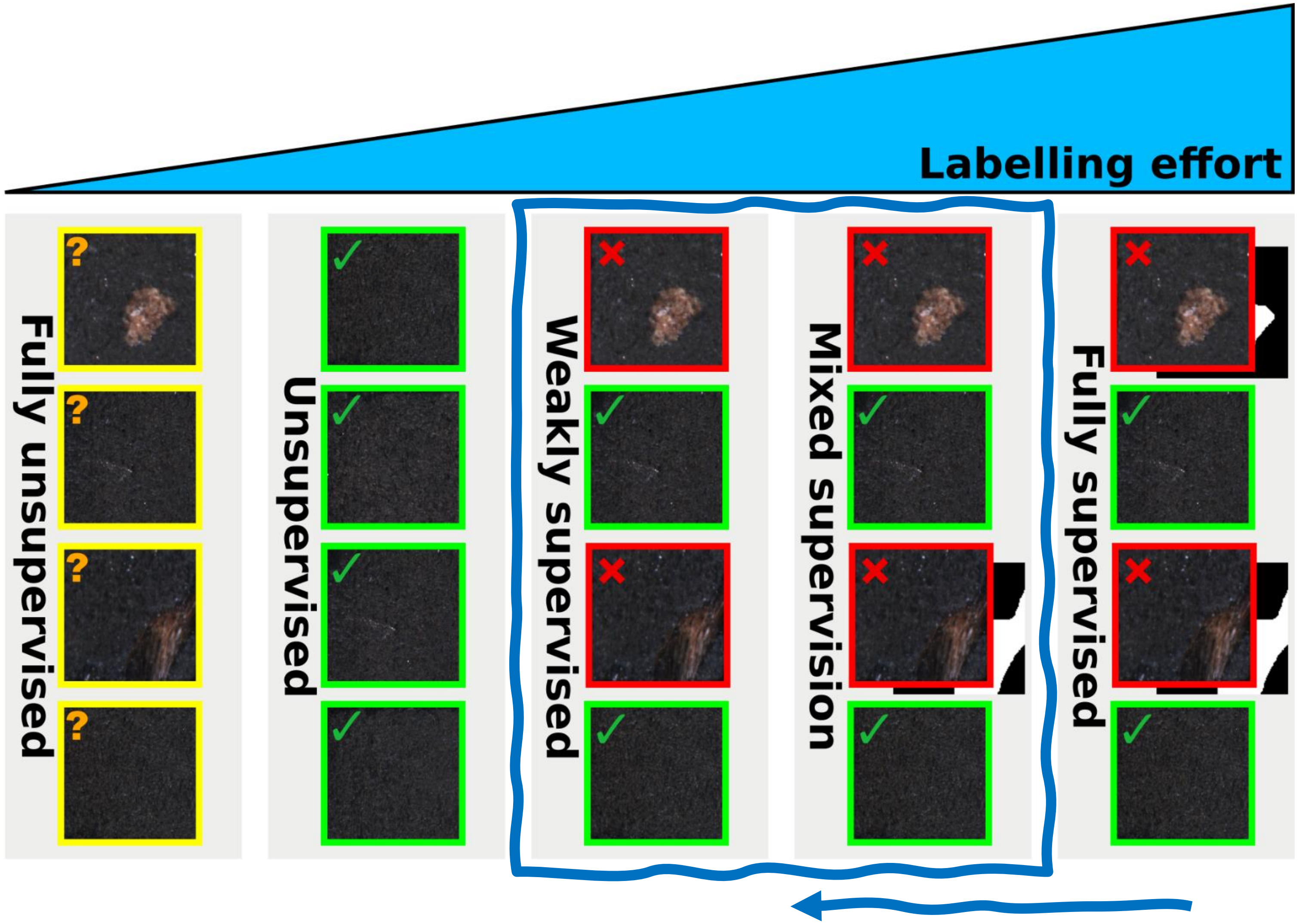
CBM 2023





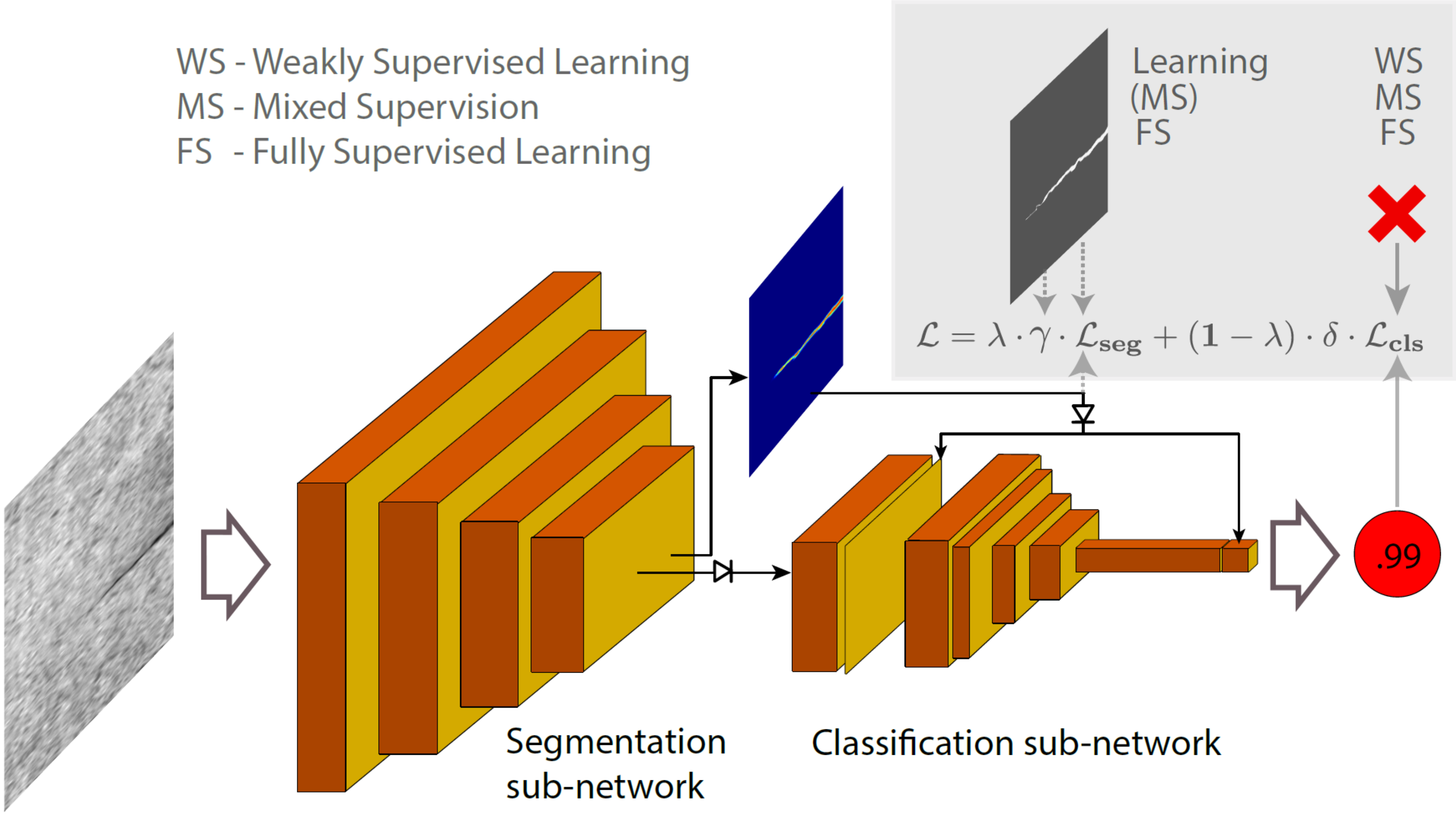


Learning regimes



Learning with mixed supervision

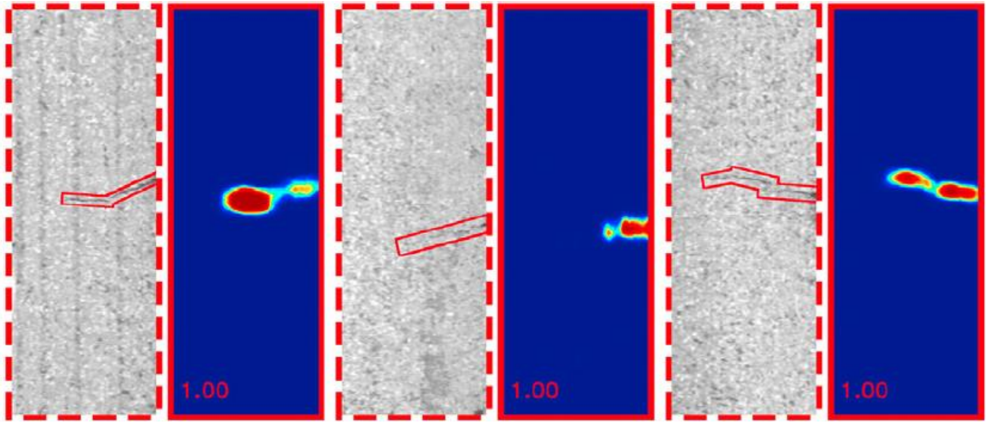
WS - Weakly Supervised Learning
MS - Mixed Supervision
FS - Fully Supervised Learning

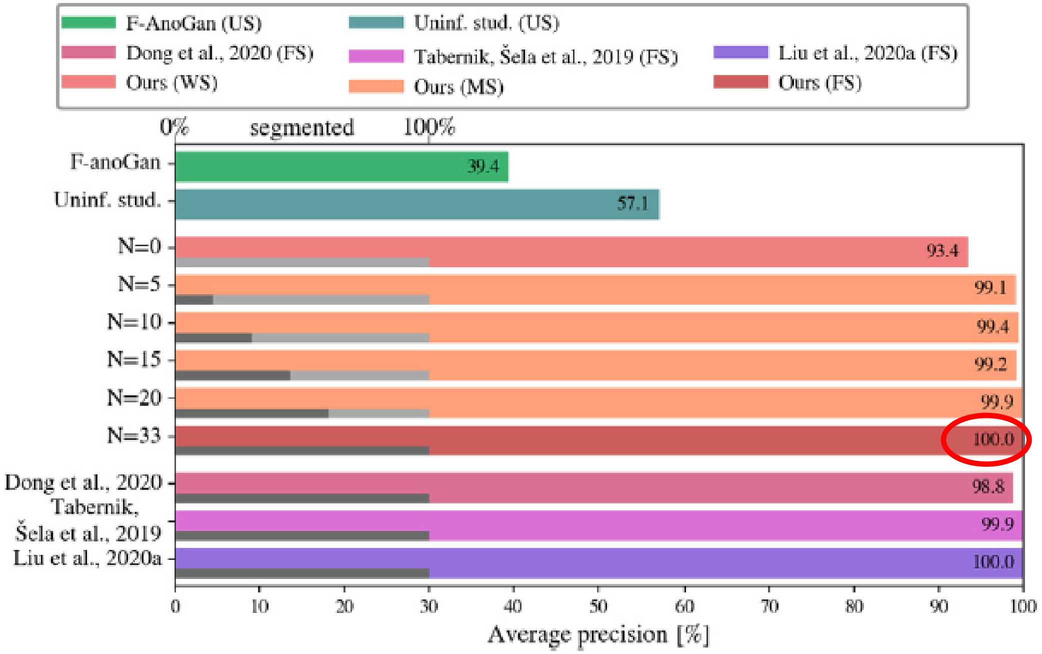


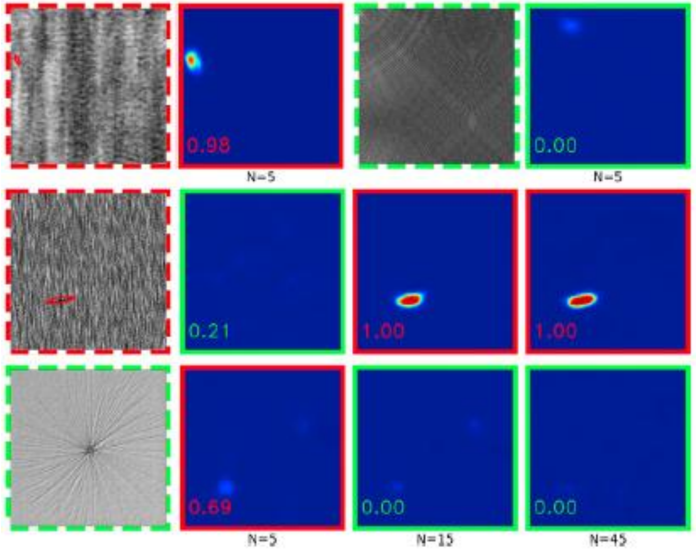
DIVID 2018-2021

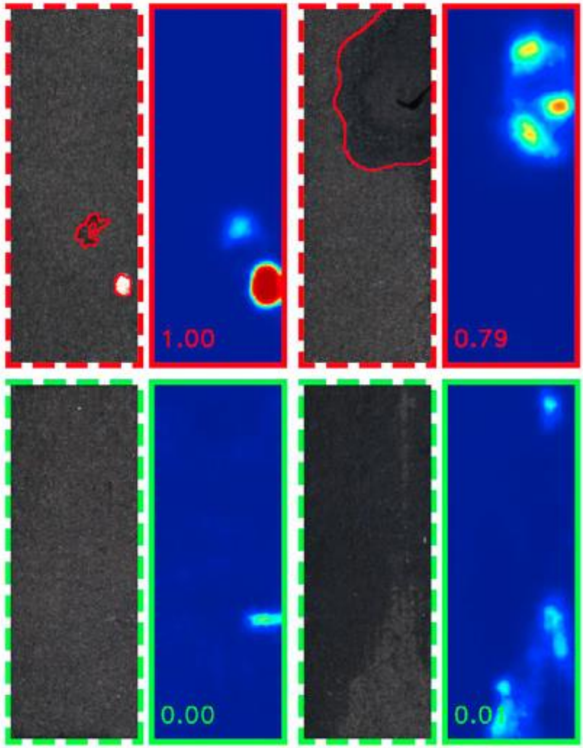
COMIND 2021

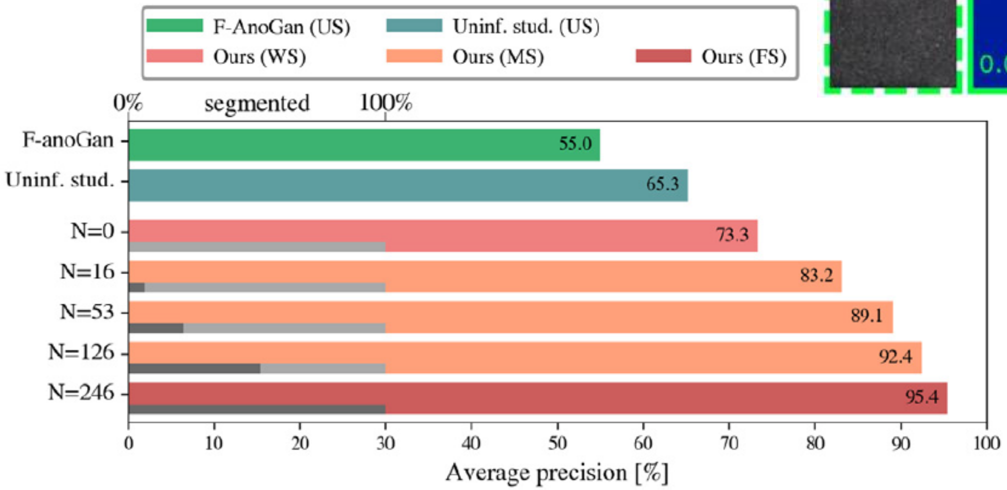
Learning with mixed supervision







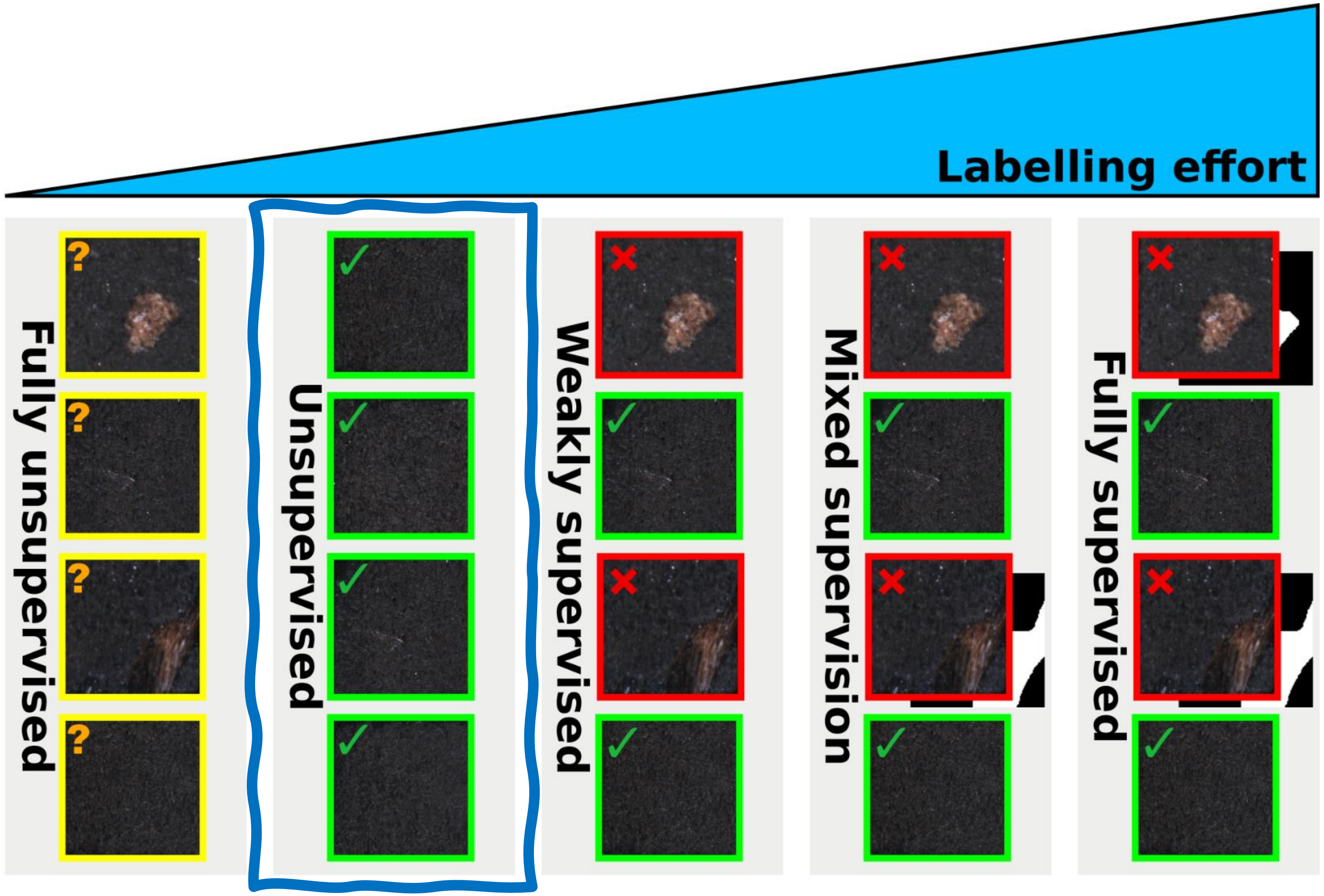




DIVID
2018-2021

COMIND 2021

Learning regimes

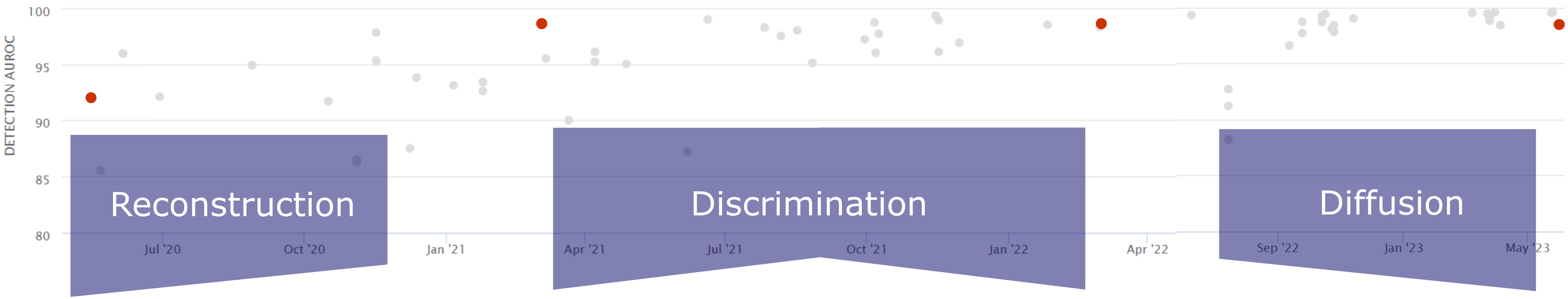


Unsupervised learning

- Only defect-free images required
- Negative-class-only learning

- Detection AUROC on MVTec AD:

[paperswithcode.com]



RIAD
PR 2021

DRAEM
ICCV 2021

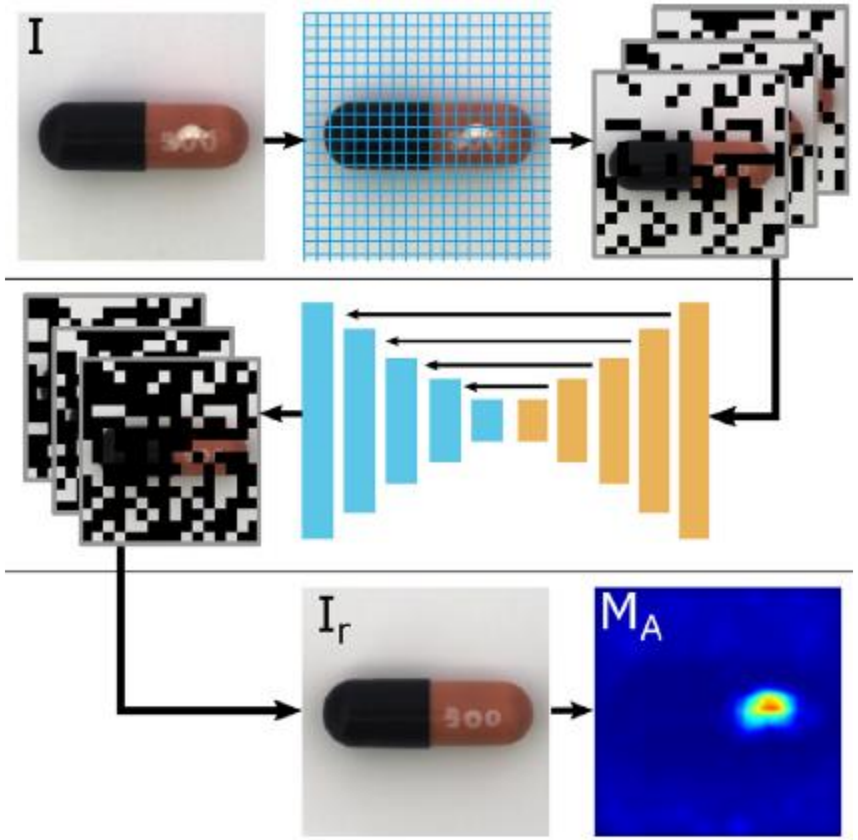
DSR
ECCV 2022

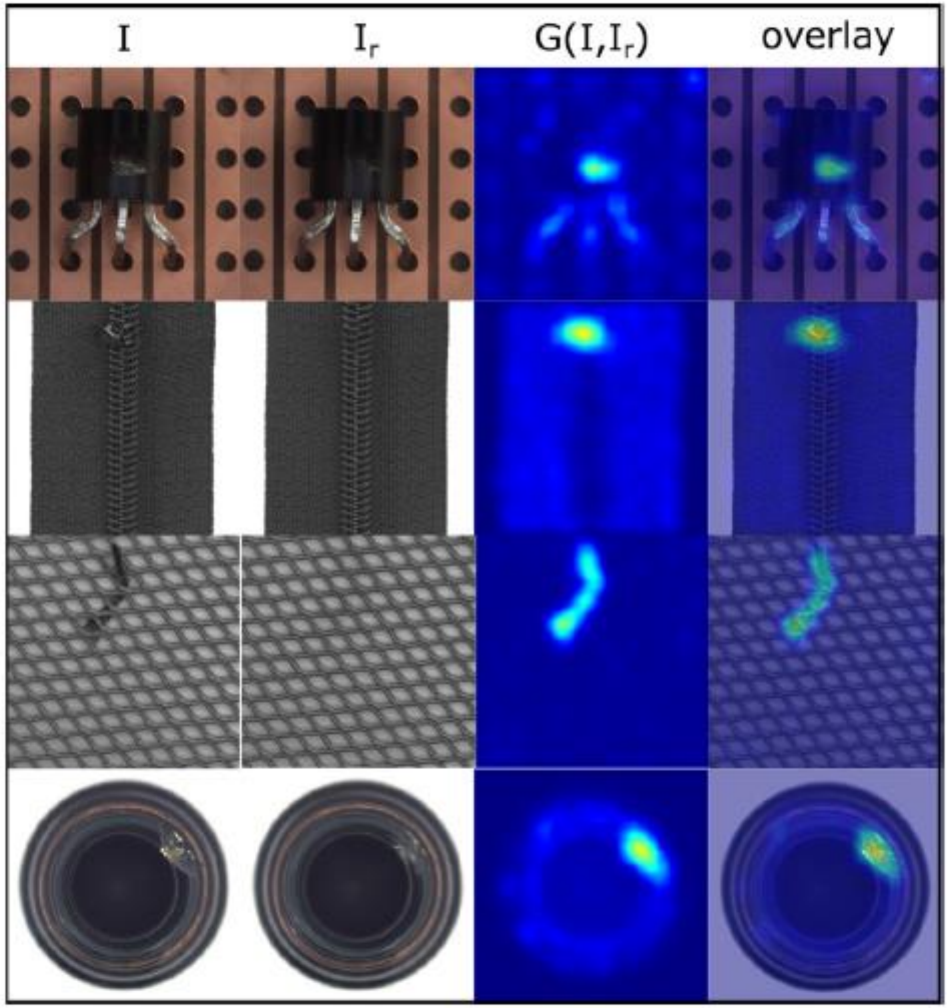
Transfussion
ECCV 2024

Unsupervised learning - RIAD

UNIVERSITY OF LJUBLJANA
FRI
Faculty of Computer and Information Science

- Reconstructive approach



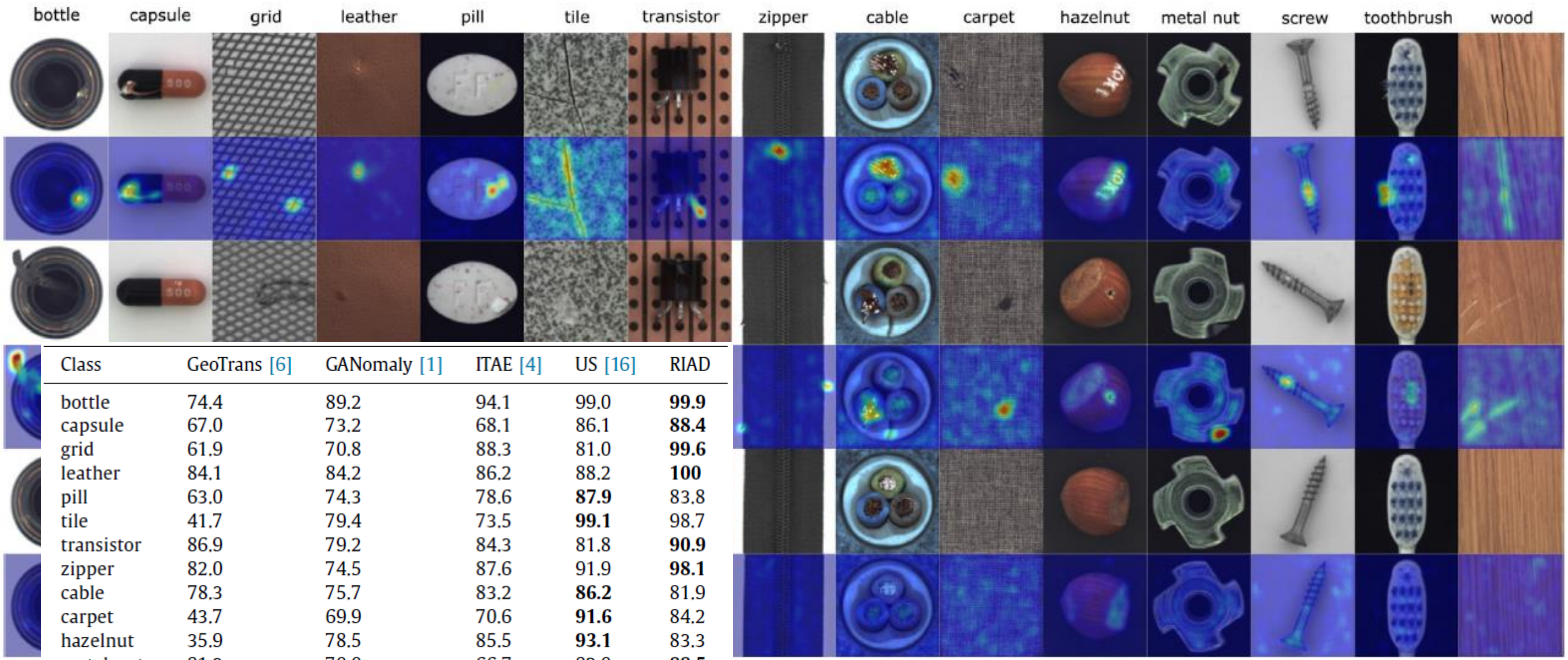


DIVID 2018-2021

CompVis 2019-2024

PR 2021

Unsupervised learning - RIAD



Class	GeoTrans [6]	GANomaly [1]	ITAE [4]	US [16]	RIAD
bottle	74.4	89.2	94.1	99.0	**99.9**
capsule	67.0	73.2	68.1	86.1	**88.4**
grid	61.9	70.8	88.3	81.0	**99.6**
leather	84.1	84.2	86.2	88.2	**100**
pill	63.0	74.3	78.6	**87.9**	83.8
tile	41.7	79.4	73.5	**99.1**	98.7
transistor	86.9	79.2	84.3	81.8	**90.9**
zipper	82.0	74.5	87.6	91.9	**98.1**
cable	78.3	75.7	83.2	**86.2**	81.9
carpet	43.7	69.9	70.6	**91.6**	84.2
hazelnut	35.9	78.5	85.5	**93.1**	83.3
metal nut	81.3	70.0	66.7	82.0	**88.5**
screw	50.0	74.6	**100**	54.9	84.5
toothbrush	97.2	65.3	**100**	95.3	**100**
wood	61.1	83.4	92.3	**97.7**	93.0
avg_{tex}	58.5	76.5	82.2	91.5	**95.1**
avg_{obj}	71.6	75.4	84.8	85.8	**89.9**
avg	67.2	76.2	83.9	87.7	**91.7**

DIVID 2018-2021

CompVis 2019-2024

PR 2021

Discriminative vs. Generative models

- Generative models

> **generative model**
> model built with machine learning that models the distribution of training examples, thereby predicting the probability of occurrence for each individual sample, it is also used for generating new samples similar to the training examples. $P(x), P(x, y)$

- Discriminative models

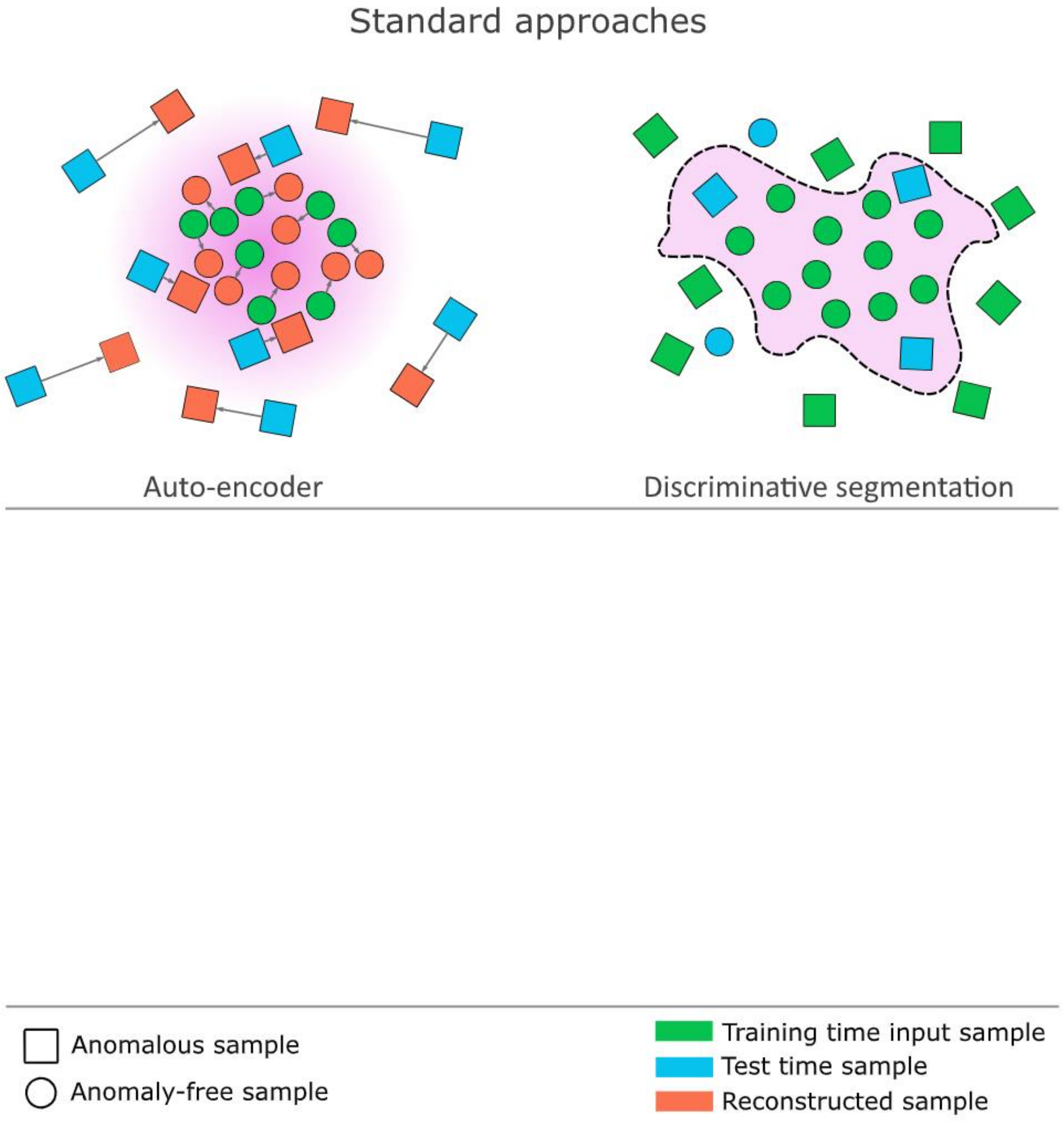
> **discriminative model**
> model, typically built with supervised learning, that models the conditional probability distribution of the target predictive value given the input instance, for example by finding a decision boundary between different classes, it is also used for classification or regression. $P(y|x)$

Unsupervised learning - DRAEM

- Reconstructive models
  - Good approximation of data
  - Unsupervised learning $"P(x)"$
  - General, task-independent
  - Enable reconstruction and outlier detection

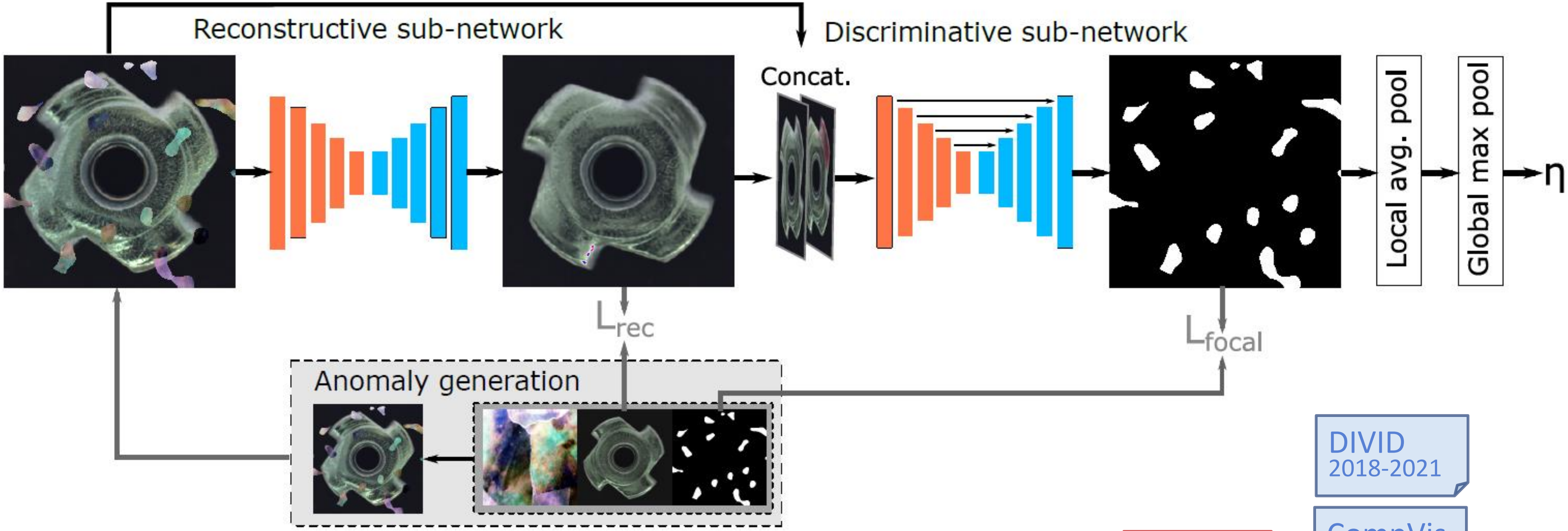
- Discriminative models
  - Supervised learning
  - Task-dependent $P(y|x)$
  - Compact representations
  - No reconstruction
  - Outlier detection as classification

⇨ Combine reconstructive model and discriminative classifier



Unsupervised learning - DRAEM

- Reconstructive and discriminative approach
- Generate synthetic anomalies

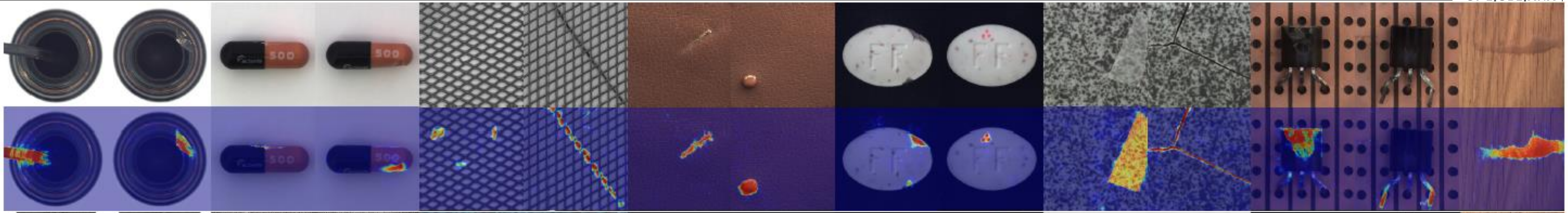


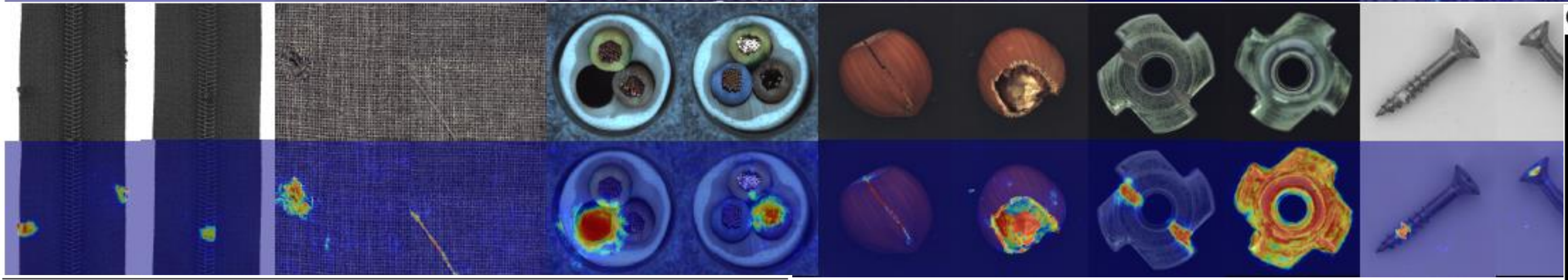
DIVID 2018-2021

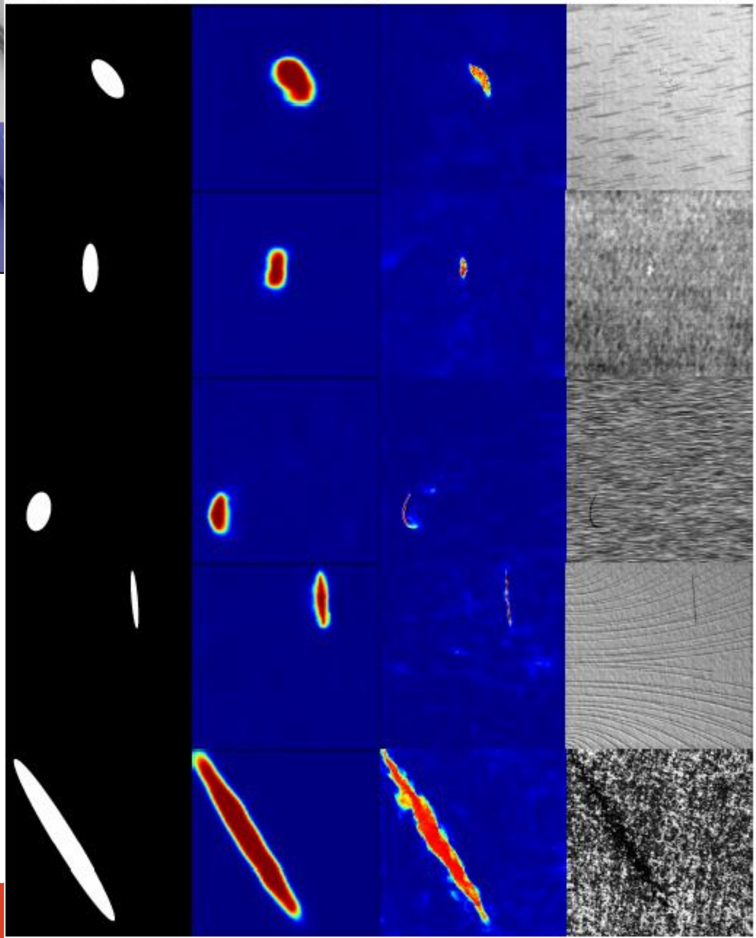
CompVis 2019-2024

ICCV 2021

Unsupervised learning - DRAEM







Class	[1]	[26]	[4]	[31]	[20]	[11]	DRÆM
bottle	79.4	98.3	99.0	99.9	**100**	99.9	99.2
capsule	72.1	68.7	86.1	88.4	92.3	91.3	**98.5**
grid	74.3	86.7	81.0	99.6	92.9	96.7	**99.9**
leather	80.8	94.4	88.2	**100**	**100**	**100**	**100**
pill	67.1	76.8	87.9	83.8	83.4	93.3	**98.9**
tile	72.0	96.1	99.1	98.7	97.4	98.1	**99.6**
transistor	80.8	79.4	81.8	90.9	95.9	**97.4**	93.1
zipper	74.4	78.1	91.9	98.1	97.9	90.3	**100**
cable	71.1	66.5	86.2	81.9	**94.0**	92.7	91.8
carpet	82.1	90.3	91.6	84.2	95.5	**99.8**	97.0
hazelnut	87.4	100	93.1	83.3	98.7	92.0	**100.0**
metal nut	69.4	81.5	82.0	88.5	93.1	**98.7**	**98.7**
screw	**100**	**100**	54.9	84.5	81.2	85.8	93.9
toothbrush	70.0	95.0	95.3	**100**	95.8	96.1	**100**
wood	92.0	97.9	97.7	93.0	97.6	**99.2**	99.1
avg	78.2	87.3	87.7	91.7	94.4	95.5	**98.0**

DIVID
2018-2021

CompVis
2019-2024

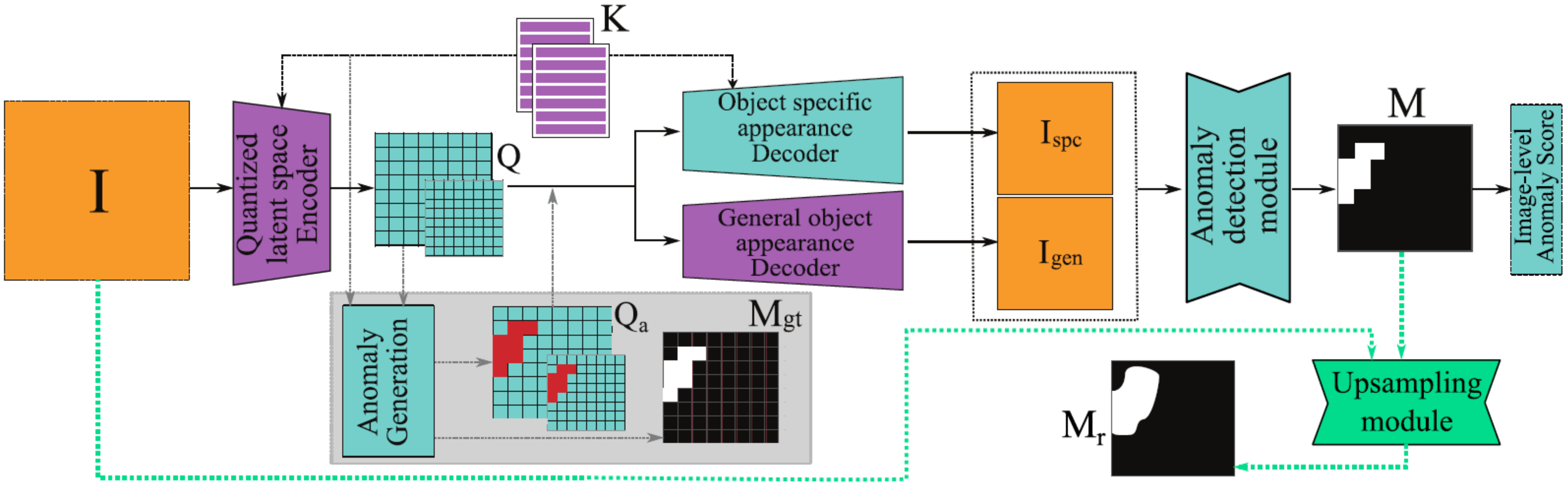
ICCV 2021

	Methods	AUROC	TPR	TNR	CA
Unsup.	RIAD [31]	78.6	79.2	69.1	70.4
	US [4]	72.5	72.6	65.3	66.2
	MAD [20]	82.4	78.7	85.7	66.2
	PaDim [11]	95.0	83.3	97.5	95.7
	DRÆM	**99.0**	**96.5**	**99.4**	**98.5**
Sup.	CADN [32]	-	-	-	89.1
	Rački *et al.* [19]	99.6	99.9	99.5	-
	Lin *et al.* [15]	99.0	99.4	99.9	-
	Božič *et al.* [6]	**100**	**100**	**100**	**100**

Unsupervised learning - DSR

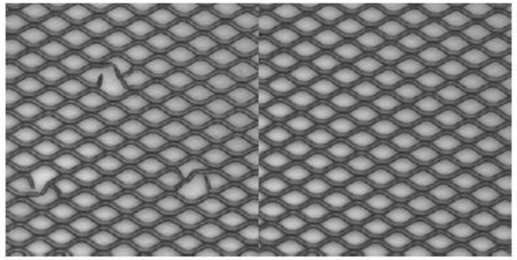
UNIVERSITY OF LJUBLJANA
FRI Faculty of Computer and Information Science

- Generate syntetic anomalies in the quantized feature space

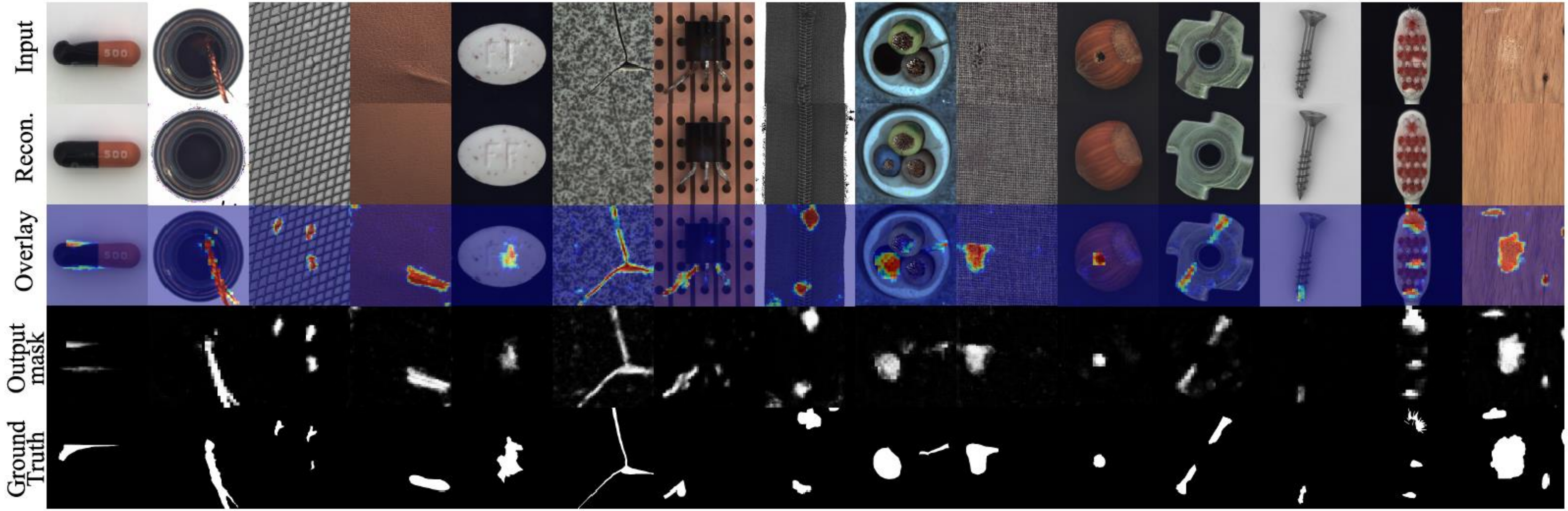


ECCV 2022

MV4.0
2021-2024



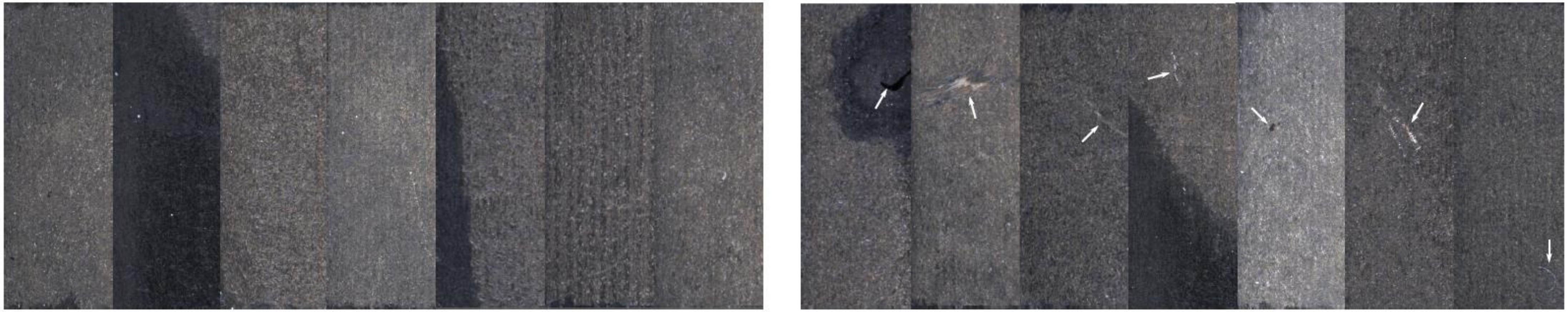
Unsupervised learning - DSR

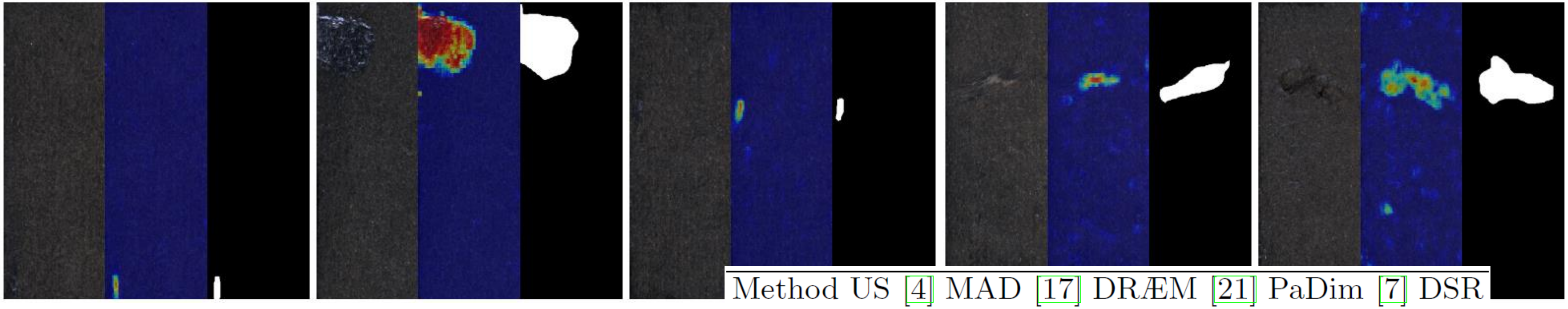


Method	bottle	capsule	grid	leather	pill	tile	trans.	zipper	cable	carpet	hazelnut	m. nut	screw	toothbrush	wood	average
[4]	99.0	86.1	81.0	88.2	87.9	99.1	81.8	91.9	86.2	91.6	93.1	82.0	54.9	95.3	97.7	87.7
[22]	99.9	88.4	99.6	**100**	83.8	98.7	90.9	98.1	81.9	84.2	83.3	88.5	84.5	**100**	93.0	91.7
[17]	**100**	92.3	92.9	**100**	83.3	97.4	95.9	97.9	**94.0**	95.5	98.7	93.1	81.2	95.8	97.6	94.4
[7]	99.8	91.5	95.7	**100**	94.4	97.4	**97.8**	90.9	92.2	99.9	93.3	99.2	84.4	97.2	98.8	95.5
[11]	98.2	98.2	**100**	**100**	94.9	94.6	96.1	99.9	81.2	93.9	98.3	**99.9**	88.7	99.4	**99.1**	96.1
[21]	99.2	**98.5**	99.9	**100**	**98.9**	99.6	93.1	**100**	91.8	97.0	**100**	98.7	93.9	**100**	**99.1**	98.0
DSR	**100**	98.1	**100**	**100**	97.5	**100**	**97.8**	**100**	93.8	**100**	95.6	98.5	**96.2**	99.7	96.3	**98.2**

Unsupervised learning - DSR

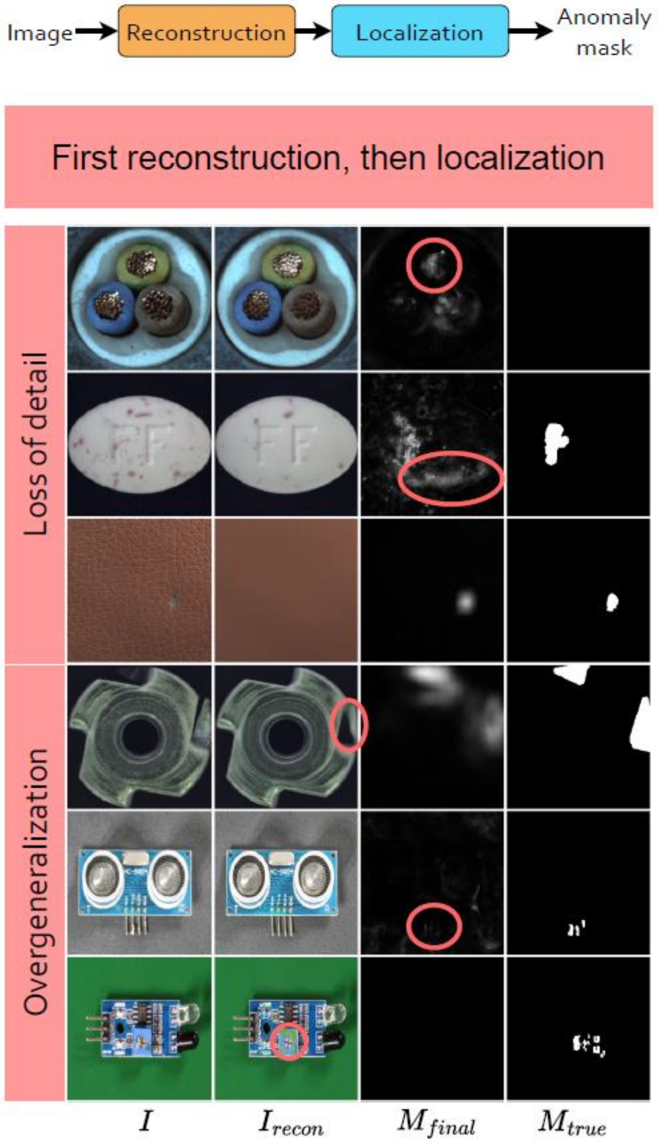
- Results on KSDD2

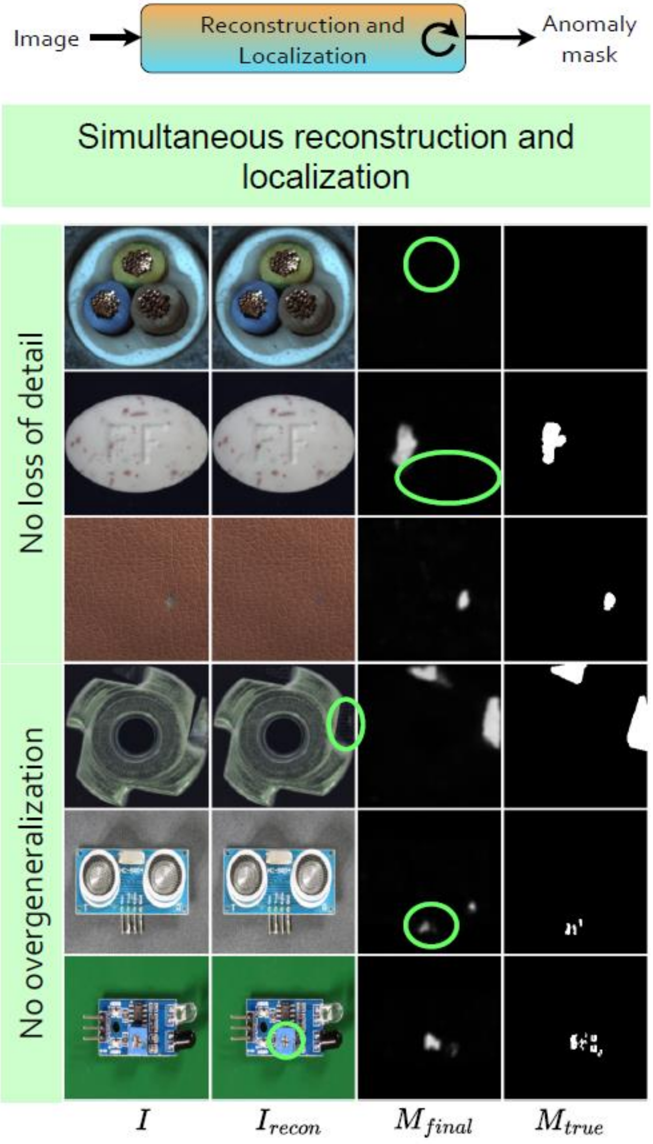




Method	US [4]	MAD [17]	DRÆM [21]	PaDim [7]	DSR
AP_{det}	65.3	79.3	77.8	55.6	**87.2**
AP_{loc}	-	-	42.4	45.3	**61.4**

Unsupervised learning - Transfusion





MV4.0
2021-2024

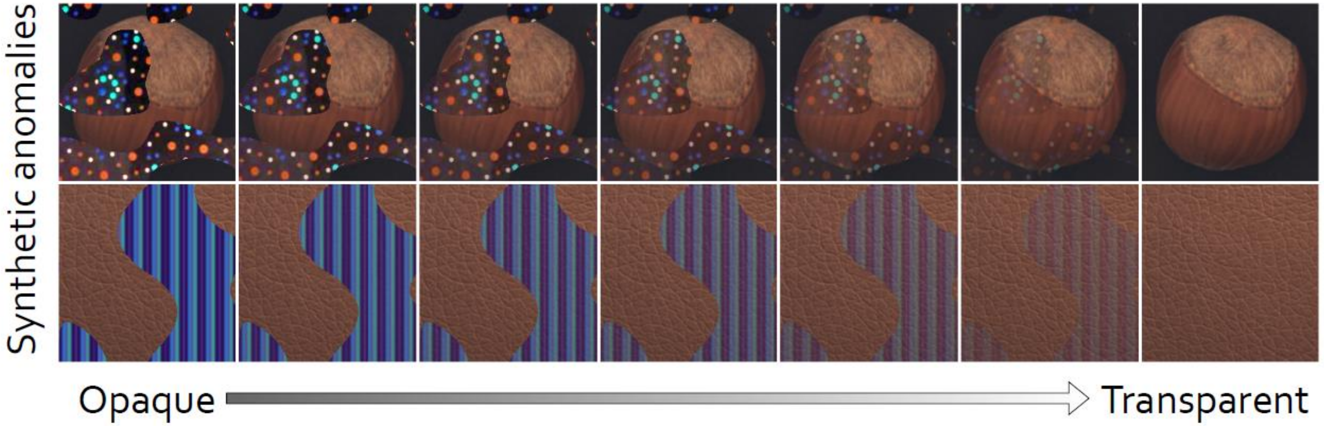
ECCV 2024

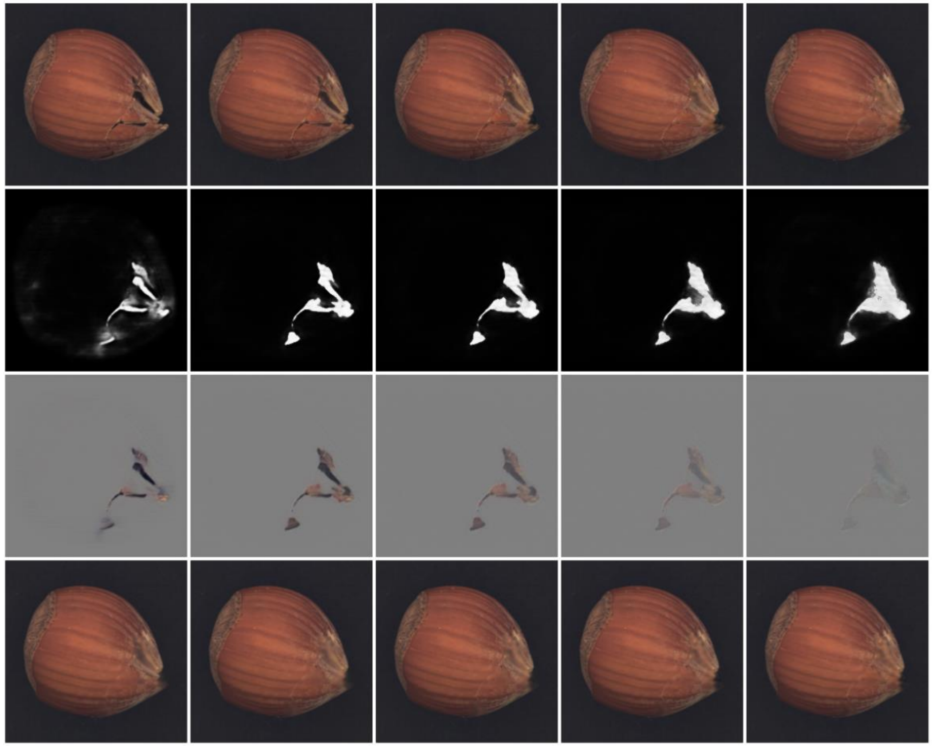
Unsupervised learning - Transfusion

- TRANSparent difFUSION
- Using Diffusion model estimate
  - Anomaly mask
  - Anomaly
  - Normal image

MV4.0
2021-2024

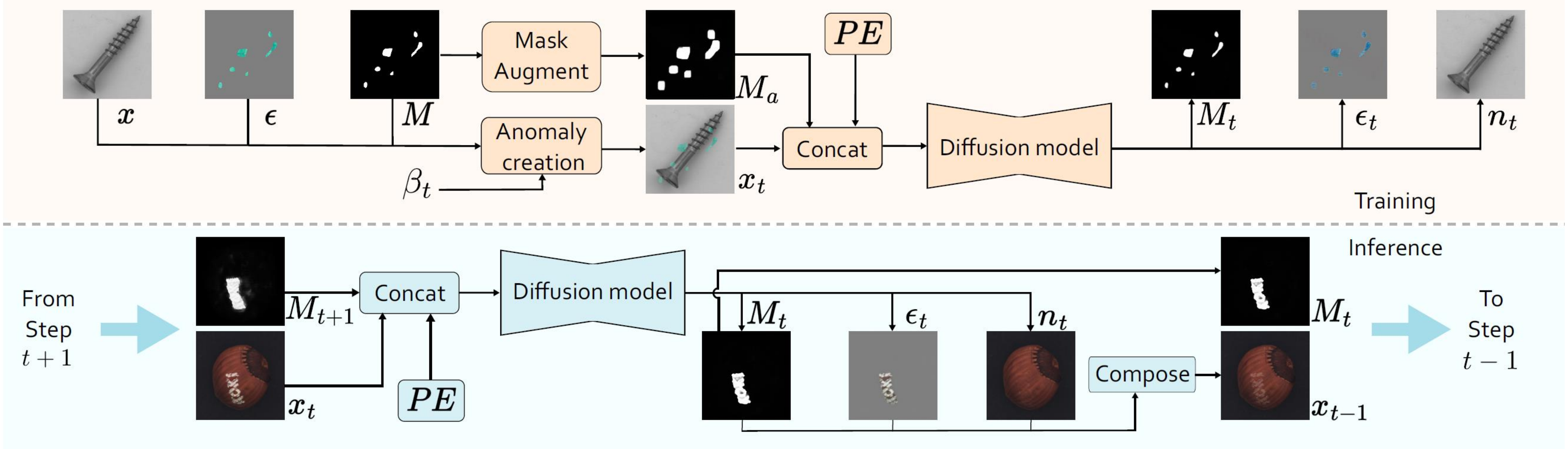
ECCV 2024

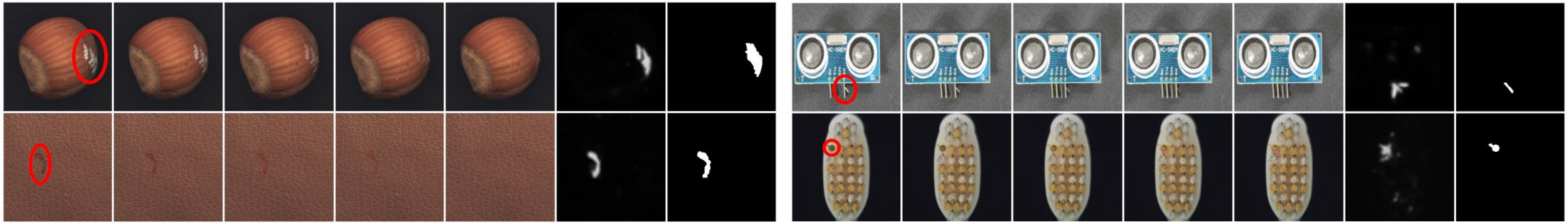




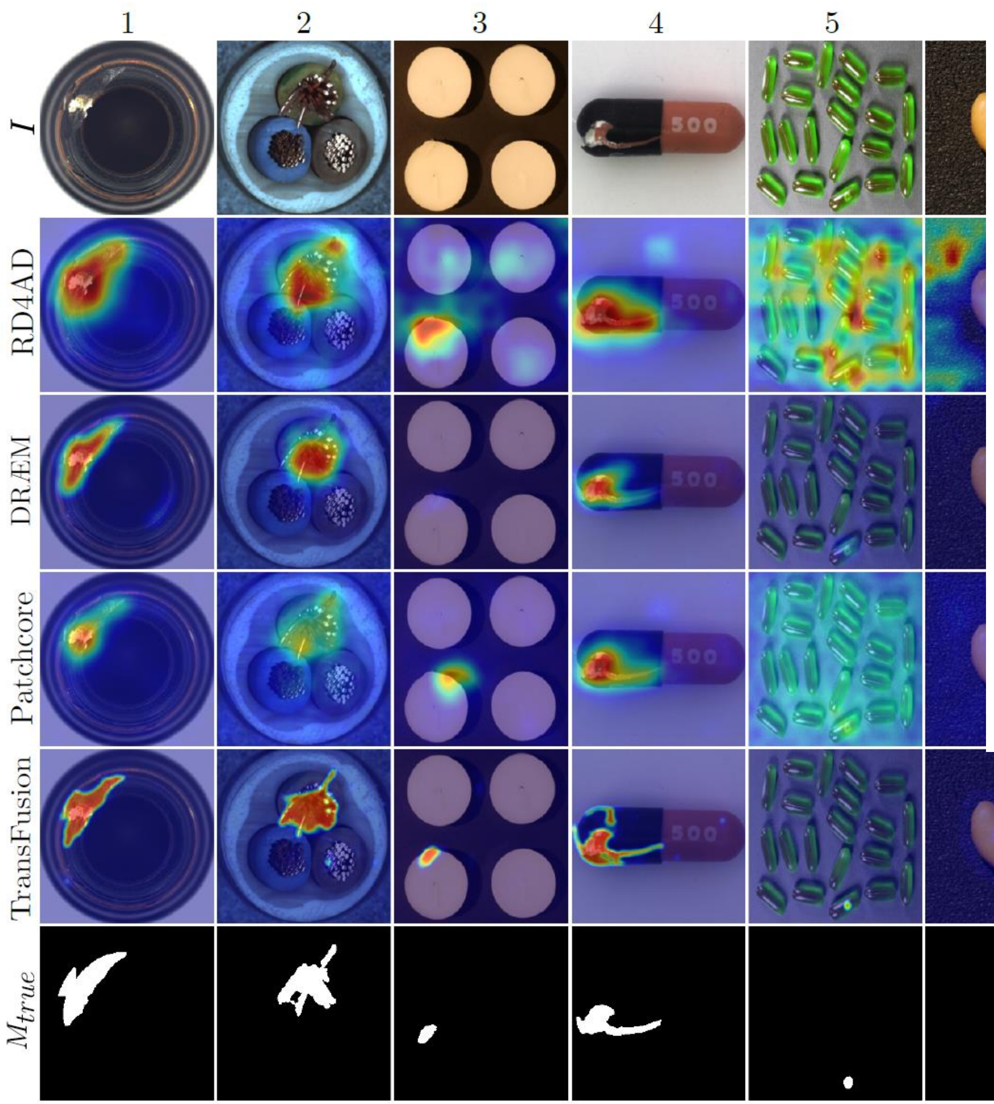
$$x_{t-1} = x_t - (\beta_t - \beta_{t-1})(M_t \odot \epsilon_t) + (\beta_t - \beta_{t-1})(M_t \odot \hat{x_0}^{(t)})$$

Unsupervised learning - Transfusion

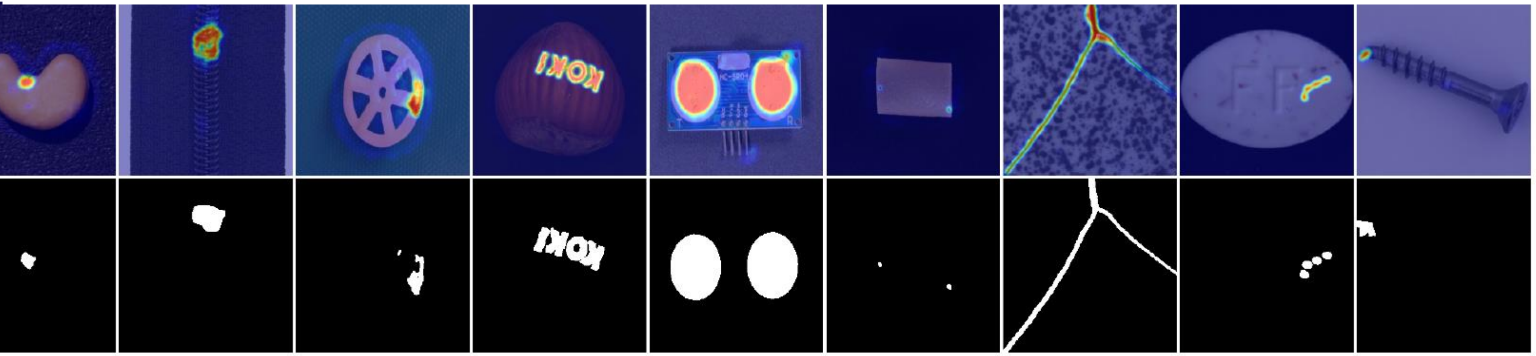




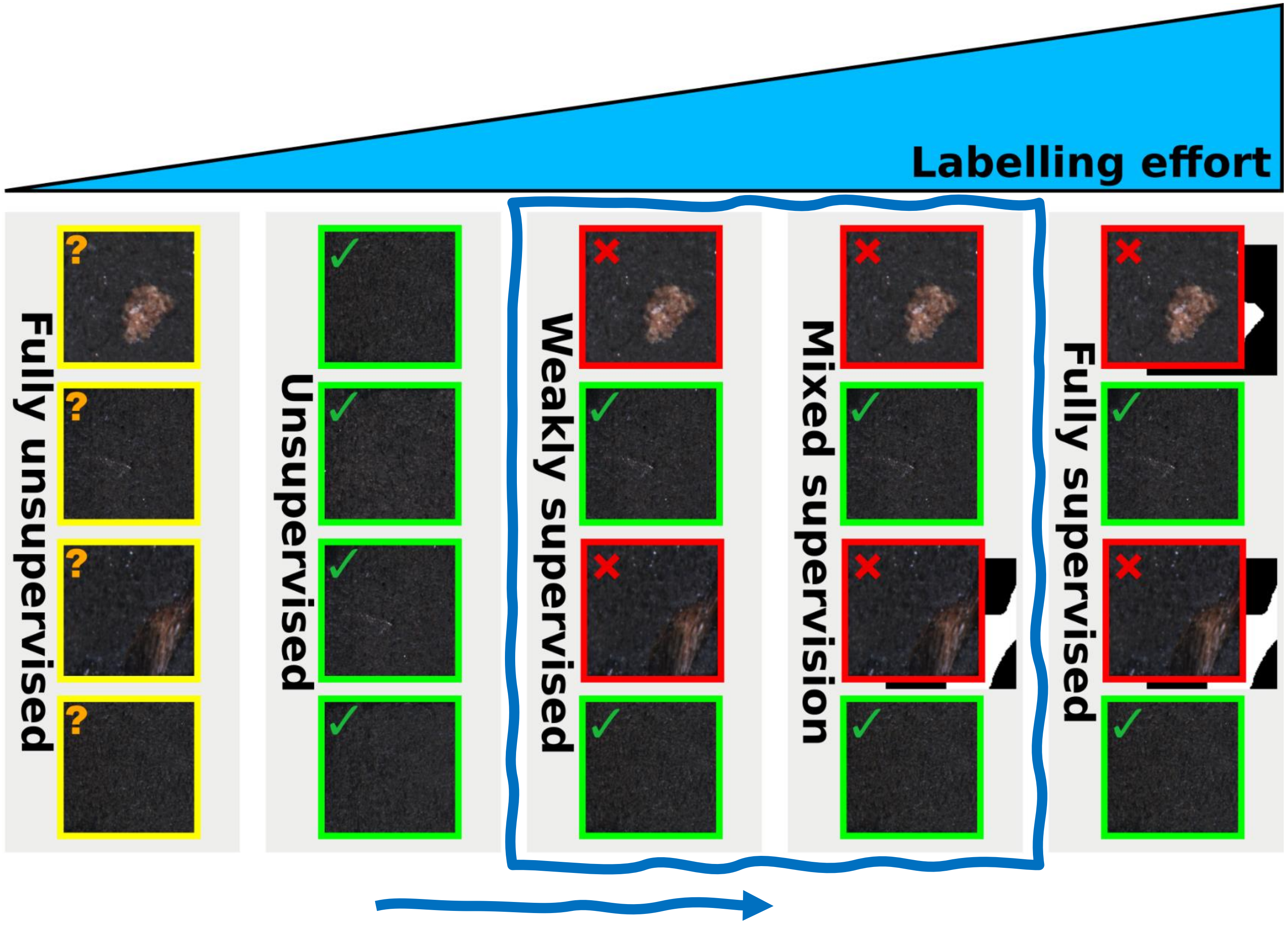
Unsupervised learning - Transfusion



Method	Venue	VisA		MVTec AD		*Average*	
		Det.	Loc.	Det.	Loc.	Det.	Loc.
AnoDDPM	CVPRW'22	78.2	60.5	83.5	50.7	80.9	55.6
DRÆM	ICCV'21	88.7	73.1	98.0	92.8	93.3	83.0
SimpleNet	CVPR'23	87.9	68.9	**99.6**	89.6	93.8	79.3
DiffAD	ICCV'23	89.5	71.2	98.7	84.8	94.1	78.0
DSR	ECCV'22	91.6	68.1	98.2	90.8	94.9	79.5
FastFlow	ArXiv'21	93.9	86.9	99.4	92.5	96.7	89.7
Patchcore	CVPR'22	94.3	79.7	99.1	92.7	97.0	86.2
AST	WACV'23	94.9	81.5	99.2	81.2	97.1	81.4
RD4AD	CVPR'22	96.0	70.9	98.5	93.9	97.3	82.4
TransFusion	-	**98.5**	**88.8**	99.2	**94.3**	**98.9**	**91.6**

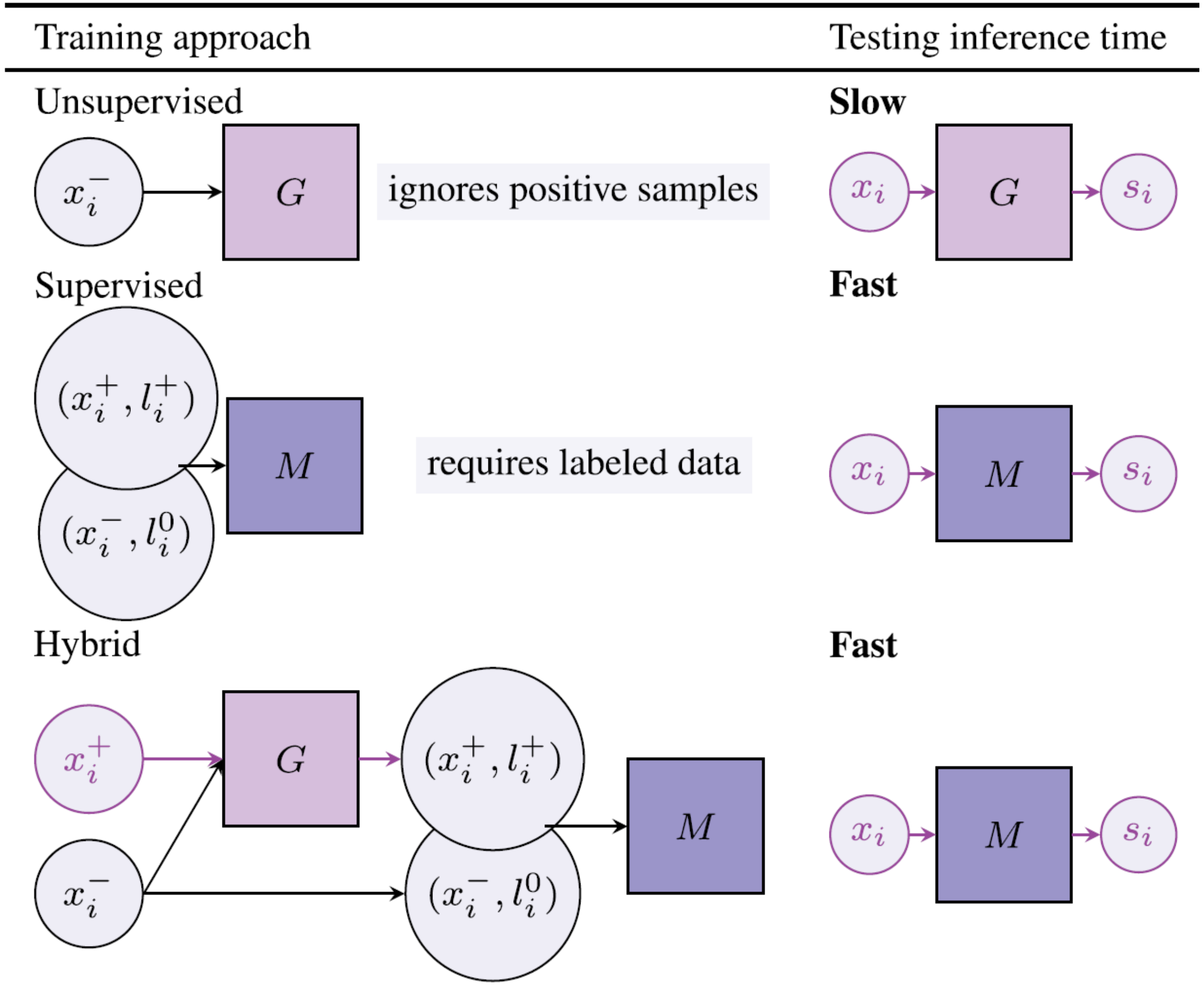


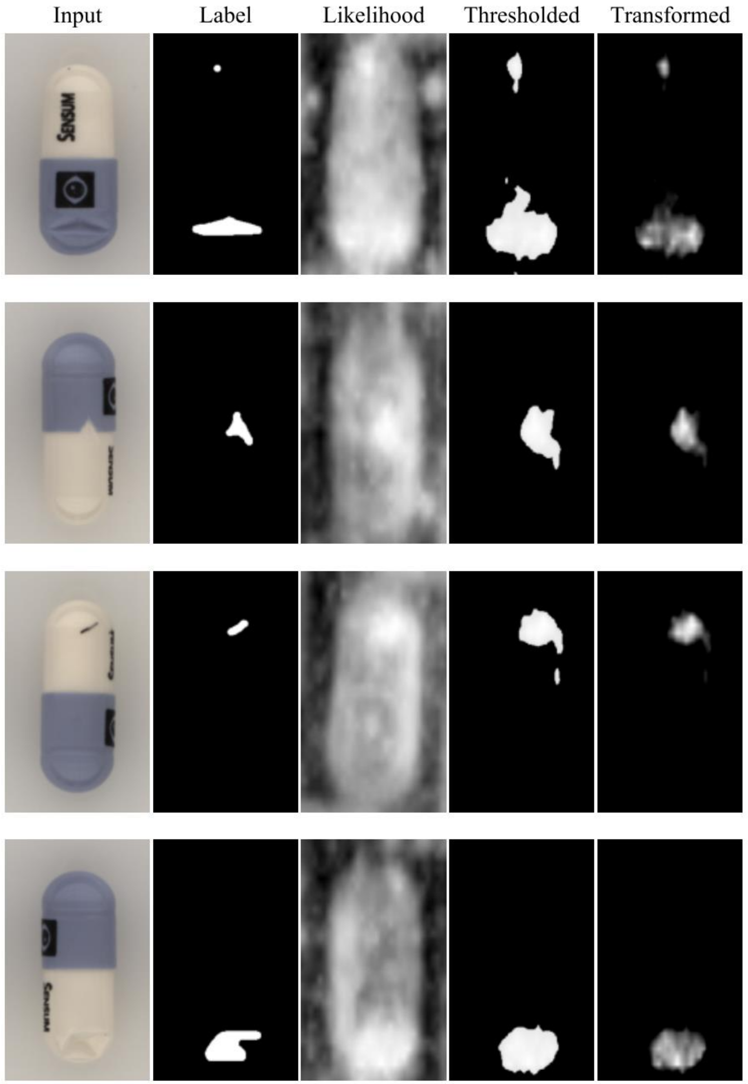
Learning regimes



Learning with mixed supervision

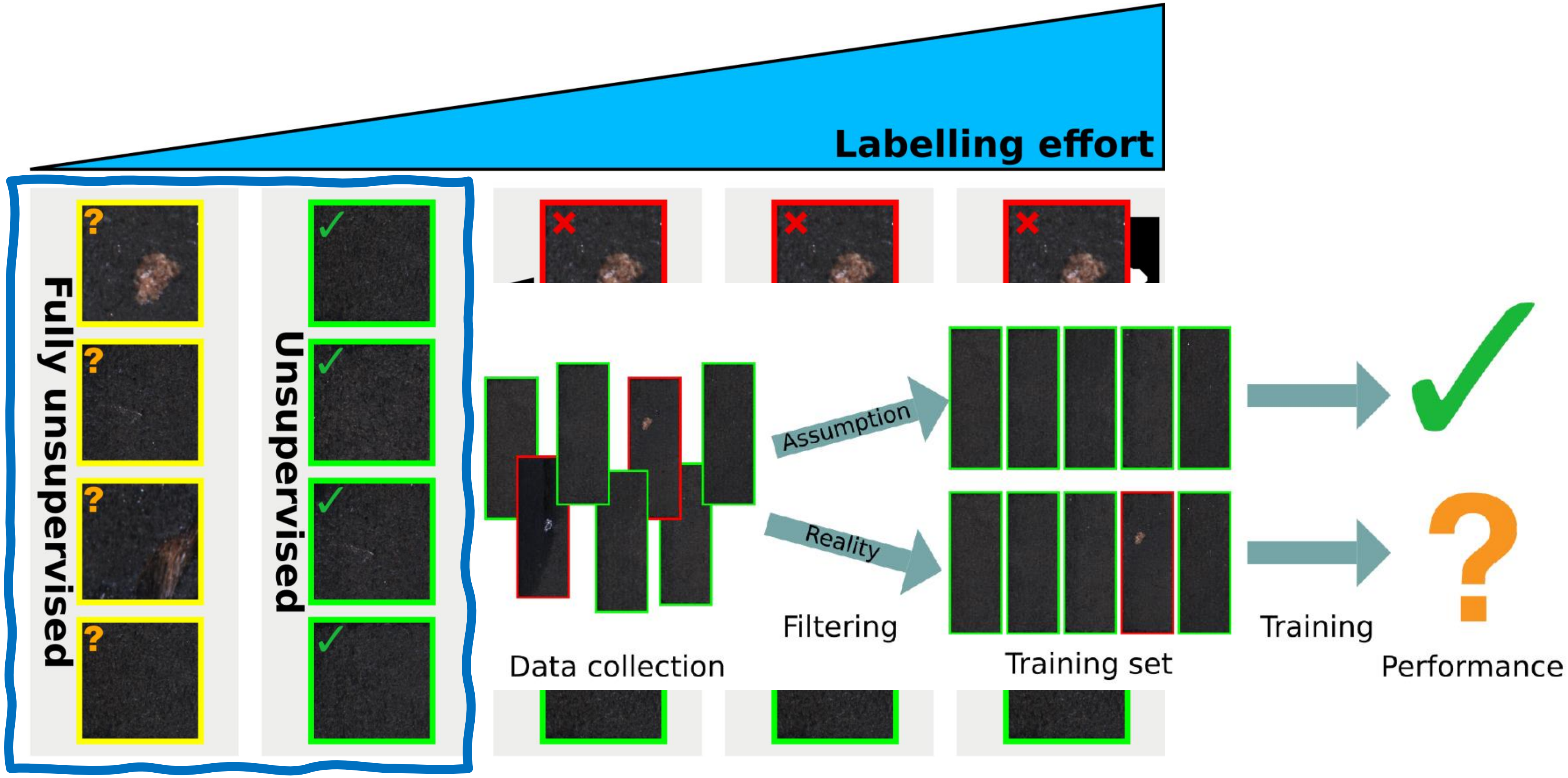
UNIVERSITY OF LJUBLJANA | FRI Faculty of Computer and Information Science



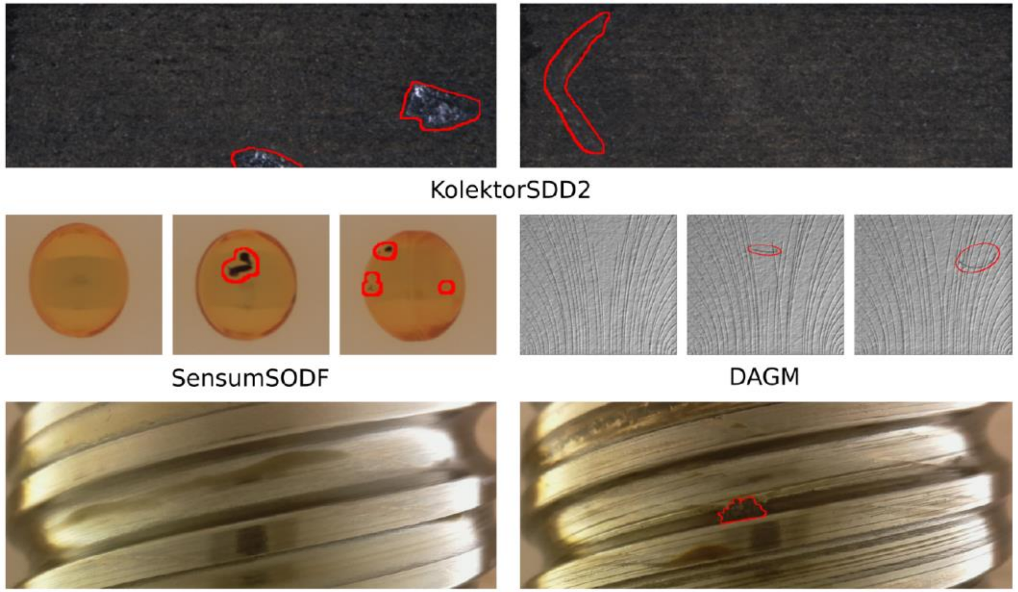


JEI 2024

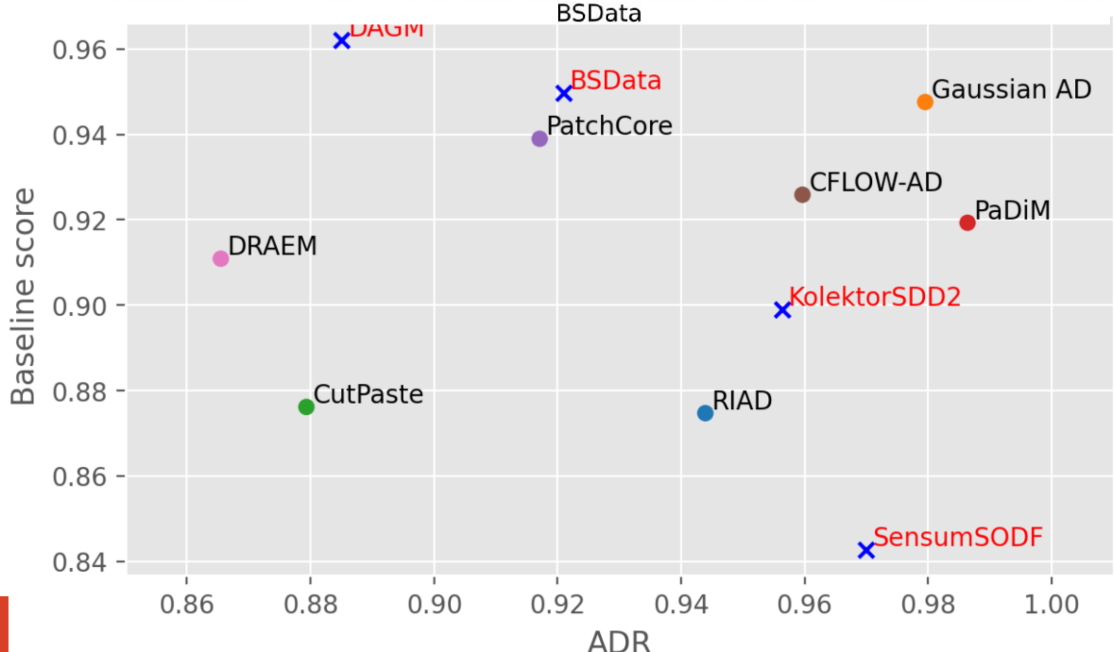
Learning regimes

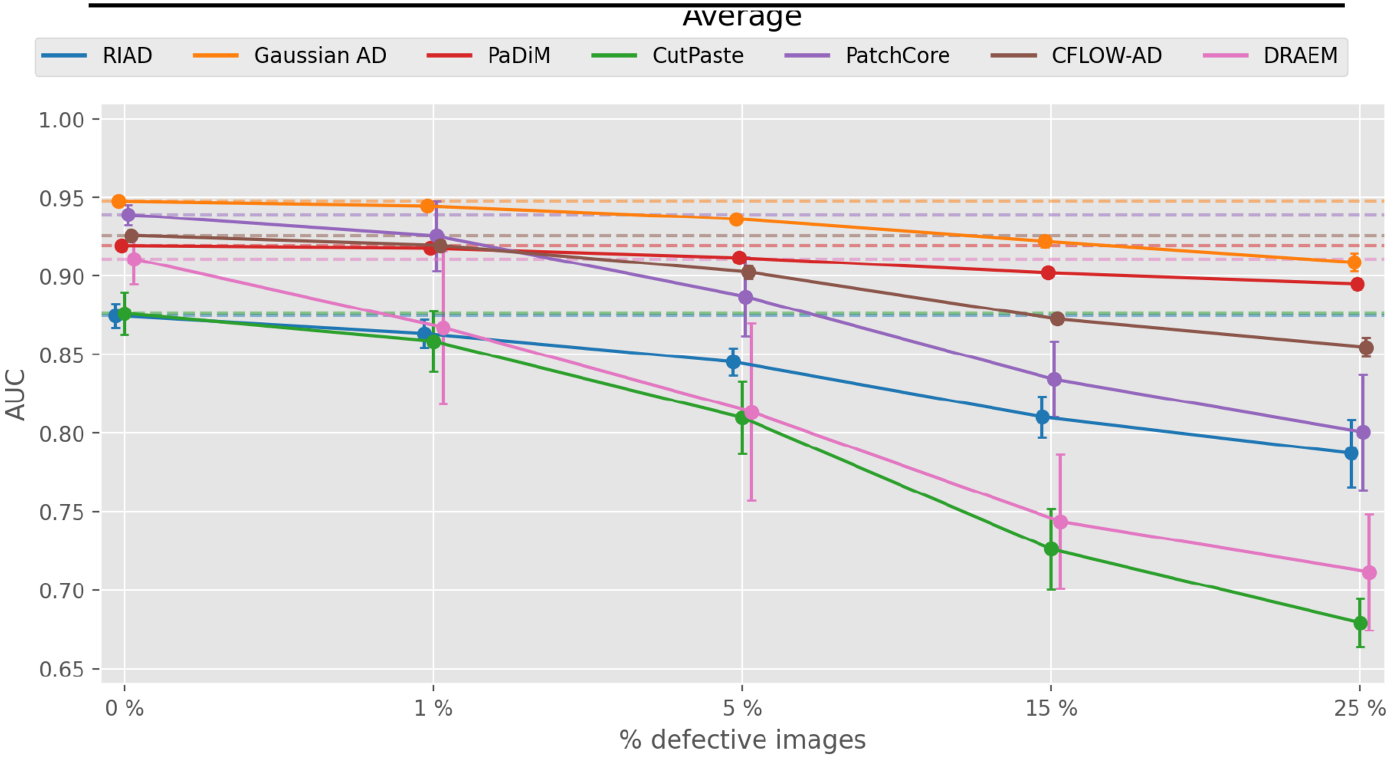


Robustness of unsupervised methods


KolektorSDD2
SensumSODF
DAGM
BSData

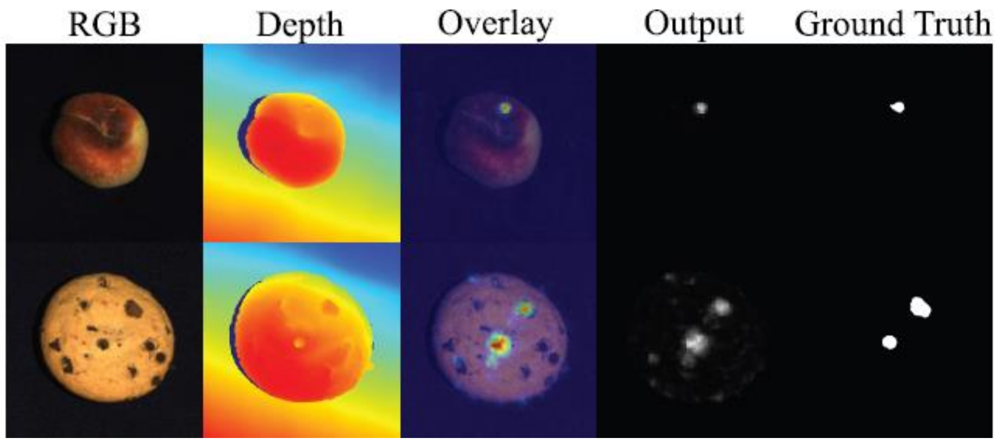
Method\Dataset	KolektorSDD2	SensumSODF	DAGM	BSData	mean
RIAD	0.952	0.992	0.870	0.962	0.944
Gaussian AD	0.989	0.986	0.965	0.978	0.980
PaDiM	0.999	0.996	1.000	0.951	0.986
CutPaste	0.888	0.925	0.837	0.866	0.879
PatchCore	0.995	0.954	0.938	0.782	0.917
CFLOW-AD	0.995	0.986	0.891	0.966	0.960
DRAEM	0.876	0.950	0.694	0.941	0.865
mean	0.956	0.970	0.885	0.921	/

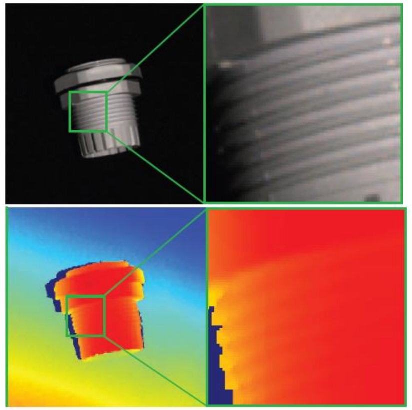


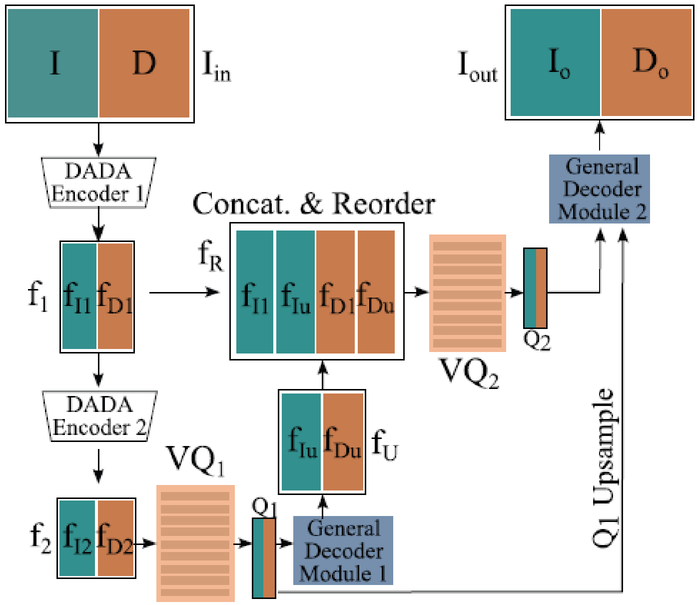


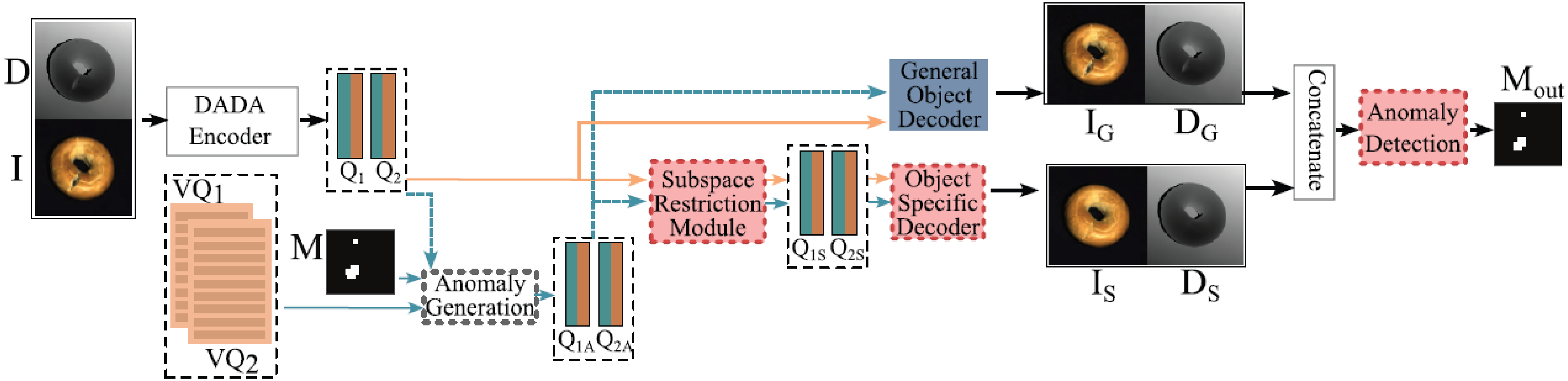
Beyond images – 3D: 3DSR

- RGB + D data
- Depth-aware descrete autoencoder (DADA)







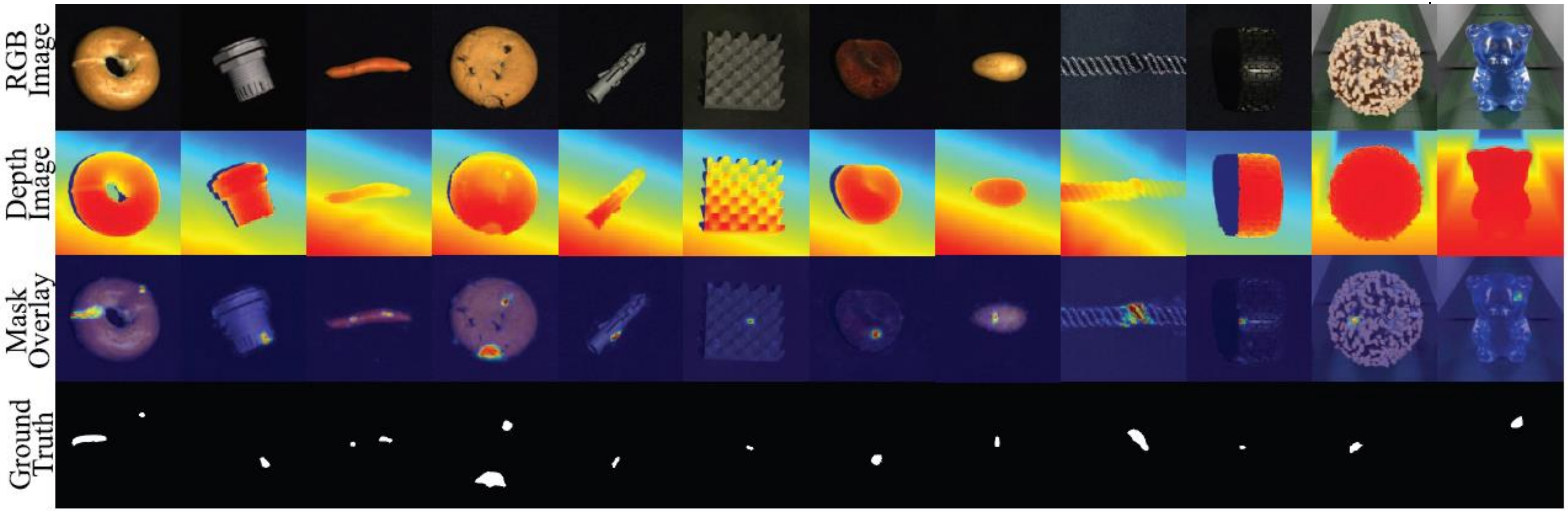


MV4.0
2021-2024

WACV 2024

3DSR results

UNIVERSITY OF LJUBLJANA
FRI Faculty of Computer and Information Science

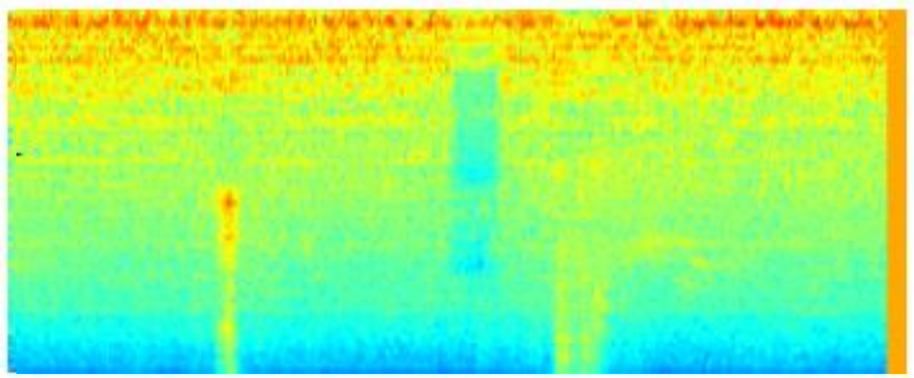


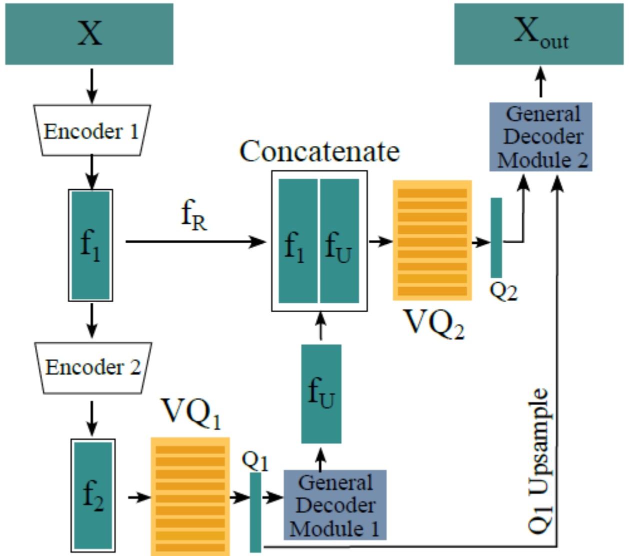
3D+RGB	Voxel AE [2]	51.0	54.0	38.4	69.3	44.6	63.2	55.0	49.4	72.1	41.3	53.8
	Depth GAN [2]	53.8	37.2	58.0	60.3	43.0	53.4	64.2	60.1	44.3	57.7	53.2
	Depth AE [2]	64.8	50.2	65.0	48.8	80.5	52.2	71.2	52.9	54.0	55.2	59.5
	PatchCore+FPFH [9]	91.8	74.8	96.7	88.3	93.2	58.2	89.6	91.2③	92.1	88.6③	86.5
	AST [17]	98.3②	87.3②	97.6②	97.1③	93.2③	88.5③	97.4②	98.1①	100①	79.7	93.7③
	M3DM [19]	99.4①	90.9①	97.2③	97.6②	96.0②	94.2②	97.3③	89.9	97.2③	85.0③	94.5②
	3DSR	98.1③	86.7③	99.6①	98.1①	100①	99.4①	98.6①	97.8②	100①	99.5①	97.8①

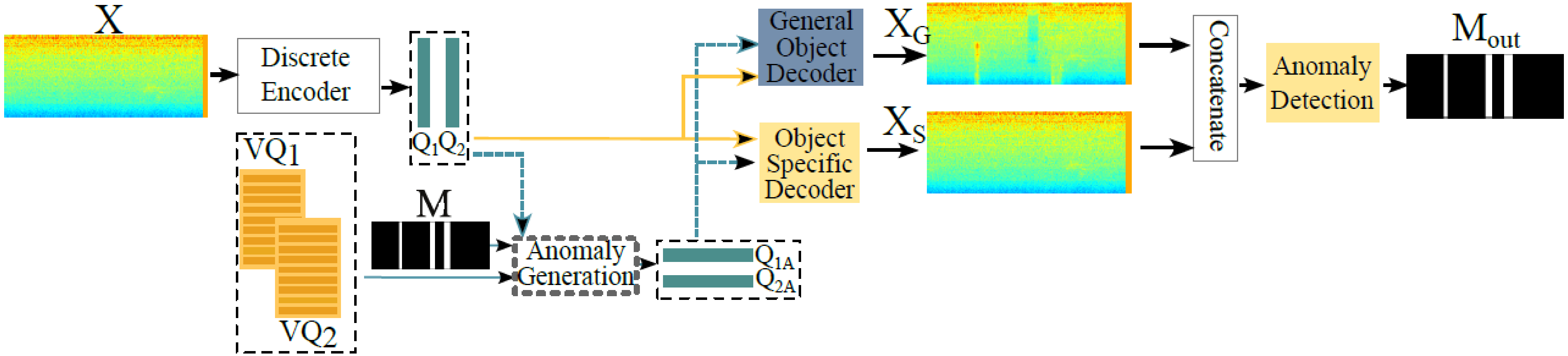
Beyond images – audio: AudDSR

UNIVERSITY OF LJUBLJANA | FRI Faculty of Computer and Information Science

- Unsupervised anomaly detection in audio
- Processing MEL spectrogram







MV4.0
2021-2024

ICASSP 2024

AuDSR results

Methods	ToyCar AUC	ToyCar pAUC	ToyConveyor AUC	ToyConveyor pAUC	Fan AUC	Fan pAUC	Pump AUC	Pump pAUC	Slider AUC	Slider pAUC	Valve AUC	Valve pAUC	Average AUC	Average pAUC
Annotation-free														
AE [1]	80.90	69.90	73.40	61.10	66.20	53.20	72.90	60.30	85.50	67.80	66.30	51.20	74.20	60.58
IDNN [13]	80.19	71.87	75.74	61.26	69.15	53.53	74.06	61.26	88.32	69.07	88.31	65.67	79.30	63.78
ANP [16]	72.50	67.30	67.00	54.50	69.20	54.40	72.80	61.80	90.70	74.20	86.90	70.70	76.52	63.82
PAE [5]	75.35	69.70	77.58	61.37	72.94	54.37	74.27	62.01	91.92	74.39	95.41	81.24	81.25	67.18
AudDSR	**91.89**	**82.90**	**78.02**	**64.60**	**73.82**	**64.98**	**85.91**	**74.32**	90.16	71.54	90.05	70.20	**84.97**	**71.45**
Annotation-reliant														
MobileNetV2 [11]	87.66	85.92	69.71	56.43	80.19	74.40	82.53	76.50	95.27	85.22	88.65	87.98	84.00	77.74
GlowAff [6]	92.20	84.10	71.50	59.00	74.90	65.30	83.40	73.80	94.60	82.80	91.40	75.00	85.20	73.90
STgram [12]	88.80	87.38	72.93	63.62	91.30	86.73	91.25	81.69	99.36	96.84	94.44	91.58	89.68	84.64
AudDSR$_{annot}$	93.60	90.65	81.57	71.23	77.46	75.39	88.52	79.16	98.56	93.00	98.90	94.76	90.12	84.59
GeCo [4]	96.62	89.33	74.69	65.82	92.73	85.19	93.09	86.89	98.61	95.26	99.06	95.52	92.47	86.34

Use all data available!

Practical considerations

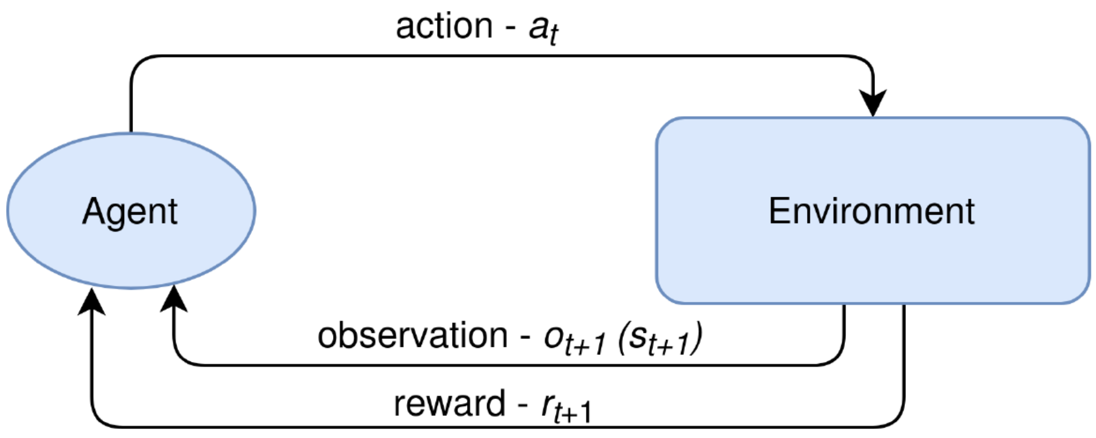
Built-in
vs.
learned

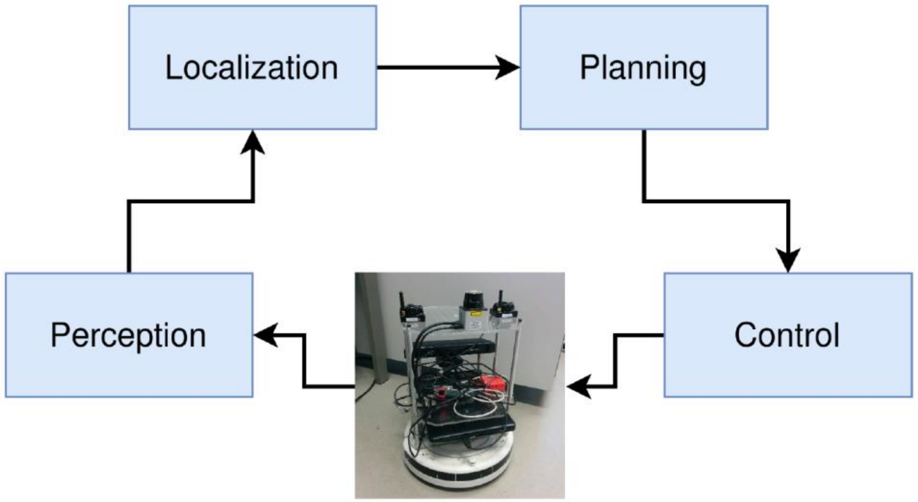
MV
vs.
DL

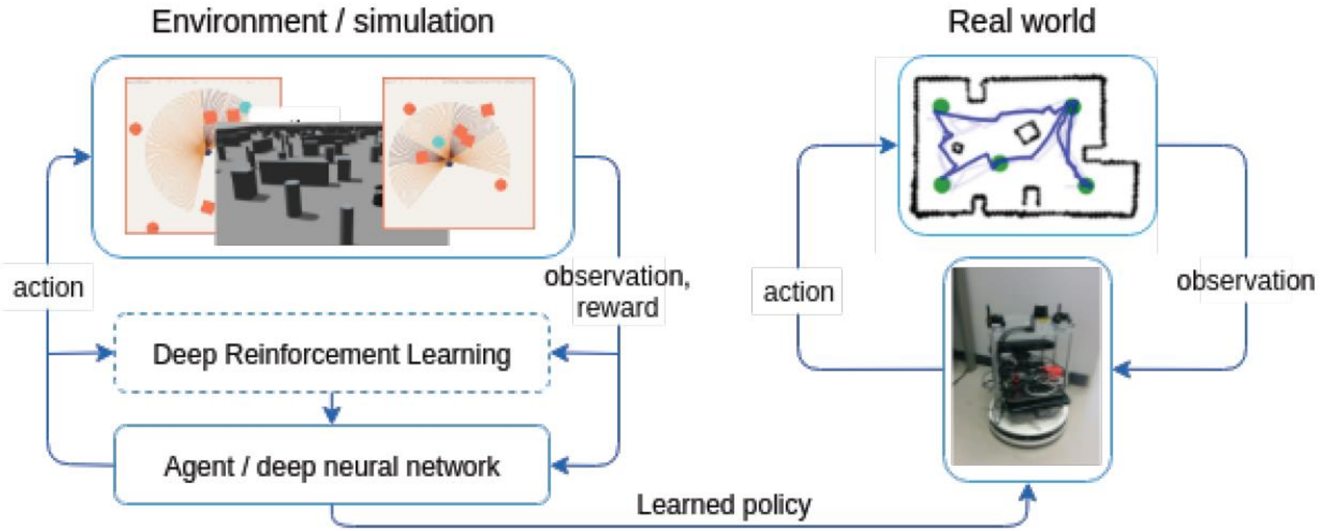
Development
and
mainainance

Deep reinforcement learning

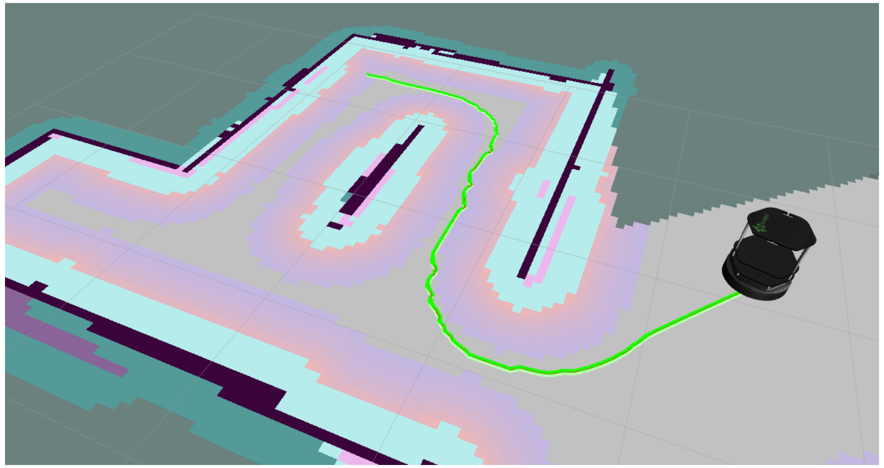
UNIVERSITY OF LJUBLJANA | FRI Faculty of Computer and Information Science







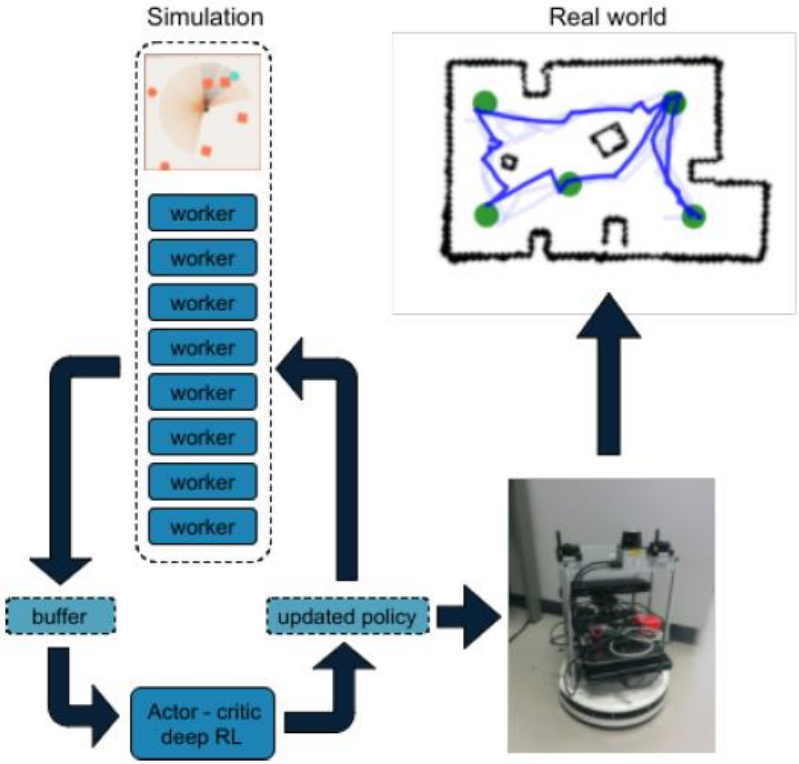
for goal-driven mapless navigation

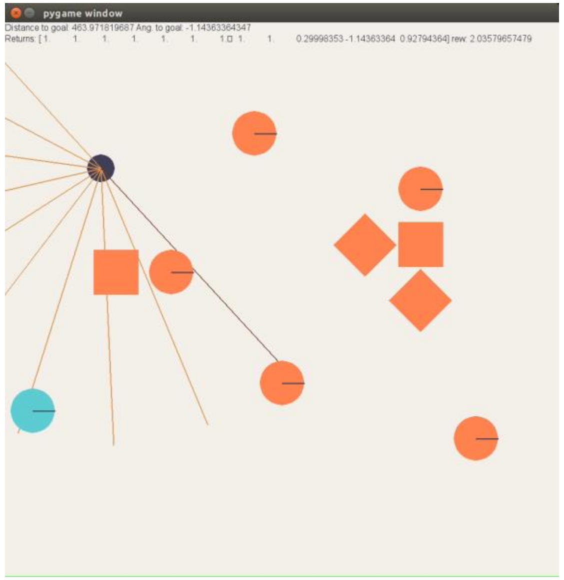


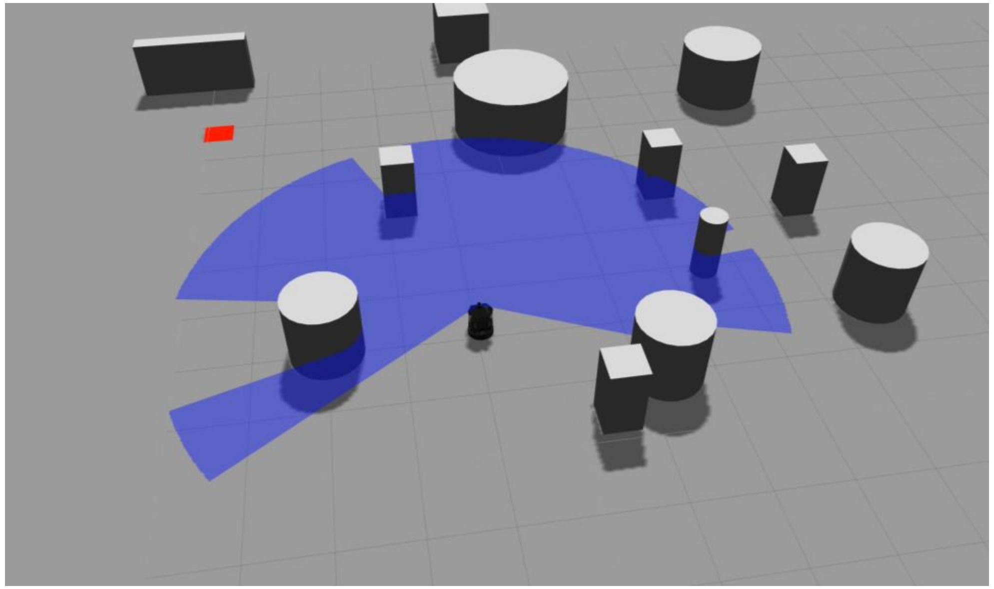
Deep reinforcement learning

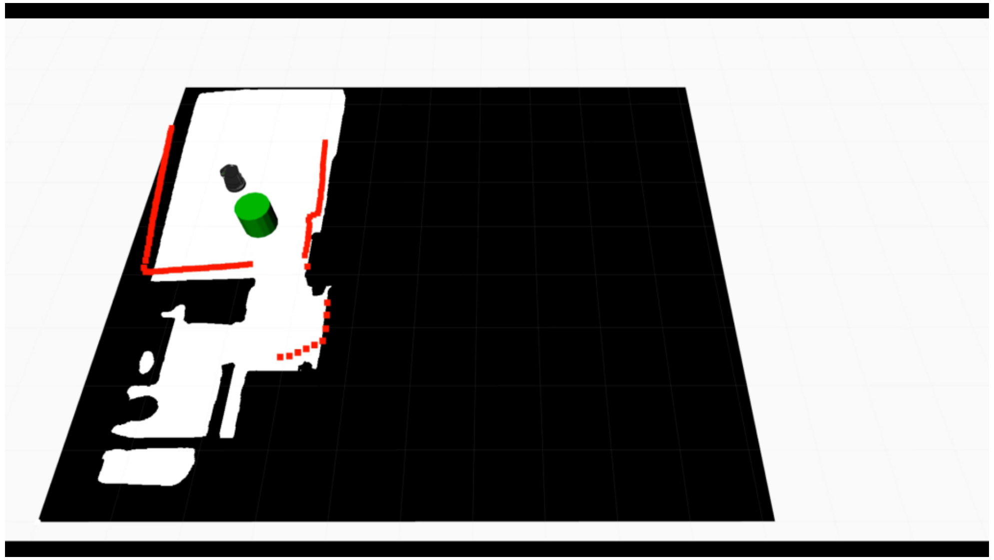
UNIVERSITY OF LJUBLJANA | FRI Faculty of Computer and Information Science

- Training in simulation
  - ~ 600 epochs, 3M steps
- Learned policy transferred to the real robot









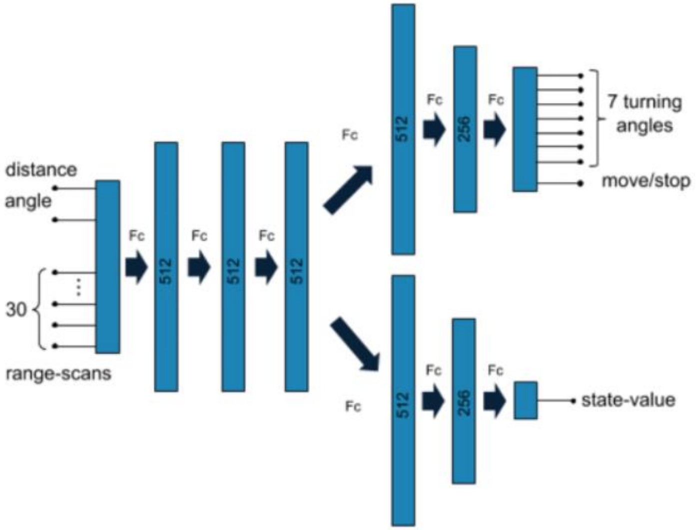
Learning only approach

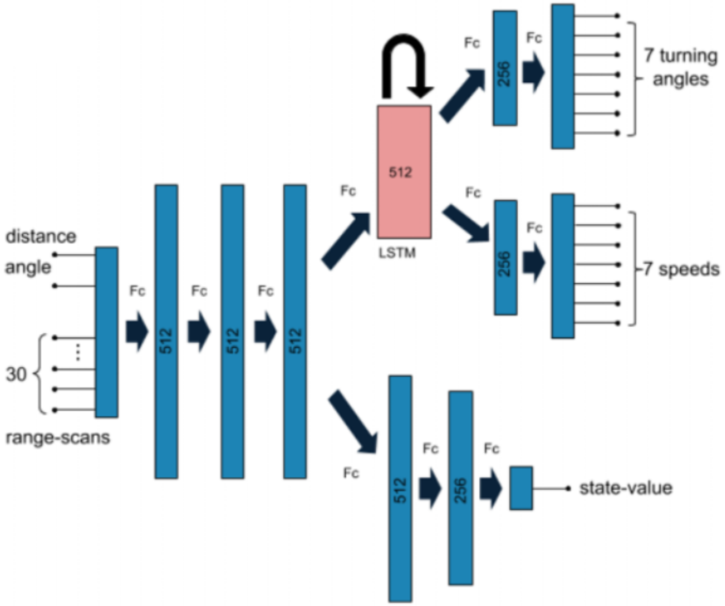
- Navigation as POMDP
- Sensor readings -> actions

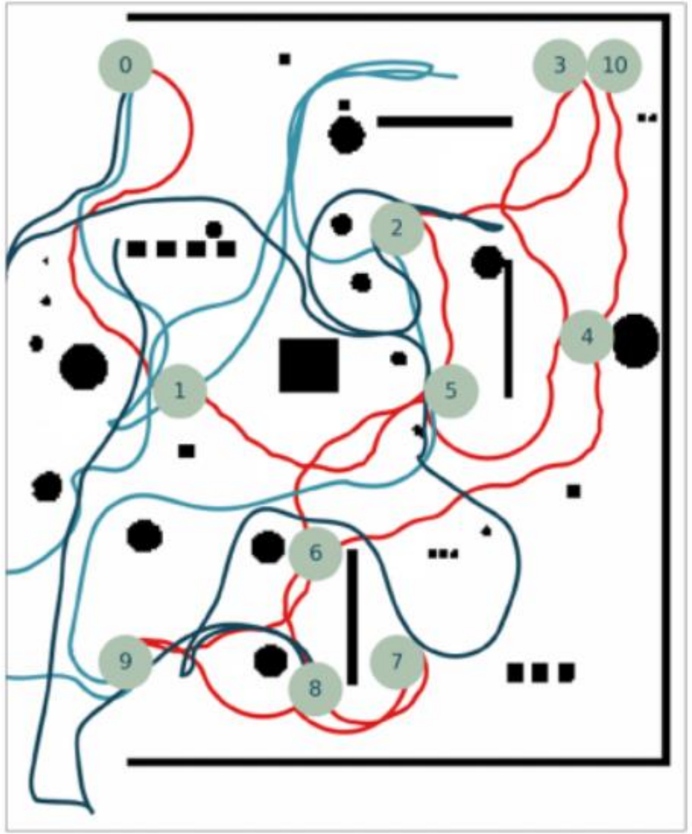
IJARS 2021

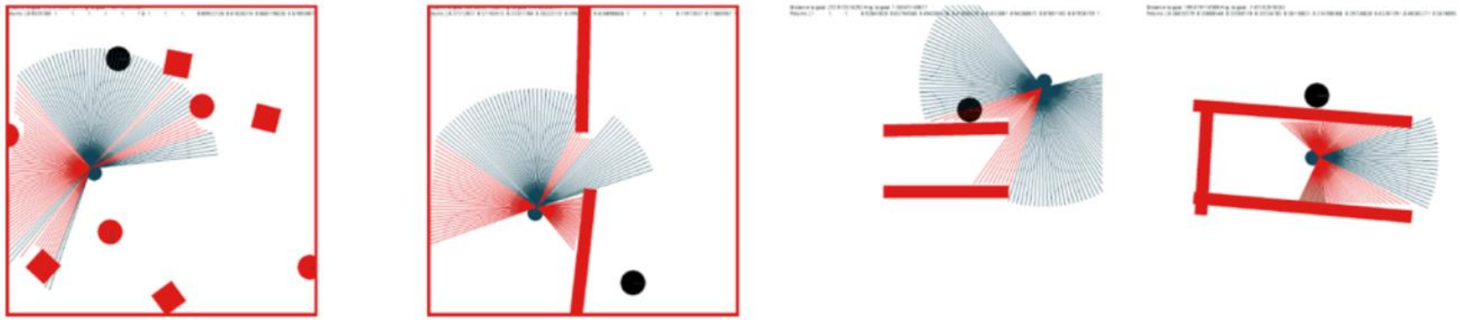
TOR 2024

CompVis 2019-2027









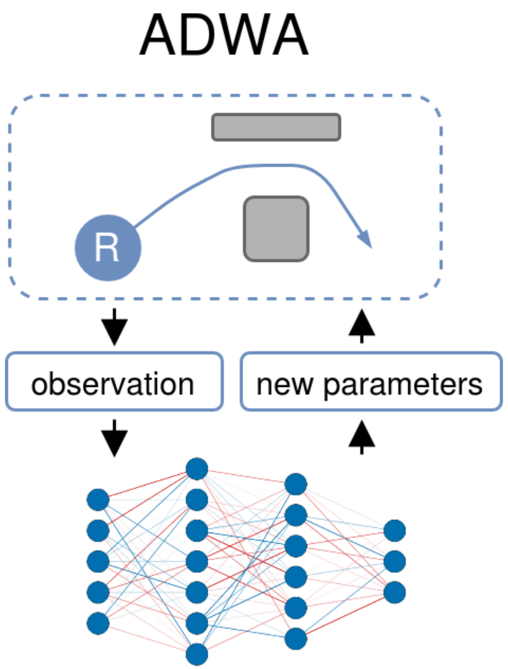
DRL for Adaptive DWA

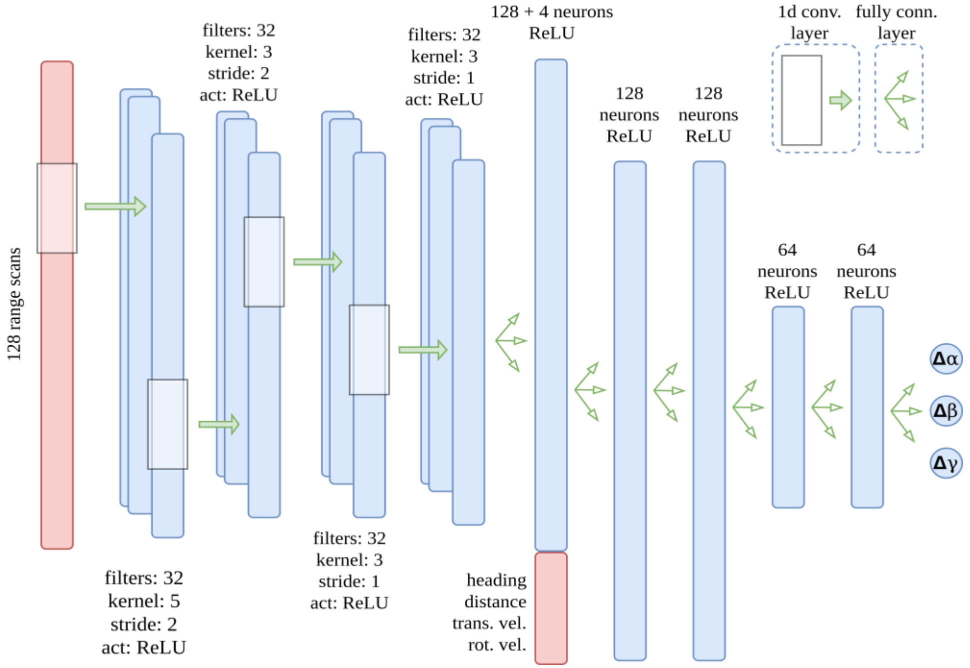
- Classic approaches (DWA)
  - Provide safety mechanisms, smooth trajectories
  - Are not optimised for specific situation
- Learning-based approaches
  - Require additional safety mechanisms
- => merry learning and DWA -> ADWA

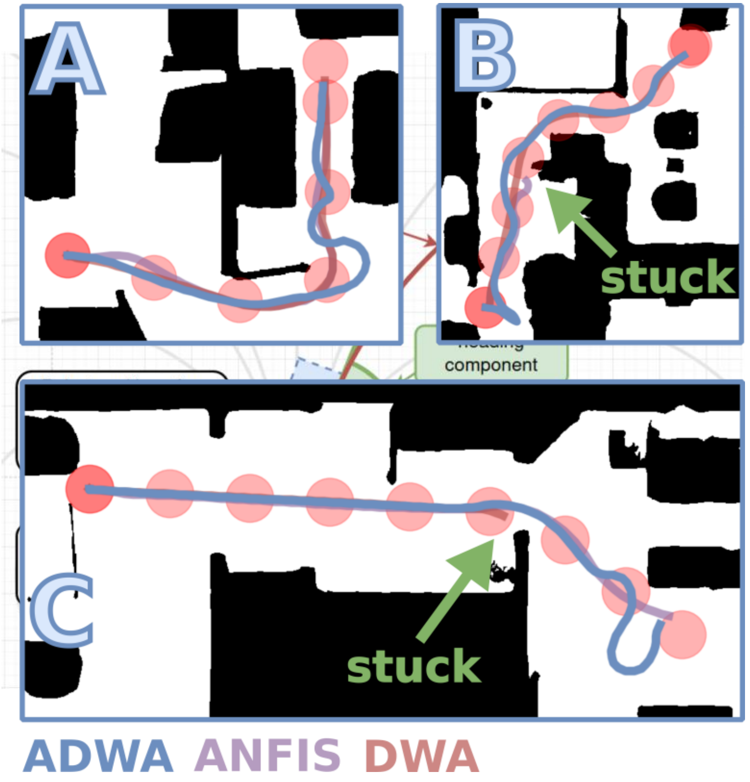
IROS 2020

CompVis 2019-2027

TOR 2024



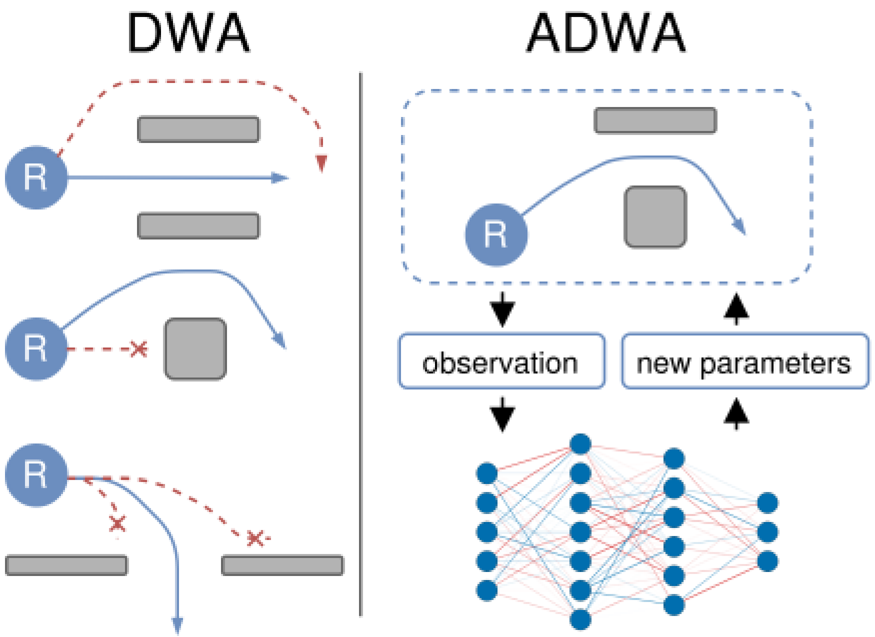


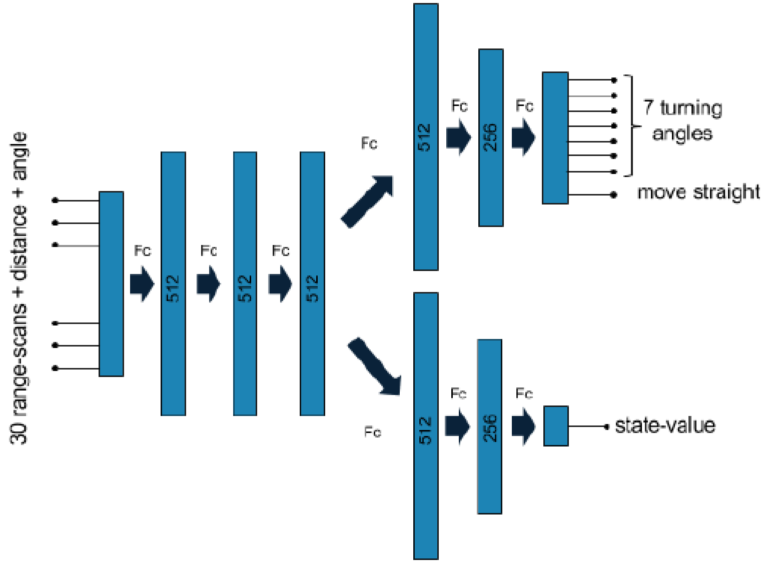


method	# completed ep.
Best DWA[7]	294
ANFIS DWA[15]	340
Ours	**520**

Built-in vs. learned

- Goal-driven mapless navgation
- Constraining the problem with background knowledge





Engineering approach

Engineering approach+ deep learning

Learning only

Function approximator

- Deep model as a function approximator
- Different training possibilities:

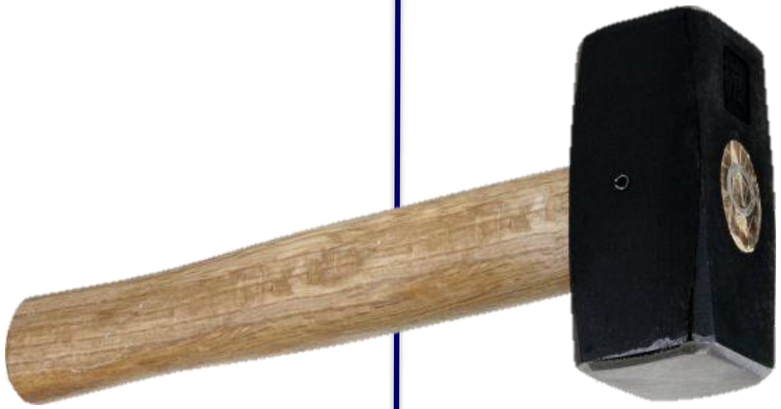
function	known	unknown
$f(x) \doteq y$	x_{tr}, y_{tr}	f
$f(x)$	x_{tr}	f
$f(x) \doteq \hat{f}(x)$	$x_{tr}, \hat{f}$	f
$f(f^{-1}(y)) \doteq y$	y_{tr}, f^{-1}	f
$f(g(x)) \doteq y$	g, x_{tr}, y_{tr}	f
$g(f(x)) \doteq y$	g, x_{tr}, y_{tr}	f

Conventional MV vs. DL

- O: „Deep learning is just a hype and a non-understandable and nonreliable black box"
- Y: „Deep learning is all you need"

- Use adequate HW (camera, lenses, illumination, background) to constrain the problem
  - garbage in garbage out
- Use good old MV techniques when they suffice
  - for less challenging or well-defined problems
  - in controlled environments
- Use MV techniques to constrain the problem
  - and make DL learning easier
  - requiring less training images
- Use DL where the problem is data-driven or hard
  - in less-controlled environments for more general tasks
  - or to speed up the development cycle

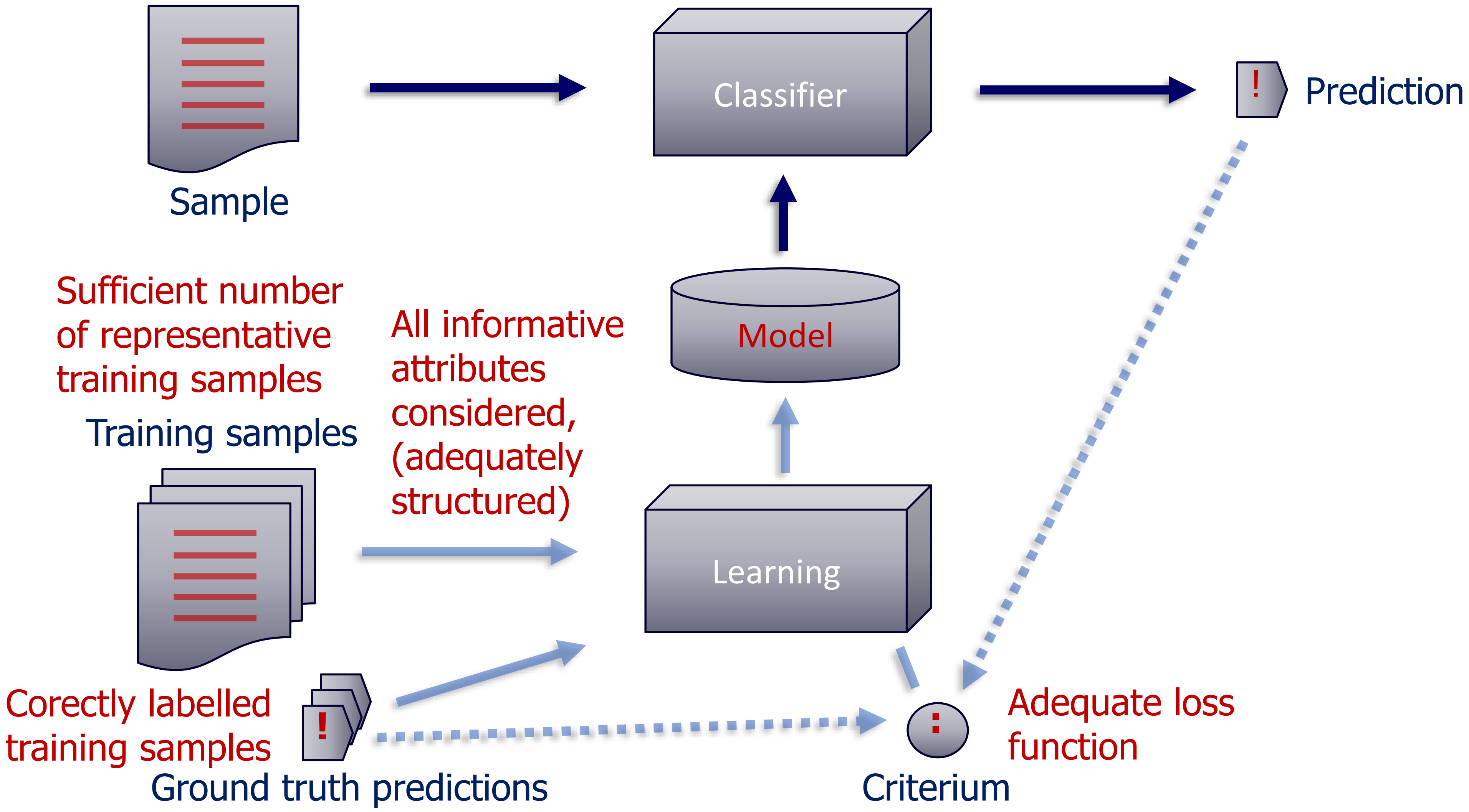
Adequate tools



Deep learning

Routine solutions	Rule-based solutions	Data-driven solutions	General intelligence

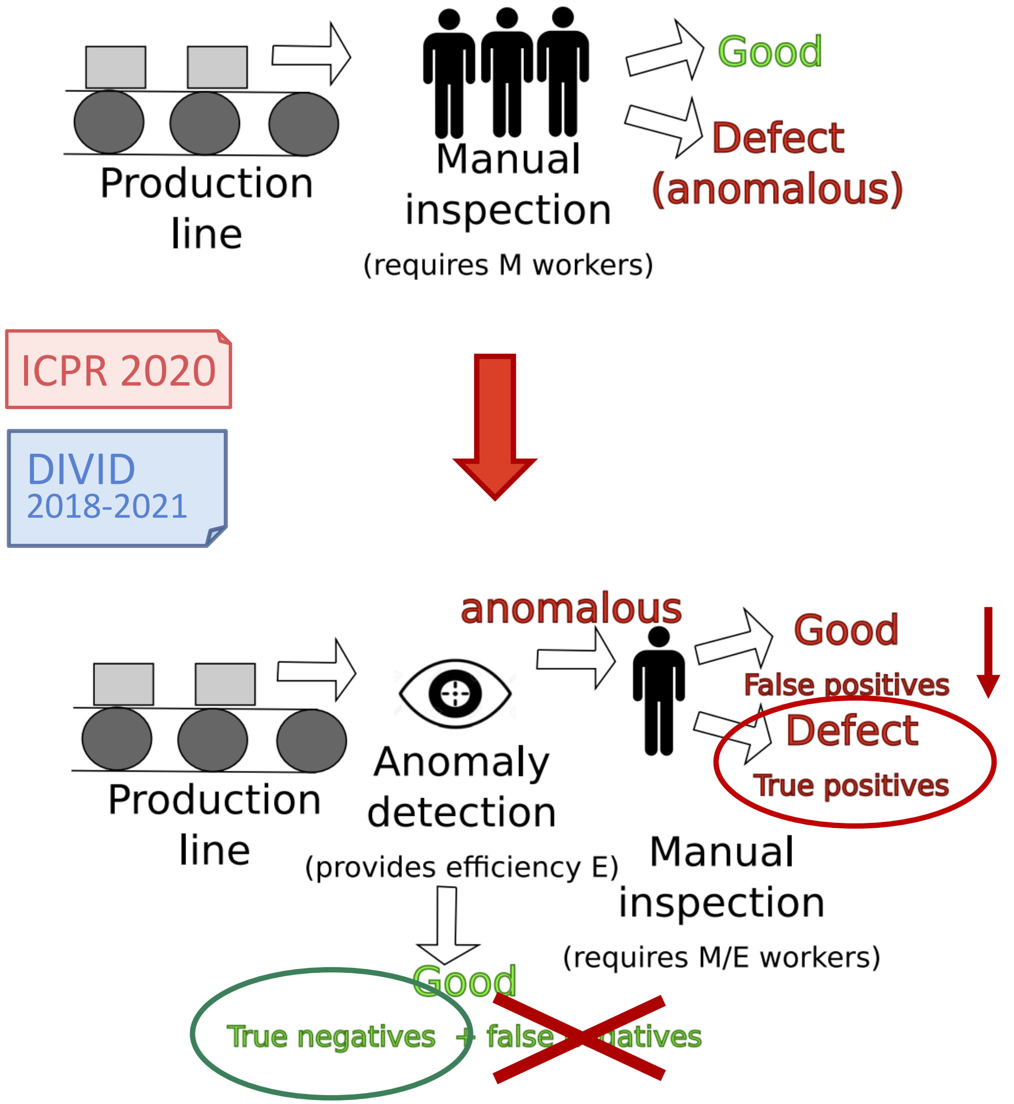
Adequate data



Development and maintainance

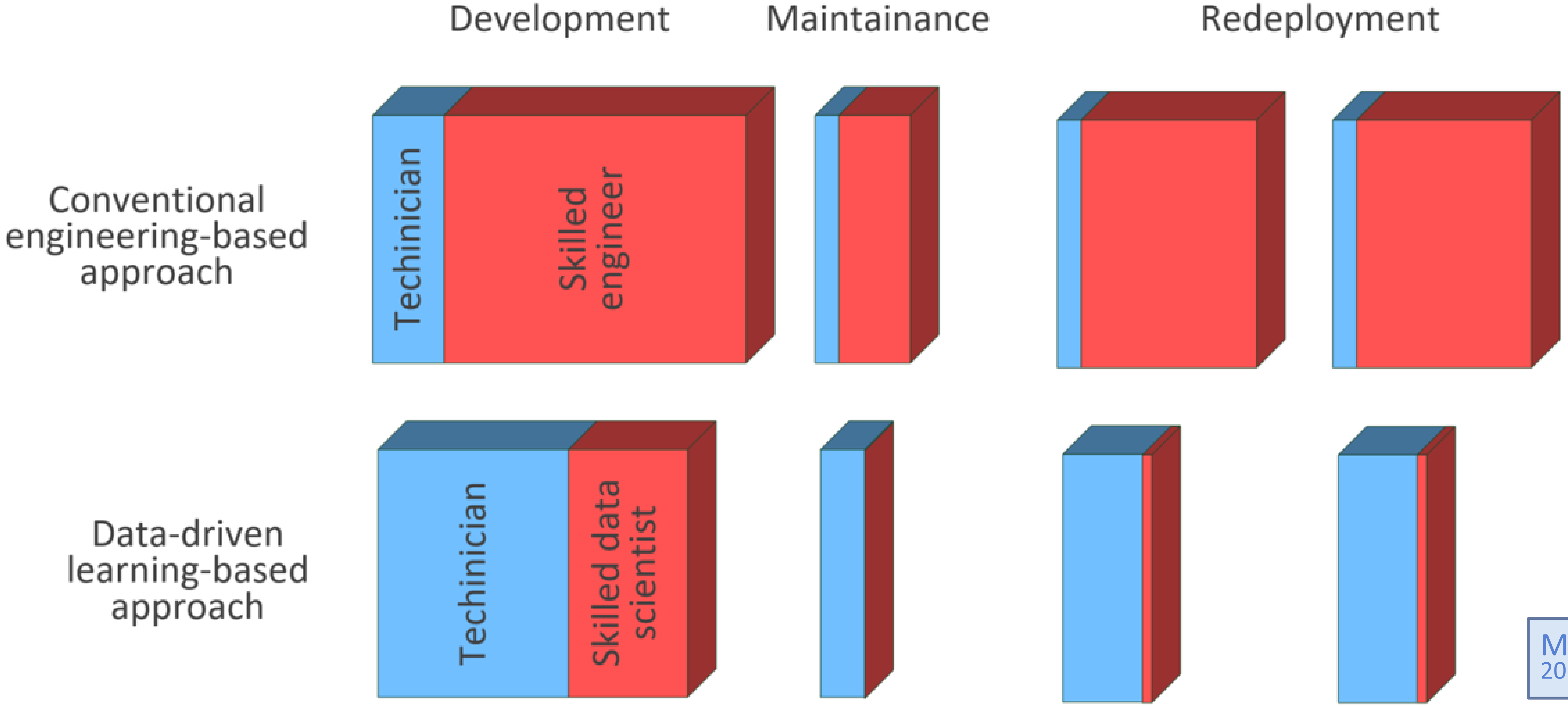
- Data, data, data!
  - Sufficient amount of representative data
  - Correctly labelled data
- Adequate design of deep architecture
  - Adequate backbone, architecture, loss function,…
  - Learning, parameter optimisation
- Efficient implementation
  - Exectution speed
  - Integration
- Development and maintainance
  - Incremental improvement of the learned model
  - Adaptation to the changes in the environment

Real world considerations



- Human in the loop
  - at least for some time
- Challenges
  - Robustness
  - Dependency on the training set
  - Domain shift
  - Non-adaptability
  - Non-interpretability
- Opportunities
  - Learning under mixed supervision
  - Explainabilty
  - Tunable parameters
  - Compability with conventional MV
  - Agility, quick adaptability

Development, maintainance and redeployment



MV4.0
2021-2024

Conclusion

UNIVERSITY OF LJUBLJANA | FRI Faculty of Computer and Information Science

- Data-driven deep-learning-based solutions
- AI/DL/CV/MV – key enabling technologies
- Wide applicability, interdisciplinarity
- Robustness
- Industry 4.0 and beyond
- New challenges, new opportunities
- Collaboration between academia and industrial partners
- Use all data available ;-)