Visual Understanding across Semantic Groups, Domains and Devices

Neural Networks and Deep Learning, 2021/2022

Dr Umberto Michieli and Dr Alberto Testolin

umberto.michieli@unipd.it

LTTM Team

https://lttm.dei.unipd.it/

Staff

- **4 Associate Professor**: Pietro Zanuttigh Simone Milani Giancarlo Calvagno Marco Cagnazzo
- **1 Assistant Professor**: Federica Battisti

- **1 postdoc**: Umberto
- **8 PhD**: Marco, Adriano, Francesco, Giulia, Mazen, Donald, Elena, Daniele
- **1 research grant**: Sofia
- MSc students

Research areas:

- Computer Vision
- Multimedia Forensics
- Image and Video Processing and Coding

Thesis Proposals

M.Sc. thesis topics **in Lab**:

- Continual Learning
- Unsupervised Domain Adaptation
- Multimodal Learning
- Federated Learning
- *You can propose your own, to be agreed on*

Vision Tasks

(focus on Semantic Segmentation)

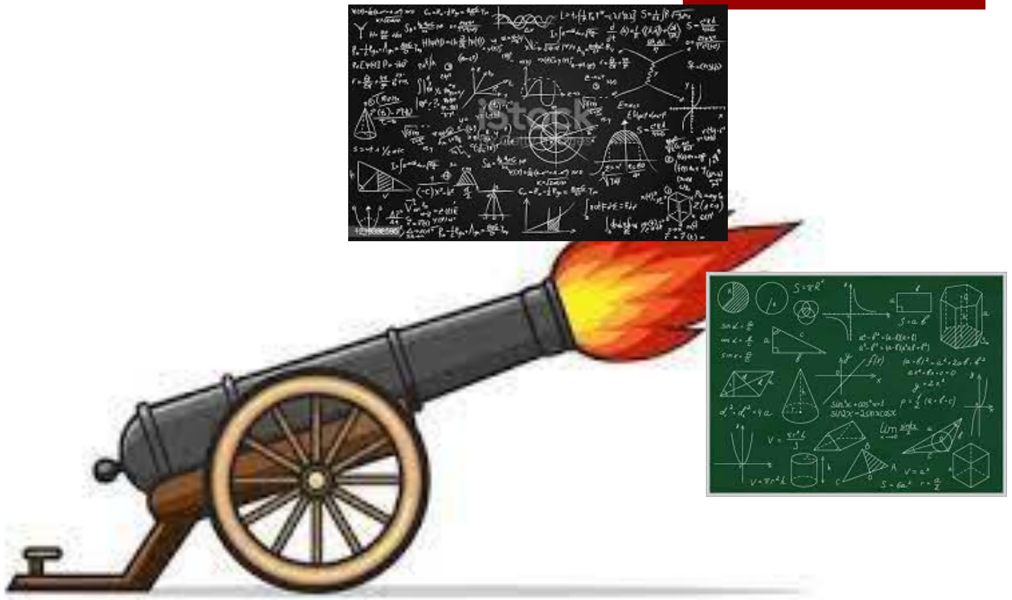
M.Sc. thesis topics **in Companies** (internships):

- 4EasyTech (Verona)
- Sisma (Schio, VI)
- Altair (Vicenza)
- Ramete (Trebaseleghe, PD) link
- Addfor (Torino)
- Nidek (Albignasego, PD)
- Microtec (Mestre, VE)

- Samsung (UK)
- Sony (Stuttgart)

Outline

1) Continual Learning
   - Knowledge Distillation
   - Latent Space Regularization
   - Replay-based Approaches
   - ➢ **Notebook** on Continual Learning (20 minutes)



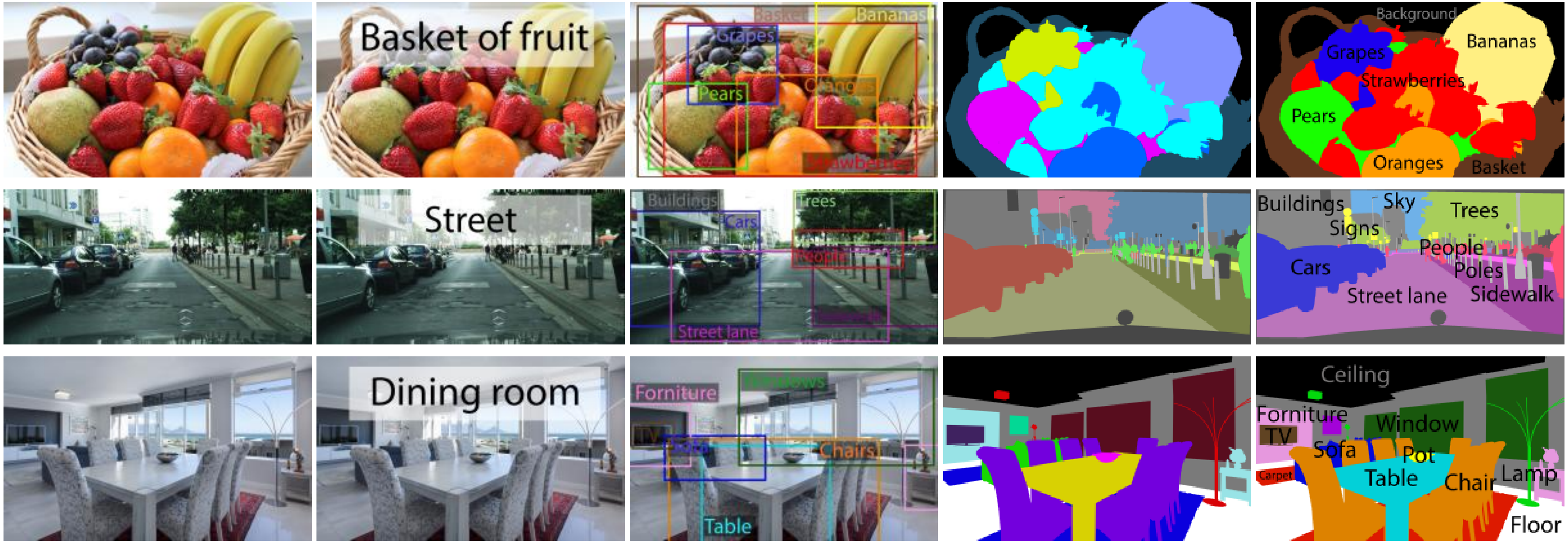
2) Unsupervised Domain Adaptation
   - Focus on Multimodal Learning

3) Federated Learning

Semantic Segmentation – Definition

Many related problems:

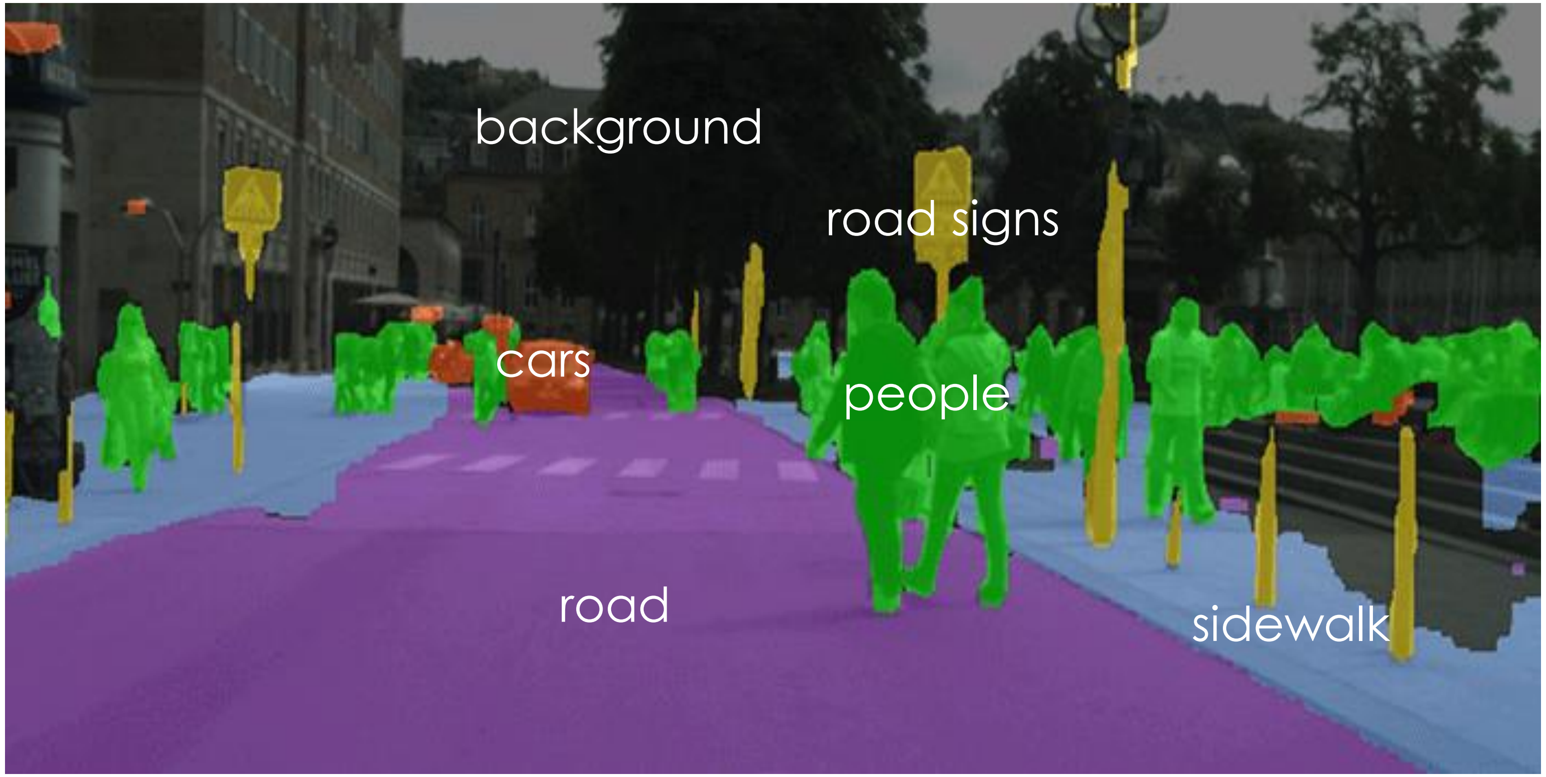
RGB	Classification	Object Localization	Segmentation	Semantic Segmentation



complexity & historical evolution

Semantic Segmentation – Definition

Assign to each pixel a label representing the class or object to which the pixel belongs



- Dense task
- Deep learning revolutionized the field (autoencoders) [1]

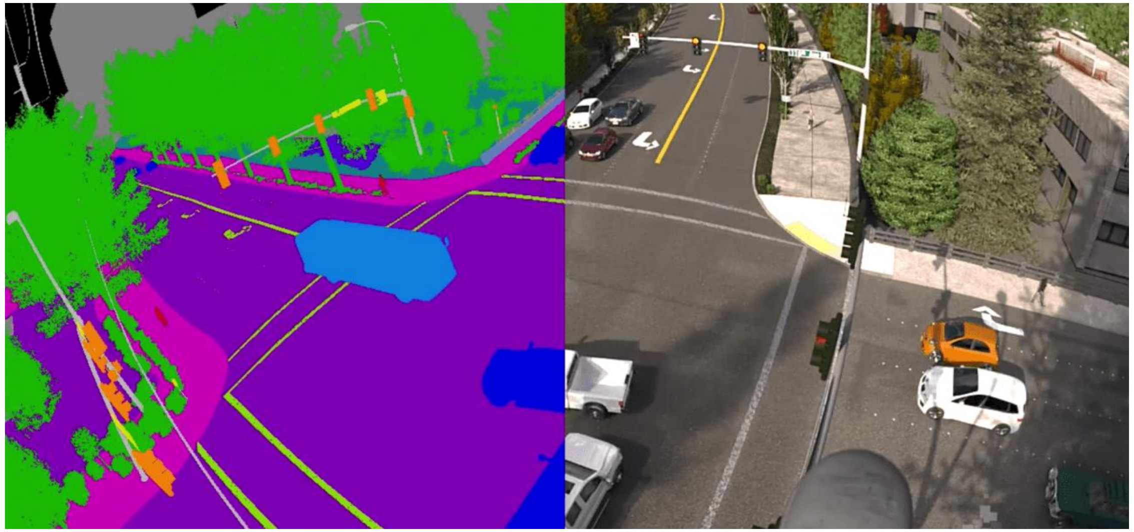
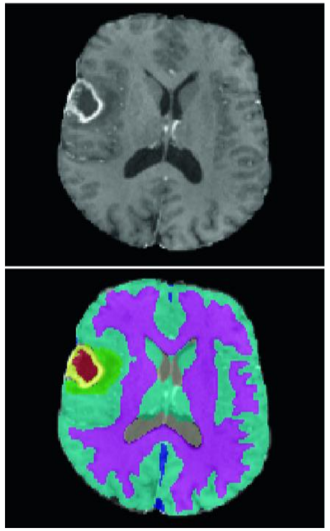
[1] Long et al., "Fully convolutional networks for semantic segmentation", CVPR 2015.

Semantic Segmentation – Applications



An important problem:

- Autonomous vehicles
- Robotics
- Video surveillance
- Medical imaging
- ...

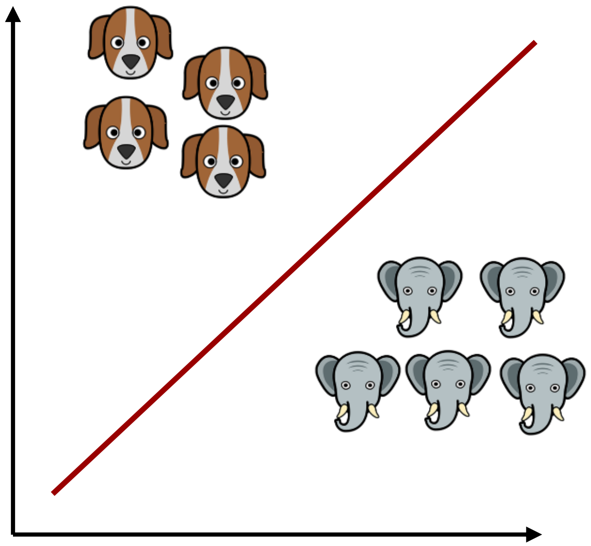



Continual Learning in Semantic Segmentation

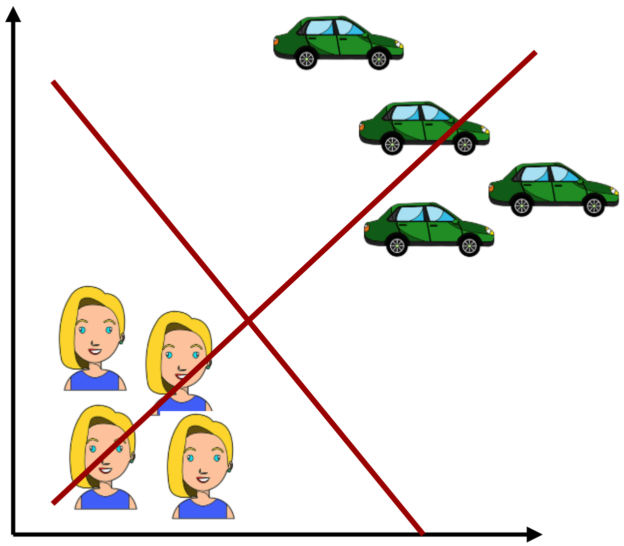
CL – Definition

Aim: learn new tasks over time without forgetting previous ones

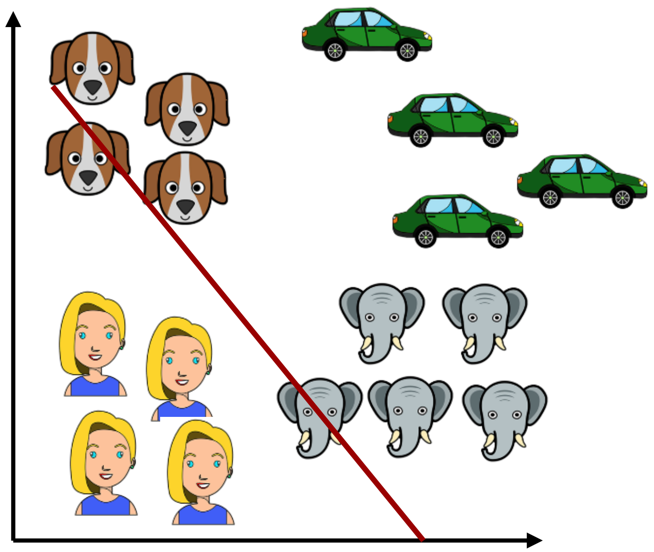
Initial task



New task



Result on all tasks



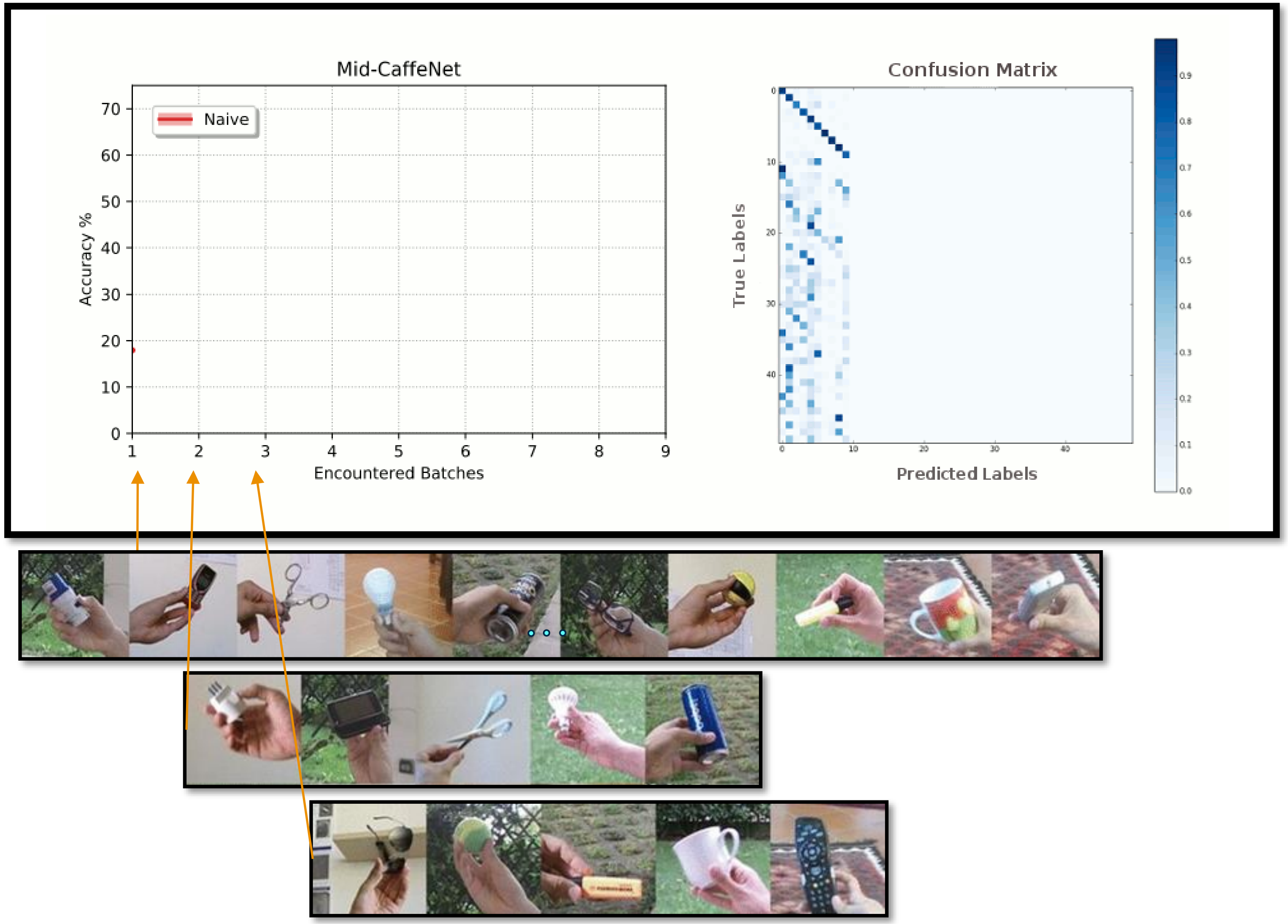
Catastrophic forgetting

In real/complex tasks:

→ we want the different tasks to interact

→ the number of parameters is of the order of millions

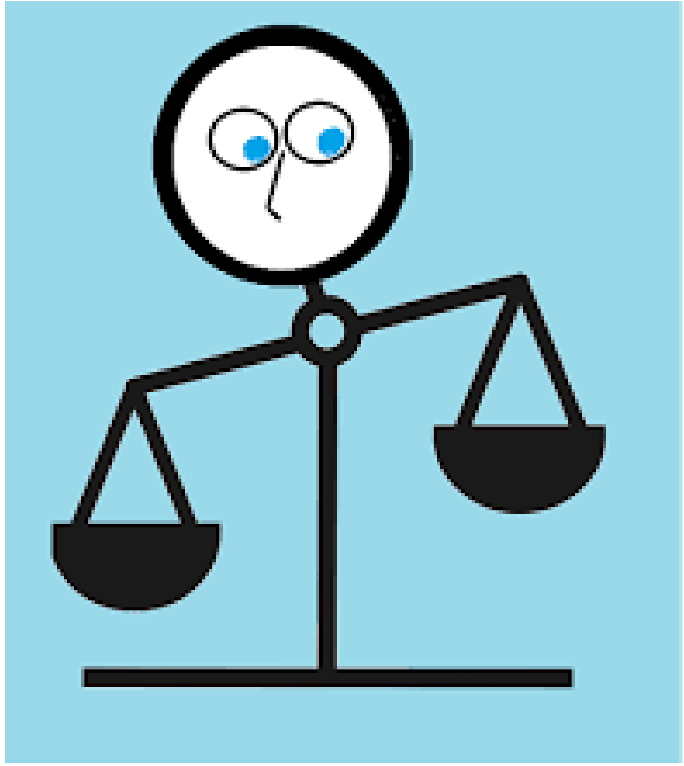
CL – Forgetting



- *A set of new objects (classes) each training stage*
- *10 the first training stage, 5 the following ones*

CORe50: a new Dataset and Benchmark for Continuous Object Recognition, V. Lomonaco & D. Maltoni. Conference on Robot Learning (CoRL), 2017.

CL – Stability-Plasticity Dilemma

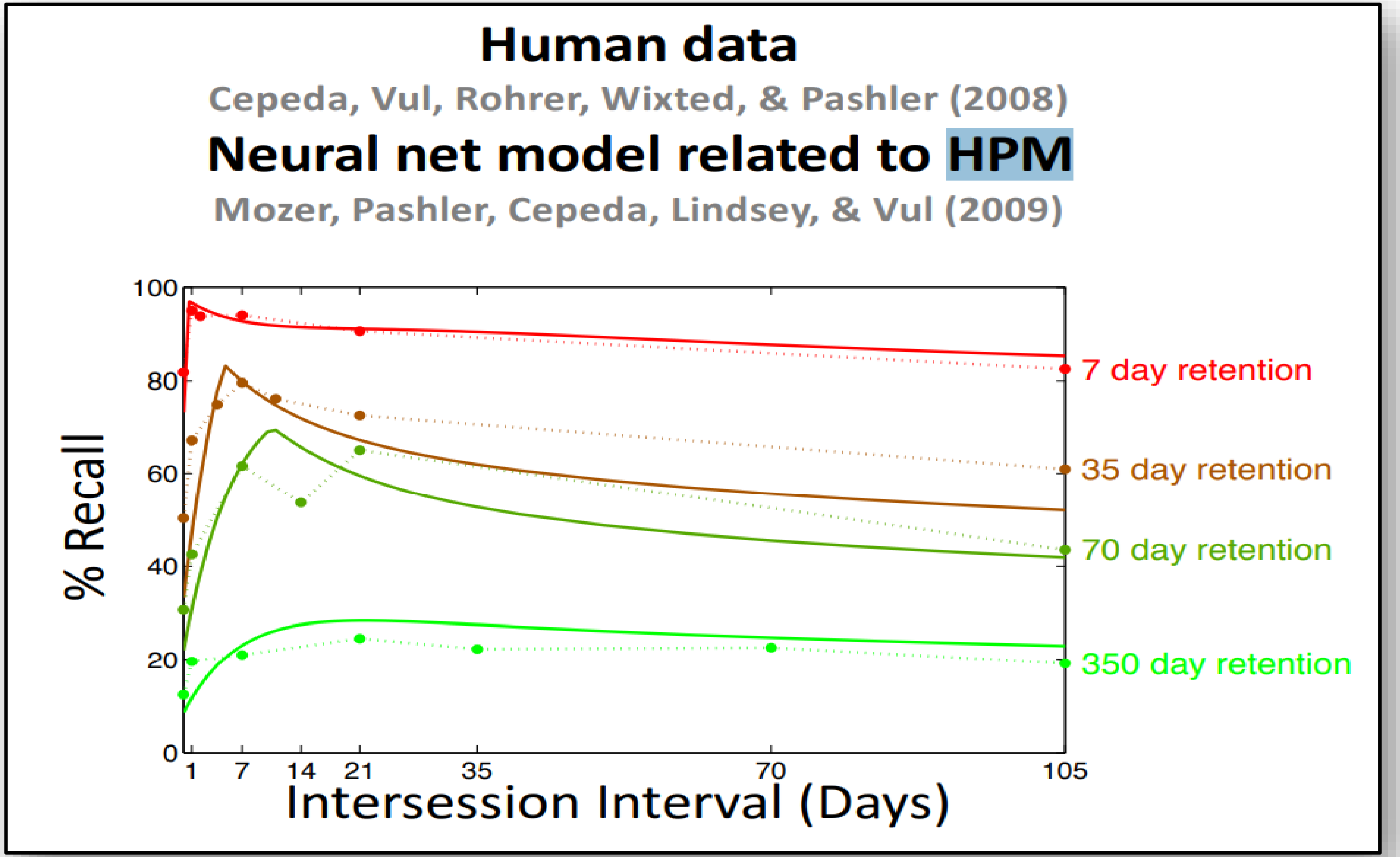


Stability-Plasticity Dilemma:

- **Stability:** *remember past concepts*
- **Plasticity:** *learn new concepts*
- **Dilemma:** *how to balance the trade-off?*

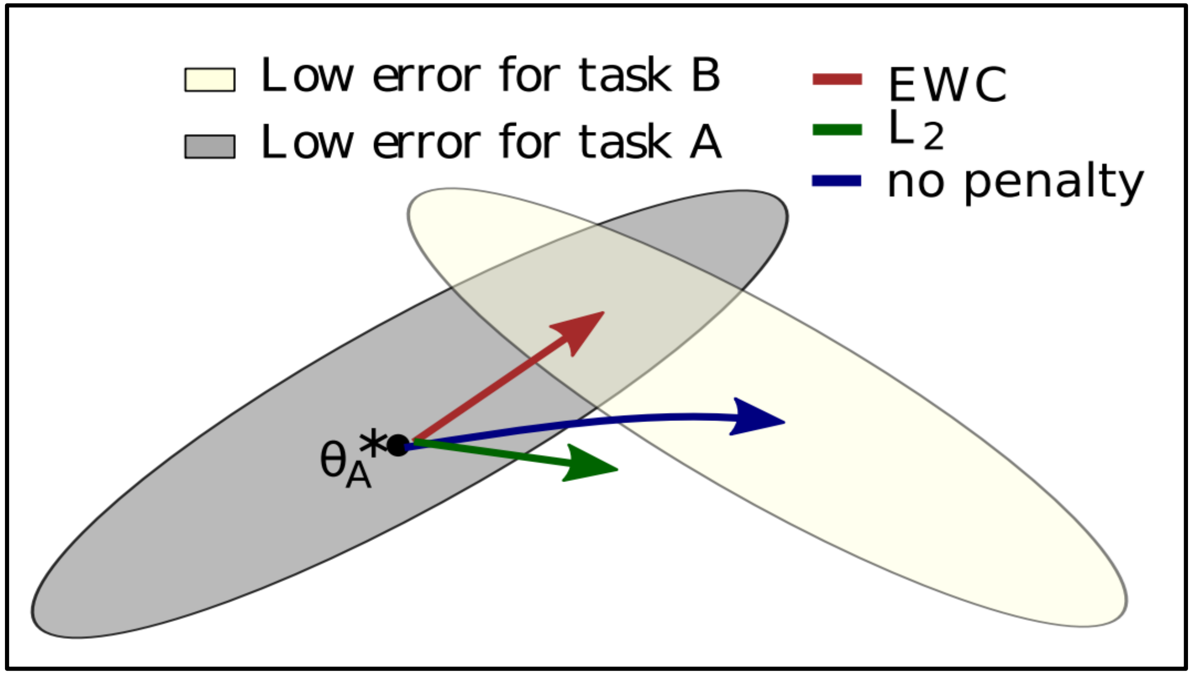
Parisi, German I., et al. "Continual lifelong learning with neural networks: A review." *Neural Networks* 113 (2019): 54-71.

CL – Stability-Plasticity Dilemma



Predicting the optimal spacing of study: A multiscale context model of memory. M. Mozer et al. NeurIPS, 2009.

CL – Stability-Plasticity Dilemma



The objective of a CL algorithm is to minimize the loss $\mathcal{L}_S$ over the entire stream of data S:

$$\mathcal{L}_S(f_n^{CL}, n) = \frac{1}{\sum_{i=1}^{n} |\mathcal{D}_{test}^i|} \sum_{i=1}^{n} \mathcal{L}_{exp}(f_n^{CL}, \mathcal{D}_{test}^i) \tag{2}$$

$$\mathcal{L}_{exp}(f_n^{CL}, \mathcal{D}_{test}^i) = \sum_{j=1}^{|\mathcal{D}_{test}^i|} \mathcal{L}(f_n^{CL}(\boldsymbol{x}_j^i), y_j^i), \tag{3}$$

where the loss $\mathcal{L}(f_n^{CL}(\boldsymbol{x}), y)$ is computed on a single sample $\langle \boldsymbol{x}, y \rangle$, such as cross-entropy in classification problems.

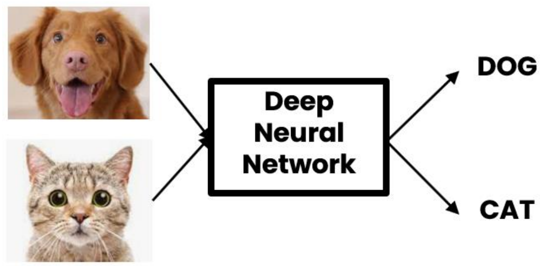
We don't have access to previously encountered data!

Total loss over S can only be approximated.

EWC: Overcoming catastrophic forgetting in neural networks. James Kirkpatrick et al. Proceedings of the national academy of sciences, 2017.

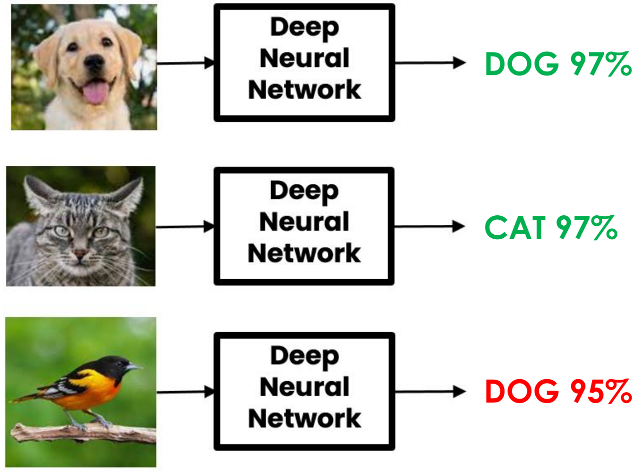
CL – Stability-Plasticity Dilemma

Training



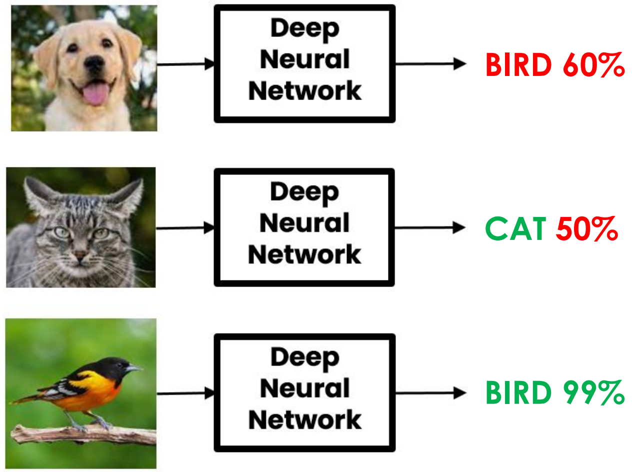
STABILITY

Test time (no update)



PLASTICITY

Test time (naïve update)



CL – Methods and Metrics

Families of **methods:**

- **Knowledge distillation** from previous models
- **Replay** (e.g., storing past data, retrieving data from other sources such as Web or GAN,…)
- **Regularization** (on weights, on features, parameter freezing,…)
- **Architectural** (e.g., ensemble of models)
- …

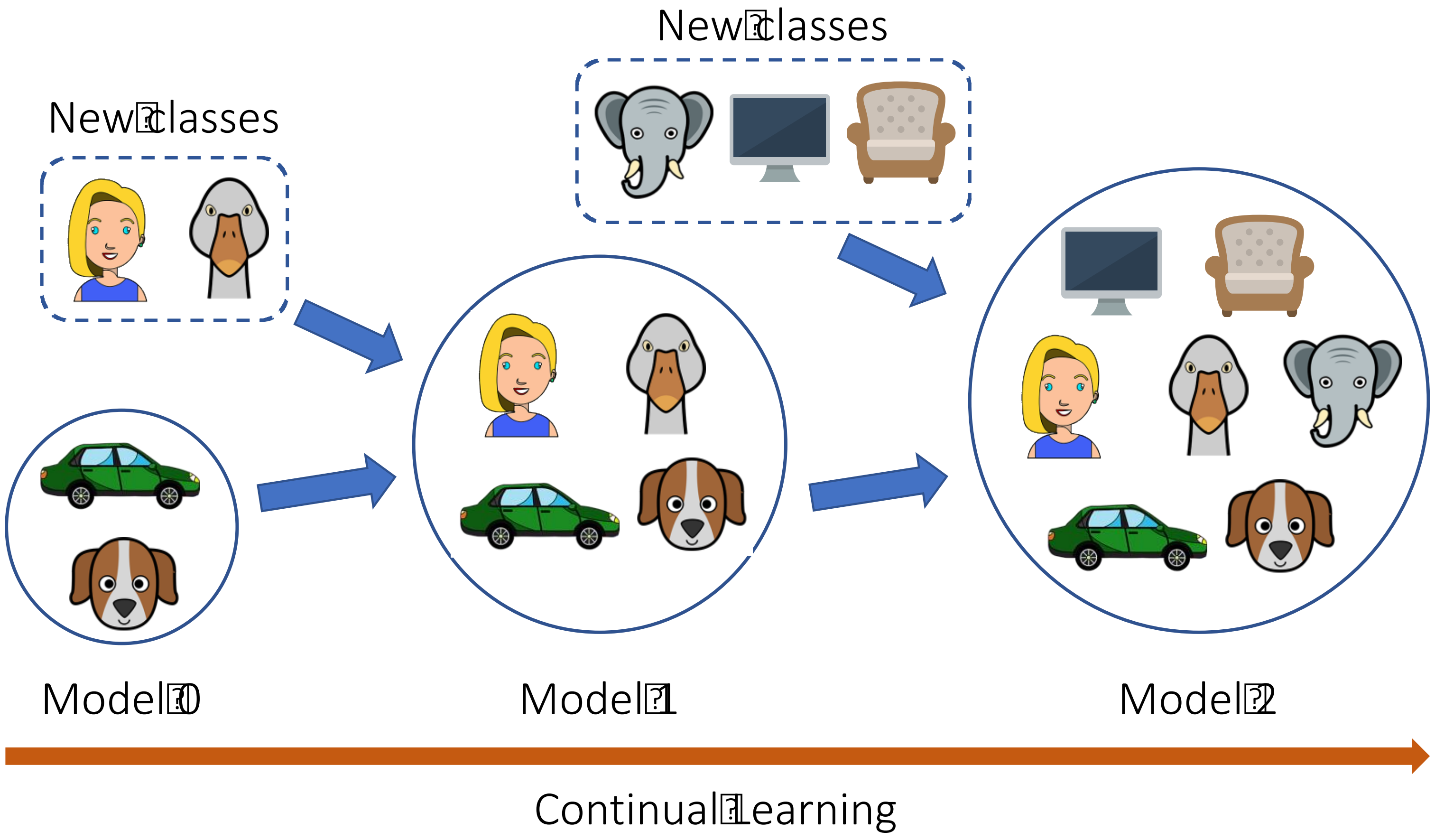
Metrics to evaluate CL algorithms:

- **Final accurcy**
- **Convergence speed**
- **Efficiency of Learning (memory, computation)**
- **Robustness**
- **…**

Parisi, German I., et al. "Continual lifelong learning with neural networks: A review." *Neural Networks* 113 (2019): 54-71.

CL SS – Definition

It's all about this: *learn to segment and label new classes w/o forgetting previous ones*



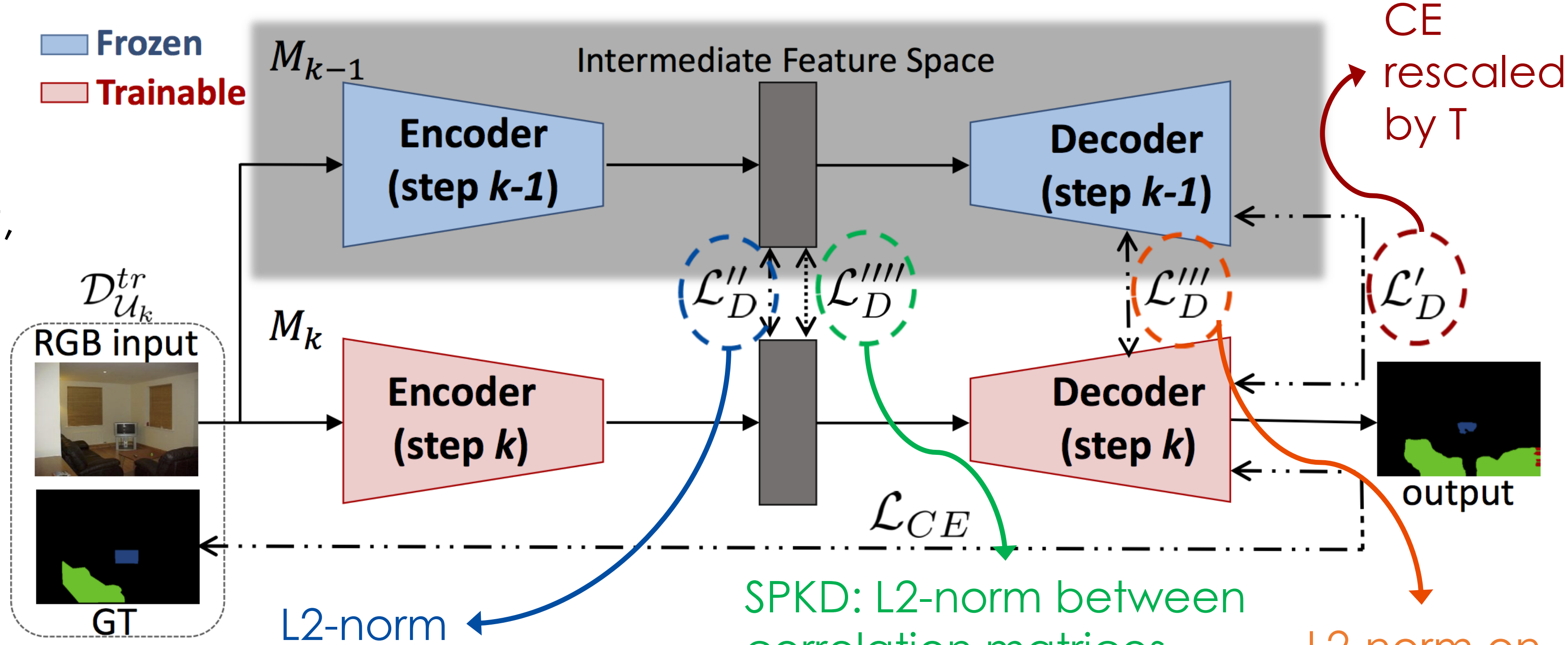
This does not happen for DNN → *Catastrophic Forgetting*

CL SS – Knowledge Distillation & Freezing

KD: a teacher model tells a student model how to respond to inputs

We designed 4 KD losses for CL in SS:

- different levels (output, features, custom)
- different functions



Michieli et al. "Incremental Learning Techniques for Semantic Segmentation", ICCVW 2019
Michieli et al. "Knowledge Distillation for Incremental Learning in Semantic Segmentation", CVIU 2021
SPKD: Tung F. et a., "Similarity-preserving knowledge distillation", ICCV 2019.

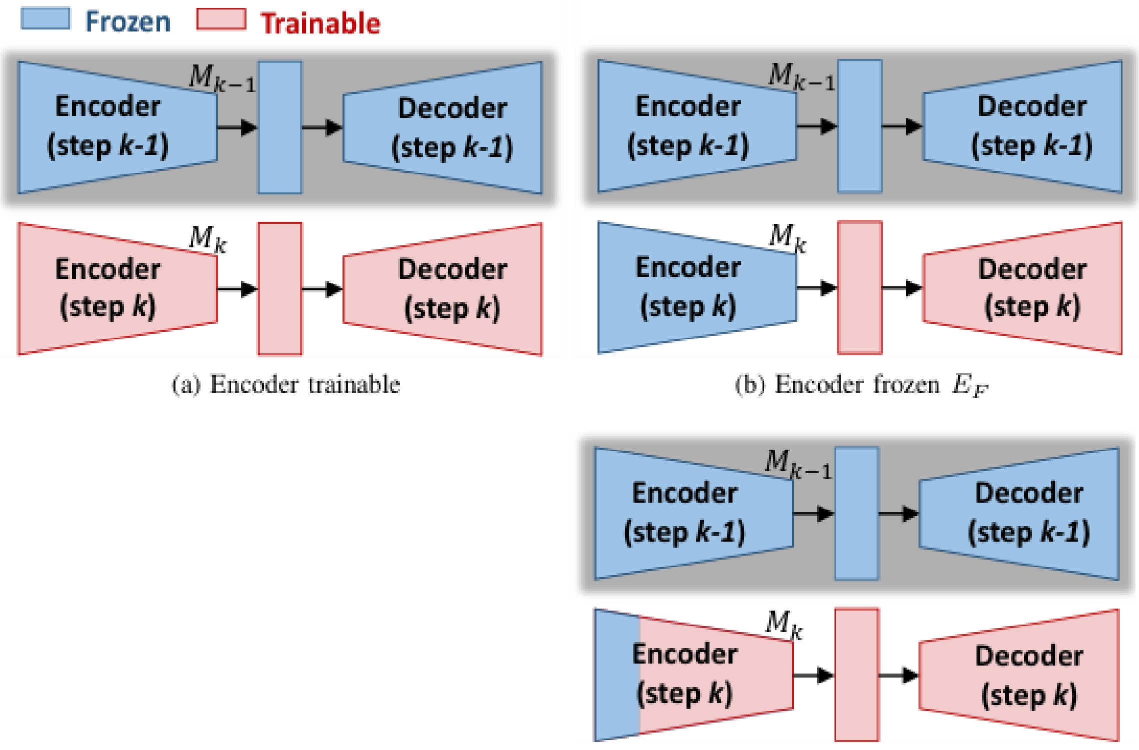
CL SS – Knowledge Distillation & Freezing

KD: a teacher model tells a student model how to respond to inputs

Freezing: important parameters for old tasks may be kept unaltered

We experimented simple freezing schemes for the new model:

- completely trainable (a)
- feature extraction capabilities frozen (b)
- first task-generic layers of features extraction frozen (c)



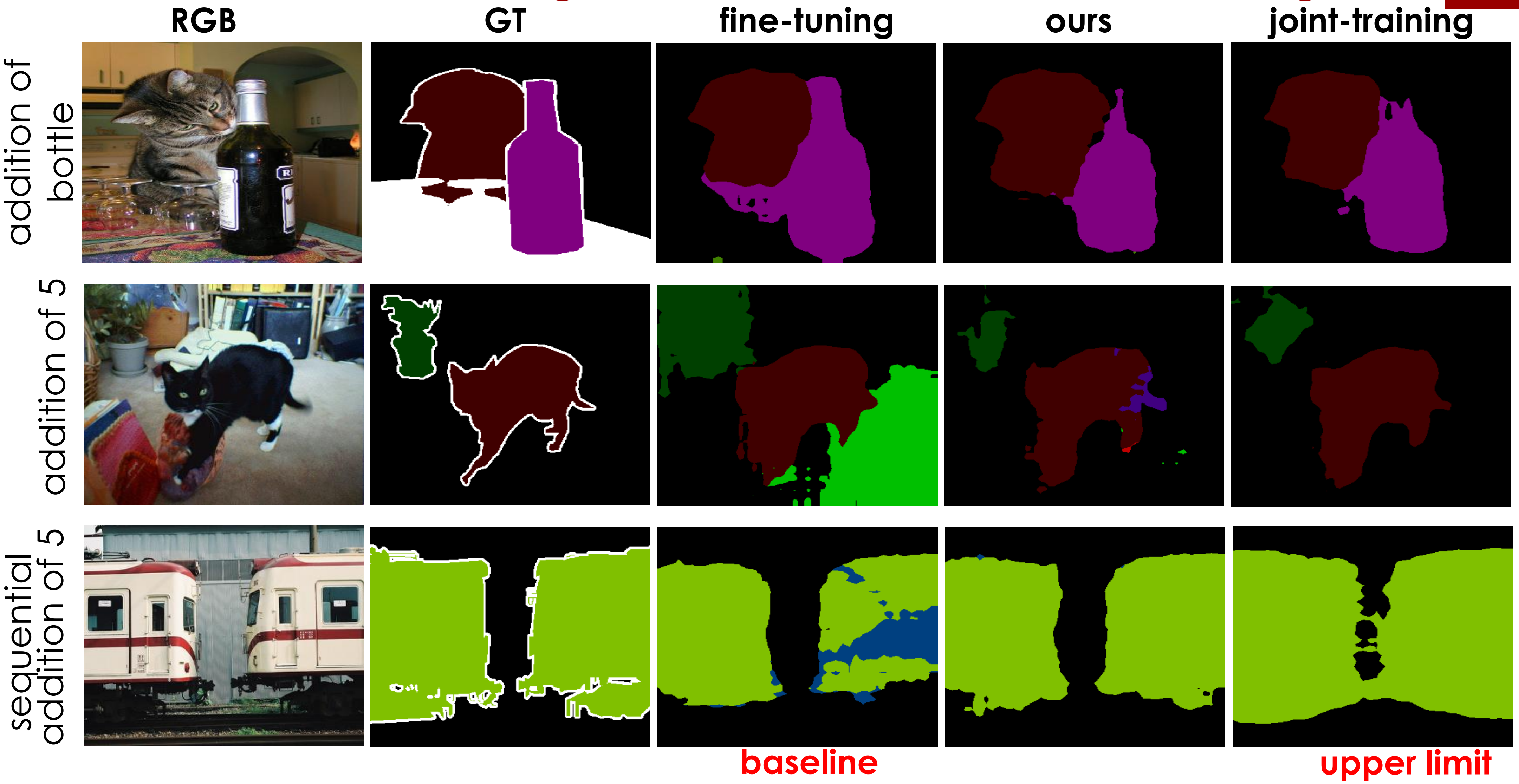
(a) Encoder trainable

(b) Encoder frozen E_F

(c) Two layers of encoder frozen E_{2LF}

Michieli et al. "Incremental Learning Techniques for Semantic Segmentation", ICCVW 2019
Michieli et al. "Knowledge Distillation for Incremental Learning in Semantic Segmentation", CVIU 2021

CL SS – Knowledge Distillation & Freezing



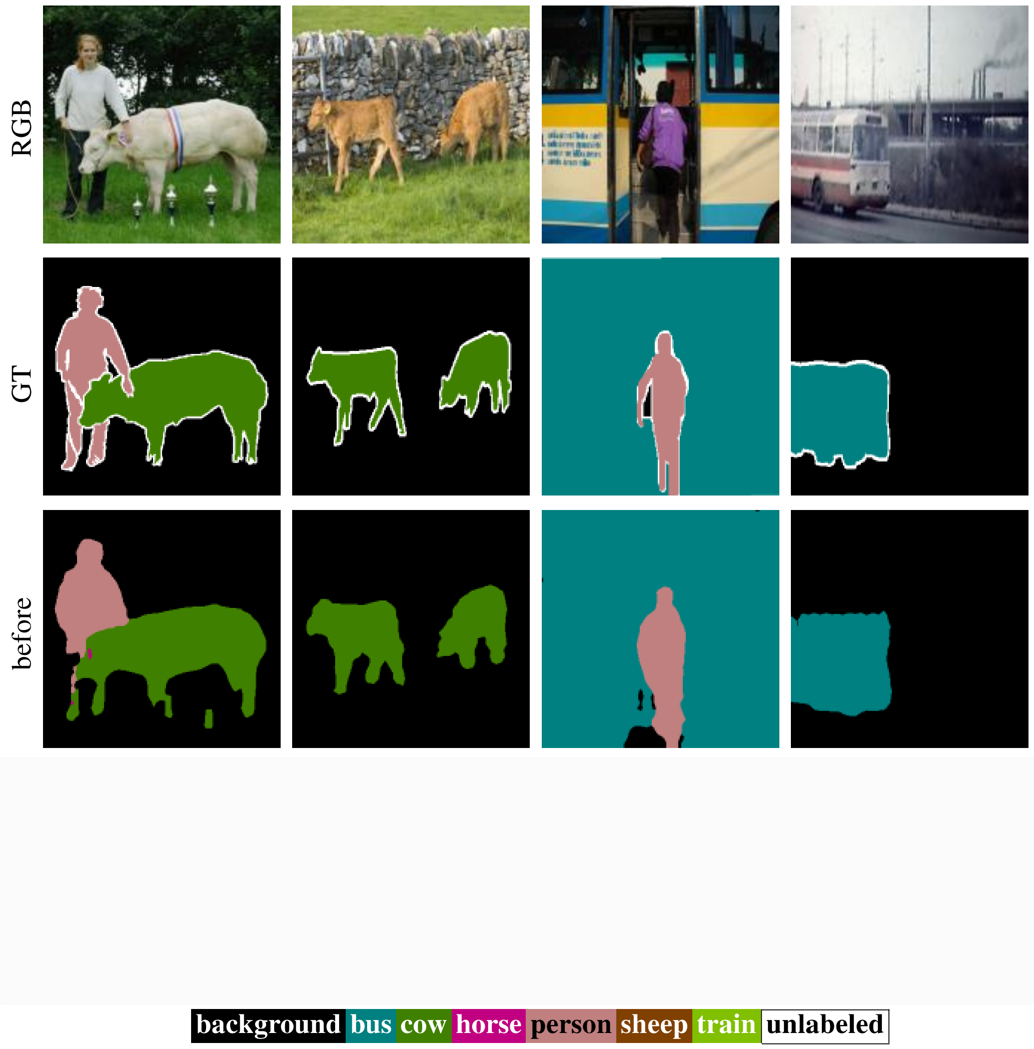
Michieli et al. "Incremental Learning Techniques for Semantic Segmentation", ICCVW 2019
Michieli et al. "Knowledge Distillation for Incremental Learning in Semantic Segmentation", CVIU 2021

CL SS – Knowledge Distillation & Freezing

Main limitation: visually similar classes are sometimes misled

Step t: we train our network with ***cow***, ***person*** and ***bus***
→ Good semantic segmentation

Step t+1: we add ***sheep*** and ***train***
→ ***cow*** misled as ***sheep***
→ ***bus*** misled as ***train***



Maximize the separation of the latent representations

Internal Feature Representations

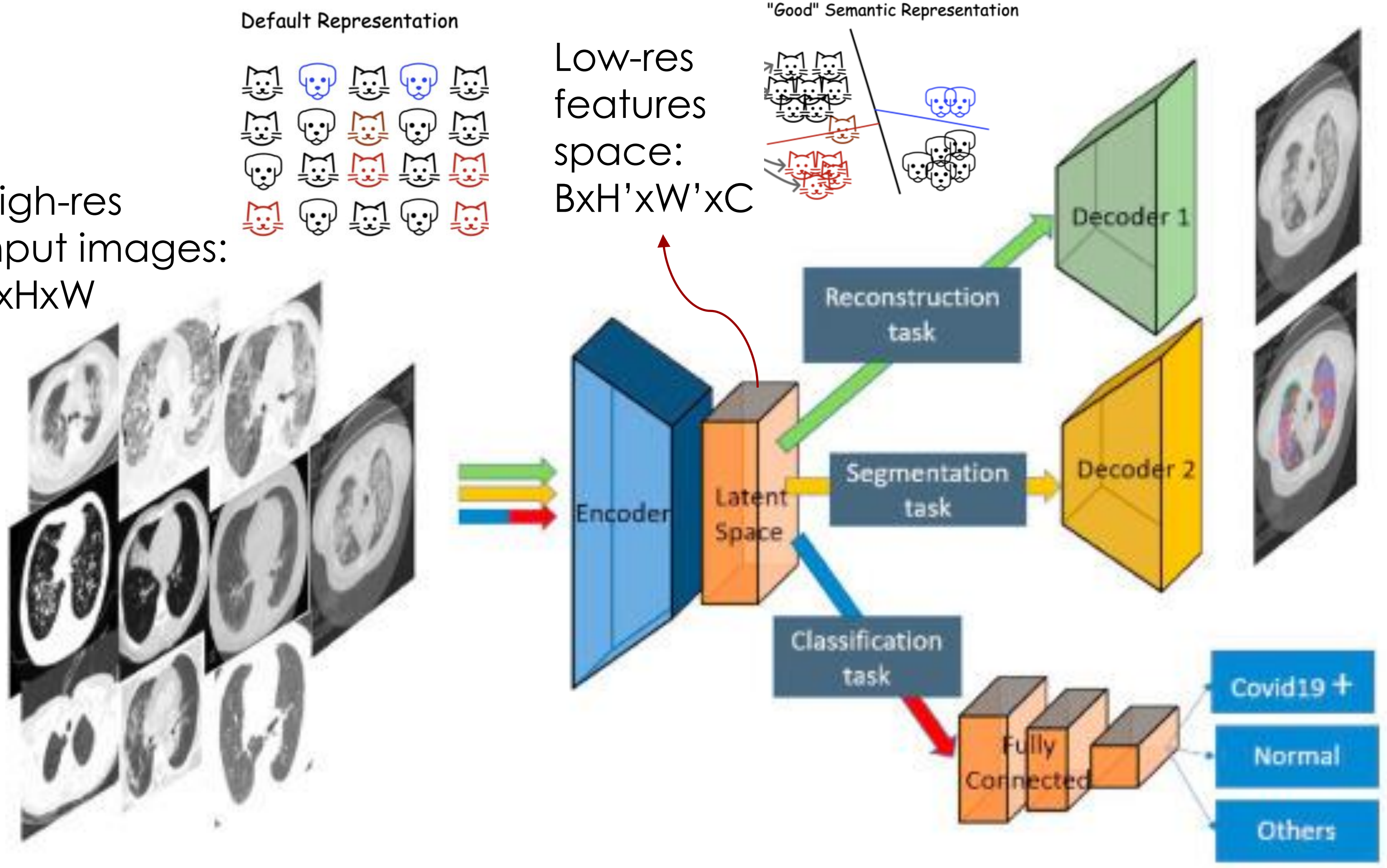
High-res input images: BxHxW

Low-res features space: BxH’xW’xC

High-resolution output: BxHxW

High-resolution output: BxHxW

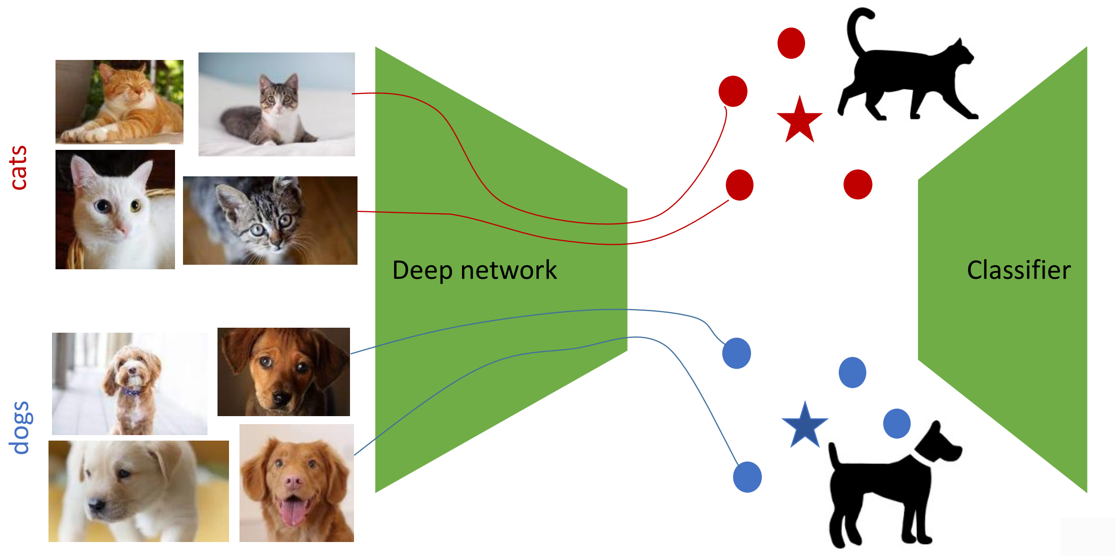
Single-label prediction

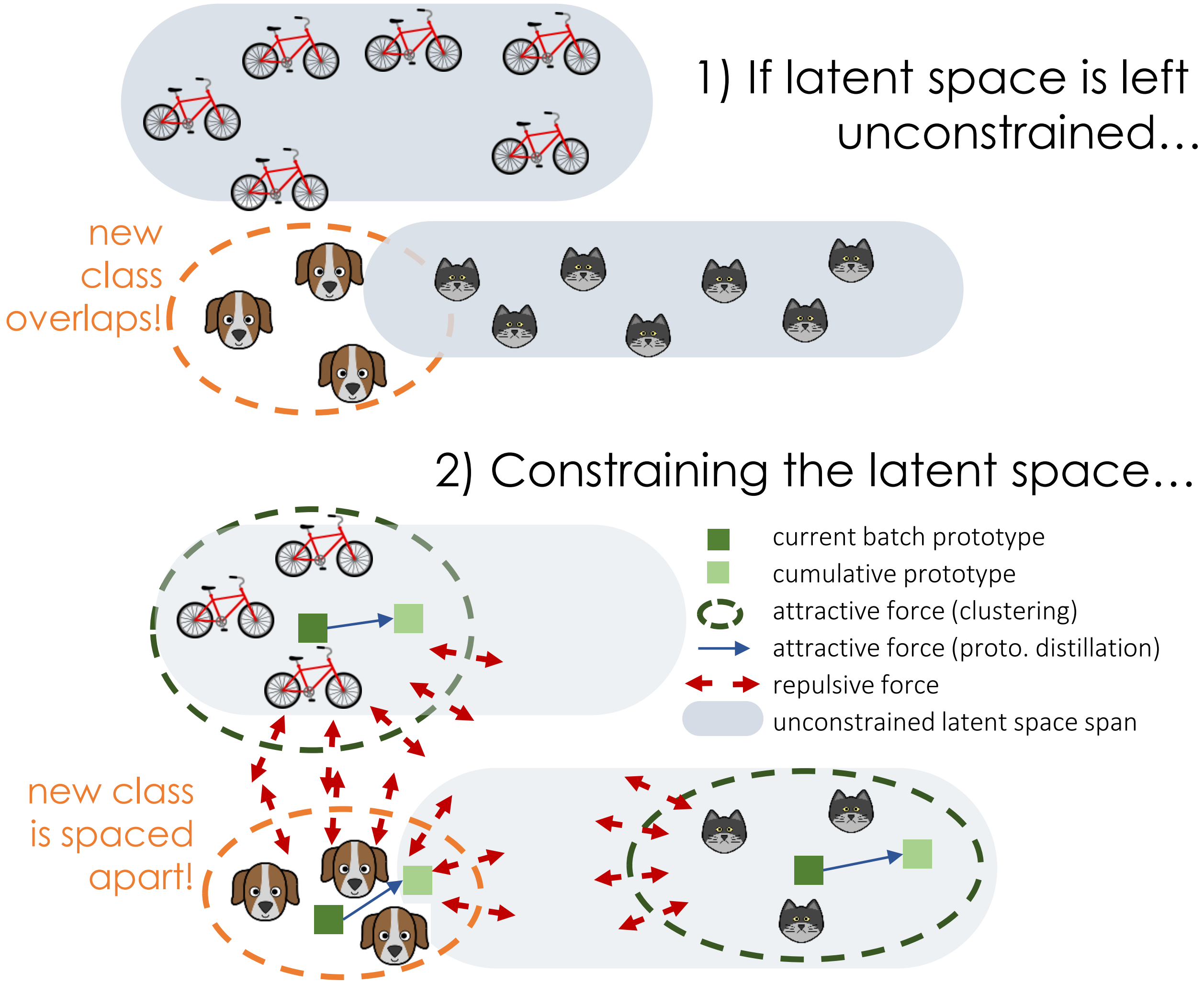


[1] Bengio, Y., et al., “Representation learning: A review and new perspectives”. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828 (2013).
[2] Girshick, R., et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014)

CL SS – Latent Space Regularization

Main limitation: visually similar classes are sometimes misled



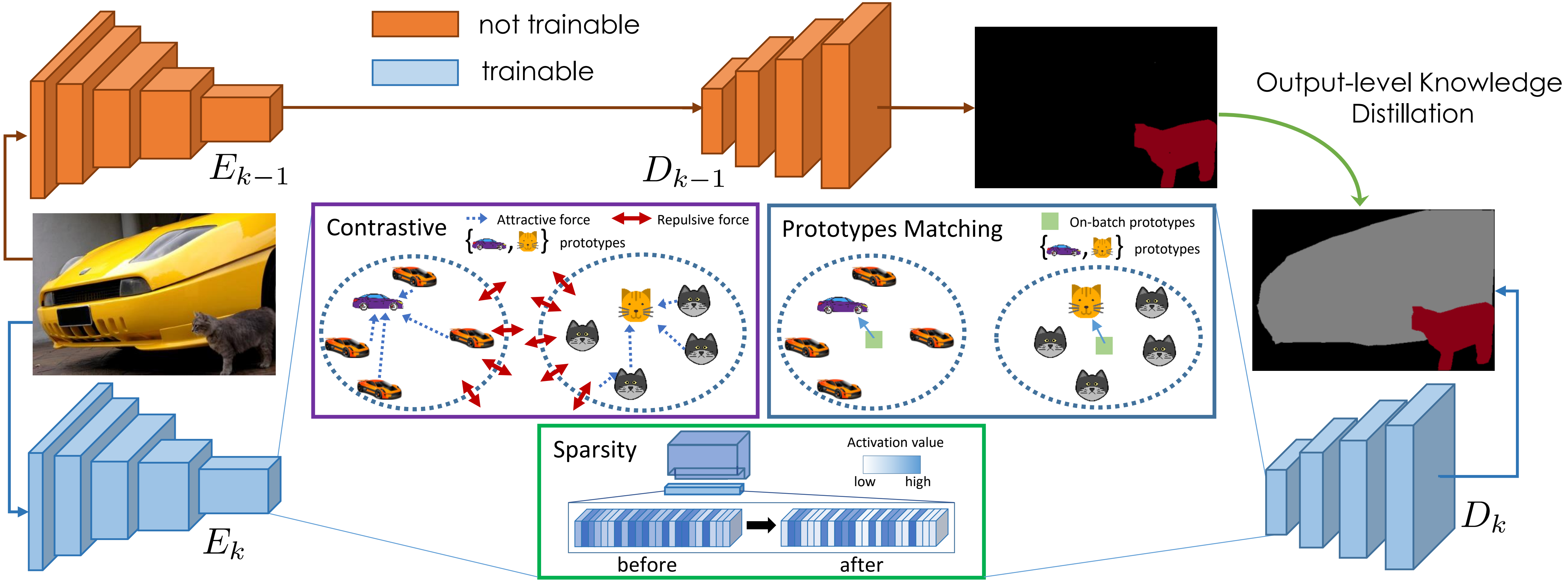


Michieli U. et al., "Continual Semantic Segmentation via Repulsion-Attraction of Sparse and Disentangled Latent Representations", **CVPR** 2021.

CL SS – Latent Space Regularization

SDR: **S**parse and **D**isentangled **R**epresentations

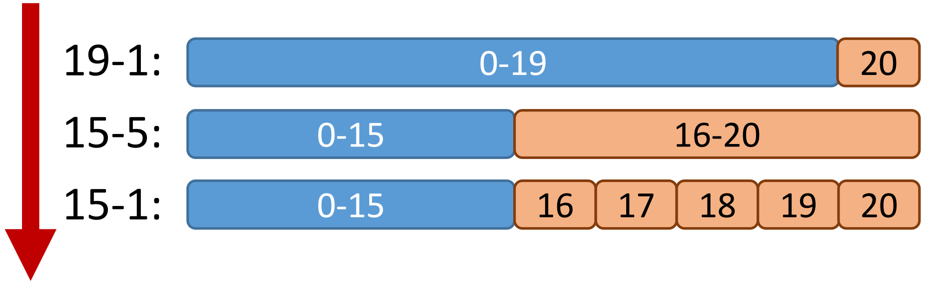
We combine task-related cross entropy loss with **4 constraints**:



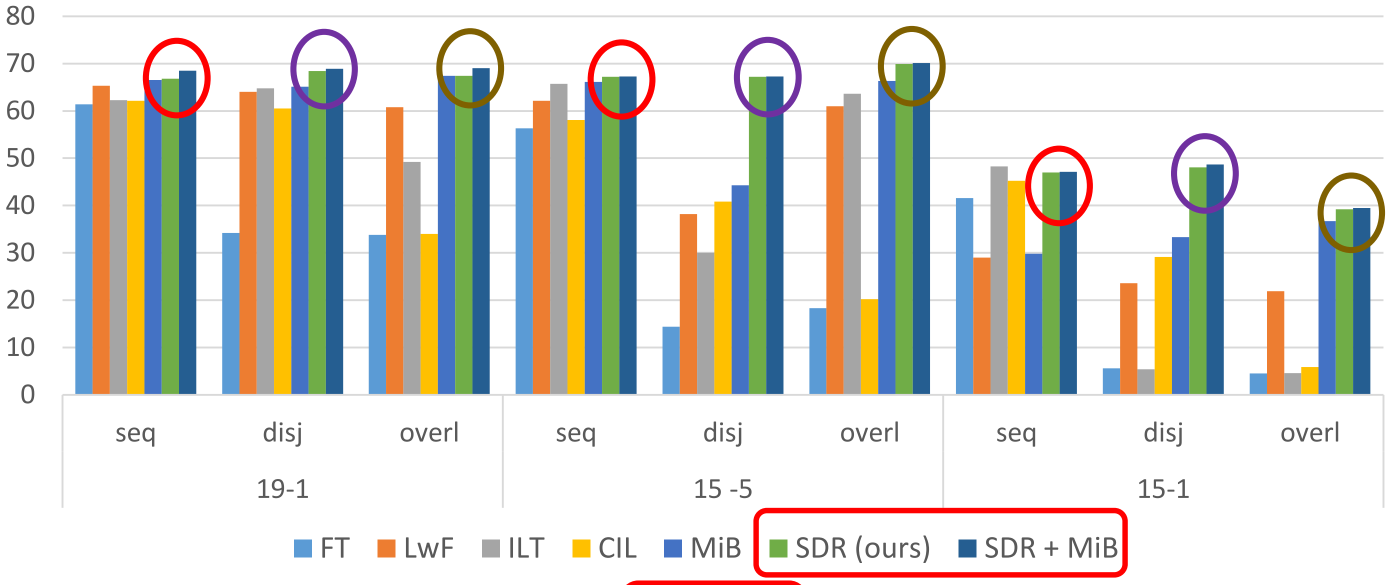
Michieli U. et al., "Continual Semantic Segmentation via Repulsion-Attraction of Sparse and Disentangled Latent Representations", **CVPR** 2021.

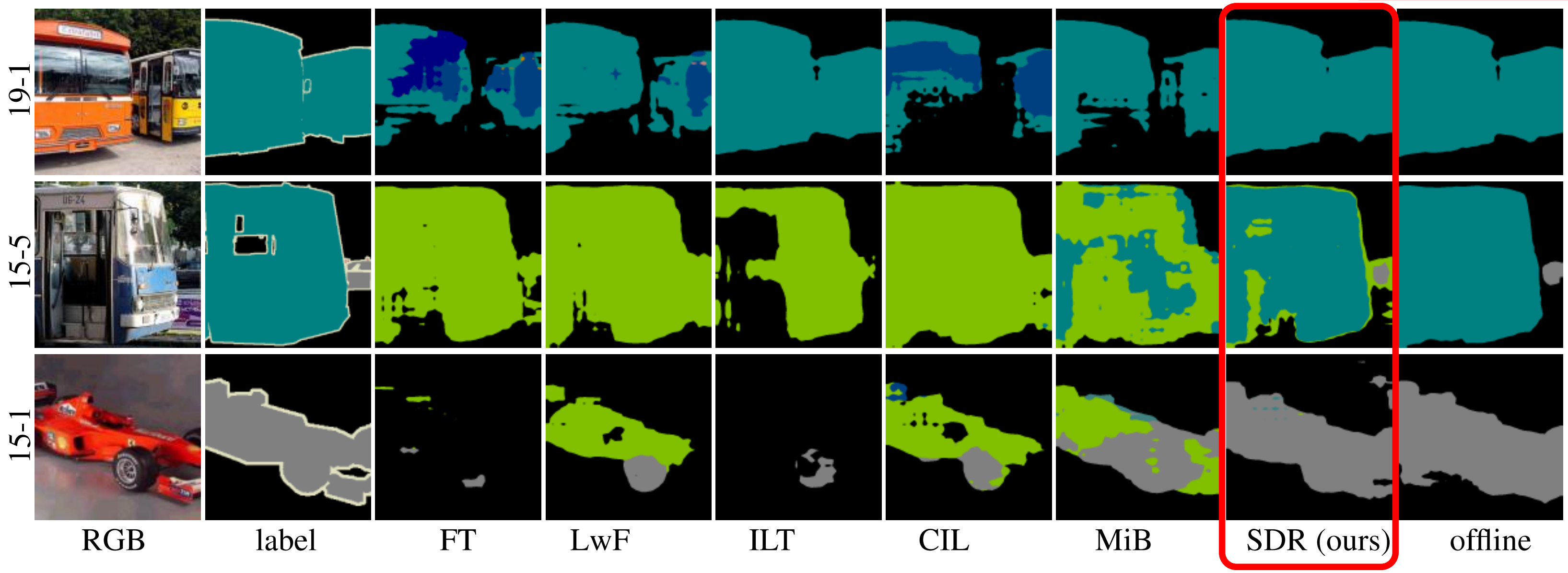
CL SS – Latent Space Regularization

Pascal VOC2012



More challenging



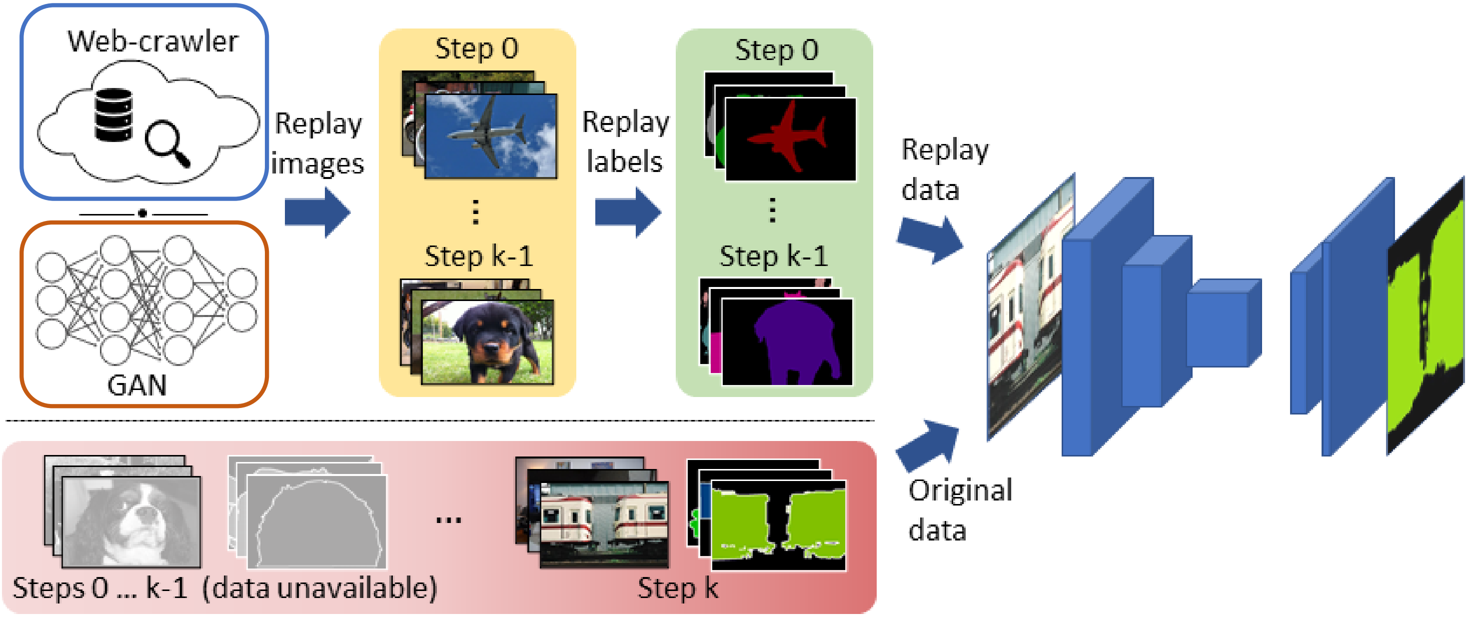


Michieli U. et al., "Continual Semantic Segmentation via Repulsion-Attraction of Sparse and Disentangled Latent Representations", **CVPR** 2021.

CL SS – Big Data Era: Replay Approach

Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", **ICCV** 2021.

CL SS – Replay Based

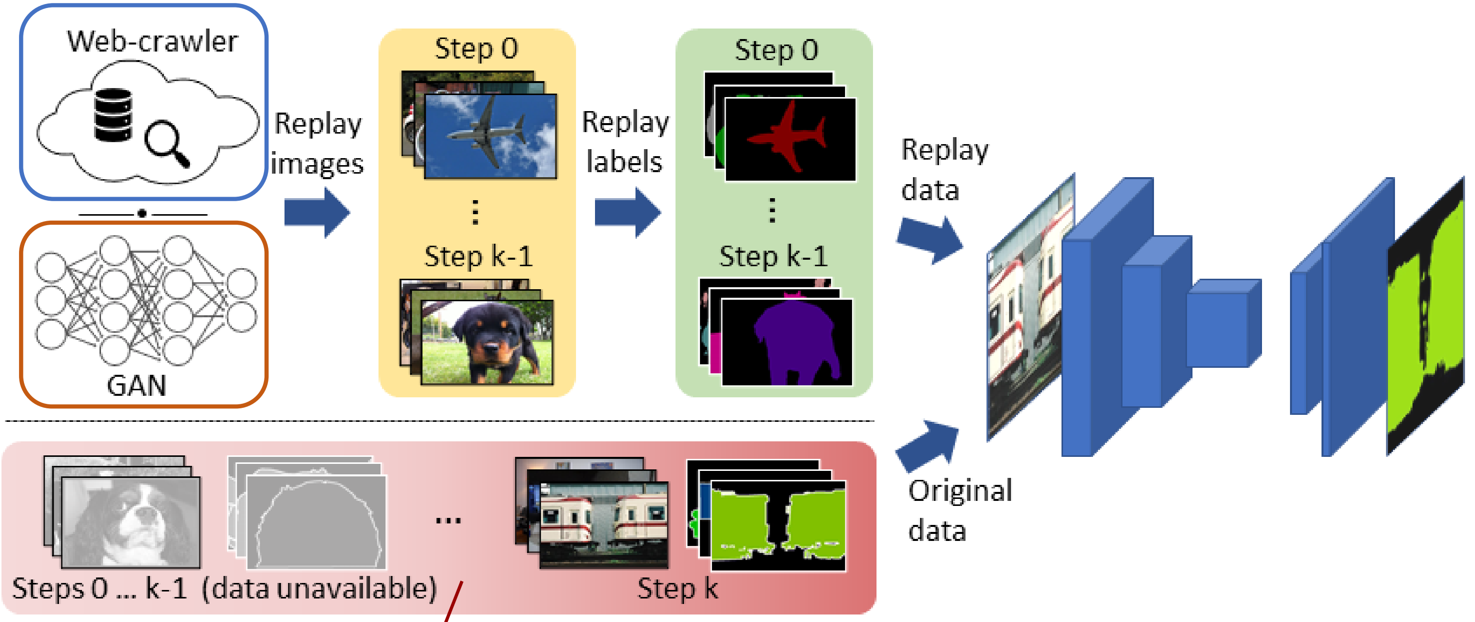


IDEA: interleave currently available data with replay samples to mitigate forgetting

Pascal

Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", **ICCV** 2021.

CL SS – Replay Based



- We use BigGAN
  Brock A., et al., "Large Scale GAN Training for High Fidelity Natural Image Synthesis" ICLR 2018
- We get images from a Web crawler

www.flickr.com

IDEA: interleave currently available data with replay samples to mitigate forgetting

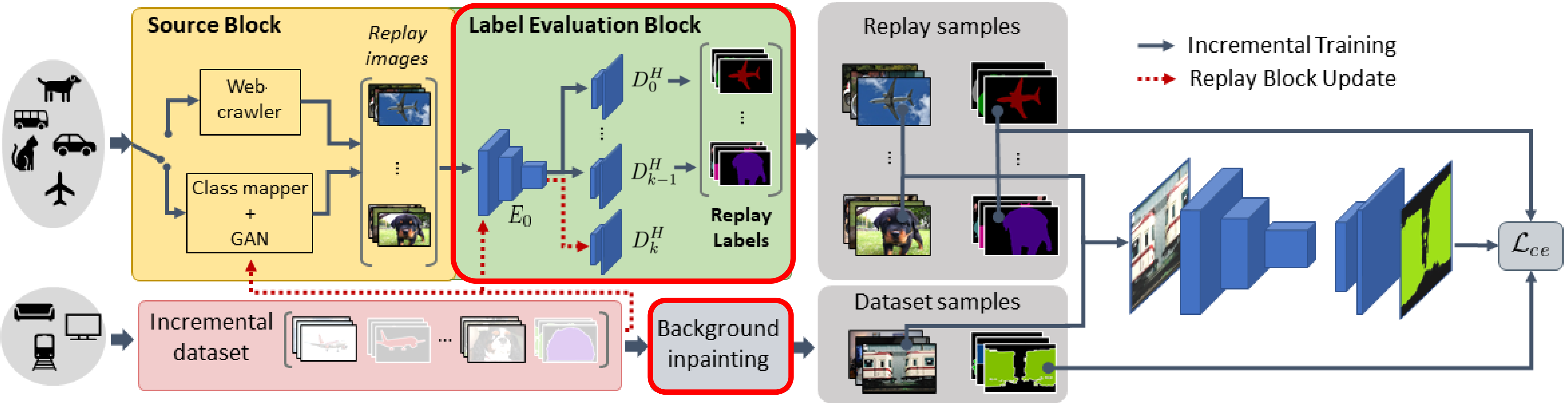
GAN

Web crawler

CL SS – Replay Based

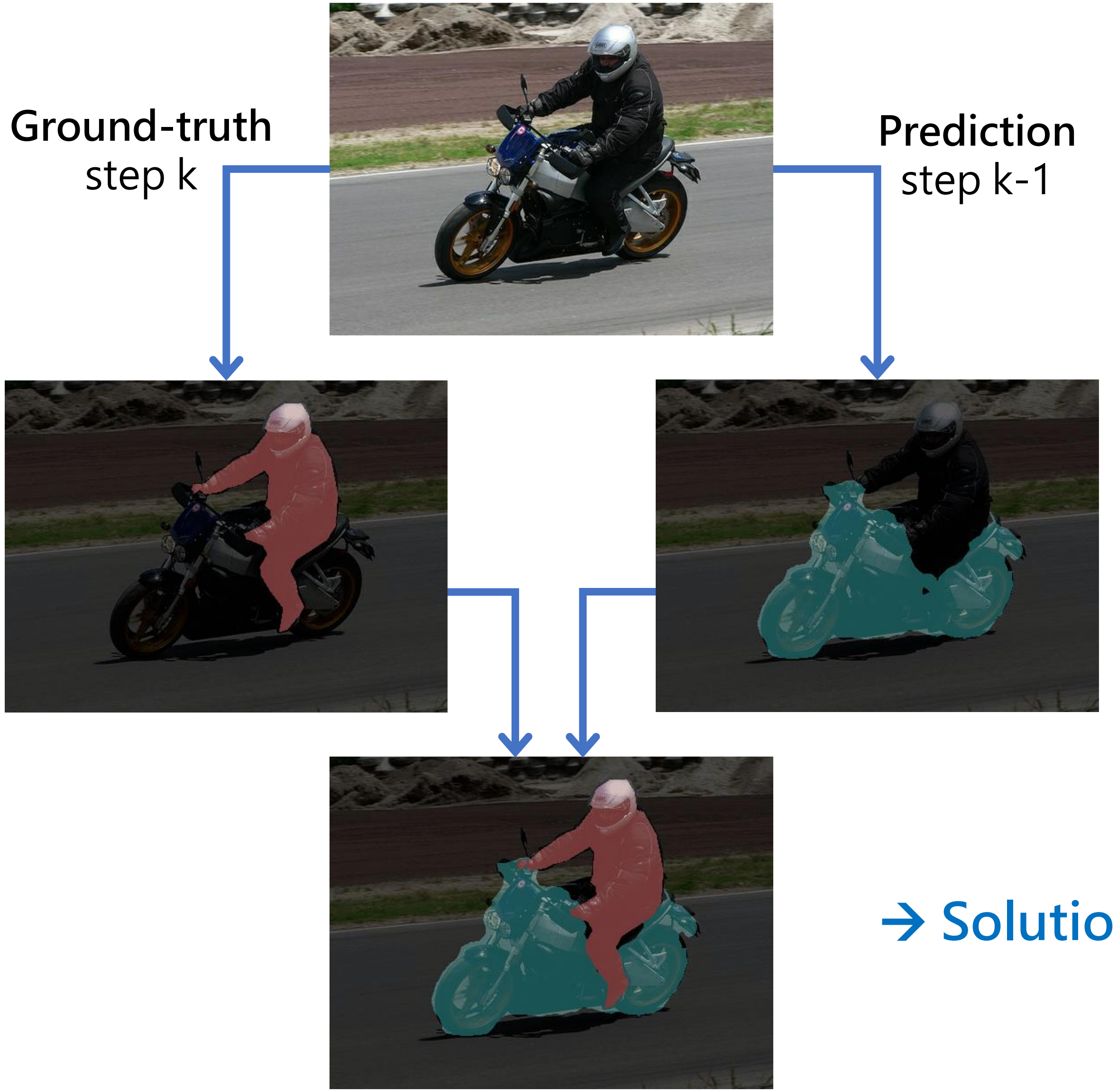
Problems:

- pseudo-labels for replay samples missing
- background-shift



Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", **ICCV** 2021.

CL SS – Replay Based



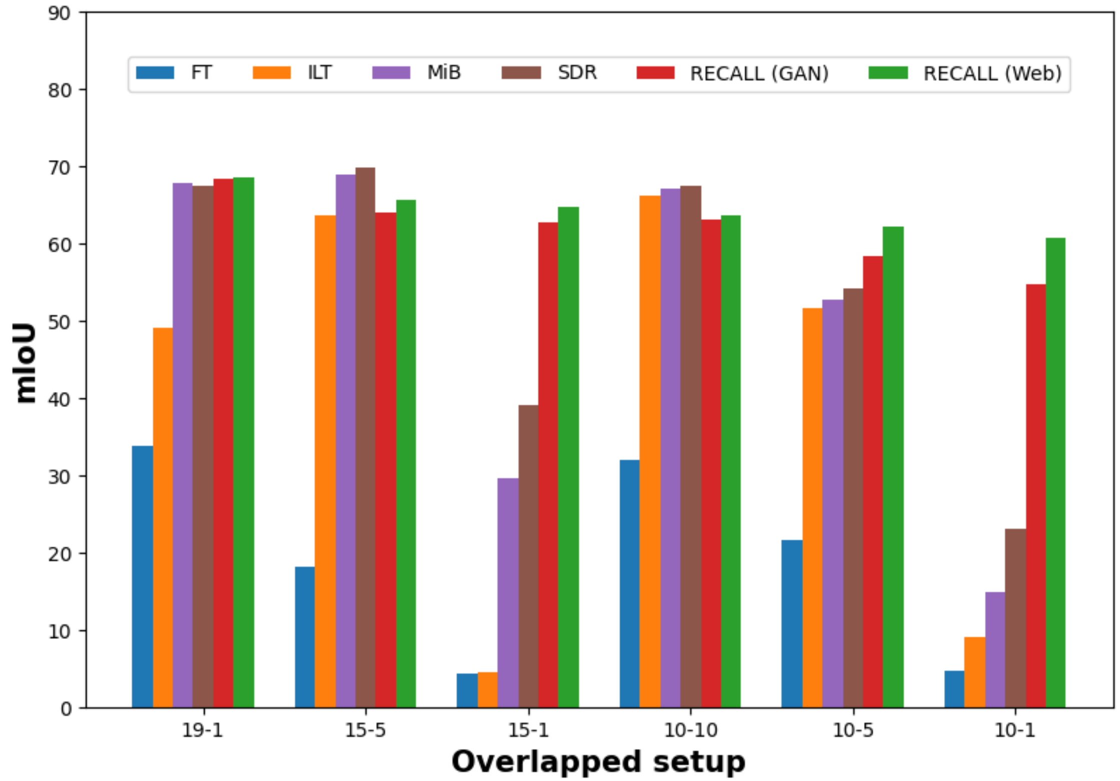
Incremental step k:

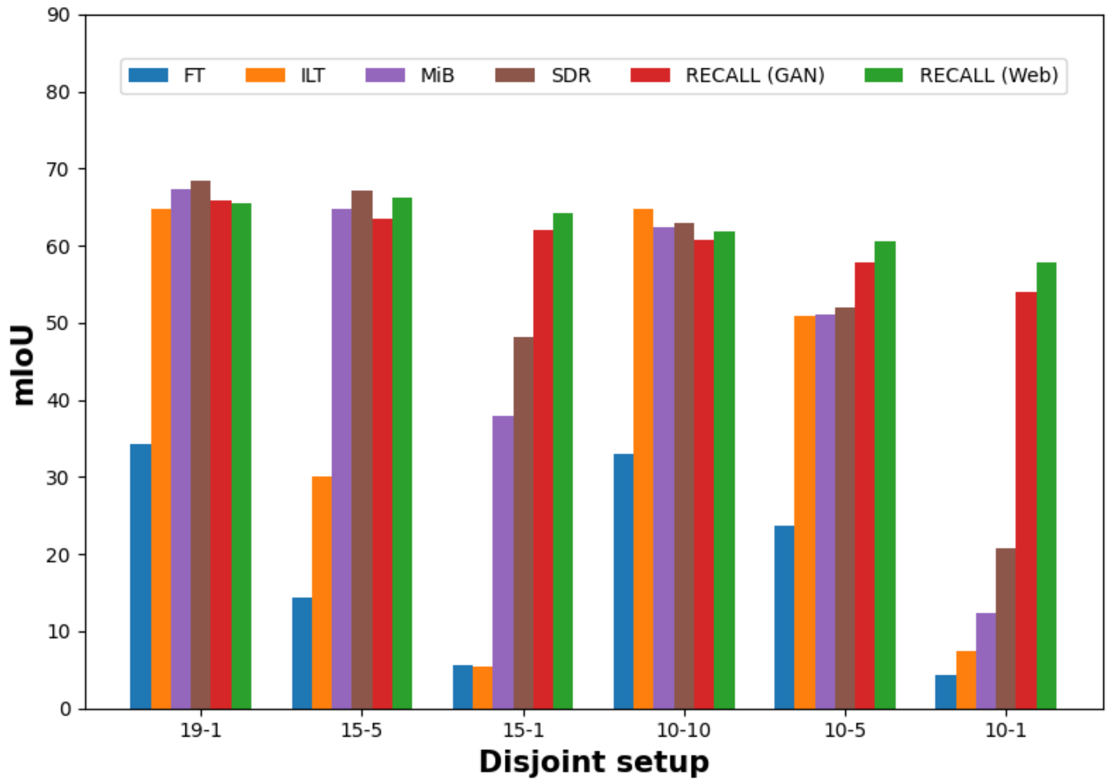
- Labels available only for new categories
- Past classes learnt in previous steps are annotated by *pseudo-labeling*

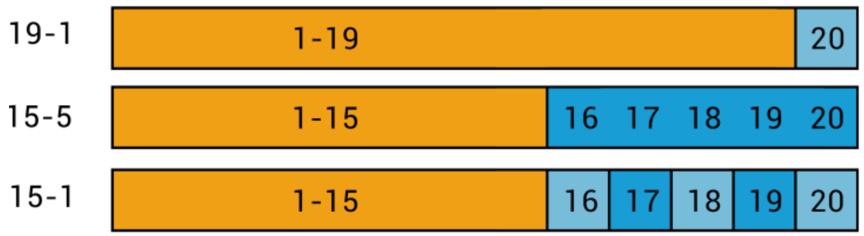
→ **Solution:** background inpainting

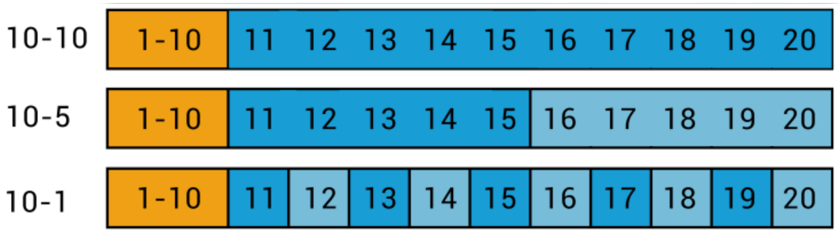
Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", **ICCV** 2021.

CL SS – Replay Based



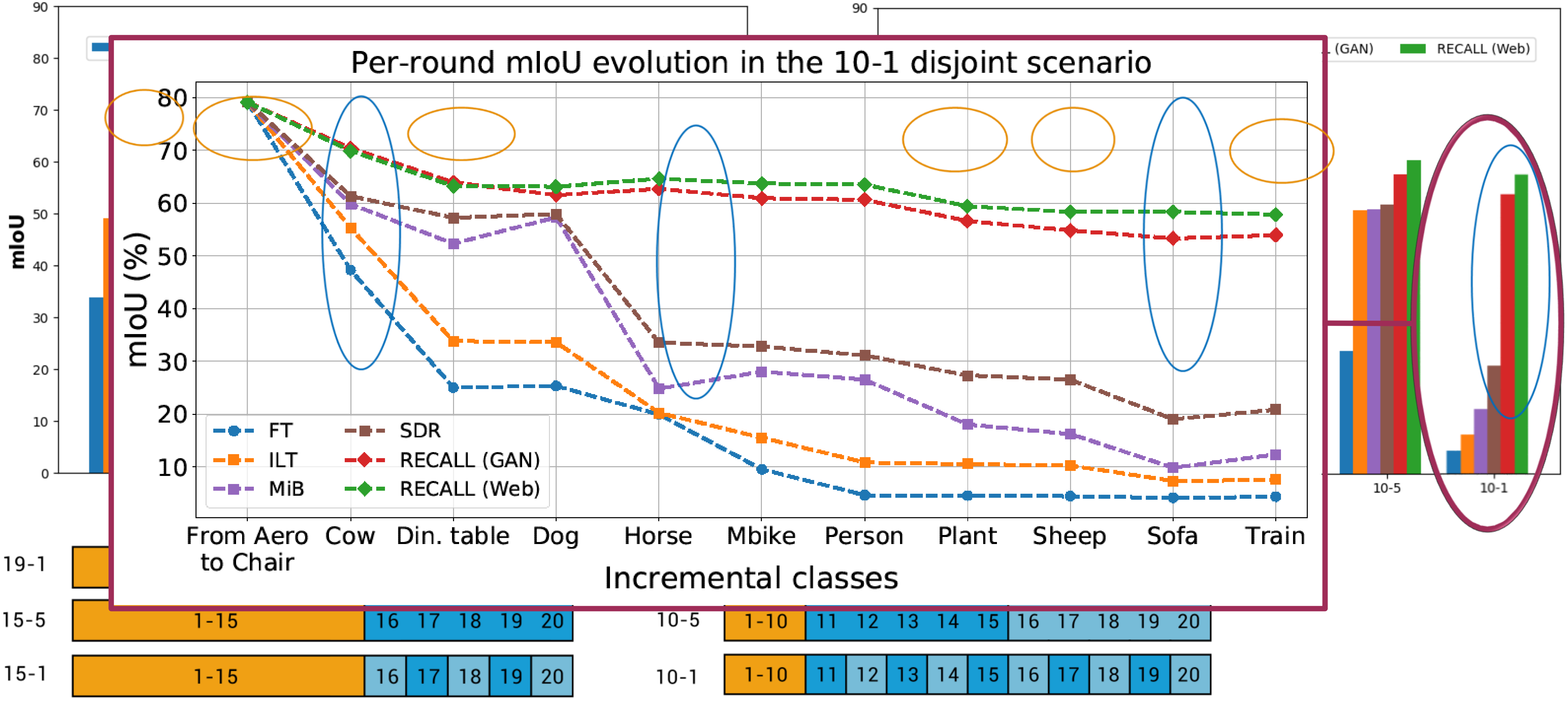






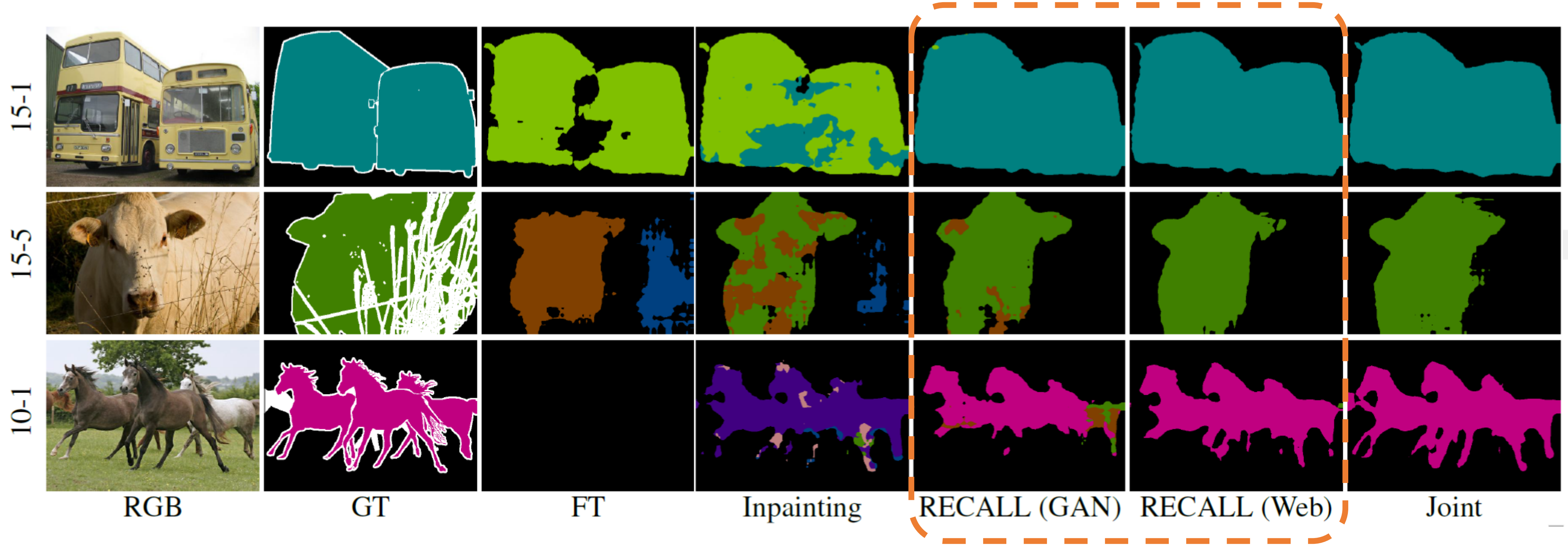
ILT: Michieli U. et al., "Incremental Learning Techniques for Semantic Segmentation", ICCVW 2019
MiB: Cermelli F. et al., "Modeling the background for incremental learning in semantic segmentation", CVPR 2020
SDR: Michieli U. et al., "Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations", CVPR 2021
RECALL: Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", ICCV 2021.

CL SS – Replay Based



ILT: Michieli U. et al., "Incremental Learning Techniques for Semantic Segmentation", ICCVW 2019
MiB: Cermelli F. et al., "Modeling the background for incremental learning in semantic segmentation", CVPR 2020
SDR: Michieli U. et al., "Continual semantic segmentation via repulsion-attraction of sparse and disentangled latent representations", CVPR 2021
RECALL: Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", ICCV 2021.

CL SS – Replay Based



- RECALL gets closer to *joint training*
- Visually similar classes are properly recognized: bus vs. train, sheep vs. cow

Maracani A. et al., "RECALL: Replay-based Continual Learning in Semantic Segmentation", **ICCV** 2021.

Takeaway on CL

- KD gives great boost of accuracy
  - **output** level
  - **feature** level
- Latent space regularization as effective technique
  - **prototype matching**
  - **contrastive learning**
  - **sparsity**
- Replay-based approaches & background inpainting
  - **GAN**
  - **Web (Flickr)**

- Complimentary to competing approaches
- Applicable to other tasks

Open Questions for Thesis

- **Web:** usage of web-crawled data (with weak supervision) to alleviate forgetting.
- **Coarse-to-Fine:** an initial training on a coarse set of classes aids a following training on a finer set of classes (hierarchically derived from coarser ones).
- **New tasks:** extend current techniques to *instance segmentation* or to account for *multiple input sources*.
- **New data:** extend current techniques to other data representations (depth, LiDAR, etc)

Outline

1) Continual Learning
   - Knowledge Distillation
   - Latent Space Regularization
   - Replay-based Approaches
   - ➢ **Notebook** on Continual Learning (20 minutes)

2) Unsupervised Domain Adaptation
   - Focus on Multimodal Learning

3) Federated Learning

Unsupervised Domain Adaptation in Semantic Segmentation

[1] Toldo M., Maracani A., Michieli U., Zanuttigh P., "Unsupervised Domain Adaptation in Semantic Segmentation: a Review", **Technologies**, 2020, 8, 35.
[2] Biasetton M., Michieli U., Agresti G., Zanuttigh P., "Unsupervised Domain Adaptation for Semantic Segmentation of Urban Scenes", **CVPRW** 2019.
[3] Michieli U., Biasetton M., Agresti G., Zanuttigh P., "Adversarial Learning and Self-Teaching Techniques for Domain Adaptation in Semantic Segmentation", **IEEE Transactions on Intelligent Vehicles** (T-IV), 2020.
[4] Spadotto T., Toldo M., Michieli U., Zanuttigh P., "Unsupervised Domain Adaptation with Multiple Domain Discriminators and Adaptive Self-Training", **ICPR** 2020.
[5] Toldo M., Michieli U., Agesti G., Zanuttigh P., "Unsupervised Domain Adaptation for Mobile Semantic Segmentation based on Cycle Consistency and Feature Alignment", **Elsevier Image and Vision Computing (IMAVIS)**, 2020.
[6] Toldo M., Michieli U., Zanuttigh P., "Unsupervised Domain Adaptation in Semantic Segmentation via Orthogonal and Clustered Embeddings", **WACV** 2021.
[7] Barbato F., Toldo M., Michieli U., Zanuttigh P., "Latent Space Regularization for Unsupervised Domain Adaptation in Semantic Segmentation", **CVPRW** 2021.
[8] Barbato F., Michieli U., Toldo M., Zanuttigh P., “Adapting Segmentation Networks to New Domains by Disentangling Latent Representations”, ArXiV 2021.

UDA – Definition

Consider 2 domain distributions:

Source data: large set, GT available, similar to target data (e.g., synthetic)

Target data: typically small set, GT expensive and error-prone (e.g., real)

AIM: Align the two distributions with no target labels

Train on source data → good accuracy on target data

SOURCE (SYNTHETIC)

GTA

~25k images
High quality
Car viewpoints

SYNTHIA

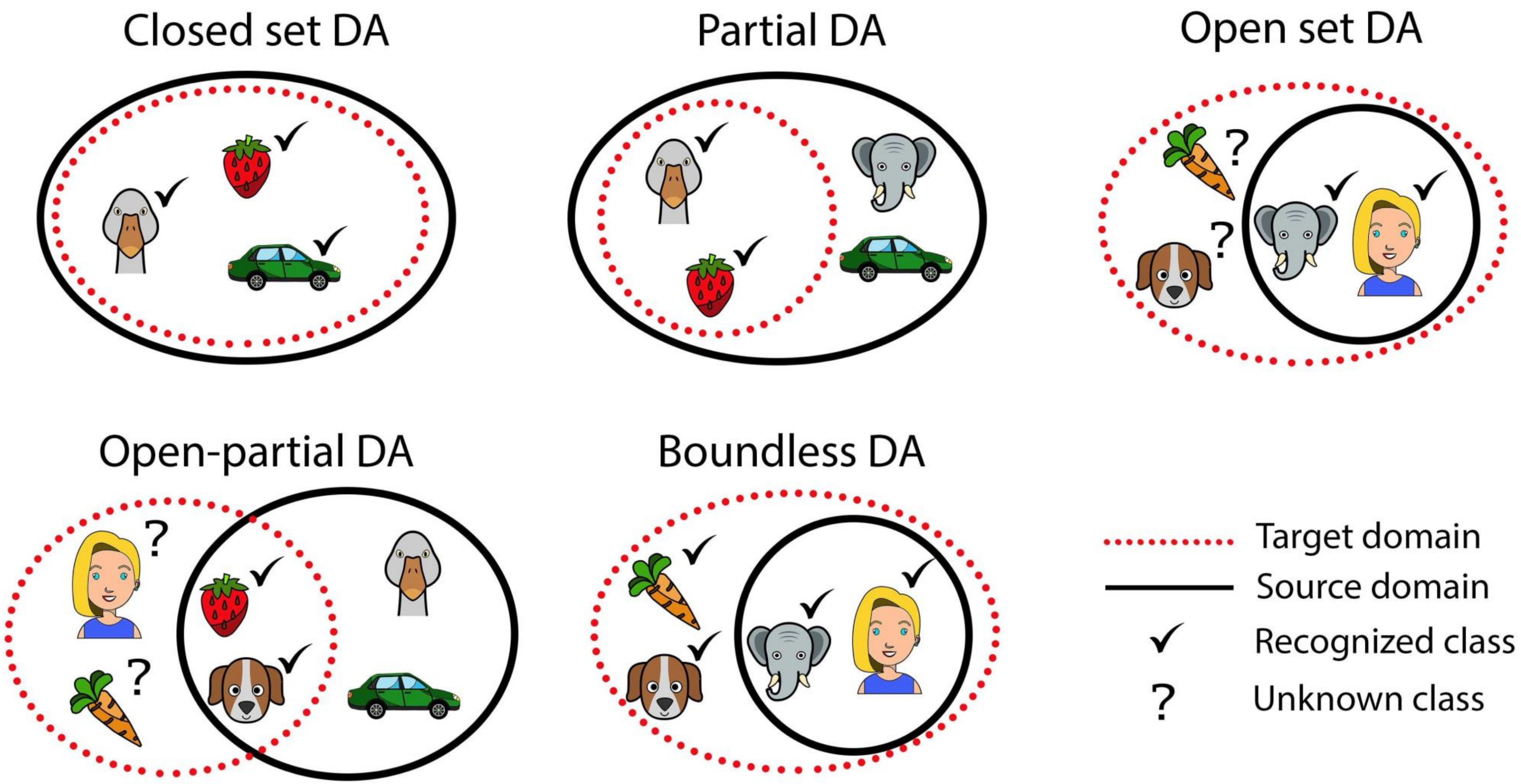
~9k images
Medium quality
Different viewpoints

TARGET (REAL WORLD)

CITYSCAPES

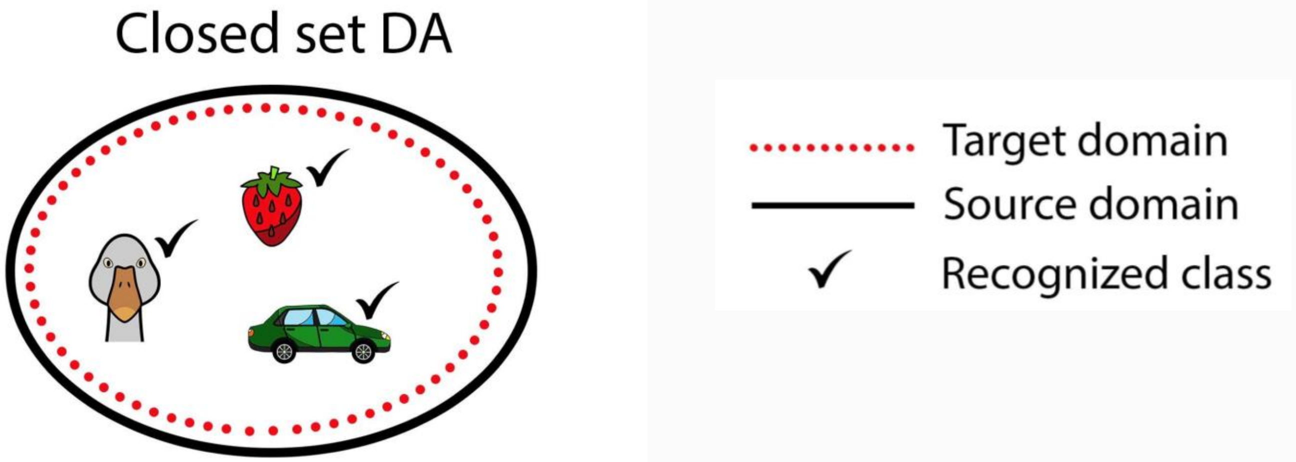
~3k images
Car viewpoints

UDA – Disambiguation



Toldo et al., "Unsupervised Domain Adaptation for Semantic Segmentation: a Survey", Technologies 2020.

UDA – Disambiguation



- All the possible classes appear in both the source **and** target domains
- Most traditional scenario

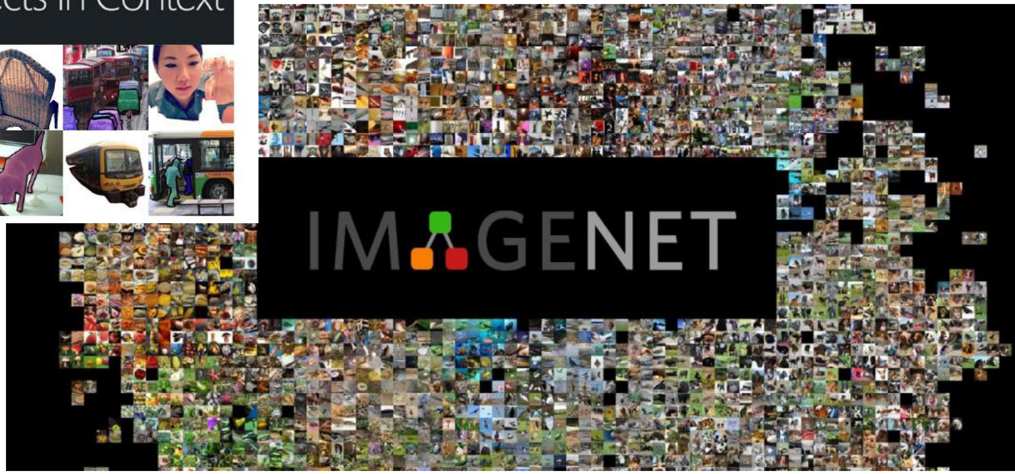
Toldo et al., "Unsupervised Domain Adaptation for Semantic Segmentation: a Survey", Technologies 2020.

UDA SS - Motivation

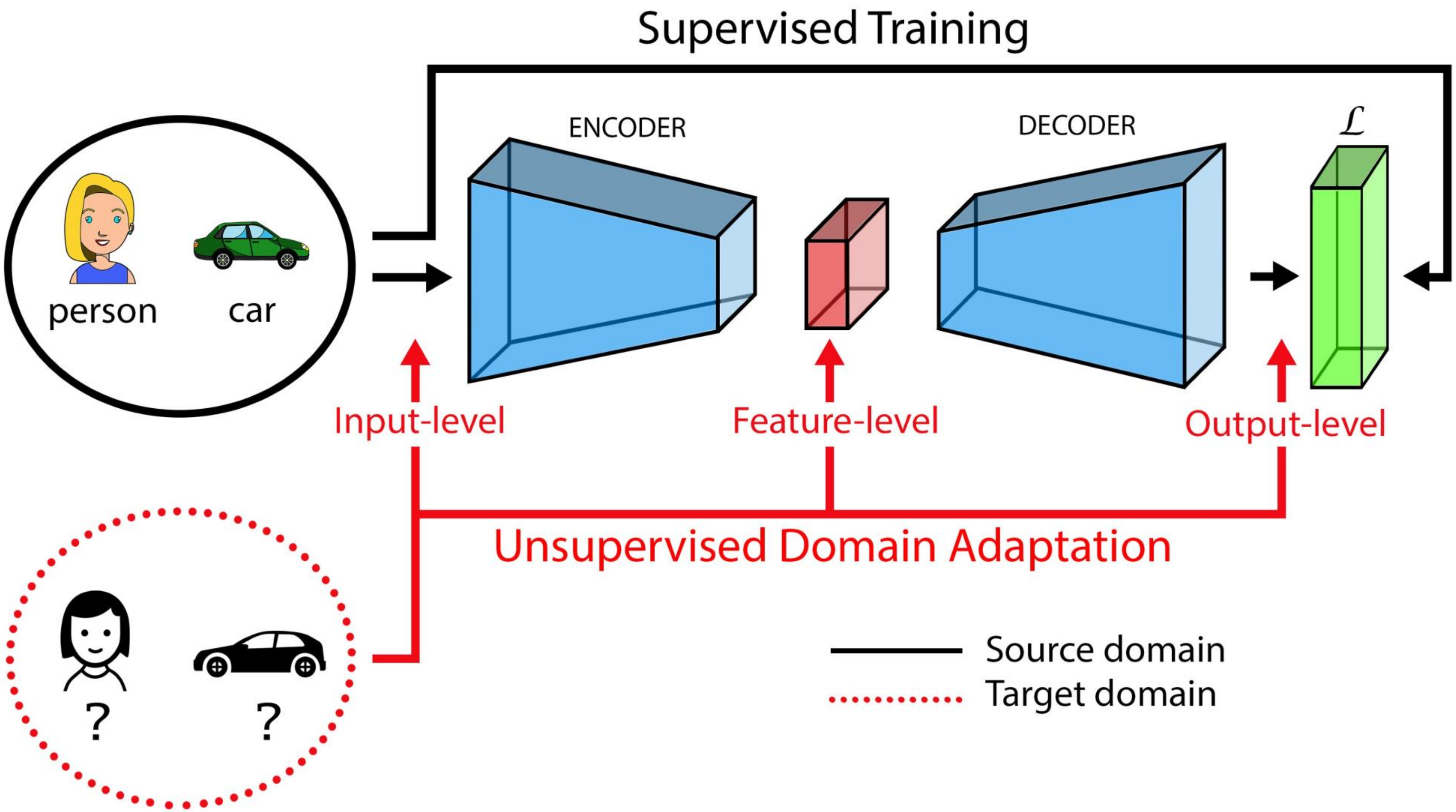
Pixel-level annotations are:

1) expensive: commercial rate ranges from 2 to 10 USD **per image**!
(...and you need thousands images...)
2) error prone
3) time-consuming
4) some problems are inherently *sample-limited* (e.g. face recognition or person re-identification)

plenty of labelled images exist in similar contexts and knowledge can be transferred



UDA SS – Definition

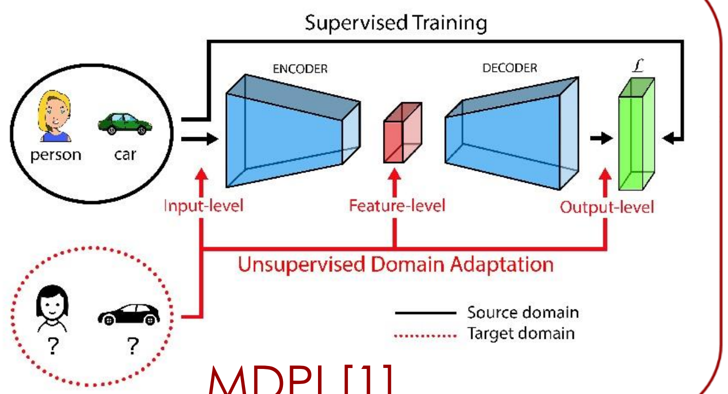


Toldo et al., "Unsupervised Domain Adaptation for Semantic Segmentation: a Survey", Technologies 2020.

UDA SS – Our Works

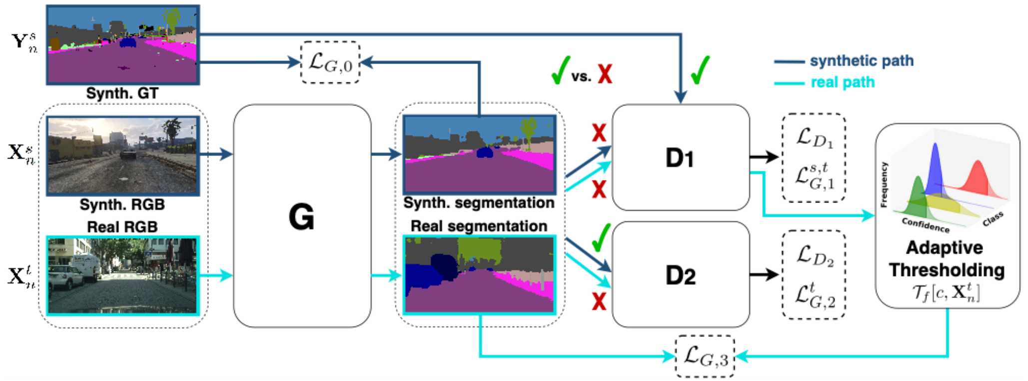
[6] Toldo M. et al., "Unsupervised Domain Adaptation in Semantic Segmentation via Orthogonal and Clustered Embeddings", **WACV** 2021.
[7] Barbato F. et al., "Latent Space Regularization for Unsupervised Domain Adaptation in Semantic Segmentation", **CVPRW** 2021.
[8] Barbato F. et al., "Adapting Segmentation Networks to New Domains by Disentangling Latent Representations", ArXiV 2021.

Review paper



MDPI [1]

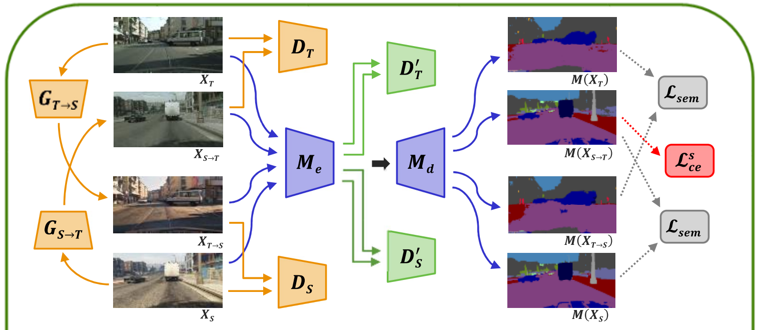
Output-level adaptation



CVPRW [2]
TIV [3]
ICPR [4]

Self-training based on confidence from predicted output

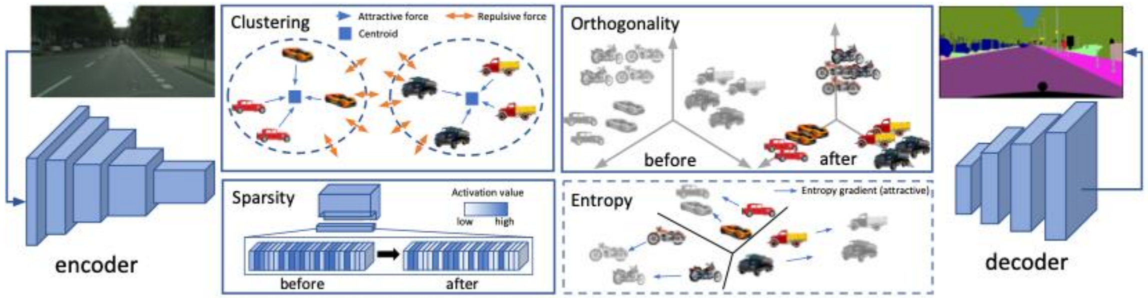
Input- and Feat-level adaptation



CycleGAN + feat level adversarial

IMAVIS [5]

Feature-level adaptation



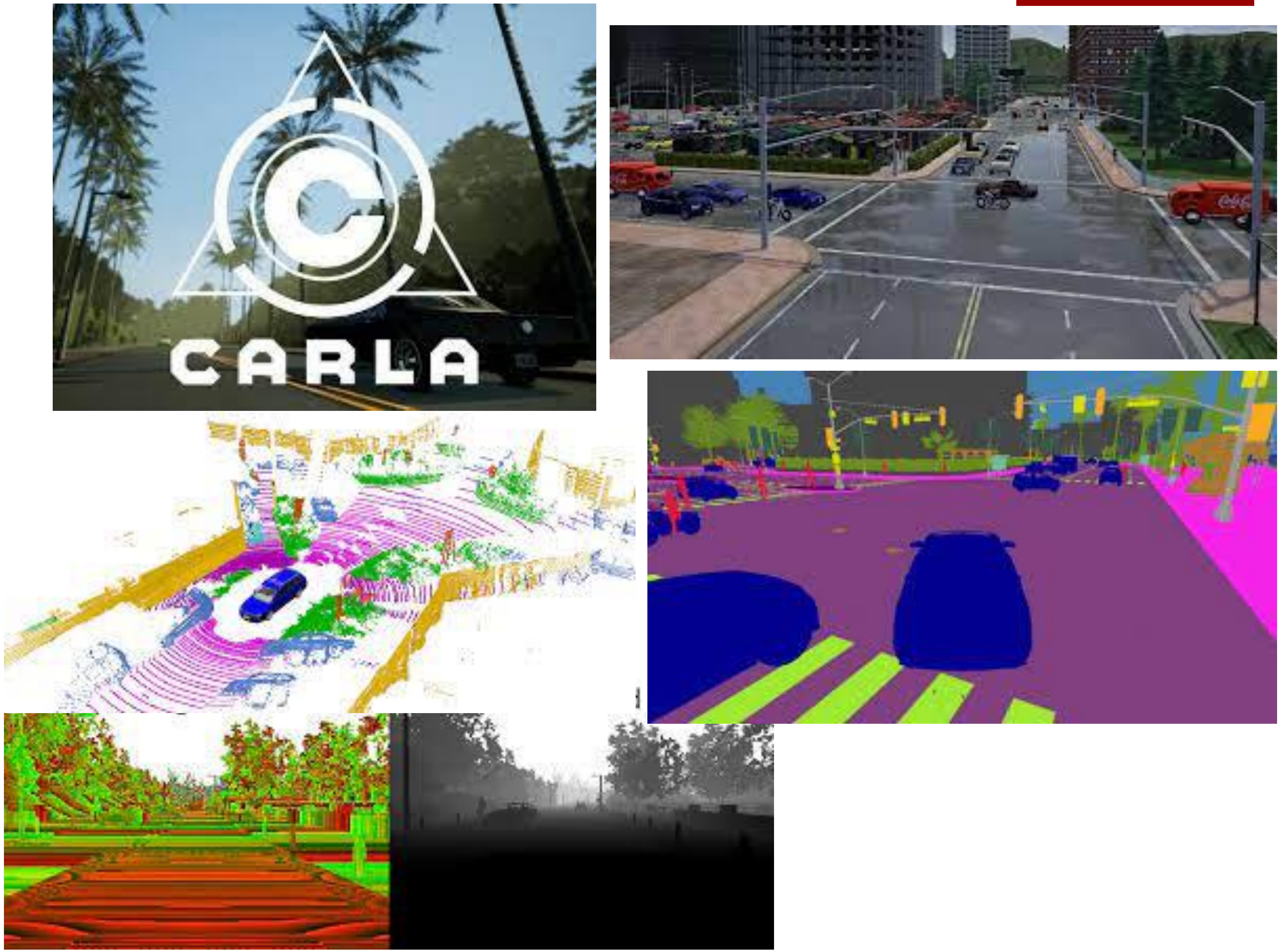
WACV [6]
CVPRW [7]
ArXiV [8]

Feature clustering (if same label), repulsion (if different label), orthogonality, sparsity, with same norm

[1] Toldo M. et al., "Unsupervised Domain Adaptation in Semantic Segmentation: a Review", **Technologies**, 2020, 8, 35.
[2] Biasetton M. et al., "Unsupervised Domain Adaptation for Semantic Segmentation of Urban Scenes", **CVPRW** 2019.
[3] Michieli U. et al., "Adversarial Learning and Self-Teaching Techniques for Domain Adaptation in Semantic Segmentation", **IEEE Transactions on Intelligent Vehicles** (T-IV), 2020.
[4] Spadotto T. et al., "Unsupervised Domain Adaptation with Multiple Domain Discriminators and Adaptive Self-Training", **ICPR** 2020.
[5] Toldo M. et al., "Unsupervised Domain Adaptation for Mobile Semantic Segmentation based on Cycle Consistency and Feature Alignment", **Elsevier Image and Vision Computing (IMAVIS)**, 2020.

UDA SS – New CARLA Dataset

- Joint effort with the SIGNET group @UNIPD
- **Large scale**: > 30.000 (unique) samples
- **27 conditions:** 3 times of day, 9 weather conditions (rainy/foggy/cloudy/…)
- **Multimodal**: 7 **RGB**, 7 **depth** cameras, 3 **LiDAR** → 15TB
- **Multiple tasks**: semantic segmentation, object detection,
  depth estimation,…



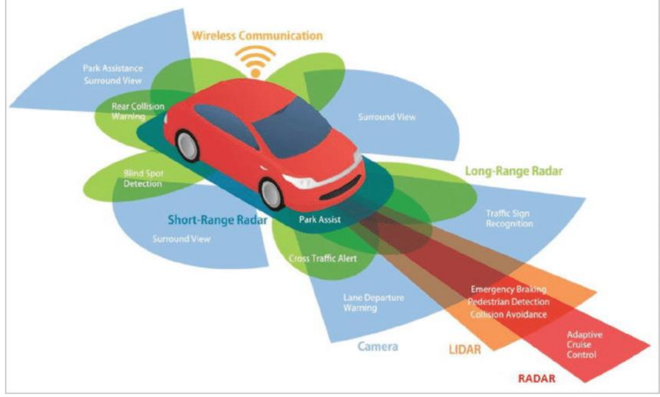
Open Questions for Thesis

- **Adaptation to different weather or time of day conditions:**

  Current datasets are highly unbalanced and contain almost always sunny images.

  What if it rains, or if it is cloudy, or if there is fog?

  We want trained networks to be invariant to the *style* of the target domain.



- **Usage of multimodal data** (e.g., RGB, depth cameras, LiDAR):

  Different sensors operate differently in different circumstances.

  If night: RGB camera is heavily affected, while LiDAR is not.

  Similarly during rain.

  How can we exploit clues from the different sensors to **build a more reliable model?**

- **Many other possibilities and you can propose your own idea**

Outline

1) Continual Learning
   - Knowledge Distillation
   - Latent Space Regularization
   - Replay-based Approaches
   - ➢ **Notebook** on Continual Learning (20 minutes)
2) Unsupervised Domain Adaptation
   - Focus on Multimodal Learning
3) Federated Learning

SAMSUNG

Samsung Research

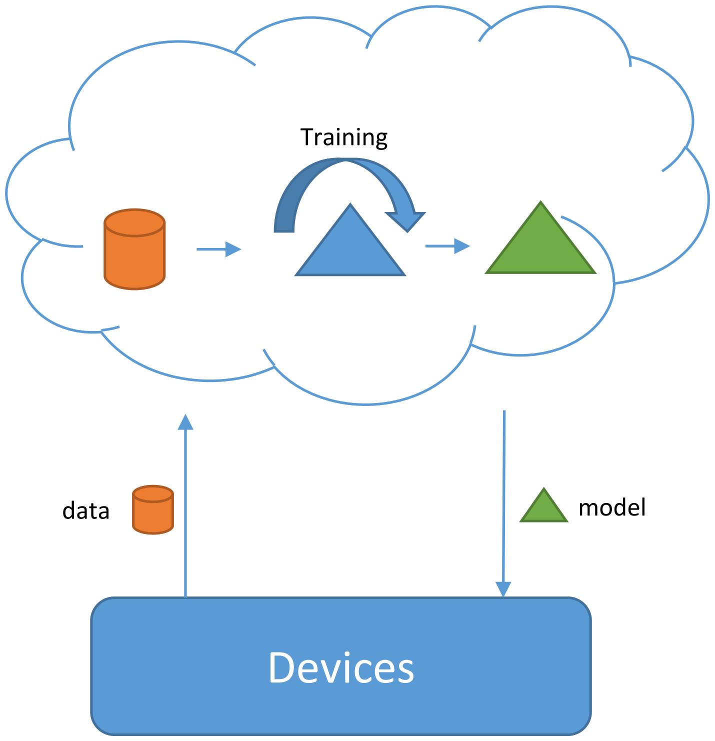
Federated Learning of Visual Feature Representations

[1] Michieli U. and Ozay M., "Are All Users Treated Fairly in Federated Learning Systems?", **CVPRW** 2021.
[2] Michieli U. and Ozay M., "Federated Learning of Visual Feature Representations", *ArXiV 2021*.

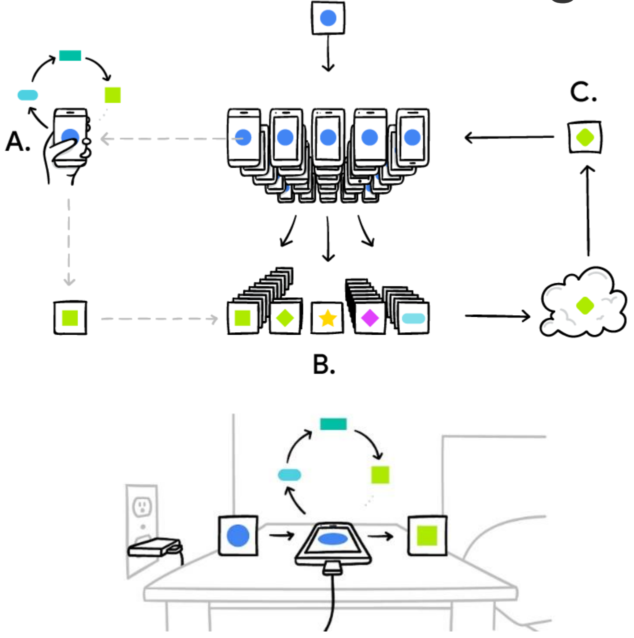
Federated Learning

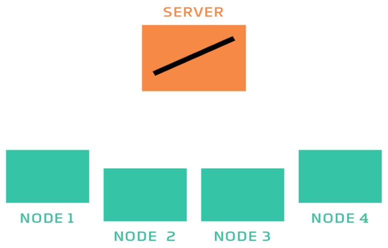
- Federated Learning: Distributed Machine Learning on **Heterogeneous Data**

Traditional (centralized) Learning



Federated Learning





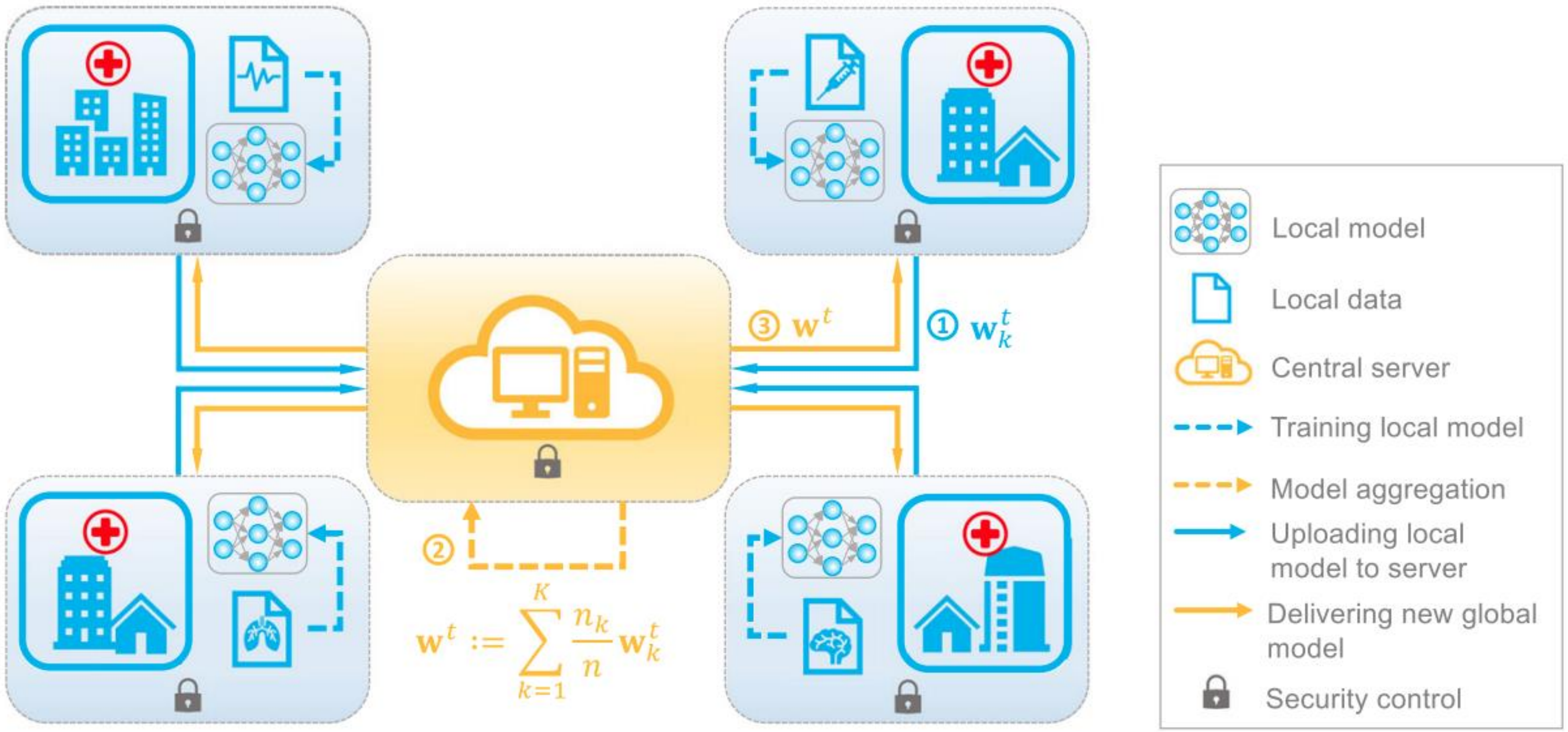
Benefits

- Ensuring Privacy
- No data breach risk
- Lower server costs
- Multiple modalities
- Data availability

Federated Learning

PROs

- avoid raw data exchanges for **privacy**
- **less communication** cost
- **more data** is available (user can opt out from data sharing)



Federated Learning

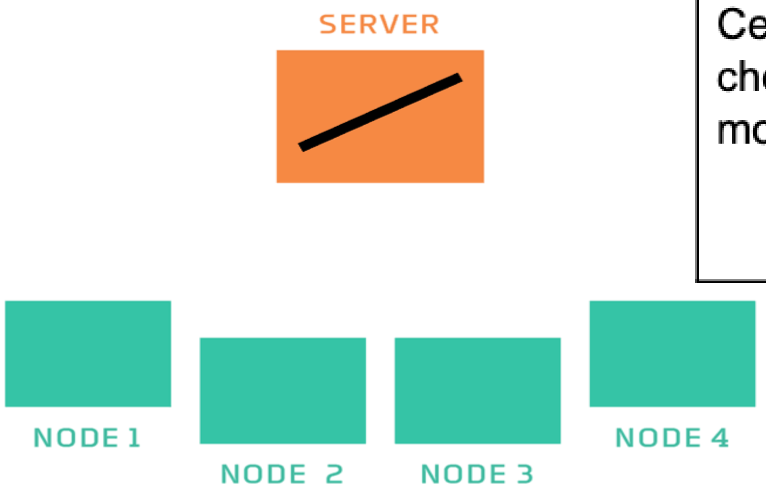
- Differences between Distributed Learning and Federated Learning

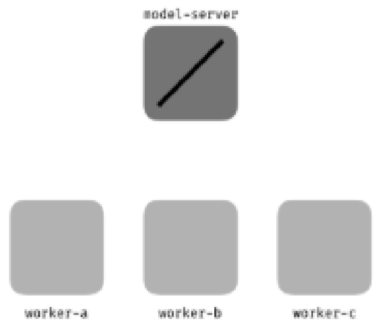
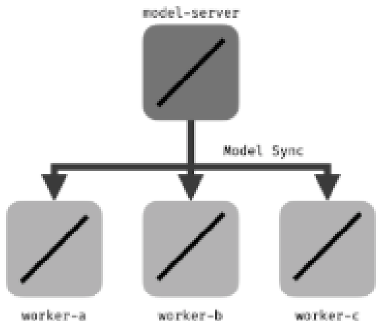
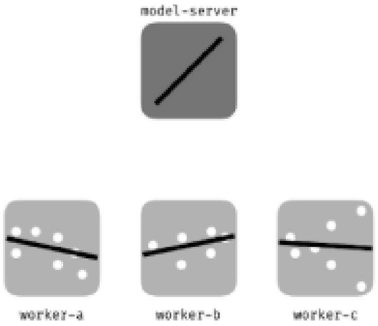
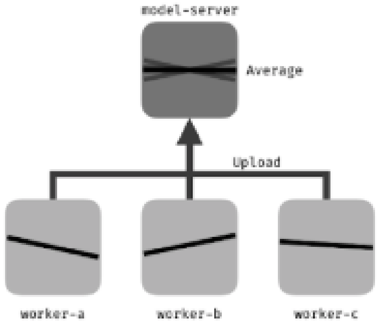
Distributed Learning	**Federated Learning**
Both aim at training a single model on multiple nodes	
Focus: Parallelizing computing power	Focus: Training on heterogeneous datasets
❖ Distributed data on each client has ~ **same size**	❖ Distributed data on each client has **not same size** (e.g., powerlaw)
❖ Data are distributed **i.i.d.**	❖ Data are distributed **non-i.i.d.**
❖ Nodes are typically (**reliable**) datacenters	❖ Clients may be **unreliable** (low battery, WiFi, etc)

Federated Learning

- Our focus will be on the *classical* **Centralized Federated Learning**

Coordination of a central server



Step 1	Step 2	Step 3	Step 4
			
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

Federated Learning

In general, federated aggregation can be expressed as:

Central model parameters

$$\boxed{\boldsymbol{w}_{t+1}} \leftarrow \sum_{k=1}^{\boxed{K}} \boxed{\boldsymbol{w}_t^k}\, \boldsymbol{a}_k^t$$

clients

Local model parameters

Attention vector

❖ FedAvg [1]: $\boldsymbol{a}_k = \frac{\boxed{n_k}}{\boxed{n}}$

samples of participant k

samples of all participants

❖ FedAtt [2]: $\boldsymbol{a}_k^t = f\left(\boldsymbol{w}_{t+1}, \boldsymbol{w}_t^k\right)$

Function of change of local weights vs. aggregate weight

❖ FairAvg [3]: $\boldsymbol{a}_k^t = 1/K$ (constant)

→ To show effect of data imbalance and non-iid-ness across clients

[1] McMahan B., et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial Intelligence and Statistics*. PMLR, 2017.
[2] Ji S., et al. "Learning private neural language modeling with attentive aggregation." 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019.
[3] Michieli U. and Ozay M., "Are All Users Treated Fairly in Federated Learning Systems?", CVPRW RCV 2021

FL – Our Works

[1]
- ❑ **Many clients** contribute little while aggregation
- ❑ **Few clients** tend to *dominate the scene*

→ If data is highly non-i.i.d. this represents a problem for convergence

Central model parameters

\# clients

Attention vector

$$\boxed{\boldsymbol{w}_{t+1}} \leftarrow \sum_{k=1}^{\boxed{K}} \boxed{\boldsymbol{w}_t^k}\ \boldsymbol{a}_k^t$$

Local model parameters

[2]
→ $\boldsymbol{a}_k^t$ influences federated training
IDEA: build an $\boldsymbol{a}_k^t$ *features-driven*

- **2 quantitative** metrics
  - Margin between prototypes → higher is better
  - Distance between features of FL approach and centralized training → lower is better
- **Qualitative** metrics based on **entropy** maps → explainability

A model is more than its set of weights

[1] Michieli U. and Ozay M., "Are All Users Treated Fairly in Federated Learning Systems?", **CVPRW** 2021.
[2] Michieli U. and Ozay M., "Federated Learning of Visual Feature Representations", *ArXiV 2021*.

Open Questions for Thesis

- **Analysis of FL algorithms** in terms of accuracy, convergence rate, communication cost,…

- **Personalized models:**

  Is it better a mediocre general model or an expert personalized model?

- **Explainability of FL models**

  Which clients or domain distributions influenced more the aggregate FL model?

  In relation to the centralized case

- **Propose new FL optimizers….**

Any Questions?

https://umbertomichieli.github.io/

https://lttm.dei.unipd.it/

umberto.michieli@unipd.it