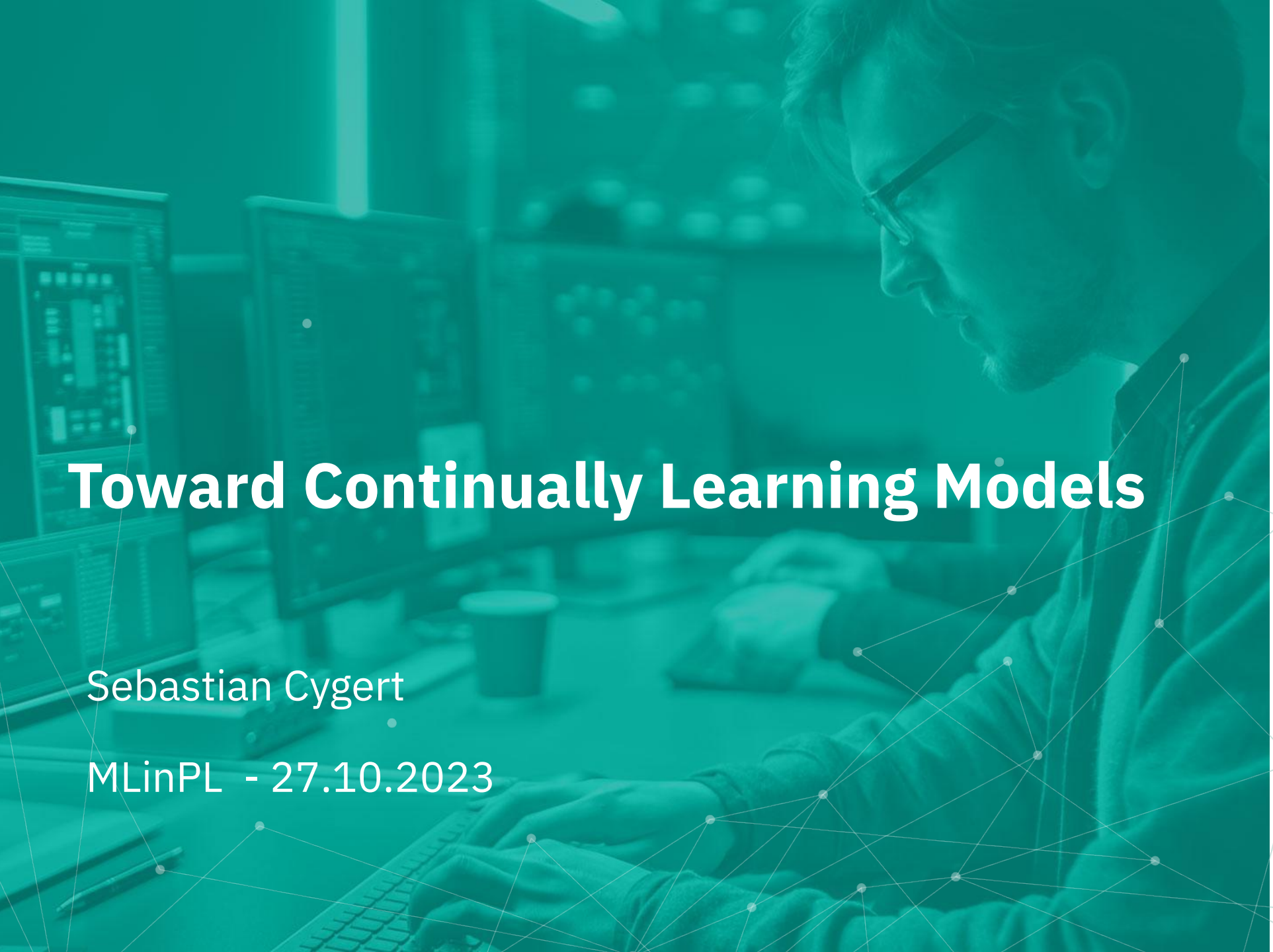


Toward Continually Learning Models

Sebastian Cygert

MLinPL - 27.10.2023

IDEAS NCBR

GDAŃSK UNIVERSITY OF TECHNOLOGY

IDEAS NCBR

Outline

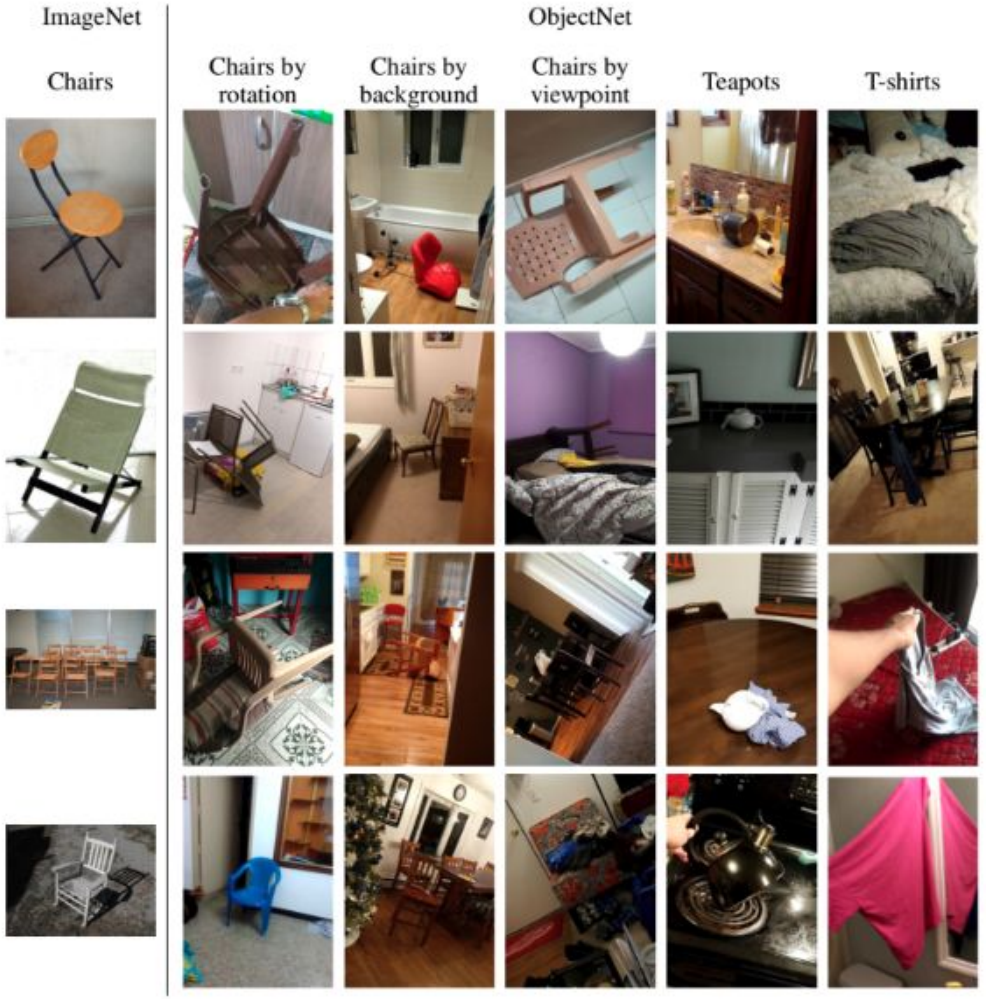
1. **Continual Learning - motivation**
2. Test-time adaptation on synthetic distribution shift
3. Proposed method for natural shifts

Challenges in the Real-World

Challenges in autonomous robots

1st ICML 2022 Workshop on Safe Learning for Autonomous Driving (SL4AD)

Shift Happens (in the real-world)



Barbu, Andrei, et al. "Objectnet: A large-scale bias-controlled dataset for pushing the limits of object recognition models." *Advances in neural information processing systems* 32 (2019).

IDEAS NCBR

Can we continue to learn after the deployment?

Continual Learning

Continual Learning

Continual learning is a machine learning paradigm that focuses on training models to acquire and retain knowledge over an extended period on a stream of data.

Continual Learning

Continual learning is a machine learning paradigm that focuses on training models to acquire and retain knowledge over an extended period on a stream of data.

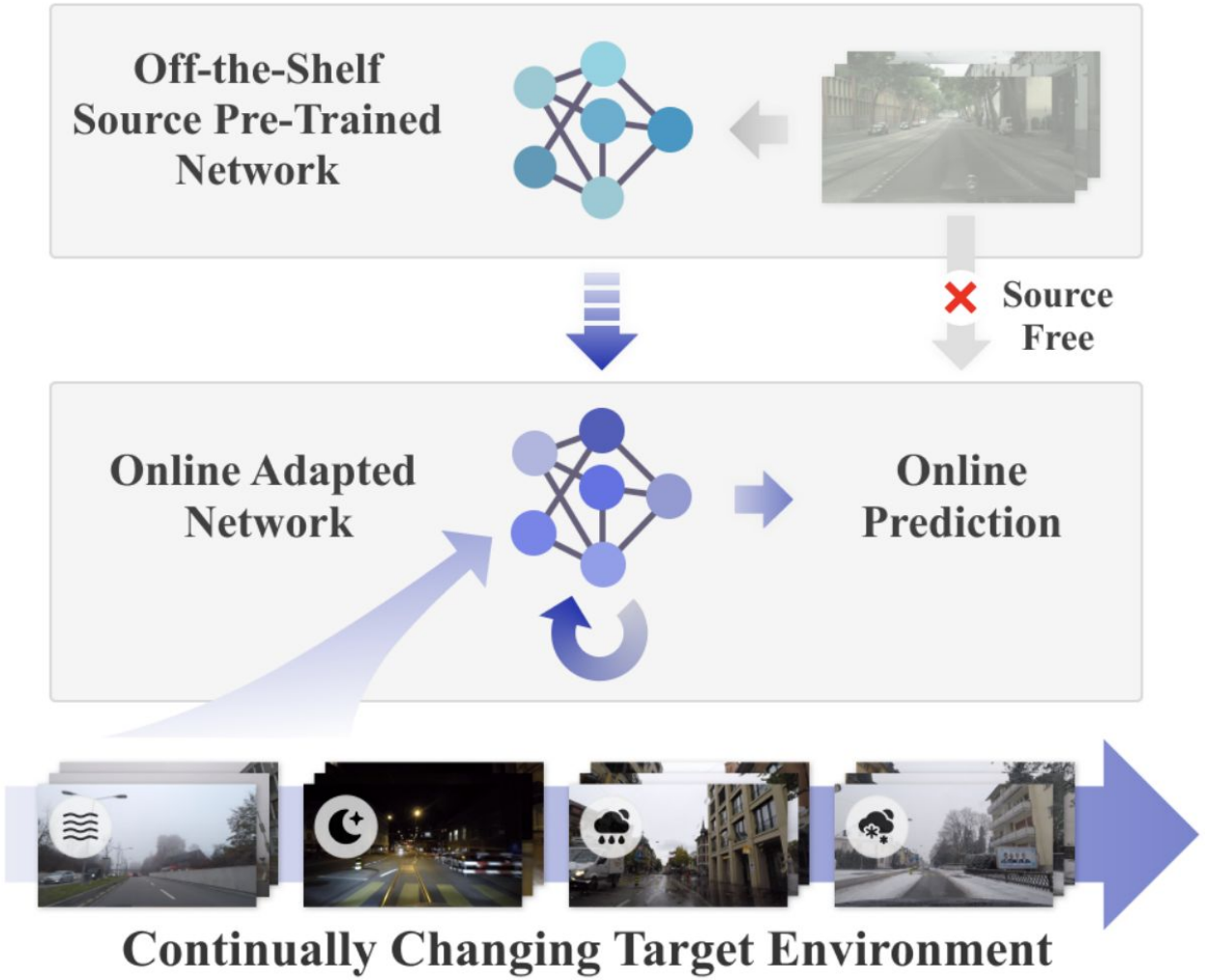
1) **Adaptability**
   - Continual learning = Continual adaptation
2) **Reduced catastrophic forgetting**
3) **Efficiency:**
   - More efficient training than standard fine-tuning when we want to add new class / new task (**e.g., medical applications**)?
4) **Human-like learning**

Outline

1. Continual Learning - motivation
2. **Test-time adaptation on synthetic distribution shift**
3. Proposed method for natural shifts

Continual Test-Time Adaptation

Continual Test-Time Adaptation



Wang, Qin, et al. "Continual test-time domain adaptation." *CVPR* 2022.

How to optimize models at test-time without access to the ground truth labels?

Entropy Minimization

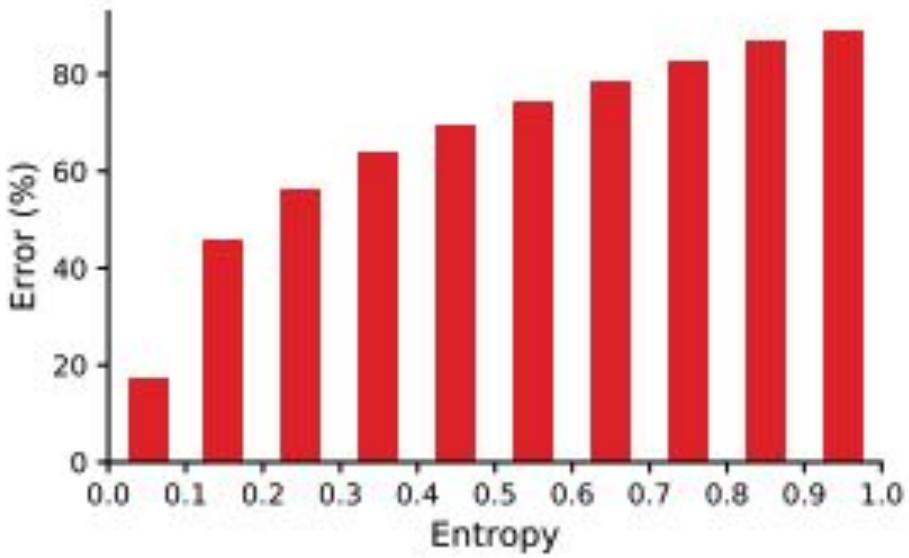


Figure 1: Predictions with lower entropy have lower error rates on corrupted CIFAR-100-C. Certainty can serve as supervision during testing.

Wang, Dequan, et al. "Tent: Fully test-time adaptation by entropy minimization." *ICLR 2021*..

Test-Time adaptation benchmarks

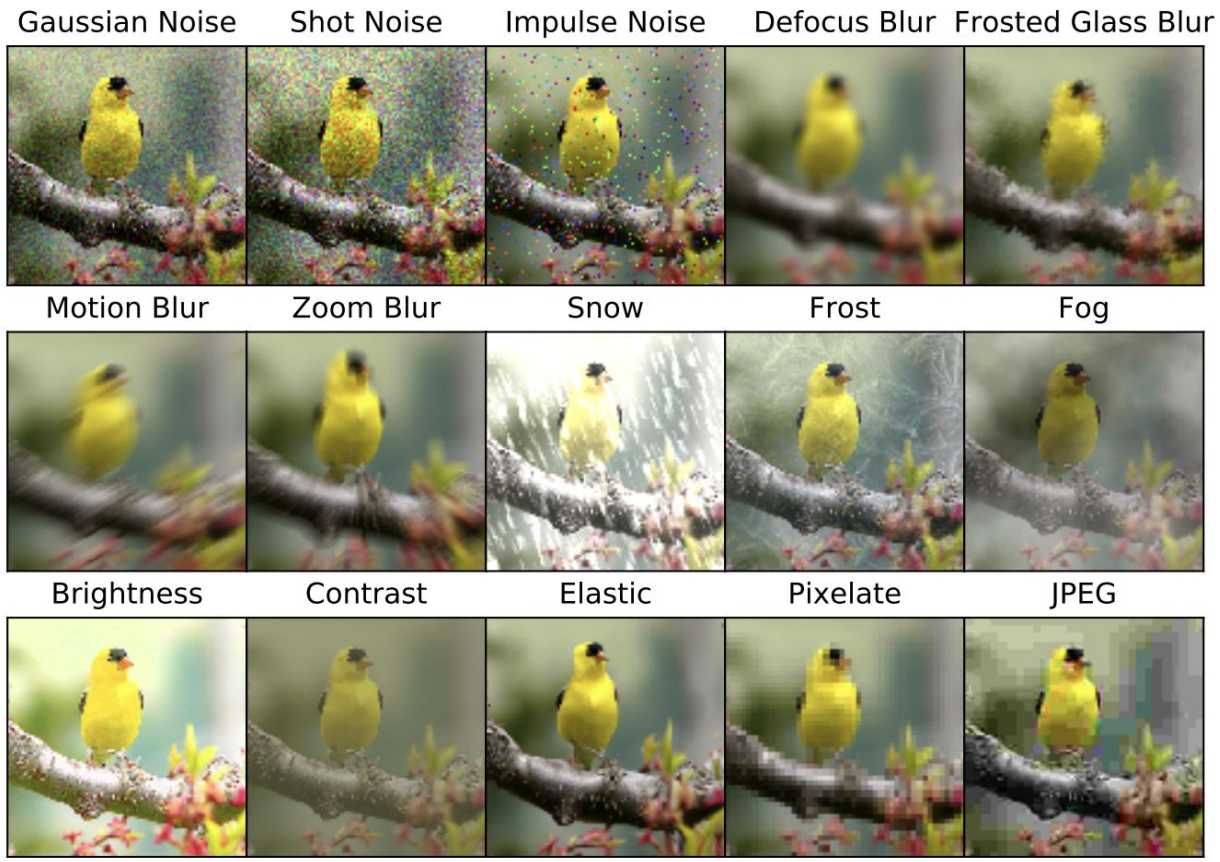


Figure 8: Examples of each corruption type in the image corruptions benchmark. While synthetic, this set of corruptions aims to represent natural factors of variation like noise, blur, weather, and digital imaging effects. This figure is reproduced from Hendrycks & Dietterich (2019).

Test-Time adaptation benchmarks

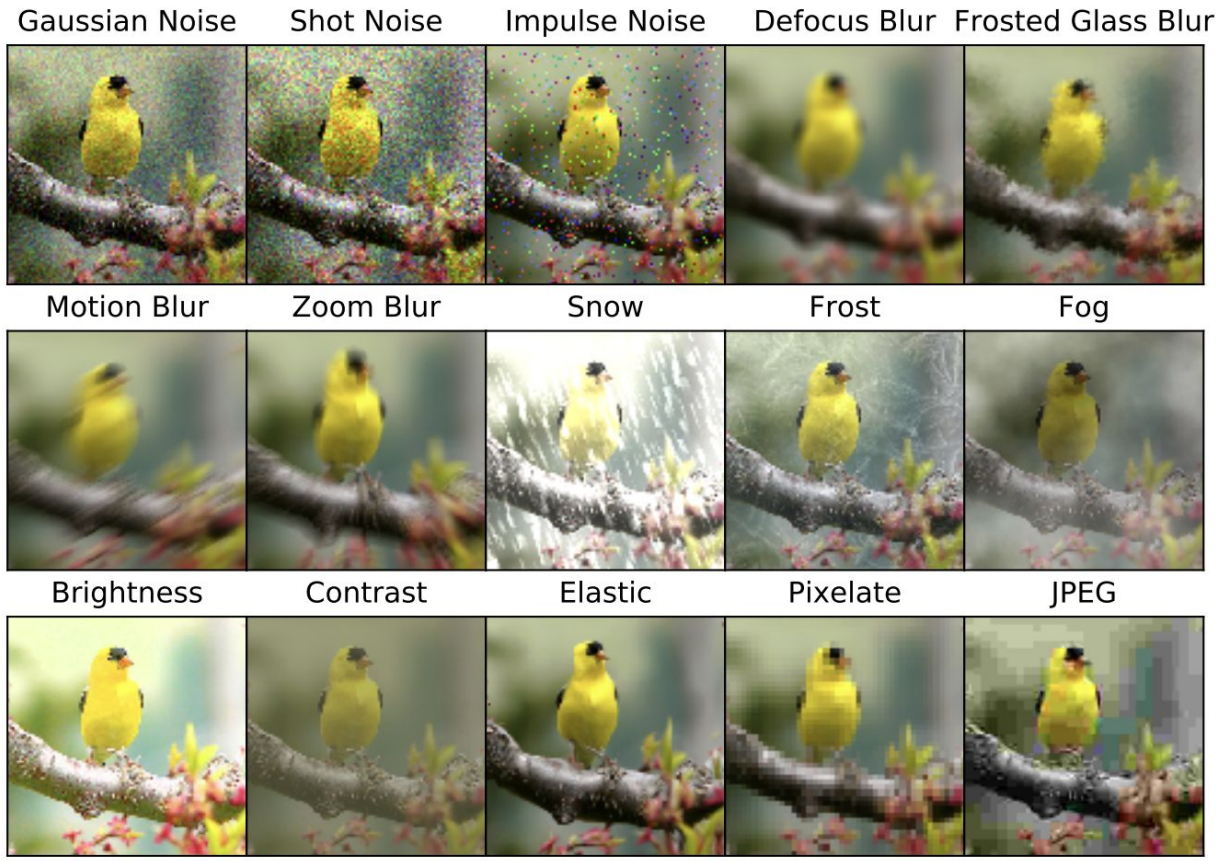


Figure 8: Examples of each corruption type in the image corruptions benchmark. While synthetic, this set of corruptions aims to represent natural factors of variation like noise, blur, weather, and digital imaging effects. This figure is reproduced from Hendrycks & Dietterich (2019).

Current SOTA:

- clean ImageNet accuracy ≈ 90%,

Test-Time adaptation benchmarks

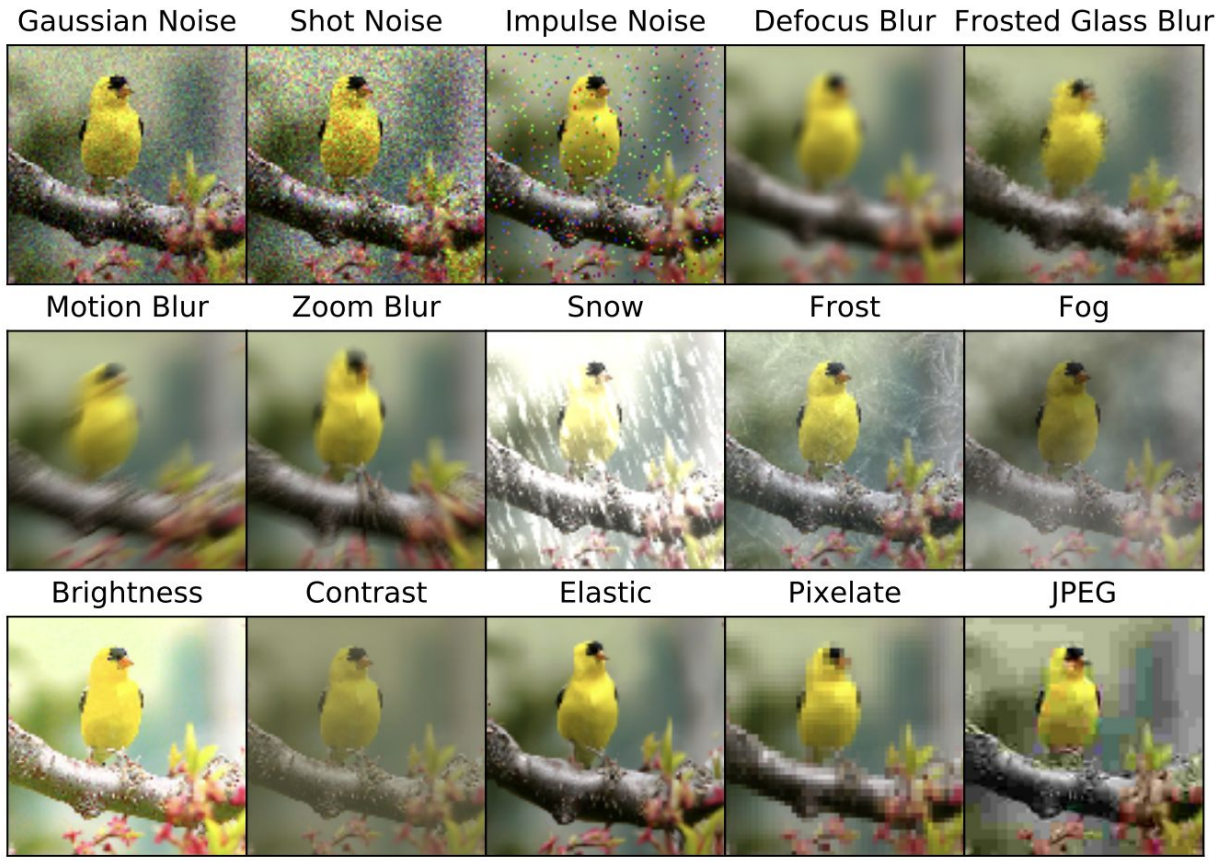


Figure 8: Examples of each corruption type in the image corruptions benchmark. While synthetic, this set of corruptions aims to represent natural factors of variation like noise, blur, weather, and digital imaging effects. This figure is reproduced from Hendrycks & Dietterich (2019).

Current SOTA:

- clean ImageNet accuracy ≈ 90%,
- accuracy on corrupted data = 56%,

Test-Time adaptation benchmarks

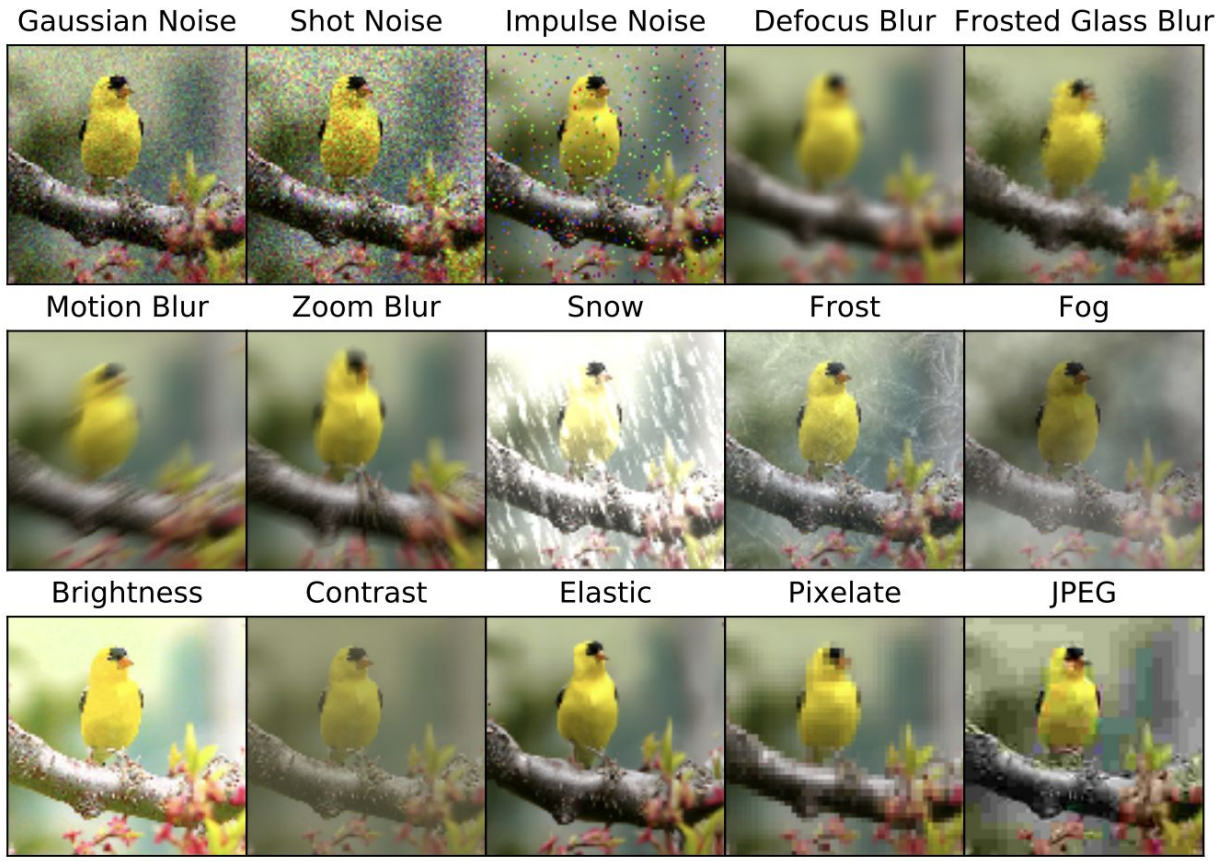


Figure 8: Examples of each corruption type in the image corruptions benchmark. While synthetic, this set of corruptions aims to represent natural factors of variation like noise, blur, weather, and digital imaging effects. This figure is reproduced from Hendrycks & Dietterich (2019).

Current SOTA:

- clean ImageNet accuracy ≈ 90%,
- accuracy on corrupted data = 56%,
- unsupervised adaptation -> 84% accuracy

Test-Time adaptation benchmarks

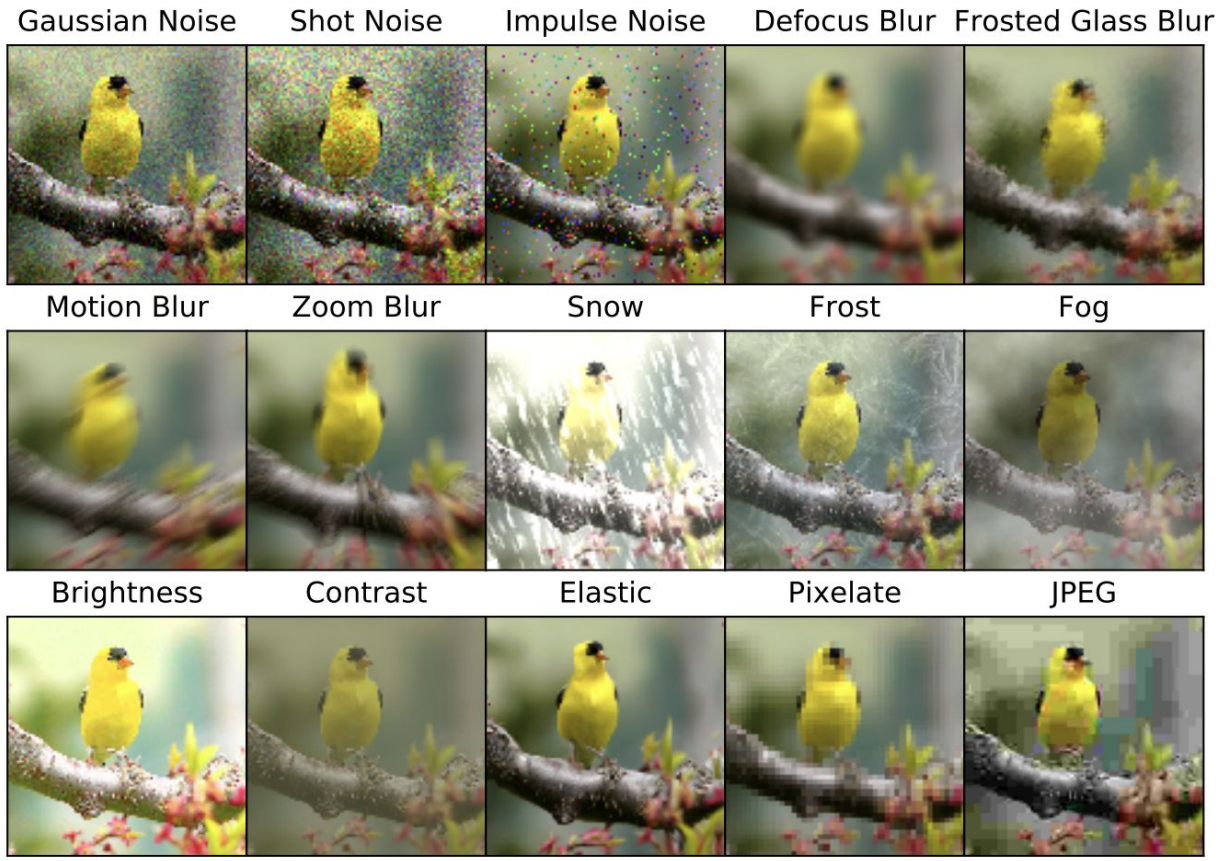


Figure 8: Examples of each corruption type in the image corruptions benchmark. While synthetic, this set of corruptions aims to represent natural factors of variation like noise, blur, weather, and digital imaging effects. This figure is reproduced from Hendrycks & Dietterich (2019).

Current SOTA:

- clean ImageNet accuracy ≈ 90%,
- accuracy on corrupted data = 56%,
- unsupervised adaptation -> 84% accuracy
- **Are we making a real progress?**

Outline

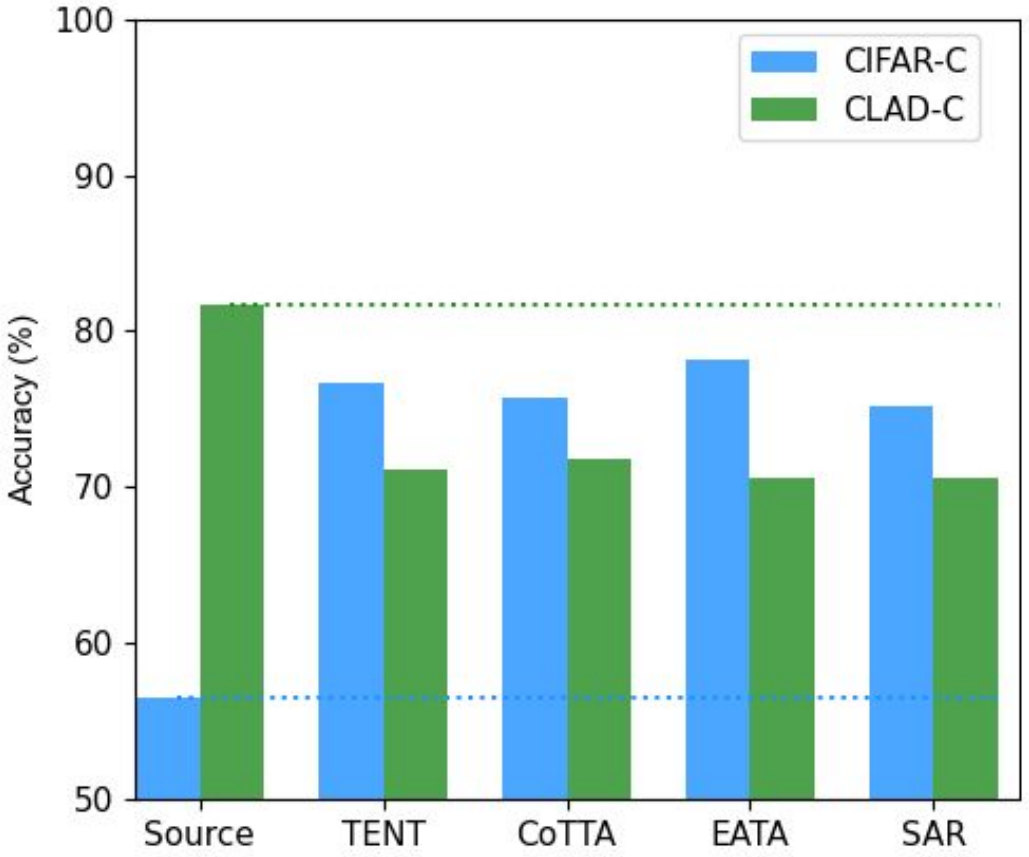
1. Continual Learning - motivation
2. Test-time adaptation on synthetic distribution shift
3. **Proposed method for natural shifts**

TTA for natural shifts

TTA for natural shifts

- varying level of distribution shift,
- varying quality,
- temporal correlation,

Test-Time adaptation on Real-World data



Sójka D., et al., AR-TTA: A Simple Method for Real-World Continual Test-Time Adaptation, ICCV 2023 CLVision workshop.

Our solution

Idea:

- **Detect** how much the data distribution has changed and **adjust** accordingly,

Our solution

Idea:

- **Detect** how much the data distribution has changed and **adjust** accordingly,

Compute shift on Batch Norm statistics

$$D(\phi^S, \phi_t^T) = \frac{1}{C}\sum_{i=1}^{C} KL(\phi_i^S || \phi_{t,i}^T) + KL(\phi_{t,i}^T || \phi_i^S)$$

Our solution

- We estimate BN statistics at time step *t* during test-time by linearly **interpolating** between saved statistics from source data and calculated values from current batch of test-time data,

$$\phi_t = (1 - \beta)\phi^S + \beta\phi_t^T$$

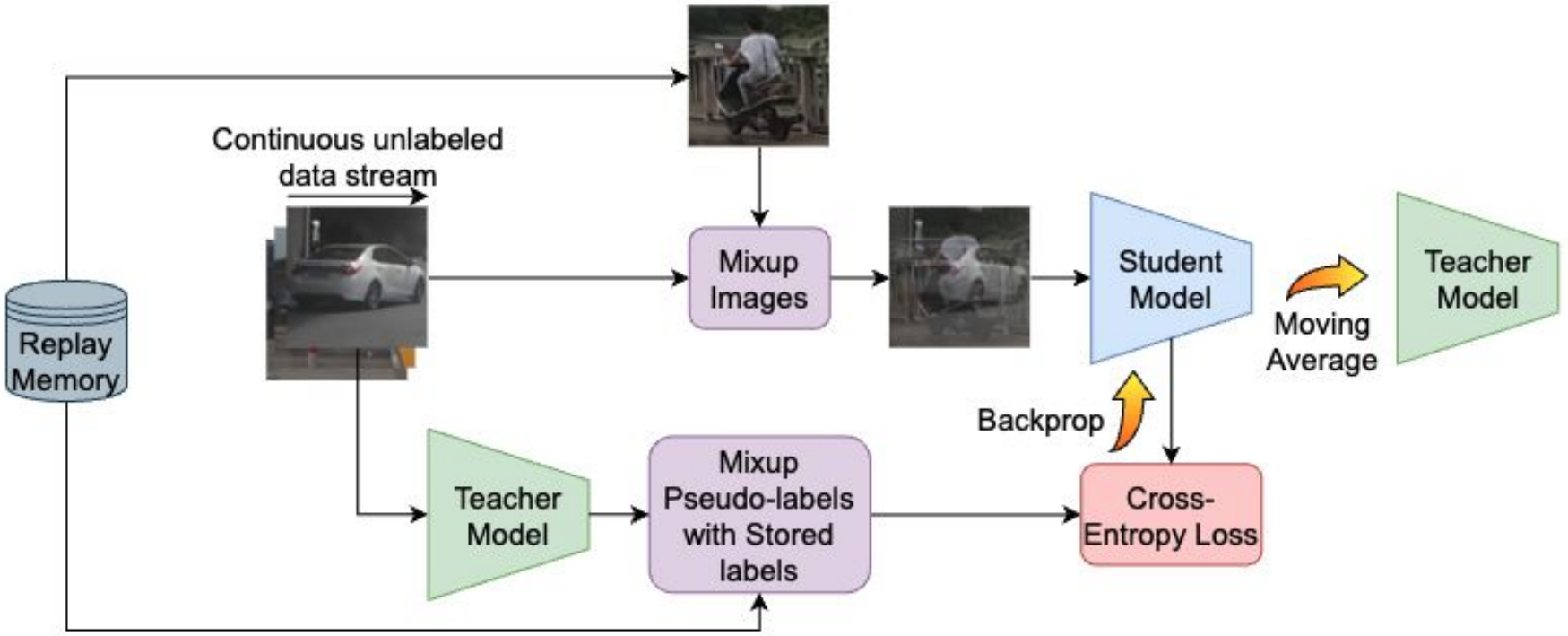
- β_t **reflects the distribution shift:**

$$\beta_t = 1 - e^{-\gamma D(\phi^S, \phi_t^T)}$$

- To provide more stability for the adaptation, we take into account previous β_{t-1} values and use an **exponential moving average** (parameter α) for β_t update.

$$\beta = (1 - \alpha)\beta_{t-1} + \alpha\beta_t$$

Architecture



Sójka D., et al., AR-TTA: A Simple Method for Real-World Continual Test-Time Adaptation, ICCV 2023 CLVision workshop.

Our Results

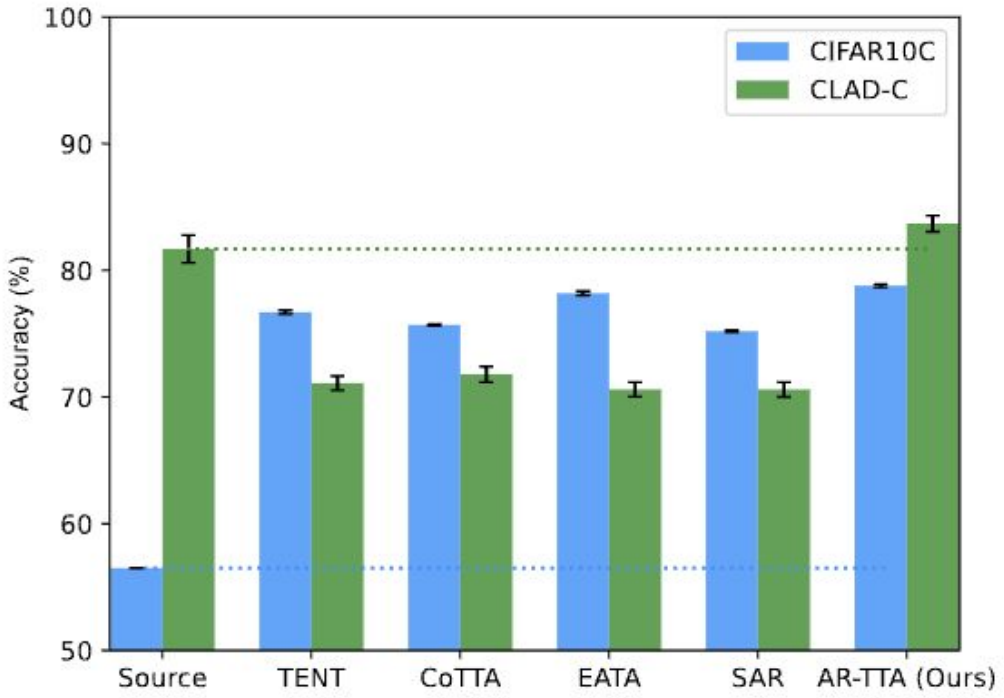
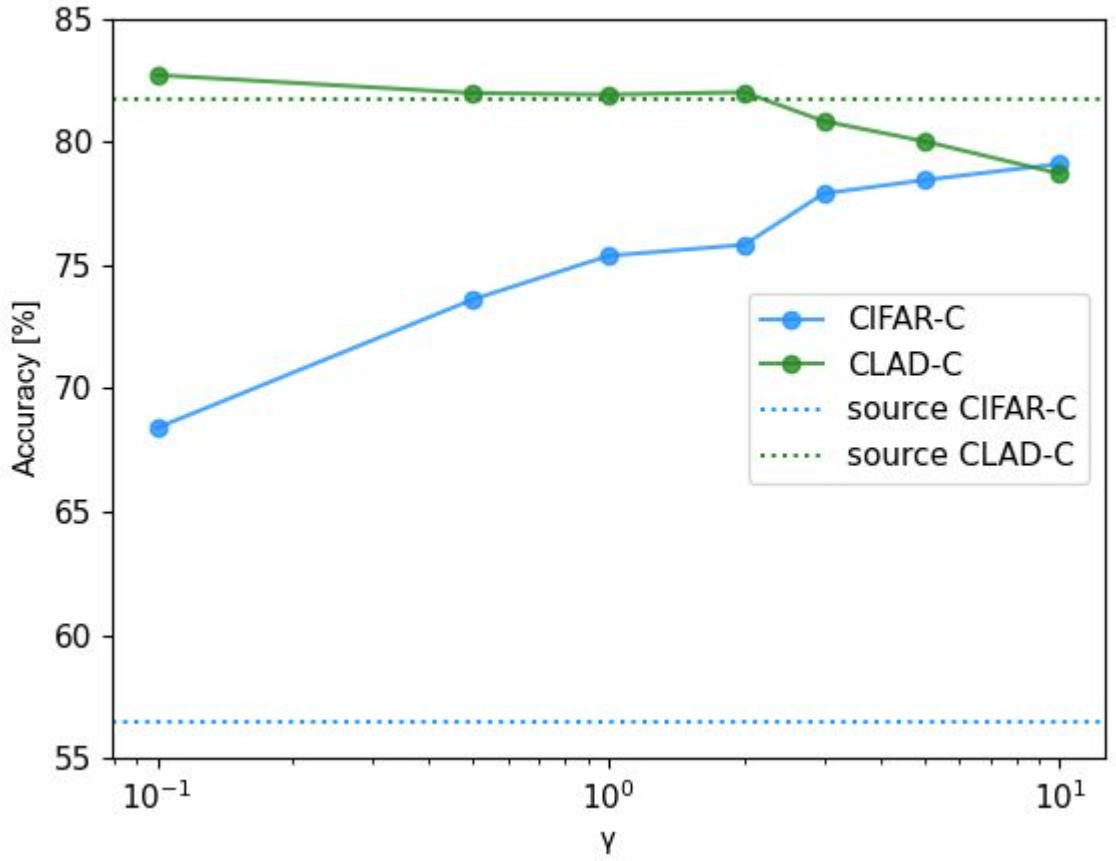


Figure 1. Continual test-time adaptation methods evaluated on synthetic (CIFAR-10C) and realistic (CLAD-C) domain shifts. Our method is the only one that consistently allows to improve over the naive strategy of using the (frozen) source model.

Sójka D., et al., AR-TTA: A Simple Method for Real-World Continual Test-Time Adaptation, ICCV 2023 CLVision workshop.

Optimizing for synthetic shift ≠ Optimizing for the Real-World



Sójka D., et al., AR-TTA: A Simple Method for Real-World Continual Test-Time Adaptation, ICCV 2023 CLVision workshop.

Hyper-parameter selection in TTA

- hp selection (learning rate, momentum, method specific parameters),
- all of the existing TTA methods assume access to the target labels

Hyper-parameter selection in TTA

- hp selection (learning rate, momentum, method specific parameters),
- all of the existing TTA methods assume access to the target labels
- good for methods comparison,
- but not very realistic

Conclusions!

Test-time adaptation improves adaptability of existing models to distribution shifts.

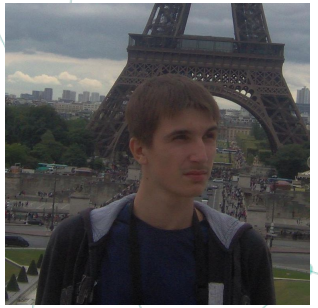
Yet, many challenges persist:

- How to select hp online?
- testing on very long sequences.

CL Team posters today @ MLinPL

AR-TTA: A Simple Method for Real-World Continual Test-Time Adaptation
[Innovation Award @ ICCV CLVision challange & ICCV 2023 CLVision Workshop]

Revisiting Supervision for Continual Representation Learning

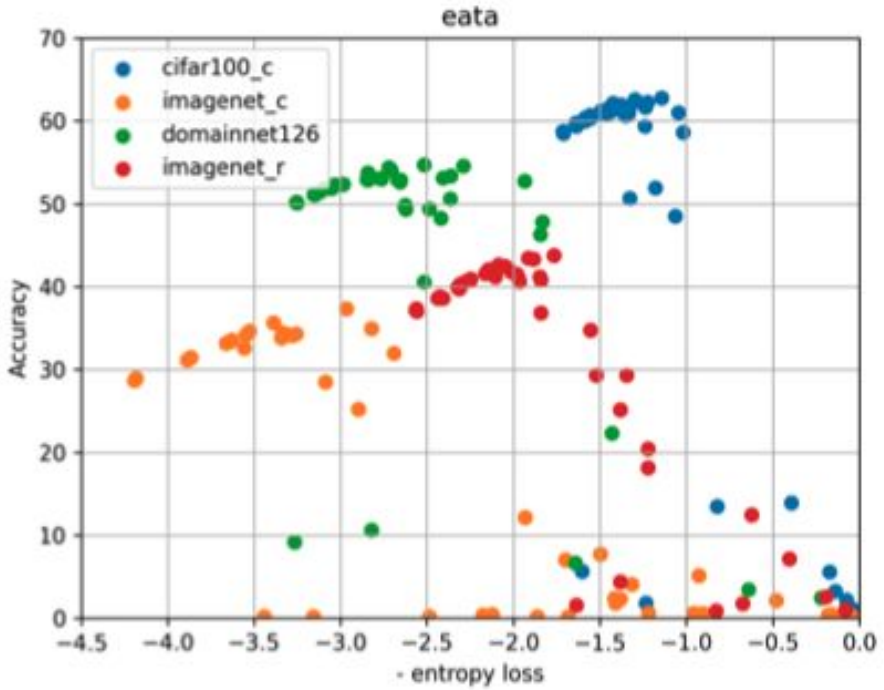


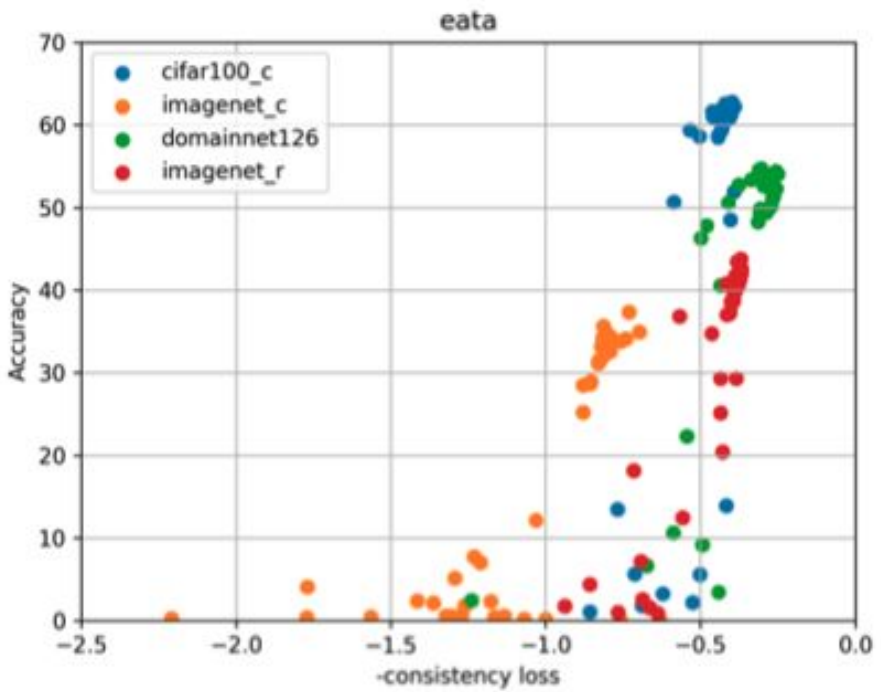
Adapt Your Teacher: Improving Knowledge Distillation for Exemplar-free Continual Learning
[WACV 2024]

Supporting Slides

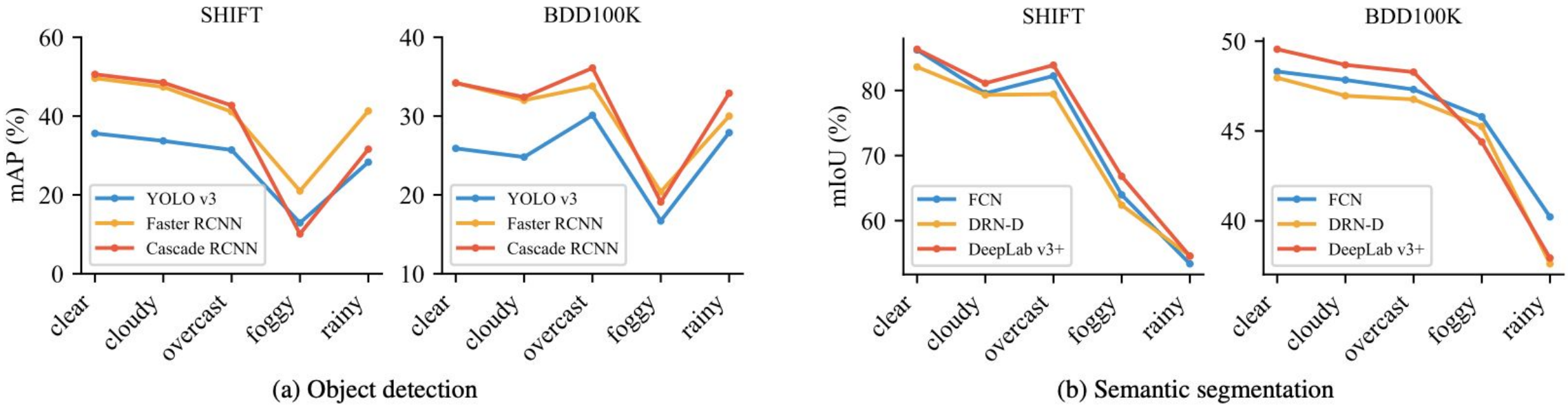
Hyper-parameter selection in TTA

- hp selection (learning rate, momentum, method specific parameters),
- all of the existing TTA methods assume access to the target labels
- this is not very realistic





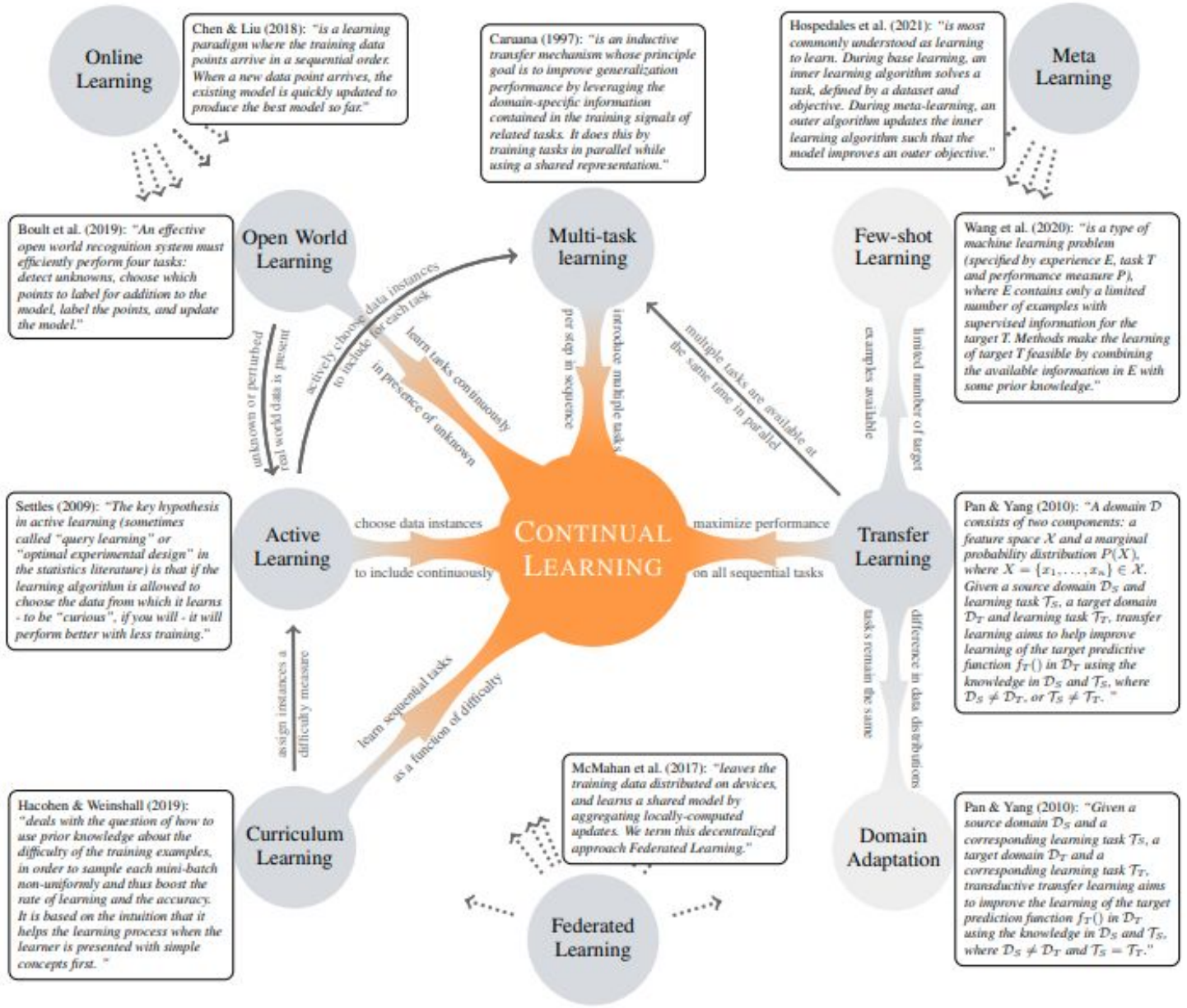
Shift Happens



Sun, Tao, et al. "SHIFT: a synthetic driving dataset for continuous multi-task domain adaptation." *CVPR* 2022.

Representation Drift

Yu, Lu, et al. "Semantic drift compensation for class-incremental learning." *CVPR,* 2020.



Mundt, Martin, et al. "CLEVA-Compass: A Continual Learning EValuation Assessment Compass to Promote Research Transparency and Comparability." arXiv preprint arXiv:2110.03331 (2021).

Recency Bias

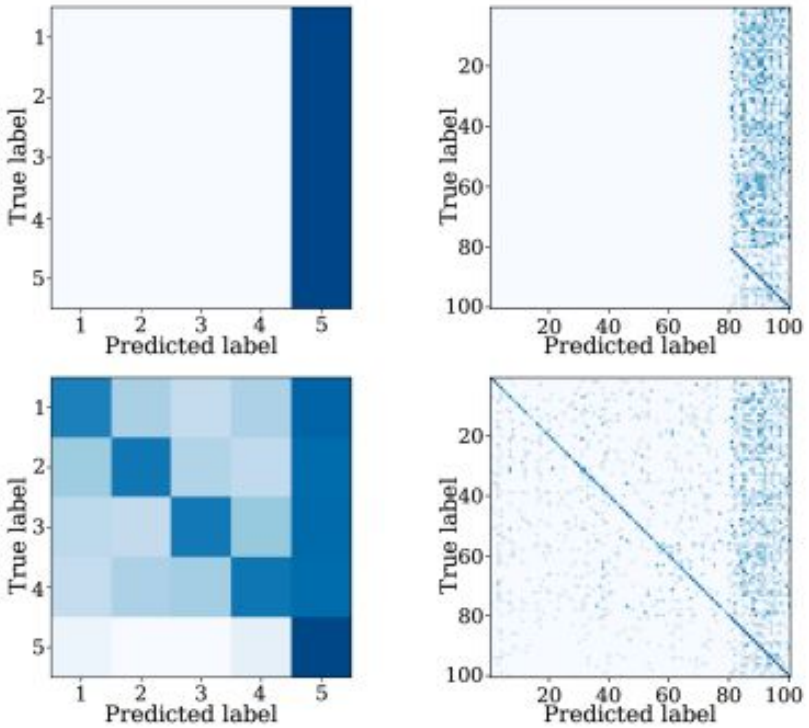
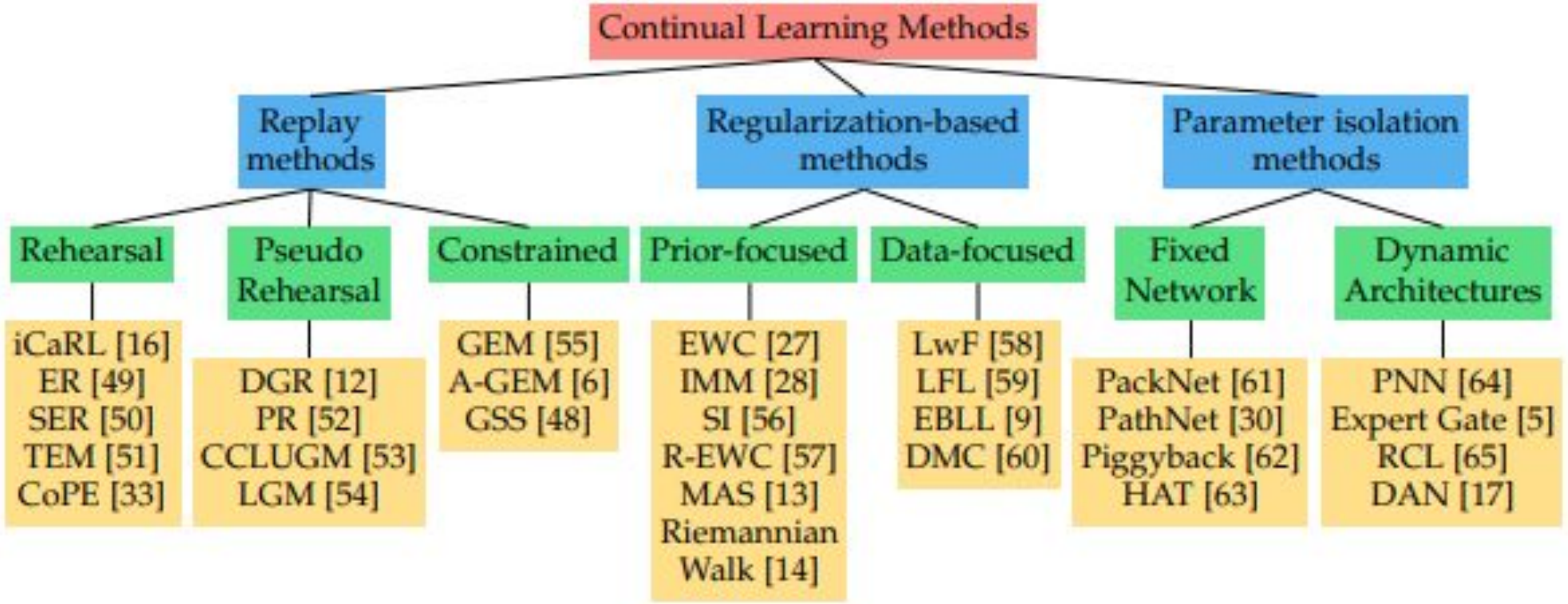


Fig. 3: Examples of task and class confusion matrices for Finetuning (top row) and Finetuning with 2,000 exemplars (bottom row) on CIFAR-100. Note the large bias towards the classes of the last task for Finetuning. By exploiting exemplars, the resulting classifier is clearly less biased.

M. Masana, et al. "Class-incremental learning: survey and performance evaluation on image classification.", TPAMI, 2022.

Approaches Overview



M. De Lange et al., "A continual learning survey: Defying forgetting in classification tasks." IEEE transactions on pattern analysis and machine intelligence (TPAMI), 2021.