Real-Time Continual Learning from Natural Video Streams

Human-centered Vision: from Body Analysis to Learning and Language Workshop

09-07-2020

Vincenzo Lomonaco, Lorenzo Pellegrini
University of Bologna
{vincenzo.lomonaco, l.pellegrini}@unibo.it
Continual Learning @ BioLab

Computer Science Department - University of Bologna

Cesena - Italy

Lab members involved in this research

- Davide Maltoni  
  Full prof.
- Vincenzo Lomonaco  
  Post doc
- Lorenzo Pellegrini  
  PhD Student
- Gabriele Graffieti  
  PhD Student
ContinualAI: a Non-profit Research Organization and Open Community on Continual Learning for AI

http://continualai.org
ContinualAI: a Non-profit Research Organization and Open Community on Continual Learning for AI

- **ContinualAI** is a non-profit research organization and the largest research community on Continual Learning for AI.

- It counts more than **900+ members** in **17 different time-zones** and from top-notch research institutions.

- Learn more about **ContinualAI** at [www.continualai.org](http://www.continualai.org)
Let me start from the end...
CORe Android App

- Demonstrates CL at the edge.
- Android smartphone with no hardware acceleration.
- Near real-time training of a MobileNet (less than 1 sec. to update the model after each 20 sec. video).

This Video: https://www.youtube.com/watch?v=Bs3tSjwbHa4
Download the App: https://github.com/lrzpellegrini/CL-CORe-App
CORe App: Details

• The App comes with a customized version of the Caffe framework along with a C++ implementation of the AR1-Free algorithm.

• MobileNetV1 (non quantized) pre-trained on ImageNet, tuned on the 10 categories of the CORe50 dataset.

• Only 100 frames are taken (≈ 5 FPS).

• Only 500 latent patterns are kept for replay (≈ 2 MB)! 
Towards CL On-The-Edge

- The **AR1-Free** algorithm and the **Latent Replay** mechanism have been designed to be extremely **light on resources**
  - Reduced memory footprint
  - Minimal latency (training time)
  - Overall low energy consumption

- The **accuracy/performance tradeoff** can be chosen based on the target platform and the problem to be solved

- Suitable for **ultra low power devices***
  - Towards Continual Learning On-The-Edge!

Our Objectives

Online Continual Learning

• Limited computation and storage, real-time updates, small non i.i.d. batches
• Training (once deployed) at the edge without network connection
• Privacy friendly (no server-side processing, no storage of row data)

Focus on real computer vision applications

• Robotics*
• Smart cameras (e.g. surveillance)
• Vision apps on mobile devices

MultiTask vs Single Incremental Task

**Multitask:** the same model learns (sequentially) multiple disjoint tasks. At inference time a task supervised signal is provided.

*What are the real applications?*

**Single Incremental task:** the same model learns (sequentially) new classes or improve its recognition accuracy over known classes. At inference time must recognize all the classes learnt so far.

*Fortunately, after initial focus on Multitask scenario, now most researchers are working in this setting.*

Single Incremental Task

1. **New Instances (NI)**
   - OpenLORIS, Core50

2. **New Classes (NC)**
   - MNIST, Cifar10, Cifar100, ImageNet100, ImageNet, Core50

3. **New Instances and Classes (NIC)**
   - Core50
CORe50

Dataset, Benchmark, code and additional information freely available at: http://vlomonaco.github.io/core50

CORe50: a Video Benchmark for CL and Object Recognition/Detection

Core50 and CLVISION challenge

1. Core50 was selected for CLVISION challenge (CVPR 2020)
   - Tree tracks: NI, NC (Multitask), NIC
   - Teams: 79 registered, 11 finalists
   - Sponsors: Intel Labs, ContinualAI, Element AI, NVIDIA

https://sites.google.com/view/clvision2020/challenge
AR-1: Architect & Regularize

From 2016 to 2020
AR-1: From 2016 to 2020

2016
- Copy-weights with Re-init (CWR), a simple but effective weights normalizing algorithm is proposed.

2017
- Two extensions of CWR called CWR+ and AR1 are proposed combining architectural with regularization approaches.

2018
- AR1*, an extension of AR1, is proposed to tackle NI, NC and NIC scenarios agnostically.

2019
- AR1 with Latent Replay mechanisms is firstly proposed and tested on the challenging NICv2-391.

2020
- AR1 with Generative Replay (work in progress...).
AR-1 (with Latent Replay)

...can we improve AR-1 accuracy by storing some data (for rehearsal) without affecting efficiency?

Practical issues with rehearsal

1. Rehearsal (or replay) requires to store some representatives of old batches
   - ICARL is one of the best-known techniques.

2. Requires extra storage
   - For example for ImageNet, if we store 20 patterns per class, the total storage is about 3.8 GB.

3. …and extra forward/backward steps!
   - When mixing new and old patterns more iterations for epoch.

Idea: storing activations at some intermediate level and not raw images.
AR-1 with Latent Replay

Data layer

- Low-level generic features
  - (unsup training or slow tuning)

Output layer (classes)

- Class specific discriminative features
- External storage (replay patterns)

Forward pass (all patterns)
Backward pass (all patterns)

Forward pass (native patterns)
Backward pass (native patterns)

Concat

External storage (replay patterns) at minibatch level

this is the solution used in the CORe demo App
Latent replay advantages

1. **Efficiency**: extra forward and backward steps take place only in the upper layers.
2. **Require less storage**.
3. **Activations can be quantized/compressed with negligible accuracy loss**.
Closing the Accuracy Gap with Latent Replay
Work in progress

**AR-1 with Latent Generative Replay**

- **Output layer (classes)**
  - Class specific discriminative features
  - Low-level generic features
    - (unsup training or slow tuning)

- **Generative Model**
  - Forward pass (all patterns)
  - Backward pass (all patterns)
  - (at minibatch level)

- **Data layer**
  - Forward pass (native patterns)
  - Backward pass (native patterns)

- **Concat**
Practical issues with generative replay

1. Training (continuously) a model to generate useful images does not scale to complex datasets
   - Ok for MNIST, CIFAR10, but not for CIFAR100, ImageNet, Core50

2. Generating features (instead of row data) was recently proved* to be a good alternative.
   - However, generation was performed at the semi-last (flat) level

3. We would like to train the generator online.
   - GAN training could be too slow

---


Challenge: generating complex activation volumes in near-real time.
Future work

Move toward unsupervised training

• Self-training by exploiting temporal coherence*.
• Openset classification (automatic discovery of new classes).
• Sparse human supervision (active learning).

Questions?

Human-centered Vision: from Body Analysis to Learning and Language Workshop

09-07-2020

Vincenzo Lomonaco, Lorenzo Pellegrini

University of Bologna

{vincenzo.lomonaco, l.pellegrini}@unibo.it