Incremental Classifier and Representation Learning

Sylvestre-Alvise Rebuffi[†,*], Alexander Kolesnikov*, Georg Sperl*, Christoph H. Lampert*
* IST Austria † University of Oxford

IST AUSTRIA
Institute of Science and Technology

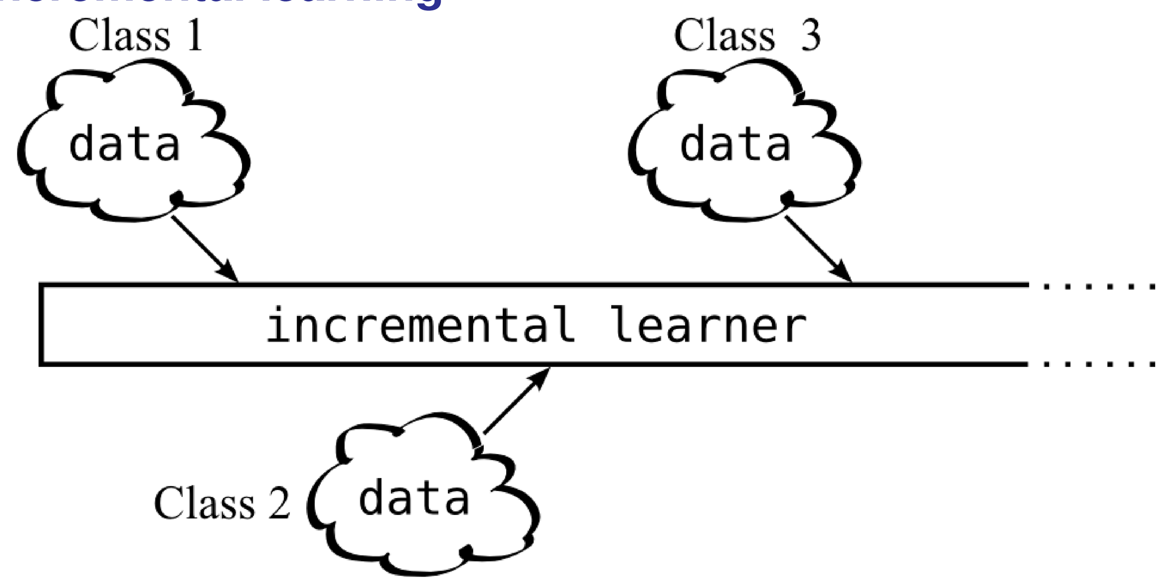
CVPR July 21-26 2017 HONOLULU

Abstract

iCaRL can learn classifiers/representation incrementally over a long period of time where other methods quickly fail. Catastrophic forgetting is avoided thanks to a combination of a nearest-mean-of-exemplars classifier, herding for adaptive exemplar selection and distillation for representation learning.

1) Motivation

Situation: class-incremental learning



Problem:

- classes appear sequentially
- train on new classes without forgetting the old ones
- no retraining from scratch: too heavy
- can't keep all the data around

2) Class-Incremental Learning

We want to/we can:

- for any number of observed classes, t, learn a multi-class classifier
- store a certain number, K, of images (a few hundreds or thousands)

We do not want to/we cannot:

- add more and more resources: fixed sized network
- store all training examples (could be millions)

The dilemma:

- fixing the data representation: suboptimal results on new classes.
- continuously improving the representation: classifiers for earlier classes deteriorate over time ("catastrophic forgetting/interference") [McCloskey, Cohen. 1989]

3) Existing Approaches

Fixed data representation:

- represent classes by mean feature vectors [Mensink *et al.*, 2012], [Ristin *et al.*, 2014]

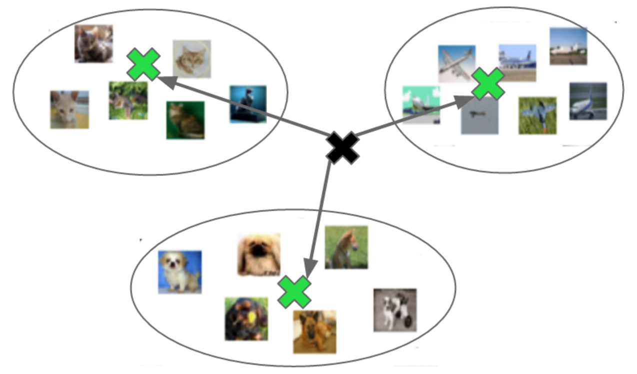
Learning the data representation:

- grow neural network incrementally, fixing parts that are responsible for earlier class decisions [Mandziuk, Shastri. 1998], …, [Rusu *et al.*, 2016]
- multi-task setting: preserve network activations by distillation [Li, Hoiem. 2016]

4) iCaRL

iCaRL component 1: exemplar-based classification.

- nearest-mean-of-exemplars classifier
- automatic adjustment to representation change
- more robust than network outputs

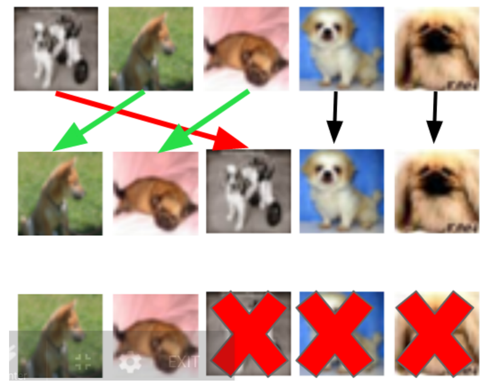


iCaRL component 2: representation learning.

- add a distillation term to loss function
- stabilizes outputs
- limited overhead, just need one copy of the network

iCaRL component 3: exemplar selection.

- greedy selection procedure by herding
- number of exemplars per class decrease as the number of class increases
- ability to remove exemplars on-the-fly through ranking of exemplars



5) Experiments

CIFAR-100:

- 100 classes, in batches of 10
- 32-layer ResNet [He *et al.*, 2015]
- evaluated by top-1 accuracy
- number of exemplars: 2000

ImageNet ILSVRC 2012:

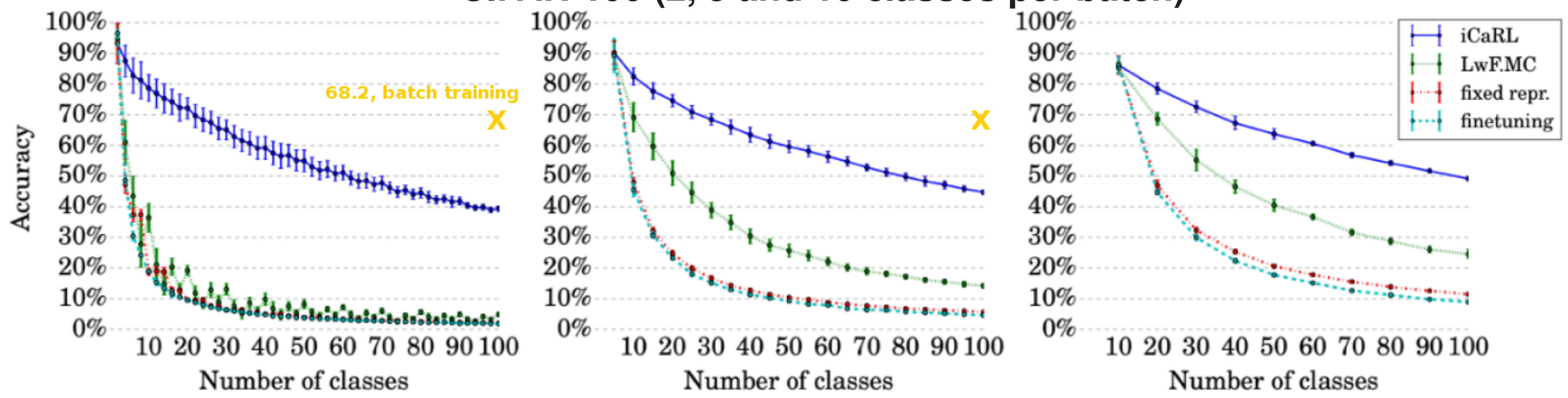
- 1000 classes, in batches of 10
- 18-layer ResNet [He *et al.*, 2015]
- evaluated by top-5 accuracy
- number of exemplars: 20000

Baselines:

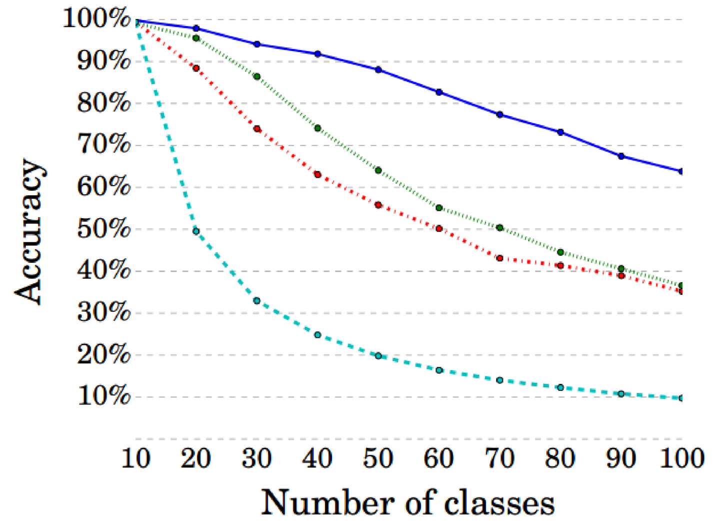
- fixed representation: freeze representation after first batch of classes
- finetuning: ordinary NN learning, no freezing
- LwF: *"Learning without Forgetting"* [Li, Hoiem. 2016], use network itself to classify

6) Results

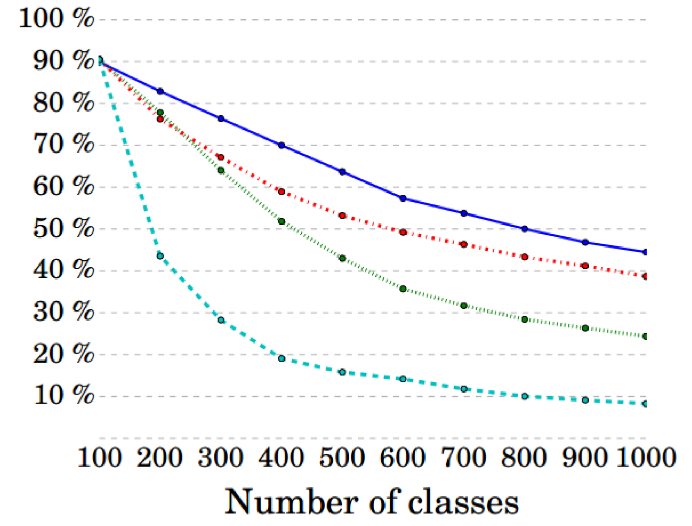
CIFAR-100 (2, 5 and 10 classes per batch)



ILSVRC 2012 (small)



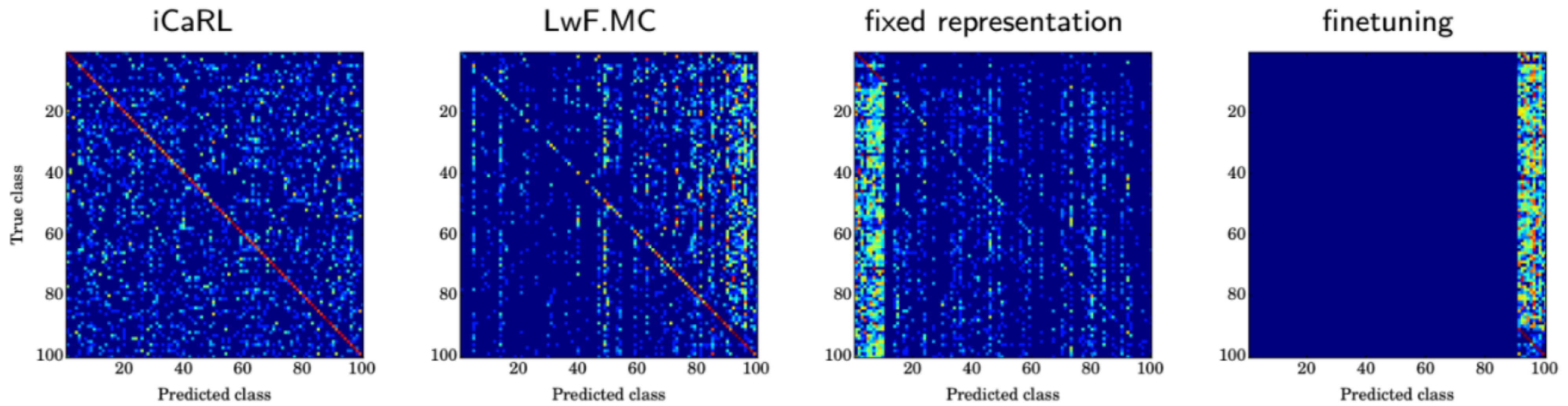
ILSVRC 2012 (full)



Discussion:

- as expected: fixed representation and finetuning do not work well
- iCaRL is able to keep good classification accuracy for many iterations
- *"Learning without Forgetting"* starts to forget earlier

Confusion matrices for CIFAR-100



Discussion:

- iCaRL: predictions spread homogeneously
- LwF.MC: prefer recently seen classes → long-term memory loss
- fixed representation: prefer first batch of classes → lack of neural plasticity
- finetuning: predict only classes among the last batch → catastrophic forgetting