Semi-Supervised Class Incremental Learning

Alexis Lechat\textsuperscript{1,2}, Stéphane Herbin\textsuperscript{1} and Frédéric Jurie\textsuperscript{2}
\textsuperscript{1}ONERA, \textsuperscript{2}Normandie Université

ICPR 2020

Artificial Intelligence, Machine Learning for Pattern Analysis PS T1.16
Class Incremental (CI) Learning

**encoder**

<table>
<thead>
<tr>
<th>Number of classes learned</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>CI</td>
</tr>
<tr>
<td>4</td>
<td>Batch</td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

**Comparison**

- **plane** vs. **bird**
- **car** vs. **cat**
Class Incremental (CI) Learning

- **plane - bird**
- **car - cat**
- **deer - dog**

ENCODER

Number of classes learned:

Accuracy:

- CI
- Batch
Class Incremental (CI) Learning

- car - cat
- deer - dog
- horse - monkey

ENCODER

Accuracy

Number of classes learned

- CI
- Batch
Class Incremental (CI) Learning

- deer - dog
- horse - monkey
- ship - truck

Graph showing accuracy over the number of classes learned for CI and batch methods.
Class Incremental (CI) Learning

horse - monkey

ship - truck

ENCODER

Accuracy

Number of classes learned
Prior works

Knowledge distillation
[Li 2017; Rebuffi 2017; Zhao 2020]

Rehearsal learning
[Rebuffi 2017; Castro 2018; Wu 2019; Zhao 2020]

Weight regularization
[Kirkpatrick 2017; Zenke 2017; Aljundi 2018]
Our approach: Semi-Supervised Incremental Learning

- Class Incremental Data Stream
  - Dog vs Cat
  - Plane vs Car
  - Bird vs Horse
  - Standard CI process with rehearsal

- Unlabeled Data Stream
  - Self-supervised training (pretext task) with cheap unlabeled data

- Model
  - Supervised learning
  - Self-supervised regularization
Objectives

• Profit from inexpensive unlabeled data to build a large self-supervised task

• Use the self-supervision as a regularization to alleviate the Catastrophic Forgetting

• Learn better representations for a more stable encoder/enhanced performances

• Further reduce the amount of labeled data needed
Our SSIL Framework

Labeled and unlabeled images

ENCODER

Discriminator y

Discriminator z

DECODER
Step 1: reconstruction

Labeled and unlabeled images

ENCODER → Y → Z → DECODER

Discriminator y

Discriminator z

Reconstruction loss
Step 2: adversarial training

Labeled and unlabeled images

ENCODER

Discriminator $y$

Adversarial loss $p(y) = \text{Cat}(y)$

$y \in \mathbb{R}^C$ with $C$ the number of clusters for self-supervised clustering

Discriminator $z$

Adversarial loss $p(z) = \mathcal{N}(0, 1)$
Step 3: supervised classification

Labeled images → ENCODER → y → Classification loss

Cluster management in y:

Unassigned clusters

Clusters assigned to encountered classes:
cat, plane, dog, car
Results: class incremental

<table>
<thead>
<tr>
<th>Method</th>
<th>MNIST</th>
<th>STL-10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Latest (%)</td>
<td>Average (%)</td>
</tr>
<tr>
<td>Oracle</td>
<td>99.4</td>
<td>99.7</td>
</tr>
<tr>
<td>Fine-Tuning</td>
<td>19.8</td>
<td>44.9</td>
</tr>
<tr>
<td>LwF</td>
<td>71.3</td>
<td>85.2</td>
</tr>
<tr>
<td>DMC</td>
<td>81.1</td>
<td>87.4</td>
</tr>
<tr>
<td>Naive Rehearsal</td>
<td>93.7</td>
<td>97.6</td>
</tr>
<tr>
<td>iCaRL</td>
<td>95.3</td>
<td>97.9</td>
</tr>
<tr>
<td>WA</td>
<td>96.0</td>
<td>98.3</td>
</tr>
<tr>
<td>Ours$^a$ (EMNIST-digits)</td>
<td>98.1</td>
<td>99.0</td>
</tr>
<tr>
<td>Ours$^a$ (EMNIST-letters)</td>
<td>95.9</td>
<td>98.5</td>
</tr>
</tbody>
</table>

$^a$ Our standard baseline on MNIST uses EMNIST-full as unlabeled data stream.
$^b$ Additional results on MNIST benchmark when using EMNIST-digits and EMNIST-letters as unlabeled data stream instead of the whole EMNIST.

**Memory size:** K=400 for MNIST and K=500 for STL-10

**Unlabeled dataset leveraged by our SSIL**

**MNIST:** EMNIST (814,255 characters, digits and letters)

**STL-10:** 100,000 unlabeled images are provided in the dataset
Results: enhanced representations

Comparison of different rehearsal strategies initialized with a self-supervised encoder (pre-trained with RotNet)
Conclusion

- SSIL achieves better performance
- SSIL requires less labeled data
- Self-supervision is an efficient regularization for incremental learning
Thank you for watching

Poster Session T1.16
ID #2885

Semi-Supervised Class Incremental Learning
Alexis Lechat, Stéphane Herbin and Frédéric Jurie