Going Beyond Continual Learning
Towards Organic Lifelong Learning

29 Jun 2023

Vineeth N Balasubramanian
Visiting Fulbright-Nehru Faculty Fellow, Carnegie Mellon University
Department of Computer Science and Engineering/Artificial Intelligence
Indian Institute of Technology, Hyderabad
Learning continually on newer classes/tasks/domains has gained importance in recent years. But do existing settings capture the real-world? How can we make lifelong learning “organic”?
**Our Group’s Research**

**Explainable, Robust DL**
- Saliency Maps (Grad-CAM++) and Attributions, AISTATS 2022, IEEE TBIOM 2021, WACV 2018
- Causality in NNs, ICML 2022, AAAI 2022, WACV 2022, ICMI 2019, NeurIPS 2019
- Antehoc Interpretability, CVPR 2022
- Attributional and Adversarial Robustness, NeurIPS 2021, ECCV 2020, AAAI 2021

**Learning with Limited Labeled Data**
- Continual Learning, CVPR 2022, WACV 2022, NeurIPS 2020, TPAMI 2021
- Few-shot/Zero-shot Learning, WACV 2021, NeurIPS 2020
- Generative Models, WACV 2022, CVPR 2018, ICCV 2017

**Thesis:** Towards learning robust reliable systems in evolving environments

Deep Learning, Machine Learning, Computer Vision
Our Group’s Research

**Explainable and Robust Learning**

- **Saliency Maps (Grad-CAM++) and Attributions**, AISTATS 2022, IEEE TBIOM 2021, WACV 2018
- **Causality in NNs**, ICML 2022, AAAI 2022, WACV 2022, ICML 2019, CVPRW 2021
- **Antehoc Interpretability**, CVPR 2022
- **Attributional and Adversarial Robustness**, NeurIPS 2021, ECCV 2020, AAAI 2021

**Learning in Data/Label-Deficient Environments**

- **Continual Learning**, CVPR 2022, WACV 2022, NeurIPS 2020, TPAMI 2021
- **Open-world Learning**, CVPR 2021

Deep Learning, Machine Learning, Computer Vision

On the Layerwise Hessian of Deep Neural Network Models, **AAAI 2021**; Submodular Batch Selection for Training Deep Neural Networks, **IJCAI 2019**; On Noise and Optimality in Neural Networks, **ICML 2018 Workshops**
In the 19th century, there was something called the **cult of domesticity** for many American **women**. This meant that most married **women** were expected to **stay in the home** and **raise children**. As in other countries, **American wives** were very much **under the control** of their **husband**, and had almost no rights. **Women** who were not married had only a few jobs open to them, such as working in **clothing factories** and serving as **maids**. By the 19th century, **women** such as **Lucretia Mott** and **Elizabeth Cady Stanton** thought that women should have more rights. In 1848, many of these **women** met and agreed to fight for more rights for **women**, including **voting**. Many of the **women** involved in the movement for **women's rights** were also involved in the movement to end **slavery**.

Tag colors:
- **LOCATION**
- **PERSON**
- **TERM**
- **DATE**
- **CONDITION**
- **PROCESS**
- **PEOPLE**
Beyond Static Supervised Learning

Why?

Self-driving car/intelligent transport system needs to adapt to new classes in new countries/environments

Robots need new skills in different environments, adapting to new situations, learning new tasks

Medical applications need to adapt to new patients, hospitals, conditions

Conversational agents need to adapt to new users, situations, tasks

Slide adapted from: Irina Rish, MILA

Thrun and Mitchell, Lifelong robot learning, RAS 1995
“Continual learning is the constant development of increasingly complex behaviors; the process of building more complicated skills on top of those already developed.”

Catastrophic Forgetting

Primary Challenge of Continual Learning

- Adapting to perform new tasks causes network to “forget” those it previously learned
- First observed by McCloskey and Cohen in 1989 in “Catastrophic forgetting in connectionist networks”

Task A=>B: locomotive - dishtowel window - reason bicycle - tree
Task A => C: locomotive - cloud window - book bicycle - couch

Catastrophic: Loss of old knowledge is sudden rather than gradual
Sir Bartlett’s Repeated Reproduction Experiment

- An attempt to study how well we remember; Bergman and Roediger replicated the results in 1999
- Recall of concepts from story is plotted against time (results directly from paper)

Train a two-layer neural network on two tasks one after the other

Task 1
Classes: \{0, 1, 2, 3, 4\}

Task 2
Classes: \{5, 6, 7, 8, 9\}

![Bar chart](chart.png)

- After Learning Task 1
- After Learning Task 2
Stability-Plasticity Tradeoff

Key challenge in CL

• Preserving previous information requires **stability**
  – Weights not changing

• Learning requires **plasticity**
  – Weights changing

• Tradeoff:
  – How to achieve stability without rigidity and plasticity without chaos?
  – How can network remain plastic in response to significant input, yet stable in response to irrelevant information?
  – How does system learn to switch between plastic and stable mode?
Categorization of CL Efforts

**Model Growing**
Increase the model capacity for every new task

**Parameter Isolation**
Explicitly identify important parameters for each task

**Regularization**
Penalize (some) parameter variations

**Knowledge Distillation**
Use the model in a previous training state as a teacher

**Rehearsal**
Store old inputs and replay them to the model

*Source: David Abati (CVPR)*
Our Efforts in CL

- **Energy-based Latent Aligner for Incremental Learning**, CVPR 2022
- **Class-Incremental Learning with Cross-Space Clustering and Controlled Transfer**, ECCV 2022
- **Meta-Consolidation for Continual Learning**, NeurIPS 2020
Alright, alright – but why go beyond continual learning?

What is organic lifelong learning?
Supervised vs Human Learning

**Standard Supervised Learning**
- Learn from a fixed dataset with fixed set of labels

**Testing**
- \( X, Y \sim D_1 \)

**Training**
- \( X \sim D_1 \)

**Learner**
- \( T_1 \)

**Continual Learning**
- Learn from available data with or without labels

**Active Learning**
- Learn interactively by asking questions

**Open-set/Open-world Learning**
- Have the ability to say “I don’t know” and ask for more information to learn a new concept and many more….

**Towards Organic Lifelong Learning**
- Refine using newer data without forgetting the past significantly
- Learn only from descriptions without access to data
Learning with Limited Supervision

Standard Supervised Learning

Multi-Task Learning

Transfer Learning

Domain Adaptation

Multi-Task Learning

Transfer Learning

Domain Adaptation

Towards Organic Lifelong Learning
Our Efforts

- **Zero-shot/Few-shot Learning**
  - CVPR 2019, WACV 2021, WACV 2020

- **Data Generation**
  - CVPR 2018, WACV 2022, WACV 2020, WACV 2019

- **Domain Generalization**
  - WACV 2022, BMVC 2021, BMVC 2020

- **Active Learning**

- **Continual Learning**
  - CVPR 2022, ECCV 2022, NeurIPS 2020, CVPR 2021, TPAMI 2021, WACV 2022

- **Domain Adaptation**
  - WACV 2022

Towards Organic Lifelong Learning
Towards Open-World Object Detection

CVPR 2021

To the best of our knowledge, first such effort in object detection

Saying “I-don’t-know” and updating model when the object becomes known

Joint work with:

Joseph K J  Salman Khan  Fahad Khan

NASSCOM AI Gamechanger 2022 Winner Award

(DL Algorithms and Architecture Research)
Open World Object Detection

Problem Setting

- Open-set classes
- Known objects (New)
- Unknown objects
- Known objects (Old)
- Tasks

Towards Organic Lifelong Learning
Open World Object Detection

Problem Setting

- Open-set classes
- Known objects (New)
- Unknown objects
- Known objects (Old)
- Tasks
Open World Object Detection

Problem Setting

- Open-set classes
- Known objects (New)
- Unknown objects
- Known objects (Old)
- Tasks
Qualitative Result

(a) Result after learning Task 2. As Task 3 classes like apple and orange have not been introduced, ORE identifies and correctly labels them as unknown. After learning Task 3, these instances are labelled correctly in sub-figure (b). An unidentified class instance still remains, and ORE successfully detects it as an unknown.
Open World Object Detection

Towards Organic Lifelong Learning
Open World Object Detection

Towards Organic Lifelong Learning
Open World Object Detection

**Overall Architecture**

- **Sort based on objectness score**
- **Label those proposals which do not have an overlapping ground-truth label as an unknown**
Open World Object Detection

Architecture

Input

Backbone

$F$

Unknown aware RPN

Roi Head

Energy based Classification Head

Regression Head

$L_{clf}$

$L_{reg}$

Towards Organic Lifelong Learning

Given a class prototype $p_i$, an input feature of class $c$, $f_c$ is clustered using ($D$ is any distance metric and $\Delta$ is a margin):

$$L_{con}(f_c) = \sum_{i=0}^{C} \ell(f_c, p_i),$$

where,

$$\ell(f_c, p_i) = \begin{cases} D(f_c, p_i) & \text{if } i = c \\ \max\{0, \Delta - D(f_c, p_i)\} & \text{otherwise} \end{cases}$$
Open World Object Detection

Architecture

Input

Backbone

$F$

RPN

Roi Head

Classification Head

Regression Head

$L_{clf}$

$L_{reg}$

Helmholtz free energy: $E(f) = -T \log \int_{l'} \exp \left( -\frac{E(f, l')}{T} \right)$,

Interesting connection with logits: $p(l|f) = \frac{\exp(g_i(f))}{\sum_{i=1}^{C} \exp(g_i(f))} = \frac{\exp(-\frac{E(f, l)}{T})}{\exp(-\frac{E(f)}{T})}$

$
\rightarrow E(f; g) = -T \log \sum_{i=1}^{C} \exp\left( \frac{g_i(f)}{T} \right).
$

Grathwohl et al, Your classifier is secretly an energy based model and you should treat it like one, ICLR 2020

Towards Organic Lifelong Learning
Open World Object Detection

Evaluation Metrics

Measuring quality of unknown detection:

Wilderness Impact: \[ \frac{P_K}{P_{K\cup U}} - 1 \]

Absolute Open-Set Error (A-OSE): Count of unknown objects that get wrongly classified as any of the known class

Measuring quality of labelled detection:

Mean Average Precision (mAP)

Dhamija et al, The Overlooked Elephant of Object Detection: Open Set, WACV 2020
Miller et al, Dropout sampling for robust object detection in open-set conditions, ICRA 2018
Open World Object Detection

Quantitative Results: PASCAL VOC Benchmark

<table>
<thead>
<tr>
<th>Task IDs (→)</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WI</td>
<td>A-OSE</td>
<td>mAP (†)</td>
<td>WI</td>
</tr>
<tr>
<td></td>
<td>(↓)</td>
<td>(↓)</td>
<td>Current known</td>
<td>(↓)</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.02004</td>
<td>7080</td>
<td>57.76</td>
<td>0.0066</td>
</tr>
<tr>
<td>Faster-RCNN</td>
<td>0.06991</td>
<td>13396</td>
<td>56.16</td>
<td>0.0371</td>
</tr>
<tr>
<td>Faster-RCNN + Finetuning</td>
<td>Not applicable as incremental component is not present in Task 1</td>
<td>0.0375</td>
<td>12497</td>
<td>51.09</td>
</tr>
<tr>
<td>ORE</td>
<td>0.02193</td>
<td>8234</td>
<td>56.34</td>
<td>0.0154</td>
</tr>
</tbody>
</table>

We also show results on incremental object detection
**Qualitative Results**

*Open World Object Detection*

*Qualitative Results* 

**Towards Organic Lifelong Learning** 

**Toothbrush** and **Book** are indoor objects introduced as part of Task 4. The detector trained till Task 3, identifies **Toothbrush** as an unknown object in sub-figure (a) and eventually learns it as part of Task 4, without forgetting how to identify **Person** in sub-figure (b).
Suitcase which was identified as unknown is eventually learned in Task 2, along with a false positive detection of Chair.
Summary

- A new and practical setting for object detection: OWOD
- A novel baseline: ORE

More results and ablations in paper

arXiv:

Code:
https://github.com/JosephKJ/OWOD

~200 citations

946 stars
24 watching
150 forks

NASSCOM AI Gamechanger 2022 Winner Award
(DL Algorithms and Architecture Research)
Real-World Applications

New Objects on the Road, We’ll Learn Them Too – IROS 2022

Towards Organic Lifelong Learning
Real-World Applications

Marine Biology

Currently implementing OWOD for Benthic, an underwater rover

http://fathomnet.org/

https://www.mbari.org/technology/benthic-rover/
Novel Class Discovery without Forgetting

Can you recognize these birds?

Joint work with:
Joseph K J, Gaurav Aggarwal, Soma Biswas, Piyush Rai, Sujoy Ghosh, Soma Biswas, Kai Han

ECCV 2022
Novel Class Discovery without Forgetting

Towards Organic Lifelong Learning
Novel Class Discovery without Forgetting

**Proposed Framework**

- **Legend**
  - $\Phi_{FE}$: Feature Extractor
  - $\Phi_{LAB}$: Labeled Head
  - $\Phi_{ULB}$: Unlabeled Head

Towards Organic Lifelong Learning
Novel Class Discovery without Forgetting

Proposed Framework

Legend:
- $\Phi_{FE}$: Feature Extractor
- $\Phi_{LAB}$: Labeled Head
- $\Phi_{ULB}$: Unlabeled Head
- $L_{FD}$: Feature Distillation Loss

Towards Organic Lifelong Learning
Novel Class Discovery without Forgetting

**Proposed Framework**

- **Frozen Feature Extractor** trained on labeled data
- **Feature Extractor** $\Phi_{FE}$
- **Pseudo Latents**
- **Labeled Head** $\Phi_{LAB}$
- **Unlabeled Head** $\Phi_{ULB}$
- **Cross Entropy Loss** $L_{CE}$
- **Feature Distillation Loss** $L_{FD}$

Legend:
- $\Phi_{FE}$: Feature Extractor
- $\Phi_{LAB}$: Labeled Head
- $\Phi_{ULB}$: Unlabeled Head
- $L_{CE}$: Cross Entropy Loss
- $L_{FD}$: Feature Distillation Loss

Towards Organic Lifelong Learning
Novel Class Discovery without Forgetting

Proposed Framework

**Legend**
- $\Phi_{FE}$: Feature Extractor
- $\Phi_{LAB}$: Labeled Head
- $\Phi_{ULB}$: Unlabeled Head
- $L_{CE}$: Cross Entropy Loss
- $L_{MI}$: Mutual Information Loss
- $L_{FD}$: Feature Distillation Loss

Towards Organic Lifelong Learning
### Results

#### Task Aware Evaluation

<table>
<thead>
<tr>
<th>Settings (→)</th>
<th>CIFAR-10-5-5</th>
<th>CIFAR-100-80-20</th>
<th>CIFAR-100-50-50</th>
<th>CIFAR-100-20-80</th>
<th>ImageNet-1000-882-30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods (↓)</td>
<td>Lab</td>
<td>Unlab</td>
<td>All</td>
<td>Lab</td>
<td>Unlab</td>
</tr>
</tbody>
</table>

| RS [19]       | 20.00 | 84.48 | 52.24 | 44.1 | 55.7 | 49.9 | 18.14 | 32.56 | 25.35 | 13.05 | 11.5 | 12.28 | 3.34 | 24.54 | 13.94 |
| NCL [61]      | 20.00 | 59.96 | 39.98 | 13.59 | 57.9 | 35.75 | 10.14 | 12.18 | 11.16 | 12.65 | 4.73 | 8.69 | 1.52 | 11.45 | 6.49 |
| UNO [16]      | 33.16 | **93.22** | 63.19 | 2.01 | 72.78 | 37.39 | 45.76 | 53.85 | 27.81 | 7.95 | 48.7 | 28.33 | 0.75 | 63.4 | 32.08 |
| Ours          | **92.72** | 90.32 | **91.52** | **65.03** | **77.03** | **71.03** | **73.18** | **55.66** | **64.42** | **84.8** | **49.67** | **67.24** | **27.46** | **79.07** | **53.27** |

#### Generalized Evaluation

| UNO [16] | 0 | 71.36 | 35.68 | 0 | 58.15 | 29.08 | 0 | 34.22 | 17.11 | 0 | 41.61 | 20.81 | 0 | 68.34 | 34.17 |
| Ours     | **79.68** | **73.66** | **76.67** | **53.23** | **60.6** | **56.92** | **62.76** | **36.42** | **49.59** | **57.85** | **42.18** | **50.02** | **21.32** | **70.99** | **46.16** |

Towards Organic Lifelong Learning
Summary

- Handling multiple unknown classes while learning continually
- A novel baseline: NCDwF

More results and ablations in paper

<table>
<thead>
<tr>
<th>TIME-1</th>
<th>TIME-2</th>
<th>TIME-3 and more***</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
<td><img src="image3.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

**Unseen Classes at a Later Time? No Problem**

**CVPR 2022**

Joint work with:

Hari Chandana K, Sumitra M, Shivam Chandhok

Towards Organic Lifelong Learning
Continual Generalized Zero-Shot Learning

**Proposed Framework**

**Cosine Similarity based GAN loss:**

\[
L_{GAN} = \mathbb{E}_{x \sim p_{data}}(X_{s, t}^{t-1} \cup R^t) \left[ \log \left[ \cos(x, D(a)) \right] \right] \\
+ \mathbb{E}_{x' \sim p_{\theta}(X'|a)} \left[ \log \left[ 1 - \cos(x', D(a)) \right] \right]
\]

**Classification losses:**

\[L_{rcl}, L_{pcl}, L_{snt} = c_e \log \frac{\exp(\cos(x, D(a_i)))}{\sum_{i \in A^t} \exp(\cos(x, D(a_i))))}, y_i \]

**Incremental bi-directional loss:**

\[L_{vat} = \min_{\theta_y} \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in c_i} \left[ \max(0, X_{\text{sim}}(\mu_{c_j}, \mu_{c_i}')) - (\tau_{\text{sim}}(a_{c_j}, a_{c_i}) + \epsilon) \right]^2 + \left[ \max(0, (\tau_{\text{sim}}(a_{c_j}, a_{c_i}) - \epsilon) - X_{\text{sim}}(\mu_{c_j}, \mu_{c_i}')) \right]^2 \]

\[L_{nuclear} = \| \mu_{c_i} - S_{c_i} \|^2 \]
Continual Generalized Zero-Shot Learning

Proposed Framework

Towards Organic Lifelong Learning
Continual Generalized Zero-Shot Learning

Proposed Framework

Towards Organic Lifelong Learning
Continual Generalized Zero-Shot Learning

Top three cosine similarity scores shared with ‘sheep’ during inference at every task is shown. Unseen classes in red, seen classes in black.

Top 3 cosine similarity scores shared with ‘dolphin’ at every task is shown. Unseen classes in red, seen classes in black.
Continual Generalized Zero-Shot Learning

Summary

- Can we classify unseen classes in a continual learning setting?
- A novel assessment of the problem and baseline

More results and ablations in paper

Code: [https://github.com/sumitramalagi/Unseen-classes-at-a-later-time](https://github.com/sumitramalagi/Unseen-classes-at-a-later-time)

Towards Organic Lifelong Learning
Our Other Ongoing Efforts

- Benchmarks for related settings
- Use of foundation models/prompts to support organic lifelong learning
- Concept-based continual learning
- ....
Open Questions and Challenges

- Variety of settings, how to unify – going from narrow AI to AGI
  - Zero-shot Learning, One-shot Learning, Few-shot Learning, Meta-learning
  - Incremental/Continual/Lifelong Learning
  - Transfer Learning, Domain Adaptation
  - ...

- Do we need general-purpose systems? Or should machines rather be narrow systems?
- When can one transfer? When can one manage to learn with limited labeled data? How to quantify?
- How to unify in a privacy-preserving/federated manner?
Acknowledgements

All students and collaborators

Questions?

vineethnb@cse.iith.ac.in
http://www.iith.ac.in/~vineethnb