Stream-51: Streaming Classification & Novelty Detection From Videos

Ryne Roady*, **Tyler L. Hayes***, Hitesh Vaidya, & Christopher Kanan

Rochester Institute of Technology
Email: tlh6792@rit.edu
Project Web: https://tyler-hayes.github.io/stream51
Personal Web: https://tyler-hayes.github.io
* denotes equal contribution

Continual AI Talk: November 27, 2020

Offline Training Does Not Suffice

Conventional offline training approaches have demonstrated weaknesses in "real-world learning."

- **Inflexible**
  - Real-world data comes from dynamic data distributions
  - Must be able to robustly handle shifts in data and novel inputs
- **Inefficient**
  - Impossible to collect data for all potential scenarios
  - Large amounts of supervised data may not be available for all domains

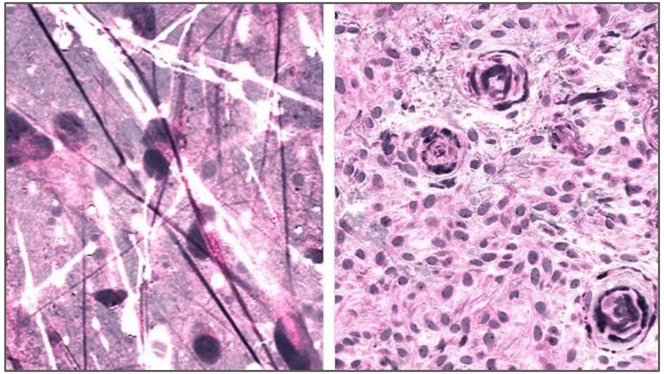
Intelligent agents should have the ability to dynamically adapt to their environment in real-time.

Real-World Learning Scenarios

Find all cars, white lines, traffic signs, and pedestrians...

While ignoring trees, billboards, telephone polls, the sky...[1]

Find all cancerous cells...



While ignoring all other tissue and slide contaminates...[2]

1. Milford, Michael, et al. "Condition-invariant, top-down visual place recognition." *ICRA-2014*.
2. Hollon, Todd C., et al. "Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks." *Nature Medicine* 2020.

Closed-World Is Dangerous and Unrealistic

Real-world consequences for misclassification

3 crashes raise questions about Tesla's Autopilot

Uber in fatal crash had safety flaws say US investigators

6 November 2019 · Share

New York Times

Warnings of a Dark Side to A.I. in Health Care

Last year, the Food and Drug Administration approved a device that can capture an image of your retina and automatically detect signs of ...

Mar 21, 2019

Who's Liable? The AV or the human driver?

Columbia researchers use game theory to help policy makers create liability rules for accidents involving self-driving cars and those driven by people

JAN 14 2020 | BY HOLLY EVARTS | IMAGES CREDIT: SHARON DI AND XU CHEN/COLUMBIA ENGINEERING

AI and Healthcare: The Battle Against Misdiagnosis

The Verge

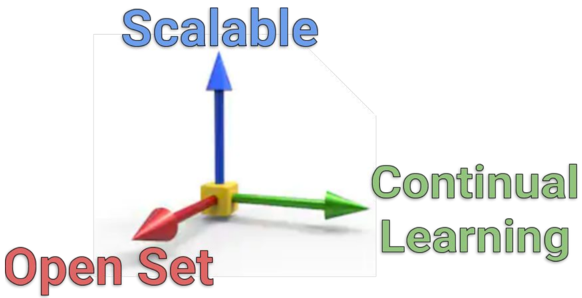
Why cancer-spotting AI needs to be handled with care

And in some areas where tech companies are pushing medical AI, this ...
But as there's no gold standard for cancer diagnosis, particularly early ...

1 hour ago

Lifelong Machine Learning Capabilities

- **Open Set Learning:**
  - Determine if inputs are novel as the system learns
- **Online Streaming Learning:**
  - Retain knowledge from previous experiences
  - Learn from large evolving data streams one sample at a time
  - Closely related to transfer learning and domain adaptation
- Learning capabilities should be closer to humans



Open Set Learning

Train a model to say "I don't know" for objects that it has not learned

Dog / Cat Open Set Classifier

Unknown

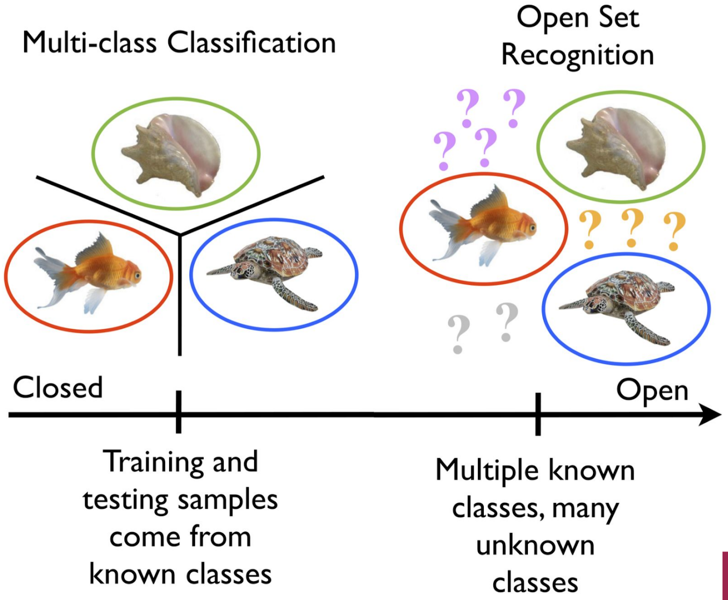
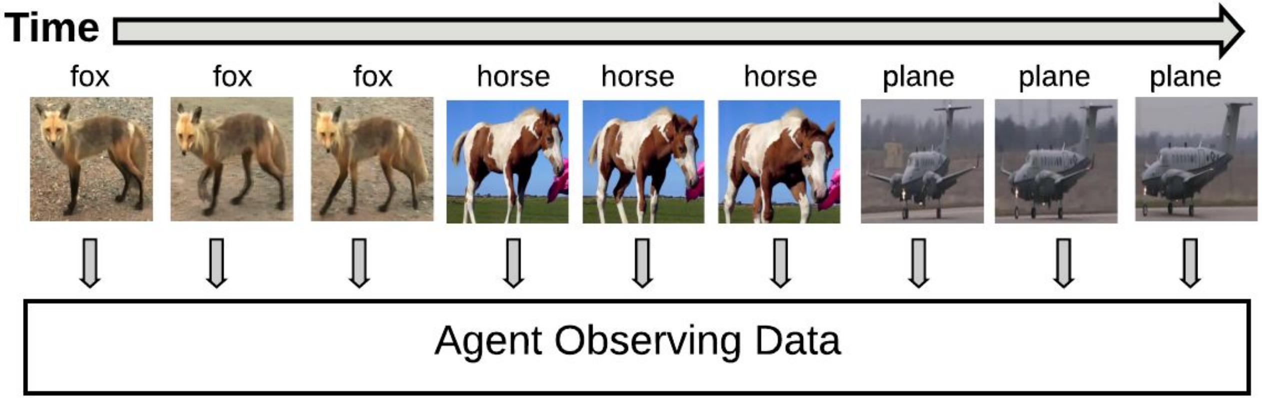


Image: https://www.wjscheirer.com/projects/openset-recognition/

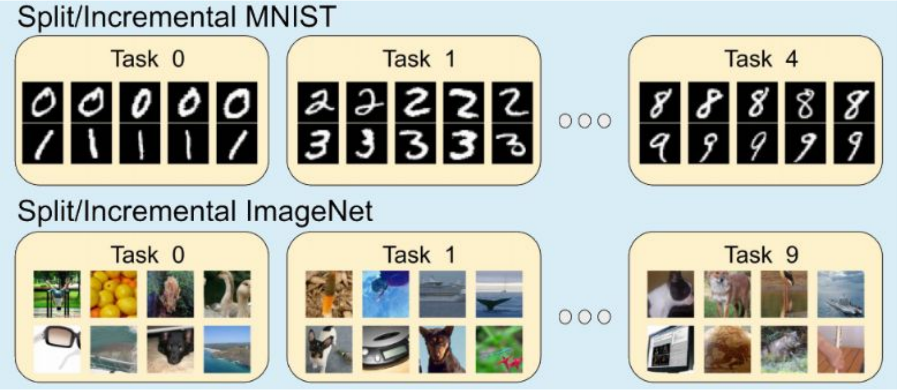
Online Streaming Learning

- Learn from **large evolving data streams** one sample at a time
  - Non-stationary data streams cause catastrophic forgetting in conventional networks
- Retain knowledge from previous experiences
- Learner is on **memory and compute budgets** and can't store all previous training data



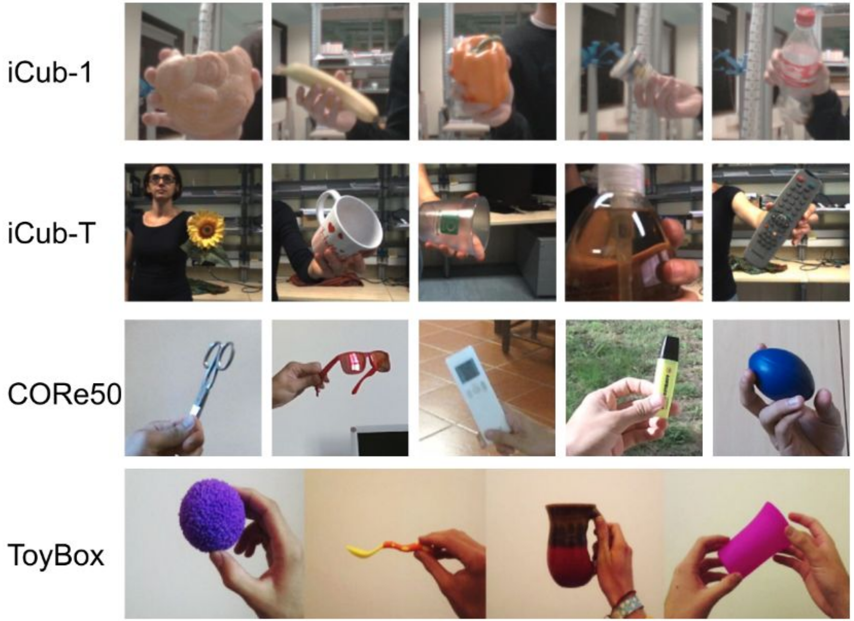
Popular Continual Learning Datasets

- Several of the most popular training paradigms involve learning from static image datasets like MNIST, CIFAR, or ImageNet
- **Problems:**
    - No temporal dependence between frames, which is more realistic
    - Classes are never revisited once they are learned
    - Agents are provided large batches of i.i.d. data for each task



Online Streaming Datasets

- More realistic training paradigms involve learning from temporally correlated video datasets like iCub, CORe50, and ToyBox
- **Problems:**
    - Each dataset is small scale with only 10-20 classes
    - Videos collected in non-natural environments (i.e., first person, hand-held objects)
    - **No open set baselines**



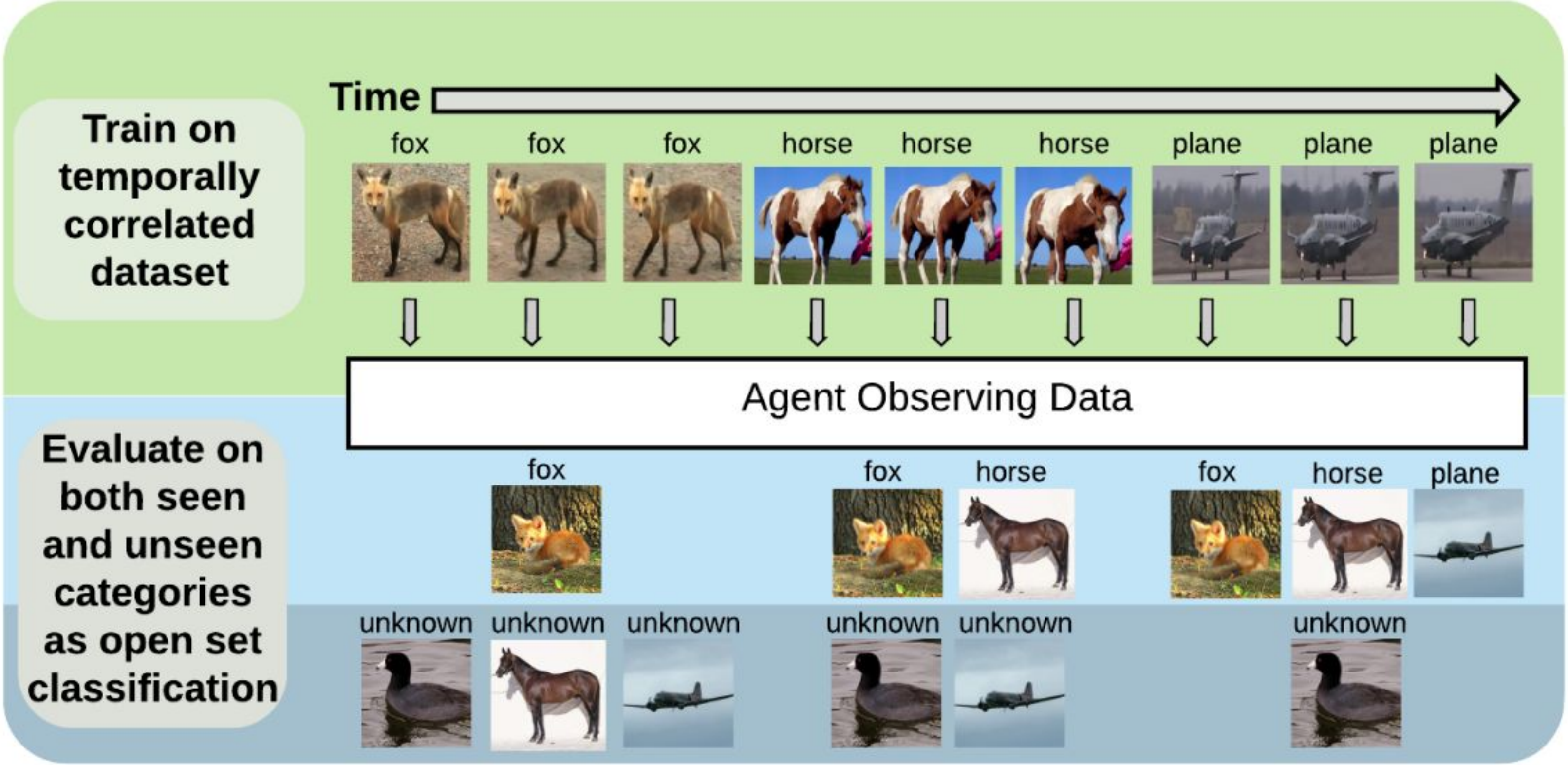

Stream-51

Stream-51 Dataset

- We present the Stream-51 dataset, consisting of 51 training classes from diverse backgrounds and environments
- We establish protocols for **streaming open set classification**
  - 5,100 total static test images (50 samples per training class and 2,550 novel samples)

Dataset	Classes	Images	Videos	Videos/ Class	Avg Frames/ Video	Acq
iCub-1 [12]	10	8,000	40	4	200	hand held
iCub-T [33]	20	400,000	2,000	100	200	hand held
CORe50 [28]	10	165,000	550	55	300	hand held
ToyBox [40]	12	2,300,000	540	45	4,200	hand held
Stream-51	51	150,736	1,136	11-37	132.69	natural/wild

Protocol for Streaming Open Set Classification



Open Set Classification Metrics

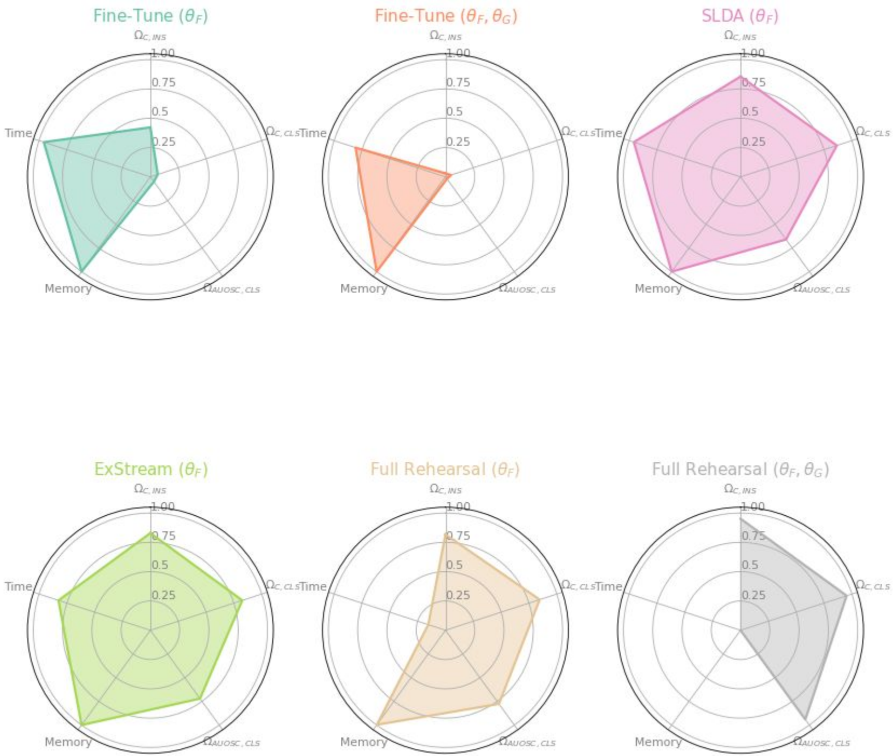
- The streaming agent must learn a classifier *H(X)=F(G(X))* to distinguish learned inputs from novel inputs
  - *S(.)* is an **acceptance score function** that uses a threshold δ to determine if inputs are novel
  - We use a **baseline softmax thresholding algorithm** for open set classification
- We extend the **area under the open set classification curve (OSC)** metric (Dhamija et al., NeurIPS-2018) to the streaming setting by normalizing performance to an offline baseline

$$\hat{y}_{t,\text{novel}} = \begin{cases} 1 & \text{if } S\left(H(\mathbf{X}_t)\right) \geq \delta \\ 0 & \text{if } S\left(H(\mathbf{X}_t)\right) < \delta \end{cases}$$

$$\Omega_{\text{AUOSC}} = \min\left(1, \frac{1}{T}\sum_{t=1}^{T} \frac{\gamma_t}{\gamma_{\text{offline,t}}}\right)$$

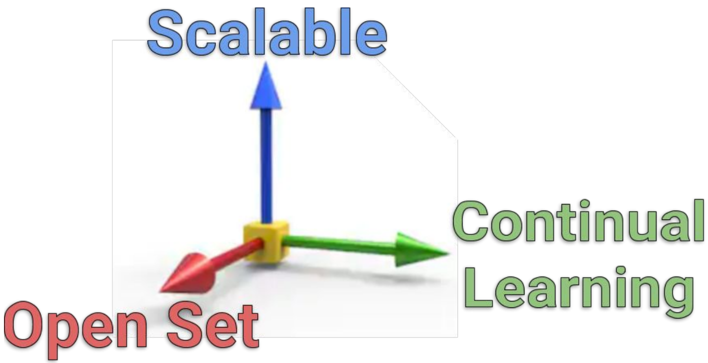
Experimental Results on Stream-51

- Overall accuracy on the dataset is 76.9% and overall AUOSC is 0.710
    - ResNet-18 initialized from Places-365
- Full rehearsal models perform well, but are **computationally expensive**
- Fine-tune models perform poorly, but are cheap in compute
- ExStream and SLDA are the top performing streaming models and **balance performance with compute**



Conclusions

- We introduced the Stream-51 dataset for learning from **temporally correlated videos from natural environments**
- We introduced protocols for streaming open set classification
- **Future Work:**
  - Evaluate more streaming models on Stream-51
  - Evaluate more open set classification methods
    - Examples: ODIN, DOC, Mahalanobis, etc.



Current Field Limitations

- **Dataset limitations:**
  - Not enough **large-scale streaming datasets**
  - CORe50 is one of the largest and most popular streaming datasets, but it only contains 10 object categories
  - Stream-51 is larger, but still limited to only 51 classes
- **Task limitations:**
  - Almost everyone is still training in the **incremental batch learning paradigm** instead of the streaming paradigm
  - Lots of groups are still evaluating on only **permuted and incremental MNIST**
    - These experiments usually do not scale up to larger, natural images
  - Lots of groups still using task labels at test time

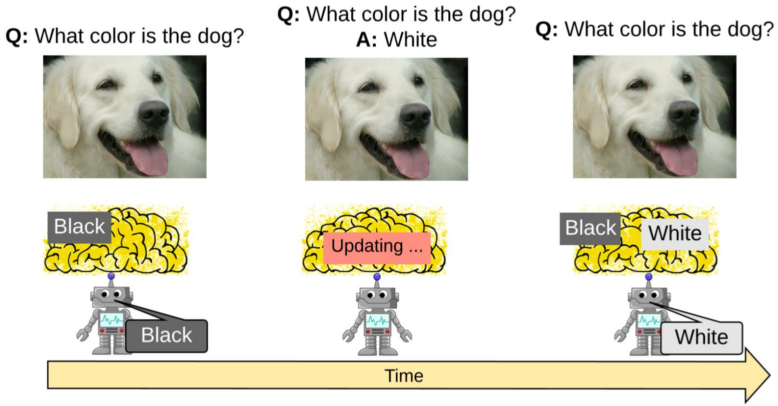
Current Field Limitations

- **No standardized evaluation procedure:**
  - Everyone uses **different datasets**
  - Some researchers train in **batches of multiple classes**, while others train class by class
- **No standardized evaluation metrics:**
  - Most researchers agree on need for **forward and backward transfer metrics**, but everyone uses their own
  - A lot of researchers do not factor **memory budgets and computational time** into metrics, which is important for deployed agents
- **Most researchers are focused on catastrophic forgetting and classification:**
  - There is more to lifelong learning including: open-world learning, curriculum learning, etc.
  - Other problems should be studied including: regression, object detection, etc.

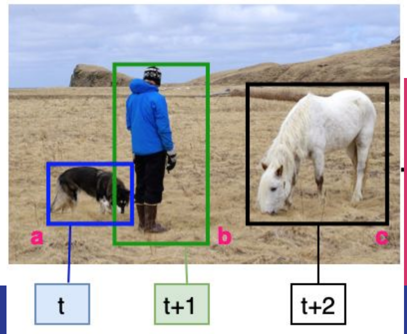
Streaming Learning for Additional Problems

- Our lab has developed streaming models and protocols for several tasks
- **Image Classification:**
  - REMIND (ECCV-2020)
  - Deep SLDA (CVPRW-2020)
  - ExStream (ICRA-2019)
  - FearNet (ICLR-2018)
- **Visual Question Answering:**
  - REMIND (ECCV-2020)
- **Object Detection:**
  - RODEO (BMVC-2020)

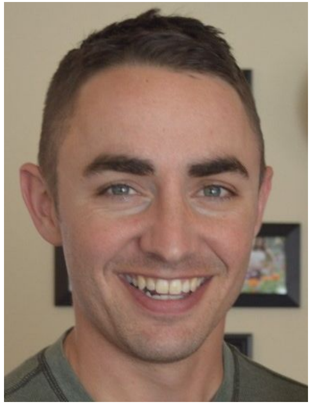
Visual Question Answering



Object Detection

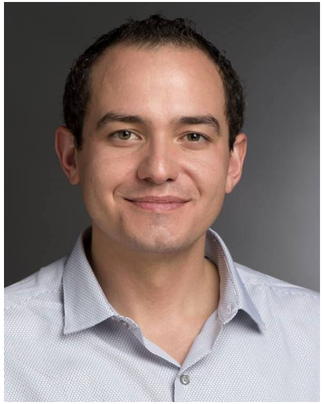


Acknowledgements: Co-Authors and Sponsors



Ryne Roady

Hitesh Vaidya



Chris Kanan

Thank You!

Questions?

Tyler Hayes
tlh6792@rit.edu
https://tyler-hayes.github.io

- We introduced the **Stream-51 dataset** for learning from temporally correlated videos and established baselines and metrics for **streaming open set classification**
- The dataset and code are publicly available:
  - https://tyler-hayes.github.io/stream51
  - https://github.com/tyler-hayes/Stream-51