KTH ROYAL INSTITUTE
OF TECHNOLOGY

Fine-Grained and Continual Visual Recognition for Assisting Visually Impaired People

Marcus Klasson

Division of Robotics, Perception, and Learning

Introduction

- Many everyday items require vision to be recognized

Grocery items ►

Household items ►

- Visually impaired people can use mobile phone apps for recognizing items

Ask questions about images ►

Get feedback to take “better” pictures ►

(a) With “Freestyle” mode. (b) With “Object” mode.

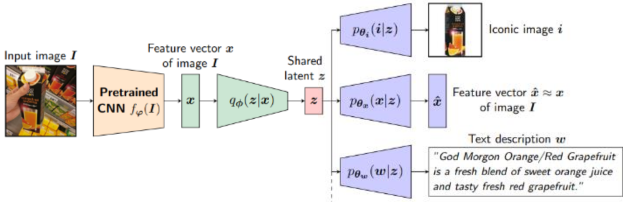
Brady, Erin, et al. "Visual challenges in the everyday lives of blind people." *Proceedings of the SIGCHI conference on human factors in computing systems*. 2013.
Bigham, Jeffrey P., et al. "Vizwiz: nearly real-time answers to visual questions." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010.
Jayant, Chandrika, et al. "Supporting blind photography." *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 2011.

Outline

- Introduction to Assistive Vision Devices
- Fine-grained Image Recognition

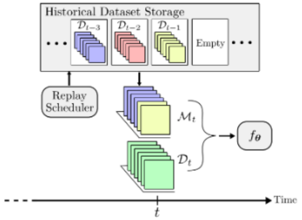
Paper A: Grocery Store dataset ►

Paper B: Grocery classification with web-scraped data ►

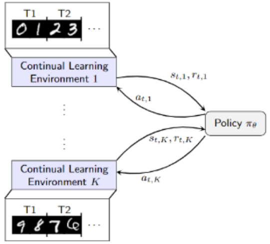


- Continual Learning

Paper C: New continual learning setting ►



Paper D: Learning replay scheduling policies ►



- Future Directions and Conclusions

Paper A: Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "A hierarchical grocery store image dataset with visual and semantic labels." *WACV*. IEEE, 2019.
Paper B: Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "Using variational multi-view learning for classification of grocery items." *Patterns* 1.8 (2020): 100143.
Paper C: Klasson, Marcus, Hedvig Kjellström, and Cheng Zhang. "Learn the Time to Learn: Replay Scheduling in Continual Learning". *Unpublished Manuscript.* 2022.
Paper D: Klasson, Marcus, Hedvig Kjellström, and Cheng Zhang. "Policy Learning for Replay Scheduling in Continual Learning". *Unpublished Manuscript.* 2022.

Introduction to Assistive Vision Devices

Assistive Tools for the Visually Impaired

- Vision impairment defined as uncorrectable loss of ability to see (WHO, 2022)
- Exist different tools for helping visually impaired people with everyday tasks

Navigation

Recognition

Communication

- Camera-based devices to assist with tasks that require visual capabilities

World Health Organization (2022). International Classification of Diseases 11th Revision (ICD-11). URL https://icd.who.int/en Accessed 2022-03-22.
Image courtesy: https://www.srf.nu/ and Vardagstips för personer med synnedsättning. from SRF (2017). Accessed: 2022-03-22.
Bigham, Jeffrey P., et al. "Vizwiz: nearly real-time answers to visual questions." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010.

Camera-Based Assistive Vision Devices

Microsoft SeeingAI (2017)

TapTapSee App (2013)

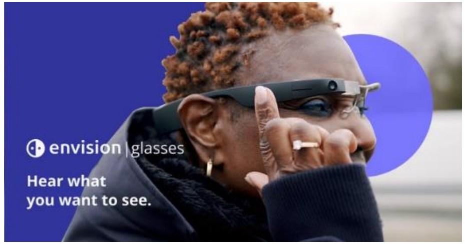
“Man in white shirt biking on street during daytime”

“Red and white horse figure”

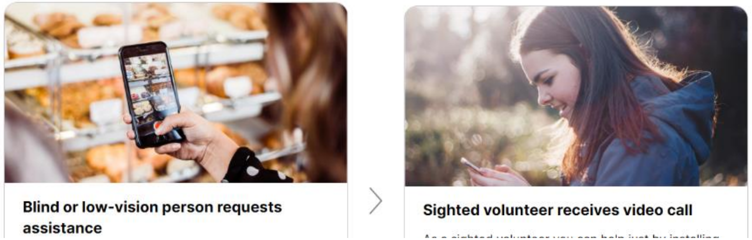
“Black metal bench near wall”

OrCam MyEye2 (2019)

Envision Glasses (2020)



Be My Eyes (2017)

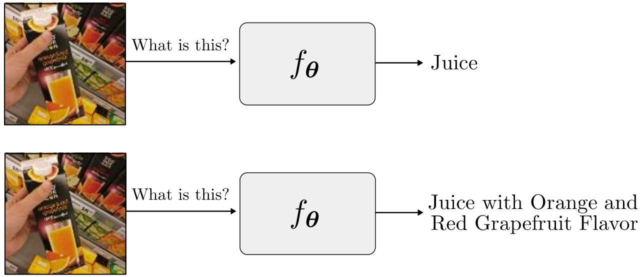


Microsoft Corporation (2017). Seeing AI. URL https://www.microsoft.com/en-us/ai/seeing-ai Accessed 2022-03-22.
Cloudsight, Inc (2013). TapTapSee. URL https://taptapseeapp.com/ Accessed 2022-03-22.
OrCam (2019). OrCam MyEye 2. URL https://www.orcam.com/sv/myeye2/ Accessed 2022-03-22.
Envision (2018). Envision App. URL https://www.letsenvision.com/envision-app Accessed 2022-03-22.
Envision (2020). Envision Glasses. URL https://www.letsenvision.com/envision-glasses Accessed 2022-03-22.
Be My Eyes (2017). Be My Eyes. URL https://www.bemyeyes.com/ Accessed 2022-03-22.

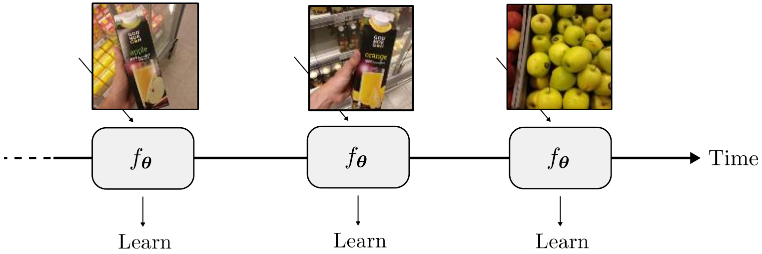
Applications for Assistive Vision

My Focus

1. Fine-Grained Image Recognition



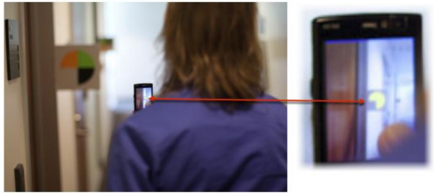
2. Continual Learning of new classes



Other Applications

Visual Question Answering

Navigation and Wayfinding



Generate Text Descriptions

Hudson, Drew, and Christopher D. Manning. "Learning by abstraction: The neural state machine." *Advances in Neural Information Processing Systems* 32 (2019).
Manduchi, Roberto. "Mobile vision as assistive technology for the blind: An experimental study." *International Conference on Computers for Handicapped Persons*. Springer, Berlin, Heidelberg, 2012.
Gurari, Danna, et al. "Captioning images taken by people who are blind." *European Conference on Computer Vision*. Springer, Cham, 2020.

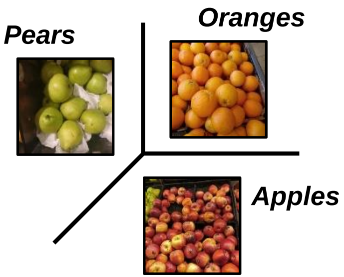
Fine-Grained Image Recognition

Paper A: A Hierarchical Grocery Store Image Dataset with Visual and Semantic Labels

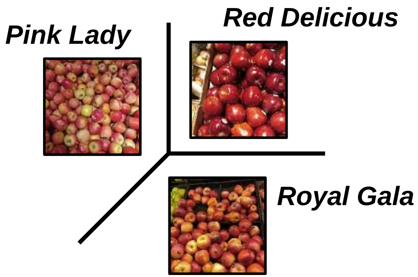
Paper B: Using Variational Multi-view Learning for Classification of Grocery Items

Fine-Grained Image Recognition

Image recognition



Fine-grained image recognition



Challenge: Learn the details that distinguishes visually similar items

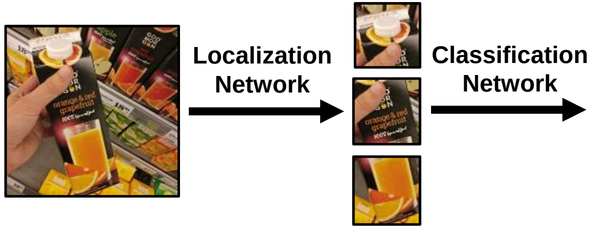
Recognition using External Information

*“**Red Delicious** is a **dark red apple** with **relatively soft pulp**…”*

“Pink Lady reminds of Royal Gala,** though… **even sweeter and crispier apple.”

*“**Royal Gala** is a **crispy… peel is thin** with a **red yellow speckled color.”***

Recognition by Localization-Classification Networks



- …
- Apple Juice
- Orange Juice
- **Orange & Grapefruit Juice**
- Red Grapefruit Juice
- …

Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "A hierarchical grocery store image dataset with visual and semantic labels." *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019.
Wei, Xiu-Shen, et al. "Fine-grained image analysis with deep learning: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

Challenge for Fine-Grained Image Recognition

- Large amounts of data is usually needed for fine-grained image recognition
- Large datasets often include web images that rarely resemble real-world cases
- Recent datasets often aim to collect real-world data

Web images in Imagenet (2009)

Grocery Store (2019)

Household items in ORBIT (2021)

Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009. Images from https://github.com/EliSchwartz/imagenet-sample-images
Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "A hierarchical grocery store image dataset with visual and semantic labels." *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019.
Massiceti, Daniela, et al. "Orbit: A real-world few-shot dataset for teachable object recognition." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Application: Grocery Shopping

- Grocery stores can be challenging environments for image classifiers

Distinguish between similar items

Misplaced items can occur

Various illuminations and occlusions

Varying backgrounds

Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "A hierarchical grocery store image dataset with visual and semantic labels." *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019.

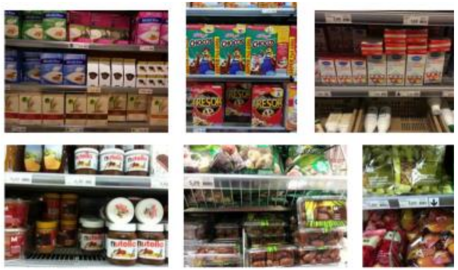
Grocery Recognition for the Visually Impaired

Image Datasets in Grocery Stores

Freiburg Groceries (2016)

MANGO (2015)

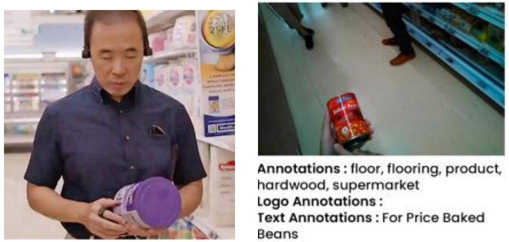
Grocery Products (2014)



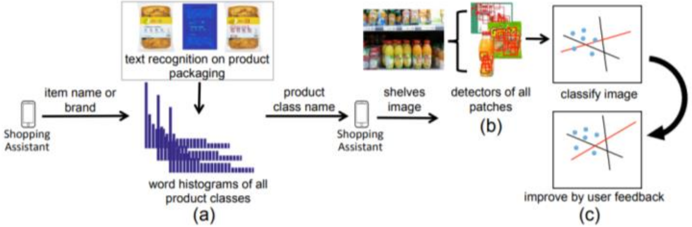
Approaches for Grocery Recognition

"Hands On" (2017)

AiSee (2020)



George et al. (2015)



Jund, Philipp, et al. "The freiburg groceries dataset." *arXiv preprint arXiv:1611.05799* (2016).
George, Marian, and Christian Floerkemeier. "Recognizing products: A per-exemplar multi-label image classification approach." *European Conference on Computer Vision*. Springer, Cham, 2014.
Waltner, Georg, et al. "MANGO-mobile augmented reality with functional eating guidance and food awareness." *International Conference on Image Analysis and Processing*. Springer, Cham, 2015.
Sosa-García, Joan, and Francesca Odone. "“Hands on” visual recognition for visually impaired users." *ACM Transactions on Accessible Computing (TACCESS)* 10.3 (2017): 1-30.
Boldu, Roger, et al. "AiSee: an assistive wearable device to support visually impaired grocery shoppers." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.4 (2020): 1-25.
George, Marian, et al. "Fine-grained product class recognition for assisted shopping." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2015.

Paper A: A Hierarchical Grocery Store Image Dataset with Visual and Semantic Labels

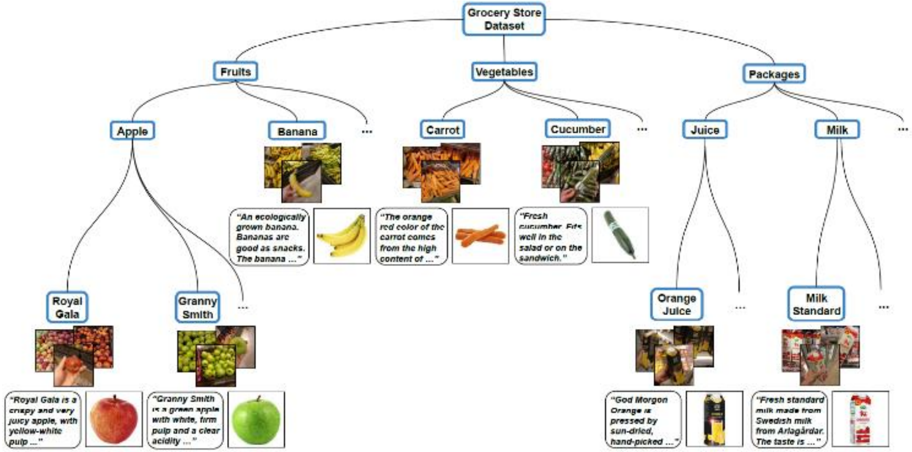
Marcus Klasson, Cheng Zhang, Hedvig Kjellström

In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*

Grocery Store Dataset

<u>Data collection inside grocery stores</u>

<u>Hierarchical labeling of items</u>



- <u>Contributions:</u>
  - 5400 images from 81 fine-grained classes
  - Includes both raw and packaged items
  - Additional web-scraped information of items downloaded from supermarket website

Publicly available: https://github.com/marcusklasson/GroceryStoreDataset

Data Types in the Grocery Store Dataset

Natural images

Iconic images

Text descriptions

Golden Delicious Apple: *"Golden Delicious has a **white juicy** pulp and a **greenish yellow** shell. The taste is **mellow and sweet**, making Golden Delicious **suitable for desserts**."*

Yellow Bell Pepper: *"The yellow pepper is **much sweeter** than the green. It also contains **more vitamins and antioxidants** than the green. Peppers are **good to eat raw in salads** and as garnish, but are **also good to fry, stew or gratinate**, for example with filling..."*

Arla Egological Skimmed Milk: *"Fresh skimmed milk made from Swedish milk from organic Arlagårdar. Skimmed milk has a **delicious full flavor** and is a **popular choice for breakfast cereals**... Milk is a natural source of, for example, **protein, calcium and vitamin B12**... The brand Arla Ko guarantees that the product is **made of 100% Swedish milk**..."*

Publicly available: https://github.com/marcusklasson/GroceryStoreDataset

Paper B: Using Variational Multi-view Learning for Classification of Grocery Items

Marcus Klasson, Cheng Zhang, Hedvig Kjellström

In *Patterns*, *1*(8), 100143

Multi-view Learning Approach

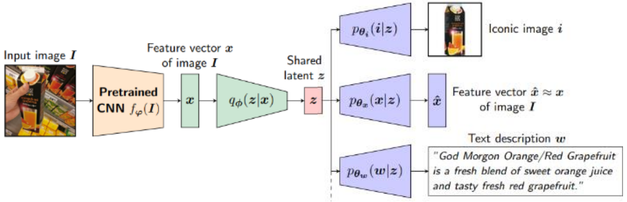
- Data views in our dataset:

(a) Royal Gala (b) Golden Delicious (c) Orange

(a) Royal Gala (b) Golden Delicious (c) Orange

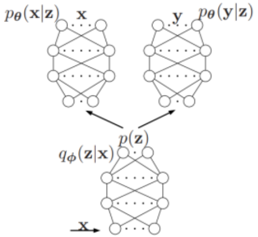
__Golden Delicious Apple:__ *“Golden Delicious has a **white juicy** pulp and a **greenish yellow** shell. The taste is **mellow and sweet**, making Golden Delicious **suitable for desserts**.”*

- Combine all views into shared representation with autoencoders to use for classification



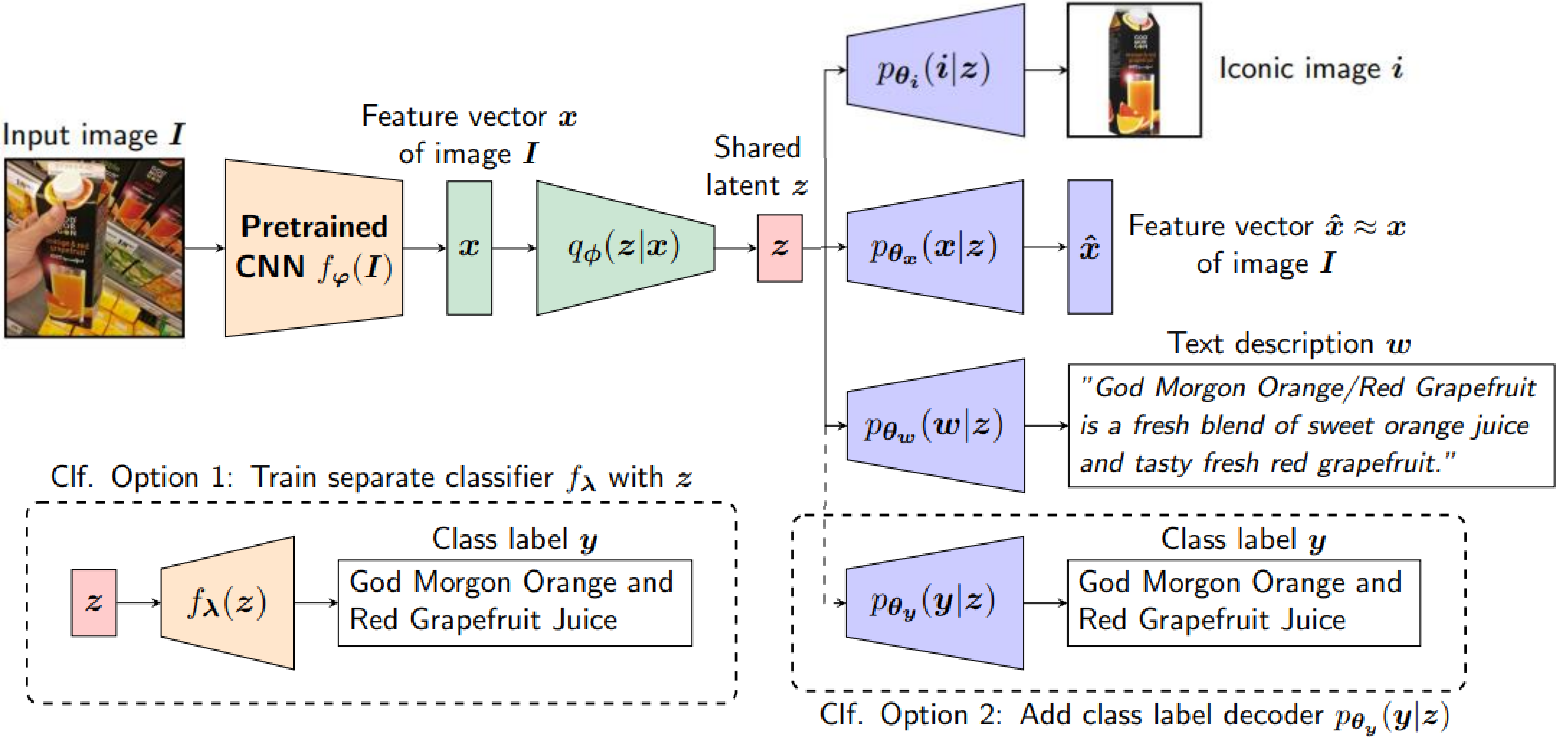
- Learn shared representation with Variational Canonical Correlation Analysis (VCCA)

__VCCA for two data views__

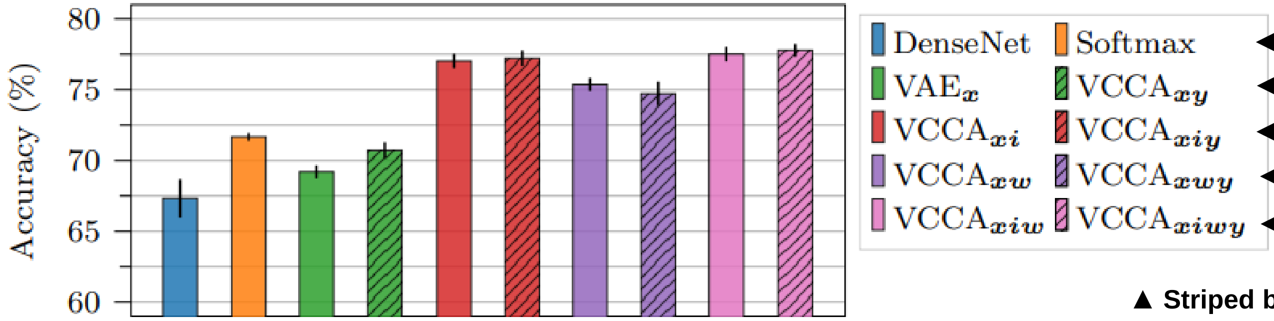


Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.
Wang, Weiran, et al. "Deep variational canonical correlation analysis." *arXiv preprint arXiv:1610.03454* (2016).

Network Architecture for VCCA



Results on Fine-Grained Recognition



◀ **Baselines using natural images**

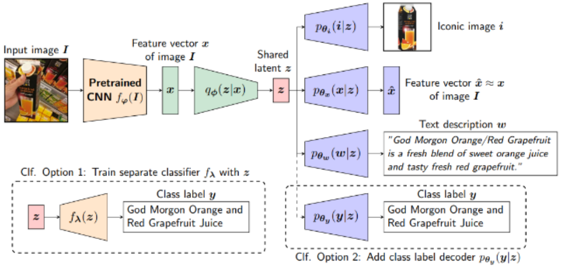
◀ **Learn representation with VAE**

◀ **Add iconic images**

◀ **Add text descriptions**

◀ **Add iconic images and text descriptions**

▲ **Striped bar indicates using class label decoder**



<u>Take-home</u>: Web-scraped information improves accuracy over using the natural images only

Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
Sharif Razavian, Ali, et al. "CNN features off-the-shelf: an astounding baseline for recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014.
Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).

Investigating the Latent Representations

Visualizing latent representations of items with PCA to observe structure differences

Insight: Iconic images separate visually different items

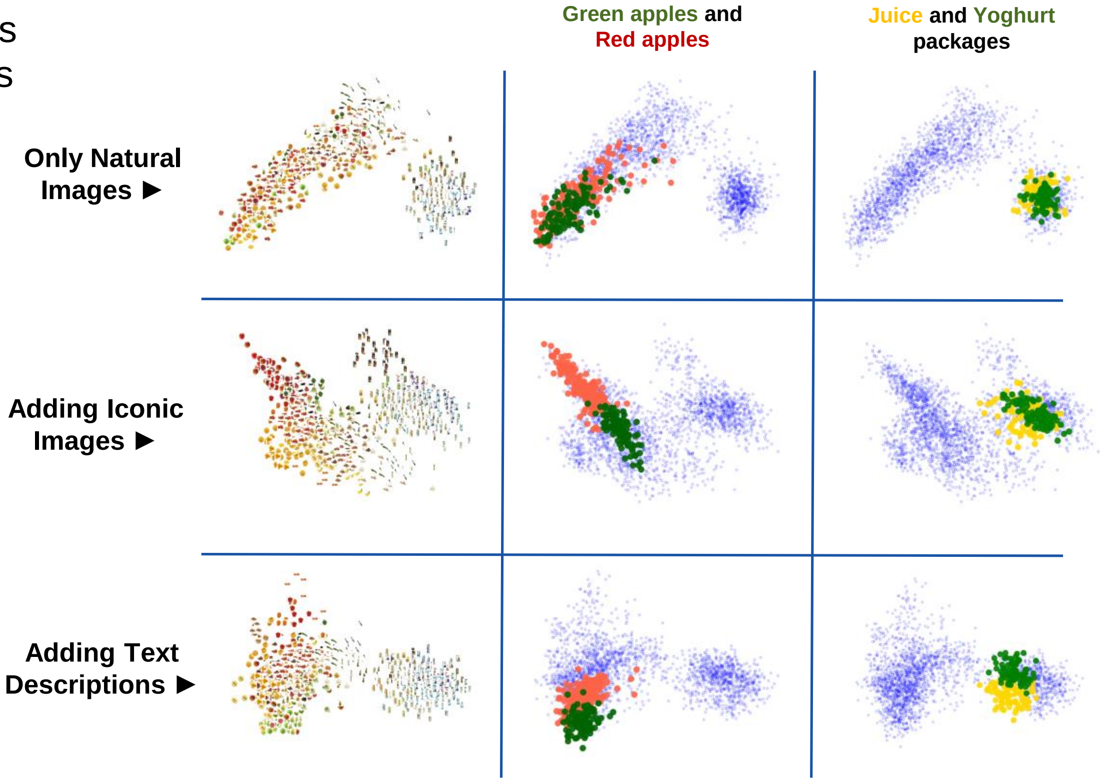
Granny Smith: ***"…green*** *apple with* ***white, firm pulp*** *and a* ***clear acidity*** *in the flavor."*

Royal Gala: ***"…crispy*** *and* ***very juicy*** *apple, with* ***yellow-white pulp****. The peel is thin with a* ***red yellow*** *speckled color."*

Insight: Text descriptions separate items with different ingredients

Tropicana Mandarin Morning: *"…is a* ***ready to drink juice*** *without pulp pressed on* ***orange, mandarin and grapes****. Not from* ***concentrate****. Mildly* ***pasteurized****."*

Valio Vanilla Yogurt: ***"…creamy vanilla yoghurt*** *original… added* ***sugar*** *than regular flavored yogurt. Great for both* ***breakfast and snacks****."*



Pearson, Karl. "LIII. On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11 (1901): 559-572.

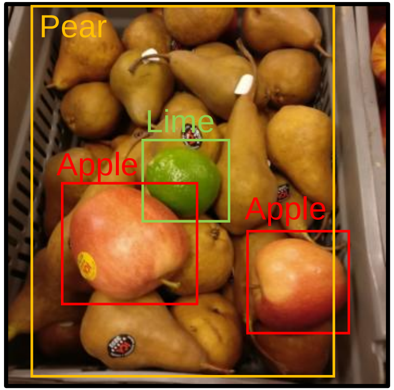
Conclusions

- Web-scraped data of groceries improves the classification accuracy over training with mobile phone images only
  - Reduces the need for collecting more images in the real-world

- Lessons learned on data collection:
  - Record videos rather than still images
  - Annotate as much as possible, e.g., presence and location of items in images, store location

- Utilize the web-scraped images and text more effectively
  - Enable fast adaptation of classifier to new items

Additional annotations

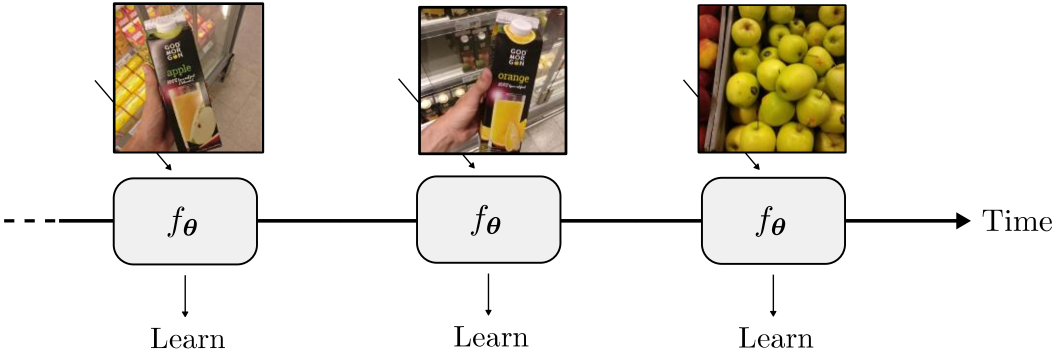


Continual Learning

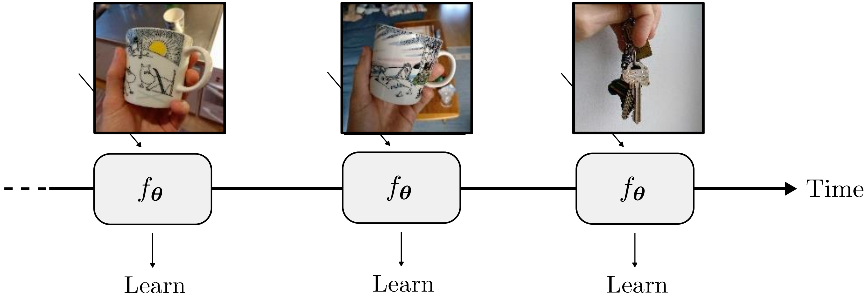
Paper C: Learn the Time to Learn: Replay Scheduling in Continual Learning

Paper D: Policy Learning for Replay Scheduling in Continual Learning

Introduction to Continual Learning



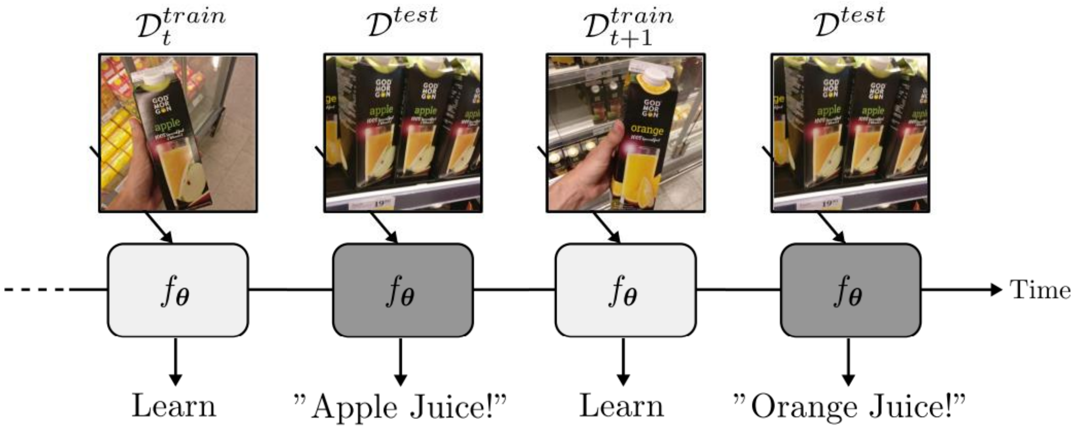
Gradually update the classifier on new tasks, i.e., a new set of object classes



Useful in many settings for assitive vision devices

De Lange, Matthias, et al. "A continual learning survey: Defying forgetting in classification tasks." *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021): 3366-3385.

Main Challenge in Continual Learning



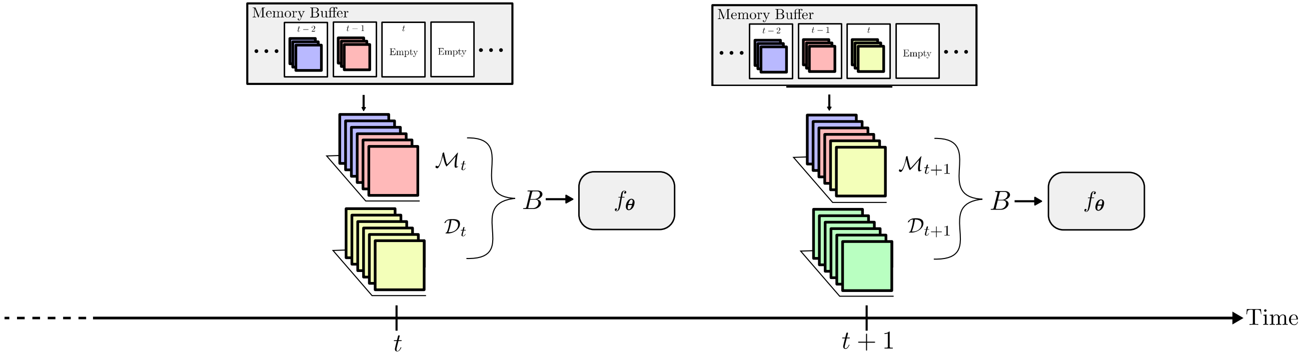
Classifier has forgotten what the apple juice packages looks like!

How do we prevent the classifier from catastrophically forgetting previously learned abilities?

De Lange, Matthias, et al. "A continual learning survey: Defying forgetting in classification tasks." *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021): 3366-3385.
French, Robert M. "Catastrophic forgetting in connectionist networks." *Trends in cognitive sciences* 3.4 (1999): 128-135.

Replay Methods in Continual Learning

- Idea: Remind model of old tasks from stored samples when learning new tasks



- Memory size is the bottleneck for good overall performance
- Various approaches:
  – Improve sample quality in the memory
  – Compress raw data into features to store more memory samples

Chaudhry, Arslan, et al. "On tiny episodic memories in continual learning." *arXiv preprint arXiv:1902.10486* (2019).
Aljundi, Rahaf, et al. "Gradient based sample selection for online continual learning." *Advances in neural information processing systems* 32 (2019).
Hayes, Tyler L., et al. "Remind your neural network to prevent catastrophic forgetting." *European Conference on Computer Vision*. Springer, Cham, 2020.
Pellegrini, Lorenzo, et al. "Latent replay for real-time continual learning." *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.

Remarks on Replay Methods

- Data storage is cheap in general

*“Today, Twitter, LinkedIn, and Facebook each **record over 12M events per second**…”* - Bailis et al. 2017

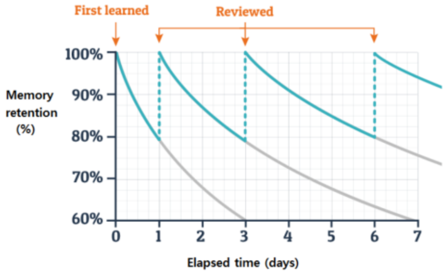
◀ 128 GB USB stick for ~200 SEK can store thousands of images

- Retraining machine learning systems on all data is often prohibited
  - Small replay memories are still required due to time constraints

- Scheduling when to rehearse learned knowledge is important in human memory retention

- Scheduling is needed for continual learning systems when the system must remember large number of tasks

<u>Spaced repetition</u>



Bailis, Peter, et al. "Macrobase: Prioritizing attention in fast data." *Proceedings of the 2017 ACM International Conference on Management of Data*. 2017.
Dempster, Frank N. "Spacing effects and their implications for theory and practice." *Educational Psychology Review* 1.4 (1989): 309-330.
Spaced repetition image from https://www.lifehack.org/851026/spaced-repetition Accessed: 2022-11-04

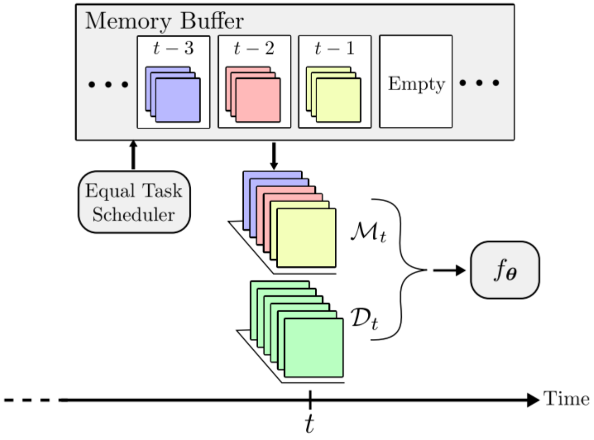
Paper C: Learn the Time to Learn: Replay Scheduling in Continual Learning

Marcus Klasson, Hedvig Kjellström, Cheng Zhang

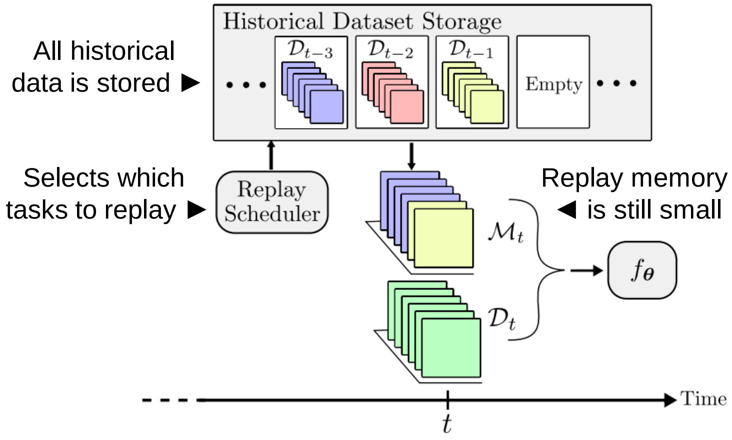
Unpublished Manuscript

Proposed Continual Learning Setting

<u>Traditional setting</u>



<u>Proposed setting</u>

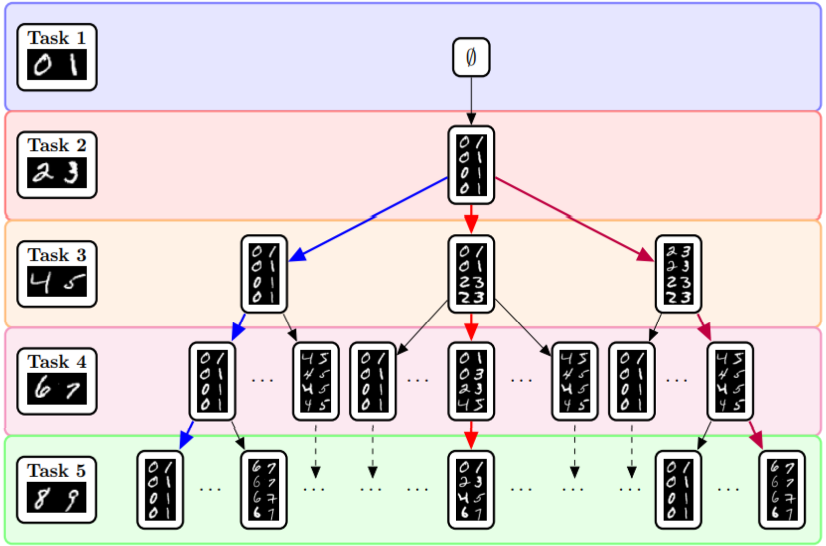


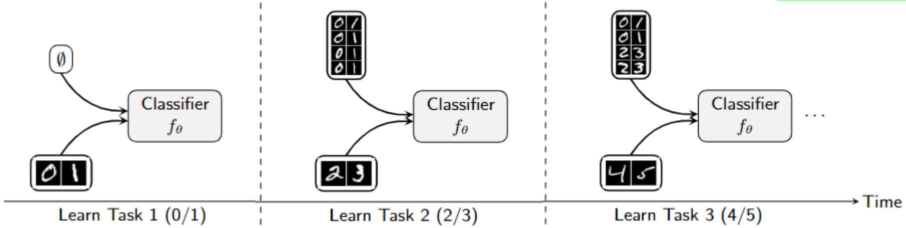
- Overall goal remains the same in new setting
  - Perform well across all tasks and mitigate catastrophic forgetting
- How to study whether replay scheduling is important in this setting?

Replay Scheduling in Continual Learning

- Establish finite number of possible replay memories at every task
  - Example with Split MNIST

- Search space is turned into a tree
  - Use Monte Carlo tree search (MCTS) to find optimal replay schedule
- Run continual learning episodes with selected replay memories





Final classification performance is used for improving the search

Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." *International Conference on Machine Learning*. PMLR, 2017.
Browne, Cameron B., et al. "A survey of monte carlo tree search methods." *IEEE Transactions on Computational Intelligence and AI in games* 4.1 (2012): 1-43.

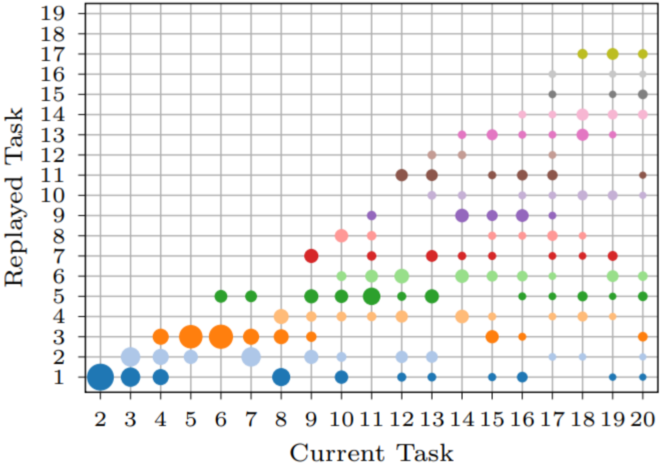
Results with MCTS Replay Schedules

Applying MCTS schedule to other replay methods

Method	Schedule	Split MNIST ACC (%)	Split MNIST BWT (%)	Split CIFAR-100 ACC (%)	Split CIFAR-100 BWT (%)
HAL	Random	96.32 ± 1.77	-3.90 ± 2.28	35.90 ± 2.47	-17.37 ± 3.76
	ETS	97.21 ± 1.25	-2.80 ± 1.59	34.90 ± 2.02	-18.92 ± 0.91
	Heur-GD	97.69 ± 0.19	-2.22 ± 0.24	35.07 ± 1.29	-24.76 ± 2.41
	MCTS	97.96 ± 0.15	-1.85 ± 0.18	40.22 ± 1.57	-12.77 ± 1.30
MER	Random	93.00 ± 3.22	-7.96 ± 4.15	42.68 ± 0.86	-35.56 ± 1.39
	ETS	92.97 ± 1.73	-8.52 ± 2.15	43.38 ± 1.81	-34.84 ± 1.98
	Heur-GD	94.30 ± 2.79	-6.46 ± 3.50	40.90 ± 1.70	-44.10 ± 2.03
	MCTS	96.44 ± 0.72	-4.14 ± 0.94	44.29 ± 0.69	-32.73 ± 0.88
DER	Random	95.91 ± 2.18	-4.40 ± 2.46	56.17 ± 1.30	-29.03 ± 1.38
	ETS	98.17 ± 0.35	-2.00 ± 0.42	52.58 ± 1.49	-32.93 ± 2.04
	Heur-GD	94.57 ± 1.71	-6.08 ± 2.09	55.75 ± 1.08	-31.27 ± 1.02
	MCTS	99.02 ± 0.10	-0.91 ± 0.13	58.99 ± 0.98	-24.95 ± 0.64

Take-home: Replay scheduling can be combined with any replay method to improve the final performance

Visualizing replay schedule from Split CIFAR-100



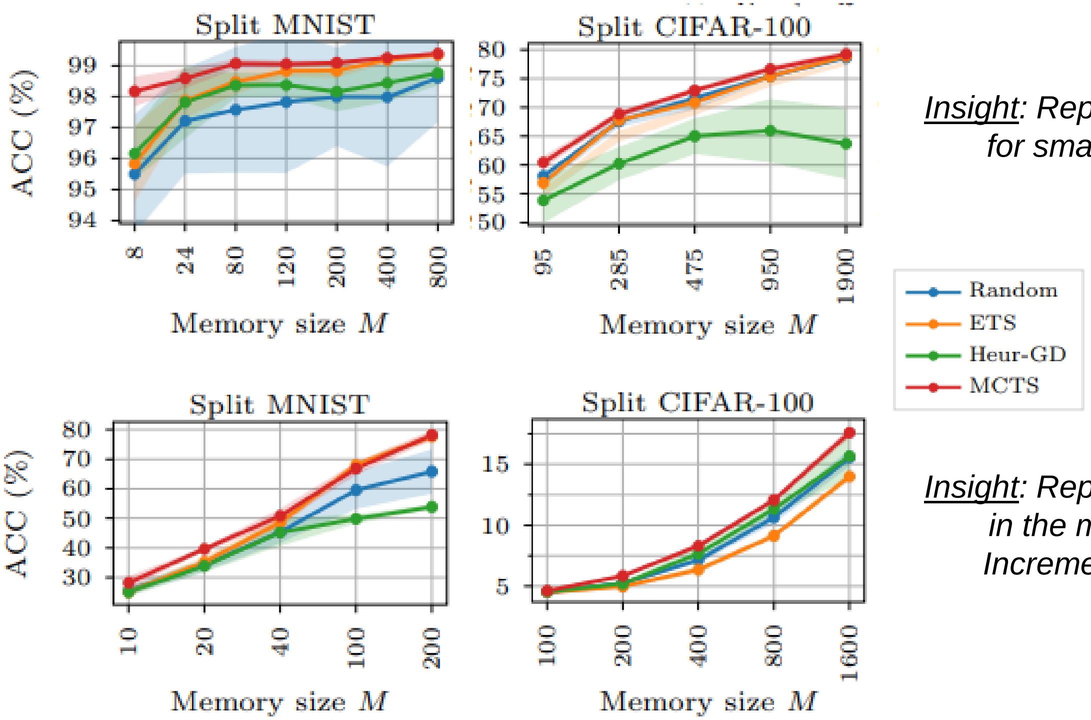
Take-home: Dynamic behavior of replay schedule is beneficial for mitigating catastrophic forgetting

Chaudhry, Arslan, et al. "Using hindsight to anchor past knowledge in continual learning." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 8. 2021.
Riemer, Matthew, et al. "Learning to learn without forgetting by maximizing transfer and minimizing interference." *arXiv preprint arXiv:1810.11910* (2018).
Buzzega, Pietro, et al. "Dark experience for general continual learning: a strong, simple baseline." *Advances in neural information processing systems* 33 (2020): 15920-15930.

Results with Varying Replay Memory Size

Task Incremental Learning ▶

Class Incremental Learning ▶



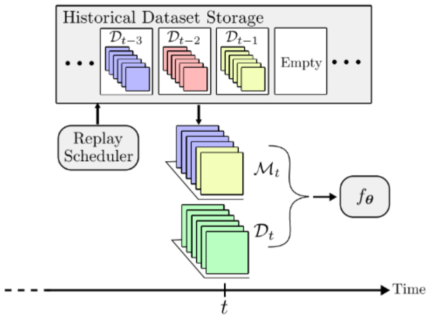
<u>Insight</u>: Replay scheduling is important for small replay memory sizes

<u>Insight</u>: Replay scheduling is important in the more challenging Class Incremental Learning scenario

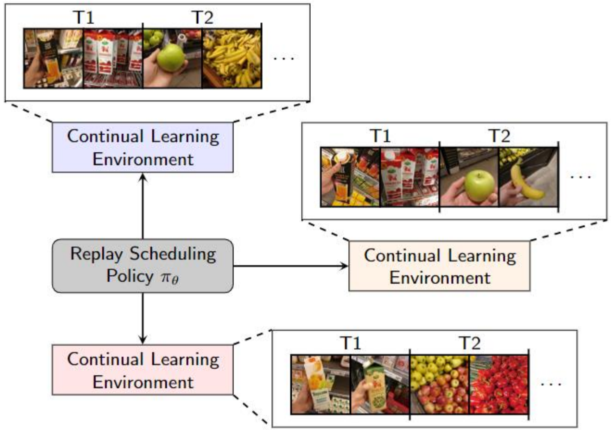
Van de Ven, Gido M., and Andreas S. Tolias. "Three scenarios for continual learning." *arXiv preprint arXiv:1904.07734* (2019).

Taking Replay Scheduling Further

- Showed: Selecting which tasks to replay is important in our proposed continual learning setting



- MCTS schedules only works for the trained dataset
  - Needs to re-run for every new datasets



- Replay scheduling policies should be adaptable to new continual learning scenarios
  - Use Reinforcement Learning to enable transferring policy to new domains

Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

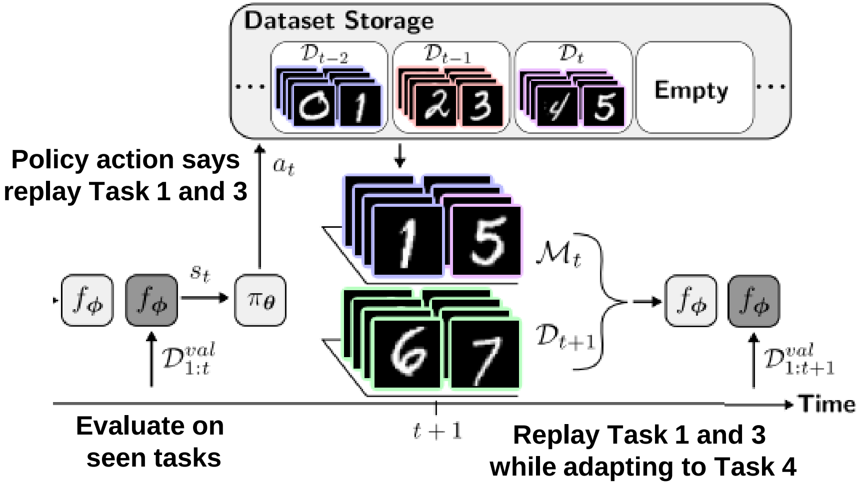
<u>Paper D</u>: Policy Learning for Replay Scheduling in Continual Learning

Marcus Klasson, Hedvig Kjellström, Cheng Zhang

Unpublished Manuscript

Learning the Replay Scheduling Policy

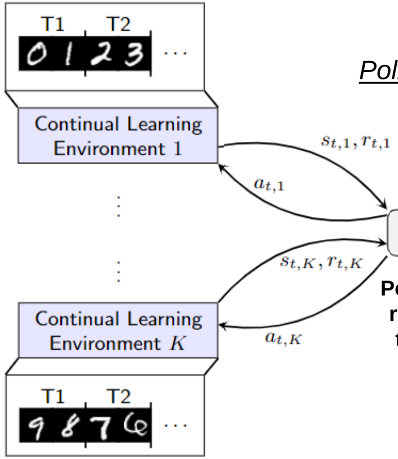
- Goal: Learn policy that selects which tasks to replay for the classifier
- Intuition: Hard and forgotten tasks should be replayed more often
  - Inform policy of task performances in the states to select replay tasks
  - Policy action tells how to compose the next replay memory



Policy Training and Evaluation

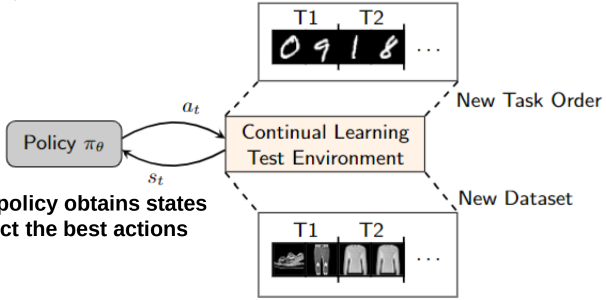
- Policy acts in several continual learning environments during training
  - Eases generalization to new domains
- Evaluate policy in new continual learning environments with unseen settings
  - New task orders
  - New datasets

Policy training



Policy obtains states and rewards to update itself to select better actions

Policy evaluation



Learned policy obtains states to select the best actions

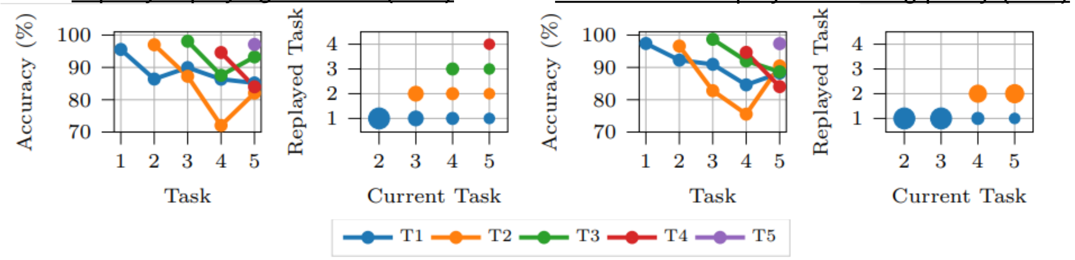
Note: Policy is applied in test environments without updating itself

Zhang, Amy, Nicolas Ballas, and Joelle Pineau. "A dissection of overfitting and generalization in continuous reinforcement learning." *arXiv preprint arXiv:1806.07937* (2018).
Kirk, Robert, et al. "A survey of generalisation in deep reinforcement learning." *arXiv preprint arXiv:2111.09794* (2021).

Results on Policy Generalization Capability

Visualize policy with bubble plot over replay proportion

Equally replaying all tasks (ETS) | With learned replay scheduling policy (A2C)



Take-home: Ours (A2C) can more flexibly focus on replaying tasks that are harder to remember (Task 2)

Ranking between methods to assess generalization

	New Task Order		New Dataset	
Method	S-MNIST	S-CIFAR-10	S-FashionMNIST	S-notMNIST
Random	4.58	5.57	4.53	4.62
ETS	4.5	6.2	4.14	4.76
Heur-GD	5.07	4.63	3.26	5.31
Heur-LD	5.27	4.21	5.68	5.68
Heur-AT	4.92	3.89	4.21	4.97
DQN (Ours)	4.0	4.47	4.97	4.08
A2C (Ours)	3.56	3.21	4.76	3.38
SAC (Ours)	4.1	3.82	4.45	3.2

Green: 1st rank
Orange: 2nd rank

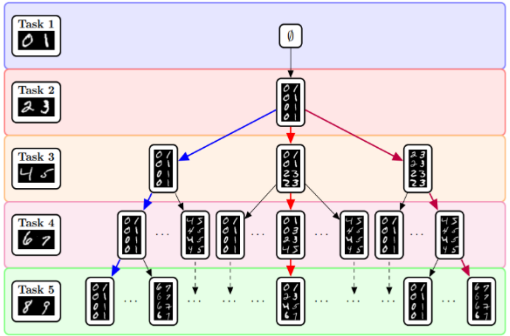
Take-home: Learned policy mostly generalizes well to test environments

- Difficulty generalizing to S-FashionMNIST
- More training environments could help

Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." *arXiv preprint arXiv:1312.5602* (2013).
Mnih, Volodymyr, et al. "Asynchronous methods for deep reinforcement learning." *International conference on machine learning*. PMLR, 2016.
Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." *International conference on machine learning*. PMLR, 2018.

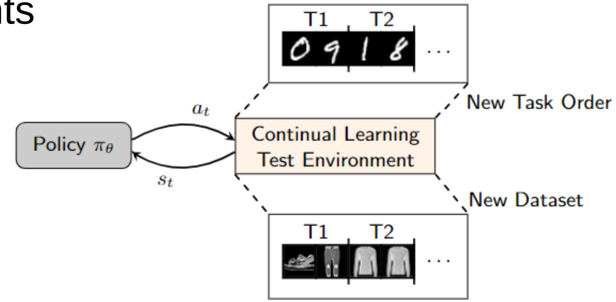
Conclusions

- Showed importance of replay scheduling in various continual learning settings



- Proposed method for learning replay scheduling policies that can be transferred to new continual learning environments

- Policy generalization to more challenging environments
  - Generate diverse training data
  - Combine with more advanced RL methods



- Evaluate on datasets with longer task horizons
  - Improve sample-efficiency

Zhang, Amy, Nicolas Ballas, and Joelle Pineau. "A dissection of overfitting and generalization in continuous reinforcement learning." *arXiv preprint arXiv:1806.07937* (2018).
Kirk, Robert, et al. "A survey of generalisation in deep reinforcement learning." *arXiv preprint arXiv:2111.09794* (2021).

Future Directions and Conclusions

Future Direction I: Federated Learning

- Federated learning with assistive vision apps
  - Locally update image recognition app on-device
  - Distribute global model from server to other users

- Example: Alice and Bob in the grocery store

Alice wants to recognize a milk package with her app

"Unknown item!"

Alice gets help with updating her app with this milk class

Later, Bob wants to recognize the same milk with his app

"Eco Milk!"

Bob's app successfully recognizes the milk thanks to federated learning

- Challenges:
  - User data quality and quantity, variabilities in hardware and network connection, etc.

Li, Tian, et al. "Federated learning: Challenges, methods, and future directions." *IEEE Signal Processing Magazine* 37.3 (2020): 50-60.
Figure courtesy to Cheng Zhang.

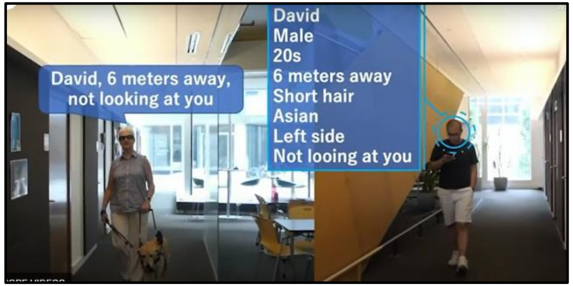
Future Direction II: Disability-First Mindset

- Research that reflects needs in the target community
  - Engagement with visually impaired people is key
  - Ensure datasets resemble scenarios for the target users

- Develop methods for handling realistic, high-variation data
  - Data-efficient methods for video data
  - Robustness to variations across different users and locations

- Discussion on privacy for users and bystanders
  - When and Where can the technology be used in practice?

ORBIT Dataset



Theodorou, Lida, et al. "Disability-first Dataset Creation: Lessons from Constructing a Dataset for Teachable Object Recognition with Blind and Low Vision Data Collectors." *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 2021.
Massiceti, Daniela, et al. "Orbit: A real-world few-shot dataset for teachable object recognition." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
Ahmed, Tousif, et al. "Privacy concerns and behaviors of people with visual impairments." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015.
Lee, Kyungjun, et al. "Pedestrian detection with wearable cameras for the blind: A two-way perspective." *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020.

Conclusions

Fine-Grained Image Recognition

Paper A: Presented a challenging fine-grained recognition dataset of grocery items

(a) Royal Gala (b) Golden Delicious (c) Orange

Golden Delicious Apple: *“Golden Delicious has a **white juicy** pulp and a **greenish yellow** shell. The taste is **mellow and sweet**…”*

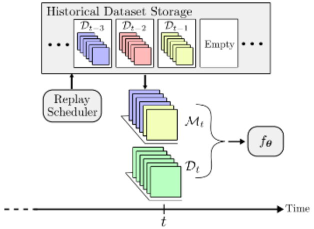
Paper B: Multi-view learning with web-scraped information can reduce the need for real-world images

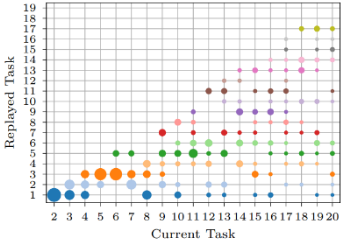
Only Natural Images

Adding Iconic Images

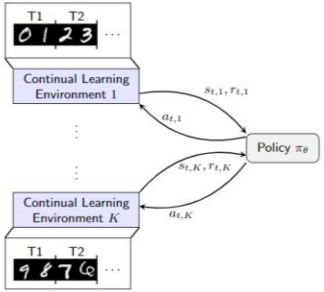
Continual Learning

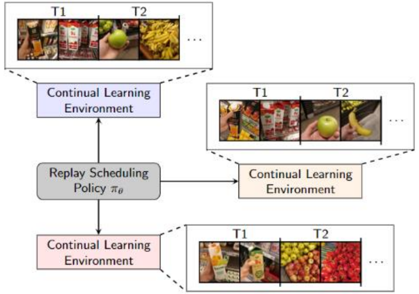
Paper C: Presented a new continual learning setting and replay scheduling which aligns well with real-world needs





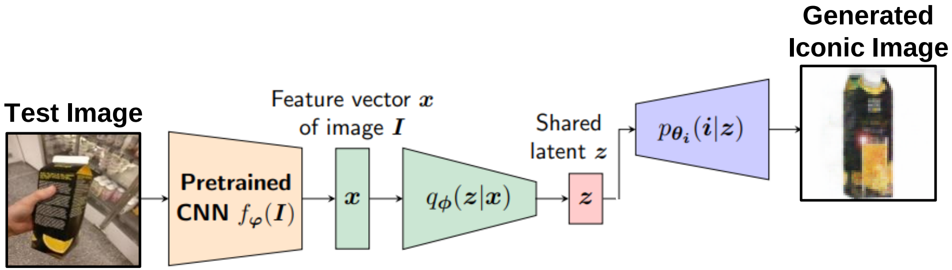
Paper D: Presented a method for learning replay scheduling policies that can be applied in new continual learning scenarios



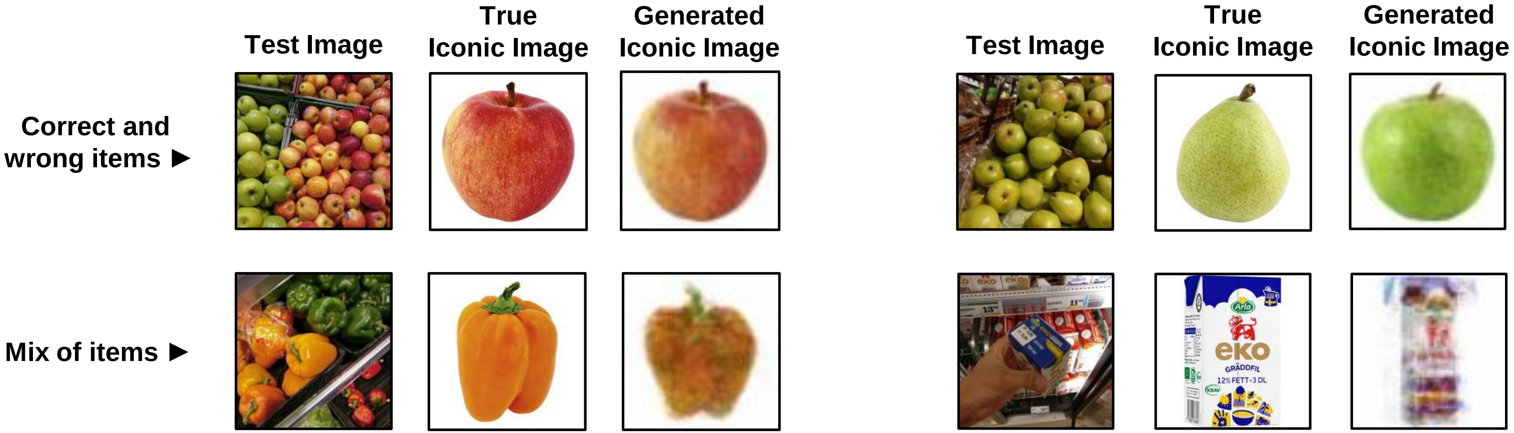


Extra Slides

Generating Iconic images



Insight: Iconic images can be used for sanity-checking the latent space.



References

References I

- Brady, Erin, et al. "Visual challenges in the everyday lives of blind people." *Proceedings of the SIGCHI conference on human factors in computing systems*. 2013.
- Bigham, Jeffrey P., et al. "Vizwiz: nearly real-time answers to visual questions." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010.
- Jayant, Chandrika, et al. "Supporting blind photography." *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 2011.
- Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "A hierarchical grocery store image dataset with visual and semantic labels." *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019.
- Klasson, Marcus, Cheng Zhang, and Hedvig Kjellström. "Using variational multi-view learning for classification of grocery items." *Patterns* 1.8 (2020): 100143.
- Klasson, Marcus, Hedvig Kjellström, and Cheng Zhang. “Learn the Time to Learn: Replay Scheduling in Continual Learning”. *Unpublished Manuscript.* 2022.
- Klasson, Marcus, Hedvig Kjellström, and Cheng Zhang. “Policy Learning for Replay Scheduling in Continual Learning”. *Unpublished Manuscript.* 2022.
- World Health Organization (2022). International Classification of Diseases 11th Revision (ICD-11). URL https://icd.who.int/en Accessed 2022-03-22.
- Synskadades Riksförbund. URL https://www.srf.nu/ . Accessed: 2022-03.02
- Vardagstips för personer med synnedsättning. . SRF (2017). Accessed: 2022-03-22.
- Bigham, Jeffrey P., et al. "Vizwiz: nearly real-time answers to visual questions." *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010.
- Microsoft Corporation (2017). Seeing AI. URL https://www.microsoft.com/en-us/ai/seeing-ai Accessed 2022-03-22.
- Cloudsight, Inc (2013). TapTapSee. URL https://taptapseeapp.com/ Accessed 2022-03-22.
- OrCam (2019). OrCam MyEye 2. URL https://www.orcam.com/sv/myeye2/ Accessed 2022-03-22.

References II

- Envision (2018). Envision App. URL https://www.letsenvision.com/envision-app Accessed 2022-03-22.
- Envision (2020). Envision Glasses. URL https://www.letsenvision.com/envision-glasses Accessed 2022-03-22.
- Be My Eyes (2017). Be My Eyes. URL https://www.bemyeyes.com/ Accessed 2022-03-22.
- Hudson, Drew, and Christopher D. Manning. "Learning by abstraction: The neural state machine." *Advances in Neural Information Processing Systems* 32 (2019).
- Manduchi, Roberto. "Mobile vision as assistive technology for the blind: An experimental study." *International Conference on Computers for Handicapped Persons*. Springer, Berlin, Heidelberg, 2012.
- Gurari, Danna, et al. "Captioning images taken by people who are blind." *European Conference on Computer Vision*. Springer, Cham, 2020.
- Wei, Xiu-Shen, et al. "Fine-grained image analysis with deep learning: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
- Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009.
- Massiceti, Daniela, et al. "Orbit: A real-world few-shot dataset for teachable object recognition." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
- Jund, Philipp, et al. "The freiburg groceries dataset." *arXiv preprint arXiv:1611.05799* (2016).
- George, Marian, and Christian Floerkemeier. "Recognizing products: A per-exemplar multi-label image classification approach." *European Conference on Computer Vision*. Springer, Cham, 2014.
- Waltner, Georg, et al. "MANGO-mobile augmented reality with functional eating guidance and food awareness." *International Conference on Image Analysis and Processing*. Springer, Cham, 2015.
- Sosa-García, Joan, and Francesca Odone. "“Hands on” visual recognition for visually impaired users." *ACM Transactions on Accessible Computing (TACCESS)* 10.3 (2017): 1-30.
- Boldu, Roger, et al. "AiSee: an assistive wearable device to support visually impaired grocery shoppers." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.4 (2020): 1-25.

References III

- George, Marian, et al. "Fine-grained product class recognition for assisted shopping." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2015.
- Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.
- Wang, Weiran, et al. "Deep variational canonical correlation analysis." *arXiv preprint arXiv:1610.03454* (2016).
- Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- Sharif Razavian, Ali, et al. "CNN features off-the-shelf: an astounding baseline for recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014.
- Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).
- Pearson, Karl. "LIII. On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11 (1901): 559-572.
- De Lange, Matthias, et al. "A continual learning survey: Defying forgetting in classification tasks." *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021): 3366-3385.
- French, Robert M. "Catastrophic forgetting in connectionist networks." *Trends in cognitive sciences* 3.4 (1999): 128-135.
- Chaudhry, Arslan, et al. "On tiny episodic memories in continual learning." *arXiv preprint arXiv:1902.10486* (2019).
- Aljundi, Rahaf, et al. "Gradient based sample selection for online continual learning." *Advances in neural information processing systems* 32 (2019).
- Hayes, Tyler L., et al. "Remind your neural network to prevent catastrophic forgetting." *European Conference on Computer Vision*. Springer, Cham, 2020.
- Pellegrini, Lorenzo, et al. "Latent replay for real-time continual learning." *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.
- Bailis, Peter, et al. "Macrobase: Prioritizing attention in fast data." *Proceedings of the 2017 ACM International Conference on Management of Data*. 2017.

References IV

- Dempster, Frank N. "Spacing effects and their implications for theory and practice." *Educational Psychology Review* 1.4 (1989): 309-330.
- Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." *International Conference on Machine Learning*. PMLR, 2017.
- Browne, Cameron B., et al. "A survey of monte carlo tree search methods." *IEEE Transactions on Computational Intelligence and AI in games* 4.1 (2012): 1-43.
- Chaudhry, Arslan, et al. "Using hindsight to anchor past knowledge in continual learning." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 8. 2021.
- Riemer, Matthew, et al. "Learning to learn without forgetting by maximizing transfer and minimizing interference." *arXiv preprint arXiv:1810.11910* (2018).
- Buzzega, Pietro, et al. "Dark experience for general continual learning: a strong, simple baseline." *Advances in neural information processing systems* 33 (2020): 15920-15930.
- Van de Ven, Gido M., and Andreas S. Tolias. "Three scenarios for continual learning." *arXiv preprint arXiv:1904.07734* (2019).
- Sutton, Richard S., and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- Zhang, Amy, Nicolas Ballas, and Joelle Pineau. "A dissection of overfitting and generalization in continuous reinforcement learning." *arXiv preprint arXiv:1806.07937* (2018).
- Kirk, Robert, et al. "A survey of generalisation in deep reinforcement learning." *arXiv preprint arXiv:2111.09794* (2021).
- Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." *arXiv preprint arXiv:1312.5602* (2013).
- Mnih, Volodymyr, et al. "Asynchronous methods for deep reinforcement learning." *International conference on machine learning*. PMLR, 2016.
- Haarnoja, Tuomas, et al. "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor." *International conference on machine learning*. PMLR, 2018.
- Li, Tian, et al. "Federated learning: Challenges, methods, and future directions." *IEEE Signal Processing Magazine* 37.3 (2020): 50-60.

References V

- Theodorou, Lida, et al. "Disability-first Dataset Creation: Lessons from Constructing a Dataset for Teachable Object Recognition with Blind and Low Vision Data Collectors." *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 2021.
- Ahmed, Tousif, et al. "Privacy concerns and behaviors of people with visual impairments." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015.
- Lee, Kyungjun, et al. "Pedestrian detection with wearable cameras for the blind: A two-way perspective." *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020.