HUME CENTER FOR NATIONAL SECURITY AND TECHNOLOGY
VIRGINIA TECH

Heterogeneous Transfer in Deep Learning for Spectrogram Classification in Cognitive Communications

Dr. Tyler Cody

Dr. Peter Beling

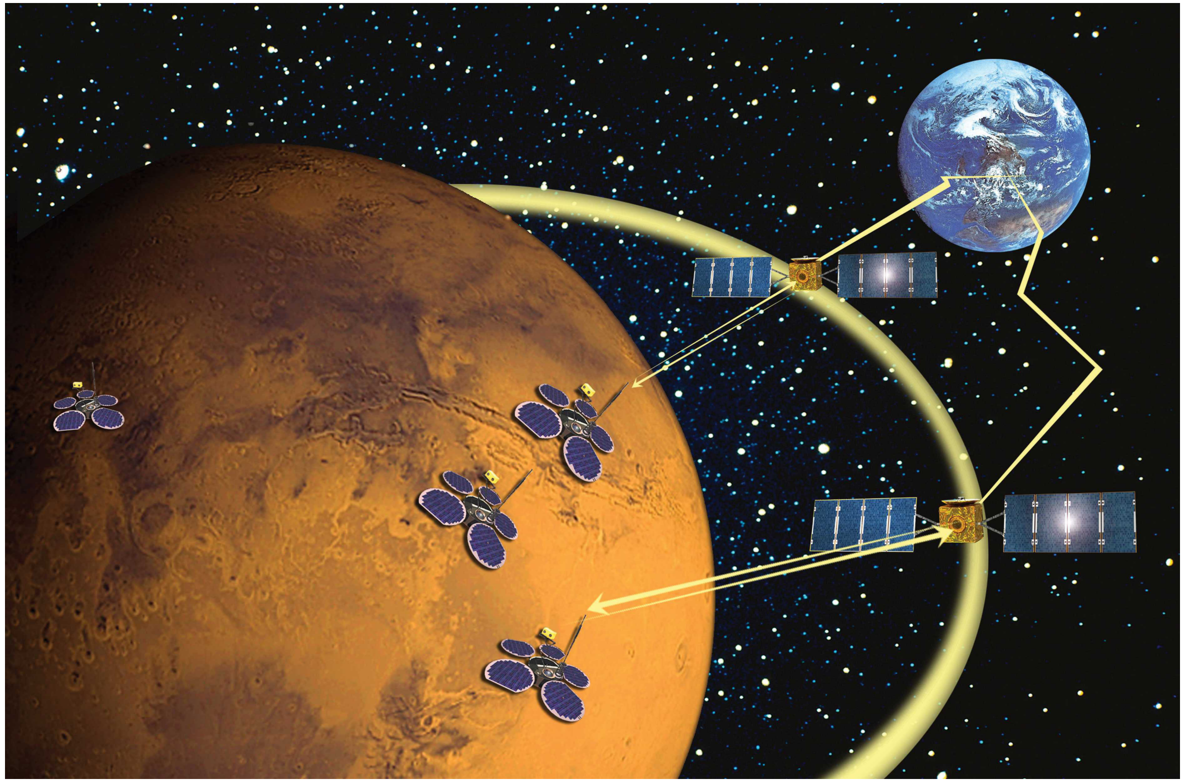
Executive Summary

Convolutional neural networks (CNNs) outperform dense neural networks (DNNs) in heterogeneous parameter and sample transfer.

Within CNNs, smaller networks can saturate offering a tradeoff whereby knowledge transfer is less useful, but training is quicker and less path-dependent.

Accuracy does not capture key trade-offs to characterizing lifecycle performance of spectrogram classifiers.

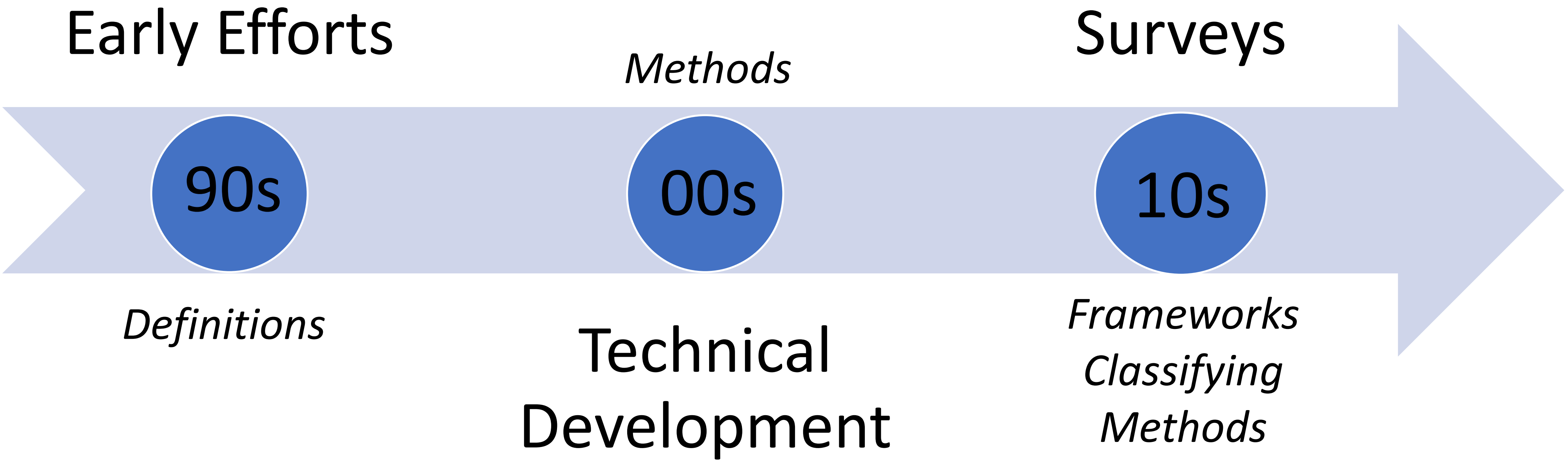
Motivation: Lifecycle Engineering for AI



Transfer Learning (TL)

“the ability of a system to recognize and apply knowledge and skills learned in previous tasks to novel tasks”

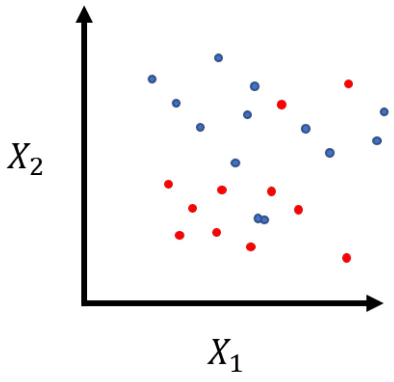
- DARPA BAA 05-29



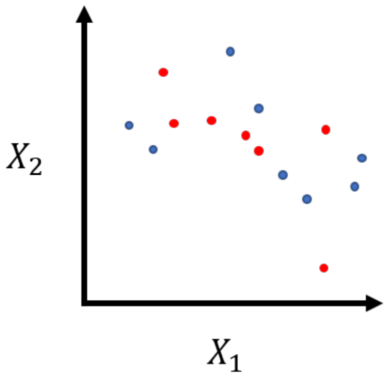
Definition of TL

Dichotomizes learning problems into their **domain** $\mathcal{D}$ and **task** $\mathcal{T}$:

- $\mathcal{D} = \{\mathcal{X},\ P(X)\}, \quad \mathcal{T} = \{\mathcal{Y},\ P(Y|X)\}$



Source



Target

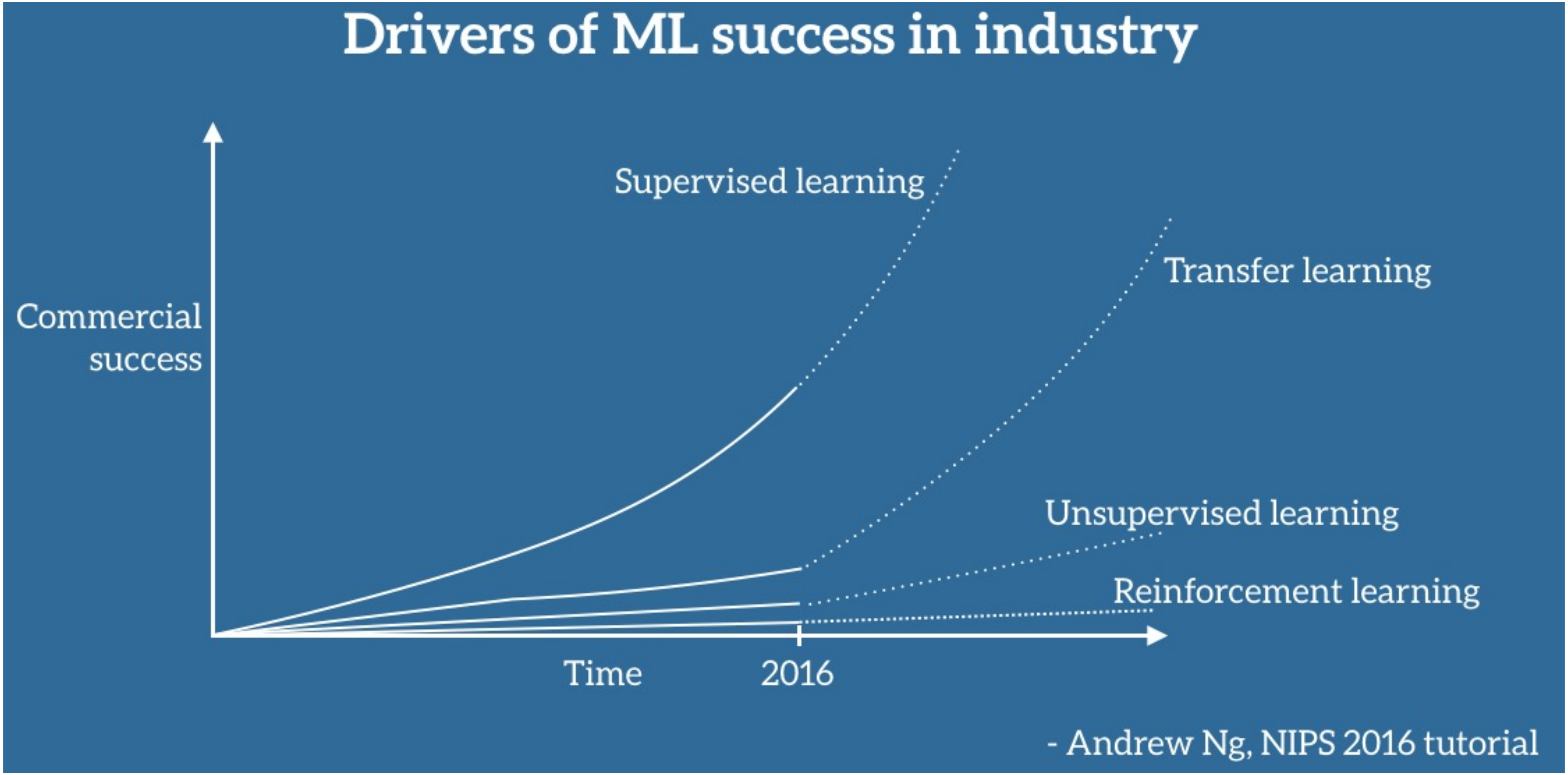
Definition

Given a source domain $\mathcal{D}_S$ and a learning task $\mathcal{T}_S$, and a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$, **transfer learning aims to help improve the learning of the target predictive function $P(Y_T|X_T)$ using knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$** , where

$$\mathcal{D}_S \neq \mathcal{D}_T \text{ (either } \mathcal{X}_S \neq \mathcal{X}_T \text{ or } P(X_S) \neq P(X_T)\text{)},$$

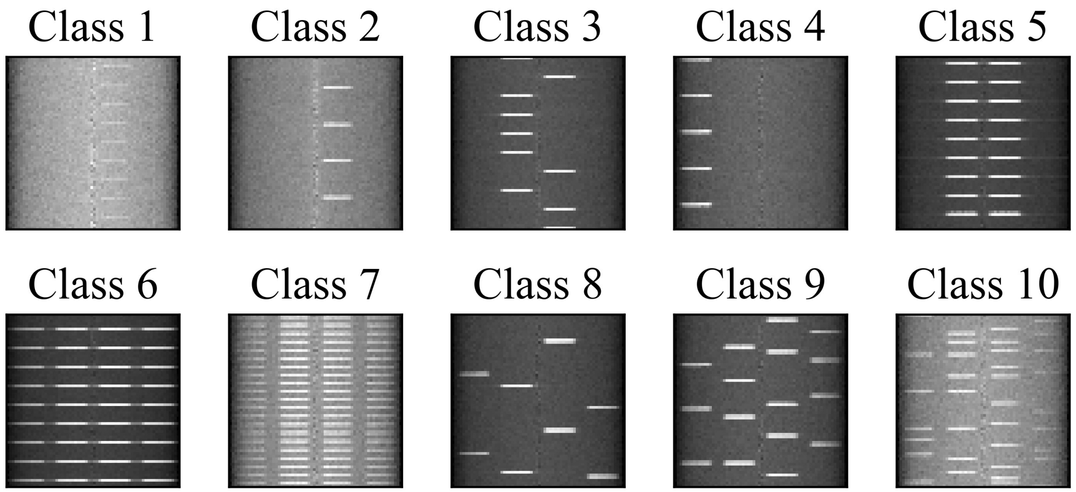
$$\text{or} \quad \mathcal{T}_S \neq \mathcal{T}_T \text{ (either } \mathcal{Y}_S \neq \mathcal{Y}_T \text{ or } P(Y_S|X_S) \neq P(Y_T|X_T)\text{)}.$$

S.J. Pan and Q. Yang, "A survey on transfer learning." IEEE Transactions on Knowledge and Data Engineering, 2010.


Drivers of ML success in industry
Supervised learning
Transfer learning
Commercial success
Unsupervised learning
Reinforcement learning
Time
2016
- Andrew Ng, NIPS 2016 tutorial

Deep Spectrogram Classification

- Variety of applications in literature including to coexist with and discern primary user, spectrum sharing scenarios, spectrum occupancy reconstruction
- Also there exists advanced deep architectures for supervised problem



F. Paisana, A. Selim, M. Kist, P. Alvarez, J. Tallon, C. Bluemm, A. Puschmann, and L. DaSilva, "Context-aware cognitive radio using deep learning," in *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2017, pp. 1–2.

Methods

Parameter and sample transfer –

- Initialize network with parameters from old network, retrain with subset of initial (source) data and (target) data of new class

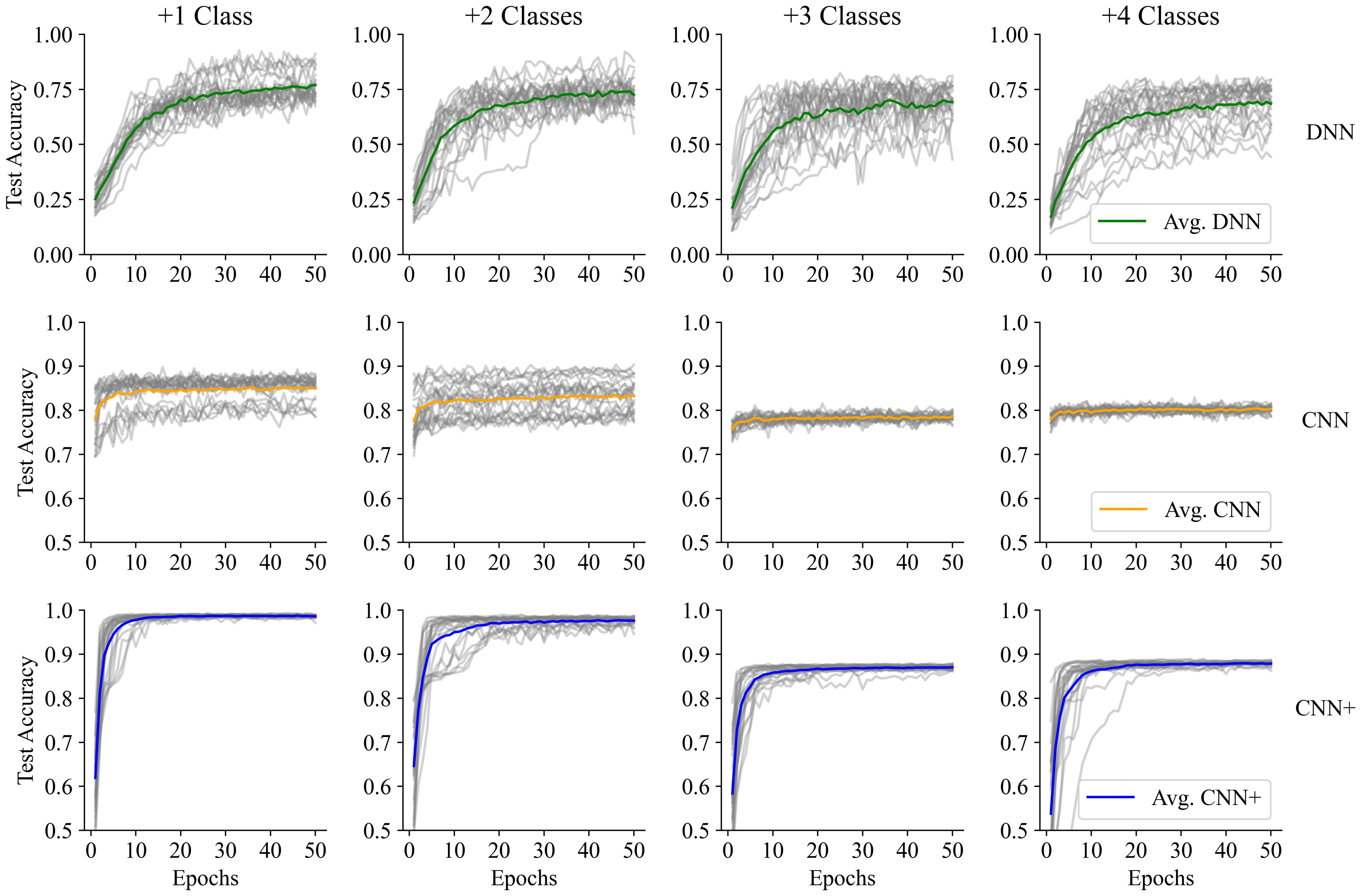
Networks –

- DNN: Fully connected network/MLP
- CNN: Convolutional neural network like Paisana et. al
- CNN+: Deeper version of CNN

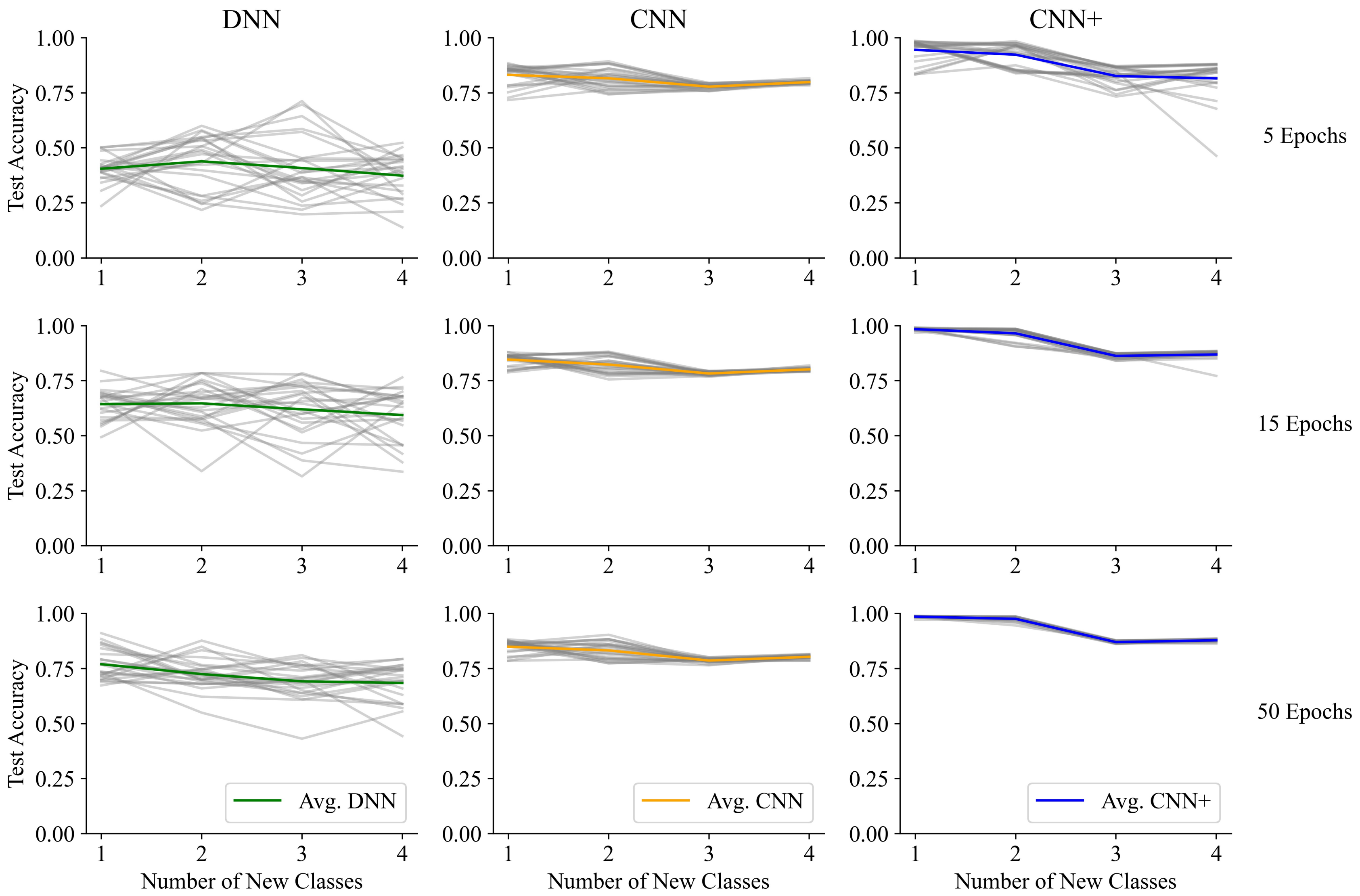
Experiment

- Initial model trained on 6 classes of spectrograms for 150 epochs with 60k 64x64 pixel images
- Additional class provided, output layer of model adjusted, model is retrained for 50 epochs with 1k images from original 6 classes and 800 images from the new class
- Repeat class addition until 10 classes total
- Initial classes fixed, run experiment for all 24 permutations of remaining classes
- Models tested using 800 images from each relevant class

CNNs dominate DNN; CNN, CNN+ have trade-offs



CNN saturates before CNN+



Conclusions

- Given appropriate model complexity, spectrogram classifiers can be fine-tuned to new label-spaces using roughly 10% the number of samples used to train them initially
- CNN and CNN+ dominate DNN in terms of accuracy, time to convergence, and sensitivity to the order in which classes are added
- CNN appears to saturate, while CNN+ does not, allowing CNN+ for greater benefit from fine-tuning at the cost of training time and of stability in performance as classes are added

Closing Remarks

- We may or may not know what changes to a model’s use, e.g. its label-space, will occur in the future
- However, we can build an empirical understanding of a model’s sensitivity to the order in which changes in label-space occur, how long a model takes to recover performance due to such changes, etc.
- And then we can design and operate cognitive communication systems with the limitations of their machine learning subsystems in mind

Contact

Tyler Cody
tcody@vt.edu

Peter Beling
beling@vt.edu

HUME CENTER FOR NATIONAL SECURITY AND TECHNOLOGY
VIRGINIA TECH

Appendix

DNN Architecture

DNN

The input 64x64 image is flattened into a 4096 dimension input layer, followed by fully-connected layers with dimension 128, 32, and 32 respectively, and lastly by an output layer with dimension corresponding to the number of classes. The model has roughly 528,000 parameters.

CNN Architecture

CNN

The input 64x64 image is convolved by 32 3x3 filters, max pooled with size 2x2, convolved with 64 3x3 filters, max pooled with size 2x2, and then flattened and passed through a dropout layer with a 0.50 rate before an output layer with dimension corresponding to the number of classes. The model has roughly 94,000 parameters.

CNN+ Architecture

CNN+

CNN+ has the same architecture as CNN except the images are convolved twice before max pooling and are passed through a fully-connected layer of dimension 128 before dropout. The model has roughly 1,450,000 parameters.