PI: [redacted]

Justification for 30 TB Storage Request

(1) Project space needs and file sizes

This storage request is in relation to our ongoing effort in training deep learning models on large datasets in non-speech audio and radar domains. We note that for this effort we have recently submitted a major OSC grant application. The effort has a critical need for storage of datasets from these domains. Specifically, these datasets are being used in and derived from the following ongoing funded projects:

(a) SONYC, a cyber-physical distributed sensing system for large-scale noise reporting in New York City
(b) PedCyc, an active transportation monitoring system that uses wireless radar sensing meshes for classifying and counting walking/jogging and biking activity on the Olentangy River Road trail through OSU campus

Towards justifying the request quantitatively, we list below the training workflow in each project, the datasets and other files used in each step, their counts and sizes:

Training Steps	Files	File Type	File Count	Total Size
L3 embedding knowledge distillation (SONYC)				
Training student model	Google audioset (environmental)	.h5	60783	11 TB
	Google audioset (music)	.h5	60606	11 TB
	Softmax output files	.npy	121389	1 GB
Feature extraction (transfer learning)	UrbanSound8K dataset	.wav	8732	7 GB
	DCASE2013 dataset	.wav	200	2 GB
	ESC 50 dataset	.wav	2000	1 GB
	Extracted features	.npy	10932	5 GB
Total Size				**22.016 TB**
Radar Target and Activity Classification (PedCyc)				
Training target RNN classifier	Radar target datasets (humans, cars, cattle, dogs)	.bbs	100000	2 TB
Training activity RNN classifier	Radar activity datasets (walking, jogging, biking)	.bbs	100000	2 TB
Total Size				**4 TB**
Radar Noise Rejection (PedCyc)				
Training target vs noise MIL classifier	Radar noise datasets (multiple environments)	.bbs	200000	4 TB
Total Size				**4 TB**
Grand Total Size				**30.016 TB**

(2) Measures we will take to optimize the storage space

- The largest datasets (SONYC) will be saved in a compressed format (.h5)
- All intermediately generated files during training would be stored temporarily in the Scratch directory, and deleted as soon as they are processed
- The datasets would be archived in BuckeyeBox (https://box.osu.edu/) as soon as the projects are completed

(3) Why we need this space

- We need storage space to store the data we already have, and the data we plan to collect, both of which have been factored into the storage estimation

(4) Length of time requested

- A minimum of 2 years, with the possibility of renewal

User Names, Emails, and Resource Usage Estimates

Name	Email	Estimated RUs
[redacted]	[redacted]	9000
		9000
		9000
		3000

Research Leveraging OSC Resources

For several years now, our research group has been exploiting the growing computational capabilities of *motes* (small, battery powered devices with compute, sensing/control, and communication capabilities), to program machine learnt functions on them. Our initial efforts focused on functions derived from so-called shallow supervised learning techniques, such as Support Vector Machines, Decision Trees, and single-layer neural networks. This yielded applications, for instance, in wildlife protection solutions that involved discriminating humans as well as large mammals such as tigers, elephants, or rhinos, on micropower radar-based motes [1, 2]. In turn, this has led us to research that attempts to fit more sophisticated functions on motes, such as joint classifiers and counters [3, 4], multi-class classifiers [5], or more robust featurization [6] that can be deployed in diverse environments while still performing accurately. In the process, we have experienced that the effort for engineering features for shallow supervised techniques scales poorly whereas the effort for deep learning scales better. However, deep learnt machines require significantly more data for training, and usually demand several orders of magnitude more computation and power at run time. To run them on edge devices, and especially on mote-scale devices, they need to be shrunk by several (typically, 2-4) orders of magnitude without sacrificing classification accuracy. This broadly frames the scope of our investigation.

A more detailed explanation of our upcoming research tasks that will leverage OSU compute facilities may be found in our recently submitted Major Project resource request submitted this month to OSC. Here we briefly describe research projects and some of their findings that have leveraged OSC resources over the past four years:

(a) SONYC: A Cyber-Physical System for Monitoring, Analysis and Mitigation of Urban Noise Pollution

In collaboration with the Center for Urban Science and Planning (CUSP), New York University, Sounds of New York City (SONYC) is an NSF project that aims at employing novel machine listening techniques for combating noise pollution in the city that can be adapted for use on sub-wearable scale devices. The objective is to deploy a low power mesh network of 200 sensors around NYC that samples the ambience continuously, and classifies interesting noises on the edge, towards enabling timely response as well as analysis of noise complaints. Since there is a dearth of strongly labeled non-speech sound events that can be used to directly train, say, a Convolutional Neural Network. Instead, we use a recently proposed technique called Look, Listen and Learn (L3) [7] to learn an embedding on a large, open-source, weakly labeled Google audio set [8], that uses audio-visual correspondence to implicitly learn associations between similar sound events, and subsequently use this embedding to train a downstream classifier using limited data. Our main goal is to compress this large embedding (**36 MB**) by 2-3 orders of magnitude,

such that it still produces good quality features at a fraction of the memory and computational cost. Using our OSC startup resources, we have verified a technique that has very strong potential in this regard: *knowledge distillation*.

In this technique, a much smaller deep architecture (for us, a specific combination of convolutional and recurrent layers) is trained as a "student" to mimic the performances of the original L3 embedding, or "teacher". In comparison with L3, which has given **79%** accuracy on a benchmark downstream dataset, our very limited training gave us as much as **66%** with a **3 MB** model, which seems to suggest that student modeling has very good potential for being competitive with the original embedding at **12x** or more reduction in size. Our plan in this direction includes training this (and other candidate student models) extensively against the original Google audio set.

(b) PedCyc: Measuring and Analyzing Active Transportation Using Wireless Sensor Networks

Supported by the Translational Data Analytics Institute at OSU, this project develops a low-cost, privacy-preserving, comprehensive solution to measuring and analyzing active transportation such as walking, jogging, and biking within built environments. Key issues in classifying and counting non-motorized travel include harder detectability compared to motorized vehicles, and lack of low cost, discreet or minimally intrusive counting technology. Our proposed method uses a micro-power doppler radar which provides rich information about movements in a ~13m radius around the sensor. Radar samples are continually fed to an embedded microcontroller that first executes an unwrapped phase-based "displacement detection" to reject noise and in-situ movements, of say trees and bushes. When one or more objects displacing through the scene are detected, the microcontroller analyzes tens of features in amplitude, phase, and time-frequency space on that data in real-time, to compute a classification of the type of displacement as pedestrians or bicycles and accordingly a regression-based count of the instances of that type in the scene. However, while we have had some success in robustly engineering features for target classification, counting and displacement detection [1, 2, 3, 6], shallow machine learning solutions generally adapt poorly to new deployments, changing environmental conditions, or across sensor generations. They also incur feature computation overheads that can impact the performance and lifetimes of the devices.

The main direction of our work in this regard is to explore the feasibility of deep learning methodologies, specifically, variants of Recurrent Neural Networks (RNNs), to eliminate feature design in favor of feature learning while being more efficient than shallow solutions. Once again using OSC resources from our startup fund, we have recently established that using deep learning can successfully alleviate the above issues in a radar target classification problem. In our experiments with FastGRNN [9], a recently published work from Microsoft Research India that uses peephole connections to solve the vanishing and exploding gradient problems in standard RNNs, we achieved **96.55%** accuracy with a **5.77 KB** single layer cell, using only a linear classifier and windowed raw signals as input. It requires only a fixed sub-cut of approximately **1.5 seconds** regardless of the duration of displacement (no need for stacking, dynamic input lengths or consensus). Our experiments show that using the RNN as a featurizer eliminates the need for feature engineering, and end-to-end classification can be performed with a mere **5%** duty cycle, yielding ~**13x** more efficiency than other methods explored. With the requested resources, we hope to continue exploring the efficacy of FastGRNNs and other comparable architectures in radar-based activity classification and counting.

With regards to the team members, while [redacted] have used the OSC startup resources to aid their research as explained above, [redacted] have recently joined these projects.

References

[1]	J. He. Robust mote-scale classification of noisy data via machine learning. Ph.D. Dissertation, OSU, 2015.
[2]	J. He, D. Roy, M. McGrath and A. Arora. Mote-scale human-animal classification via micropower radar. Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (SENSYS), pp. 328-329, 2014.
[3]	J. He and A. Arora. A regression-based radar-mote system for people counting. Proceedings of IEEE International Conference on Pervasive Computing and Communications (PerCom), Budapest, 2014.
[4]	T. Damoulas, J. He, R. Bernstein, C.P. Gomes and A. Arora. String kernels for complex time-series: Counting targets from sensed movement. 22nd International Conference on Pattern Recognition (ICPR), pp. 4429-34, 2014.
[5]	J. Salamon and J.P Bello. Deep convolutional neural network and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 2017.
[6]	D. Roy, C. Morse, M. McGrath, J. He, and A. Arora. Cross-environmentally robust intruder detection in radar motes. Proceedings of 14th International Conference on Mobile Ad-hoc and Sensor Systems (MASS), 2017.
[7]	R. Arandjelović and A. Zisserman. Look, Listen and Learn. In ICCV, 2017.
[8]	Google Audioset. https://research.google.com/audioset/
[9]	A. Kusupati, M. Singh, K. Bhatia, et al. FastGRNN: A fast, accurate, stable and tiny kilobyte sized gated recurrent neural network. In NIPS, 2018 (to appear).