# 2022 LEAD Capstone Poster Session

## Machine Learning Initiatives for UTSW

Daniel J. Stoddard

Assistant Professor

Biophysics

# Abstract


- Cryo-Electron Tomography (Cryo-ET) is a 3D imaging technique of cellular samples that is poised to revolutionize our understanding of disease pathology and progression at the molecular level, yielding extensive insight into basic cellular function. However, identifying proteins and structures of interest in Cryo-ET data, or characterizing larger-scale changes to cellular architecture is an immense challenge. Though this problem will take years to solve, machine learning and adaptive neural networks are almost certainly the tools that will be used to solve it. Neural Networks can be trained on simulated data and real-world cryo-ET data being generated at CEMF. Deep neural networks can be used to evaluate sample quality and increase the efficiency of data collection, saving valuable instrument and bench time. This technology can be used to segment complex imaging data, annotate proteins/structures of interest, and highlight anomalies within the data that are easily missed by human observation. Many datasets also contain valuable information applicable to projects that are unrelated to the initial research question. Maximizing the use of data is imperative wherever possible. Over the next 3-5 years UTSouthwestern needs the hardware and personnel to implement these revolutionary new technologies to make the most of our already robust structural biology programs.

# Objectives

- Establish necessary infrastructure to utilize DEEP learning technology for cryoEM data
- Recruit necessary expertise to take advantage of these advances
- Once established, adapt strategies for specific projects
- Expand to other applications in the future (radiology, oncology etc.)

# Background Information

- Cryo Electron Tomography (cryo-ET) images frozen-hydrated cellular samples in 3D using a high-powered transmission electron microscope.
- UTSW has one of the best Cryo-ET programs in the world.
- Machine learning has made astounding gains in recent years, very soon neural networks will be able to segment, annotate, and analyze complex images, such as cellular cryo-ET data.
- Cryo-ET data analysis will dramatically advance our understanding of the cellular context of disease, and basic biology alike, revealing for new drug targets and potential treatments for some of the worlds most prevalent pathologies.
- To take advantage of these emerging technologies, minimal infrastructure and personnel can be integrated into UTSWs already robust computational infrastructure.

# Project Plan

- Admittedly, this project may be best implemented 2-4 years from now, coincident with the development of DEEP learning platforms specifically tailored for cryoEM data.
- Initially:
- Hire 1-3 computational biologists
- Purchase 2-4 high-end GPUs/servers and integrate thim into our computing architecture.
- Work on 2-3 pilot projects showing the capabilities of ML for cryoET
- Expand focus throughout the structural biological community at CEMF

# Application of What You Learned at LEAD

- This plan stems directly from the PAR (Problem, Action, Response) system espoused by our LEAD mentors.
- The problem is a good one, we have too much useful data to analyze by brute force methods!
- The action is to utilize the tools (not necessarily to create them) to analyze this rich and complex data.
- The result is greater utilization of our structural biological data, potentially leading to insights and breakthroughs in drug discovery, treatment, and fundimental biological research.

# Proposed Budget

- Personnel: $120,000-300,000/year
- Infrastructure: $50,000-$100,000 ~10-20K/year(maintenance and licensing) thereafter
- Total budget: $180,000-$500,000 depending on scale

# Innovation and Significance

DEEP learning has broad application in Bioinformatics, Molecular dynamics simulations, cell biology, Radiology, the list goes on. When developed, these analyses will accelerate both drug discovery and basic research alike. UTSouthwestern stands to benefit more than most from these advances because our institution is involved in so many disciplines that will be improved with the advent of Deep learning applications. In fact, one may be hard pressed to find disciplines that will not utilize DEEP learning on some level in the coming years. Lets put ourselves in a position to take maximal advantage of these approaching advances to image analysis.

# References


- Moebel, E., Martinez-Sanchez, A., Lamm, L. et al. Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. Nat Methods 18, 1386–1394 (2021). https://doi.org/10.1038/s41592-021-01275-4
- Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun Biol 4, 874 (2021). https://doi.org/10.1038/s42003-021-02399-1
- Georgi Dimchev, Behnam Amiri, Florian Fäßler, Martin Falcke, Florian KM Schur,
- Computational toolbox for ultrastructural quantitative analysis of filament networks in cryo-ET data,Journal of Structural Biology,Volume 213, Issue 4,2021,107808,ISSN 1047-8477,
- https://doi.org/10.1016/j.jsb.2021.107808.
- Lamm L, Righetto RD, Wietrzynski W, Pöge M, Martinez-Sanchez A, Peng T, Engel BD. MemBrain: A deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms. Comput Methods Programs Biomed. 2022 Sep;224:106990. doi: 10.1016/j.cmpb.2022.106990. Epub 2022 Jul 1. PMID: 35858496.