# Neural Network for Nanoscience: Scanning Electron Microscope Image Recognition

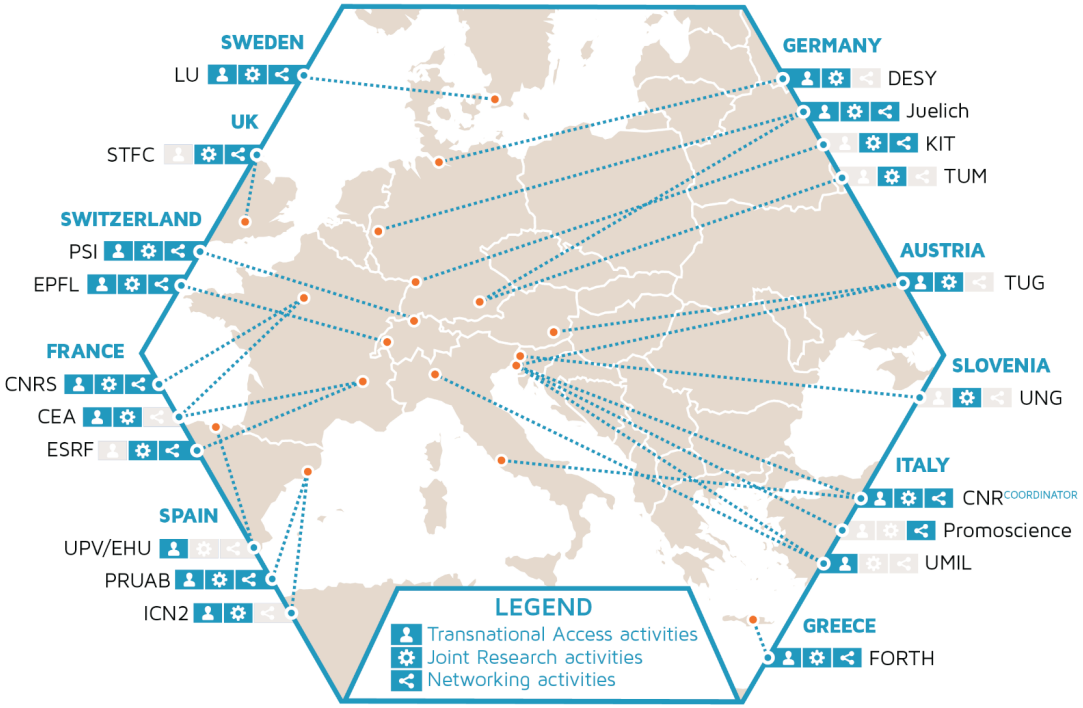
Giuseppe Piero Brandino - R&D

eXact-lab s.r.l.

# NFFA-Europe project

## Nano Foundries and Fine Analysis

- Network for multidisciplinary research at the nanoscale, from synthesis to nanocharacterization to theory and numerical simulation.
- 20 partners
- The widest range of tools for research at the nanoscale



# Nanoscience needs

- Manage and store high volume/variety of data generated
- High Performance data access
- Data Sharing: reuse and interoperability
- Identify and organize Metadata associated to data
- Metadata search engine: data accessible and searchable

# Nanoscience needs

- Manage and store high volume/variety of data generated
- High Performance data access
- Data Sharing: reuse and interoperability
- Identify and organize Metadata associated to data
- Metadata search engine: data accessible and searchable

# Datascience in NFFA

- Develop the first distributed Information and Data Repository Platform (IDRP) for the nanoscience community
- Employ deep learning to classify scientific images
- Integrate a SEM image classifier to provide additional metadata
- Produce a semi-automatic workflow from the instrument to the repository
- Allow semantic search on the database

# Glossary

- Training from scratch: the weights of all the layers are re-initialized
- Transfer learning: the weights of all layers have been trained on a different dataset
- Fine tuning: all weights are initialized to the last check point and are allowed to vary (with small learning rate)
- Feature extraction: last layer(s) re-initialized, previous layers frozen to checkpoint

# Supervised learning

We created and manually annotated the first sample of classified SEM images (for a total of ~20.000 images)

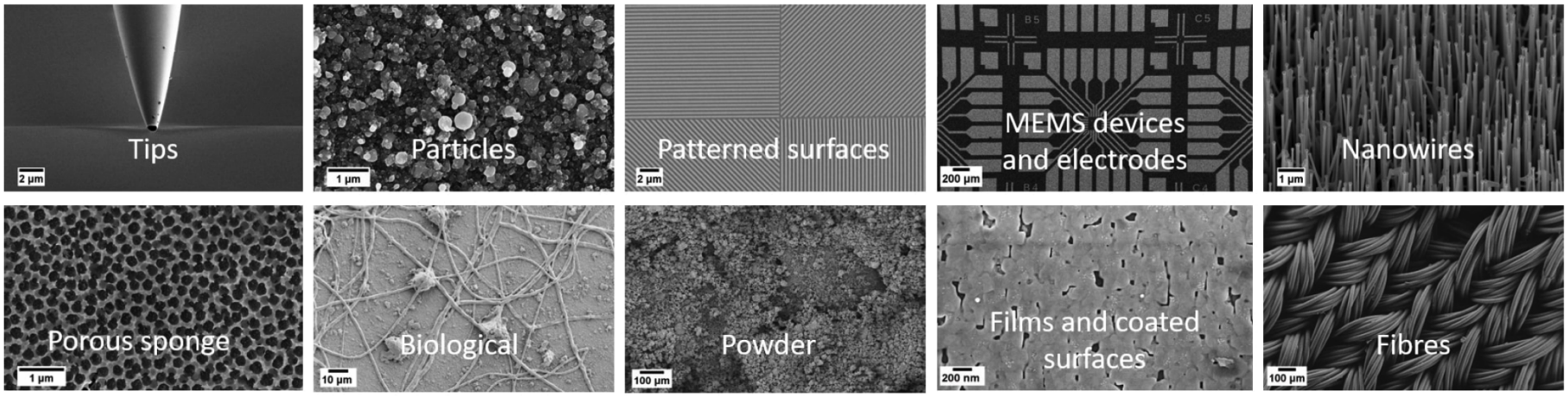



**Figure 1.** Categories chosen for SEM images. The dimensionality of nanoscience objects provided the basis for the choice. Other categories, such as `Biological` and `Tips` were added as these were common images found in the SEM database.

Modarres et al. 2017, Scientific Report, in press

# Machine learning tools

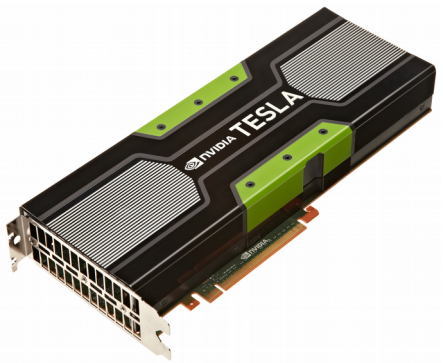
- Tensorflow, Tensorflow-Slim
- Network architectures
  - Inception v3, v4
  - ResNet v2
  - DenseNet (preliminary results)
- 85/15 splitting of the dataset

# Computational resources

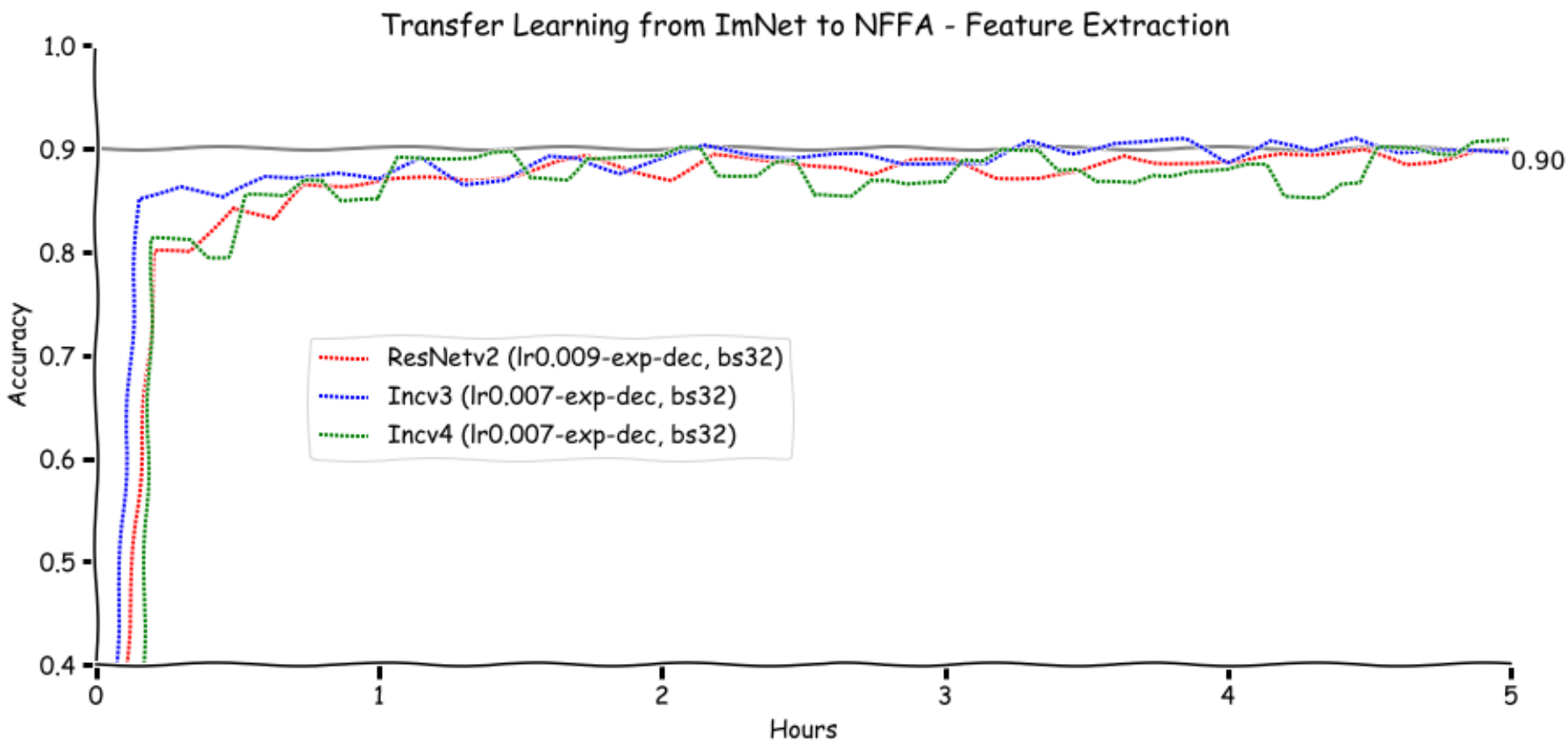
Cosilt Cloud Computing Environment


Eurotech AURORA sytem with 8 GPU nodes

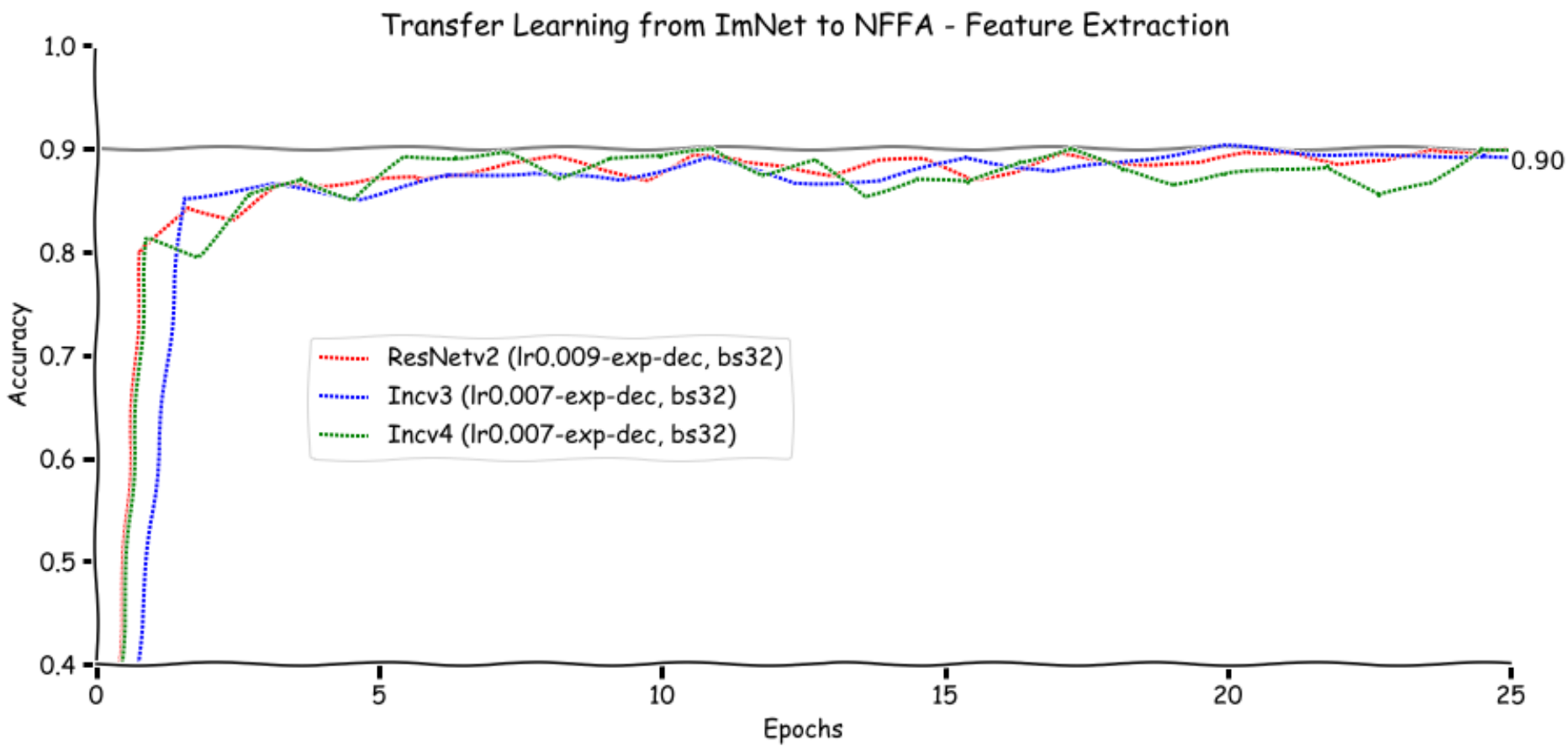
- 2 Intel Xeon E5-2697 (12 cores each)
- 2 NVIDIA K20s GPUs (2496 CUDA cores, 5GB ram)



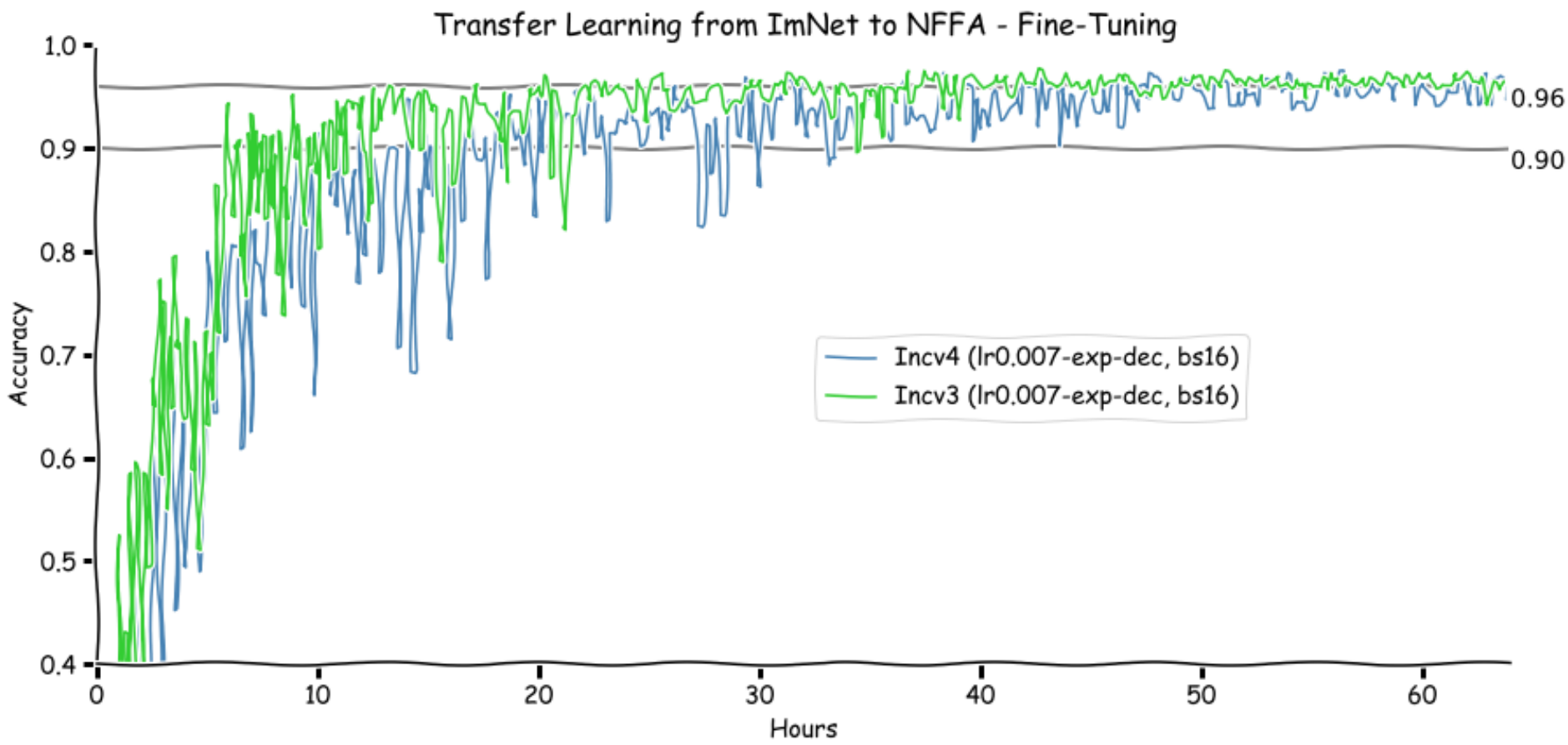
# Tran.learn.+feature extraction



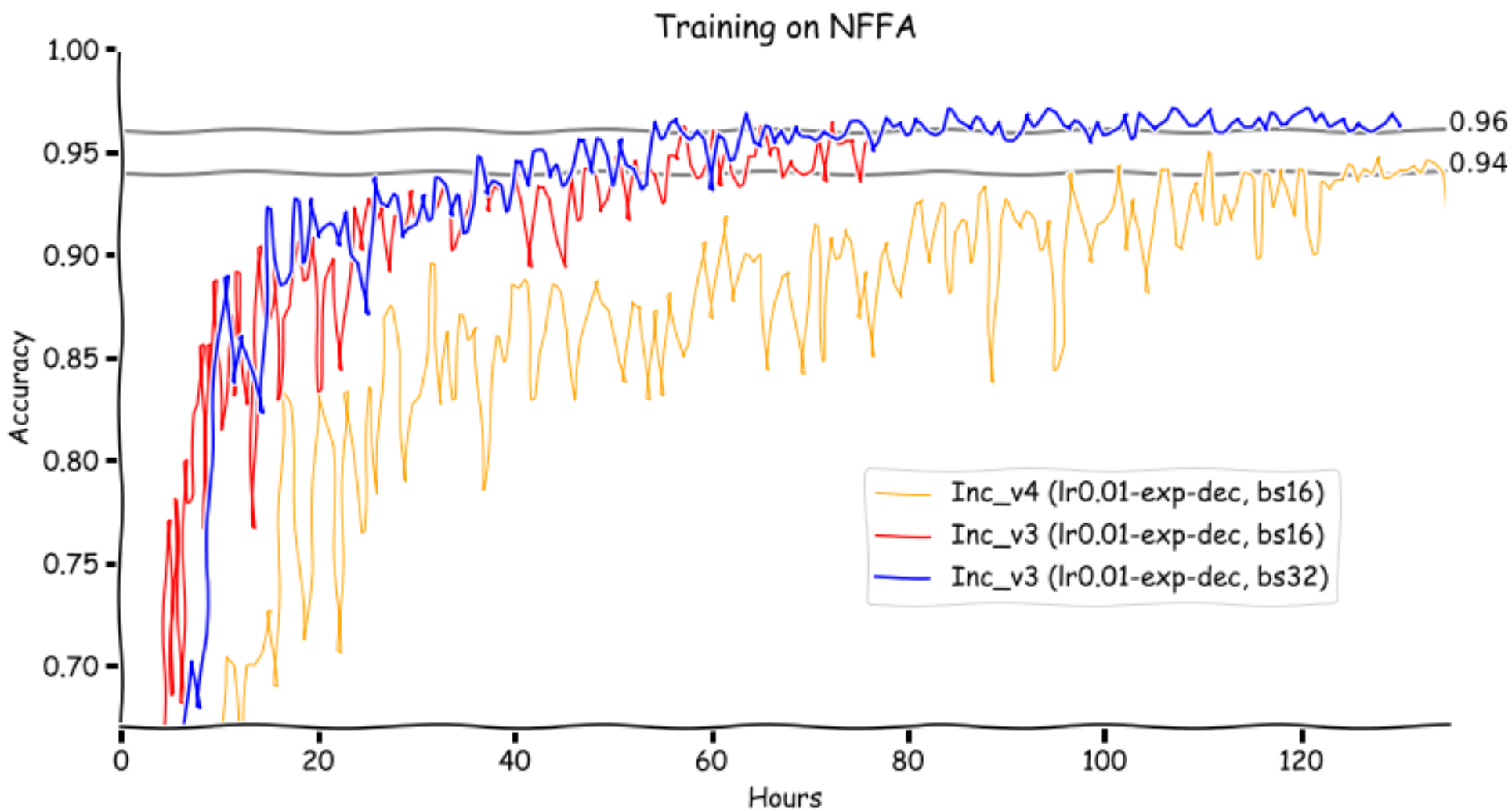
# Tran.learn.+feature extraction



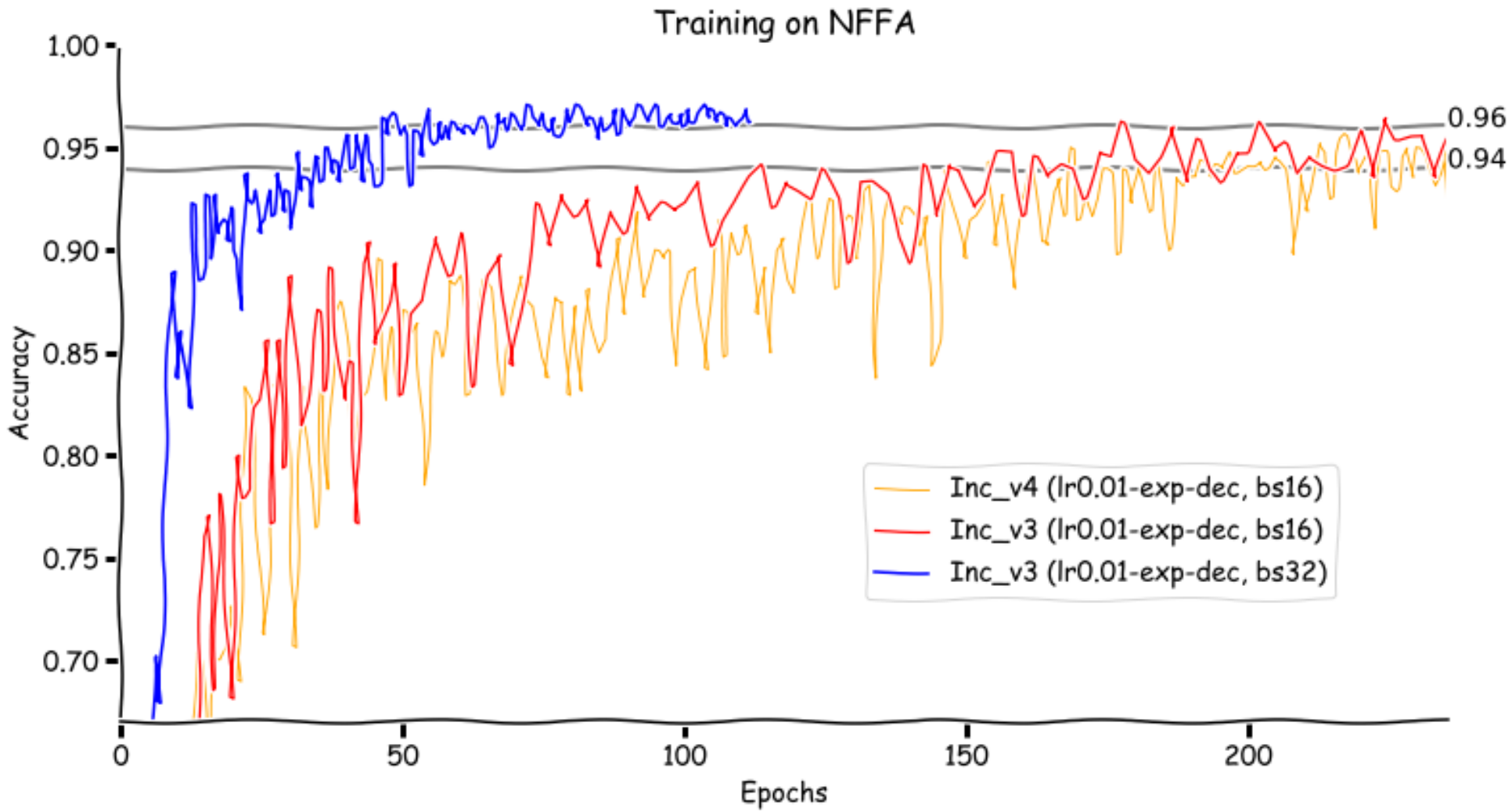
# Tran.learn.+ Fine Tuning



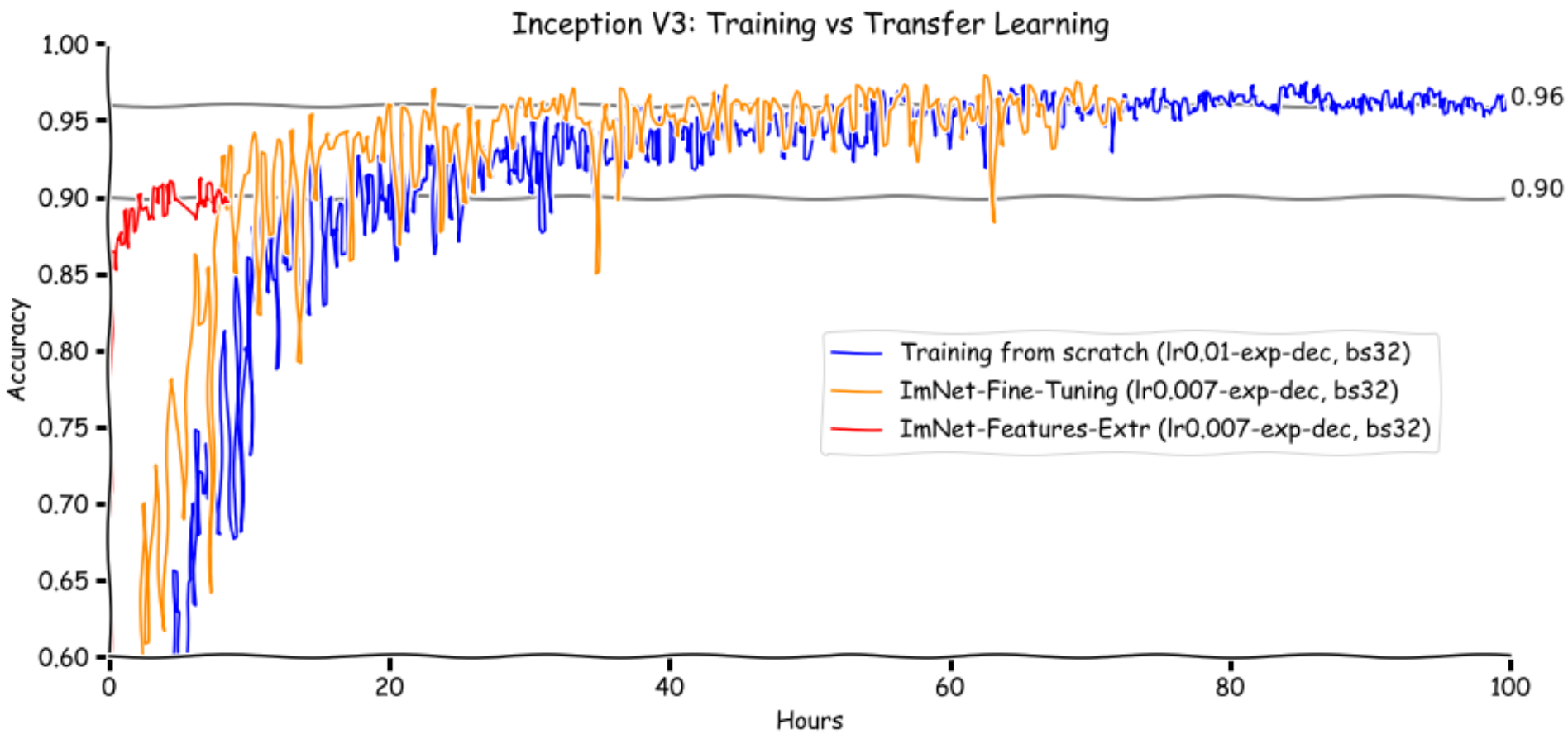
# Train from scratch



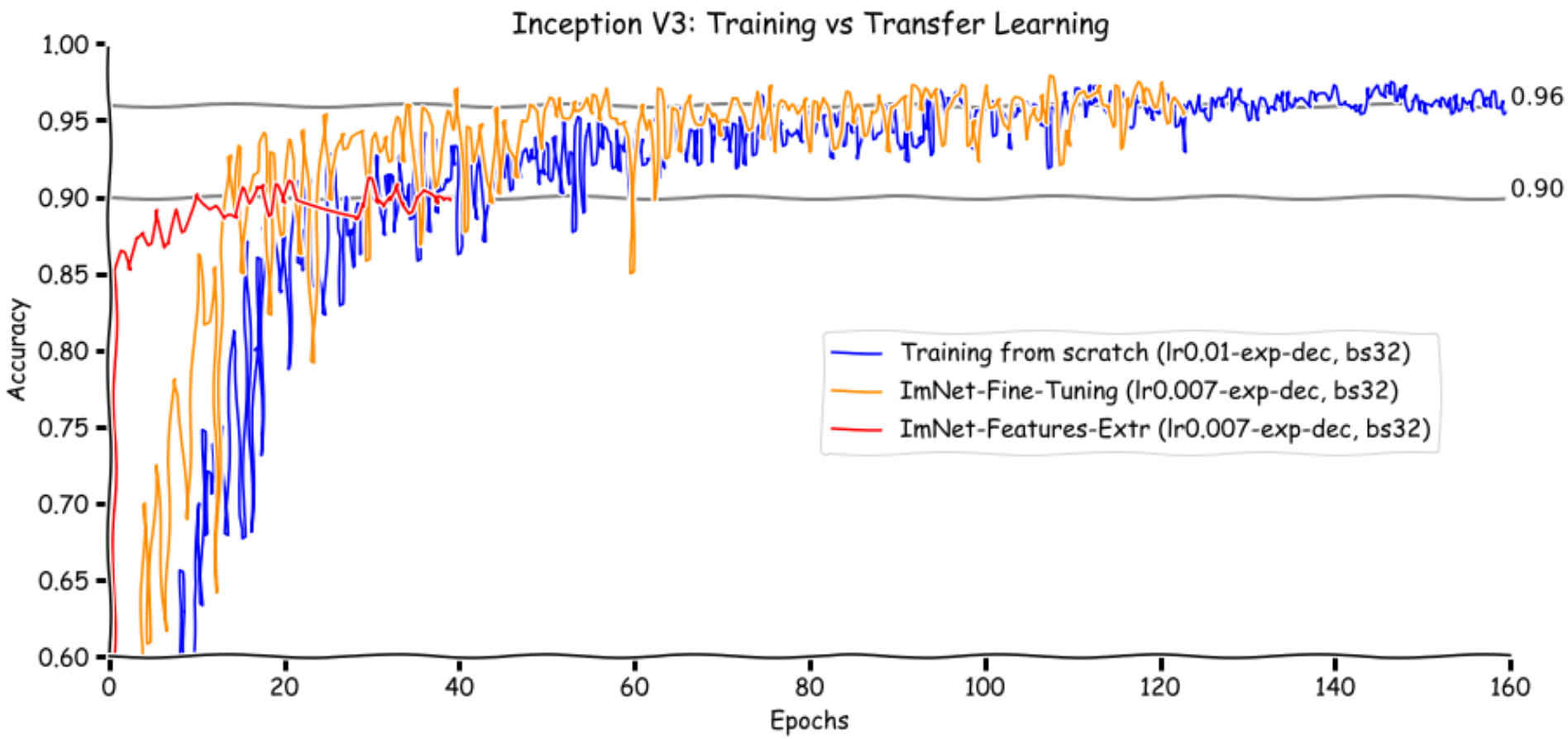
# Train from scratch



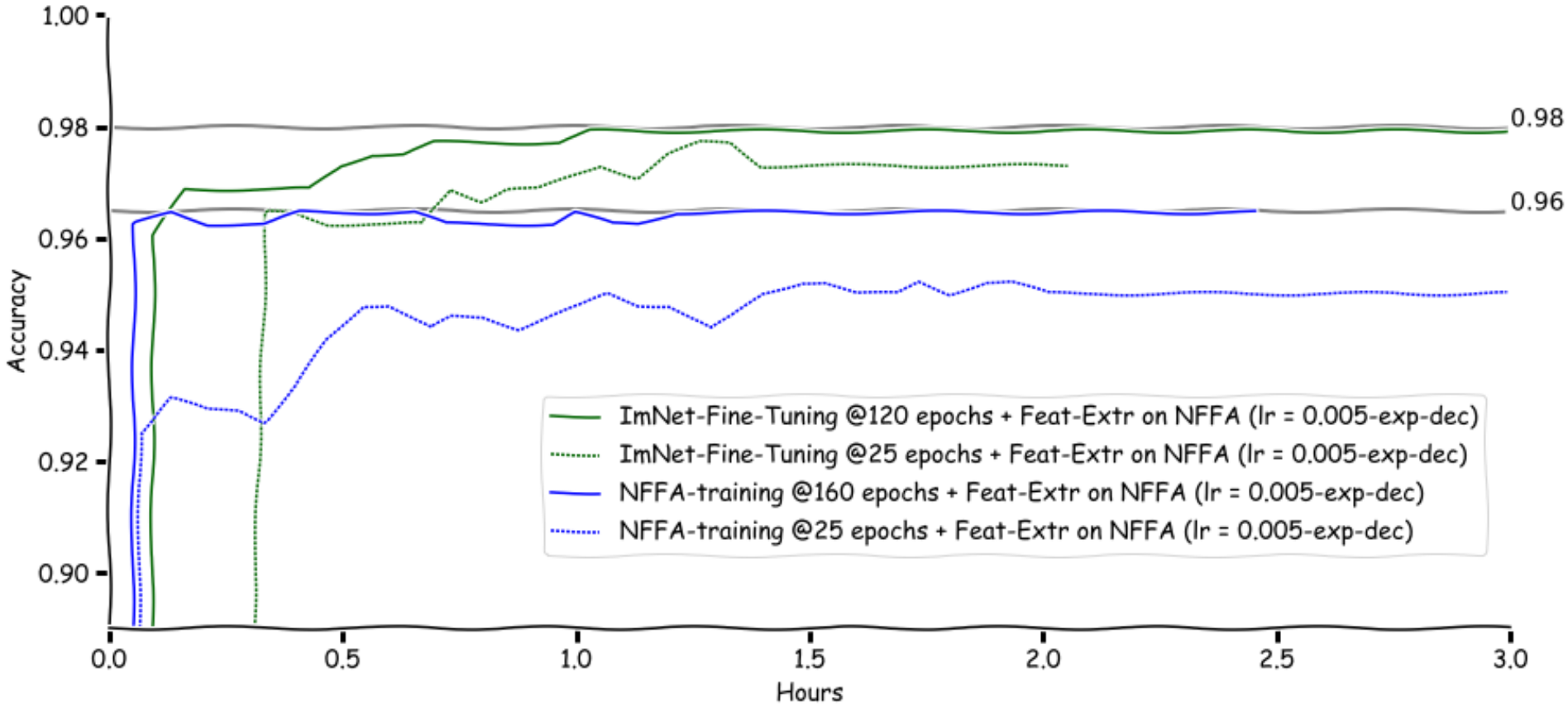
# Traning vs Transfert learning



# Traning vs Transfert learning



# Refining further

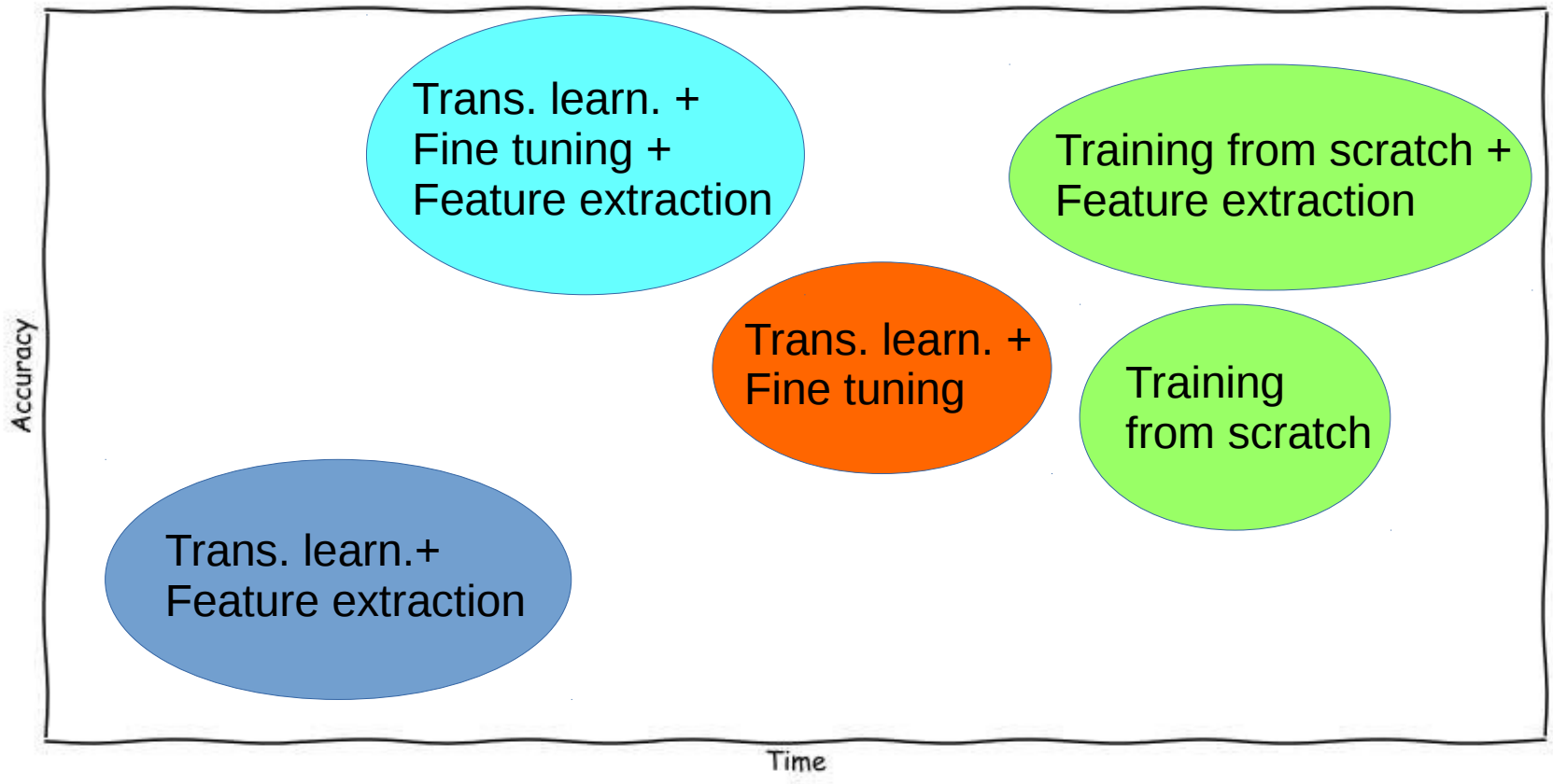



**Transfert Learning + Fine-Tuning + Feature Extraction**

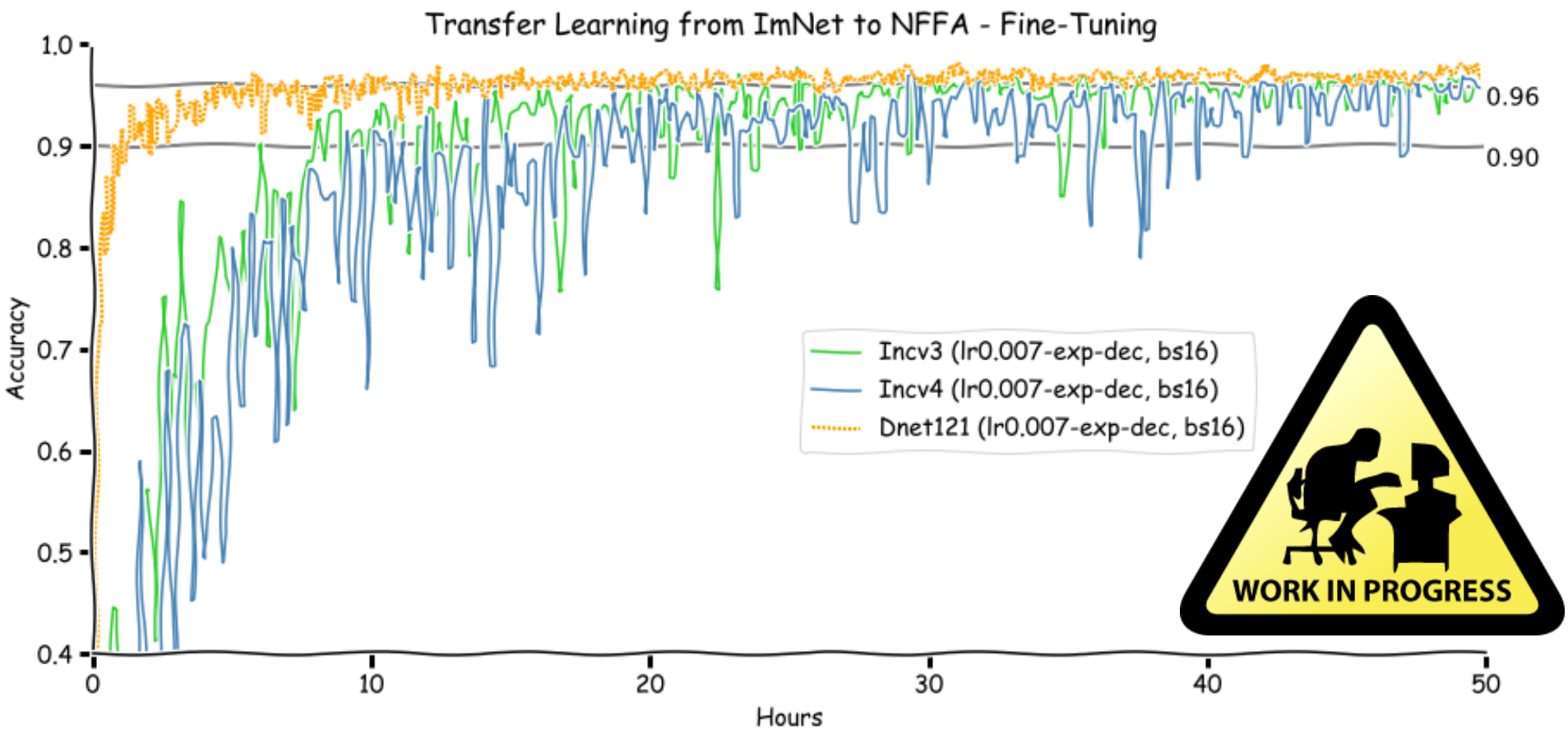
↓

**Better accuracy and faster time to solution!**

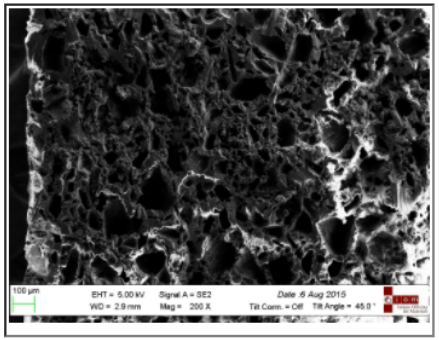
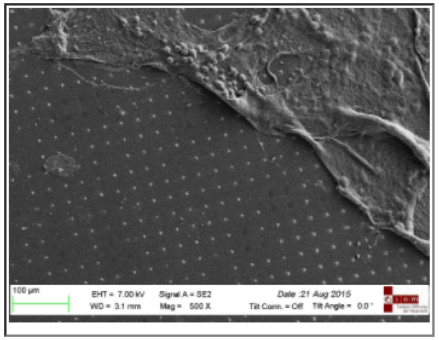
# Accuracy vs. time



# DenseNet

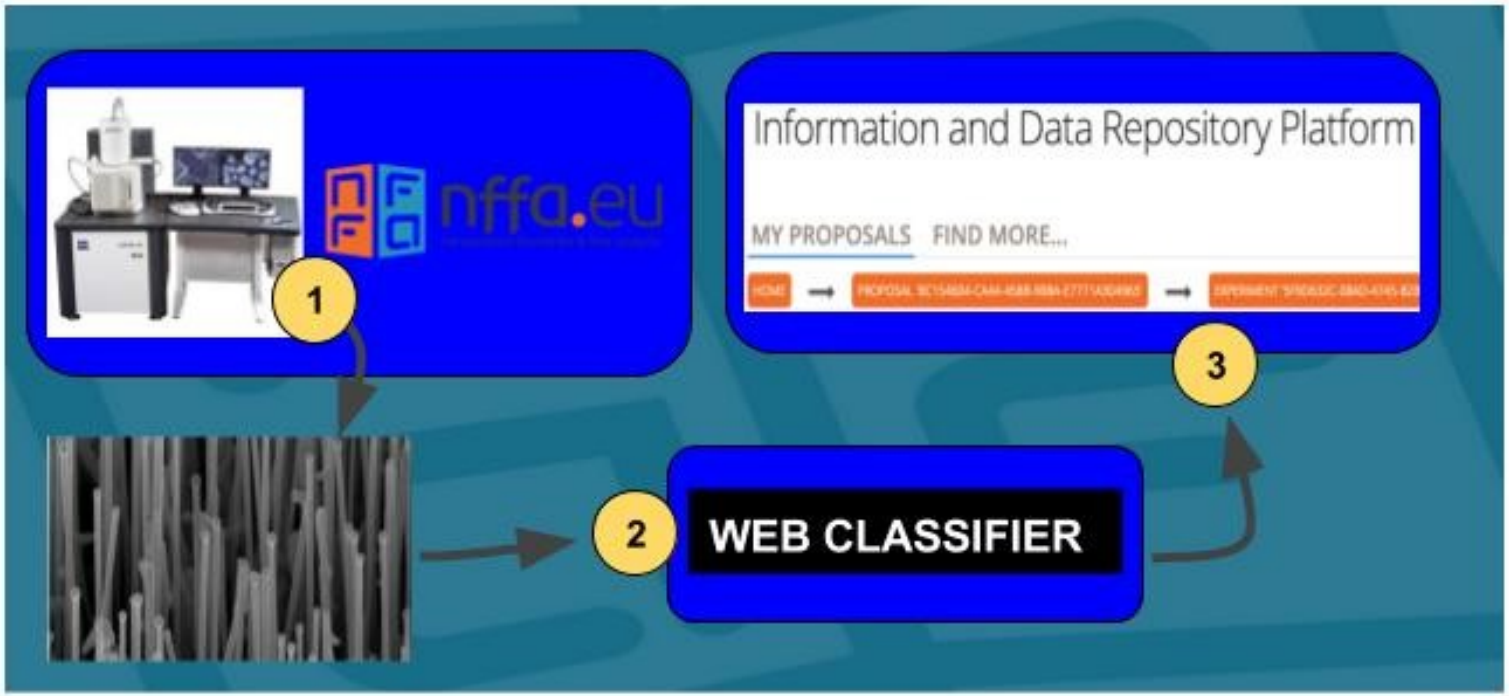


# SEM images web classifier

| Image | Processing page | Nanowires | MEMS devices electrodes | Biological | Films Coated Surface | Porous Sponge | Patterned surface |
|---|---|---|---|---|---|---|---|
|  | 237 | 0.2% | 4.1% | 0.03% | 0.9% | 93.6% Predicted | 0.31% |
|  | 238 | 00.07% | 0.00% | 99.9% Predicted | 0.00% | 0.00% | 0.00% |

Submit changes

# The complete Workflow



# Conclusions

- Very high accuracy (>96%) on a specific dataset
- Best approach (both for accuracy and time to solution) → Transf. Learning + Fine Tuning + Feature Extraction
- Automatic metadata generation
- Semi-automatic workflow for SEM scientist, which allows to further improve the traning