CS-502 Project Presentation

Agenda

- Benchmark Overview
- Datasets
- Algorithms
- Project Options
- Project Deliverables

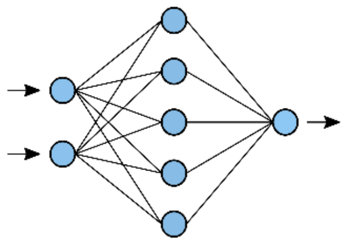
Few-Shot Bench: Overview

Datasets

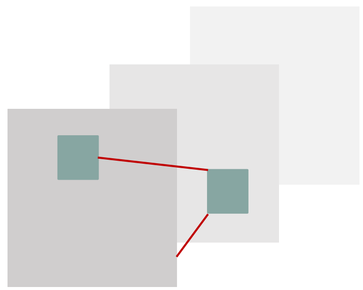
Tabula Muris
SwissProt

Backbones

Fully-connected layers



Convolutional blocks



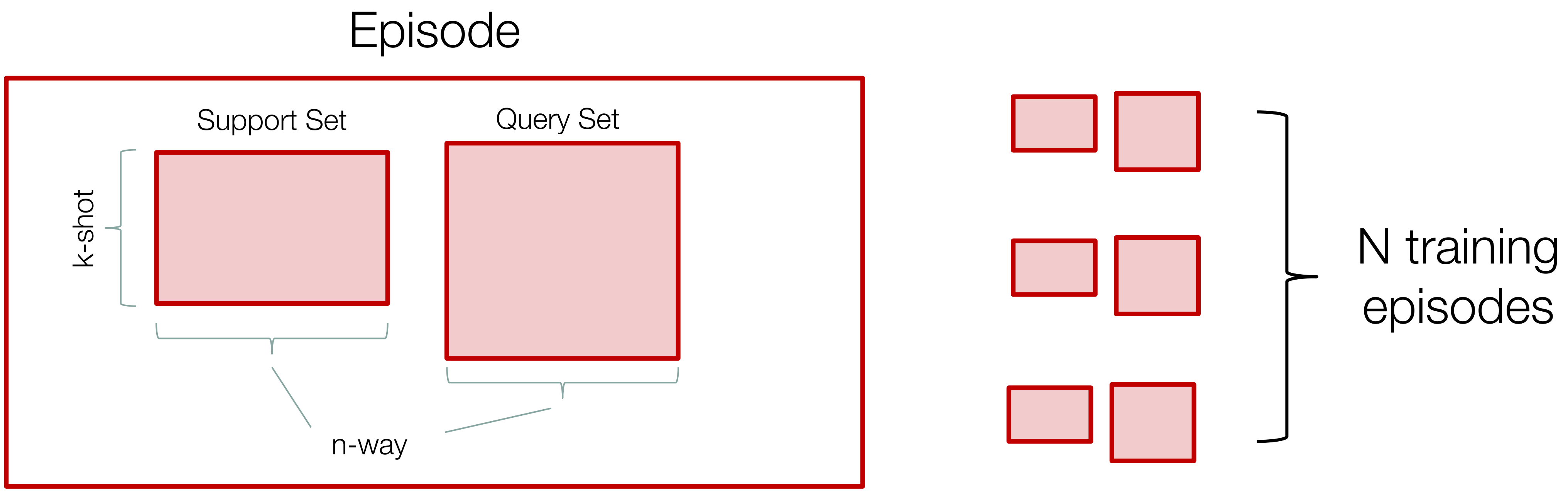
Algorithms

Baseline-Finetune (Transfer learning)
MAML
ProtoNet
MatchingNet

```
python train.py exp.name={exp_name} method=maml dataset=tabula_muris n_shot=5
```

Episodic Training for Few-Shot Learning

Key idea: Train model by mimicking few-shot regime



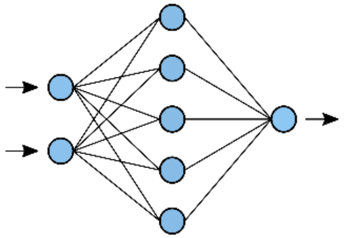
Few-Shot Bench: Datasets

Datasets

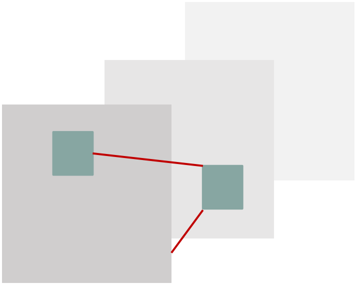
Tabula Muris
SwissProt

Backbones

Fully-connected layers



Convolutional blocks



Algorithms

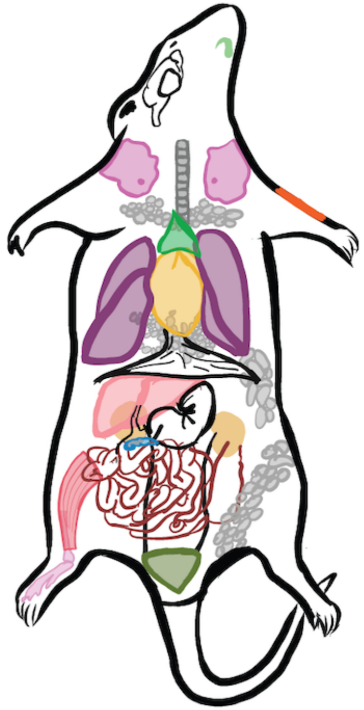
Standard NN
Baseline-Finetune (Transfer learning)
MAML
ProtoNet
MatchingNet

Two datasets:

- Tabula Muris: Cell type annotation task across tissues
- SwissProt: Gene function prediction from the sequence information

Datasets: Tabula Muris

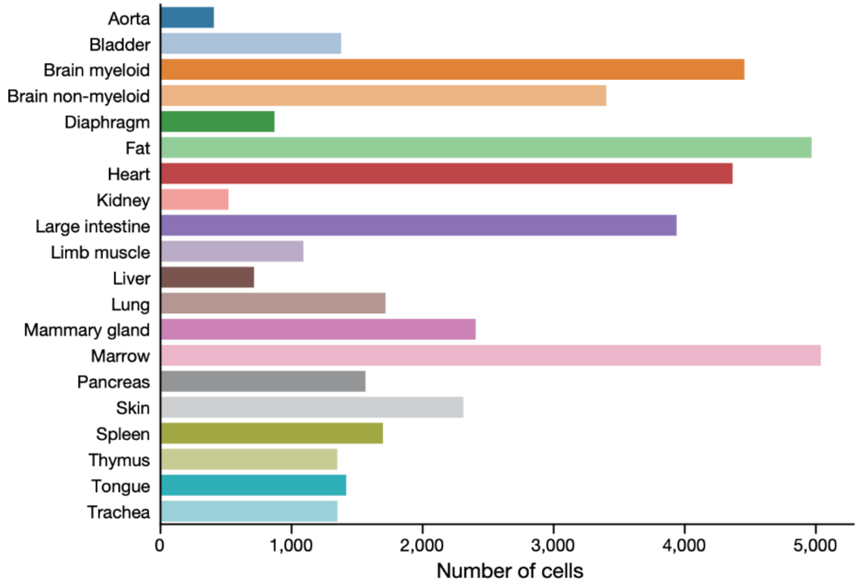
Task: few-shot cell-type annotation across tissues



Single-cell gene-expression profiles from 105,960 cells

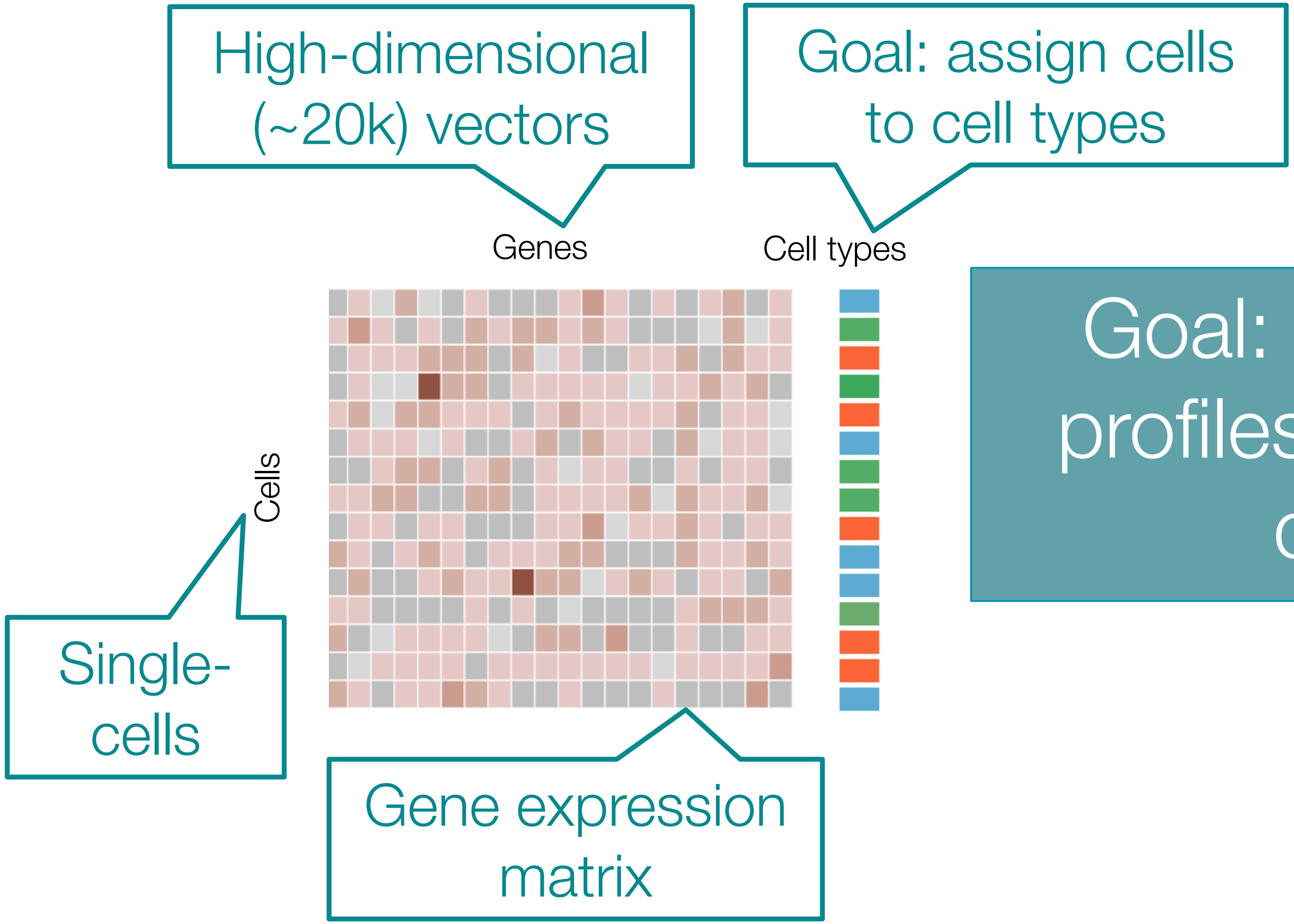
Features are expression levels of 2,866 genes

124 cell types across 23 organs of a model mouse organism



Cao et al. [Concept Learners for Few-Shot Learning.](#) *ICLR 2021.*

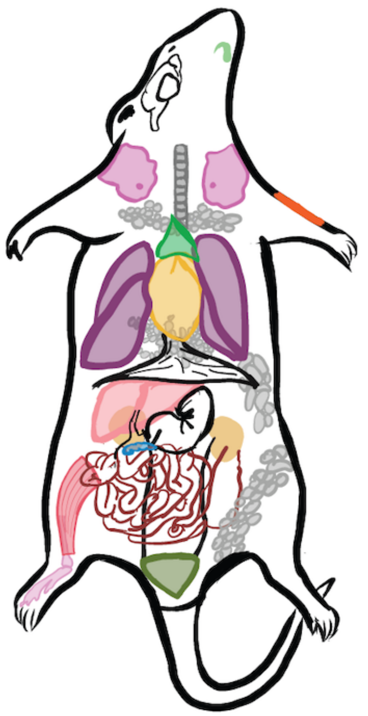
Cell Type Annotation Task



Goal: Given gene expression profiles of cells, assign cells to different cell types

Datasets: Tabula Muris

Generalize to cell types in a new tissue given only a few-labeled examples per class



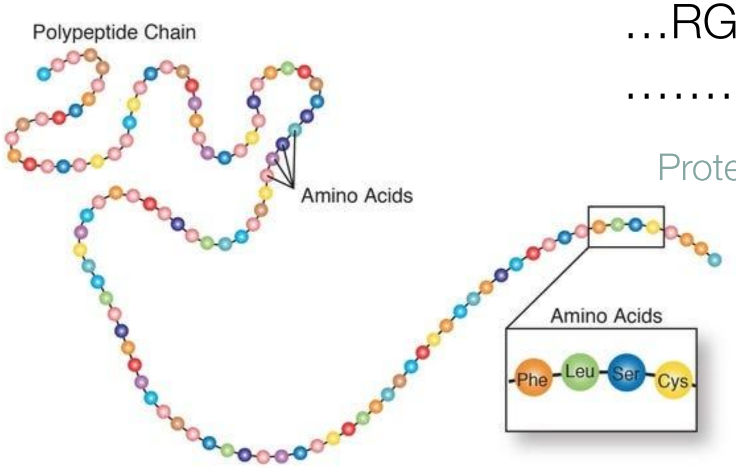
Data Splits by tissue		
Training tissues	BAT, Bladder, Brain, Diaphragm, GAT, Heart, Kidney, Limb Muscle, Liver, MAT, Mammary Gland, SCAT, Spleen, Trachea	59 cell types
Validation tissues	Skin, Lung, Thymus, Aorta	47 cell types
Test tissues	Large Intestine, Marrow, Pancreas, Tongue	37 cell types

Datasets: SwissProt

Input Data: Protein comprising of Amino Acid Sequence

.DRDFFERGSS……..
…RGSHHVAQLER…
…………AAGLFRUC

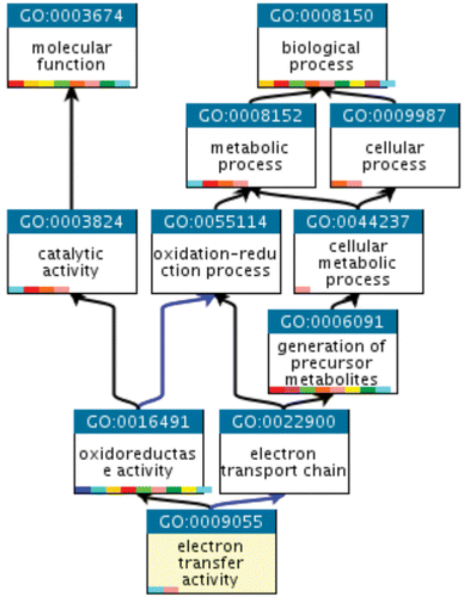
Protein Sequence Entry



To Predict: Gene Ontology Label

GO:0008150 biological process,
GO:0009987 cellular process,
GO:0001376 immune system process
GO:0023052 signaling

Gene Ontology Annotation



The Gene Ontology Consortium (GOC), an international team of experts, crafted these labels to categorize the role, location, and function of genes across various species.

The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research.*

Task: Protein Function Prediction through GO label

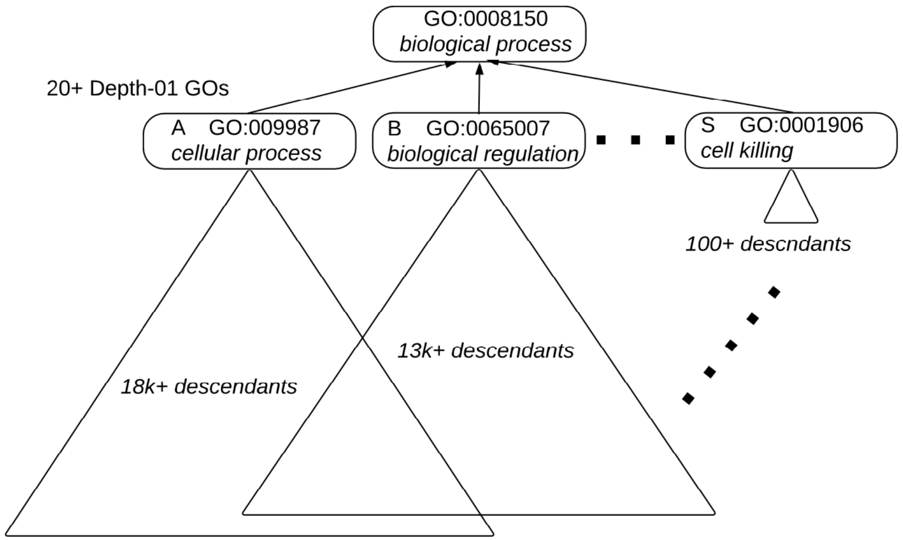
Approach and Model Architecture

Challenges (addressed)

GO labels form a DAG. Potentially many valid labels per sequence. For instance, ancestors of a given label are also valid labels.

We resolve this by choosing the most specific label. This has already been done for you.

Each sequence gets a unique label.

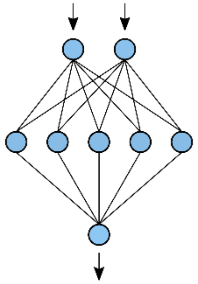


GO DAG

ESM-2

[0.025, 0.00.. ..] ..] ..]
[0.005, 0.00.. ..] ..] ..]
[0.175, 0.00.. ..] ..] ..]
[0.025, 0.00.. ..] ..] ..]
[0.050, 0.00.. ..] ..] ..]

ESM embeddings of length 1280



Fully-connected layers

GO:0019827

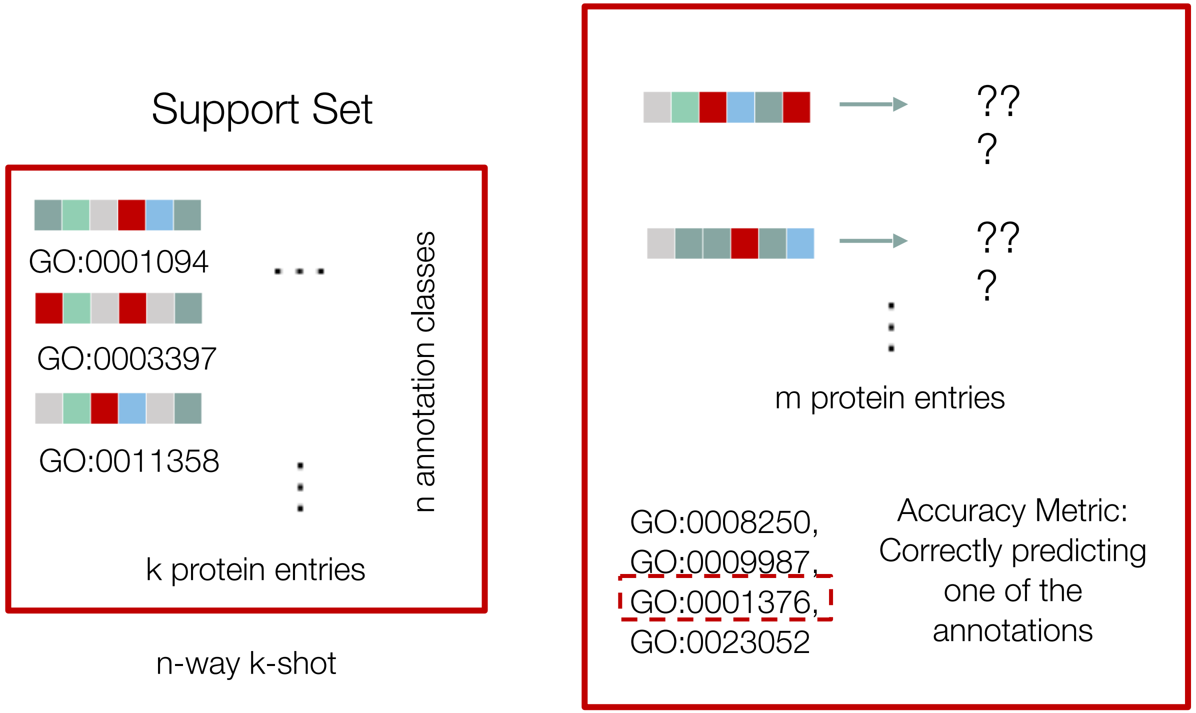
D. V. Klopfenstein et al. [GOATOOLS: A Python library for Gene Ontology analyses.](#) *Nature Scientific Reports 2018.*

Datasets: Swissprot

Sample Episode

Data Splits by GO labels

Training:	GO:0000097 GO:0000272 GO:0001676 GO:0001897	97 labels
Val:	GO:0046493 GO:0048580 GO:0048660	25 labels
Testing:	GO:0075512 GO:0090257	14 labels



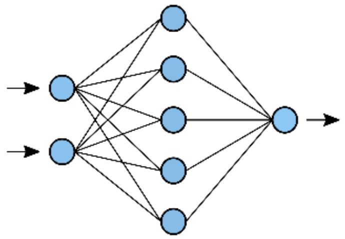
Few-Shot Bench: Backbones

Datasets

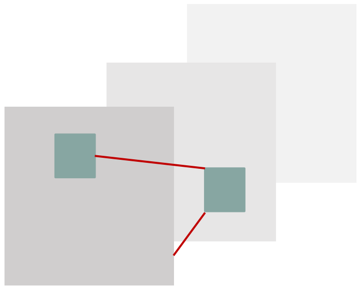
Tabula Muris
SwissProt

Backbones

Fully-connected layers



Convolutional blocks



Algorithms

Standard NN
Baseline-Finetune (Transfer learning)
MAML
ProtoNet
MatchingNet

Currently supported backbones:

- Fully-connected networks
- Convolutional ResNets
- You can also implement your own backbone!

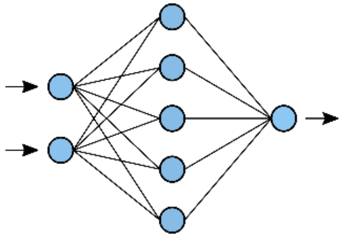
Few-Shot Bench: Algorithms

Datasets

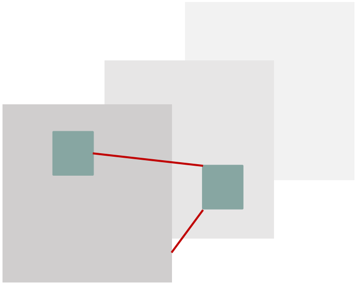
Tabula Muris
SwissProt

Backbones

Fully-connected layers



Convolutional blocks



Algorithms

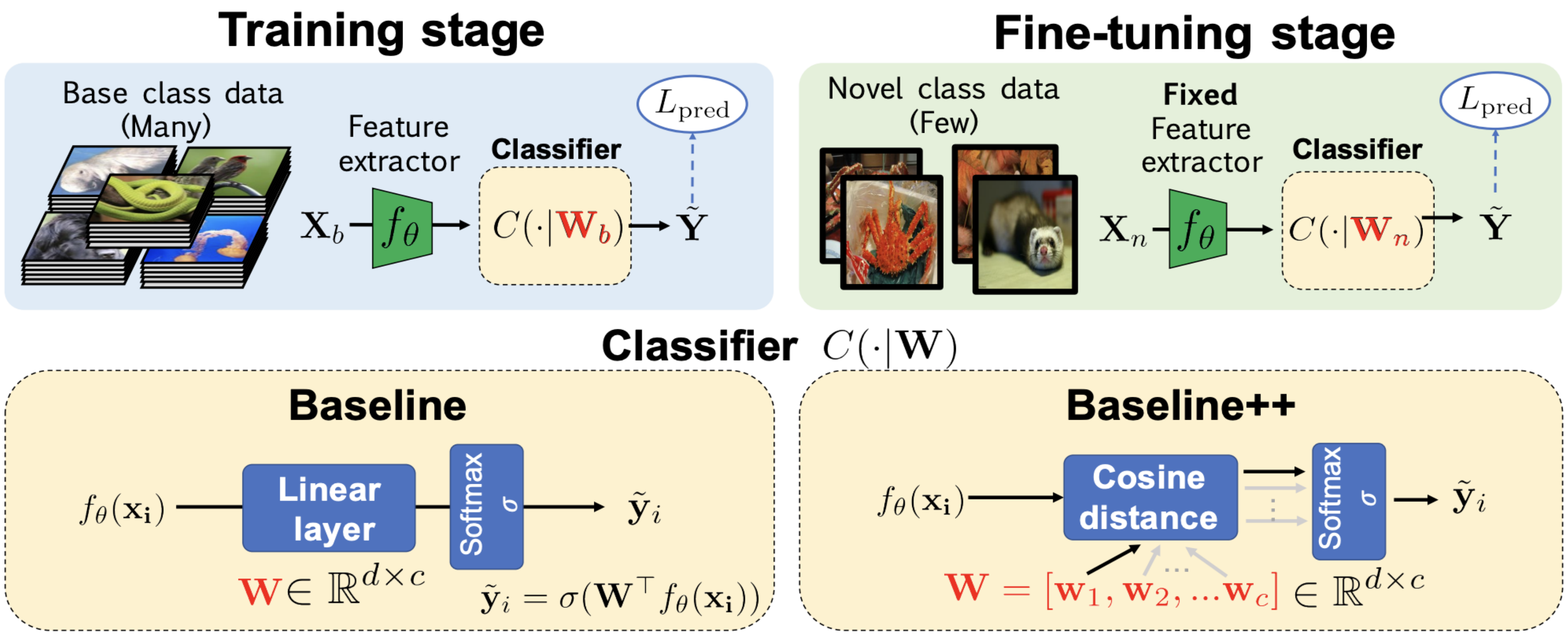
Standard NN
Baseline-Finetune (Transfer learning)
MAML
ProtoNet
MatchingNet

Five algorithms:

- Standard NN, baseline-finetune, MAML, ProtoNet, MatchingNet
- Most algorithms will be covered in the lecture!

Algorithms: Baseline and Baseline++

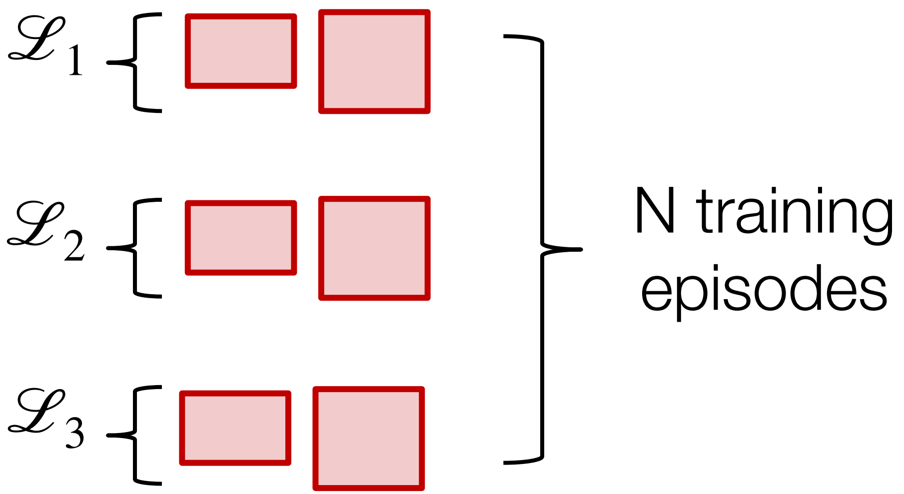
Key idea: Pretrain feature extractor and perform transfer

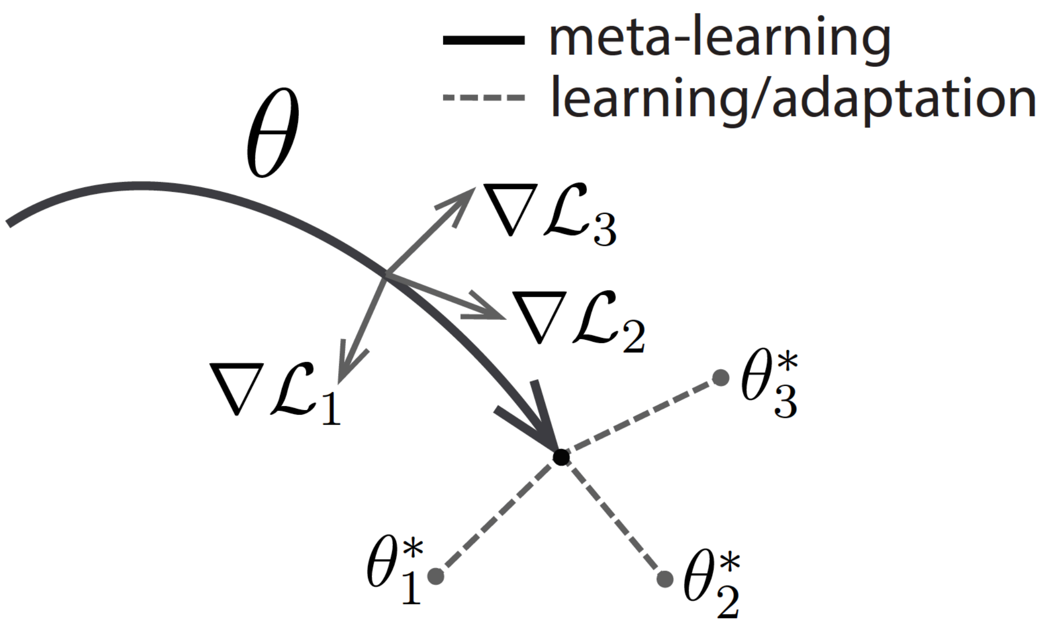


Chen et. al. A Closer Look at Few-Shot Classification. *ICLR 2019.*

Algorithms: MAML

Key idea: Learn good initialization for fast adaptation by SGD

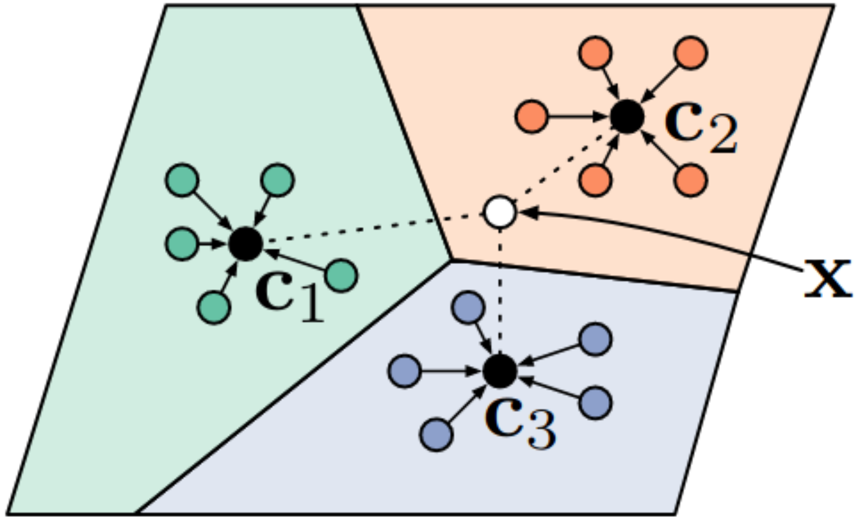




Finn et. al. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *ICML 2017.*

Algorithms: ProtoNet

Key idea: Assign query points to nearest class prototypes



Construct class prototypes using support set

$$c_k = \frac{1}{|S_k|} \sum_{(x_s,\ y_s) \in S_k} f_\theta(x_s)$$

Assign query point to the nearest prototype

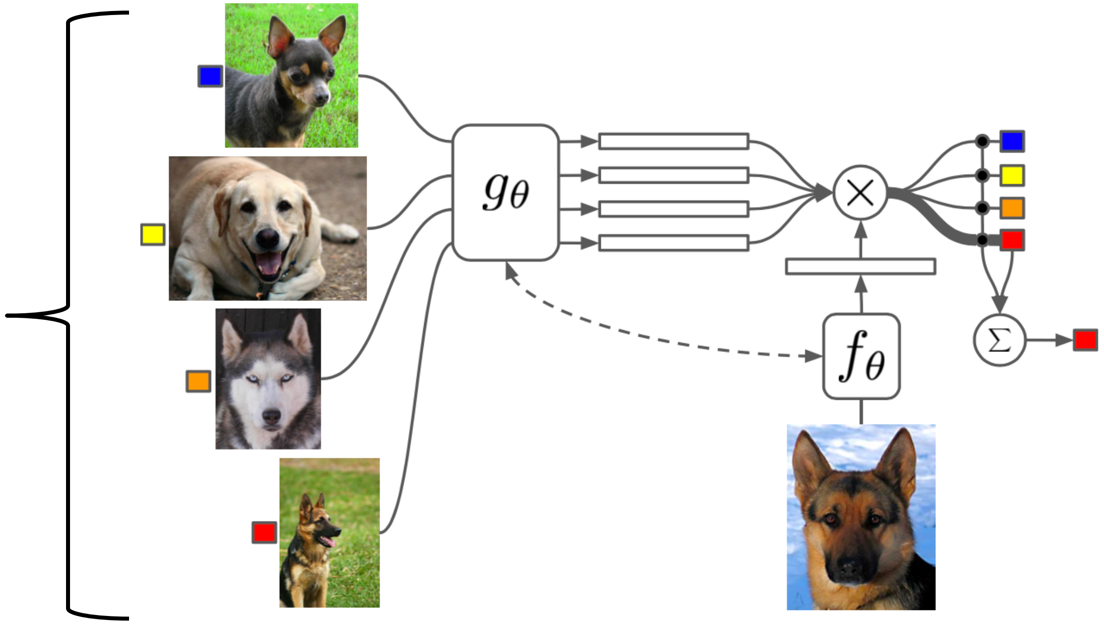
$$\mathrm{p}(y = k \mid x) \propto \exp(-d(x,\ c_k))$$

Snell et al. Prototypical Networks for Few-shot Learning. *NeurIPS 2017.*

Algorithms: MatchingNet

Key idea: Prediction on query point is the weighted average of support labels

Support examples



Query point

$$y_q = \sum_{(x_s,\ y_s) \in S} a\left(x_q;\ x_s\right) y_s$$

Vinyals et. al. [Matching Networks for One-shot Learning.](#) *NeurIPS 2016.*

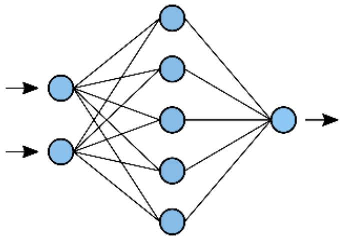
Project option #1: New dataset of your choice

Datasets

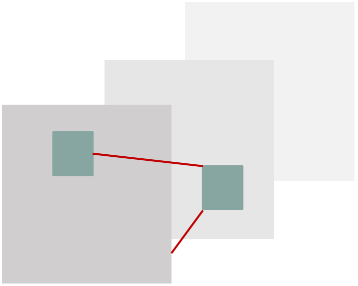
Tabula Muris
SwissProt

Backbones

Fully-connected layers



Convolutional blocks



Algorithms

Standard NN
Baseline-Finetune (Transfer learning)
MAML
ProtoNet
MatchingNet

Extend the benchmark with a new bio dataset:

- Motivate the few-shot learning task
- Provide meaningful train/val/test splits
- Evaluate and study the performance of existing algorithms in the benchmark on a new dataset

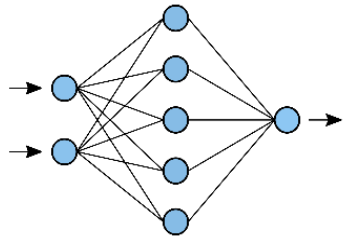
Project option #2: New algorithm of your choice

Datasets

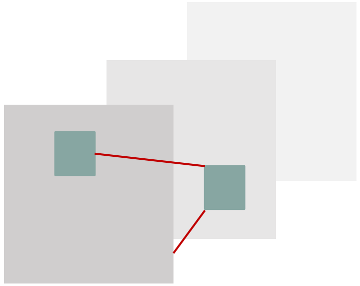
Tabula Muris
SwissProt

Backbones

Fully-connected layers



Convolutional blocks



Algorithms

Standard NN
Baseline-Finetune (Transfer learning)
MAML
ProtoNet
MatchingNet

Extend the benchmark with a new few-shot learning algorithm:

- Implement a new algorithm of your choice
- Compare it to the existing algorithms in the benchmark
- Ablate the components of the implemented algorithm

Project option #2: New algorithm of your choice

Possible options to implement:

- Lee et al. Meta-Learning with Differentiable Convex Optimization. *CVPR 2019.*
- Sung et al. Learning to Compare: Relation Network for Few-Shot Learning. *CVPR 2018.*
- Bertinetto et al. Meta-learning with differentiable closed-form solvers. *ICLR 2019.*
- Oreshkin et al. Task dependent adaptive metric for improved few-shot learning. *NeurIPS 2018.*
- Hu et al. Leveraging the Feature Distribution in Transfer-based Few-Shot Learning. *ICANN 2021.*
- Shalam et al. The Self-Optimal-Transport Feature Transform. *Arxiv 2022.*
- Mishra et al. A simple neural attentive meta-learner. *ICLR 2018.*
- and many more…

Check PapersWithCode
for SOTA methods!

https://paperswithcode.com/task/few-shot-image-classification

Project option #3: Research in the wild

You can pick any biomedical task of your choice and try to solve it using the deep learning method of your choice!

- It does not need to be a few-shot learning task
- Motivate the setting and problem well
- Apply any architectures/approaches you want

General advice to start with the FewShotBench

- Start with running existing algorithms on SwissProt
  - It is cheaper to run!
- If you decide to work on the new dataset that is graph structured, you can implement other backbones (i.e. GNNs).
  - Use PyTorch Geometric!

General Advice

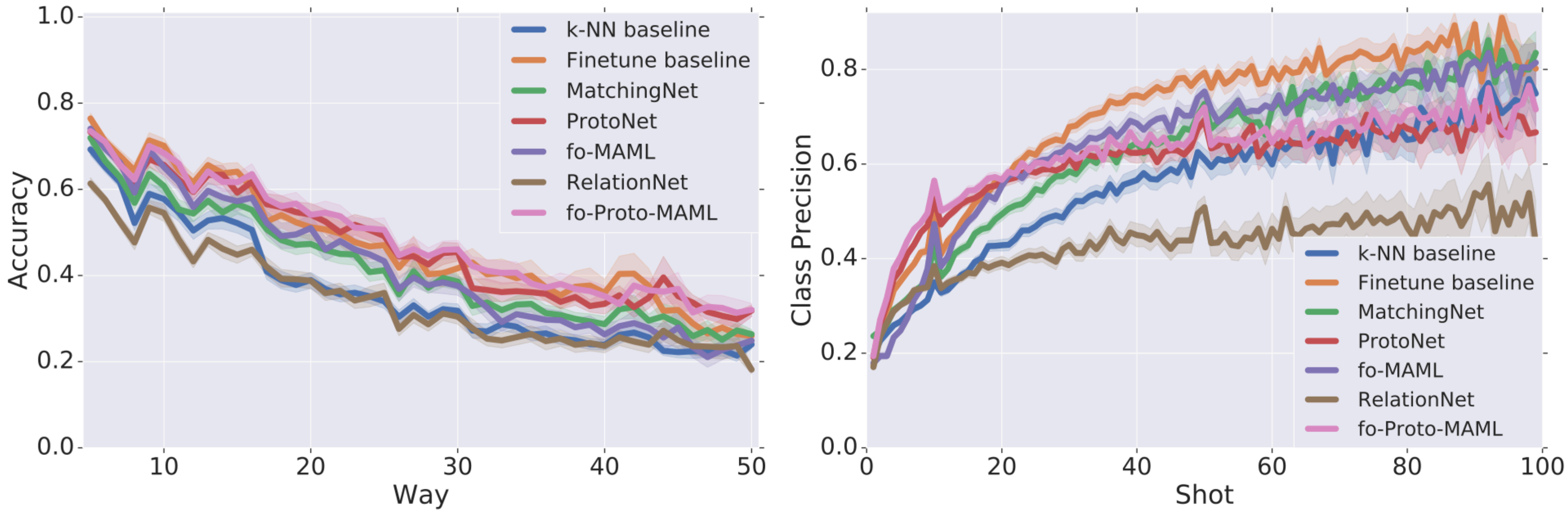
- In case you are running out of Google credits:
  - You can add your credit card in the account and Google will give 200$ more without crediting you!

Project Deliverables

4 page
project report

GitHub repo with
- implementations
- data loading (Option 1, 3)
- instructions to run

Ablations example for the report: MetaDataset



(a) Ways analysis (b) Shots analysis

Triantafillou et. al. [Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples.]() *ICLR 2020.*