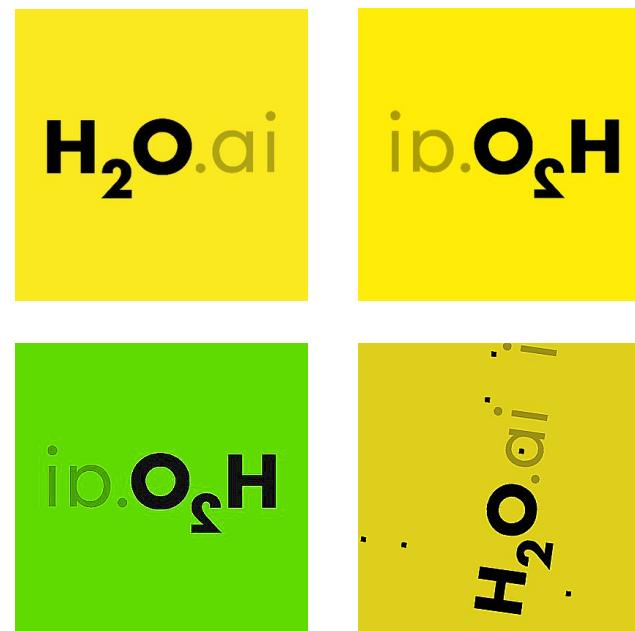


General Pipeline for Computer Vision Problems

Yauhen Babakhin, Data Scientist @ H2O.ai



About me



b.e.s.

Data Scientist at H2O.ai
Minsk, Belarus
Joined 5 years ago · last seen in the past day

[Profile](#) [Edit Profile](#)

[Home](#) [Competitions \(17\)](#) [Kernels \(1\)](#) [Discussion \(39\)](#) [Datasets](#) [...](#)

Competitions Grandmaster 

Current Rank	Highest Rank
90 of 118,043	26

 5	 2	 5
---	---	---

[TGS Salt Identification Cha...](#)  1st  a year ago · Top 1% of 3229

[Quick, Draw! Doodle Recog...](#)  4th  9 months ago · Top 1% of 1316

[Airbus Ship Detection Chal...](#)  6th  10 months ago · Top 1% of 882

Kernels Contributor 

Discussion Contributor 

Unranked

 0	 0	 1
---	---	---

[Products Clustering](#)  17  3 years ago

[1st Place Solution with Code](#)  370  votes · a year ago

[6th Place Solution \(41st in t...\)](#)  47  votes · 10 months ago

[12th Place Approach](#)  29  2 years ago

H2O.ai

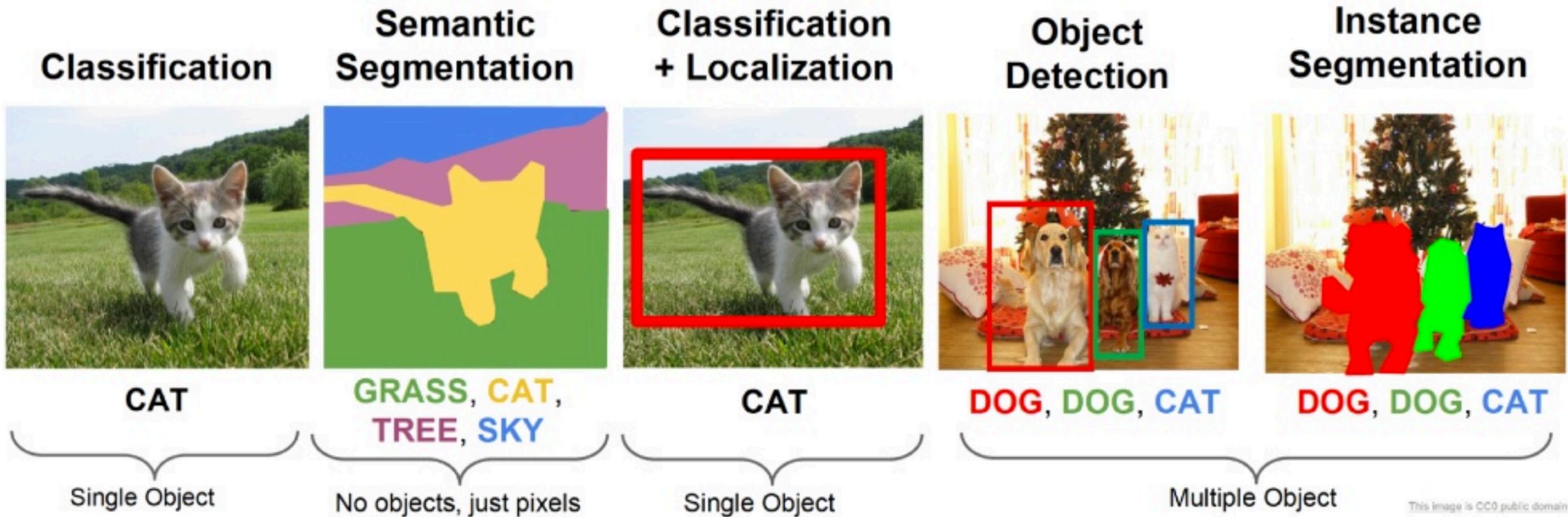
Contents

Prepare the data

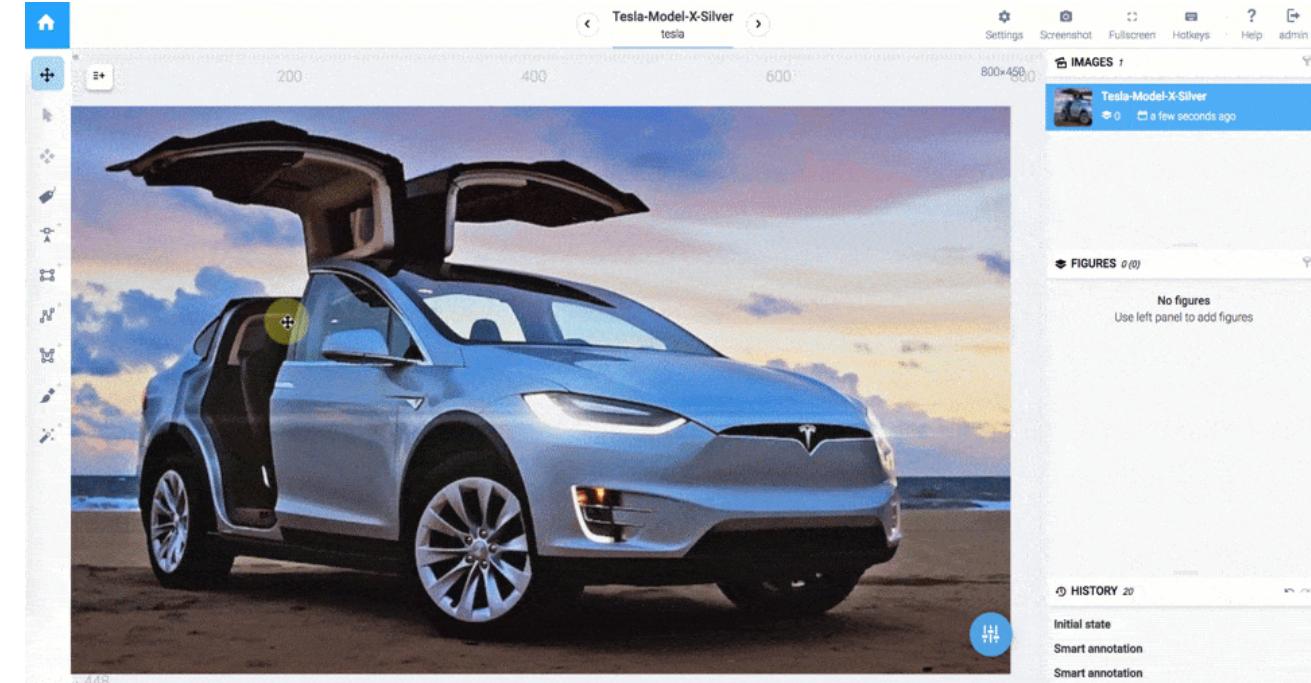
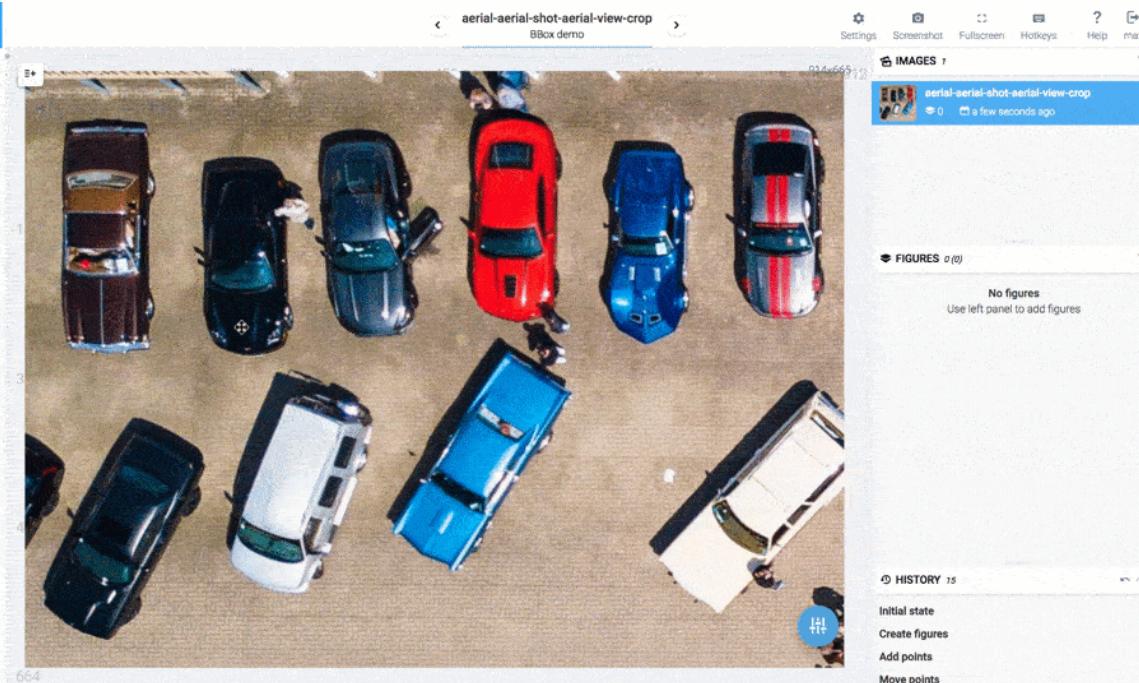
Initialize the
training

Train the model

Define the problem

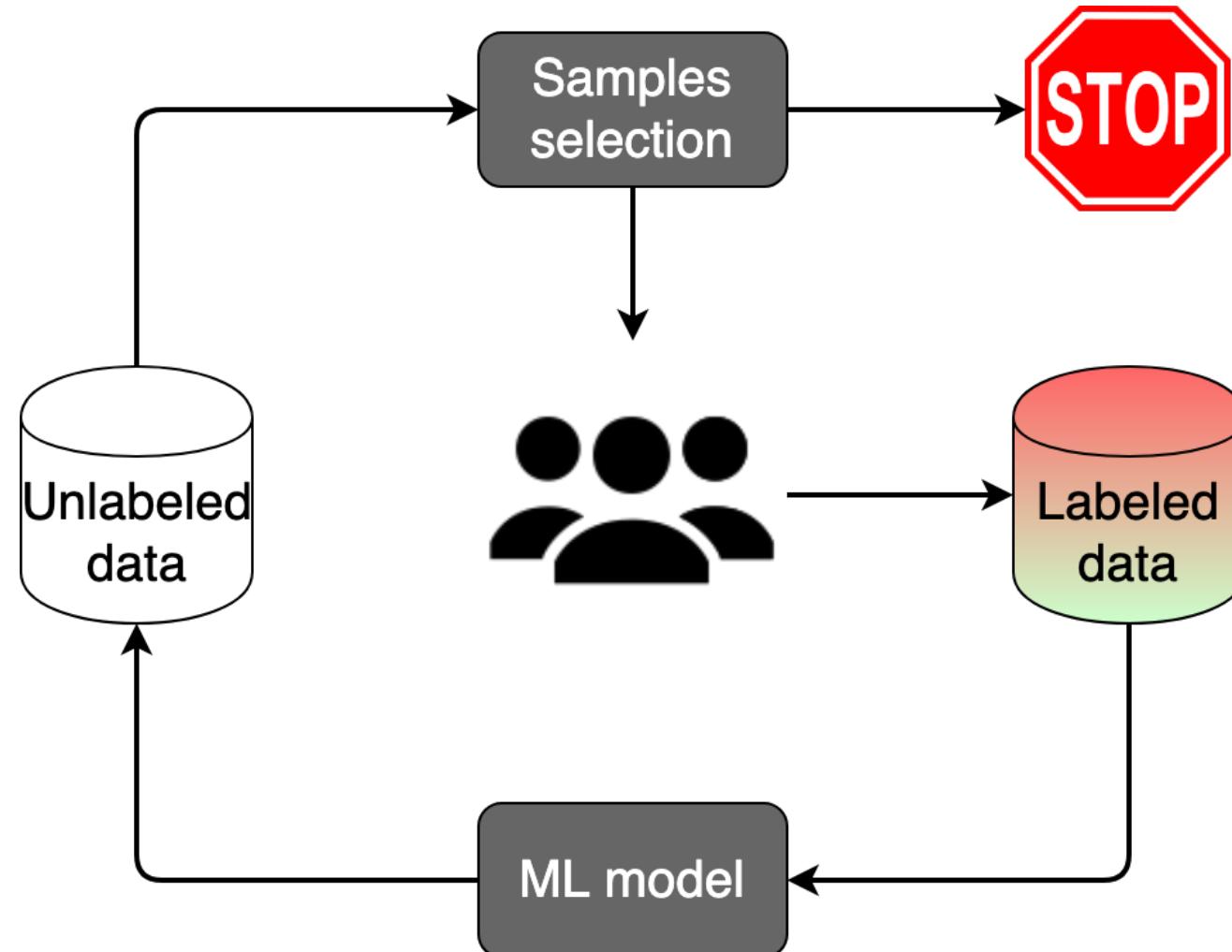


Label the data

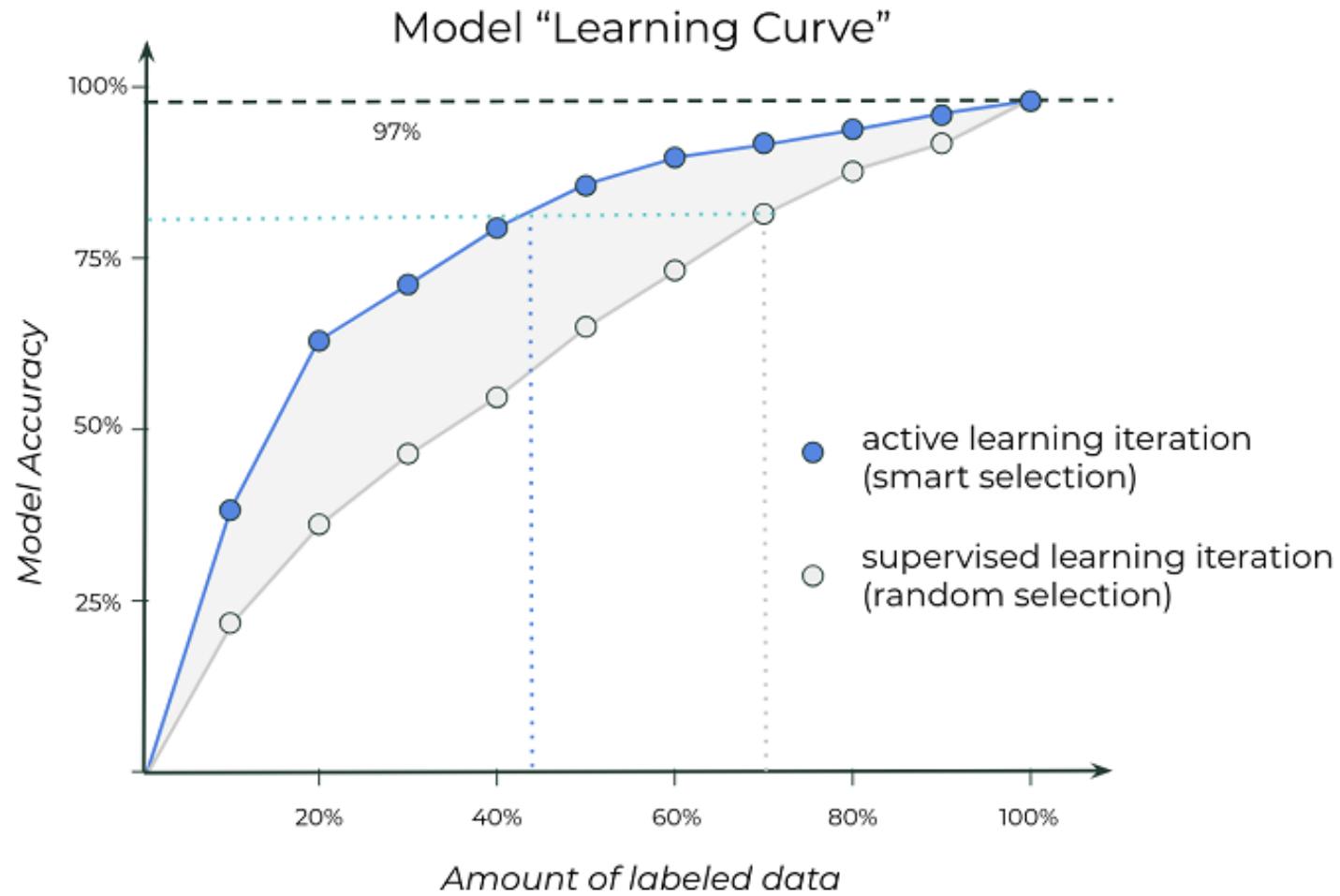


<https://hackernoon.com/%EF%B8%8F-advanced-annotation-tools-in-deep-learning-training-data-for-computer-vision-with-supervisedly-847f8699a9cb>

Active learning + Human-in-the-loop

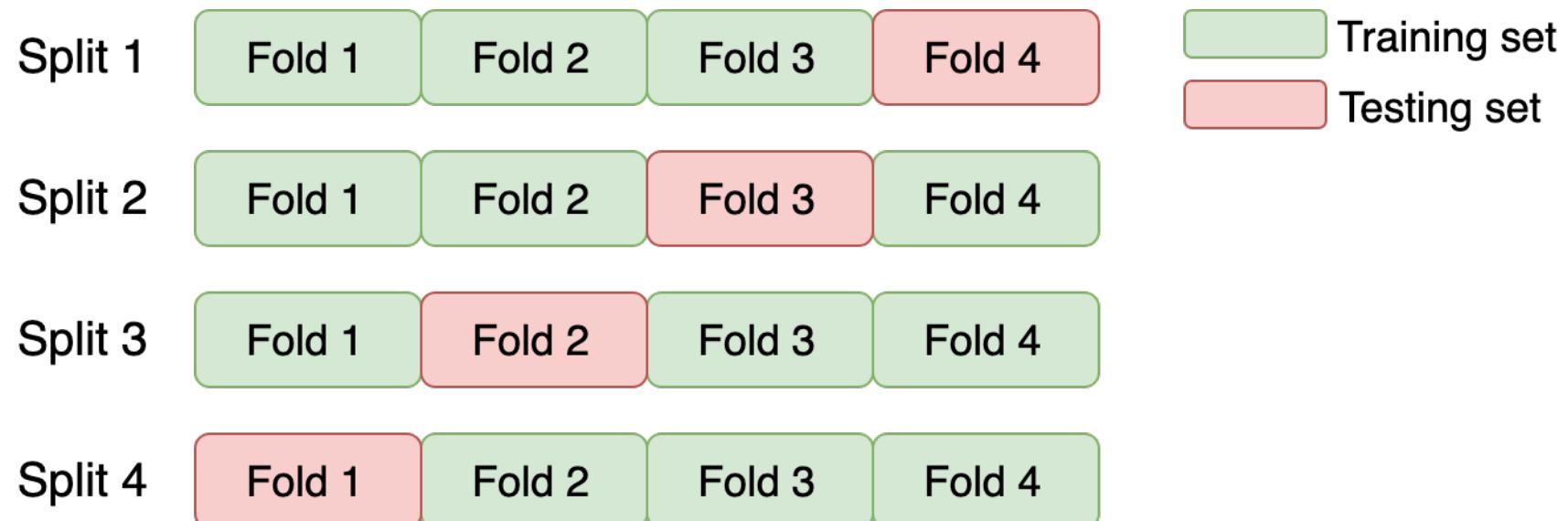


Active learning advantage



Validation strategy

- Evaluation metric
- Simple holdout vs. cross-validation



Data is ready-to-use

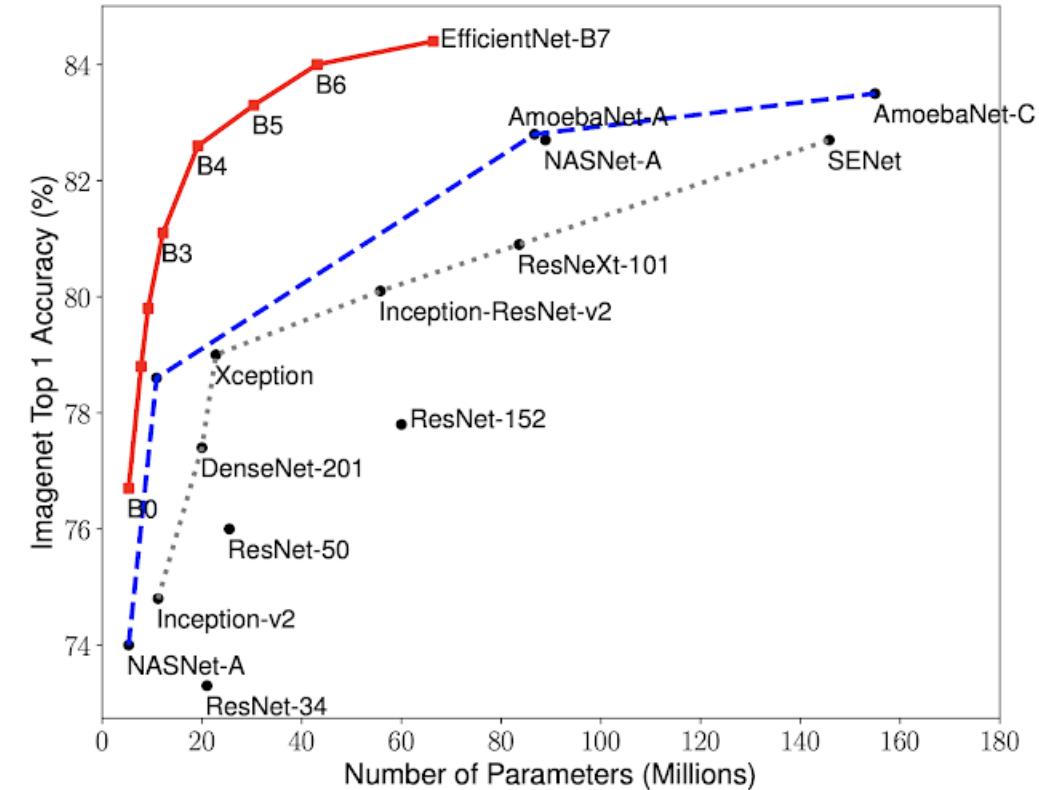
Prepare the data

Initialize the
training

Train the model

Neural net architecture

- Custom architecture
- Famous architectures zoo
- Neural Architecture Search (NAS)



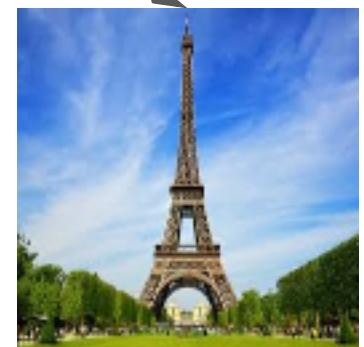
Weights initialization

- Random initialization
- ImageNet pretrained



Cropping strategies

- Resize
- Random crop
- Resize → Random crop
- Random crop → Resize



Augmentations



MixUp augmentations

	ResNet-50	Mixup [48]	Cutout [3]	CutMix
Image				
Label	Dog 1.0	Dog 0.5 Cat 0.5	Dog 1.0	Dog 0.6 Cat 0.4
ImageNet	76.3	77.4	77.1	78.6
Cls (%)	(+0.0)	(+1.1)	(+0.8)	(+2.3)

Before we start training

- Optimizer: Adam (AdamW, RAdam), SGD, RMSProp

Before we start training

- Optimizer: Adam (AdamW, RAdam), SGD, RMSProp
- Loss (differs from the metric)

Before we start training

- Optimizer: Adam (AdamW, RAdam), SGD, RMSProp
- Loss (differs from the metric)
- Batch size + initial input shape

Training is initialized

Prepare the data

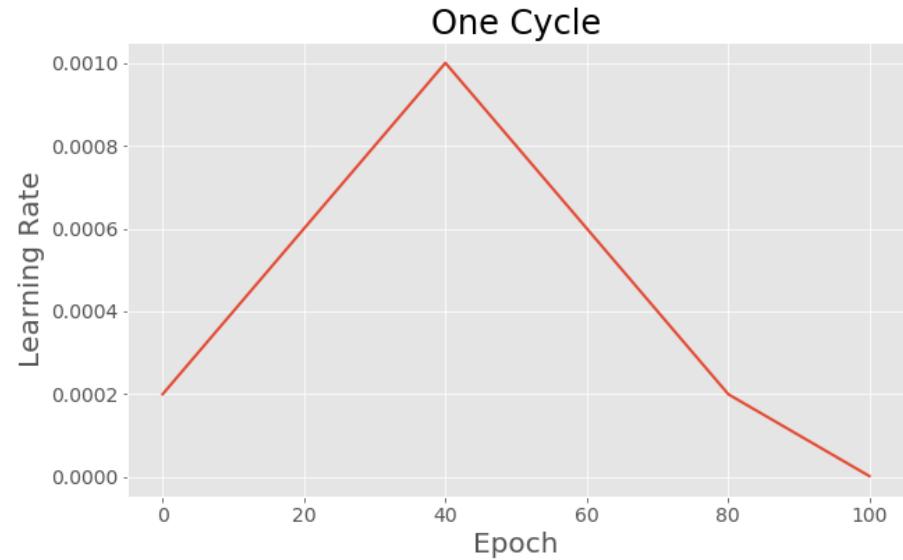
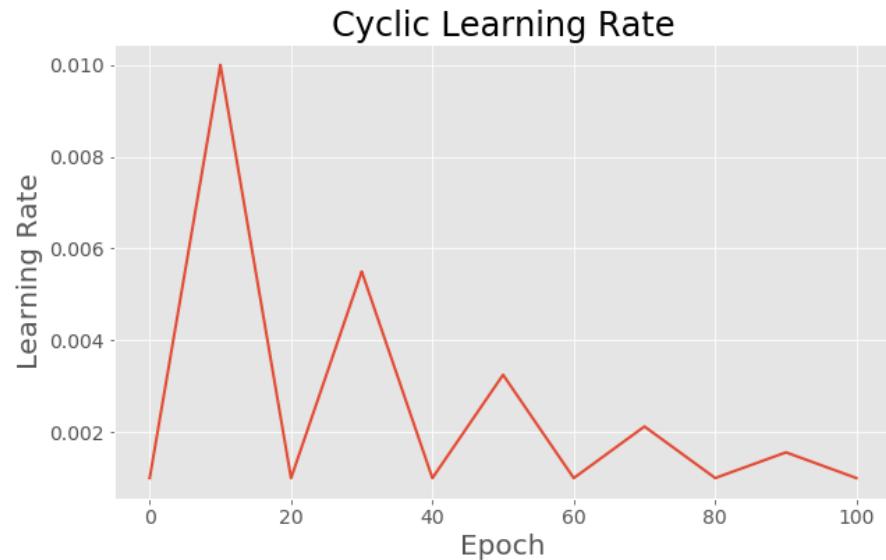
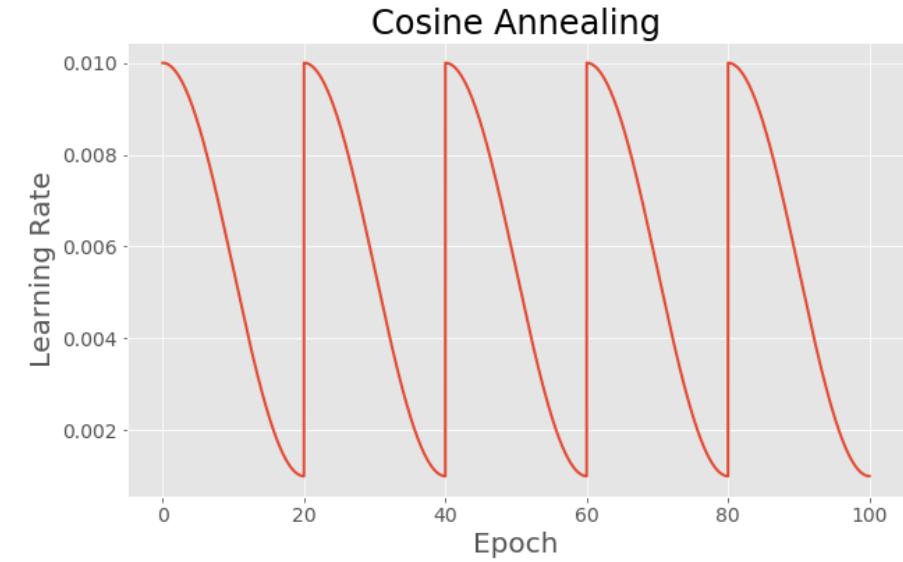
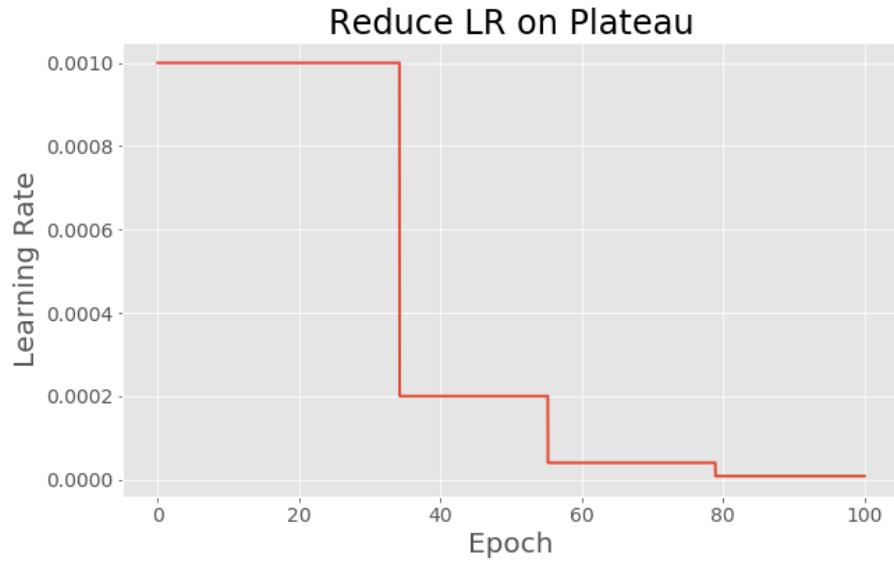
Initialize the
training

Train the model

Baseline model

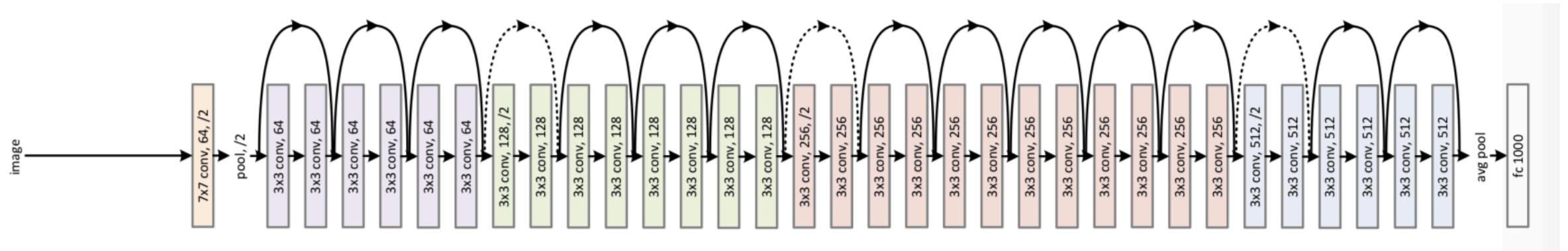
- Basic architecture, basic hyperparameters
- Debug:
 - Visualize everything (train generator, augmentations, test generator, predictions)
 - Overfit to the small portion of the train data

Learning rate schedule

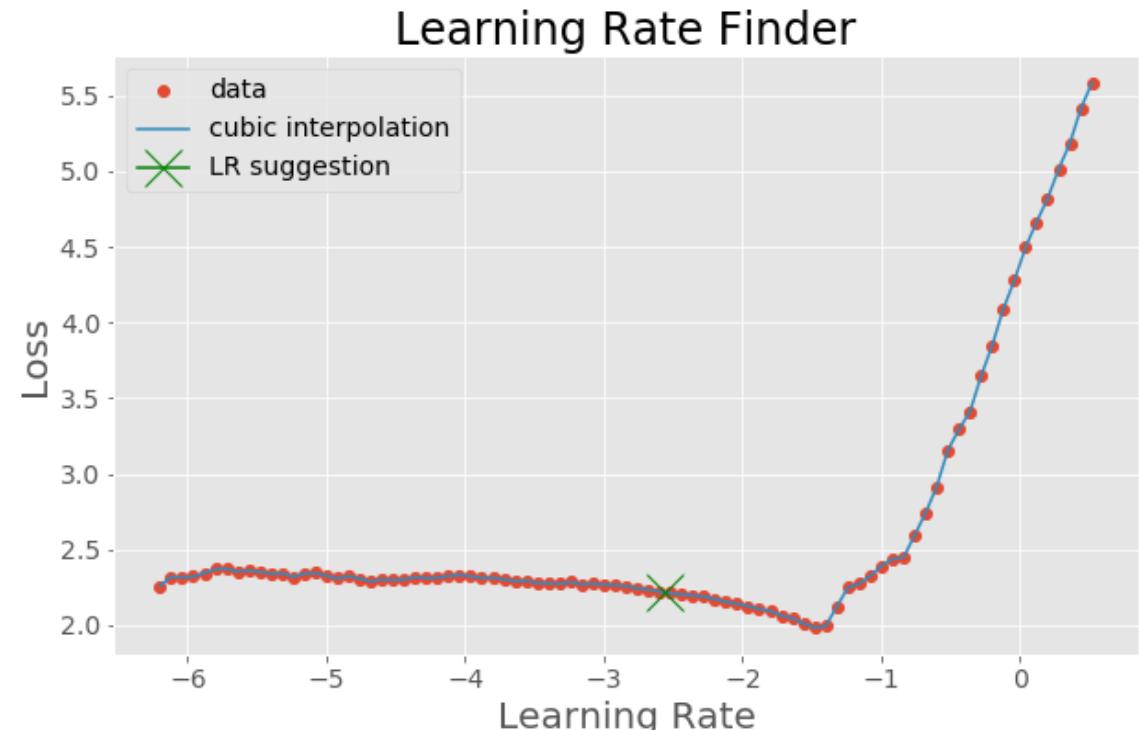
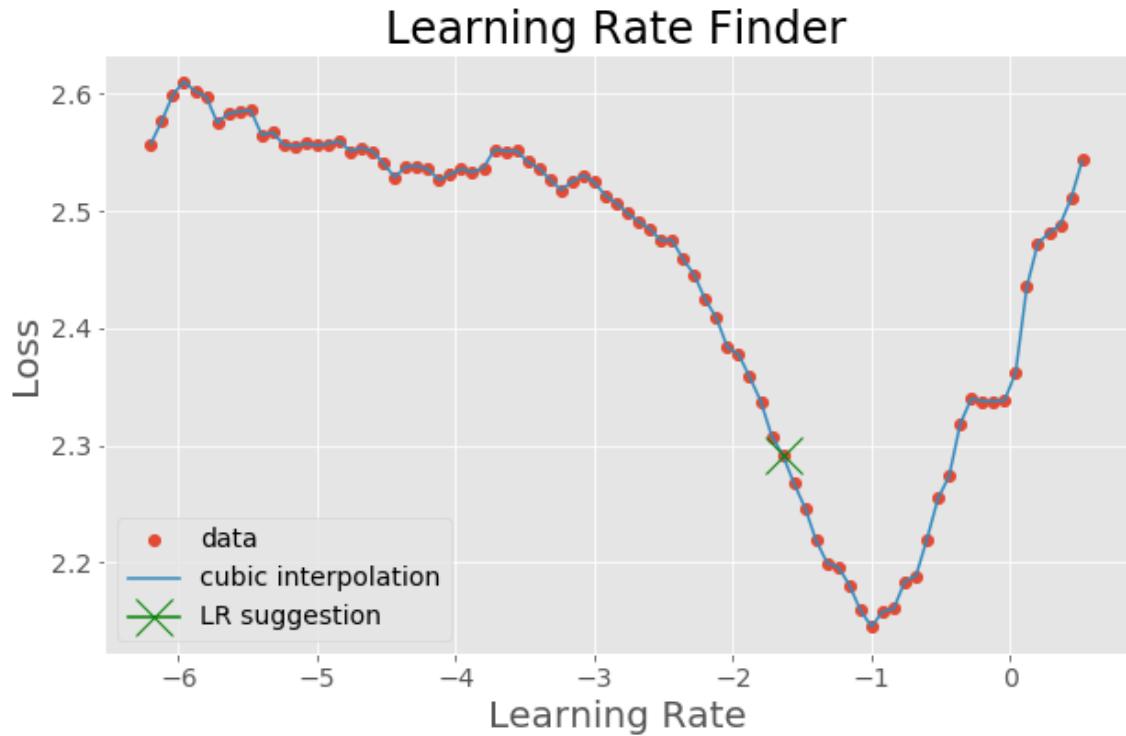


Learning rate add-ons

- Freeze pretrained weights initially
- Discriminative learning rate (different learning rates for different parts of the network)

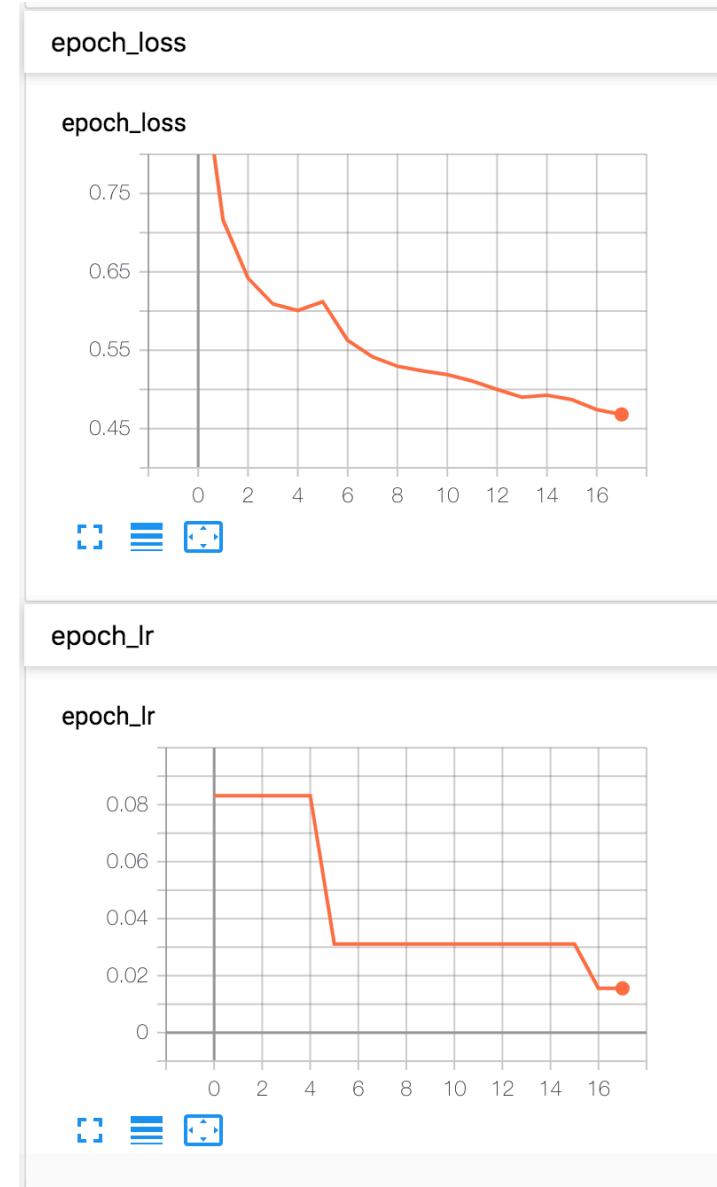


Find initial learning rate

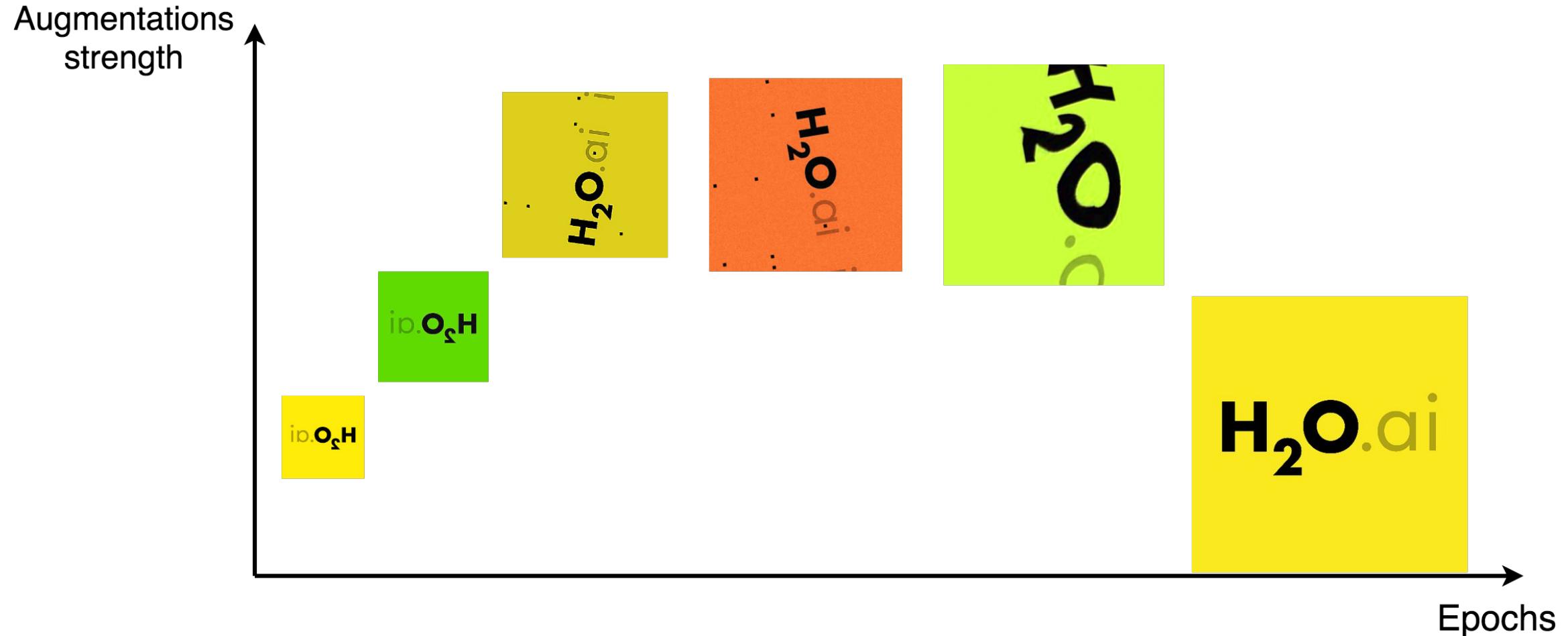


Follow the training process

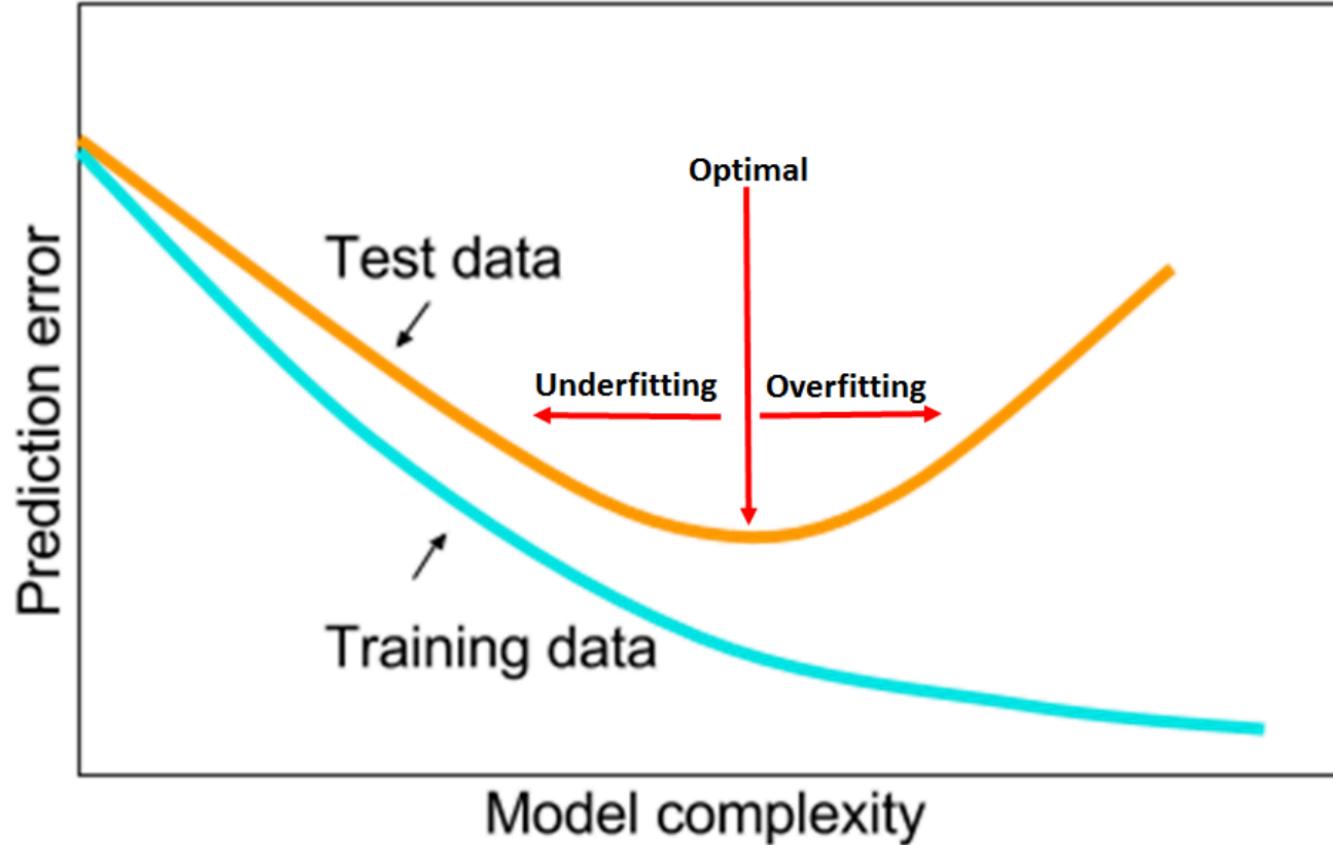
- Track loss, metrics, learning rate, etc.
- Visualize predictions dynamics
- Save model checkpoints



Speed up convergence

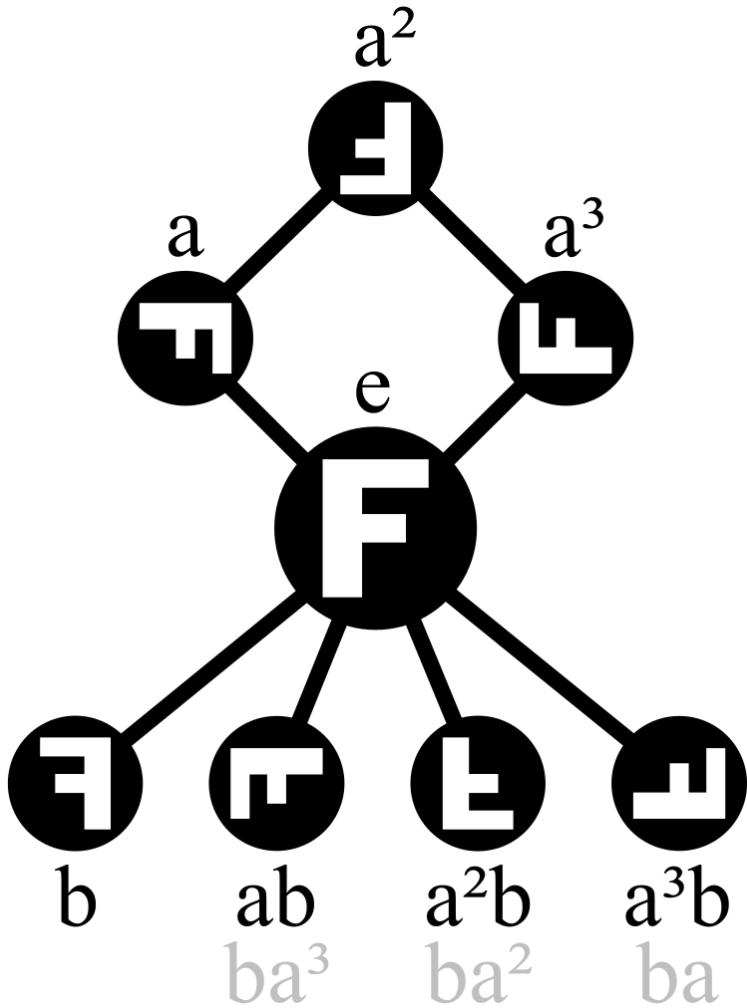


Early stopping



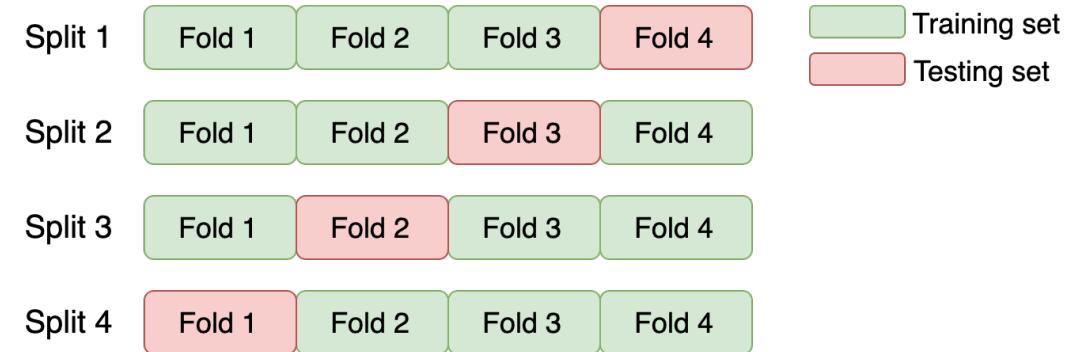
Test time augmentations (TTA)

- Crops
- Flips
- Rotations
- D4 group



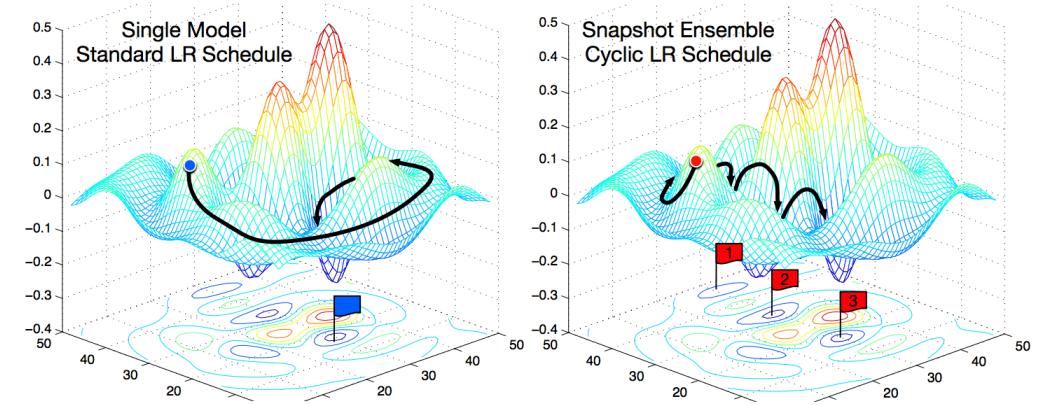
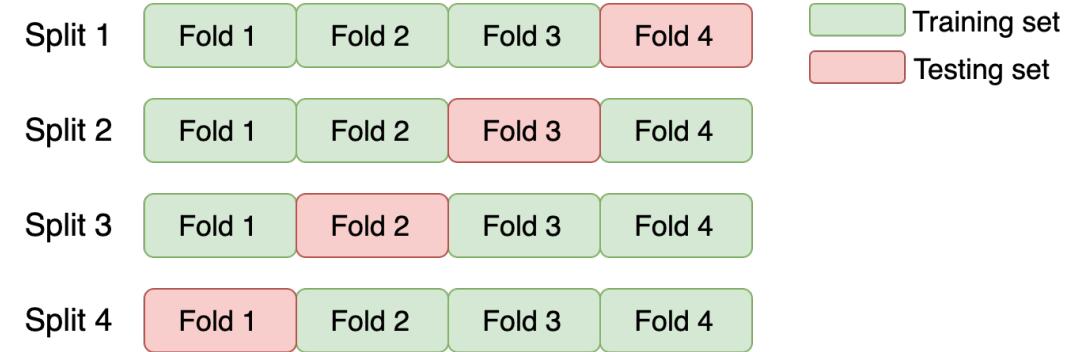
Ensembling

- Multiple folds



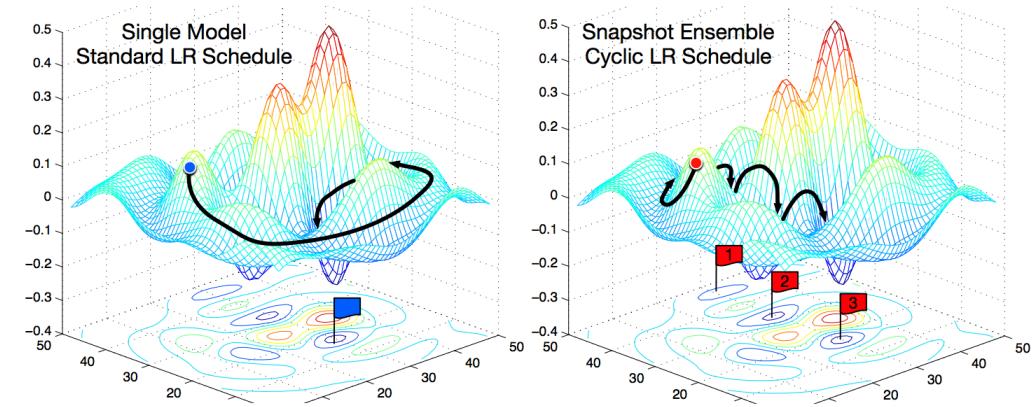
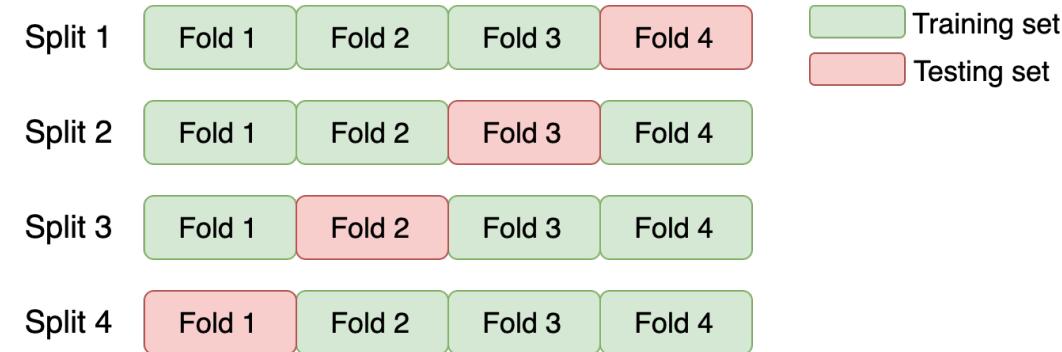
Ensembling

- Multiple folds
- Multiple snapshots



Ensembling

- Multiple folds
- Multiple snapshots
- Blending of multiple models
- Out-of-fold predictions + stacking



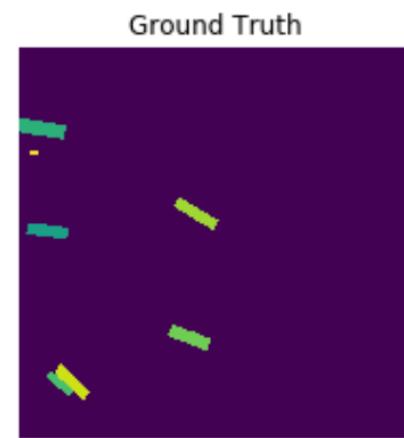
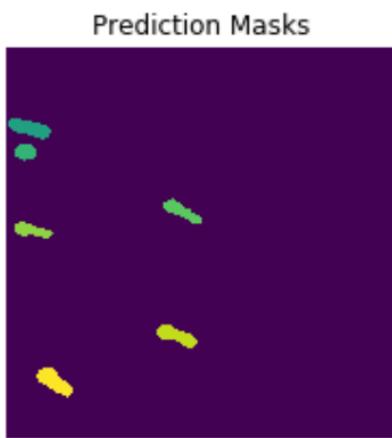
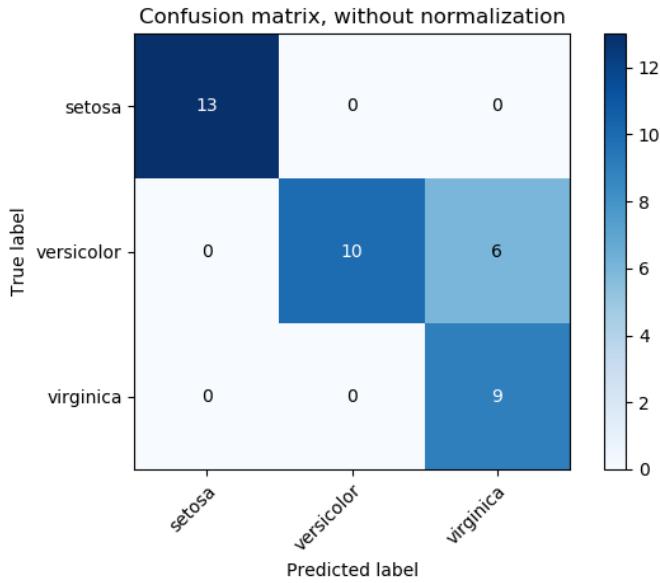
Model is trained

Prepare the data

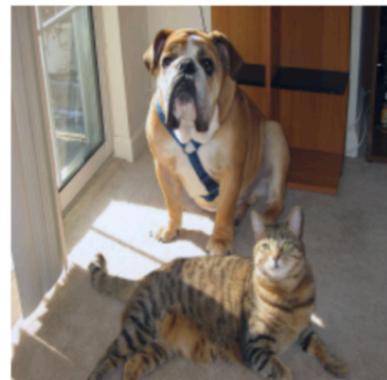
Initialize the
training

Train the model

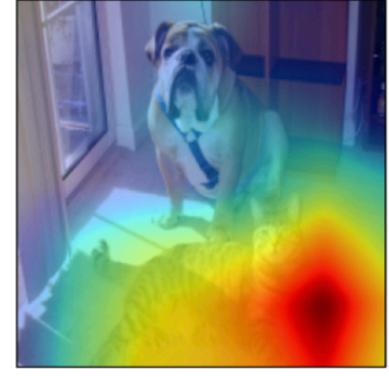
Analyze test predictions



(a) Original Image



(g) Original Image

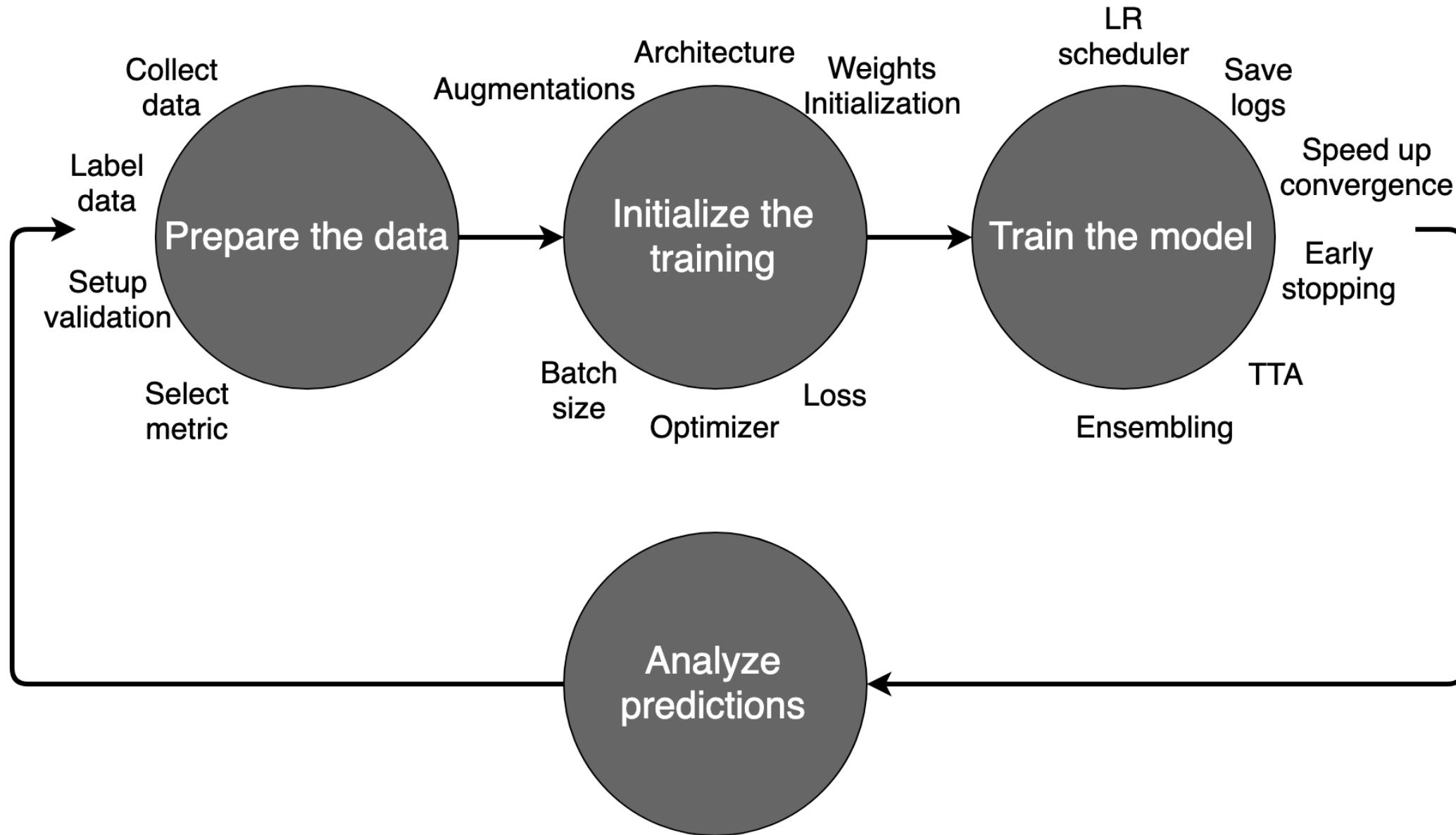


(f) ResNet Grad-CAM ‘Cat’



(l) ResNet Grad-CAM ‘Dog’

Overall pipeline



What to do if it fails

- Evaluate by eyes
- Train longer
- Debug again
- Get more data

H2O.ai Driverless AI

- Automatic image classification
- Automatic image semantic segmentation



Thank you!