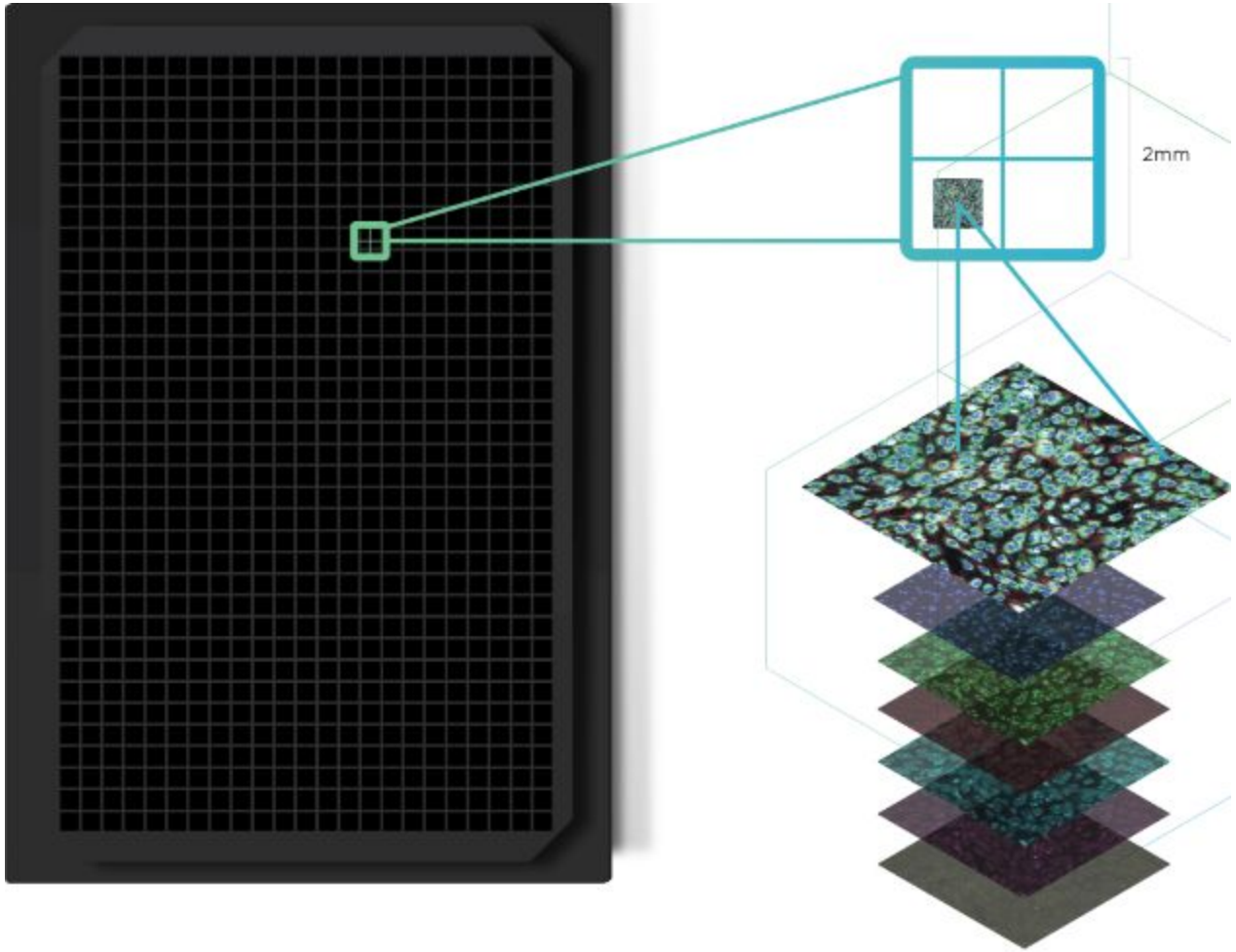
Recursion Cellular Image Classification

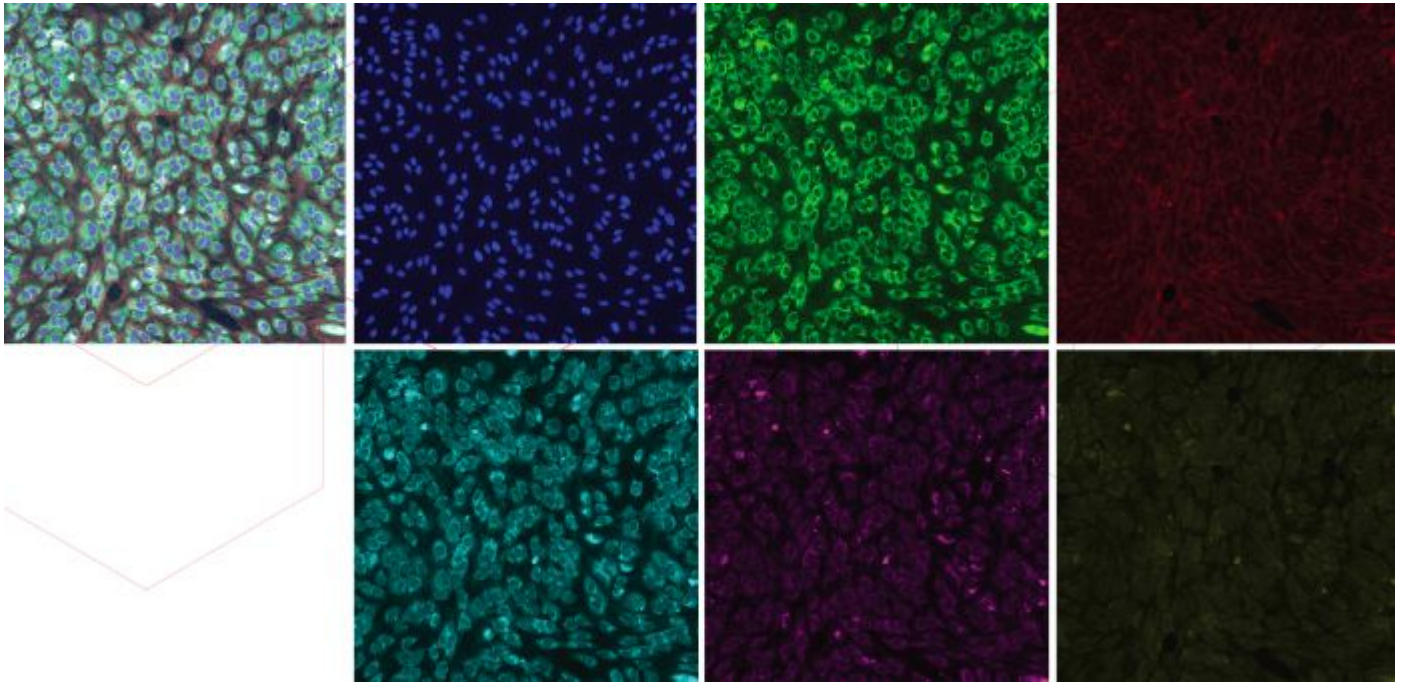
Kirill Brodt

Task description

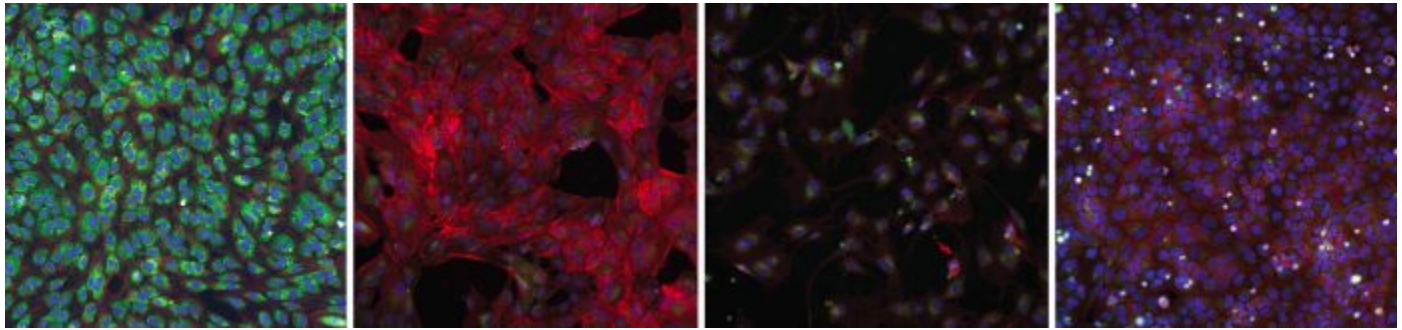
- Ordinary image classification task (or slightly not?)
- 1108 classes
  - 6 channels of 512x512 arrays (6x512x512)
  - 4 cell types of siRNAs (HepG2, HUVEC, RPE and U2OS) divided per experiments (batch effect problem)
  - 4 plates with 277 wells (and 277 classes)
  - positive and negative controls (31 another classes)
- Accuracy metric


2mm

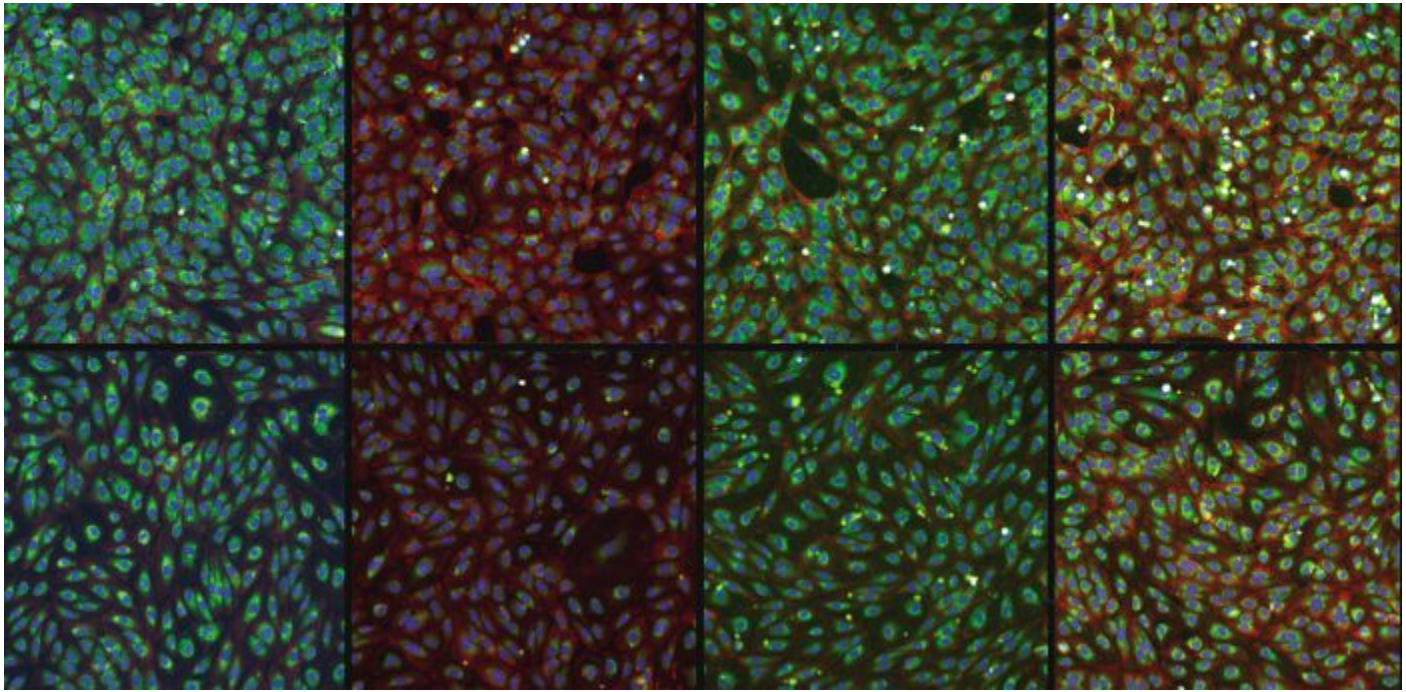
6 channels



4 cell types (HUVEC, RPE, HepG2, U2OS)



Batch effect



Solution (PyTorch)

- DenseNet121 from public kernel per each cell type (4 models) on 6x512x512
- Mixed precision (O1 level)
- BCE loss (yes, not xEntropy)
- Batch size 8
- Adam lr=1e-3 with reduce by 10 on 320 and 350 epochs (375 epochs in total)
- Pretrain on controls first 100 epochs
- Mean predictions over 2 sites
- RandomRotate90 + HorizontalFlip augmentations
- Best checkpoint on validation (splitted by experiments)
- 24h on 1 Tesla V100

Train score about 1, validation -- **random** (0.001 ~ 1/1108)

Key ingredient: eval mode and inference per experiment

- Two weeks of debugging and eureka! **Turn off eval mode** and get 0.54 public and 0.75 private from nothing!

- Due batch effect the model couldn't "~~generalize~~" learn good features for all samples through all experiments
  - Need some good standardization of all samples
  - With eval mode Batch Normalization Layer do not normalizes over batch and uses learned statistics over all experiments (recall batch effect)
- **Inference per experiments** (trick: infinite bagging)

Tricks

- Virtual batch 128=16*8 (in other words batch accumulation)
  - 0.72 (+18) pub and 0.86 (+11) pvt (0.54 pub and 0.75 pvt)
- In one experiment there are only one class
  - 0.76 (+4) pub and 0.89 (+3) pvt
- Actually, classes divided equally in 4 plates (277 classes/plate)
  - 0.86 (+10) pub and 0.94 (+5) pvt
- Finetune with MixUp on all data (new sample is a convex combination of two old)
  - 0.88 (+2) pub and 0.95 (+1) pvt

N. B.: Same local validation and private scores

What wasn’t better my validation

- Densete201, ResNet50, SEResNeXt50
- Cross entropy loss
- Label smoothing
- CosFace loss (metric learning)
- Random crops (448x448)
- Standardization by experiments, plates and channels
- TTA with flips, rotations, etc..
- RAdam
- Cyclical LR with Cosine annealing
- Using controls during training
    - sample **p**ositive and **n**egative siRNAs in batch and make new features cat([x, x*p, x*n, x*p*n]), where x is ordinary siRNA
- Pseudo Labeling

Mistakes review

- Know what's under the hood
- Sufficient data -- no need bacchanalia