Recursion Cellular Image Classification

overview of 84th place solution (bronze medal)

Георгий Сурин
ods.ai @formemorte

Competition overview

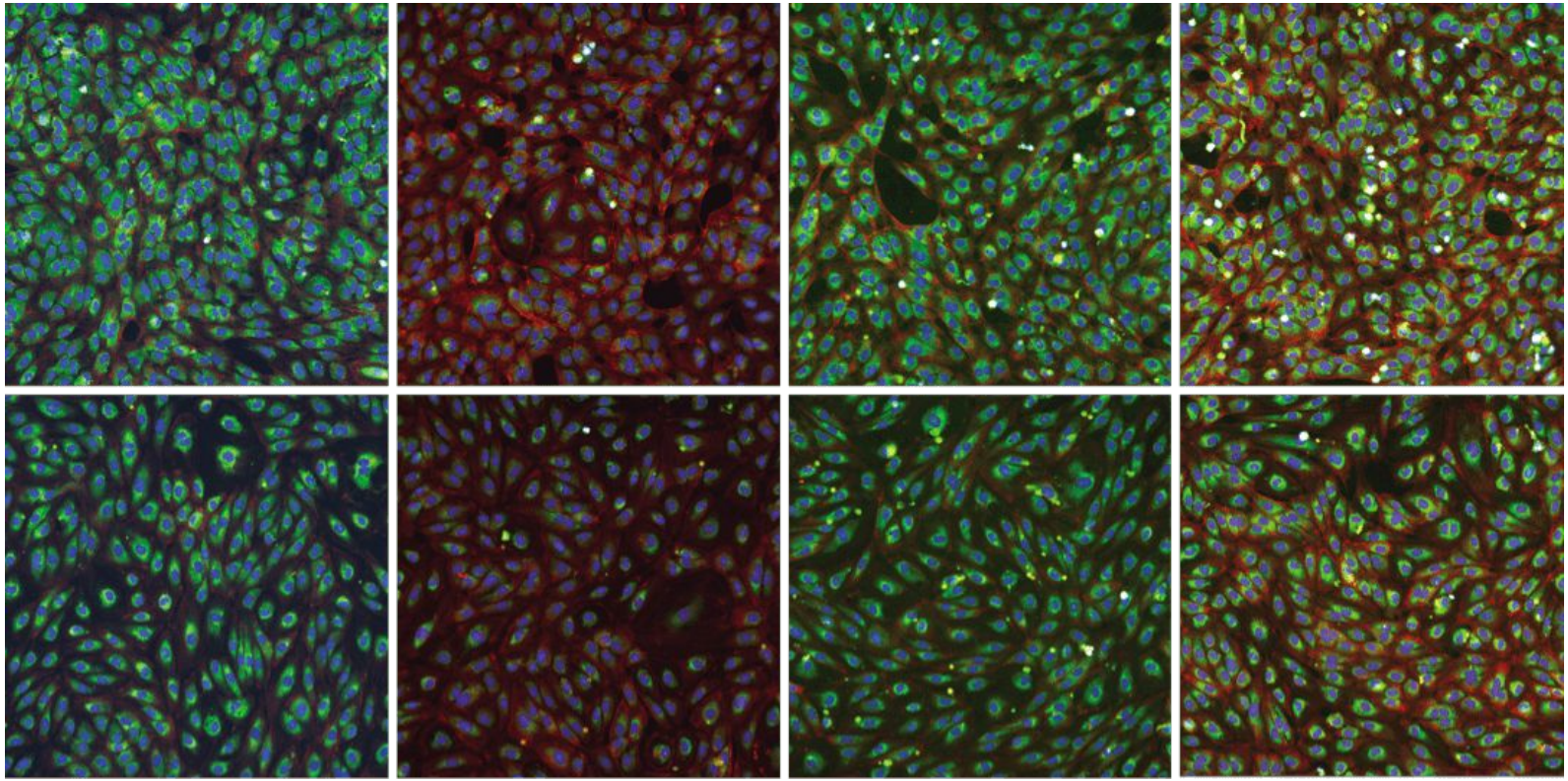
The same siRNAs (effectively genetic perturbations) have been applied repeatedly to multiple cell lines, for a total of 51 experimental batches. Each batch has four plates, each of which has 308 filled wells. For each well, microscope images were taken at two sites and across six imaging channels. Not every batch will necessarily have every well filled or every siRNA present.

File Description

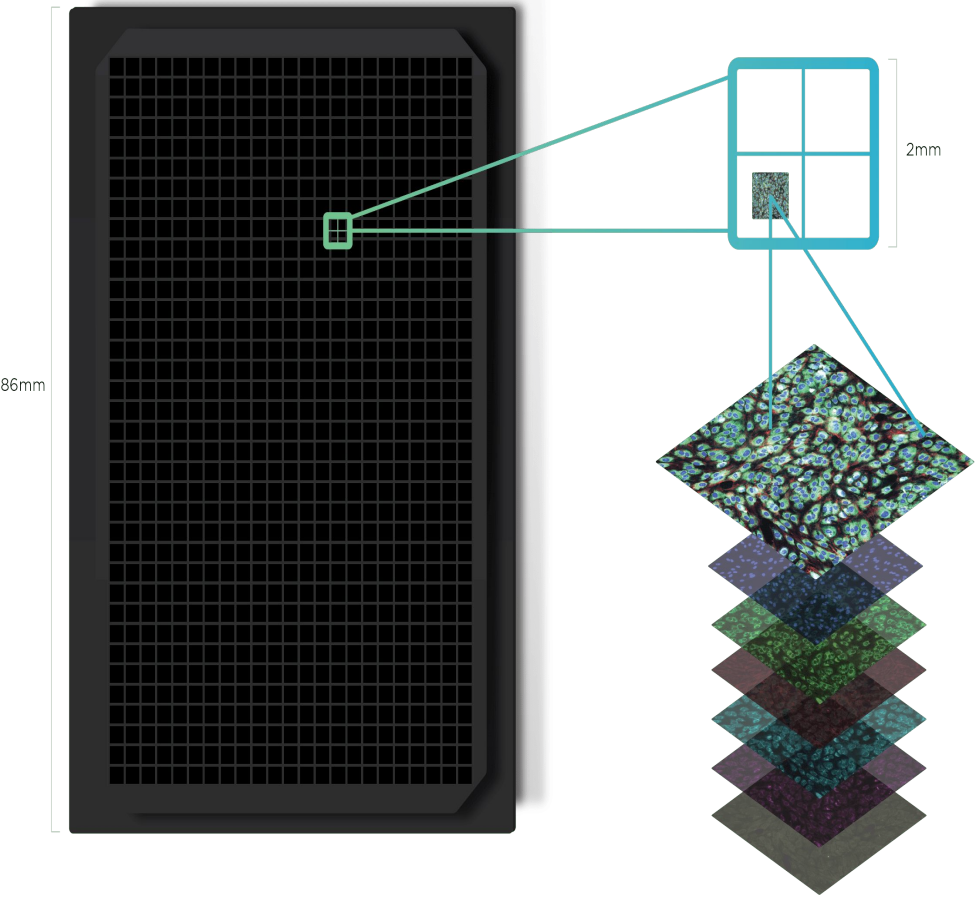
- [train/test].zip: the image data. The image paths, such as `U2OS-01/Plate1/B02_s2_w3.png`, can be read as:
    - Cell line and batch number (U2OS batch 1)
    - Plate number (1)
    - Well location on plate (column B, row 2)
    - Site (2)
    - Microscope channel (3)

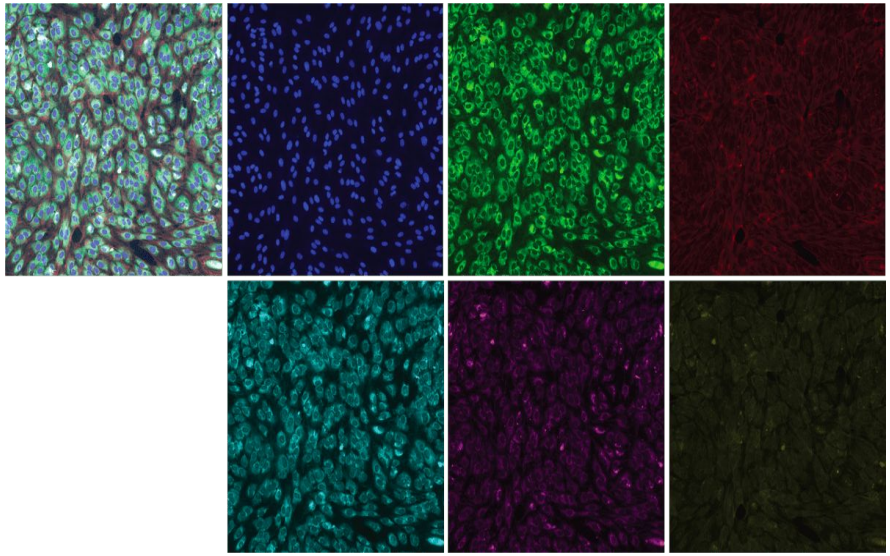
      Please note that the [train/test].csv and [train/test]_controls.csv combined describe the images found in [train/test].zip. You will only be making predictions on the images listed in test.csv, not on all the images found in test.zip.
- [train/test].csv
    - `id_code`
    - `experiment`: the cell type and batch number
    - `plate`: plate number within the experiment
    - `well`: location on the plate
    - `sirna`: the target
- [train/test]_controls.csv In each experiment, the same 30 siRNAs appear on every plate as positive controls. In addition, there is at least one well per plate with untreated cells as a negative control. It has the same schema as [train/test].csv, plus a `well_type` field denoting the type of control.
- pixel_stats.csv Provides the mean, standard deviation, median, min, and max pixel values for each channel of each image.
- sample_submission.csv A valid sample submission.

Data

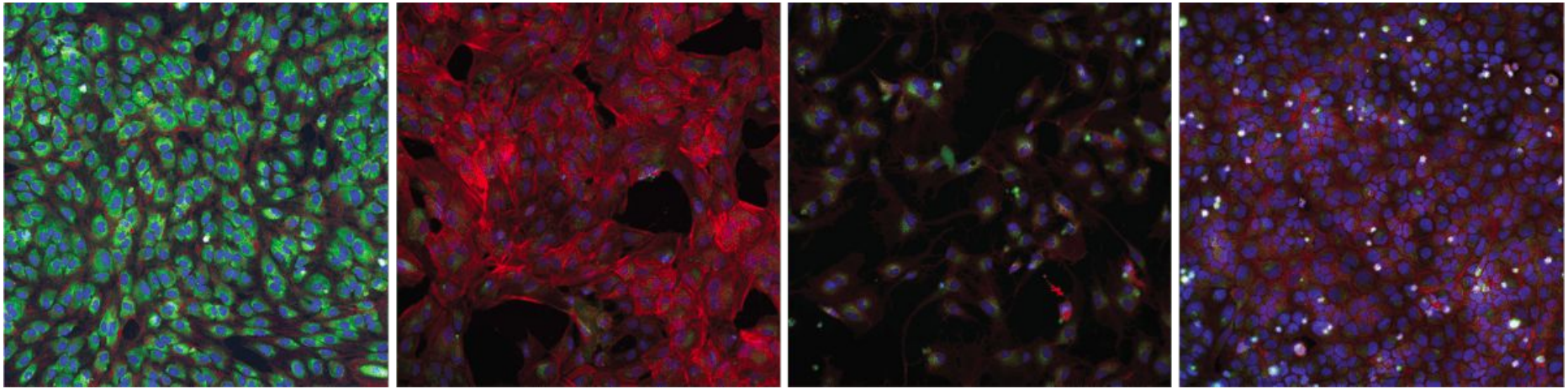


Images of two different genetic conditions (rows) in HUVEC cells across four experimental batches (columns).





ядра (blue), эндоплазматические ретикулы (green), актин (red), ядрышки (cyan), митохондрии (magenta), аппарат Гольджи(yellow).



Images of the same siRNA across four cell types: HUVEC, RPE, HepG2, U2OS

Settings

Random split 95%/5%

augmentation (for augmentations was used Albumentations):

- D4

optimizer: SGD momentum=0.9, wd=5e-4, warmup 5 epoch to lr=0.01, lr_decay /10 on 15, 25 epoch

loss: BCE

model Resnet34 for experiments / SeResneXt50 for real training + dropout(0.5) before head

mixup + label smoothing (from bag of tricks https://arxiv.org/pdf/1812.01187.pdf)

Albumentations: https://github.com/albu/albumentations