

Секретный докладчик



kaggle™

Human Protein Atlas Image Classification

Что такое kaggle ?

**НЕТ ВРЕМЕНИ
ОБЪЯСНЯТЬ!**

НО Я ОБЪЯСНИЮ

Что такое kaggle ?

- Соревновательная платформа
- Крупнейшая площадка по соревнованиям в сфере DS
 - Boosters.pro
 - MIBootCamp.ru
- 300 завершенных соревнований по различным направлениям
- 16 активных соревнований
- Большое и дружное сообщество
- Возможность создавать публичные варианты решений и делиться ими
- Возможность объединяться в команды
- М - мотивация



Почему нужно решать kaggle

kaggle Search Competitions Datasets Kernels Discussion Learn ...

Competitions Documentation InClass

General InClass Sort by Grouped All Categories Search competitions

1 Entered Competition

 **VSB Power Line Fault Detection**
Can you detect faults in above-ground electrical lines?
Featured · 2 months to go · signal processing, tabular data, binary classification

 **Two Sigma: Using News to Predict Stock Movements**
Use news analytics to predict stock price performance
Featured · Kernels Competition · 6 months to go · news agencies, time series, finance, money

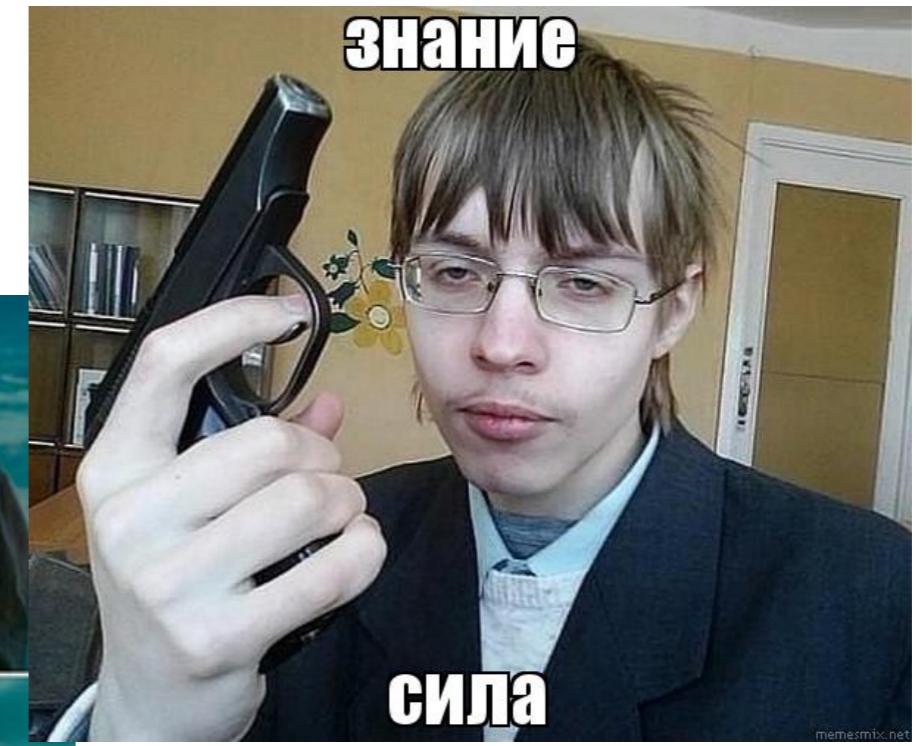
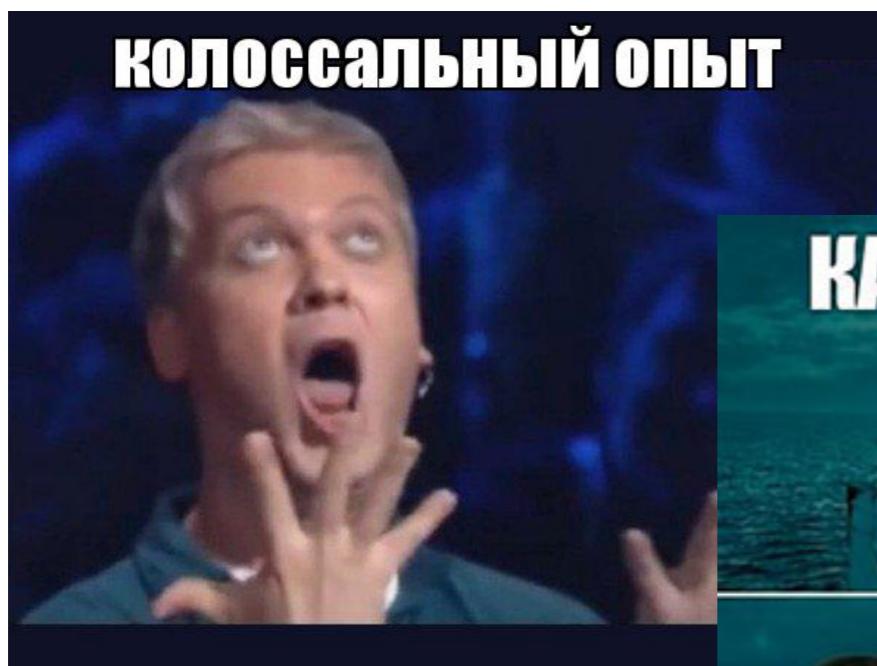
 **LANL Earthquake Prediction**
Can you predict upcoming laboratory earthquakes?
Research · 4 months to go · earth sciences, physics, signal processing

\$25,000 573 teams

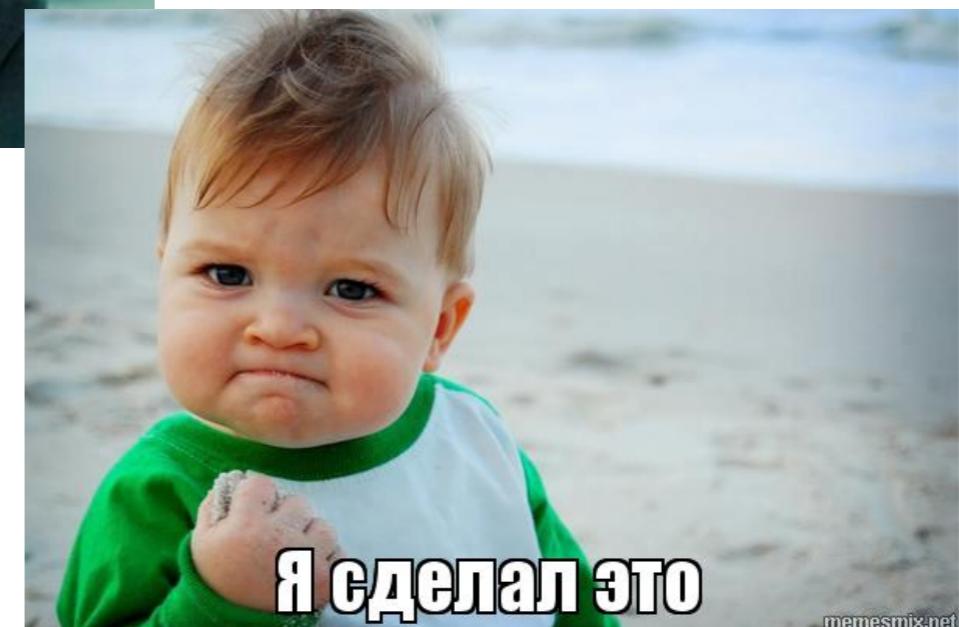
\$100,000 2,902 teams

\$50,000 667 teams

Почему нужно решать kaggle



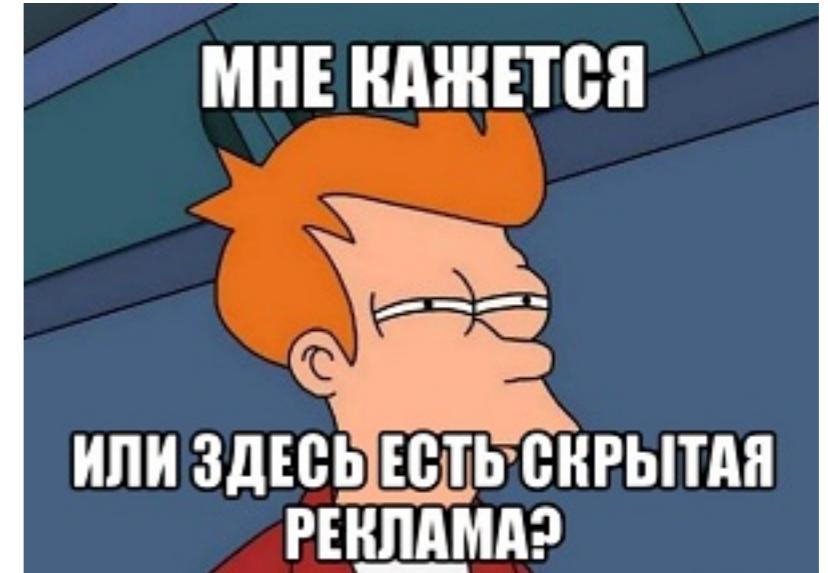
memesmix.net



memesmix.net

Оказывается, мы уже решаем **kaggle**

- Каждую (почти) субботу
- В 16:00 по местному времени
- В офисе **7bits** на Таубе 5
- Зарешки Kaggle!
- Telegram Чат для более подробной информации
@kaggleClubOmsk

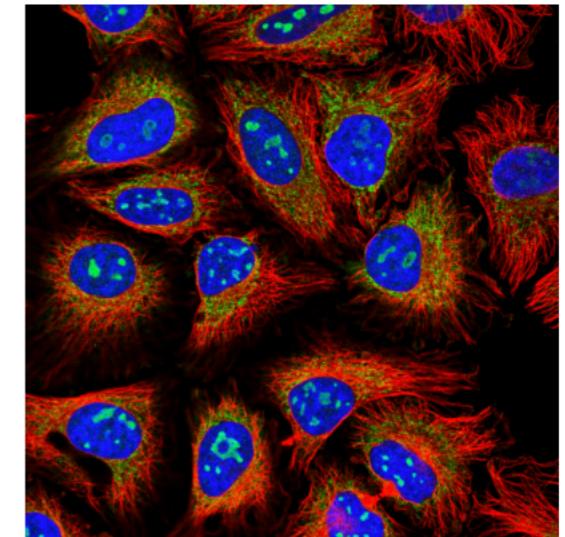


Human Protein Atlas Image Classification

Overview

Description

In this competition, Kagglers will develop models capable of classifying mixed patterns of proteins in microscope images. [The Human Protein Atlas](#) will use these models to build a tool integrated with their smart-microscopy system to identify a protein's location(s) from a high-throughput image.



Evaluation

Prizes

Timeline

Special Prize Instructions

Proteins are “the doers” in the human cell, executing many functions that together enable life. Historically, classification of proteins has been limited to single patterns in one or a few cell types, but in order to fully understand the complexity of the human cell, models must classify mixed patterns across a range of different human cells.

Images visualizing proteins in cells are commonly used for biomedical research, and these cells could hold the key for the next breakthrough in medicine. However, thanks to advances in high-throughput microscopy, these images are generated at a far greater pace than what can be manually evaluated. Therefore, the need is greater than ever for automating biomedical image analysis to accelerate the understanding of human cells and disease.

Nature Methods has indicated interest in considering a paper discussing the outcome and approaches of the challenge. The Human Protein Atlas team would like to invite top performing teams to join as co-authors in the writing of this paper.

Top performing teams will also be eligible to compete for the special prize. Additional information for both competition and co-authoring for *Nature Methods* will become available through the Discussion posts once the competition is complete.



Human Protein Atlas Image Classification

What files do I need?

You will need to download a copy of the images. Due to size, we have provided two versions of the same images. On the data page below, you will find a scaled set of 512x512 PNG files in [train.zip](#) and [test.zip](#). Alternatively, if you wish to work with full size original images (a mix of 2048x2048 and 3072x3072 TIFF files) you may download [train_full_size.7z](#) and [test_full_size.7z](#) from [here](#) (warning: these are ~250 GB total).

You will also need the training labels from [train.csv](#) and the filenames for the test set from [sample_submission.csv](#).



Description

Submissions will be evaluated based on their [macro F1 score](#).

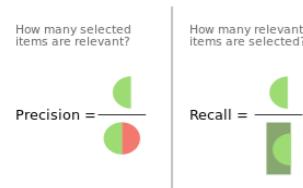
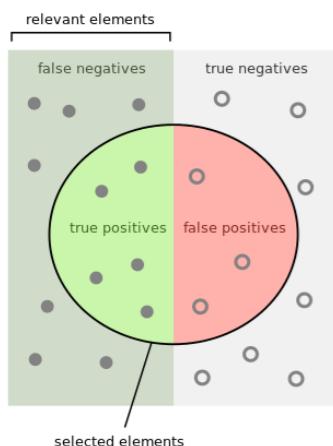
Evaluation

Submission File

Prizes

Timeline

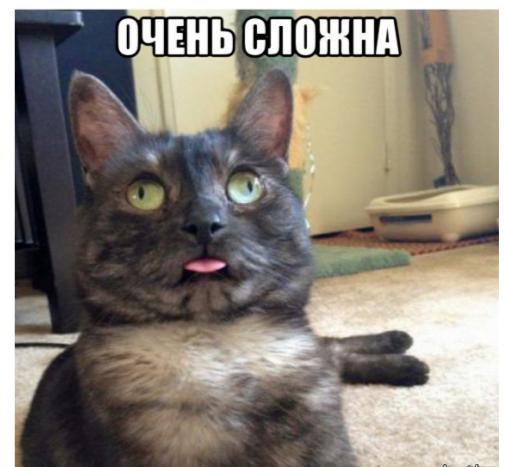
For each Id in the test set, you must predict a class for the [Target](#) variable as described in [the data page](#). Note that multiple labels can be predicted for each sample.



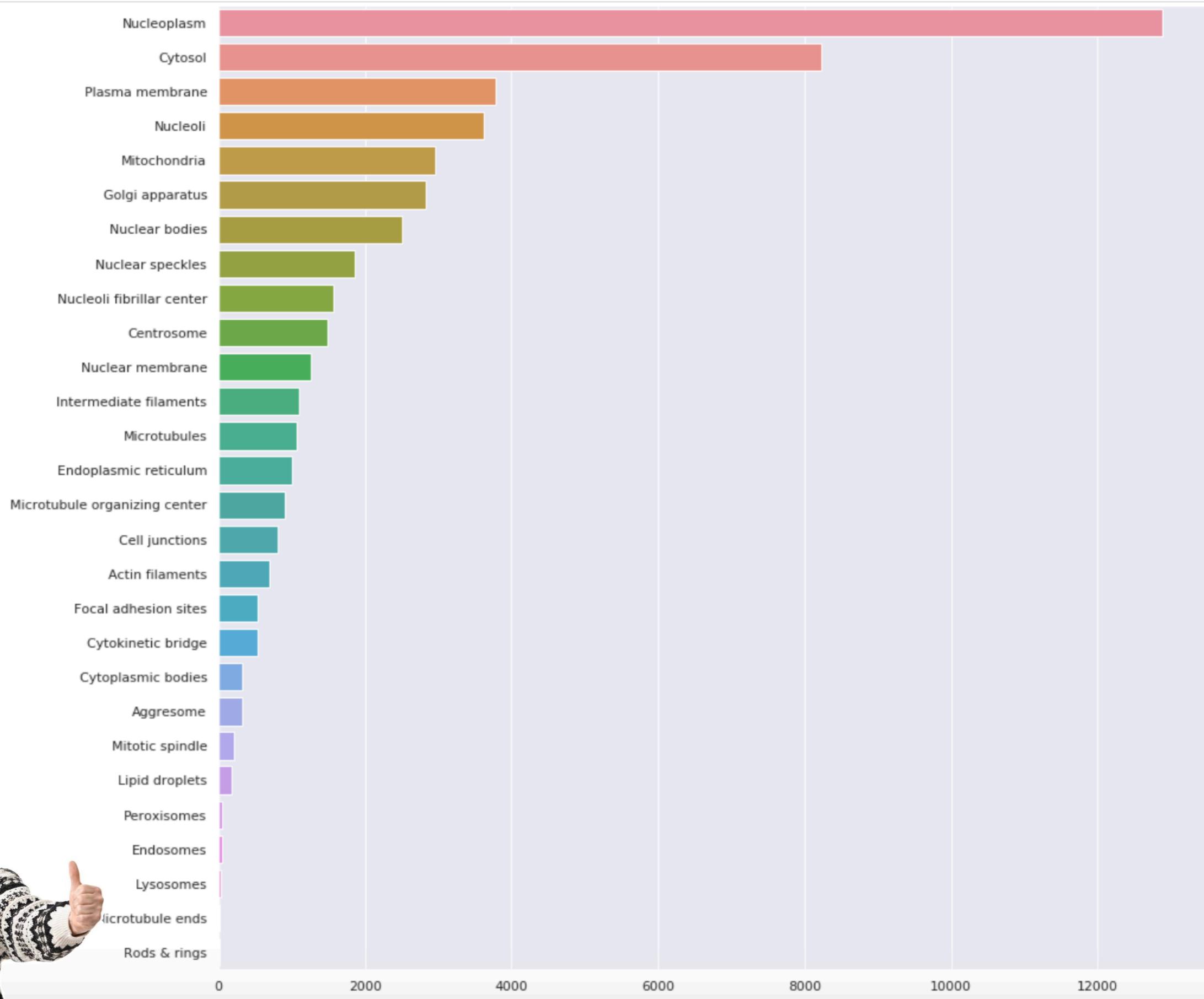
$$\text{Macro - Precision} = \frac{\text{Precision1} + \text{Precision2}}{2}$$

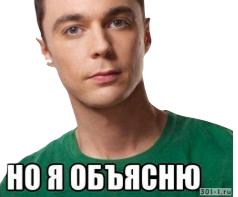
$$\text{Macro - Recall} = \frac{\text{Recall1} + \text{Recall2}}{2}$$

$$\text{Macro - F - Score} = 2 \cdot \frac{\text{Macro - Precision} \cdot \text{Macro - Recall}}{\text{Macro - Precision} + \text{Macro - Recall}}$$



Human Protein Atlas Image Classification



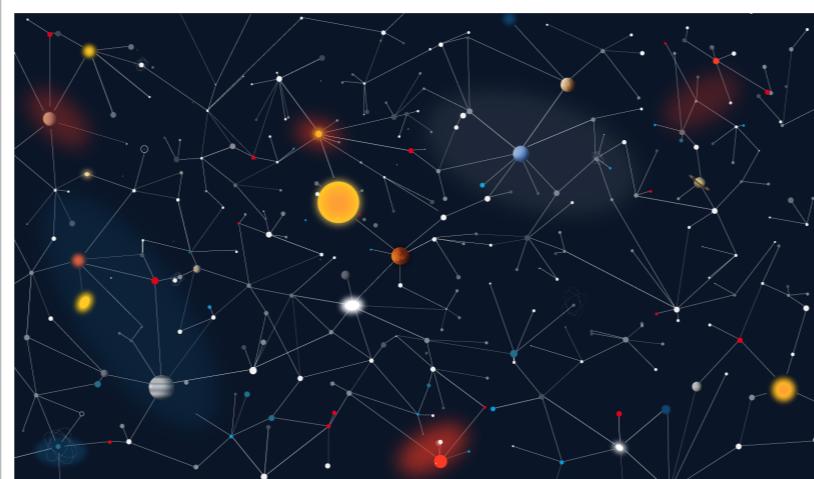


Human Protein Atlas Image Classification

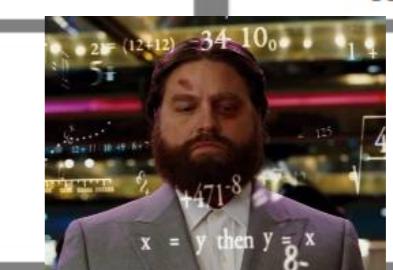
Процесс обучения нейросети



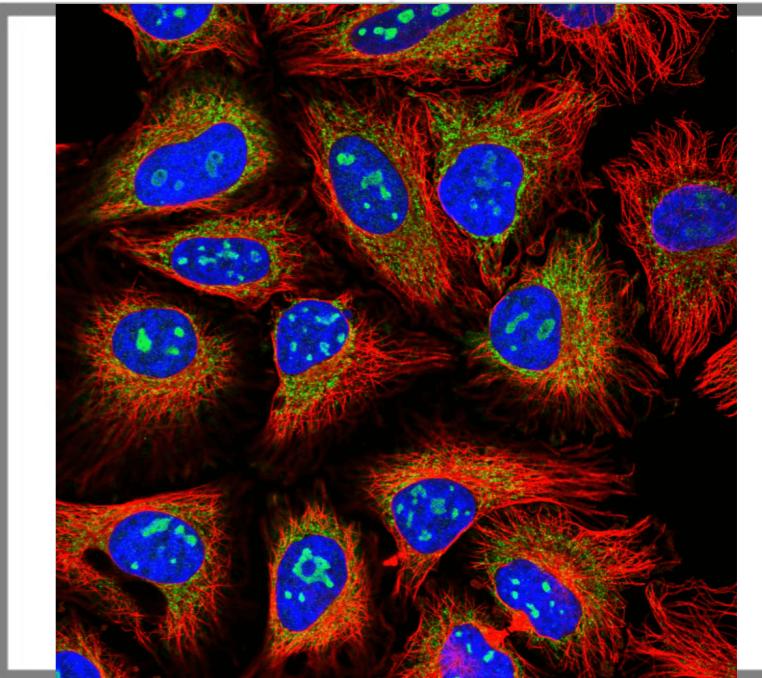
распространение
сигнала по нейросети



ответ
сети



ошибка
мала



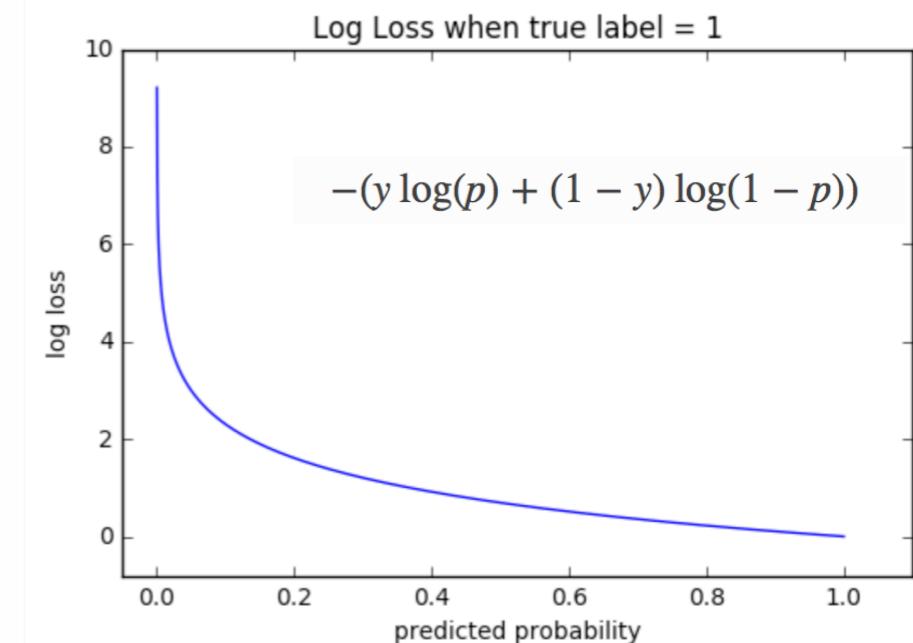
ошибка
велика



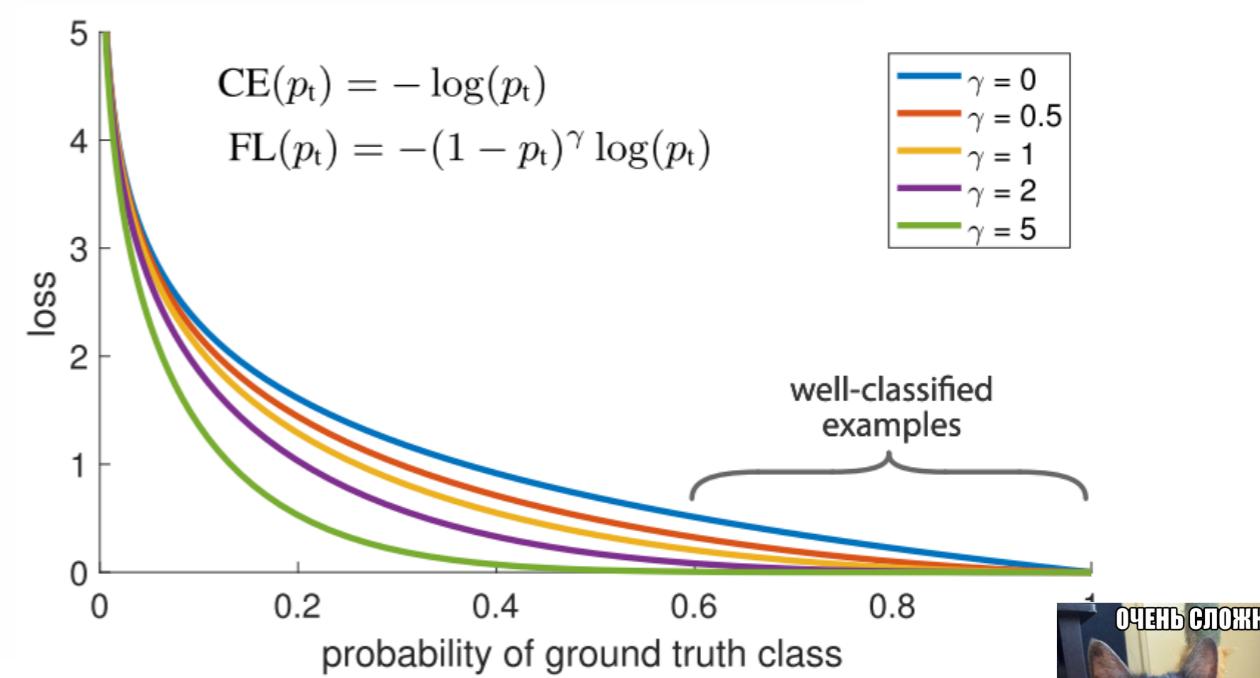
Human Protein Atlas Image Classification



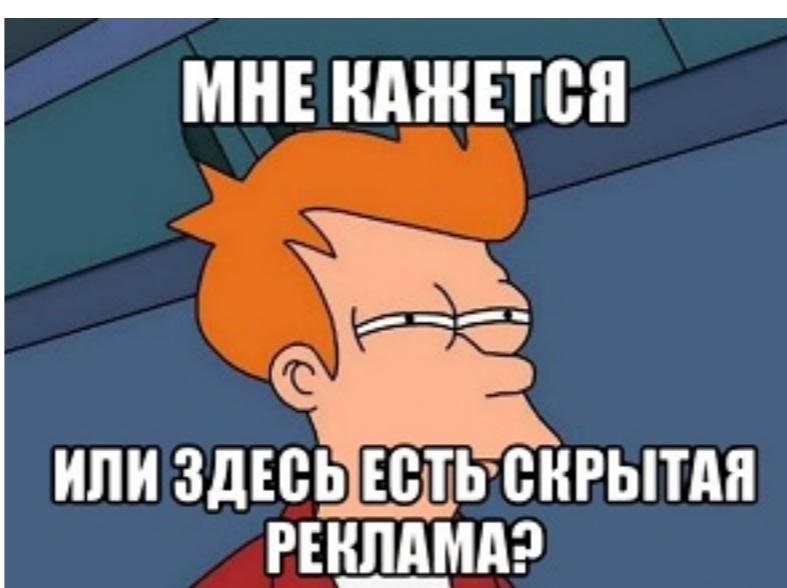
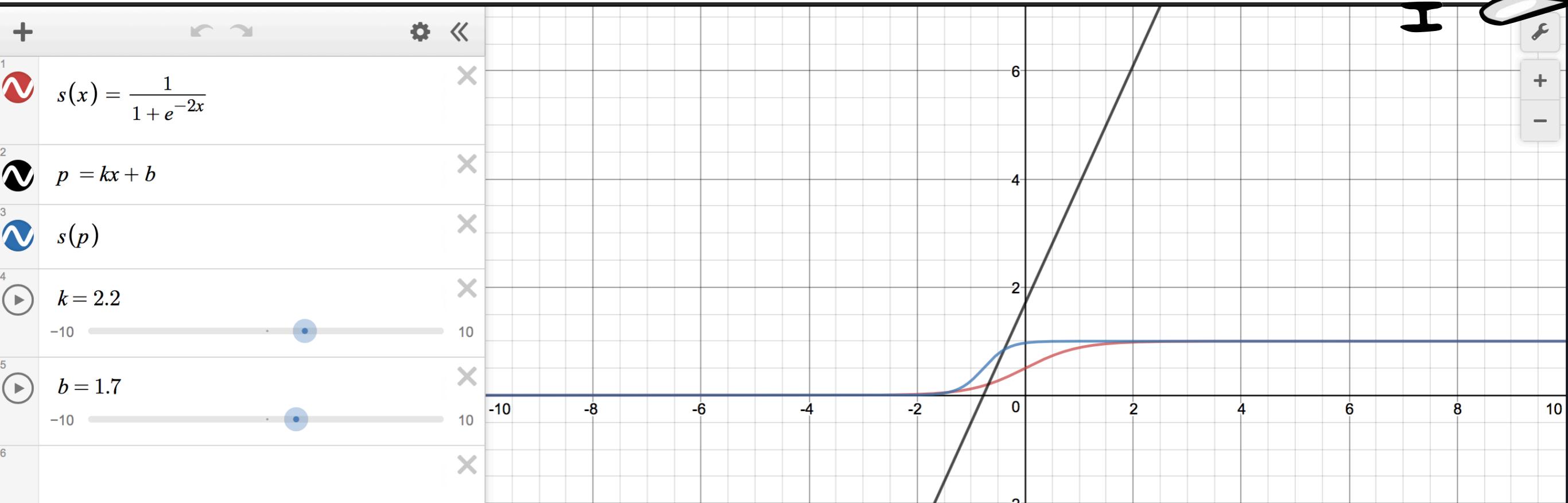
LogLoss



FocalLoss

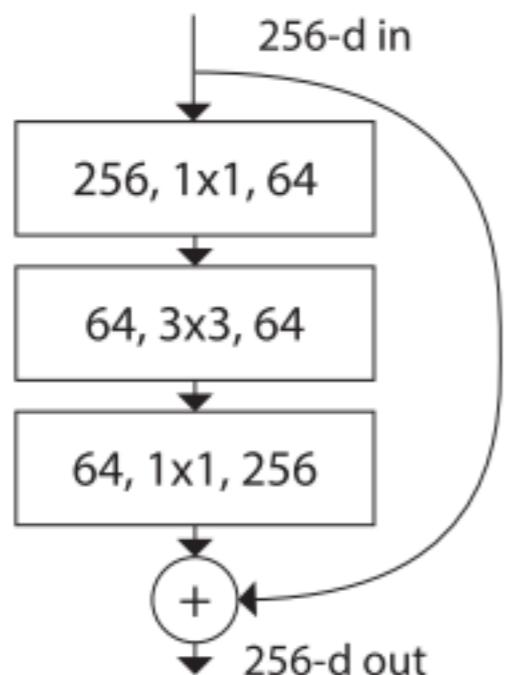


ничоси

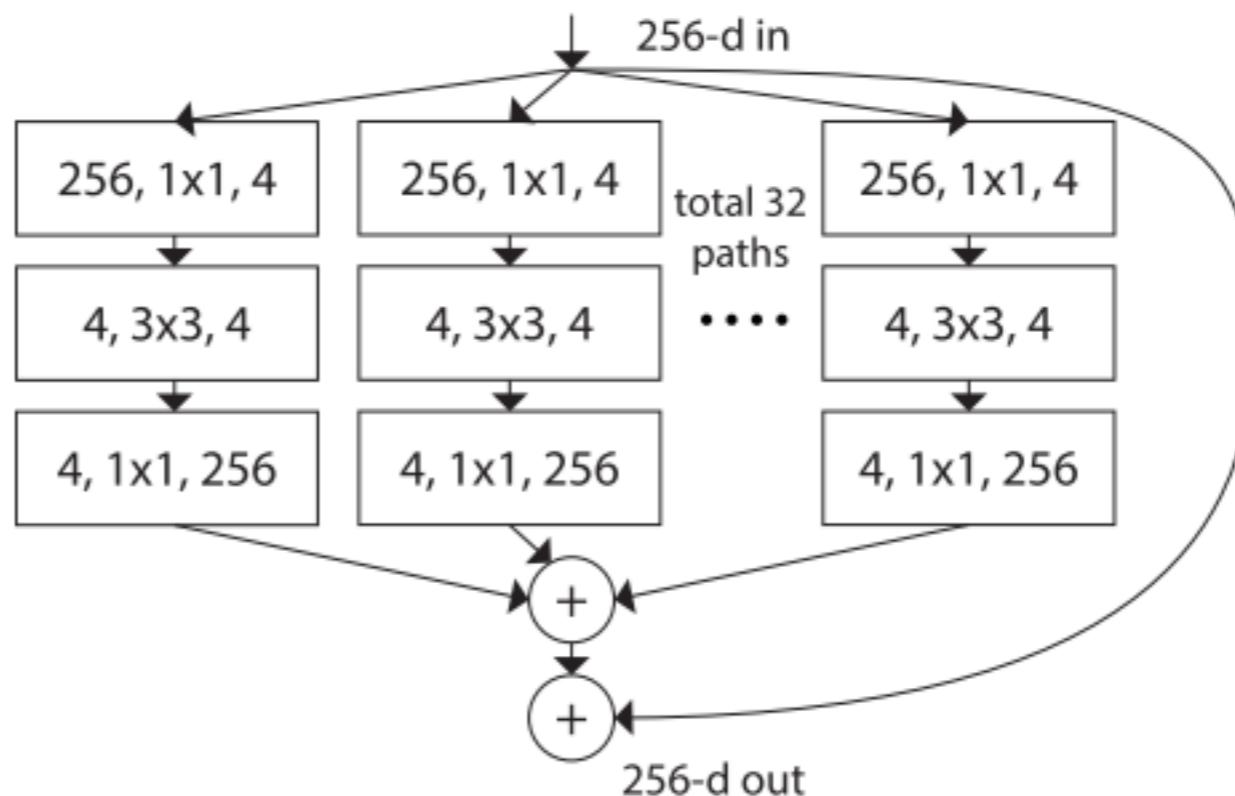


Human Protein Atlas Image Classification

ResNet

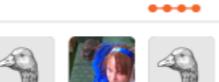


ResNext



PYTORCH

Human Protein Atlas Image Classification

451	▼ 42	KMAI_seek		0.47427	5	1mo
452	▲ 518	Lukasz Nalewajko		0.47400	17	15d
453	▼ 32	JB_nb		0.47387	23	
454	▲ 419	hpa		0.47		
455	▲ 549	kiyomomo		0.4		
456	▼ 44	AndrewWang		0.47332		
457	▼ 44	Menoua Keshishian		0.47326	11	
458	▼ 44	Rebecca, Brie		0.47262	13	1mo
459	▼ 43	Chris Oosthuizen		0.47260	22	1mo
460	▼ 43	Saronio		0.47253	9	1mo
461	▼ 43	yyob		0.47241	32	17d
462	▲ 197	437		0.47227	29	15d
463	▼ 44	Tim Joseph		0.47226	26	1mo
464	▲ 546	BII Tampere		0.47214	22	15d
465	▼ 45	Shubhra Aich		0.47209	34	2mo
466	▼ 44	without GPU		0.47153	24	1mo

Human Protein Atlas Image Classification

711	▲ 133	john_ting		0.44420	5	2mo
712	▲ 146	stormstone		0.44420	7	2mo
713	▲ 146	sneffort		0.44420	9	1mo
714	▲ 146	KMAI_Pat and Mat		0.44420	11	2mo
715	▲ 146	Divya Komaravolu		0.44420	6	1mo
716	▲ 149	Prabhat Kanaujia		0.44420	5	1mo
717	▲ 8	Bac Nguyen Cong		0.44400	50	1mo
718	▼ 193	xpw1		0.44392	6	1mo
719	▼ 43	Yuan Tian		0.44391	6	25d
720	▼ 274	梦想还是要有的		0.44386	25	16d
721	▼ 190	MSJose		0.44383	6	3mo
722	▲ 74	Riccardo Gallina				
723	▼ 101	samirthapa				
724	▼ 138	rentiansky				
725	▼ 120	Vishy				
726	▼ 213	chenhao				
727	▲ 237	Chaoxi Xu		0.44372	2	2mo
728	▼ 262	without GPU		0.44363	24	1mo

Human Protein Atlas Image Classification



protein_classification_03.csv a month ago by Vyaches	0.41002	0.44397
protein_classification_v-2.csv a month ago by Vyaches	0.43447	0.46760
protein_classification_t-2.csv a month ago by Vyaches	0.43372	0.46634
protein_classification_f-2.csv a month ago by Vyaches	0.44988	0.45555
protein_classification_c-2.csv a month ago by Vyaches	0.44363	0.47153

