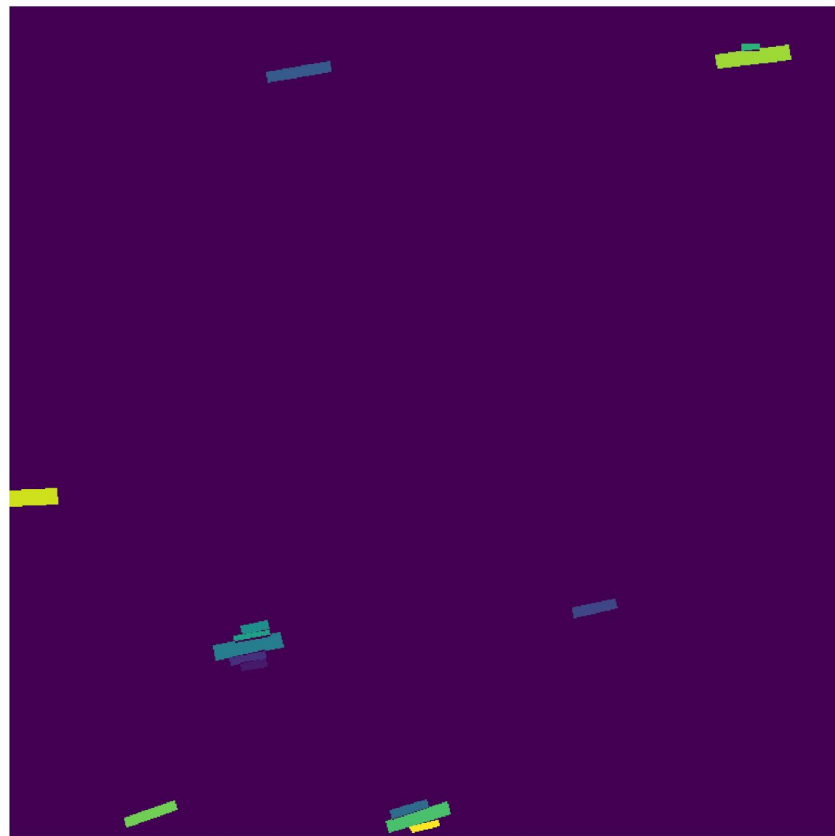
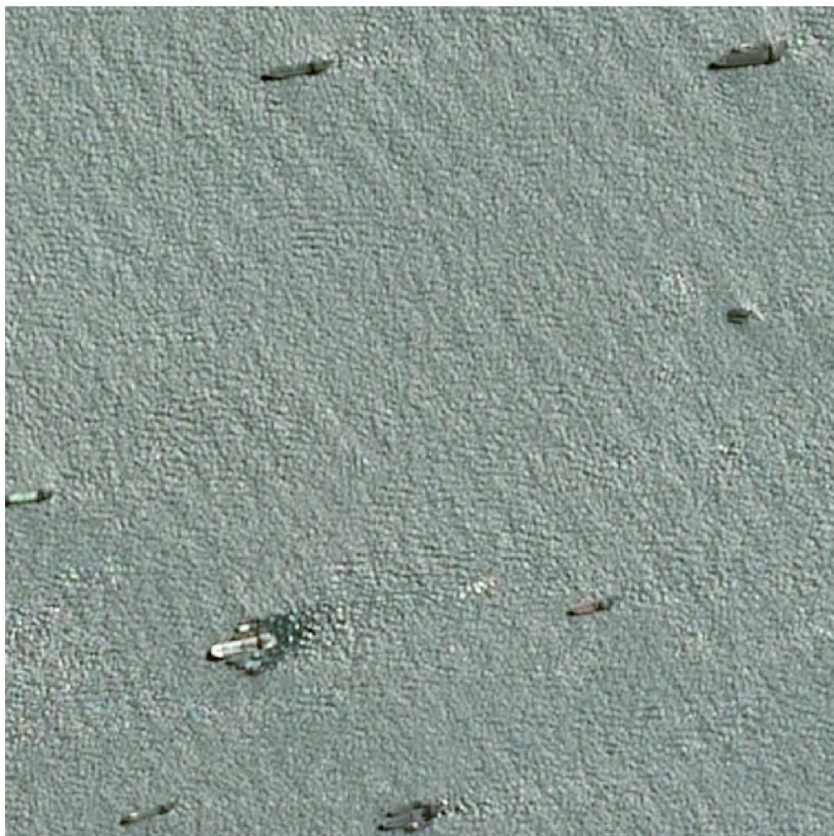


Хэши-беляши: kaggle Airbus lea**k**

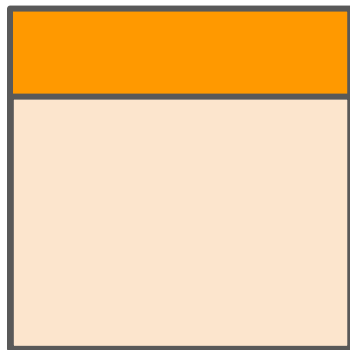
Артур Кузин
Lead Data Scientist, Dbrain

Постановка задачи: Instance segmentation



Размер датасета

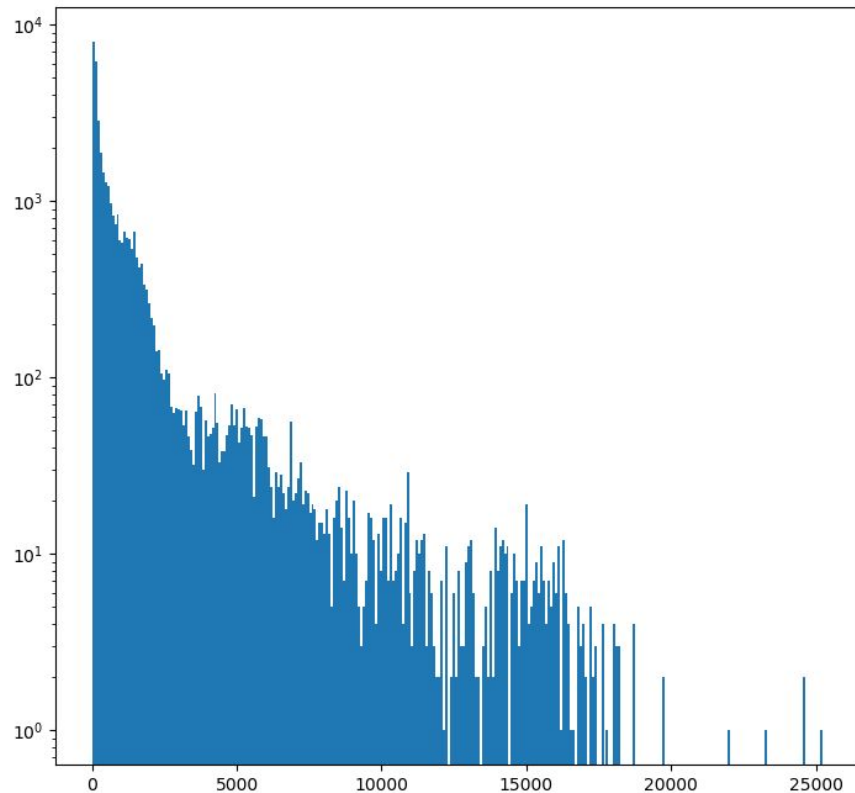
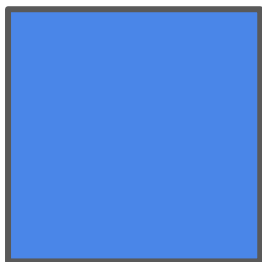
Train
104k



нет
кораблей

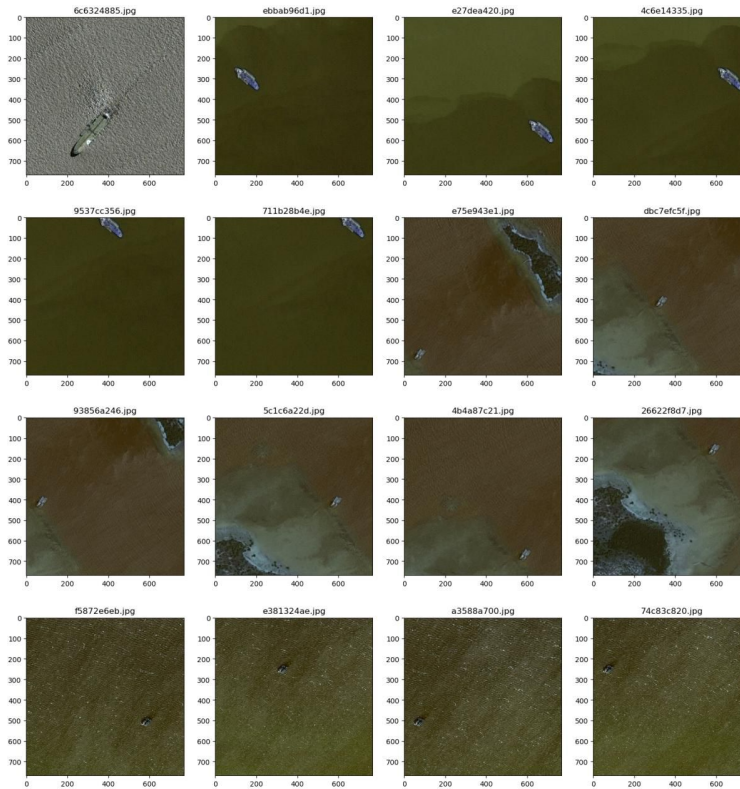
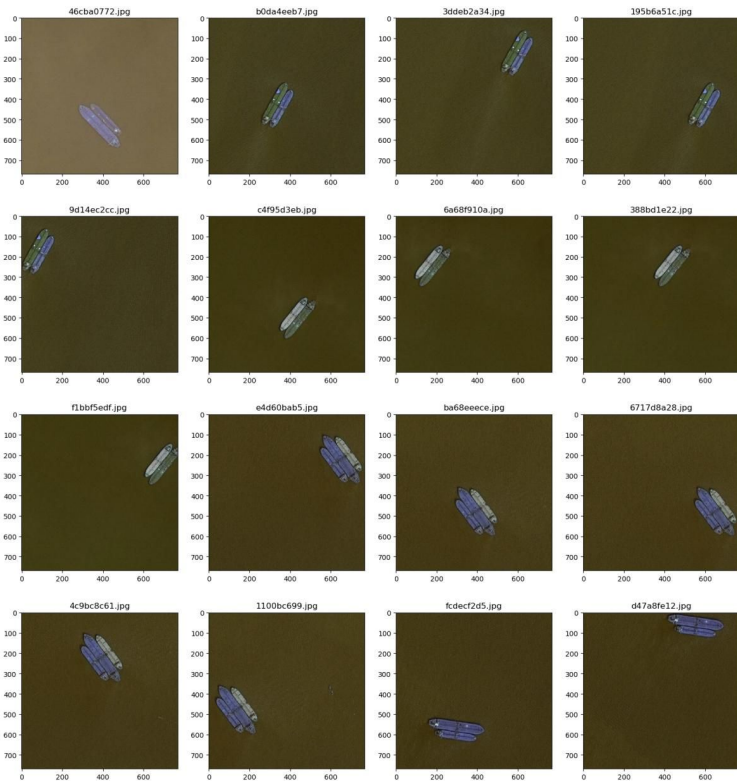
768 x 768

Test
85k



EDA

```
df.sort_values(['min_area', 'max_area'])
```



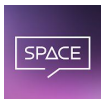
Изобразительные лики



Обучать модель на
репрезентативных
фолдах с учетом
дубликатов



Искать ответы теста в
трейне по SIFT'ам и
хэшам-беляшам



<https://youtu.be/Twli7zoiIPs>



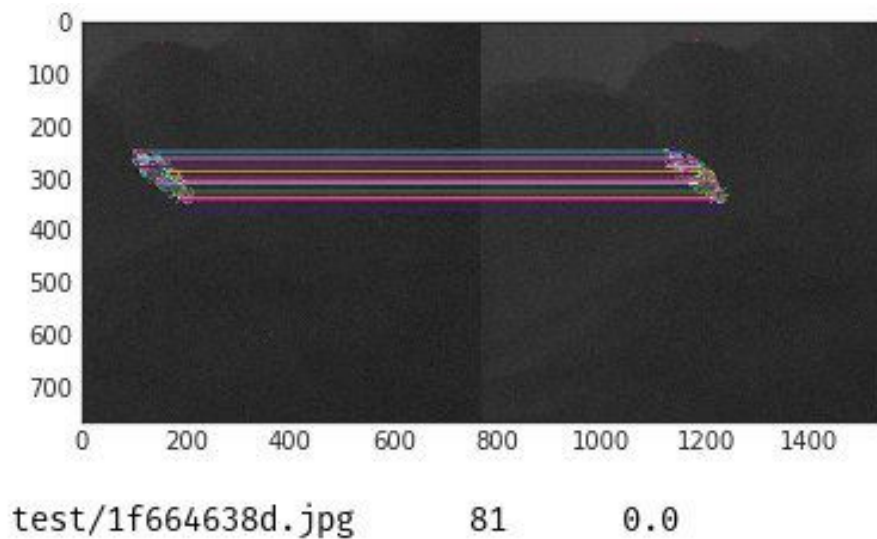
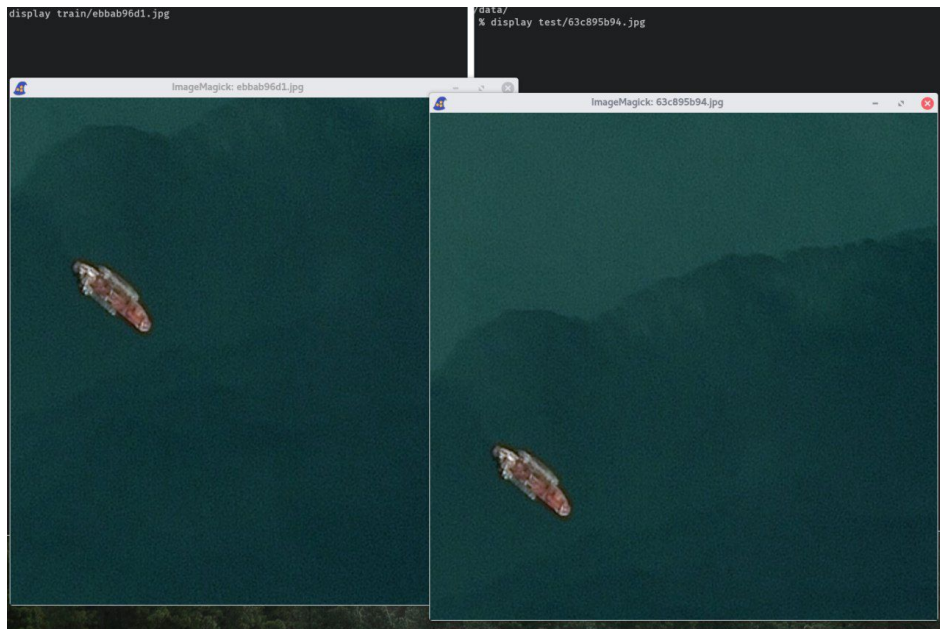
Konica-Minolta-1



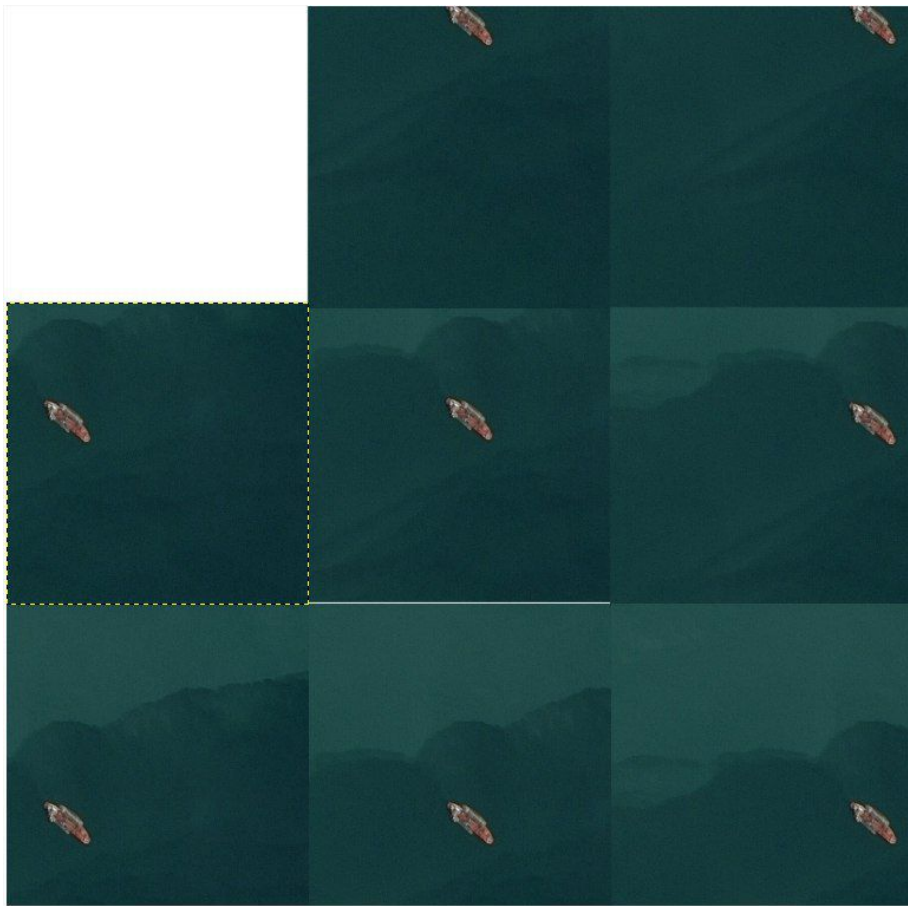
N+1 fish



Поиск по SIFT'ам



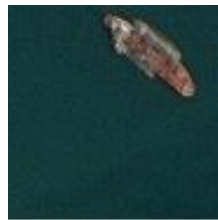
5 - train, 3 - test



Annoy: test 2 train



768x768



256x256



32x32



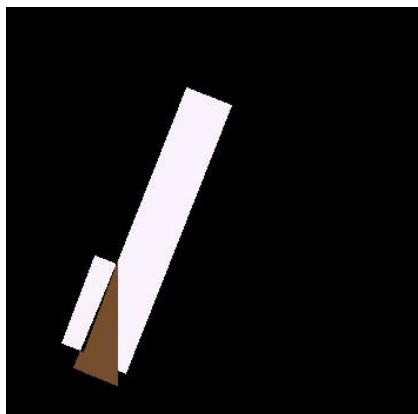
3096



github.com/spotify/annoy

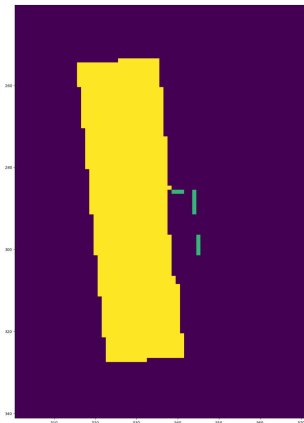


Небольшие технические трудности



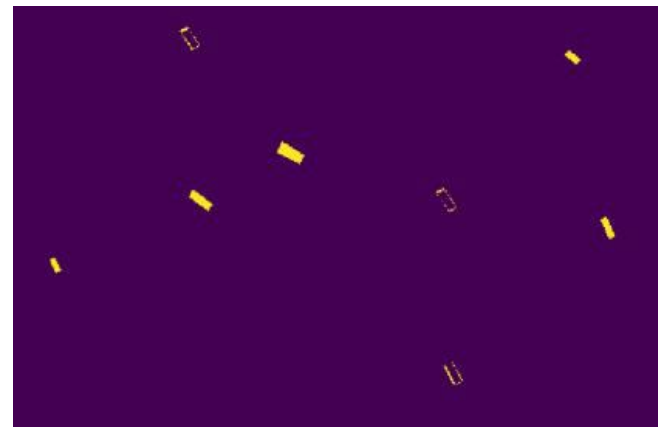
Name

subm3_belyash6_all.csv



Submitted

a few seconds ago



Wait time

3 seconds

Failed

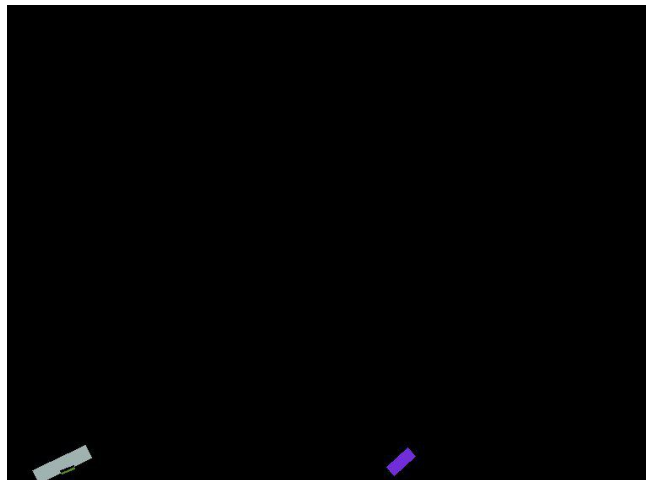
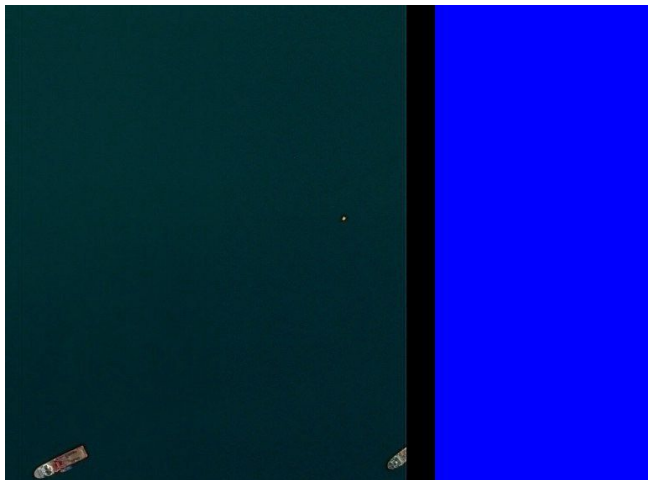
2 Exceptions:

The same pixel may not be assigned to two different objects.

The same pixel may not be assigned to two different objects.



Оракул и silent режим



Submission and Description

Public Score

[mix2.csv](#)

0.494

3 minutes ago by [Evgeny Nizhibitsky](#)

[add submission details](#)

[mix1.csv](#)

0.499

6 minutes ago by [Evgeny Nizhibitsky](#)

[add submission details](#)



Andres Torrubia strikes again!



Andres Torrubia

Severe leak

posted in [Airbus Ship Detection Challenge](#) 3 days ago



96

Hi all,

As you may have seen we got 0.98+ LB with just one sub. Although we had a lot of good ideas for the challenge my partner Pavel found a leak which renders this competition almost useless.

The images in test are just shifted versions of images in train. The problem also happens in train, it looks like airbus had bigger images and then the 768x768 are random crops of the bigger images; but it looks they didn't check whether there were any overlaps.

e.g. train/dd38da47f.jpg and train/02776139a.jpg

How to find:

- Run nearest neighbors on all images
- For each image take N closest neighbors and find where it overlaps

1

▲ 4

Pavel Yakubovskiy



0.994

18

41m

Перезапуск



inversion

Data Leak and Next Steps

posted in [Airbus Ship Detection Challenge](#) 3 days ago



94

First and foremost, let me express my sincere apologies for the frustration this situation has caused. I'll provide a full explanation of what happened below, but first want to give a high-level summary of the next steps.

- A new set of Test images will be provided for the competition in a few weeks
- In the meantime, the competition will still be "open" in the sense that the current data will be available and you can submit to the leaderboard
- To ensure everyone is on an even playing field, the current Test masks will be released (timing: Wed, Aug 29, by EOD)
- Once new test images are collected, we'll "reset" the data by moving current Test images into the Train folder, upload the new Test data, and reset the leaderboard
- Once the new data has been uploaded, we will extend the competition deadline to account for the delay

