

Детектирование объектов на спутниковых снимках

1. The 2nd Tellus Satellite Challenge
1st place -- 9k\$
2. DIUx xView 2018 Detection Challenge
1st place -- 28k\$

Сергиевский Николай

The 2nd Tellus Satellite Challenge

Площадка: <https://signate.jp/> / <https://signate.jp/competitions/153>

Timeline: 18.01.2019 - 14.02.2019

59 competitors

Rank	Username	Public	Private	Submission(s)	Submitted
1	Nick Sergievskiy	0.56657	0.47559	31	2019-02-14 17:58:02
2	peterzhang	0.55917	0.45325	69	2019-02-14 12:36:02
3	starfish	0.46261	0.41213	50	2019-02-14 22:59:02
4	tascj	0.42856	0.40918	31	2019-02-14 19:40:02
5	katsuntu	0.26607	0.33826	6	2019-02-13 08:05:02

Описание задачи

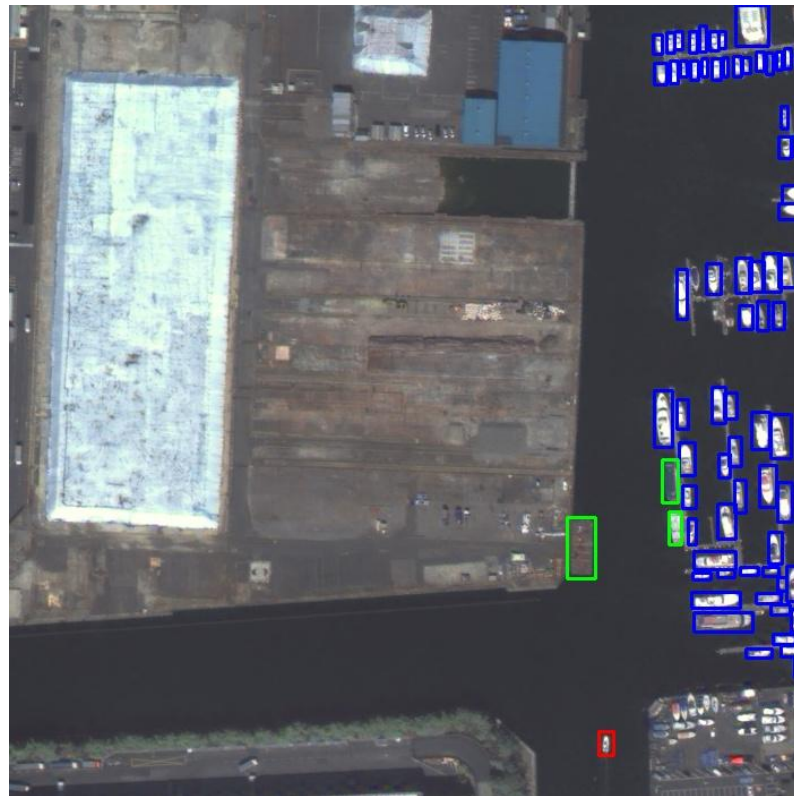
- Детектирование объектов
- 3 класса
- Разрешение спутниковых снимков: 50cm per 1px
- Метрика: самурайский mAP
 - $\text{IoU} > 0.6$
- Тренировочный набор:
 - 20 изображений
 - разрешение 15Kx15K (в среднем)
 - ship_not_moving 6245, barge 2356, ship_moving 224
- Тестовый набор:
 - 21 изображение



БОЛЬШЕ статистики!

	num	min (5% quantile)	max (95% quantile)
ship_not_moving	6245	7	56
barge	2356	17	88
ship_moving	224	12.35	101.3

- Объектов меньше 16 пикселей -- 2346
- Соотношение сторон
 - $0.5 < ar < 2$ -- 3122
 - $ar < 0.4$ OR $ar > 2.5$ 1979+513=2492
 - $ar < 0.33$ OR $ar > 3$ -- 915



Подготовка данных

- Разбить тренировочный набор на 5 фолдов
- Нарезать на фрагменты NxN с перекрытием
 - 700x700 с перекрытием 80 пикселей
 - Игнорировать объекты площадь которых стала меньше 30%
 - Пропускать пустые изображения
- Аугментация (офлайн)
 - D4 (поворот на [0, 90, 180, 270])

В итоге:

- Train -- 3800
- Val -- 1200

Post Processing

2xTTA as single

1. фрагменты 700x700 без перекрытия
2. фрагменты 750x750 без перекрытия
 - 2.1. поворот 90 градусов
3. Объединение через NMS (или голосование)

Выбор модели детектирования

Зоопарк подходов

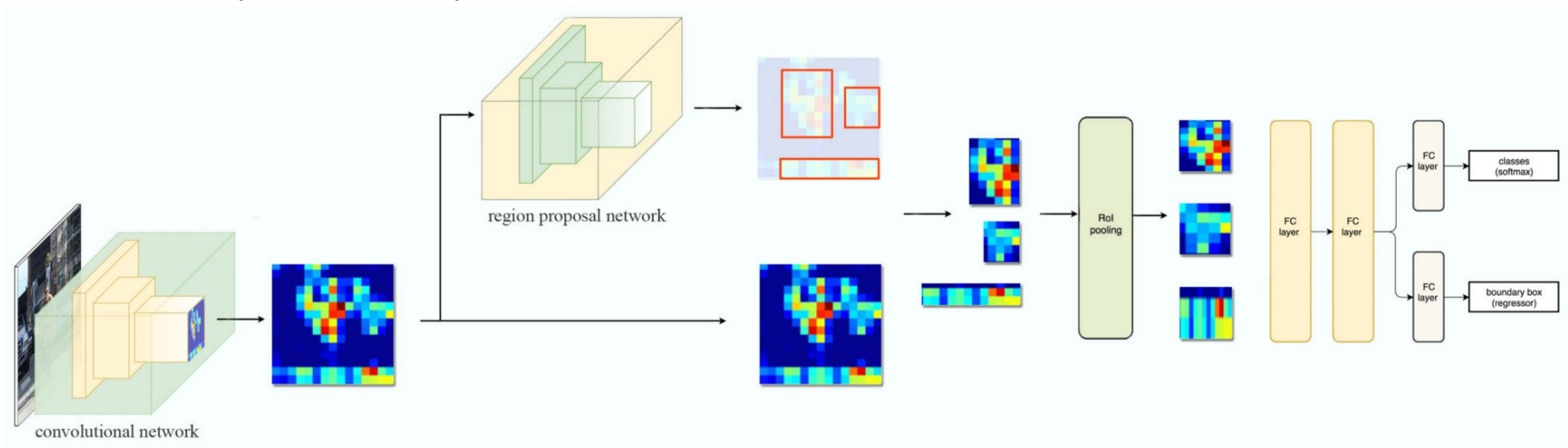
- Faster RCNN
 - MaskRCNN
- R-FCN
- SSD
- YOLO
- RetinaNet

Полезные навороты

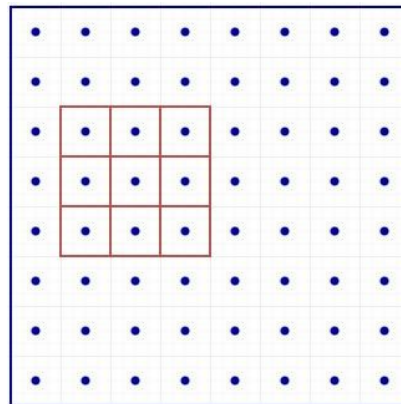
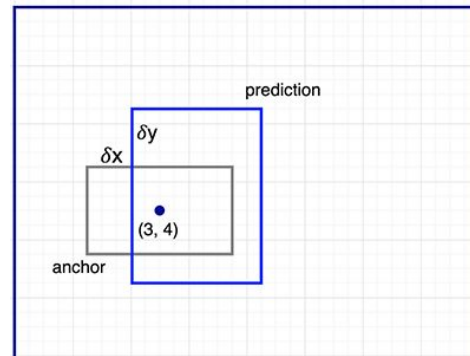
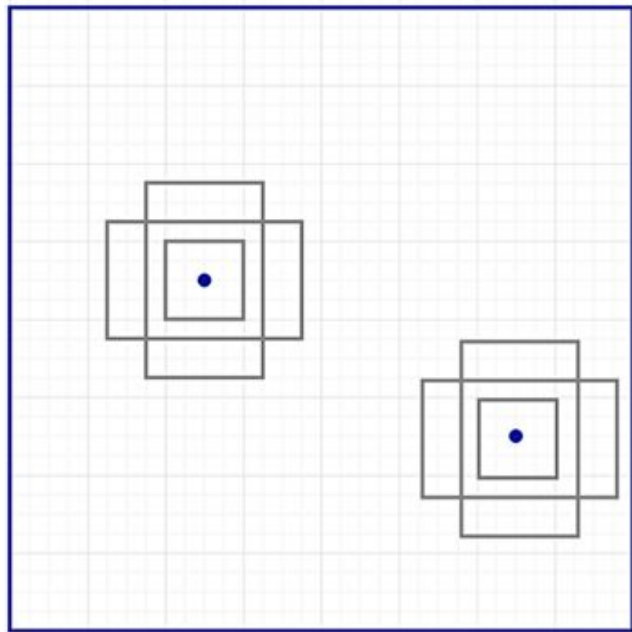
- FPN
- DeformableConvolution
- Cascade (RCNN)
- GN or SyncBN

Faster-RCNN

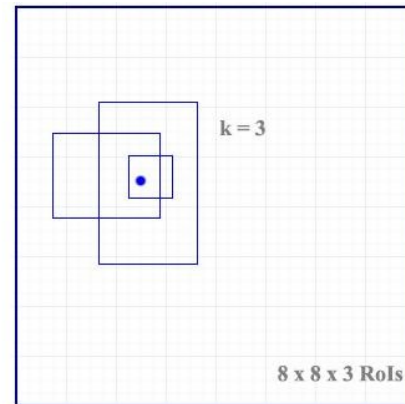
1. Feature Generation. Backbone
2. RPN
3. Head (Fast RCNN)



anchors and RPN



8 x 8 feature maps

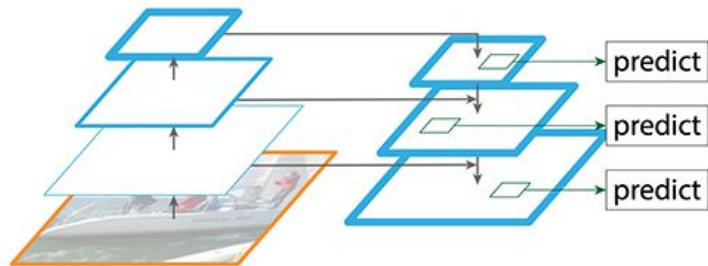


8 x 8 x 3 RoIs

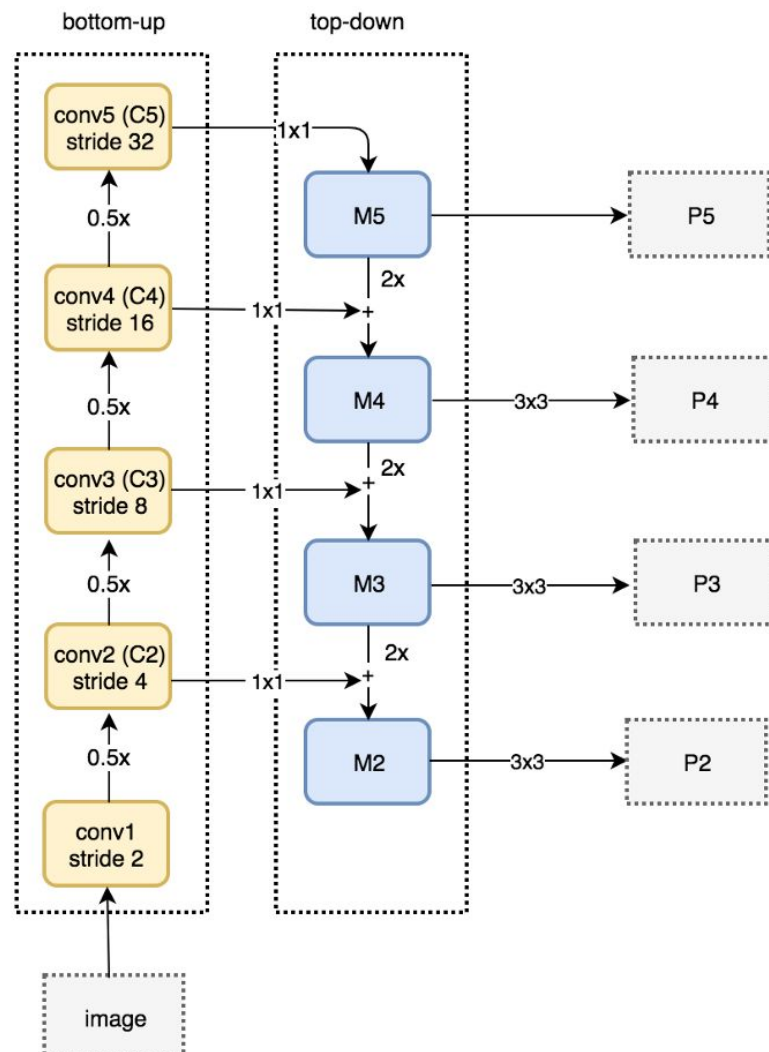
3 proposals for each location

Теория.

FPN



ResNet



CascadeRCNN

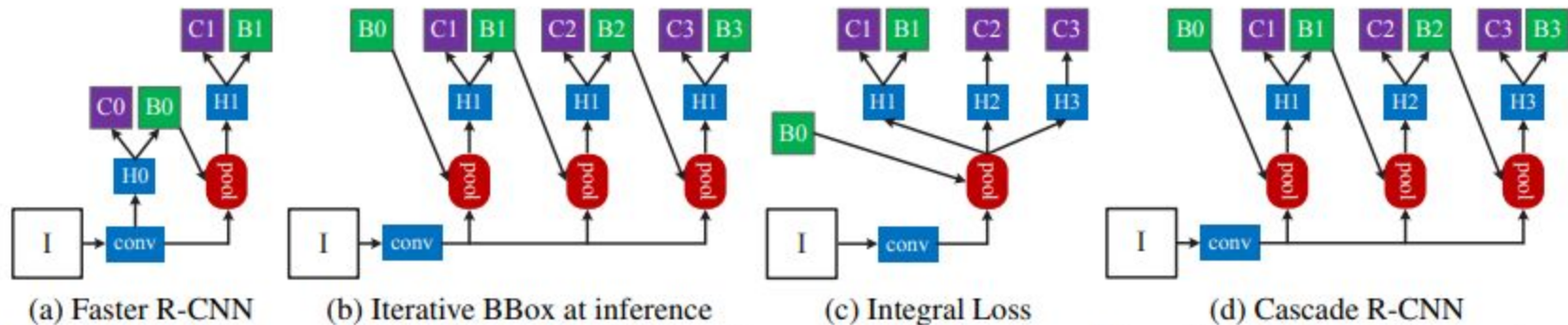


Figure 3. The architectures of different frameworks. “I” is input image, “conv” backbone convolutions, “pool” region-wise feature extraction, “H” network head, “B” bounding box, and “C” classification. “B0” is proposals in all architectures.

Library







Library: <https://github.com/open-mmlab/mmdetection>

- PyTorch 0.41 and PyTorch 1.0
- FPN
- MaskRCNN
- CascadeRCNN
- Deformable
- Retina
- GN



COCO
Segmentation 2018 AP 

 MMDet	0.490
 Megvii (Face++)	0.488
 FirstShot	0.470
 UCenter	0.467

Baseline

FPN FasterRCNN / Resnet18

(3.5h -- train time)

- anchor_size=32
- scales from 650 to 1200 (im_size=700)
- flip (horizontal)
- color jittering `brightness_delta=32,`
`contrast_range=(0.5, 1.5),`
`saturation_range=(0.5, 1.5),`
`hue_delta=18),`

Results

0.32372 | 0.32958 (Public LB | Private LB)

0.328 {'ship_not_moving': 0.214, 'ship_moving': 0.172, 'barge': 0.598} -- IoU=0.6

0.465 (+0.137) --IoU=0.5

Model v2

FPN CascadeRCNN / Resnet50

(7h -- train time)

- anchor_size=16

Results

0.49558 | 0.41298 (Public LB | Private LB)

0.498 {'ship_moving': 0.314, 'barge': 0.737, 'ship_not_moving': 0.441}

Model v3

FPN CascadeRCNN / Resnet50

(8h -- train time)

- `anchor_ratios=[0.33, 0.5, 1.0, 2.0, 3.0],`

Results

0.51190 | 0.45238 (Public LB | Private LB)

0.542{'ship_not_moving': 0.536, 'ship_moving': 0.554, 'barge': 0.537}

TTA

Основной TTA 3x2TTA.

- 600, 650 -> 1000
- 700, 750 -> 1000, flip
- 800, 850 -> 1000

Results

0.52284 | 0.48825 (Public LB | Private LB)

Ensemble

1. ResNet50_dconv
 anchor_ratios=[0.33, 0.5, 1.0, 2.0, 3.0]
2. ResNet50
3. anchor_size=32
4. anchor_ratios=[0.33, 0.5, 1.0, 2.0, 3.0]
5. ResNeXt101

* default anchor_size=16 for other models

Results

0.56657 | 0.47559 (Public LB | Private LB)

Выводы

- Scale jittering +
- Аугментация +
- Cascade +
- для маленьких объектов `anchor_size=16` -- +1%
- Метрика `NMS=0.7` -- +1.5%
- `Anchor aspect ratio` -- +2-4%
- Triks: `conf*0.8`, для объектов на краю +0.4%

Overfit:

- Resnet50_dconv -5.8% на Private LB
- ResNeXt -6.2% на Private LB
- ТТА одной модели на Private LB лучше чем финальный ансамбль +1.3%
- Ансамбль только ResNet50 +1.9% (0.54764 | 0.49449)
-

Должно помогать но не успел:

- Претренеровка на другом датасете
 - airbus +4%
- Одновременно учить на нескольких датасетах (xView, DOTA, airbus)

xView: Objects in Context in Overhead Imagery

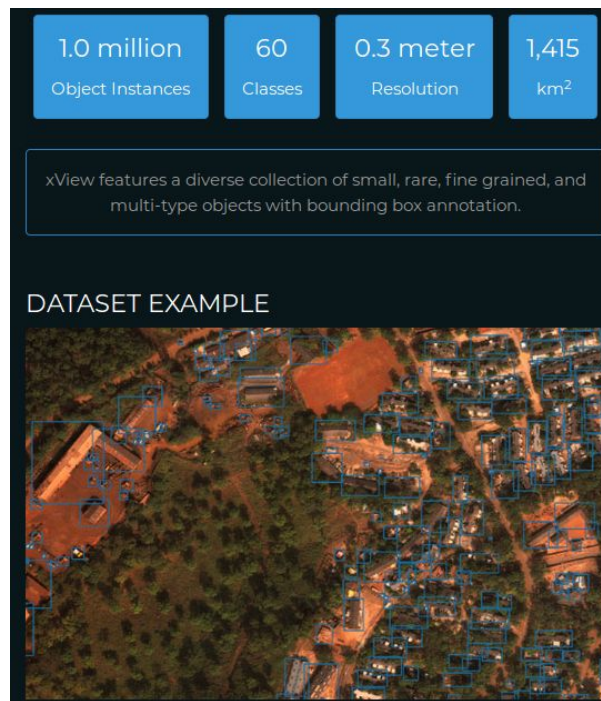
<http://xviewdataset.org/>

<https://challenge.xviewdataset.org>

Timeline: 9.05.2018 - 17.08.2018

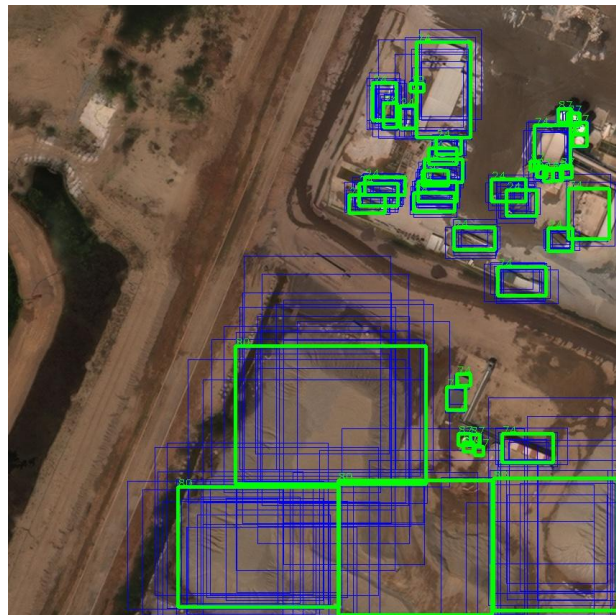
<https://arxiv.org/abs/1802.07856>

xView: Objects in Context in Overhead Imagery



Описание задачи

- Детектирование объектов
 - 60 классов
 - Метрика: mAP
 - $\text{IoU} > 0.5$
 - Тренировочный набор:
 - 742 изображений
 - разрешение 3Kx3K (в среднем)
 - Тестовый набор:
 - 104 изображение
-
- Несбалансированные классы 100 vs 200 000
 - Аппроксимированная семантическая сегментация



Baseline

Official model:

- SSD tf (detectionAPI)
- Backbone: resnet50
- 300x300 crops

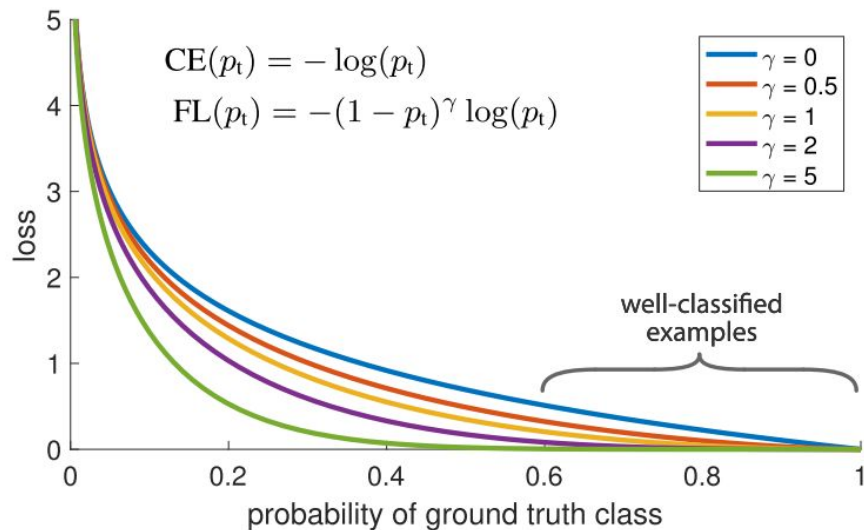
FPN FasterRCNN / Resnet50 (18h -- train time)

- anchor_size=[24, 32]
- scales from 500 to 900 (im_size=700)
- flip (horizontal)
- Color Jittering

	Additional anchors (scale 1.2x)	Baseline (SSD)
11_Fixed-wing_Aircraft	52.222	12.18
12_Small_Aircraft	64.306	37.71
13_Cargo_Plane	78.433	66.91
15_Helicopter	46.535	58
17_Passenger_Vehicle	1.229	4.71
18_Small_Car	62.432	40.83
19_Bus	43.524	37.73
20_Pickup_Truck	7.356	0.78
21_Utility_Truck	7.571	28.46
23_Truck	14.993	4.69
24_Cargo_Truck	20.659	9.72

Focal Loss

Retina Net: ResNet50 FPN with FocalLoss



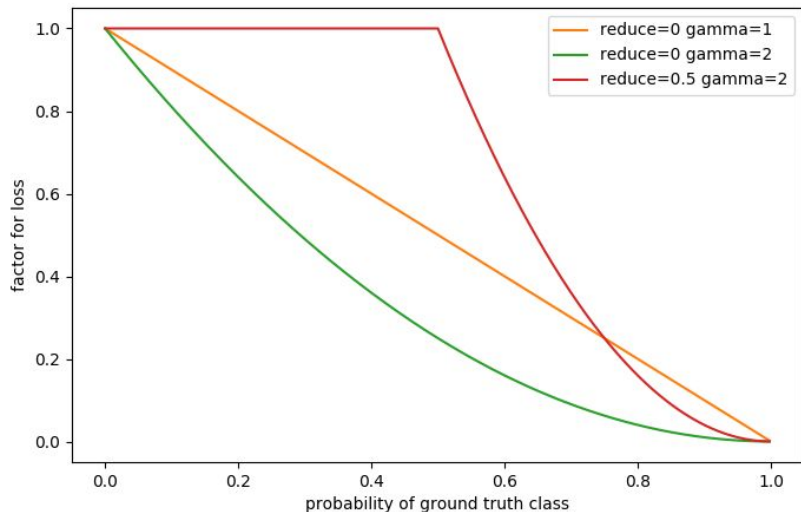
Model name	mAP
baseline (SSD)	21.78*
FPN-50	24.37
FPN-50 additional anchors	25.62
Retina Net	9.70

2 Stage Focal Loss

- Увеличить количество гипотез для RPN и Head (x2)
- Применить Focal Loss для RPN (Sigmoid)
- Применить Focal Loss для Head (Softmax)

Model name	mAP
baseline (SSD)	21.78*
FPN-50	24.37
FPN-50 additional anchors	25.62
Retina Net	9.70
FPN-50 Focal loss	19.01

Reduced Focal Loss



$$f(x) = \begin{cases} 1 & : pt < th \\ \frac{(1 - pt)^\gamma}{th^\gamma} & : pt \geq th \end{cases}$$

Model name	mAP	Recall	mRecall
baseline (SSD)	21.78*		
FPN-50	24.37	0.762	0.672
FPN-50 additional anchors	25.62	0.774	0.685
Retina Net	9.70		
FPN-50 Focal loss	19.01	0.43	0.66
FPN-50 Reduced Focal loss	26.92	0.542	0.741

IoU Reduced Focal loss random skip most frequent

- IoU Reduced Focal Loss

$$K(iou) = \begin{cases} 1 & \text{if } iou \leq th_{iou} \text{ low} \\ |aiou + b| & \text{if } iou > th_{iou} \text{ low and } iou > th_{iou} \text{ high} \\ K_{high} & \text{if } iou = th_{iou} \text{ high} \end{cases}$$

- Random skip most frequent classes

Model name	mAP	Recall	mRecall
baseline (SSD)	21.78*		
FPN-50	24.37	0.762	0.672
FPN-50 additional anchors	25.62	0.774	0.685
Retina Net	9.70		
FPN-50 Focal loss	19.01	0.43	0.66
FPN-50 Reduced Focal loss	26.92	0.542	0.741
FPN-50 IoU Reduced Focal loss	27.44	0.554	0.748
FPN-50 IoU Reduced Focal loss random skip most frequent	28.32	0.517	0.775

Ensemble

1. масштаб: 1.1 модель: "scale_anchors", поворот 90 градусов
2. масштаб: 0.9 модель: "focal_v10"
3. масштаб: 0.8 модель: "focal_v10", поворот 90 градусов
4. масштаб: 1.2 модель: "baseline_tf"
5. масштаб: 1 модель: "baseline_tf"
6. масштаб: 0.7 модель: "baseline_tf"
7. масштаб: 0.6 модель: "baseline_tf"

Public LB mAP= **0.317438** { "map/medium": "0.352039", , "map/large": "0.401615",
"map/common": "0.303794", "map/rare": "0.346584", "f1": "0.040269", "map/small":
"0.232972" }

Спасибо

Reduced Focal Loss: 1st Place Solution to xView object detection in Satellite Imagery

<https://arxiv.org/abs/1903.01347>