



## Задача «Разработка модели обнаружения для деперсонализации объектов на изображениях»

### Описание:

В современных условиях пошатнувшейся информационной безопасности пользователи различных сервисов ежедневно подвержены риску утечки персональных данных. Специалисты по кибербезопасности InfoWatch заявили, что в 2021 году они зафиксировали 1729 эпизодов утечек данных по всему миру. В результате утечек оказались скомпрометированными около 8,5 млрд записей персональных и платежных данных. Несколько месяцев назад в нашей стране широкую огласку получила массовая утечка данных пользователей популярной службы доставки еды.

Мировой и отечественный опыт показывает, что утечка персональных данных является буквально неизбежным явлением в современном цифровом мире. В России более 13 млн камер по всей стране ежедневно обрабатывают персональные данные миллионов граждан (в частности, изображения лиц и государственные регистрационные номера транспортных средств). Таким образом, нарастает актуальность в постоянном хранении данных в деперсонализированном виде. Деперсонализация — процесс удаления связи между субъектом данных и совокупностью идентифицирующих его данных.

Участникам чемпионата предстоит разработать модель обнаружения изображений лиц и силуэтов людей, а также номеров автомобилей. Решение задачи позволит обезличить персональные данные граждан и не допустить их использование злоумышленниками.

### Условия задачи:

Участникам необходимо научиться детектировать на изображениях следующие объекты: головы, людей, автомобили, лица, автомобильные номерные знаки.

Train содержит 2002 изображения, test – 855. В train находятся 2 папки: images и labels. В папке labels находятся .txt файлы разметки. В названии .txt файла явно указан класс. Файл разметки формата .txt содержит в себе данные о bbox в формат yolo.

Подробнее ознакомиться можно здесь —  
<https://github.com/ultralytics/yolov5/issues/2293#issuecomment-785534291>



## Метрика:

Решения участников оцениваются путём вычисления mean Average Precision (mAP@.5). Итоговое значение вычисляется как среднее по всем классам. При отправке решений необходимо соблюдать следующее соответствие классов:

car': 0, 'head': 1, 'face': 2, 'human': 3, 'carplate': 4

Пример отправки решений – файл simple\_solution.csv. Обязательно должен содержать следующие поля: 'ImageID', 'LabelName', 'Conf', 'XMin', 'XMax', 'YMin', 'YMax'

Подробнее с метрикой можно ознакомиться:

[https://storage.googleapis.com/openimages/web/evaluation.html#object\\_detection\\_eval](https://storage.googleapis.com/openimages/web/evaluation.html#object_detection_eval)

<https://github.com/ZFTurbo/Mean-Average-Precision-for-Boxes>

## Формулы:

$TP$  = True positive

$TN$  = True negative

$FP$  = False positive

$FN$  = False negative

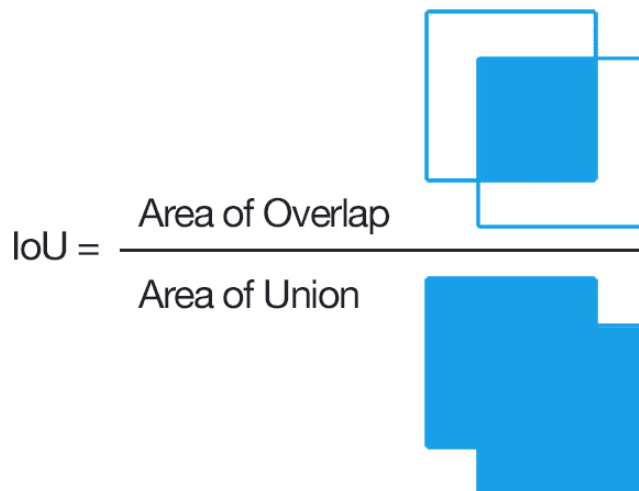
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$



IoU (Intersection over union) – перекрытие между двумя областями.


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Средняя точность (AP – Average Precision) рассчитывается как средневзвешенное значение precision для каждого порога, вес – это увеличение recall по сравнению с предыдущим порогом.

Mean Average Precision – это среднее значение AP для каждого класса.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i$$

### Дополнительно:

Для понимания организаторами статуса подготовки решения, необходимо, чтобы тизер решения был прикреплен в личный кабинет как минимум за неделю до завершения чемпионата. Обратите внимание, что тизер можно изменять в личном кабинете до конца соревнования.



### Правила чемпионата:

1. С момента открытия датасета до момента завершения приема решений репозиторий участника, в котором он ведет разработку по задаче текущего чемпионата, должен оставаться закрытым.
2. Участник обязан открыть доступ к репозиторию на чтение по ссылке (которая была прикреплена в ЛК в поле «Ссылка на код (гитхаб)»), не позднее чем в течение 12 часов с момента окончания дедлайна отправки решений на региональном чемпионате.
3. Согласно п. 5.8 Положения в процессе верификации решений организаторы и технические эксперты, проверяющие решения участников, в праве назначить интервью с участниками чемпионата. Участник получит приглашение и ссылку на интервью не позднее чем за 12 часов до публикации итогового лидерборда. Пропуск интервью участником является поводом для дисквалификации.
4. Организаторы вправе дисквалифицировать участника за непредоставление одного из артефактов решения задачи: тизера, скринкаста, презентации, ссылки на репозиторий, а также в случае выявления плагиата кода.
5. Участник, получивший 2 дисквалификации за сезон конкурса, попадает в чёрный список, с дальнейшим отстранением от участия в чемпионатах до конца сезона.
6. При загрузке решения участник подтверждает, что его решение соответствует условиям задачи и правилам проведения чемпионата.