

## РАЗРАБОТКА МЕТОДА ПРЕДСКАЗАНИЯ ТОЧНОСТИ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ДЕТЕКЦИИ ОБЪЕКТОВ

*Великанов Максим Сергеевич*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: maksim.velikanov@graphics.cs.msu.ru*

*Научный руководитель — Ватолин Дмитрий Сергеевич*

Автоматические средства анализа видео являются неотъемлемым компонентом модернизации процессов, для которых все еще необходимо присутствие человека. Например, встроенные видеодомофоны стали использоваться в системах безопасности для контроля общественного порядка и охраны объектов, для оптимизации строительных работ, отслеживания грузоперевозок, удаленного медицинского обслуживания. Просмотр больших объемов видео — утомительное занятие для человека, поэтому на практике применяются алгоритмы 1) обнаружения (детекции) объектов и 2) распознавания объектов, такие как распознавание лиц, автомобильных номеров.

С ростом спроса на автоматизацию обработки видеопотока становятся актуальными специализированные алгоритмы сжатия видео, ориентированные на машинную обработку. Для построения таких алгоритмов необходима метрика качества, оценивающая, насколько сильные искажения внес в изображение алгоритм сжатия. В классических видеокодеках применяются метрики PSNR, SSIM [1], но результат этих метрик плохо коррелирует с точностью обнаружения объектов. В данной работе предлагается алгоритм оценки качества изображения, предсказывающий точность алгоритмов детекции объектов.

Для экспериментов был выбран детектор объектов YOLO [2], так как он имеет высокую скорость работы и используется на практике. В качестве набора данных взят Microsoft Common Objects in Context 2017 validation [3], часто используемый для сравнения методов детекции. В подготовительную работу входило сжатие этого набора данных разными видеокодеками с разными параметрами качества, для обучения модели и анализа влияния сжатия на качество детекции. Для сжатия использовались зарекомендовавшие себя стандарты сжатия AVC (H.264), HEVC (H.265), VVC (H.266), AV1, JPEG, с 4 параметрами качества. Количество исходных изображений — 3125, сжатых — 65625.

На созданном наборе данных с 20 вариантами сжатия был запу-

шен YOLO, для каждого объекта из истинной разметки было подсчитано качество его детекции по метрике Intersection over Union. Разработанная модель обучалась на предсказание данного показателя.

Модель принимает на вход 2 изображения: исходное и сжатое некоторым кодеком, а также ограничивающую рамку объекта. На выходе предсказывается Intersection over Union детектора YOLO для данного объекта. Архитектура модели основывается на энкодере моделей классификации — VGG16, более быстрой ResNet18, либо на имеющемся энкодере YOLO. Для агрегации признаков для каждой интересующей рамки применяется Region of Interest pooling. Сравнивались несколько возможных архитектур модели, наилучшие корреляции предсказаний модели с истинными значениями IoU были достигнуты с энкодером YOLO — 0.57 по Спирмену, 0.38 по Пирсону.

### Иллюстрации

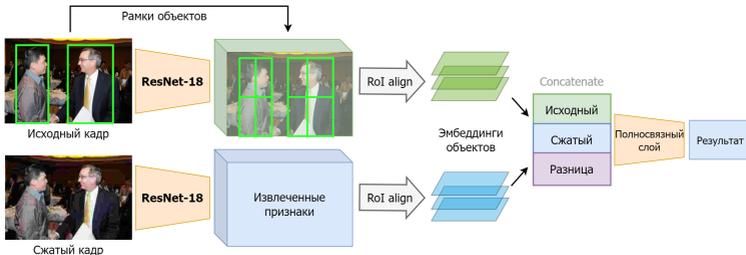


Рис. 1. Архитектура предложенного метода с операцией RoI pooling

### Литература

1. Wang Z., Bovik A., Sheikh H., Simoncelli E. Image quality assessment: from error visibility to structural similarity //IEEE transactions on image processing, 2004, P. 600–612.
2. Bochkovskiy Alexey., Chien-Yao W., Hong-Yuan M.. Yolov4: Optimal speed and accuracy of object detection //arXiv preprint arXiv:2004.10934, 2020.
3. Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick L. Microsoft coco: Common objects in context //European conference on computer vision, 2014, P. 740–755.