**Ministry of Science and Higher Education**

**of the Russian Federation**

**ITMO University**

Faculty of Digital Transformations

Educational program 01.04.02 Applied mathematics and informatics\_\_\_\_

Subject area (major) Big Data and Machine Learning\_\_\_\_

REPORT

on practical training Research Internship

Task topic: Investigation of the influence of environmental factors on the psychoemotional state of a person

Student: Yamoldin Alexander, J4134c

Head of Practice from the trainee’s host organization: Basov Oleg, National Center for Cognitive Development, Senior Researcherhead of the laboratory «Center Information Optical Technologies»

Head of Practice from ITMO University: Alexandra Klimova, Faculty of Digital Transformation, assistant professor

|  |  |
| --- | --- |
|  | Practice completed with grade \_\_\_\_  Commission member signatures:  \_\_\_\_\_\_\_\_\_\_\_\_\_\_ *full name*  (signature)  \_\_\_\_\_\_\_\_\_\_\_\_\_\_ *full name*  (signature)  \_\_\_\_\_\_\_\_\_\_\_\_\_\_ *full name*  (signature)  Date \_\_\_\_\_\_\_\_ |

St. Petersburg  
2022

**Annotation**

26 pages, 5 figures, 53 references

MACHINE LEARNING, CO2 PREDICTION, HVAC ENERGY CONSUMPTION, INDOOR AIR QUALITY, ENERGY EFFICIENTY, WELL-BEING

**Object of research**: Indoor air quality.

**Target of research:** The aim of the work is to create a model that provides a comfortable temperature and carbon dioxide level of indoor environment by using a combined method of ventilation.

**Methodology:** Application of Machine Learning and Artificial Intelligence for Dynamic Carbon Dioxide Level and Indoor Temperature Prediction.

**Results:** In this work, the approaches used to solve the problem of dynamic control of the indoor environment have been considered. In many articles there were found weaknesses, which should be considered in the future master's dissertation.

Оглавление

[List of Abbreviations and Designations 4](#_Toc106486814)

[1. Experiment and data collection 5](#_Toc106486815)

[1.1 Experiment 5](#_Toc106486816)

[1.2 Data collecting 7](#_Toc106486817)

[1.3 Data labeling 11](#_Toc106486818)

[2. Learning and understanding basic concepts of Computer Vision 14](#_Toc106486819)

[2.1 Метрики оценки 14](#_Toc106486820)

[2.2 Краткая теория свёрточных нейронных сетей 17](#_Toc106486821)

[2.3 Object Detection 21](#_Toc106486822)

[2.3 YOLO model 23](#_Toc106486823)

[2.4 Transfer learning 24](#_Toc106486824)

[2.5 Object tracker 25](#_Toc106486825)

# List of Abbreviations and Designations

# Experiment and data collection

## 1.1 Experiment

Экспериментальные данные было решено собирать в 408 лаборатории национального центра когнитивных разработок ИТМО. Общий вид лаборатории представлен на рис.1.

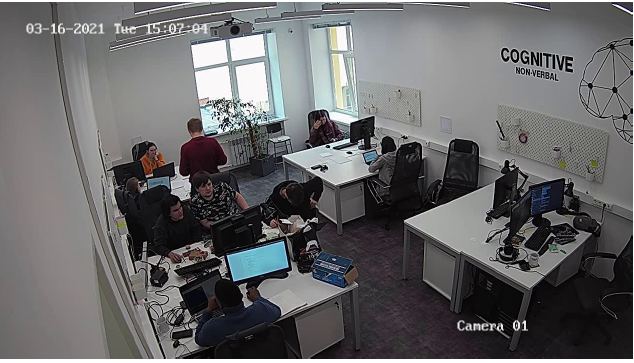


Рисунок 1. Лаборатория 408 национального центра когнитивных разработок ИТМО

Данная лаборатория оснащена двумя стационарными IP камерами, подвешенной диаметрально противоположно в ближнем левом и дальнем правом верхнем углу аудитории таким образом, чтобы не мешать и не раздражать присутствующих в аудитории испытуемых, но при этом иметь полный обзор пространства аудитории, а также положения людей с двух диаметрально противоположных сторон. В лаборатории статичный задний фон и рабочие места испытуемых, что позволяет впоследствии производить детекцию людей более простой для нейронной сети. Кроме того, данная аудитория оборудована датчиками измерения освещенности окружающей среды (## написать датчик) и датчиком измерения показателей окружающей среды Bosch BME680. Данный сенсор является цифровым 4 в 1 устройством, способным измерять загрязнённость воздуха, влажность воздуха, атмосферное давление и температуру окружающей среды [datascheet Bosch]. Сенсор Bosch требует предварительной калибровки на базовый уровень [datasheet bosch], поэтому перед началом работы датчик был откалиброван на базовый уровень загрязнённости лаборатории.

Для эксперимента были отобраны 6 испытуемых. Каждый испытуемый был проанализирован на предмет отсутствия хронических и иных заболеваний, которые бы могли помешать чистоте эксперимента. Кроме того, каждый испытуемый был проинформирован об его участии в эксперименте и дал письменное согласие на обработку персональных данных. Гендерное разбиение испытуемых в ходе экспериментов было в отношении 5 мужчин к 1 девушке (рис.2).

Picture 2. Gender distribution of persons tested

Средний возраст испытуемых составлял 24 года.

## 1.2 Data collecting

После всех приготовлений и настройки оборудования производились эксперименты. Всего было записано 17 экспериментальных дней: 16.03.2022 – 22.03.2022, 29.03.2022 и 01.04.2022 – 10.04.2022. Собираемые данные включали в себя записи с ближней и дальней IP камер видеонаблюдения (рис.3 и рис.4), а также данные с цифровых сенсоров: температура (С), сопротивление газа (ohm), относительная влажность (%), атмосферное давление (hPa), altitude (m), luminance window (lux), luminance door (lux). Пример фрейма данных представлен на рис. 5.

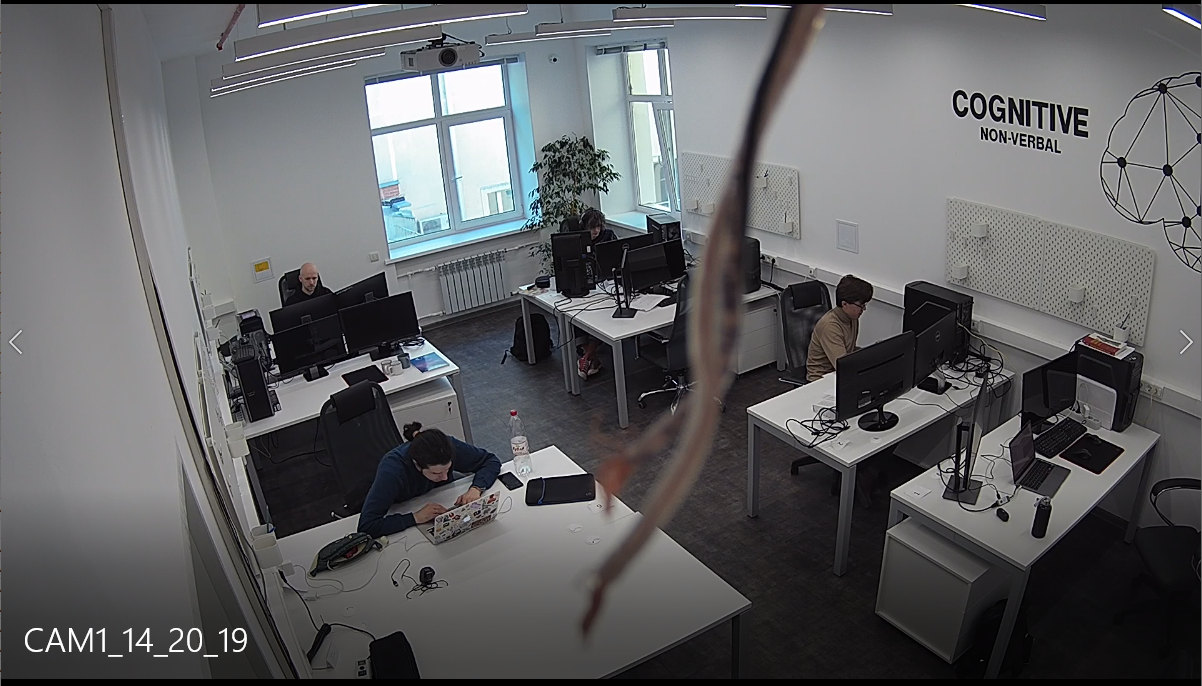


Рисунок 3. Кадр с ближней камеры

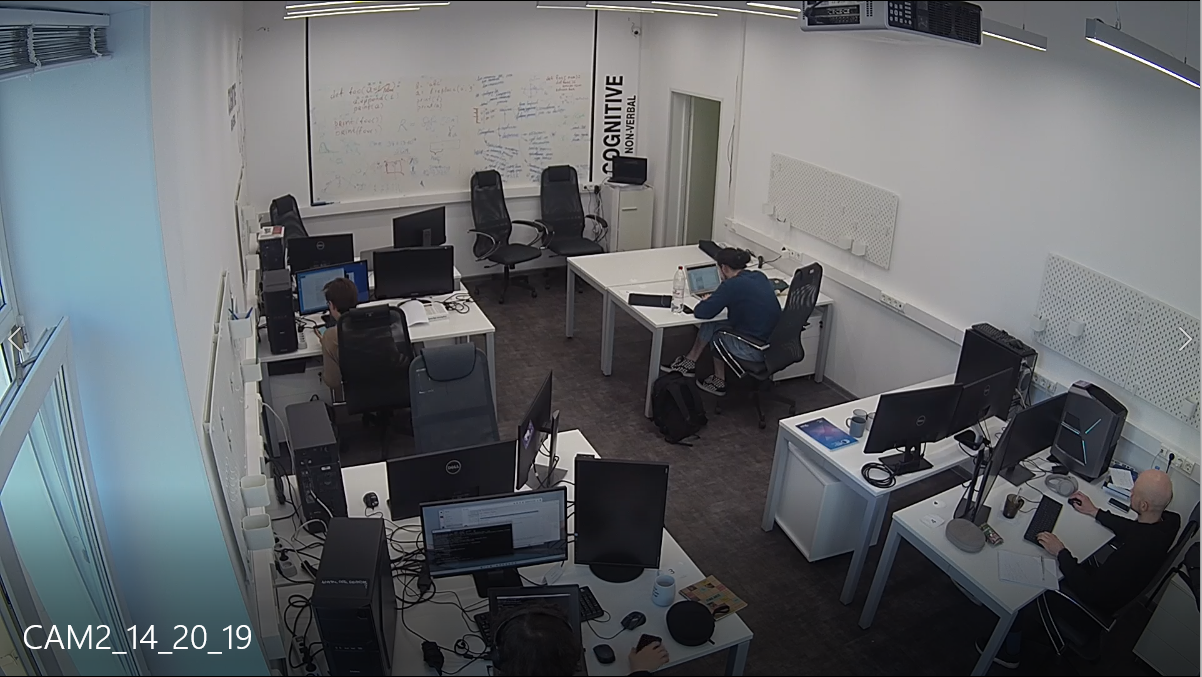


Рисунок 4. Кадр с дальней камеры 01.04.2022 14:20

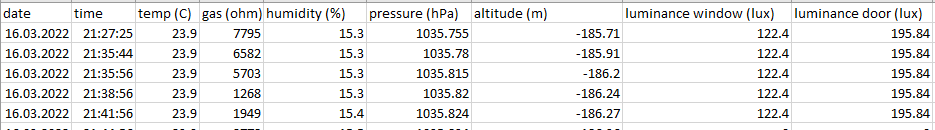
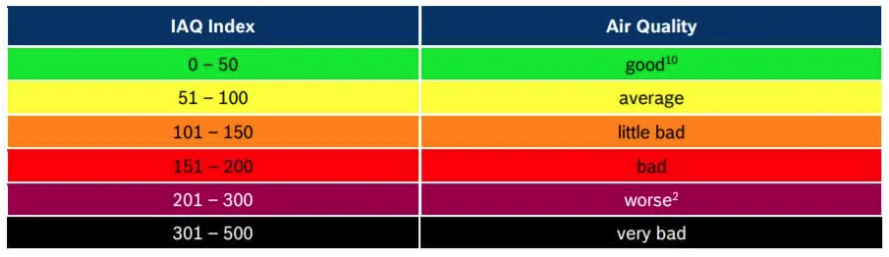


Рисунок 5. Пример фрейма данных, полученных с датчиков

Всего было собрано 129.5 часов видеоматериала формата .mp4, а также 55634 записи показателей сенсоров.

В данной работе предполагается оценка психоэмоционального стостояния человека в зависимости от факторов окружающей среды. Поэтому необходимо перевести показатели сенсоров датчика Bosch BME 680 к единому показателю Indoor Air Quality (IAQ). Опираясь на [https://ae-bst.resource.bosch.com/media/\_tech/media/datasheets/BST-BME680-DS001.pdf] the , the , the (Pic. ).



Перевод показателей, полученных с сенсора датчика в единицы IAQ [https://github.com/thstielow/raspi-bme680-iaq] происходит на основании закона уравнения идеального газа

|  |  |
| --- | --- |
|  | (0) |

где, – давление газа, Па;

– объем, ;

- the amount of substance of the gas, moles;

- the absolute temperature, K;

- the gas constant, in SI units .

И уравнения August-Roche-Magnus

|  |  |
| --- | --- |
|  | (0) |

где, – давление газа, kPa;

- the absolute temperature, K.

Then evaluate the maximum absolute water density

|  |  |
| --- | --- |
|  | (0) |

where, – is the temperature, C.

Then calculating the absolute humidity

|  |  |
| --- | --- |
|  | (0) |

where, hum – is relative humidity.

Using relative humidity influence of the present water vapor concentration on the gas resistance can be compensated. The dependency appears to be exponential. Thus, the bare VOC resistance is obtained by

|  |  |
| --- | --- |
|  | (0) |

where, – is the gas resistance, ohm.

The air quality is calculated as the ratio between and the ceiling value gas\_ceil, which is further squared for a steeper slope at higher air qualities and capped:

|  |  |
| --- | --- |
|  | (0) |

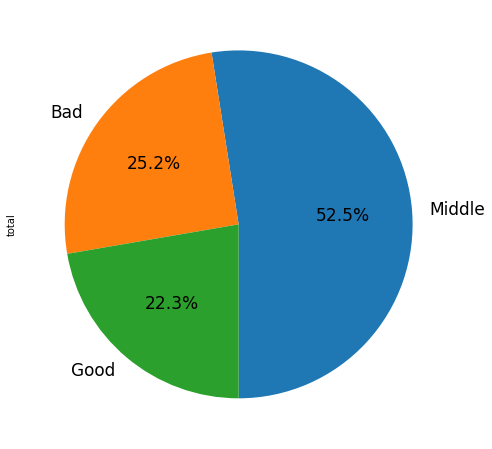
where, – max value for clear gas, ohm.

So can be measured

|  |  |
| --- | --- |
|  | (0) |

After that we can implement IAQ metrics on our dataset with the borders below for the , the , the we have an air quality distribution (pic.)

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 12398 | 29201 | 14032 |



Picture . Pie distribution of IAQ condition during the data collection

We can see that more than 50% of experimental data in collect on middle IAQ condition.

## 1.3 Data labeling

Для обучения нейронной сети нужно разметить данные. Нейронные сети могут работать только с изображениями. Работа с видео заключается по сути в покадровом дроблении видеозаписи на отдельные изображения и пропускание этих изображений сквозь слои нейронной сети. Для разбивки видео покадрово я использую open-source инструмент компании DVDVideoSoft. Сама же непосредственная разметка данных происходила на сервисе Roboflow Annotate. Поскольку данный сервис позволяет размечать изображения с помощью BBox, а также сразу их подготавливать в пригодный для YOLO формат.

В нашем исследовании будет проводится бинарное разделение эмоциональных состояний человека: положительные и отрицательные. Каждое из состояний характеризуется своим набором невербальных признаков, таким как жесты и позы. В различных психо-эмоциональных состояниях человек выделяет различные невербальные признаки при взаимодействии с предметами или другими людьми.

Положительные невербальные признаки и соответствующие им жесты и позы представлены в таблице:

Таблица . Положительные эмоциональные состояния человека и соответствующие им жесты.

|  |  |
| --- | --- |
| Эмоциональное состояние собеседника | Жесты, позы |
| Искренность | Раскрытые руки ладонями вверх |
| Обдумывание решения | Человек сидит на краешке стула, склонившись вперед, голова слегка наклонена и опирается на руку |
| Доброжелательность | Мягкая улыбка, небольшой наклон головы в сторону собеседника, выражение глаз |
| Заинтересованность | Взгляд искоса сопровождается слегка поднятыми бровями или улыбкой  Человек наклоняется в сторону собеседника (любезность, внимание) |

Отрицательные невербальные признаки и соответствующие им жесты и позы представлены в таблице:

Таблица . Отрацительные эмоциональные состояния человека и соответствующие им жесты.

|  |  |
| --- | --- |
| Эмоциональное состояние собеседника | Жесты, позы |
| Чувство собственной вины, напряженное восприятие ситуации | Руки спрятаны (за спину, в карманы) |
| Защита, оборона | Руки скрещены на груди |
| Усталость, эмоциональное или физическое напряжение | Зевок |
| Желание оградиться от окружение | ПТоза со скрещивающимися ногами |

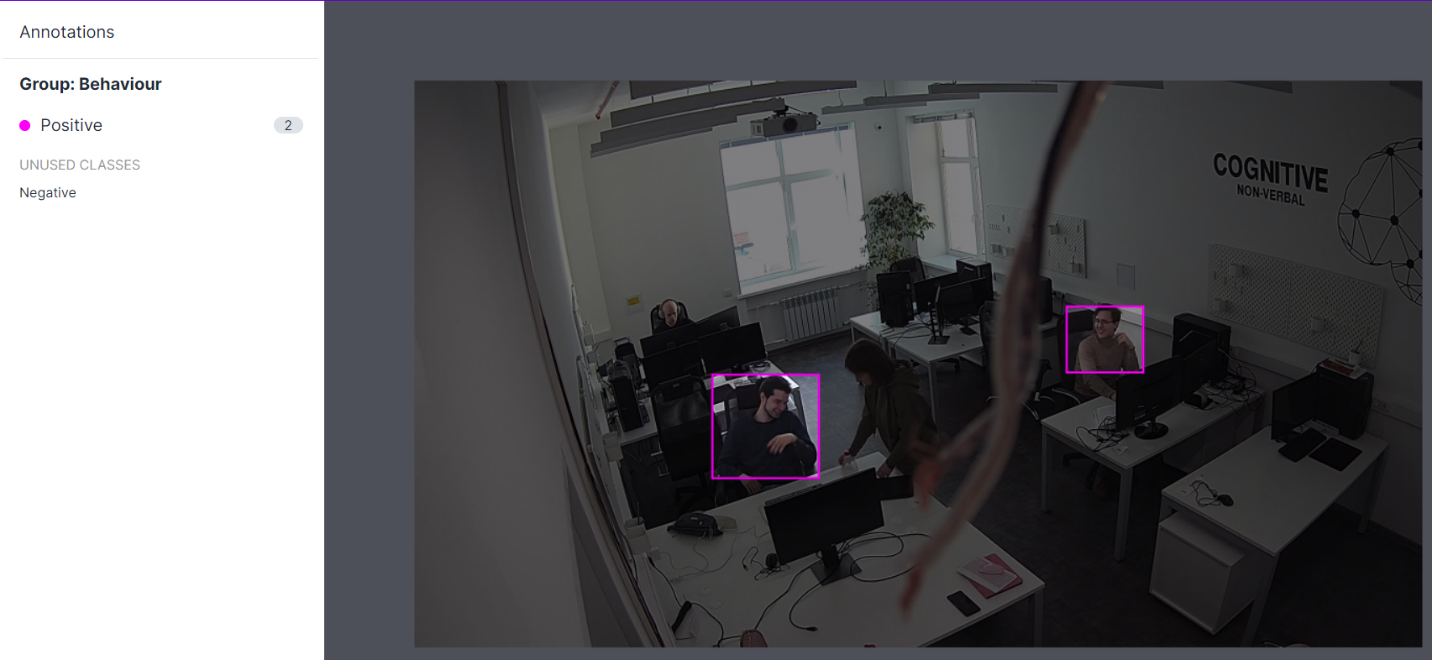
В качестве примера разметки положительного класса рассмотрим рис.





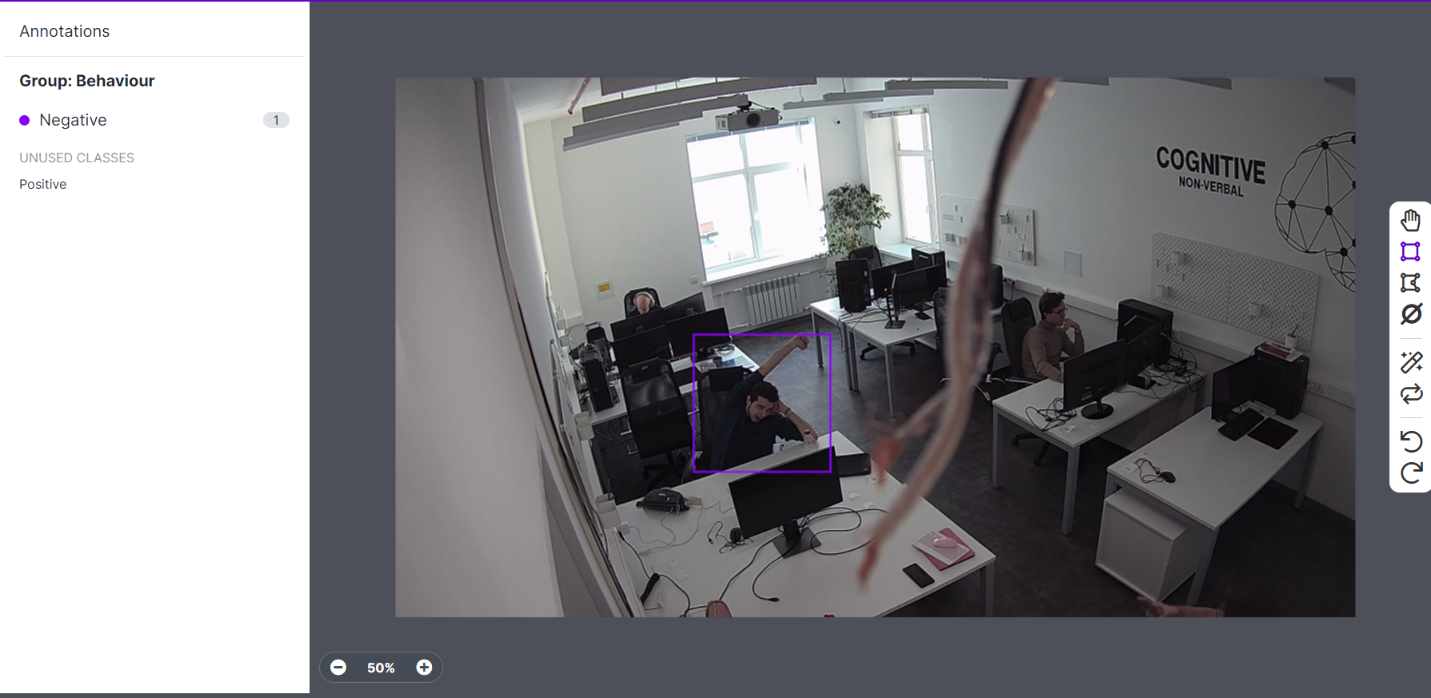
Picture – Positive emotions of people. Man 1 and 2 are smiling.

Таким образом мы можем отнести эмоции людей при ручной разметке к положительному классу рис.



Picture – Hand-labeling smiling people to positive class

В качестве примера разметки отрицательного класса рассмотрим рис.



Picture - Hand-labeling yawn people to negative class

Эмпирически выявлено [https://www.youtube.com/watch?v=M9ufLM2bvRM&list=PLVFGVo0DNh5ek970Foh\_9Dw2Anx2ps8it&index=4], что в среднем для обучения YOLO требуется от 500 фотографий на каждый паттерн. С учетом этого, для данной работы необходимо разметить вручную 4000 фотографий.

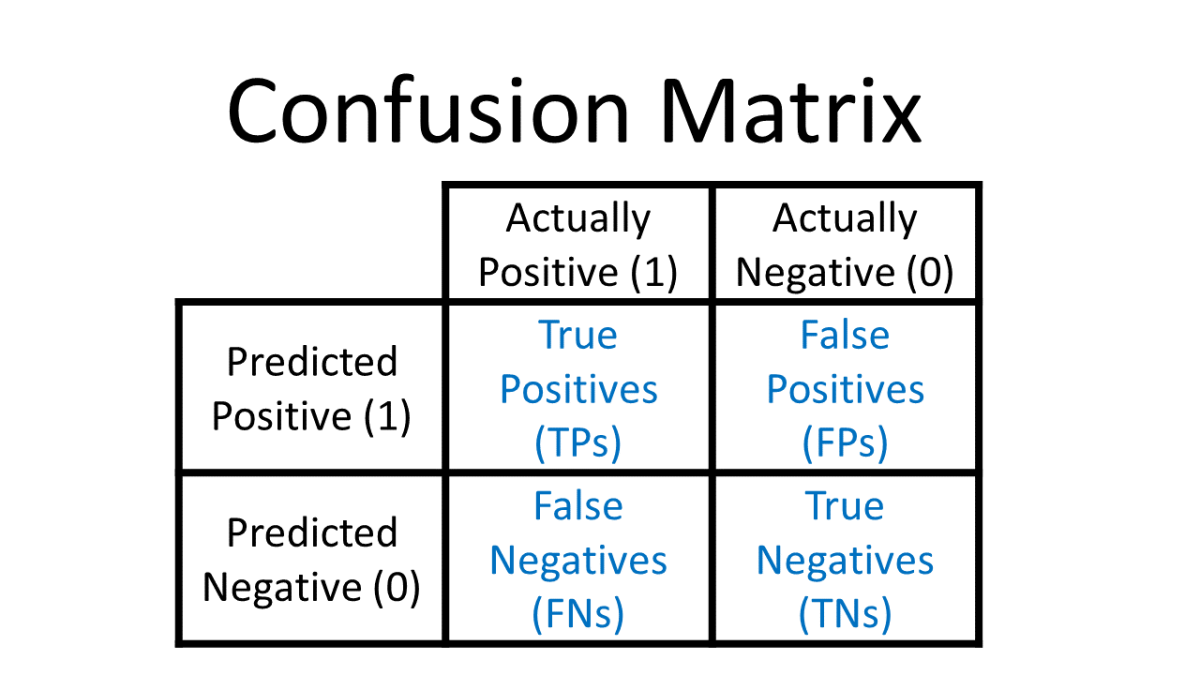
Эта задача в данный момент находится на этапе решения.

# Learning and understanding basic concepts of Computer Vision

## 2.1 Метрики оценки

MAP (Mean Average Precision) – является одной из самых распространённых метрик измерения точности декодеров. Для описания MAP введём понятие следующих метрик: precision, recall, IoU, а также понятие confusion matrix.

Confusion matrix представляет собой диаграмму точности прогнозирования классификатора для двух и более классов (рис. 6).



Picture 6. Confusion matrix[https://glassboxmedicine.files.wordpress.com/2019/02/confusion-matrix.png?w=1200]

True Positive – это результат, при котором модель правильно предсказывает положительный класс.

True Negative – это результат, при котором модель правильно предсказывает отрицательный класс.

False Positive – это результат, при котором модель неверно предсказывает положительный класс.

False Negative – это результат, при котором модель неверно предсказывает отрицательный класс.

Таким образом, метрика Precision может быть введена следующим образом:

|  |  |
| --- | --- |
|  | (1) |

где, TP – True Positive;

FP – False Positive.

Precision измеряет точность предсказаний классификатора.

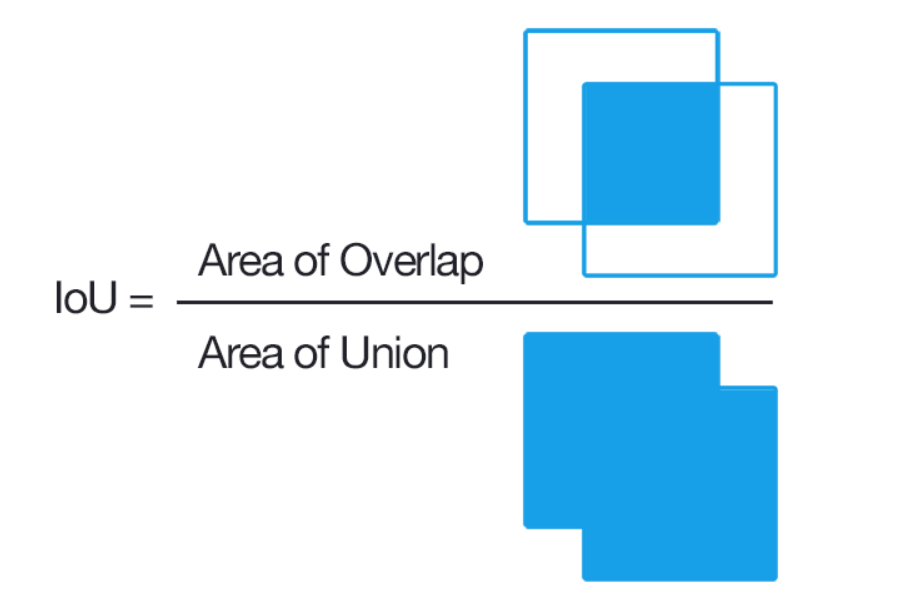
Следующей вводимой нами метрикой является Recall. Recall показывает, насколько хорошо классификатор определяет объекты «положительного» класса.

|  |  |
| --- | --- |
|  | (2) |

где, TP – True Positive;

FN – False Negative.

Последней вводимой метрикой является метрика IoU (Intersection over Union) – метрика, которая учитывает покрытие прогнозируемого BBox и истинного (ground true) BBox. IoU это число от 0 до 1, показывающее, насколько у двух объектов совпадает внутренний объем (рис.7).



Picture 7. Evaluating Intersection over Union

Системы обнаружения объектов делают прогнозы в терминах ограничивающей рамки (BBox) и метки класса (class label) используя значение IoU для заданного порога (threshold). Например, если порог IoU равен 0,5, а значение IoU для прогноза составляет 0,7, то прогноз классифицируется как истинно положительный (TF). С другой стороны, если IoU равно 0,3 то, как ложноположительный (FP).

Average Precision вычисляет среднюю точность для Recall в диапазоне от

0 до 1.

|  |  |
| --- | --- |
|  | (3) |

Тогда Mean Average Precision может быть вычислено как среднее для Average Precision

|  |  |
| --- | --- |
|  | (4) |

где, Q – количество запросов.

## Краткая теория свёрточных нейронных сетей

Термины «глубокое обучение» или «глубокая нейронная сеть» относятся к искусственным нейронным сетям (ANN). За последние несколько десятилетий они стали одним из самых мощных инструментов анализа фото/видео данных. Самым популярным нейросетевым инструментом анализа фото-видео данных являются свёрточные нейронные сети (CNN). Это название взято из математической линейной операции между матрицами, называемой сверткой.

CNN, представляет собой специализированный тип модели нейронной сети, предназначенный для работы с данными двумерного изображения, хотя их можно использовать с одномерными и трехмерными данными. Одной из главных проблем которую решает CNN это распознавание пространственно-независимых паттернов, устойчивых к изменению масштаба, поворотам, смене ракурса и прочим искажениям. Например, в приложении для распознавания людей не нужно обращать внимание на то, где на изображениях они расположены. Единственная задача – обнаружить их независимо от их положения в данном изображении. Еще один важный аспект CNN – это получить абстрактные функции при продвижении в более глубокие слои. В среднем точность распознавания таких сетей превосходит полносвязные нейросети на 10–15%. Так же увеличивается скорость обучения за счет распараллеливания процесса свертки по каждой карте, а также обратной свертки при распространении ошибки по сети [ Li Z. et al. A survey of convolutional neural networks: analysis, applications, and prospects //arXiv preprint arXiv:2004.02806. – 2020.].

В общем случае CNN состоит из трёх видов слоёв: convolutional, pooling and full-connection layers (рис.8) [Архитектура сверточной нейронной сети [Электронный ресурс]. – 2018. – URL: https://studfile.net/preview/6871496/page:8/].

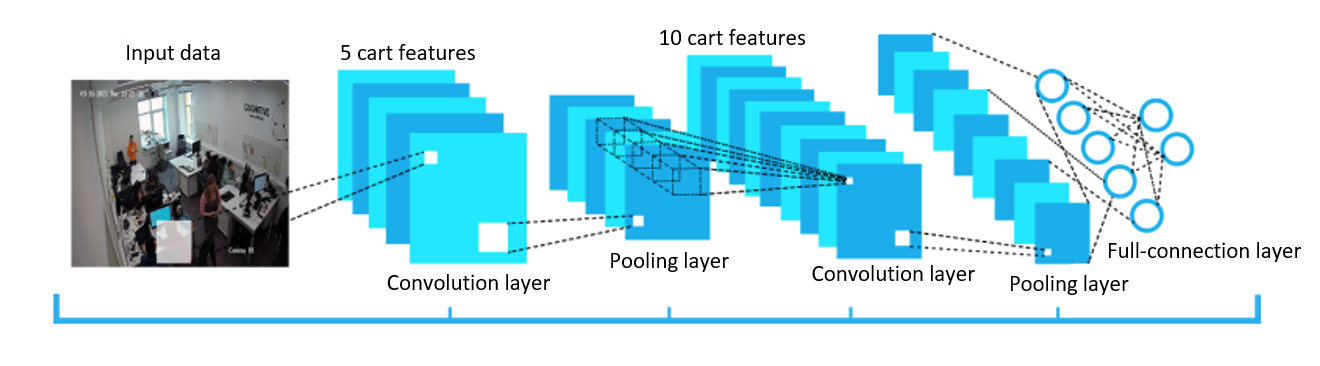


Рисунок 7. Архитектура свёрточной нейронной сети

Входные данные каждого значения пикселя нормализуются MinMax scaler в диапазон [0, 1] следующим образом

|  |  |
| --- | --- |
|  | (5) |

|  |  |
| --- | --- |
|  | (6) |

где,  – функция нормализации данных;

– функция дисперсии данных;

c – цвет пикселя ϵ [0, 255];

– наименьшее значение найденного пикселя на изображении;

– наибольшее значение найденного пикселя на изображении.

Основным элементом свёрточной нейронной сети является свёрточный слой, который выполняет операцию, называемую свёрткой [. Свёртка в Deep Learning простыми словами [https://neerc.ifmo.ru/wiki/index.php?title=Сверточные\_нейронные\_сети]. Свёртка – это линейная операция, выполняющая умножение набора весов на входные данные. Свёрточный слой состоит из  ядер с аддитивными составляющими смещения  для каждого ядра, и вычисляет свёртку выходного изображения предыдущего слоя с помощью каждого из ядер, прибавляя каждый раз составляющую смещения. Затем ко всему выходному изображению применяют функцию активации . Обычно, входной поток для свёрточного слоя состоит из  каналов. Например, для входного слоя = 3 при анализе red/green/blue изображения. Ядра также расширяют таким образом, чтобы они состояли из каналов, таким образом, формула для одного канала выходного изображения выглядит так:

|  |  |
| --- | --- |
|  | (7) |

где, K – ядро смещения;

I – двумерное изображение;

x – размерность двумерного изображения по первой оси;

y – размерность двумерного изображения по второй оси;

– функция активации;

b – составляющая смещения;

- размерность матрицы К по первой оси;

- размерность матрицы К по второй оси;

– количество каналов;

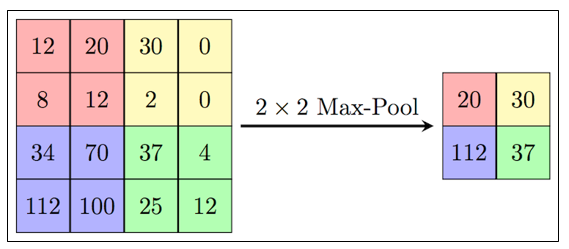
Использование ядра позволяет снизить вычислительную нагрузку и агрегировать дополнительную информацию из-за использования взвешенных сумм.

Вслед за свёрточным слоем принято использовать pooling layer. Это используется для уменьшения зависимости от точного позиционирования объектов путём уменьшения пространственного разрешения. Снижение пространственного разрешения это просто уменьшение количество пикселей карты функций за счет pooling layer [https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/].

Наиболее популярными являются два вида pooling:

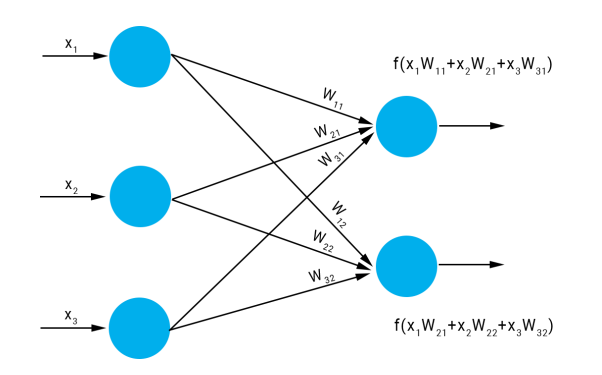
- Average pooling – это вариант pooling, при котором в качестве выходных данных берется среднее значение пикселей, попавших в ядро

-Max pooling (рис.8) – это вариант pooling, при котором в качестве выходных данных берется максимальное значение пикселей, попавших в ядро.



Picture 8. *Формирование новой карты подвыборочного слоя на основе предыдущей карты сверточного слоя. Операция Max Pooling [https://habr.com/ru/post/348000/]*

В конце свёрточной нейронной сети располагается полносвязный слой (рис.9).



Picture 9. Full connection layer

.Полносвязный слой – это слой, в котором каждый нейрон получает входные данные от всех нейронов своего предыдущего слоя[Dense Layer [Электронный ресурс]. – 2020. – URL: <https://machinelearningknowledge.ai/keras-dense-layer-explained-forbeginners/>]. В этом слое выполняется умножение матрицы на вектор, а значения, используемые в матрице, являются параметрами, которые можно обучать. Результатом работы полносвязного слоя является вектор размерности m при m классовой классификации.

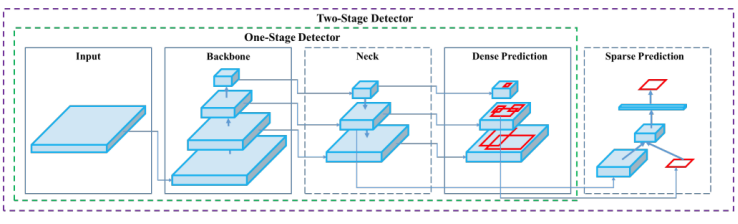
## 2.3 Object Detection

Обнаружение объектов является одной из наиболее важных задач компьютерного зрения. С помощью компьютерного зрения возможно обнаруживать экземпляры визуальных объектов определённого класса (например: люди, растения, транспортные средства). Отличительной особенностью детектирования является обнаружение точных координат искомых объектов в виде bounding box (BBox). Обнаружение объектов является одной из фундаментальных проблем компьютерного зрения, лежащей в основе многих других задач, таких как: сегментация, мониторинг безопасности и автономное вождение. Большинство современных детекторов используют лучшие сети глубокого обучения в качестве основы [Jiao L. et al. A survey of deep learning-based object detection //IEEE Access. – 2019. – Т. 7. – С. 128837-128868.].

Существующие детекторы объектов изображений можно разделить на две категории: One-stage Detectors and Two-stage Detectors.

Examples of One-Stage Detectors are RetinaNet, YOLO, SSD.

Examples of Two-Stage Detectors are R-CNN, Fast R-CNN, Faster R-CNN.



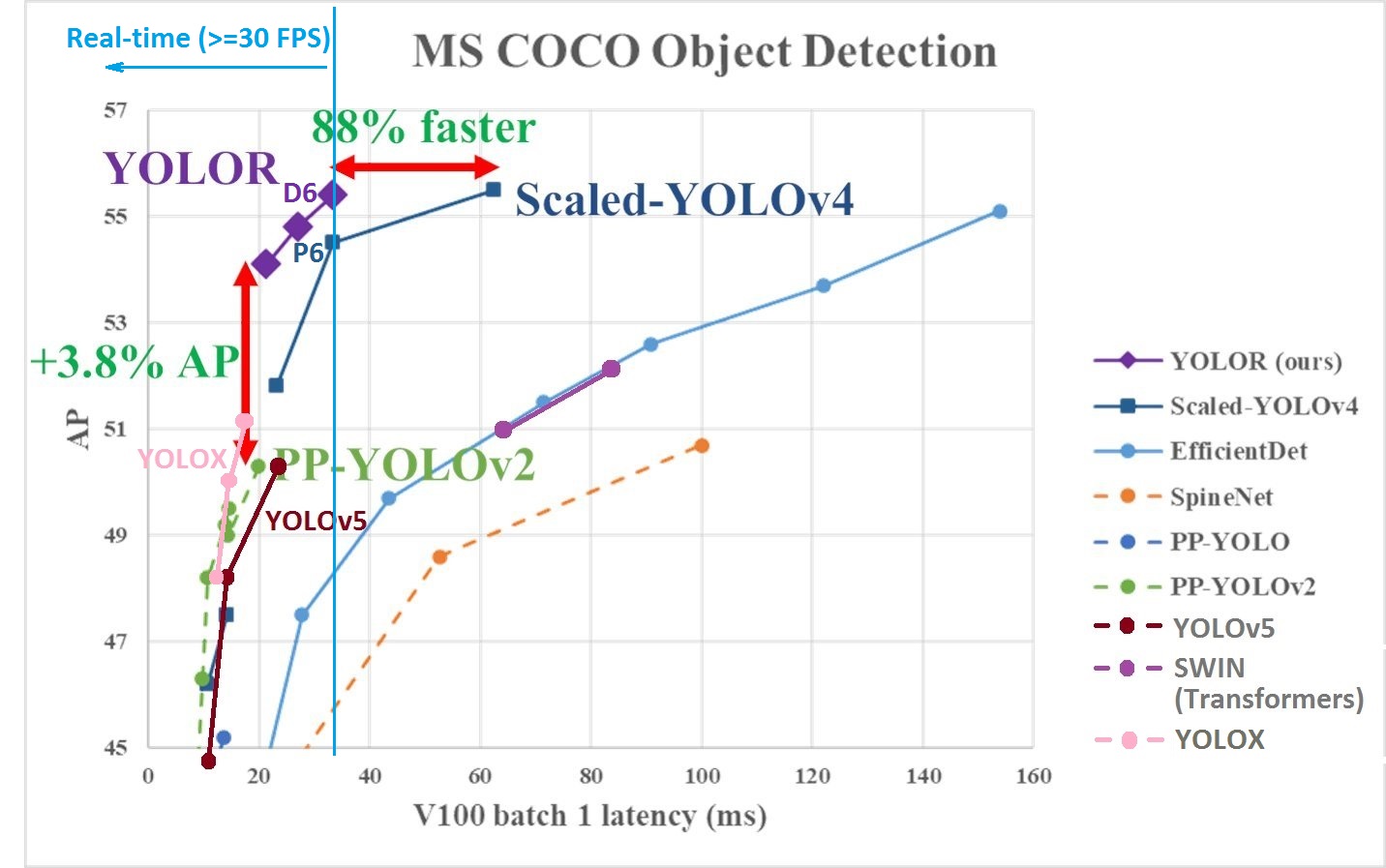
Picture 6. The structure of One-Stage and Two-Stage Detectors[Bochkovskiy A., Wang C. Y., Liao H. Y. M. Yolov4: Optimal speed and accuracy of object detection //arXiv preprint arXiv:2004.10934. – 2020].

Основное различие между этими двумя методами заключается в том, нужно ли создавать региональное предложение. Авторы двухступенчатых детекторов разработали специальную сеть, в которой используется выборочный поиск для извлечения всего 2000 регионов из изображения. region proposal network, RPN расположена после последнего сверточного слоя. Эти 2000 предложений областей-кандидатов преобразуются в квадрат и передаются в сверточную нейронную сеть, которая выдает 4096-мерный вектор признаков в качестве выходных данных. Эта сеть позволяет генерировать предлагаемые регионы на основе последней сверточной карты признаков. После этого этапа используется стандартная схема (пулинг целевого региона (RoI pooling), полносвязные слои, а затем классификация и регрессия).

Двухступенчатые детекторы обладают высокой точностью локализации и распознавания объектов, тогда как одноступенчатые детекторы обеспечивают высокую скорость вывода. Алгоритмы одноэтапного обнаружения объектов не требуют создания предложения региона. Он может напрямую классифицированный объект и его координаты. Двухэтапные алгоритмы обнаружения объектов должны генерировать область предложения перед их классификацией и позиционированием.

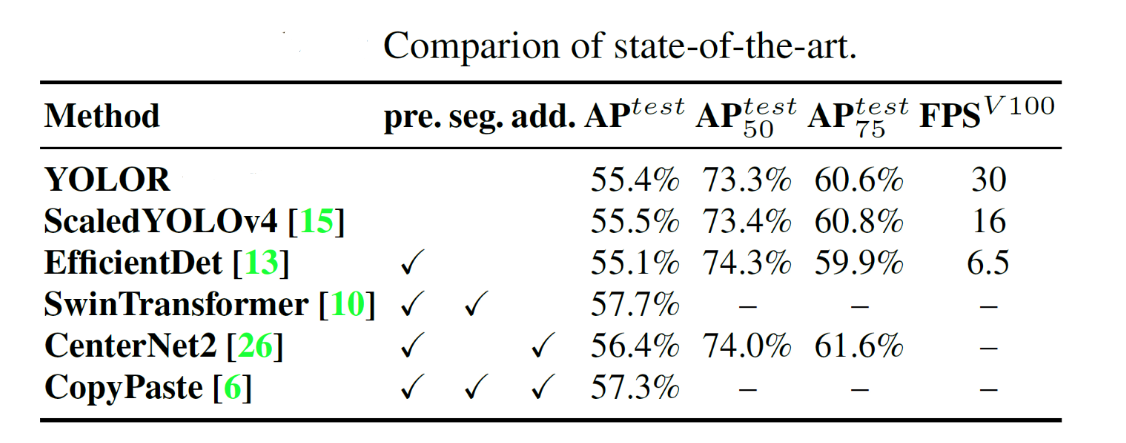
Чтобы напрямую оценить значения координат каждой точки BBox, нужно рассматривать эти точки как независимые переменны. Для решения этой проблемы исследователи предложили использовать вместо стандартной функции потерь, функцию IoU потерь.

В настоящее время тестирование моделей детектирования объектов проводится на Microsoft COCO dataset [https://www.microsoft.com/enus/research/wpcontent/uploads/2014/09/LinECCV14coco.pdf]. Данный датасет содержит 328 000 изображений с 2.5 million labeled instances of 91 objects types that would be easily recognizable by a 4 year old. По состоянию на 2022 год наилучшим образом на задаче детектирования объектов изображения справляется YOLOR[https://arxiv.org/pdf/2105.04206.pdf] and Scaled-YOLOv4[https://arxiv.org/pdf/2011.08036.pdf] (pic.7).



Picture 7. Models comparing on MS COCO dataset [https://arxiv.org/pdf/2105.04206.pdf]

В данной работе будут рассматриваться наилучшие мировые решения, доступные на момент написания данной работы. Опираясь на работу [https://arxiv.org/pdf/2105.04206.pdf](статья уолоР)(pic.8) становится очевидно, что наилучшим выбором для решаемой нами задачи будет ScaledYOLOv4. ScaledYOLOv4 проигрывает в FPS, зато имеет наибольшую точность распознавания. Поскольку в этой работе анализируется изменение IAQ в помещении, что является сильно инерциальным параметром, то нам не нужно требовать работы нейронной сети непосредственно в реальном времени. Поэтому я отдал приоритет в сторону качества распознавания объектов.



Picture 8. Comparison state-of-the-art object detection models

## YOLO model

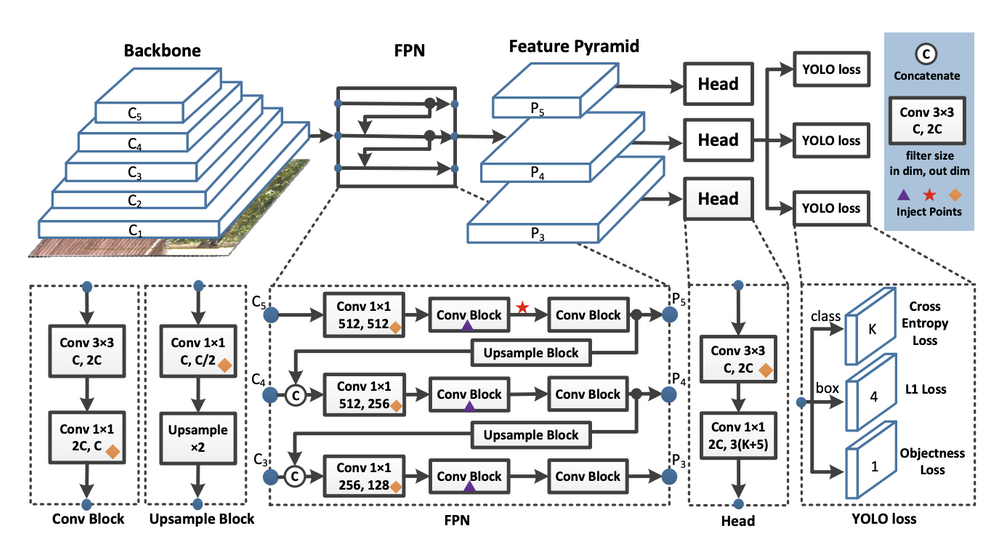
Модель YOLO (You Only Look Once) была впервые опубликована (Джозефом Редмоном и др.) В 2015 году. Первоначальная сеть YOLO использует модифицированный GoogLeNet[https://arxiv.org/pdf/1409.4842.pdf] в качестве опорной сети. Позже была создана новая модель под названием DarkNet-19, которая следует общей конструкции фильтров 3 × 3, удваивая количество каналов на каждом этапе объединения. Фильтры 1 × 1 также используются для периодического сжатия представления объекта по всей сети. YOLOv4 сделана на основе YOLOv3 с добавлением следующих изменений: в качестве основной сети (backbone) в YOLOv4 используется сверточная нейронная сеть CSPDarknet53, промежуточный этап (neck) SPP (spatial pyramid pooling) и (PAN) Path Aggregation Network, полносвязный итоговый классификатор (Dense Prediction) YOLOv3.

Основные модели (Backbone) были предварительно обучены как классификаторы изображений, прежде чем были адаптированы для задачи обнаружения. Для адаптации сети классификации в сеть для детектирования объектов необходимо удалить нескольких последних слоев сети и добавить сверточный слой с фильтрами для получения предсказаний BBox. На входное изображение накладывается сетка, разделяющая его на S×S областей. Для каждой области нейронная сеть определяет 5 ограничивающих рамок объекта, уровень достоверности (confidence score) обнаружения рамки, которая отражает степень уверенности модели в том, что поле содержит объект

|  |  |
| --- | --- |
|  | (8) |

В результате работы сети выбирается область с наибольшим уровнем достоверности ответа сети.

В отличие от родительской YOLOv4(pic.9) модели the Scaled-YOLOv4 authors apply the concepts laid out in the [**Cross-Stage Partial Networks**](https://arxiv.org/pdf/1911.11929) [https://arxiv.org/pdf/1911.11929.pdf]



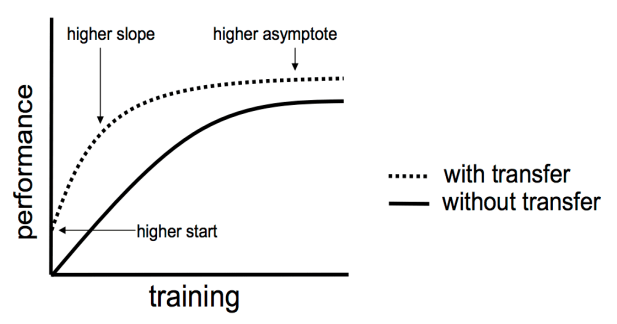
Picture 9. YOLOv4 architecture

Кроме этого To detect large objects in large images, the authors increase the depth and number of stages in in the CNN backbone and neck. This allows them to first scale up input size and number of stages, and dynamically adjust width and depth according to realtime inference speed requirements.

## 2.4 Transfer learning

Transfer learning – это метод машинного обучения, при котором уже обученная модель, переобучается на другую связанную задачу. Трансферное обучение позволяет значительно быстрее прогрессировать при обучении на новом датасете, что многократно улучшает производительность при моделировании задачи. Трансферное обучение популярно в глубоком обучении, учитывая огромные ресурсы и наборы данных, на которых обучаются модели глубокого обучения. Трансферное обучение работает в глубоком обучении только в том случае, если паттерны исходной модели, являются общими для модели переобучения. Сначала обучается базовая сеть на базовом наборе данных и задаче, а затем изученные паттерны передаются во вторую целевую сеть для обучения на целевом наборе данных и задаче [A Gentle Introduction to Transfer Learning for Deep Learning [Электронный ресурс]. – 2017. – URL: <https://machinelearningmastery.com/transferlearning-for-deep-learning/>].

Использование transfer learning обладает тремя преимуществами (рис. 10)



Picture 10. Transfer learning approach

Опираясь на рисунок, можно заметить, что при использовании transfer learning мы имеем. Более высокий старт по сравнению с исходной моделью. из-за чего сокращается время на обучение и параллельно задача начальной инициализации весов отбрасывается. Кривая производительности имеет больший тангенс угла наклона, поэтому достижение асимптотического максимума обобщающей способности модели происходит быстрее. Более того сама асимптота производительности лежит выше, чем на исходной модели. Это означает, что для данной модели у нас достигается более высокий конвергентный навык [Transfer learning and fine-tuning [Электронный ресурс]. –

2021. – URL: <https://www.tensorflow.org/tutorials/images/transfer_learning>].

## 2.5 Object tracker

Слежение за объектами, является сложной задачей. Трудности с отслеживанием объектов могут возникать из-за резкого движения объекта, изменения паттернов внешнего вида как объекта, так и сцены, перекрытия объектами, а также движения камеры. Отслеживание. В этом разделе будут исследованы несколько подходов по слежению объектов, разобраны алгоритмы, описаны плюсы и минусы.[ Bochinski E., Eiselein V., Sikora T. High-speed tracking-by-detection withoutusing image information //2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). – IEEE, 2017. – С. 1- 6.]

Процесс отслеживание объектов:

− Принимается начальный набор обнаруженных объектов (например, входной набора координат BBox)

− Создание уникального идентификатора для каждого из начальных обнаружений

− Отслеживание каждого из объектов, когда они перемещаются по кадрам в видео, с сохранением назначения уникальных идентификаторов. Идеальный алгоритм отслеживания объектов должен удовлетворять следующими параметрам:

− Использовать фазу обнаружения объекта только один раз (то есть при первоначальном обнаружении объекта). − Быть довольно быстрым – намного быстрее, чем запуск самого детектора объектов.

− Уметь обрабатывать, когда отслеживаемый объект исчезает или выходит за границы видеокадра.

− Быть устойчивы к окклюзии

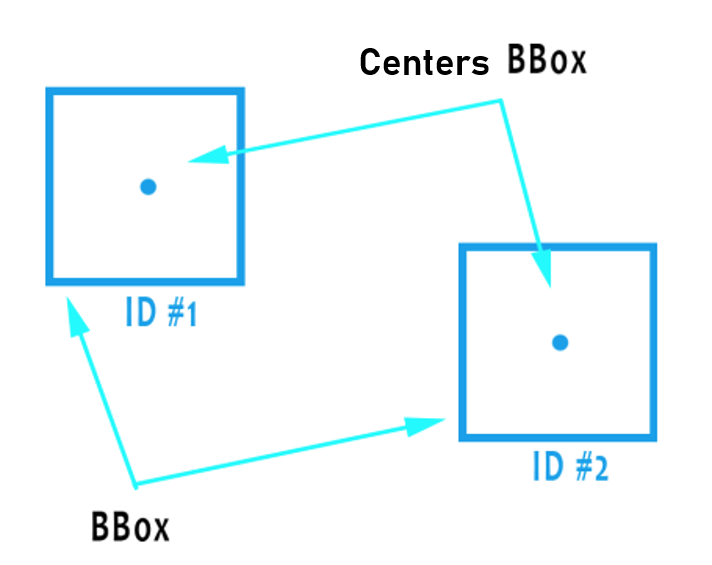
− Уметь подбирать объекты, которые были потеряны между кадрами.

Алгоритм отслеживания центров:

Алгоритм отслеживания объектов называется отслеживанием центров, поскольку он основан на евклидовом расстоянии между существующими центрами объектов и новыми центрами объектов между последующими кадрами в видео.

Алгоритм отслеживания центров – это многоэтапный процесс.

Шаг 1: принимаются координаты BBox и вычисляется его центр (рис.11)



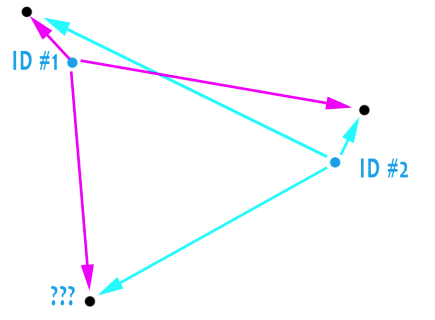
Picture 11. Example of centers and body BBox

Алгоритм отслеживания центров принимает набор координат ограничивающего прямоугольника (x, y) для каждого обнаруженного объекта в каждом отдельном кадре. Обрамляющие рамки (BBox) могут быть созданы любым типом детектора объектов (определение порога цвета, выделение контура, каскады Хаара, HOG, Faster R-CNN, нейронные сети SSD, YOLO и т. Д.), при условии, что они вычисляются для каждого кадра в Видео.

Как только детектор определяет координаты BBox, нужно вычислить координаты его центра (x, y). На рисунке 11 показано принятие набора координат обрамляющего прямоугольника и вычисление центра.

Поскольку это начальный набор BBox представленных алгоритму, то им присваиваются уникальные идентификаторы

Шаг 2: вычисление евклидова расстояния между новыми BBox и существующими объектами (рис. 12)



Picture 12. Adding new objects to tracking

Для каждого последующего кадра в видеопотоке применяется шаг 1, вычисления центров объектов; однако вместо того, чтобы назначать новый уникальный идентификатор каждому обнаруженному объекту, сначала нужно определить, можно ли связать центры новых объекта с центрами старого объекта. Чтобы выполнить этот процесс, вычисляется евклидово расстояние для каждой парой существующих объектов и входных объектов. На рисунке 12 видно, что на следующей итерации обнаружено три объекта. Две близкие друг к другу пары – это два существующих объекта.

Шаг 3 обновление (x, y) -координаты существующих объектов

Алгоритм предполагает, что потенциально объект будет перемещаться между последующими кадрами, но расстояние между центрами для кадров Ft и Ft + 1 будет меньше, чем все другие расстояния между объектами. На рисунке 12 видно, как алгоритм отслеживания центров выбирает ассоциацию центов, которые минимизируют их соответствующие евклидовы расстояния. При этом остается одна лишняя точка без ассоциаций.

Шаг 4: Регистрация нового объекта

Если обнаружено больше входных данных, чем отслеживаемых существующих объектов, необходимо зарегистрировать новый объект. Регистрация означает добавление нового объекта в список отслеживаемых объектов и назначение ему нового идентификатора.

Затем алгоритм возвращается к шагу 2 и повторяет последовательность шагов для каждого кадра в видеопотоке

Шаг 5: отмена регистрации старых объектов.

Алгоритм отслеживания объекта должен уметь обрабатывать ситуации, когда объект был потерян или покинул поле зрения. Регистрация старых объектов отменяется, когда они не могут быть сопоставлены с какими-либо существующими объектами для всего N последующих кадров.

Алгоритм отслеживания с использованием фильтра Калмана

Это продвинутый вариант трекера, помимо расстояния между центрами объектов используются их параметры сходства и скорость объекта для прогнозирования.

Особенности вида за пределами компонента обнаружения игнорируются при отслеживании, для оценки движения ограничивающего прямоугольника используются только положение и размер.

Для прогнозирования движения и ассоциации данных движения используются два классических, но чрезвычайно эффективных метода, фильтр Калмана и венгерский алгоритм [Bewley A. et al. Simple online and realtime tracking //2016 IEEE international conference on image processing (ICIP). – IEEE, 2016. – С. 3464-3468.]

Межкадровые перемещения каждого объекта аппроксимируются с помощью линейной модели постоянной скорости, которая не зависит от других объектов и движения камеры. Состояние каждой цели моделируется следующим образом:

|  |  |
| --- | --- |
|  | (9) |

где, – горизонтальное расположение пикселей центра;

– вертикальное расположение пикселей центра;

– площадь ограничивающего прямоугольника;

– соотношение сторон ограничивающего прямоугольника.

Когда срабатывает детектор, обнаруженный BBox, используется для обновления состояния трекера, где переменные скорости определяются с помощью Kalman filter framework [Peksa J. Prediction Framework with Kalman Filter Algorithm //Information. – 2020. – Т. 11. – №. 7. – С. 358.].

При назначении обнаружения существующим целям геометрия BBox каждой цели оценивается путем прогнозирования ее нового местоположения в текущем кадре. Матрица издержек классификации (Cost-matrix) вычисляется как расстояние пересечения по объединению (IoU) между каждым обнаружением и всеми предсказанными ограничивающими рамками для существующих целей.

Расстояние IOU ограничивающих прямоугольников неявно учитывает кратковременное перекрытие. В частности, когда цель перекрывается другим объектом, обнаруживается только окклюдер.

Когда объекты входят в область видимости или покидают ее, необходимо создавать или уничтожать уникальные идентификаторы. Для создания новых объектов рассматривается любое обнаружение с перекрытием меньше IoUmin как указание на существование не отслеживаемого объекта. Трекер инициализируется с использованием геометрии ограничивающего прямоугольника с нулевой скоростью [Алгоритм трекинга объектов в реальном времени с обработкой ошибок [Электронный ресурс]. – 2013. – URL: https://cyberleninka.ru/article/n/algori tm-trekinga-obektov-v-realnom-vremeni-s-obrabotkoy-oshibok/viewer]

Затем новый объект проходит испытательный срок, цель должна быть связана с последующими обнаружениями, чтобы накопить достаточно доказательств и не допустить отслеживание ложных срабатываний.

Треки завершаются, если они не обнаруживаются в течении TLostкадров. Это предотвращает неограниченный рост количества объектов слежения и ошибок локализации, вызванных предсказаниями на большие промежутки времени без поправок.

Практическая часть данной работы касается методики проведения эксперимента и локации проведения эксперимента, а именно 408 лабораторию Национального центра конгнитивных разработок ИТМО. Для данного исследования использовались 2 диаметрально расположенные IP камеры, а также датчик освещенности совместно с датчиком контроля параметров окружающей среды Bosch BME680, который позволяет измерять такие параметры, как сопротивление и влажность воздуха, температура и атмосферное давление. В качестве испытуемых в данном эксперименте принимало участие 6 человек. Средний возраст подопытных составил 24 года, гендерное распределение было 5 к 1 в пользу мужчин.

Всего сбор данных проводился в течении 17 экспериментальных дней. За это время удалось собрать 129.5 часов видео с и 55634 записи показателей сенсора. Важно отметить, что это объем сырых данные и после data mining вероятно их объем уменьшится.

Поскольку сенсор Bosch BME680 напрямую не вычисляет Indoor Air Quality нам пришлось вручную переводить показатели сенсора в индекс IAQ. Формула для перевода была выведена на основе закона уравнения идеального газа и уравнения Магнуса. Получили, что 52.5% экспериментальных данных были сделаны в воздухе среднего качества, 25.2% в воздухе низкого качества и 22.3% в высококачественном воздухе.

Поскольку нейронные сети не имеют возможно напрямую работать с видео данными мы разбили видеозаписи покадрово и взяли каждый 30-й кадр в целях уменьшения количества необходимых к разметке данных при сохранении большей части информации. Раскадровка видео проводилась open-source tool from DVDVideoSoft.

Мы приняли решение использовать в данном исследовании бинарную классификацию эмоциональных состояний человека: положительное и отрицательное. Для положительного состояния мы выделили такие поведенческие маркеры как: Sincerity, Consideration of the decision, Benevolence, Interest. К отрицательному состоянию мы отнесли следуюшие маркеры: Feelings of self-blame, strained perceptions of the situation, Defence or protection, Fatigue or emotional or physical tension, Desire to distance oneself from one's surroundings. Для каждого поведенческого состояния мы выделили уникальные невербальные признаки и принялись за ручную разметку данных. Ручная разметка данных проводится с помощью условно-бесплатному инструменту Roboflow Annotate.

Empirically, it has been found that YOLO training requires an average of 500 photos per pattern. Taking this into account, 4000 photos need to be manually partitioned for this job.

На текущий момент наше исследование находится на этапе ручной разметки изображений по невербальным признакам.