1 сл. Здравствуйте, меня зовут Пчелкина Ольга, я студентка 4 курса направления Информатика и вычислительная техника. Тема моей выпускной квалификационной работы – Разработка модуля опасных действий человека на основе нейросетевого подхода для интеллектуальной системы. Данный модуль разработан для системы видеоаналитики - EYECONT компании Mallenom Systems.

2 сл. В рамках данной работы необходимо было разработать алгоритмы и программное обеспечение для обнаружения людей на видеоизображениях, проверки использования средств индивидуальной защиты, проверки правильности применения средств индивидуальной защиты. Для этого были решены следующие задачи:

* Выбрать нейросетевые моделей;
* сформировать выборочные данные;
* обучить нейронные сети;
* реализовать API для взаимодействия с модулем и его конфигурирования;
* реализовать интерфейс для демонстрации и оценки работы модуля.

3 сл. На этом слайде представлено сравнение с аналогами по нескольким параметрам. Основными конкурентами данной системы являются - Система промышленной видеоаналитики CenterVision компании Центр2 и Детектор отсутствия касок (Macroscop)).

4 сл. Здесь вы можете увидеть стек применяемых в ходе работы технологий.

5 сл. Для реализации разрабатываемого модуля были выбраны нейросетевые модели:

* YOLOv4 для обнаружения людей
* MixNets для классификации СИЗ (MixNet\_S – для классификации правильности ношения и обнаружения красной каски. Классификатор обнаружения красной каски необходим для алгоритма. MixNet\_L – для классификации CИЗ и положения человека (лицом или спиной к камере)).

6 сл. YOLO или You Only Look Once — архитектура сверточных нейронных сетей, которая используется для распознавания множественных объектов на изображении. Главная особенность этой архитектуры по сравнению с другими состоит в том, что большинство систем применяют сверточные нейронные сети несколько раз к разным регионам изображения, в YOLO CNN применяется один раз ко всему изображению сразу. На вход модели подается некоторое входное изображение, после чего оно пропускается через модифицированную нейронную сеть GoogleNet, состоящую из 20 слоев, выдающую нам карту признаков. Следом идут четыре слоя Convolution с функцией ReLU и два слой Fully Connected, первый из которых также с применением функции ReLU, получая на выходе вектор. Полученный вектор преобразовывается в тензор.

Архитектура нейросети YOLOv4 (You Only Look Once version 4) является одной из самых эффективных и точных нейросетей для обнаружения объектов на изображениях. Она состоит из более 160 слоев и проходит через различные этапы обработки данных. Рассмотрим основные слои YOLOv4:

Слои входных данных: они отвечают за получение изображений для обработки. На вход подается изображение размером 608х608.

Convolutional Layers: обычные сверточные слои, которые извлекают признаки из изображения. В YOLOv4 используются в основном слои Darknet-53 (53 слоя свертки).

Residual Layers: резуидальные сверточные блоки, которые облегчают обучение и повышают производительность сети.

SPP Layer: Spatial Pyramid Pooling Layer. Этот слой выполняет пирамидальную пулинг операцию на слоях признаков, позволяя нейросети иметь лучшую обобщающую способность. (пулинг – отбрасывание ненужных признаков, пирамедальная - последовательная)

Upsampling Layers: слои, которые увеличивают размерность предыдущих слоев, используя повторное исполнение.

Concatenation Layers: слои, которые объединяют выходы из различных слоев.

Detection Layers: слои, которые применяются для обнаружения объектов на изображении. Они сгенерируют группы якорей в соответствии с размерами предсказаний, при этом используются различные решения для улучшения точности.

Loss Layers: функции потери, которые сравнивают предсказания с истинными значениями.

Характеристики YOLOv4:

160 слоев

Многимасштабная обработка изображений

Автоматический выбор границ определения объектов

Эффективное использование аппаратных ресурсов

В целом, архитектура YOLOv4 представляет собой глубокую нейросеть, которая использует многоэтапный процесс для обработки изображений и обнаружения объектов. Она эффективна и точна, и в настоящее время считается одной из лучших нейросетей для компьютерного зрения и обработки изображений.

7 сл. MixConvs —концепция объединения ядер нескольких размеров в один слой, которая является основным строительным блоком для MixNets. На данном слайде приведена архитектура MixNet-M, показывающая различные группы ядер. Вначале используются меньшие ядра, аналогичные текущим тенденциям с современными сетями. Однако более крупные ядра затем постепенно интегрируются по мере прохождения данных через слои. MixNet-L - это просто MixNet-M с коэффициентом глубины 1,3.

Коэффициент глубины (depth multiplier) является параметром, который используется для сокращения числа параметров и операций в свёрточных слоях. Это позволяет ускорить вычисления и снизить требования к памяти, что особенно важно при работе с мобильными устройствами и веб-браузерами.

Установка коэффициента глубины в 1 соответствует применению стандартных сверточных операций. При увеличении коэффициента глубины увеличивается количество параметров и вычислений, что может улучшить точность, но также сделать нейронную сеть более трудоемкой.

MixNets и EfficientNets - это два разных подхода к созданию нейросетей, которые могут быть эффективны в различных задачах машинного обучения. Однако, MixNets предлагает следующие преимущества по сравнению с EfficientNets:

Меньшее число параметров. MixNets использует меньше параметров, чем EfficientNets, что делает ее более легкой и быстрой в работе.

Повышенная точность при обучении на маленьких наборах данных. MixNets показала лучшие результаты в классификации изображений на меньших наборах данных, т.к. использует оптимизированные функции свертки.

Более легкая архитектура. MixNets использует свертки с меньшим числом фильтров, что позволяет ей работать с менее ресурсной аппаратурой.

Лучшая робастность к искажениям. MixNets обнаруживает более малые признаки, что позволяет нейросети быстро и точно обучаться на изображениях с различными искажениями.

Несмотря на то, что MixNets может быть лучше в некоторых задачах, в других задачах, EfficientNets могут показывать лучшие результаты. Поэтому, выбор нейросети зависит от типа задачи и набора данных.

8 сл. Архитектура MixNet-S.

9 сл. В ходе обучения была получена следующая точность:

* YOLOv4 – детектор людей. Точность mAP: 0.894. Точность F1-score: 0.863.
* MixNet\_S – классификатор красной каски. Точность F1-score: 1.

10 сл. MixNet\_S – классификатор СИЗ. Точность F1-score: 0.977.

11 сл. MixNet\_S – классификатор правильности ношения СИЗ. Точность F1-score: 0.45.

F1 score и mAP - это две метрики, используемые для оценки производительности нейронных сетей при задаче классификации и обнаружении объектов.

F1 score - это гармоническое среднее между точностью (precision) и полнотой (recall) модели. Он показывает, насколько точно и полно модель может идентифицировать классы объектов. F1 score является одной из наиболее популярных метрик в задаче классификации.

mAP (Mean Average Precision) - это метрика, используемая для оценки точности модели в обнаружении объектов. Он измеряет среднюю точность обнаружения объектов на нескольких уровнях порога срабатывания. mAP является одной из наиболее популярных и точных метрик в задаче обнаружения объектов.

12 сл. Алгоритм определения наличия СИЗ и правильности их ношения:

* если обнаружена красная каска проверяются антистатические перчатки, термостойки комбинезон, защитный лицевой щиток
* если обнаружена обычная каска, проверяется обычный комбинезон
* правильность ношения СИЗ, добавляется в событие только если есть соответствующий СИЗ (**SimpleOveralls** /**HeatProofOveralls** (*Простая / Термостойкая спецодежда*) - OverallsNotFastened (*Комбинезон не застегнут*), SleevesRolledUp (*Рукава закатаны*), BootsOverTrousers (*Сапоги поверх брюк*). **DielectricGloves** (*Диэлектрические перчатки*) - GlovesUnderOveralls (*Перчатки под спецодеждой*), GlovesAreCurved (*Края перчаток загнуты*))

13 сл. Архитектура модуля – микросервисная. Обращение к сревисам нейронных сетей и ядра реализуется с помощью REST API. Клиент посылает запросы – сервер отдает ответ в виде JSON или MultiPart(Json + JPEG).

14 сл. На данном слайде представлена демонстрация загрузки нейронных сетей для данного модуля, а также создание контекста в тестовом приложении.

15 сл. Демонстрация работы модуля на видео, где нет нарушений.

16 сл. Демонстрация работы модуля на видео с нарушением ношения СИЗ (Отсутствие диэлектрических перчаток).

17 сл. Спасибо за внимание.