1. **1 Data mining**

Источником для сбора тренировочных данных в настоящей работе являются видео файлы, записанные с IP камеры в лаборатории когнитивной невербалике ИТМО. (рис. 1.)



Рис 1. – Пример изображения с IP камеры университета ИТМО

Сырые данные представляют собой видео файлы, разделенные по датам и времени записи. Общая файловая структура видео файлов представлена на рис 2.

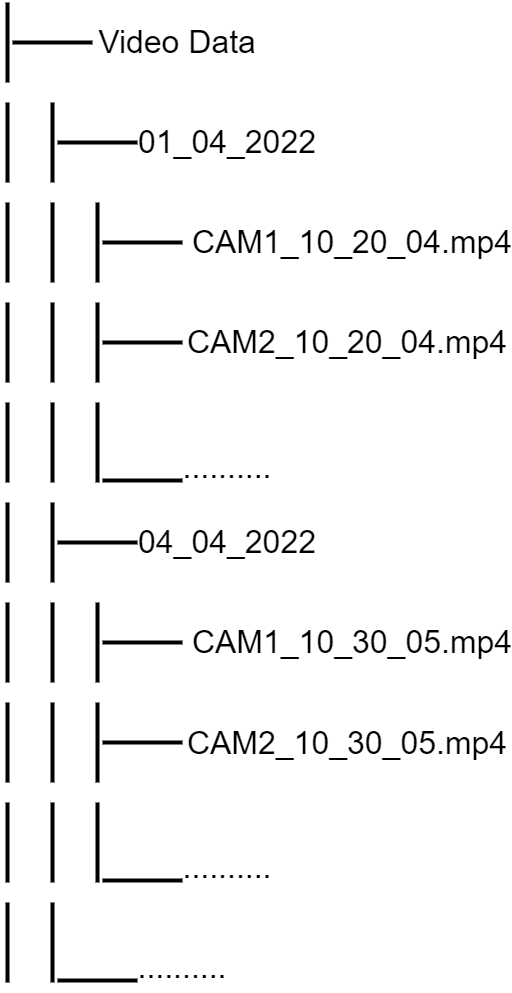


Рис. 2 – Общая файловая структура видео файлов

Где,

Названия родительских директорий, находящихся на верхнем уровне корневой системы, представлены в формате dd.mm.yyyy и указывают на дату видеофиксации, где

dd – days,

mm – months,

yyyy – years.

В субдиректориях датированных директорий находятся видеофайлы, формата .mp4[MPEG-4 File Format, Version 2 (Full draft). Sustainability of Digital Formats. Washington, D.C.: Library of Congress. 2 March 2012. Retrieved 13 December 2021. (https://www.loc.gov/preservation/digital/formats/fdd/fdd000155.shtml)] следующей структуры:

CAM1/CAM2 – предоставляет информацию с какой камеры происходила видеофиксация. САМ1 – фронтальная камера, CAM2 – тыловая камера, установленные в лаборатории когнитивной невербалики.

Пример изображения с фронтальной камеры представлен на рис 3



Рис3. – Изображение с фронтальной камеры CAM1

Пример изображения с тыловой камеры CAM2 представлен на рис. 4



Рис 4 – Изображение с тыловой камеры CAM2

После обозначения той или иной камеры в названии видеофайла идут цифры формата hh.mm.ss, обозначающие дату съемки видео, где

hh – hours,

mm – minutes,

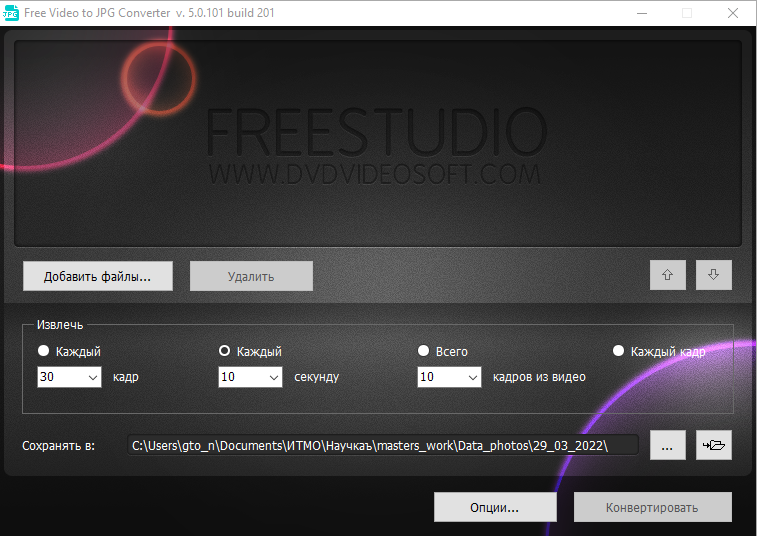
ss – seconds.

Например, название 01\_04\_2022/CAM1\_10\_20\_04.mp4 расшифровывается как,

Видео формата mp4, снятое на фронтальную камеру CAM1 1 апреля 2022 года в 10 часов 20 минут 04 секунды.

Сырые данные представляют из себя видео. To train a neural network we need to partition the data. Because neural networks can only work with images. Working with video is essentially fractioning the video recording into individual images and running these images through layers of neural network.

I used an open-source tool from DVDVideoSoft Free Video to JPG[https://www.dvdvideosoft.com/ru/products/dvd/Free-Video-to-JPG-Converter.htm] Converter to break down the video frame by frame.



Pic 3 – Interface of Video to JPG converter

Чтобы разделить видео на фотографии необходимо выбрать интервал разбиения.

В этой программе есть опции разделения видео покадрово. Данное программное обеспечение позволяет это сделать в двух вариациях. В первом варианте, в качестве JPG[T.81 – DIGITAL COMPRESSION AND CODING OF CONTINUOUS-TONE STILL IMAGES – REQUIREMENTS AND GUIDELINES" (PDF). CCITT. September 1992. Retrieved 12 July 2019.] изображения будет сохраняться каждый N-кадр, где число N – периодичность извлечения кадров. Система позволяет пользователю производить гибкую настройку режимов извлечения кадров изображения. Полный диапазон извлечения кадров лежит в промежутке [1:10000]. Второй вариант подразумевает лимитирование общего количества извлеченных из видео кадров. Пользователю, как и в первом варианте предоставляется возможность гибкой настройки общего количества извлечённых кадров в диапазоне [10:500].

Дополнительно в программе предусмотрена опция сохранения кадра видео каждые N-seconds. Для разбивки видео в нашей работе использовалась именно эта опция, видео разбивалось покадрово с частотой 1 кадр каждые 10 секунд и сохранялось в директорию Data\_photos.

Структура хранения файлов изображений раскадровки видео представлена на рис 4.

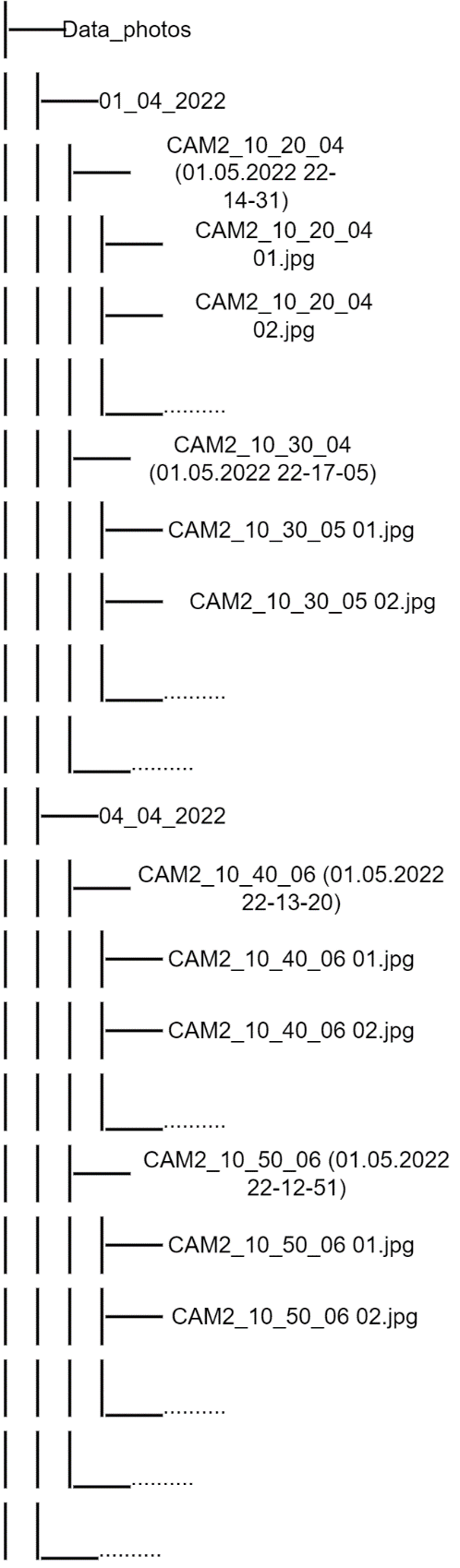


Рис 4 – Директория хранения раскадровки

Названия родительских директорий, находящихся на верхнем уровне корневой системы, представлены в формате dd.mm.yyyy и указывают на дату видеофиксации, где

dd – days,

mm – months,

yyyy – years.

В субдиректориях датированных директорий находятся директории с названиями в формате CAM{x}\_hh\_mm\_ss (dd.mm.yyyy hh-mm-ss), где

{x} - предоставляет информацию с какой камеры происходила видеофиксация. {x} = 1 – фронтальная камера, {x} = 2 – тыловая камера, установленные в лаборатории когнитивной невербалики,

hh\_mm\_ss – время в которое производилась съемка видео (hh – часы, mm – минуты, ss – секунды). В общем случае часть CAM{x}\_hh\_mm\_ss с точностью до расширения видеофайла совпадает с названием соответствующей видеозаписи в сабдерикторях структуры Video data.

(dd.mm.yyyy hh-mm-ss) – дата и время в которое происходила экстракция изображений из видео (dd – days, mm – months, yyyy – years, hh – hours, mm- months, ss – seconds).

Внутри директории находятся извлечённые из видео кадры формата .jpg. Структура имени аналогично первой части имени (до скобок) их родительской директории за исключением двух цифр в конце. Эти цифры показывают порядковый номер извлеченного кадра из видео. Благодаря этому у нас есть возможность вычислить точное время происходящего в кадре, используя эмпирическую формулу вычисления текущего времени кадра:

где,

time – текущее время кадра,

start\_time - время в которое производилась съемка видео,

step\_delay – шаг экстракции изображений (in our case step\_delay = 10 seconds)

ordinal number – последние две цифры в названии изображения.

В результате экстракции каждого 10-го кадра из видео получили 10856 изображений, которые требуют ручной разметки.

1. **1 Data labeling**

В машинном обучении процесс идентификации необработанных данных (строки таблиц, аудио файлы, видео, тексты и т.д.) называется разметкой данных. Разметка данных заключается в добавлении некоторого количества информативных меток для обеспечения контекста. Благодаря контекстным меткам алгоритм машинного обучения может получить представления об корректном и некорректном результате работы и вследствие этого корректировать внутренние веса. Данный процесс называется обучением.

Обучение на размеченных данных называется – обучение с учителем. Это один из самых распространённых методов тренировки нейронных сетей и алгоритмов машинного обучения, он применяется в различных сферах, включая обработку естественного языка, распознавание речи, компьютерное зрение. В нашем случае метки указывают алгоритму есть ли на изображении человек в плохом или хорошем психоэмоциональном состоянии. Психоэмоциональное состояние человека определяется по невербальным признакам.

Ручная разметка изображений проводилась с помощью графического инструмента аннотирования изображений с открытым исходным кодом LabelImg [https://github.com/heartexlabs/labelImg]. It is written in Python and uses Qt for its graphical interface. Annotations are saved as XML files in PASCAL VOC format, the format used by ImageNet. Besides, it also supports YOLO and CreateML formats. Интерфейс программы представлен на рис 5.

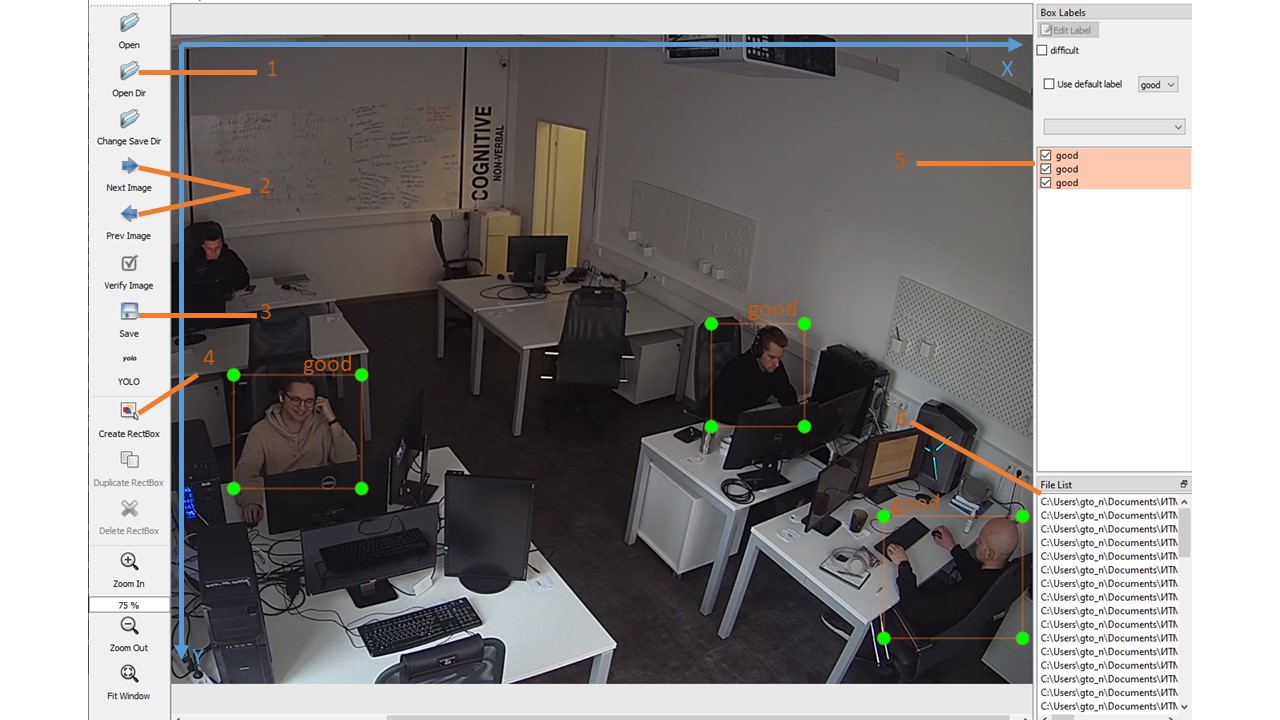


Рис 5 – Интерфейс программы LabelImg

Где,

1. Кнопка выбора директории с размечаемыми изображениями
2. Кнопки переключения между изображениями
3. Кнопка сохранения изображения. Сохранение происходит в ту же директорию, где находится изображение
4. Кнопка активации функционала создания BBox
5. Панель с размеченными лейблами на изображении
6. Список всех изображений, доступных к разметке

Необходимо обратить внимание, что координатные оси X и Y имеют начало в левом верхнем углу с направлением осей слева на право по X и сверху вниз по Y.

После разметки изображения в формате yolo в текущей директории появится одноимённый .txt[ Lewis, John (2006). Computer Science Illuminated. Jones and Bartlett. [*ISBN*](https://en.wikipedia.org/wiki/ISBN_(identifier)) [*0-7637-4149-3*](https://en.wikipedia.org/wiki/Special:BookSources/0-7637-4149-3)] файл для каждого .jpg изображения со следующей структурой: <class\_label> <Xleft> <Yleft> <Xright> <Yright>, где

<class\_label> - целое число, обозначающее id класса. Порядок и перечень классов указывается в автоматически создаваемом файлом classes в этой же директории. В нашем случае 0 – good PES, 1- bad PES;

<Xleft> <Yleft> <Xright> <Yright> - значения с плавающей точкой, имеющие 6 знаков после запятой и обозначающие значения координат верхнего левого и правого нижнего угла BBox относительно высоты и ширины изображения. Значение лежит в пределах (0:1].

К разметке данных необходимо подходить крайне ответственно, поскольку от корректности, точности и аккуратности размеченных данных будет зависеть финальный результат.

In our research, a binary distinction will be made between positive and negative human emotional states. Each of the states is characterized by its own set of non-verbal signs, such as gestures and postures. In different psycho-emotional states, a person highlights different non-verbal signs when interacting with objects or other people.

The positive non-verbal signs and their corresponding gestures and postures are presented in the table 2:

Table 1 - Positive human emotional states and their corresponding gestures [[Isaac Wang & Jaime Ruiz (2021) Examining the Use of Nonverbal Communication in Virtual Agents, International Journal of Human–Computer Interaction, 37:17, 1648-1673, DOI: 10.1080/10447318.2021.1898851](https://www.tandfonline.com/doi/pdf/10.1080/10447318.2021.1898851?src=getftr) ].

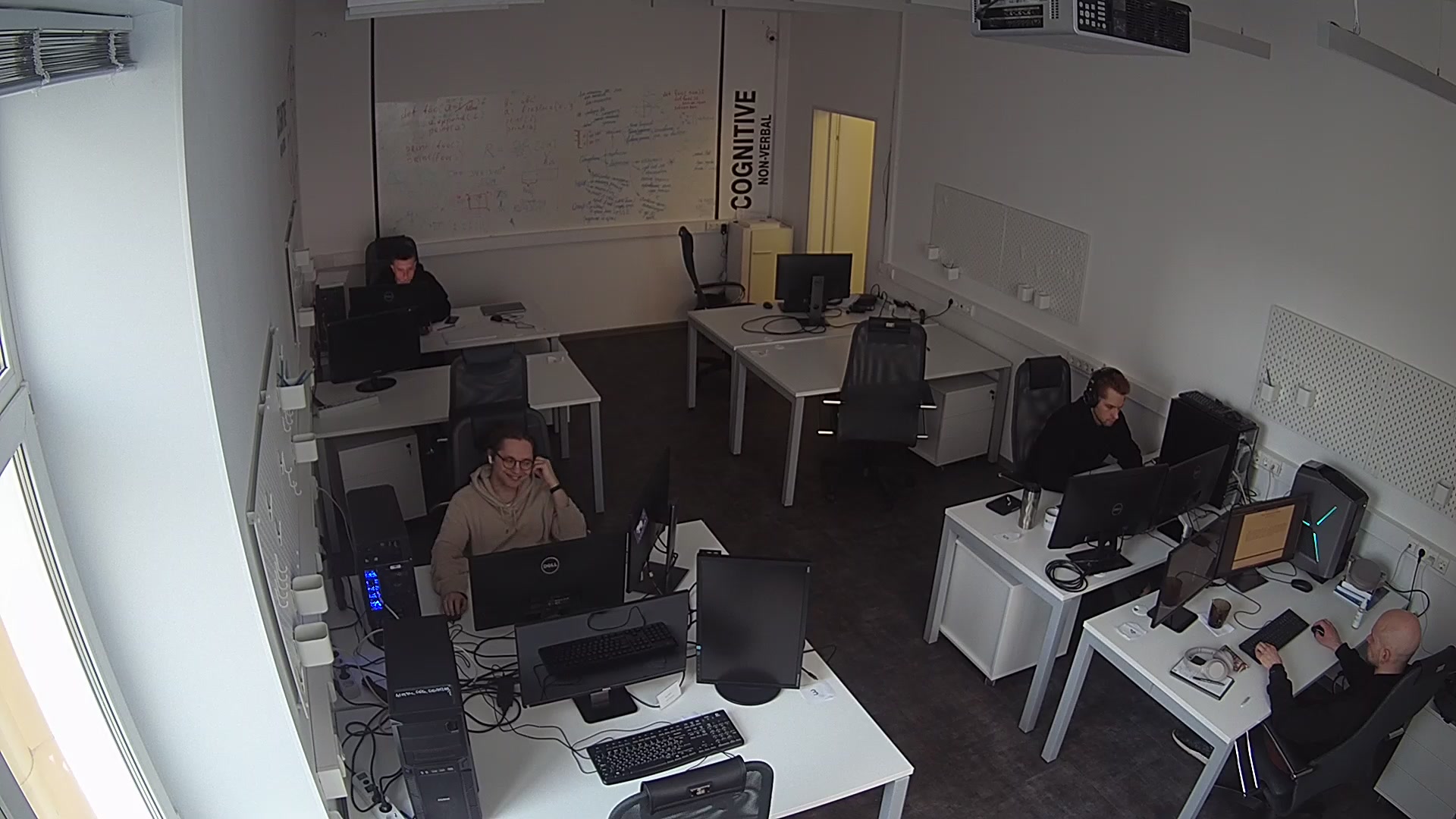
|  |  |
| --- | --- |
| The emotional state of the person | Gestures, poses |
| Sincerity | Open hands, palms up. |
| Consideration of the decision | Person sitting on the edge of a chair, leaning forward, head slightly tilted and resting on the hand |
| Benevolence | A soft smile, slight inclination of the head towards the other person, an expression in the eyes. |
| Interest | A squinting gaze is accompanied by a slightly raised eyebrow or a smile.  Leaning towards the other person (courtesy, attention) |

Negative non-verbal signs and their corresponding gestures and postures are presented in Table 3:

Table 2 - Negative emotional states of a person and their corresponding gestures [https://www.elitarium.ru/sostoyanie-sobesednik-napryazhenie-rechi-golos-ulybka-nervoznost-ustalost-zhestikulyaciya-vzglyad-volnenie-obman-lozh-strah-neuverennost-nastroenie/].

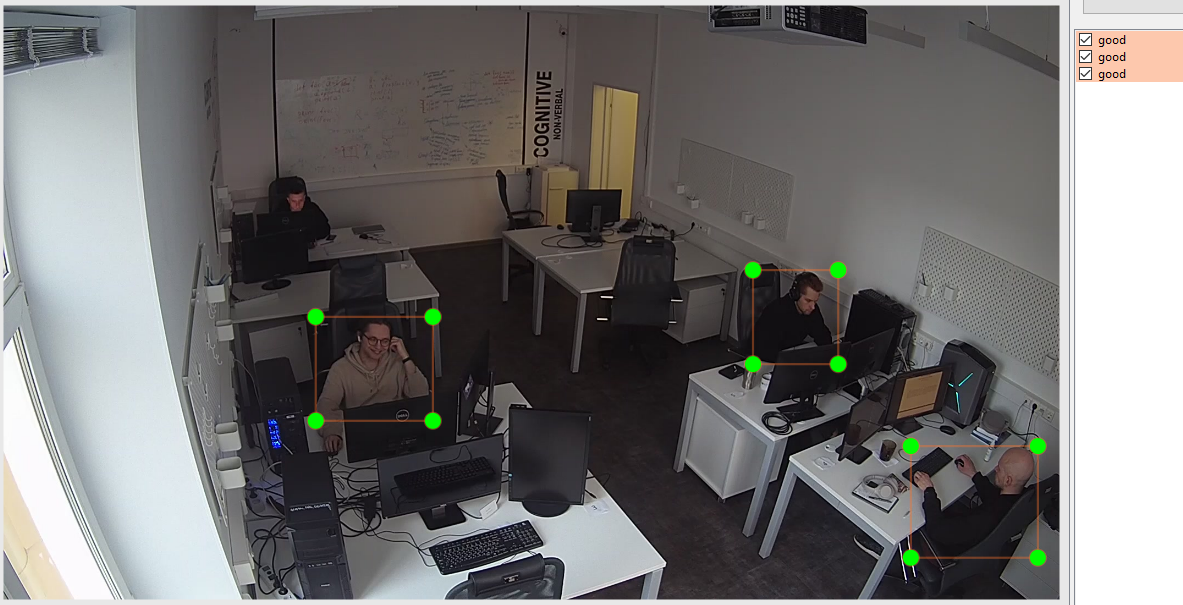
|  |  |
| --- | --- |
| The emotional state of the person | Gestures, poses |
| Feelings of self-blame, strained perceptions of the situation | Hands hidden (behind back, in pockets) |
| Defence or protection | Arms crossed on chest |
| Fatigue or emotional or physical tension | Yawn, hand propping up the head, hand scratching eyes |
| Desire to distance oneself from one's surroundings | Cross-legged pose |

As an example of positive class markup, consider pic. 6.



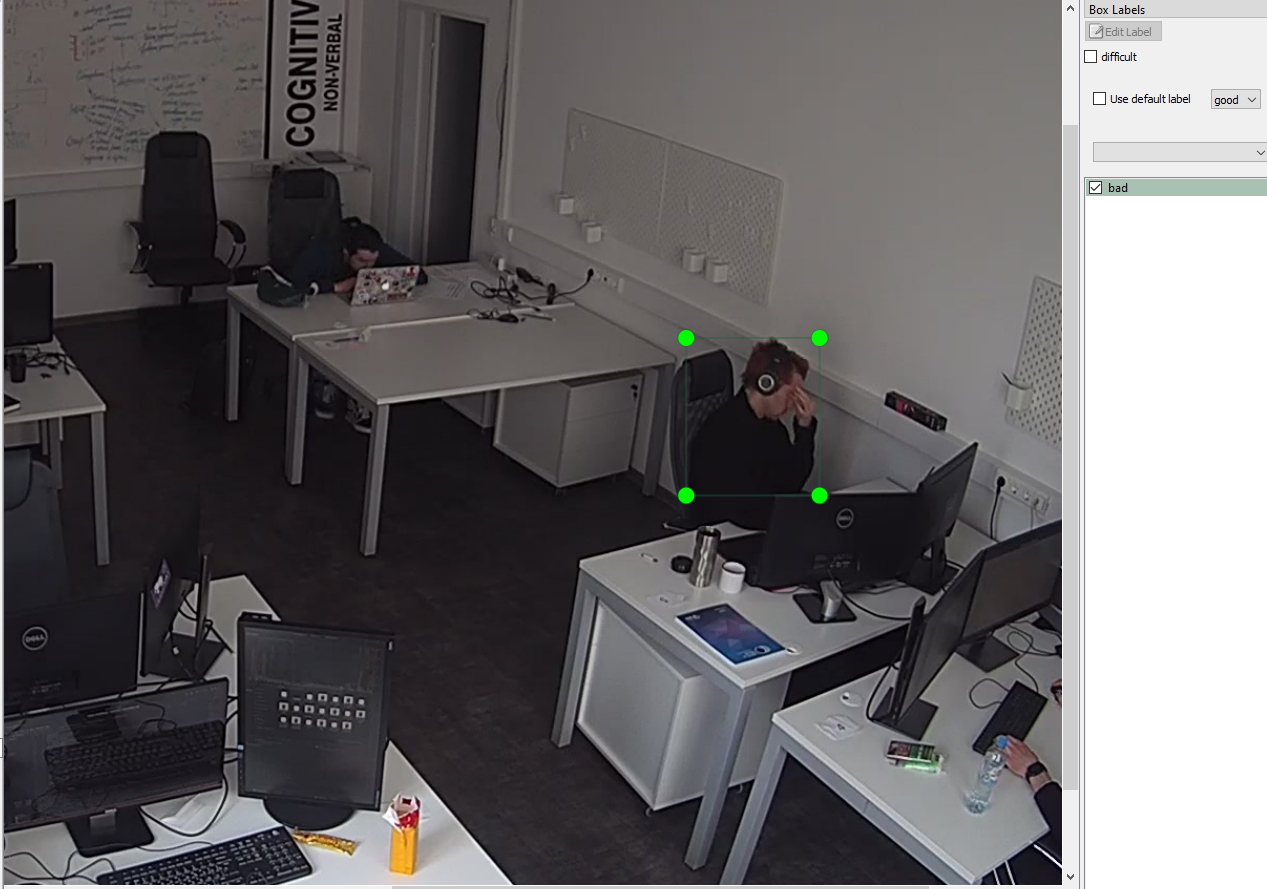
Picture 6 – Positive emotions of people. Man is smiling and others have open hands.

Thus we can classify people's emotions at manual marking as positive pic. 7.



Picture 7 – Hand-labeling smiling people to positive class

As an example of negative class markup, consider pic. 8



Picture 8 - Hand-labeling man which scratching his eyes to negative class

В ходе ручной разметки из общего количества изображений 10856 в итоговую разметку попало 861 изображение, что составило всего около 8% первоначальной выборки. Это обусловлено тем, что в сырых данных в видеоформате присутствуют моменты отсутствия испытуемых на рабочем месте. Кроме этого, определение психоэмоционального состояния человека происходит по конечному набору невербальных признаков, которые испытуемые не всегда выказывают.

Отдельной сложностью являются смешанные моменты, когда в кадре присутствует несколько испытуемых с различным набором невербальных признаков, отвечающим различным психоэмоциональным состояниям человека. Для поддержания строкой детерминированности классового разделения изображений, изображения с мультизначным набором невербальных признаков не берутся в анализ.

Классовое разбиение размеченных изображений составляет:

370 – к лейблу Good (0 class)

491 – к лейблу Bad (1 класс)

Графически это представлено на рис. 9.

Классовое отношение:

Что говорит о небольшом дизбалансе классов в сторону Bad оценки.

Ручная оценка корреляции Пирсона[Pearson, Karl (20 June 1895). ["Notes on regression and inheritance in the case of two parents"](https://books.google.com/books?id=60aL0zlT-90C&pg=PA240). *Proceedings of the Royal Society of London*. **58**: 240–242. [Bibcode](https://en.wikipedia.org/wiki/Bibcode_(identifier)):[1895RSPS...58..240P](https://ui.adsabs.harvard.edu/abs/1895RSPS...58..240P)] между психоэмоциональным состоянием человека и качеством воздуха выдала значение 0.81, что свидетельствует о наличии сильной корреляционной связи между величинами [https://medstatistic.ru/articles/correlacia.pdf].

Выводы:

В результате проделанной работы экспериментальные данные с IP камер лаборатории когнитивной невербалики были собраны и агрегированы в одну директорию с древовидной структурой и разделением поддиректорий по датам с вложенными в них видеозаписями, разделенными временными метками.

Поскольку сырые данные представляют из себя видео, а обучение нейронной сети возможно только на изображениях, необходимо разделить видео на кадры. Для этого был найден и успешно установлен open-source tool from DVDVideoSoft Free Video to JPG Converter, позволяющий ращбить видео на кадры с заданной пользователем частотой. В своей работе частота разбиения видео составляла 1 кадр в 10 секунд. Таким образом, после разбиения всех файлов видеозаписей получилось 10856 изображений, которые были размещены в древовидной структуре каталогов с корневыми директориями в виде даты, поддиректориями, представляющими собой время съемки видео и тип используемой камеры, внутри которых помещены порядково пронумерованные файлы изображений, которые предстоит разметить вручную. Для восстановления точного времени кадра была эмпирически выведена формула, учитывающая порядковый номер изображения, шаг раскадровки и время записи видео.

Для ручной разметки изображений использовался графический инструмент аннотирования изображений с открытым исходным кодом – LabelImg. Инструмент позволяет размечать файлы сразу в формате yolo, создавать в текущей директории .txt документы с координатами BBox и лэйблами для каждого изображения. Подопытные размечались исходя из выказываемых ими невербальных признаков и разделялись на две группы, в нулевую группу (0 label = good) попадали люди с невербальными признаками, отвечающими за хорошее психоэмоциональное состояние, полный набор которых представлен в таблице 1, в основном для анализа выбирались люди, сидящие прямо с вытянутыми руками. В первую группу (1 label = bad) попадали люди с выраженными невербальными признаками, отвечающими за отрицательное психоэмоциональное состояние, в основном в анализе присутствовали люди с руками у лица и явно выраженной скукой. Полный перечень невербальных признаков, отвечающих за отрицательное невербальное состояние представлен в таблице 2.

В ходе ручной разметки из общего количества изображений 10856 в итоговую разметку попало 861 изображение, что составило всего около 8% первоначальной выборки. Это обусловлено тем, что в сырых данных в видеоформате присутствуют моменты отсутствия испытуемых на рабочем месте. Кроме этого, определение психоэмоционального состояния человека происходит по конечному набору невербальных признаков, которые испытуемые не всегда выказывают.

Для поддержания строкой детерминированности классового разделения изображений, изображения с мультизначным набором невербальных признаков не берутся в анализ.

Классовое разбиение размеченных изображений составляет: 370 – к классу good (0 label), 491 – к классу bad (1 label). Классовое отношение составляет 0.75, что говорит о небольшом дизбалансе классов в сторону bad оценки.

Кроме этого, оценка корреляции Пирсона между психоэмоциональным состоянием человека и качеством воздуха составила 0.81, поэтому можно говорить о сильной статистической взаимосвязи между качеством этими величинами

Этот вывод является одним из ключевых результатов нашей работы. На независимом эксперименте нам удалось показать, что существует сильная взаимосвязь между психоэмоциональным состоянием человека и качеством воздуха окружающей среды, проявляющееся в отображении невербальных признаков человека.