УДК 004.8; 004.032.26; 159.952
DOI: 10.25559/SITITO.16.202002.500-509

Подход к оценке состояний внимания и проектирование моделей распознавания на базе нейронных сетей

Я. Н. Артамонова[1*], И. М. Артамонов[2]
[1] ООО «Нейрокорпус», г. Москва, Россия
121415, Россия, г. Москва, ш. Варшавское, д. 1, стр. 1-2
[*] ceo@neurocorp.ru
[2] ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)», г. Москва, Россия
125993, Россия, г. Москва, ш. Волоколамское, д. 4

Аннотация

В статье рассматривается подход к цифровизации феномена внимания. В работе приведены ссылки о том, что внимание улучшает любую деятельность. Психолого-педагогические исследования показывают, что особое положительное влияние внимания оказывает на деятельность обучения. Выбор направления исследования и разработки технологий диагностики внимания обусловлен прикладными задачами и ожиданиями повышения эффективности и скорости освоения программ обучения, отказа от неэффективных методик, оперативной реакции на трудности в освоении учебной программы и повышение легкости восприятия материалов. Авторы, основываясь на экспертном анализе видео данных, сформулированных требований к методике, рассматривают возможность использования методов компьютерного зрения и алгоритмов распознавания изображений на основе нейронных сетей для анализа внимания по наблюдаемым паттернам выразительных движений. Показаны модель обработки видеоданных, структура модели нейронной сети определения общих паттернов внимания. Предложенные в статье подходы и модели позволяют рассмотреть возможности современных информационных технологий в такой области как образование. Рассмотреть применение алгоритмов на основе нейронных сетей в образовательной деятельности, повысить эффективность обучающих программ, в особенности он-лайн и дистанционных программ, за счет оперативной обратной связи ведущим обучающих мероприятий и возможности в реальном времени корректировать учебный процесс и методические материалы.

Ключевые слова: распознавание внимания, видеоанализ поведения, дистанционный мониторинг, нейронные сети.

Финансирование: настоящая работа основана на исследованиях, выполненных при финансовой поддержке Фонда содействия инновациям (Договор № 1ГС1НТИС5/43222 от 06.09.2018).

Для цитирования: Артамонова, Я. Н. Подход к оценке состояний внимания и проектирование моделей распознавания на базе нейронных сетей / Я. Н. Артамонова, И. М. Артамонов. – DOI 10.25559/SITITO.16.202002.500-509 // Современные информационные технологии и ИТ-образование. – 2020. – Т. 16, № 2. – С. 500-509.

An Approach to Assessing Attention States and Designing Recognition Models Based on Neural Networks

Y. N. Artamonova[a*], I. M. Artamonov[b]
[a] Neurocorpus LLC, Moscow, Russia
1/1-2 Warsaw Highway, Moscow 121415, Russia
[*] ceo@neurocorp.ru
[b] Moscow Aviation Institute (National Research University), Moscow, Russia
4 Volokolamsk Highway, Moscow 125993, Russia

Abstract

The article discusses the approach to digitalization of the phenomenon of attention. The work provides links that attention improves any activity. Psychological and pedagogical studies show that a particular positive effect of attention has on the activities of learning. The choice of the direction of research and development of attention diagnostics technologies is determined by applied tasks and expectations of increasing the efficiency and speed of mastering training programs, abandoning ineffective methods, promptly responding to difficulties in mastering the curriculum and increasing the ease of perception of materials. The authors, based on expert analysis of video data, formulated requirements for the methodology, consider the possibility of using computer vision methods and image recognition algorithms based on neural networks to analyze attention according to the observed patterns of expressive movements. The model of video data processing, the structure of the neural network model for determining common patterns of attention are shown. The approaches and models proposed in the article allow us to consider the possibilities of modern information technologies in such an area as education. To consider the use of algorithms based on neural networks in educational activities, to increase the effectiveness of training programs, especially on-line and distance learning programs, through prompt feedback from leading training events and the ability to adjust the learning process and teaching materials in real time.

Keywords: Attention recognition, video analysis, distance monitoring, neural networks.

Funding: This work is based on research funded by the Fund for the Promotion of Innovations (Contract No. 1ГС1НТИС5 / 43222 dated 06.09.2018).

For citation: Artamonova Y.N., Artamonov I.M. An Approach to Assessing Attention States and Designing Recognition Models Based on Neural Networks. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(2):500-509. DOI: https://doi.org/10.25559/SITITO.16.202002.500-509

Введение

Каждый из нас хотя бы раз в жизни сталкивался с призывом: «Обратите внимание!», «Будьте внимательны!», «Внимание, пожалуйста!», «Attention, please!», «Achtung!», «Atención!», «Attenzione!», «Pozor!», «Dikkat!».

Однако, оказывается, что «внимание не представляет самостоятельного психического процесса»[1]. Один из основателей психологии Р. Вудвортс пишет: «Каждый эксперимент, который ставит перед испытуемым задание выполнить что-либо, требует от него внимания, будь то эксперимент по памяти, времени реакции, ощущению, восприятию или решению задачи». «Внимание не представляет самостоятельного психического процесса, так как не может проявляться вне других процессов. Мы внимательно или невнимательно слушаем, смотрим, думаем, делаем. Таким образом, внимание является лишь свойством различных психических процессов»[2].

С другой стороны, показатели внимания, к которым относятся объем, концентрация, устойчивость, длительность, переключение, могут быть выделены и изучены. Нарушения внимания в нейропсихологии являются одним из диагностических признаков. Известны крайние случаи изменения: при синдроме дефицита внимания и гиперактивности (СДВГ), различных заболеваниях мозга. Опытный нейропсихолог только на основании картины нарушения свойств внимания может сформировать гипотезу о локализации зон поражения в центральной нервной системе. Дифференциальная диагностика позволяет различать болезнь или снижение качества внимания на фоне усталости, незрелости или болезни[3].

П. Я. Гальперин сформулировал решение этого парадокса так: «Не всякий контроль есть внимание, но всякое внимание означает контроль. Контроль лишь оценивает деятельность или ее результаты, а внимание их улучшает»[4]. Современные психолого-педагогические исследования показывают, что особое положительное влияние внимания оказывает на деятельность обучения[5].

Возможно ли, использовать методы компьютерного зрения для оценки степени внимания человека в ситуации? Существуют ли наблюдаемые выразительные движения (позы, мимика) людей, выделяя и анализируя которые, можно сделать косвенные качественные выводы о том внимателен человек или нет?

У. Джеймс, один из основателей психологии, описывая феномен внимания пишет: «Тот или иной орган наших чувств должен приспособиться к наиболее ясной рецепции объекта внимания путем настройки мускульного аппарата» и далее… «Приспособление органа чувств. Происходит оно не только в сенсорном, но и в умственном внимании к объекту. …Когда мы присматриваемся или прислушиваемся к чему-либо, то непроизвольно приспосабливаем глаза и уши, а также поворачиваем голову и тело…»[6].

Иными словами, в качестве вводного тезиса эксперимента по видеоанализу внимания важно, что:

- выразительные движения внимания существуют;
- внимание проявляется «в настройке мускульного аппарата» для лучшего восприятия, что проявляется в повороте головы и тела в сторону объекта;
- приспособление может происходить рефлекторно или произвольно;
- эти позы и повороты возможно наблюдать или зафиксировать на видео для последующего анализа.

На данном уровне исследования не различалось непроизвольное (рефлекторное) внимание, когда человек реагирует на громкий звук, свет, внезапное движение или новый объект, и произвольное, когда человек реагирует мотивировано волевым усилием направляя свое внимание на объект изучения. Мы принимаем упрощенную гипотезу Т. Рибо, что верно в обе стороны: человек обладая интересом осознанно направляет свое внимание на объект, или объект, захватившей внимание, вызывает дальнейший интерес[7].

Цель исследования

(1) исследование подходов к оценке внимания по различным паттернам выразительных движений; (2) проектирование модели обработки данных; (3) разработка аналитического модуля технологии распознавания выразительных движений внимания на базе нейронных сетей.

Основная часть

С целью выбора направления исследований потребовался предварительный анализ видеоданных, которые при соблюдении требований по деперсонализации, ограниченному и ответственному доступу были предоставлены партнерами, заинтересованными в их интерпретации и анализе. Партнеры представили для анализа, подготовки моделей, обучения и тестирования нейросетей видеоданные, общим объемом более 1831 час, полученных с 162 точек съемки в разных аудиториях. Видеозаписи использовались для первичного визуального анализа, формулирования рабочей гипотезы и создания задания для разметки дата-сета для нейросети. В данной статье рассматривается только часть выполненной работы интегральную модель обработки данных видеоанализа движений головы.

---

[1] Купцова О. В. Внимание как особый психический процесс / О. В. Купцова // Проблемы современной науки и образования. – 2017. – № 20. – С. 83-85. – URL: https://www.elibrary.ru/item.asp?id=29201991 (дата обращения: 18.04.2020).

[2] Общая психология. Тексты. В 3 т. Т. 3: Субъект познания. Книга 4 / Ред.-сост.: Ю. Б. Дормашев, С. А. Капустин, В. В. Петухов. – М.: Когито-Центр, 2013. // Тема 20. Психология Внимания. – С. 10-489.

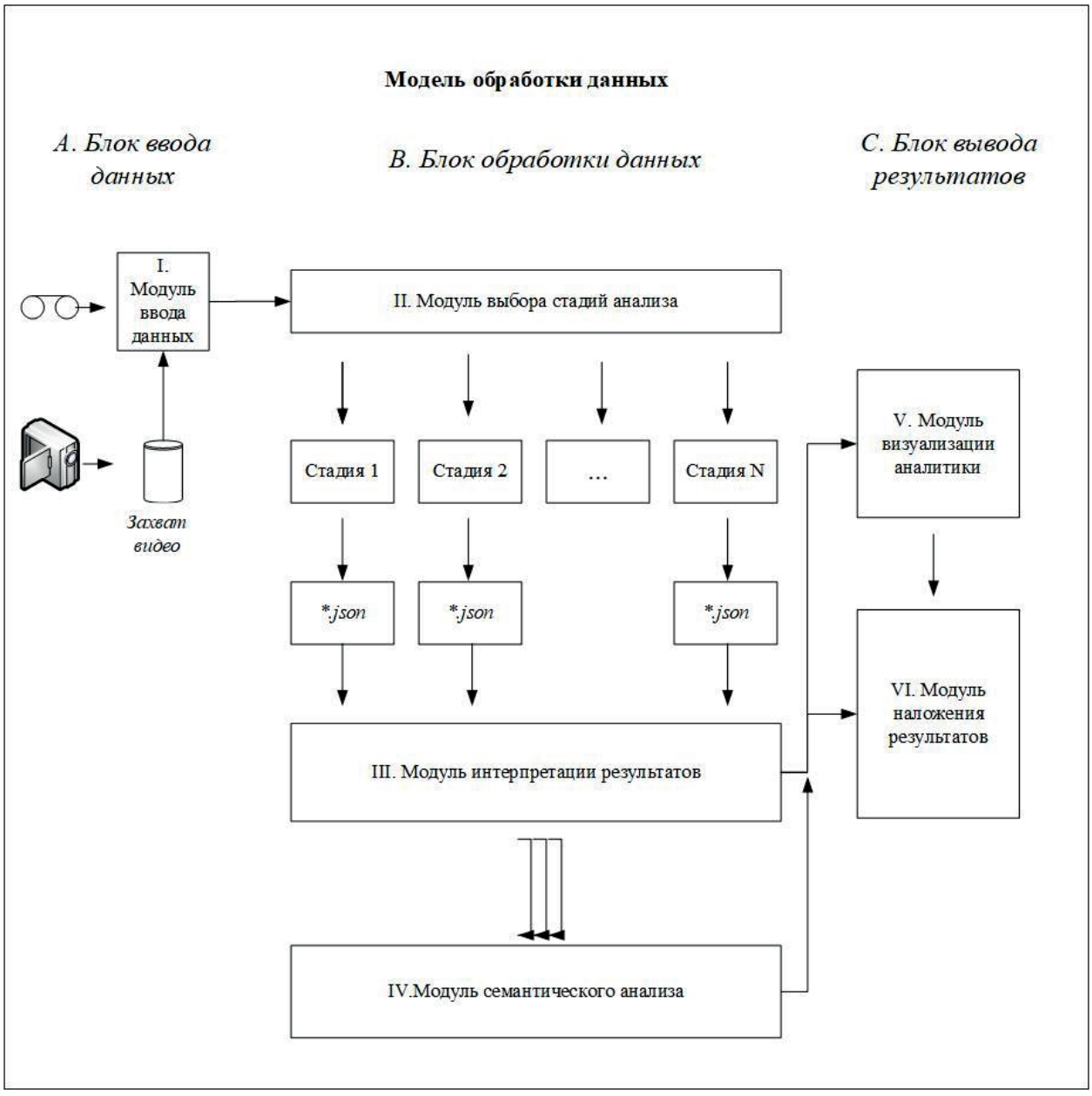
[3] Вассерман, Л. И. Методы нейропсихологической диагностики / Л. И. Вассерман, С. А. Дорофеева, Я. А. Меерсон Я.А. – С.-Пб.: Стройлеспечать, 1997. – URL: http://clinicpsy.ucoz.ru/Library/vasserman_l.i-metody_nejropsikhologicheskoj_diagno.pdf (дата обращения: 18.04.2020).

[4] Гальперин, П. Я. К проблеме внимания // Хрестоматия по вниманию / Под ред. А. Н. Леонтьева, А. А. Пузырей, В.Я. Романова. – М.: Изд-во МГУ, 1976.

[5] Айзенберг, Б. И. Распределение внимания в мыслительной деятельности учащихся массовой и вспомогательной школы: дис. ... канд. психол. наук. – М.: МГПИ им. В.И.Ленина, 1986.

[6] Общая психология. Тексты. В 3 т. Т. 3: Субъект познания. Книга 4 / Ред.-сост.: Ю. Б. Дормашев, С. А. Капустин, В. В. Петухов. – М.: Когито-Центр, 2013. // Тема 20. Психология Внимания. – С. 25.

[7] Общая психология. Тексты. В 3 т. Т. 3: Субъект познания. Книга 4 / Ред.-сост.: Ю. Б. Дормашев, С. А. Капустин, В. В. Петухов. – М.: Когито-Центр, 2013. // Тема 20. Психология Внимания. – С.48-49.



Р и с. 1. Модель обработки данных

F i g. 1. Data processing model

Методы решения задач.

Выбор методов решения задачи полностью обусловлен спецификой и новизной исследуемой области, а именно мониторингом внимания в психологическом контексте. Решать ее предполагается прикладными математическими методами, а именно проектированием нейронных сетей, алгоритмами Искусственного Интеллекта (ИИ). В качестве основного метода оценки внимания рассматривались различные методы компьютерного зрения, позволяющие фиксировать ключевые движения человека.

Существенно, что речь идет о распознавании движения человека в естественной среде, без каких-либо маркеров на его теле (одежде). Данные фиксируются «обычной» IP-видеокамерой с разрешением FullHD, не используются специальное оборудование: инфракрасные камеры, датчики глубины и т.п. Сохраняется естественная для человека среда обучения, минимизируются отвлекающие факторы (помехи).

Экспериментальным путем выведены рекомендованные параметры размещение видеокамер: камера не менее FullHD (2К), с широкоугольным объективом, с фокусным расстоянием 2,8М для малых аудиторий (до 30 кв.м) и 4М для средних аудиторий (от 30кв.м). Точки сбора данных размещаются на фронтальной стене помещения: для анализа сидящих фигур - на высоте от 1,8-2,3м; для анализа занятия, на котором фигуры могут перемещаться по аудитории - на высоте от 2-2,5м.

Расположении камеры, ранее установленное для мониторин-

га безопасности, сзади со спины, по-прежнему присутствует в анализе, однако при таком расположении камер, значительная часть информация теряется даже для восприятия человеком:

- не выделяется лицо человека; полностью теряется оценка направленности взгляда, слежения за ведущим; исключен анализ эмоционального состояния ученика.
- затруднена диагностика позы рук и корпуса, т.к. вид со спины частично закрывает руки спереди.

При этом сохраняется возможность анализа:

- общей картины динамики в классе, перемещение объектов, сбора около доски;
- поворотов назад и отвлечений к соседу, особенно сзади: т.к. большую часть времени лицо не фиксируется, при повороте назад, оно, наоборот, попадает в кадр;
- фиксация лица ведущего, однако расстояние по большей диагонали от камеры при данном качестве видео потока, позволяет фиксировать только общую мимику.

В процессе проверки данных требований в ходе тестов на реальных данных дополнительно выяснилось:

Необходимость использовать как проводные, так и беспроводные способы передачи данных, не нарушая дизайн помещения и не усложняя процедуру необходимостью прокладки сети. Однако, возникли и ограничения - на сегодня подключения к wi-fi происходит на частоте 2,4 ГГЦ. В силу того, что это стандартный рекомендованный диапазон для большинства приборов, этот диапазон сильно зашумлен. При передаче данных с камер по протоколу H.264, при подключении более одной камеры происходит: на визуально замечаемое время потеря кадров при передаче данных, появление артефактов – полос и/или квадратов, засвечивающих кадр и не допускающих его дальнейшую обработку. Соответственно, при передаче «тяжелого» видеопотока данных потребовалось дополнительное специальное перепрограммирование камер и устранение препятствующих эффектов при передаче по Wi-Fi без потери качества.

Было выявлено, что переход на разрешение 4M не дает прибавки качества распознавания по сравнению со стандартным Full HD (2M), более того матрицы, устанавливаемые в Full HD (2M) камеры в среднем дают при той же самой оптике более качественное изображение для распознавания. В связи с этим, несмотря на то, что мы использовали разные типы камер, Full HD (2M) является достаточным.

Отдельно было проведено сравнительное исследование работы проводного и беспроводного соединения: до 6-ти камер качество не меняется, если больше 6-ти камер с потоком более 16Гб (устанавливается на камере), то в этом случае беспроводное соединение уступает по качеству проводному и требуются донастройки системы.

Часть промежуточных шагов, использовавшихся при отладке моделей (например, корректировка распознавания поз), опущена, и далее описана итоговая интегральная модель обработки данных, которая позволяет использовать данные интегрально, а не как набор отдельных показателей. На рис. 1. представлена предлагаемая модель обработки данных.

Модель обработки данных, включает в себя:

I. Блок ввода данных

1) Модуль ввода данных

На входе: два типовых источника видео, данных: stream-поток с видеокамеры, либо уже готовая видеозапись. В случае stream-поток параллельно он записывается. Время обработки stream-потока данных близко к реальному времени. При развертывание полного функционала время обработки увеличивается.

II. Блок обработки данных

2) Модуль выбора схемы анализа.

Модуль выбора количества стадий обработки. К стадиям обработки относятся, к примеру: сегментация людей, детекция лиц, детекция анфас/бэк детекция поз, детекция направления головы. По мере развития системы, необходимости расширения функционала и разработки новых алгоритмов планируется добавление новых стадий.

Выбор количества стадий производится на основании результатов и возможностей отработки каждой стадии. По умолчанию, в случаях, когда человек располагается к камере анфас и его части тела видны и детектируемы, проходит максимально возможный маршрут обработки данных. В случае, если показаний для запуска стадии не выявлено – она пропускается. К примеру, если сработали алгоритмы face-detection, то далее возможен переход к распознаванию направлений головы, если человек стоит спиной – данная стадия пропускается. По итогам каждой стадии – вывод результатов работы в виде *.json файла.

3) Модуль интеграции результатов.

Собираются данные, выявленные по отношению к определенному идентификатору id человека. Соответственно, критерием отнесения к id является близость векторов лица, фигуры и т.п. идентификаторов.

4) Модуль семантического анализа

Фиксируется близость к определенной позе и наименования ее, так называемая семантический анализ. Используются как простые наименования поз: сидит, стоит, поднимает руку, ходит. Так и интегральные специфические показатели: активен/пассивен, вовлечен/не вовлечен.

По возможности, предполагается семантическая сегментация специфических занятий: слушает лекцию, собирает робота.

III. Блок вывода результатов

5) Модуль визуализации аналитики.

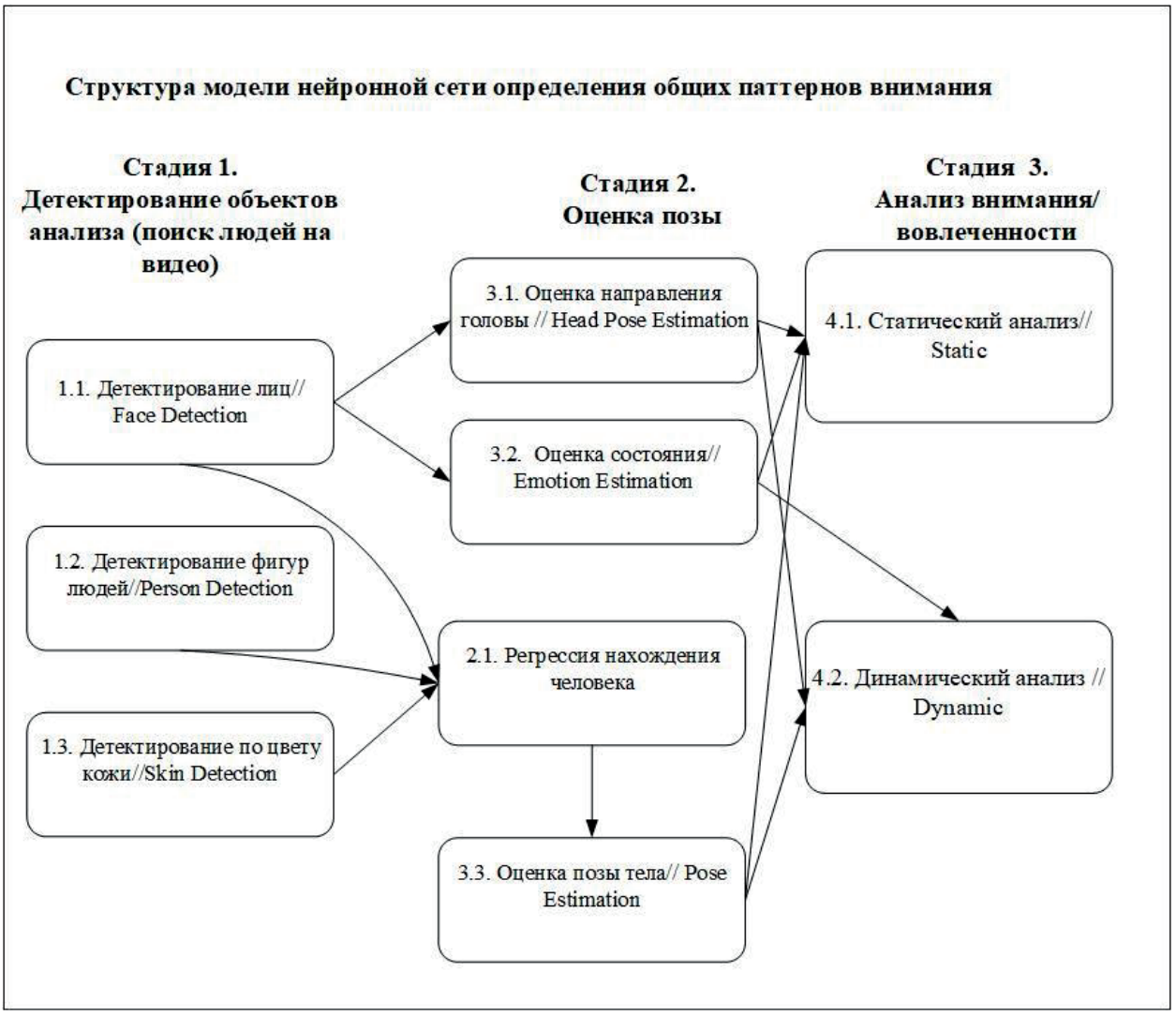
Вывод графиков и визуализация полученной статистики данных в виде пост-отчетов, интегральных отчетов по всем полученным данным за выделенный период.

6) Модуль наложения данных

В настоящее время обработка по отдельным стадиям занимает от 0,1 – 0,2 децисекунды, что воспринимается как обработка в реальном времени. В связи с практически отсутствующим запаздыванием в обработке, возможен обратный вывод и наложение графиков и графических образов на видеопоток.

Таким образом, разработана и представлена структура модели обработки данных и ее элементы.

Итоговая модель нейронной сети спроектирована как стадии совместной обработки данных. В виду сложности и многослойности анализируемого феномена внимания, более корректно в названии модели использовать множественное число – модель «нескольких нейронных сетей». Настройка и выбор стадий происходит гибко. На рис. 2 представлена схема структуры модели нейросети.



Р и с. 2. Структура модели нейронной сети определения общих паттернов внимания

F i g. 2. The structure of a neural network model for defining general patterns of attention

Стадия 1. Детектирование объектов анализа (поиск людей на видео).

- Детектирование лиц (Face Detection)
- Детектирование фигур людей (Person Detection)
- Детектирование по цвету кожи (Skin Detection)

По итогам распознавания лиц и поиска людей на видео, предварительно выяснилось, что стандартные алгоритмы, к примеру каскады Хаара, во многих случаях срабатывают лучше, чем нейросетевой. Визуально, качество распознавание и количество опознанных объектов значимо выше. Такими условиями стали: не очень качественное изображение с камер после использования сжатия, алгоритмы показывают более высокое качество

Стадия 2. Детектирование лиц

- Регрессия нахождения человек, фиксация его id.

Стадия 3. Оценка позы

- Оценка направления головы (Head Pose Estimation)
- Оценка состояния (Emotion Estimation)
- Оценка позы тела (Pose Estimation)

Стадия 4. Анализ внимания/вовлеченности

- Статический анализ (Static)/ оценка состояний
- Динамический анализ (Dynamic)

Статистический анализ: общее состояние группы людей в кадре. Динамический анализ – изменение движения и позиций людей по отношению друг к другу и стационарным предметам, к примеру: стол, стул.

Определение последовательности движения на видео и/или позы на одном кадре с помощью нейронных сетей состоит из следующих основных этапов:

1. Входной кадр проходит через сверточную нейронную сеть, выделяющую местонахождение субъекта.
2. Происходит поиск и привязка к ключевым точкам.
3. Изображение декодируется целиком.
4. На изображение накладывается сетка;

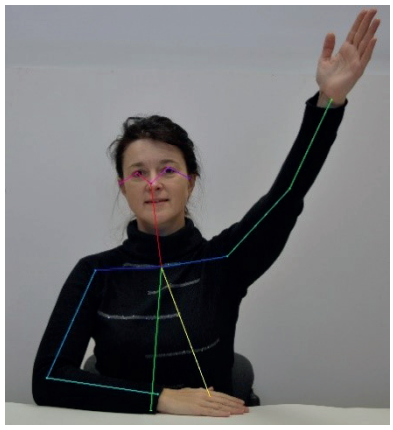
Аналитические расчеты. Аналитические расчеты включают интегральные показатели в целом по группе с учетом времени наблюдения

% Внимания_in _time=∑People_attention/∑People
% Внимания_in_room= Mediana {% Внимания_in _time_j}
При анализе видео, исходное изображение проходит процесс детектирования и захвата лиц в кадре. Сигнатура лица используется только для созданий внутреннего словаря Face ID с целью дополнительного отнесения позы к фигуре, как уникальному объекту. Однако, сохранения данных по FaceID между экспериментальными сессиями не предусмотрено в виду законодательных ограничений на обработку персональных данных. Данные собираются и оцениваются интегрально по группе. Это позволяет:
1) учитывать статистику в условиях неполных данных (голова скрыта или частично видна);
2) реализовать «мерцающий» трекинг (когда голова временно пропадает из поля обзора камер);
3) аппроксимировать данные при временных разрывах данных.
Итоговая модель нейронной сети спроектирована как стадии совместной обработки данных. В виду сложности и многослойности анализируемого феномена внимания, более корректно в названии модели использовать множественное число – модель «нескольких нейронных сетей». Настройка и выбор стадий происходит гибко.

Полученные результаты

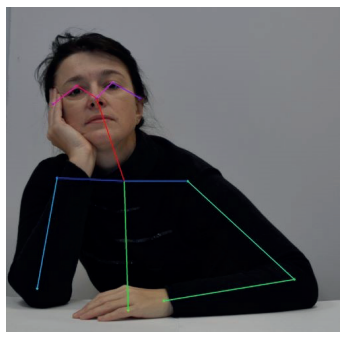
Задача выделения паттернов выразительных движений внимания и их семантическая интерпретация в реальном времени в реальных условиях является типовой для человека и совершенно нетривиальной задачей для нейронных сетей. Для ее решения потребовалась использовать различные подходы как нейросетевые, так и методы предварительной обработки видео.
Результатом аппроксимации движений головы, является два массива данных: данные о векторах и данные о ключевых точках. На основании данных векторов строится статическая аналитика (наличие или отсутствие состояния). Данные о смещении точек фигуры используются при анализе динамики изменений по конкретной фигуре.
С точки зрения хранения и обработки данных, передача данных производится с помощью словарей и массивов Numpy Python, и близких к ним файлов/объектов JSON, что позволяет использовать noSQL СУБД MongoDB.
В качестве примера, промежуточный вектор интегральной модели совместного использования нейросетевых моделей может содержать следующую информацию:
1. Код камеры
2. Номер кадра
3. Время (установленное на камере)
4. Интегральные вектора в формате:
4.1. ID объекта в кадре
4.2. Координаты объекта (голов(ы), человека) в кадре, в виде углов координаты прямоугольника в виде Face bbox ((xi1,yi1), (xi2,yi2))
4.3. Код модуля распознавания (например, head_position для направления головы)
4.4. Дополнительный служебный идентификатор (код сетки /версии программы и т.д.)
Векторы с наборами значений в виде «код: значение», для положения головы – углы в 3х координатах: (Xi;Yi;Zi)

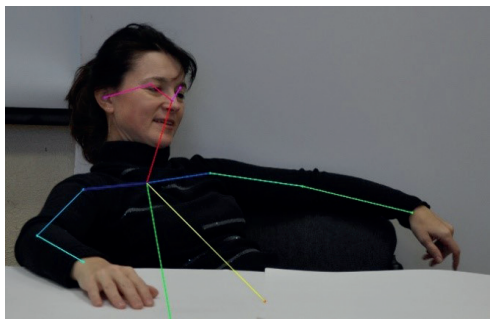
По итогам совместного использования нейросетевых моделей в рамках интегральной модели, можно сделать вывод, что оно в значительной степени увеличило качество работы алгоритма, в том числе – визуально замечаемое распознавание по итогам нанесенной поверх данных в кадре разметки.
а) Позы активности и вовлечения в работу. При активном вовлечении в работу, есть значимый маркер мониторинга – поднятая рука, обозначающая готовность ответить на вопрос. На рис. 3 показана аппроксимация позы вовлеченности в работу - поднятой руки



Р и с. 3. Аппроксимация позы «Поднятая рука»
F i g. 3. Raised Hand Pose Approximation

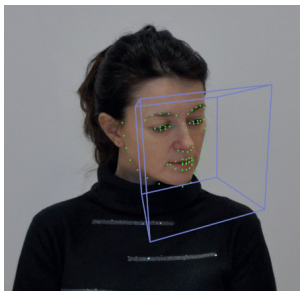
б) Позы усталости/невнимания. При усталости позы меняются. Слушателю требуется поддержка головы сначала рукой, а затем часто при скуке и усталости слушатели начинают опираться на стол всем телом или откидывание тела назад. На рис. 4 показана аппроксимация позы усталости





Р и с. 4. Аппроксимация поз усталости/невнимания
F i g. 4. Fatigue / Inattention Pose Approximation

Одним из значимых показателей внимания является направленность взгляда ученика или более в общем случае оценка направленность головы. Смотрит ли он на педагога во время объяснения материала, или в материалы, критичным является отклонение более чем на 90 градусов в направлении окна или соседа. В модели предполагается оценка направления головы.

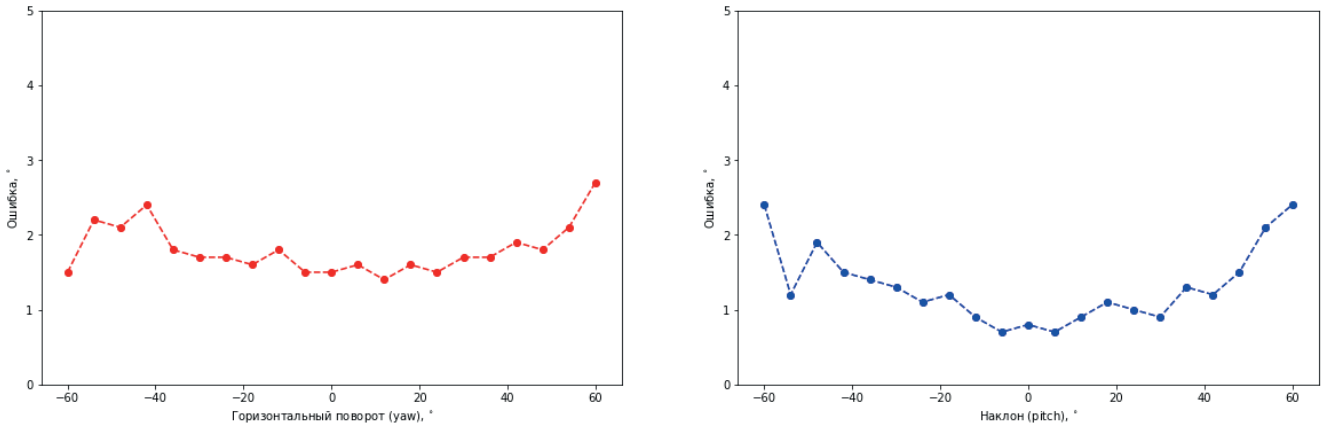


Р и с. 5. Аппроксимация направления головы (Head Pose Estimation)
F i g. 5. Head Direction Approximation

Для оценки качества полученных результатов нами была выполнена оценка по открытому набору данных BIWI head pose, имеющему аннотированные углы поворота головы данных[8]. Наклоны вправо / влево (roll) не учитывались в связи с тем, что они, с точки зрения оценки внимания психологом, равнозначно интерпретируемы положительно и слабо влияют на направление внимания. хотя указанный набор данных содержит информацию о поворотах головы в диапазоне -75/+75 градусов, мы ограничились диапазоном -60/+60 градусов, которые покрывает основные варианты практического применения. В качестве ошибка рассматривалась разница в реальном и полученном углах.

Оценка точности угла поворота головы нейросетевой моделью распознавания представлена на графике на рис. 6.

На графиках видно, что ошибка составляет от 1 до 3-х градусов, для целей оценки поворота головы ученика данная точность более чем достаточна.



Р и с. 6. Графики оценки точности угла поворота головы нейросетевой моделью
F i g. 6. Graphs for estimating the accuracy of the head rotation angle by a neural network model

Заключение

Предложенные в статье подходы к оценке внимания по наблюдаемым паттернам выразительных движений позволяют рассмотреть возможности реализовать цифровизацию в такой области как образование, рассмотреть применение подходов распознавания на базе нейронных сетей в образовательную деятельность, повысить эффективность обучающих программ, в особенности он-лайн и дистанционных программ, за счет оперативной обратной связи ведущим обучающих мероприятий и возможности в реальном времени корректировать учебный процесс и методические материалы.

References

[1] Pulli K., Baksheev A., Kornyakov K., Eruhimov V. Real-time computer vision with OpenCV. *Communications of the ACM*. 2012; 55(6):61-69. (In Eng.) DOI: https://doi.org/10.1145/2184319.2184337

[2] Huang R., Pedoeem J., Chen C. YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. In: *2018 IEEE International Conference on Big Data (Big Data)*. Seattle, WA, USA; 2018. p. 2503-2510. (In Eng.) DOI: https://doi.org/10.1109/BigData.2018.8621865

[3] Xiao B., Wu H., Wei Y. Simple Baselines for Human Pose Estimation and Tracking. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (ed.) Computer Vision – ECCV 2018. ECCV 2018. *Lecture Notes in Computer Science*. 2018; 11210:472-487. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-030-01231-1_29

[4] Zafeiriou S., Zhang C., Zhang Z. A Survey on Face Detection in the wild: past, present and future. *Computer Vision and Image Understanding*. 2015; 138:1-24. (In Eng.) DOI: http://dx.doi.org/10.1016/j.cviu.2015.03.015

[5] Cao Z., Hidalgo G., Simon T., Wei S. -E., Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021; 43(1):172-186. (In Eng.) DOI:

[8] Fanelli, G. Random Forests for Real Time 3D Face Analysis / G. Fanelli, M. Dantone, J. Gall, A. Fossati, L. Van Gool. – DOI 10.1007/s11263-012-0549-0 // International Journal of Computer. – 2013. – Vol. 101, Issue 3. – Pp. 437-458. – URL: https://doi.org/10.1007/s11263-012-0549-0 (дата обращения: 18.04.2020).

https://doi.org/10.1109/TPAMI.2019.2929257

[6] Sun K., Xiao B., Liu D., Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA; 2019. p. 5686-5696. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2019.00584.

[7] Chen S., Yang R.R. Pose Trainer: Correcting Exercise Posture using Pose Estimation. *arXiv*:2006.11718 [cs.CV]. 2020. (In Eng.)

[8] Andriluka M., Pishchulin L., Gehler P., Schiele B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH; 2014. p. 3686-3693. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2014.471

[9] Toshev A., Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH; 2014. p. 1653-1660. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2014.214

[10] Pishchulin L., Andriluka M., Gehler P., Schiele B. Poselet Conditioned Pictorial Structures. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR; 2013. p. 588-595. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2013.82

[11] Cao Z., Simon T., Wei S., Sheikh Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI; 2017. p. 1302-1310. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2017.143

[12] Lin T.Y. et al. Microsoft COCO: Common Objects in Context. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (ed.) Computer Vision – ECCV 2014. ECCV 2014. *Lecture Notes in Computer Science*. 2014; 8693:740-755. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-10602-1_48

[13] Yang Z., Li Y., Yang J., Luo J. Action Recognition With Spatio-Temporal Visual Attention on Skeleton Image Sequences. *IEEE Transactions on Circuits and Systems for Video Technology*. 2019; 29(8):2405-2415. (In Eng.) DOI: https://doi.org/10.1109/TCSVT.2018.2864148

[14] Gupta A., Agrawal D., Chauhan H., Dolz J., Pedersoli M. An Attention Model for Group-Level Emotion Recognition. In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*. Association for Computing Machinery, New York, NY, USA; 2018. p. 611-615. (In Eng.) DOI: https://doi.org/10.1145/3242969.3264985

[15] Guo X., Polanía L.F., Barner K.E. Group-level emotion recognition using deep models on image scene, faces, and skeletons. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. Association for Computing Machinery, New York, NY, USA; 2017. p. 603-608. (In Eng.) DOI: https://doi.org/10.1145/3136755.3143017

[16] Girdhar R., Ramanan D. Attentional Pooling for Action Recognition. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA, USA; 2017. p. 1-12. (In Eng.)

[17] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV; 2016. p. 770-778. (In Eng.) DOI: https://doi.org/10.1109/CVPR.2016.90

[18] Rosner T.M., D'Angelo M.C., MacLellan E., Milliken B. Selective attention and recognition: effects of congruency on episodic learning. *Psychological Research*. 2015; 79:411-424. (In Eng.) DOI: https://doi.org/10.1007/s00426-014-0572-6

[19] Cheng Z., Bai F., Xu Y., Zheng G., Pu S., Zhou S. Focusing Attention: Towards Accurate Text Recognition in Natural Images. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice; 2017. p. 5086-5094. (In Eng.) DOI: https://doi.org/10.1109/ICCV.2017.543

[20] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. San Diego, CA, USA; 2015. Available at: http://arxiv.org/abs/1409.1556 (accessed 18.04.2020). (In Eng.)

[21] Wang L., Xiong Y., Wang Z., Qiao Y., Lin D., Tang X., Van Gool L. Temporal Segment Networks for Action Recognition in Videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019; 41(11):2740-2755. (In Eng.) DOI: https://doi.org/10.1109/TPAMI.2018.2868668

[22] Schuldt C., Laptev I., Caputo B. Recognizing human actions: a local SVM approach. In: *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*. Vol. 3. Cambridge; 2004. p. 32-36. (In Eng.) DOI: https://doi.org/10.1109/ICPR.2004.1334462

[23] Kuehne H., Jhuang H., Garrote E., Poggio T., Serre T. HMDB: A large video database for human motion recognition. In: *Proceedings of the 2011 International Conference on Computer Vision (ICCV '11)*. IEEE Computer Society, USA; 2011. p. 2556-2563. (In Eng.) DOI: https://doi.org/10.1109/ICCV.2011.6126543

[24] Pham H.H., Salmane H., Khoudour L., Crouzil A., Zegers P., Velastin S.A. Spatio-Temporal Image Representation of 3D Skeletal Movements for View-Invariant Action Recognition with Deep Convolutional Neural Networks. *Sensors*. 2019; 19(8):1932. (In Eng.) DOI: https://doi.org/10.3390/s19081932

[25] Wang L., Qiao Y., Tang X. MoFAP: A Multi-level Representation for Action Recognition. *International Journal of Computer Vision*. 2016; 119(3):254-271. (In Eng.) DOI: https://doi.org/10.1007/s11263-015-0859-0

Поступила 18.04.2020; принята к публикации 13.07.2020; опубликована онлайн 30.09.2020.
Submitted 18.04.2020; revised 13.07.2020; published online 30.09.2020.

Об авторах:

Артамонова Яна Николаевна, генеральный директор, ООО «Нейрокорпус» (121415, Россия, г. Москва, ш. Варшавское, д. 1, стр. 1-2), ORCID: http://orcid.org/0000-0002-4947-1562, ceo@neurocorp.ru

Артамонов Игорь Михайлович, начальник отдела информационных сетей, ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)» (125993, Россия, г. Москва, ш. Волоколамское, д. 4), ORCID: http://orcid.org/0000-0001-8343-4821, iartamonov@gmail.com

Благодарности: команда выражает особую благодарность доктору физико-математических наук, профессору НИИ системных исследований РАН Виталию Львовичу Дунин-Барковскому.

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Yana N. Artamonova, Research Manager, Neurocorpus LLC (1/1-2 Warsaw Highway, Moscow 121415, Russia), ORCID: http://orcid.org/0000-0002-4947-1562, ceo@neurocorp.ru

Igor M. Artamonov, Head of the Information Networks Department, Moscow Aviation Institute (National Research University) (4 Volokolamsk Highway, Moscow 125993, Russia), ORCID: http://orcid.org/0000-0001-8343-4821, iartamonov@gmail.com

Acknowledgement: The team is particularly grateful to Vitaly Lvovich Dunin-Barkovsky, Doctor of Physical and Mathematical Sciences, Professor of the Federal State Institution Scientific Research Institute for System Analysis of the Russian Academy of Sciences.

All authors have read and approved the final manuscript.