

Видеорегистрация присутствия студентов в аудитории на основе нейросетевого распознавания лиц

А. А. Суханов, Д. Э. Баев, О. И. Федяев

E-mail: studysukhanov@mail.ru, azo.cw@yandex.ru, olegfedyayev@yahoo.com

Аннотация:

В данной статье описана архитектура свёрточной нейронной сети VGGFace, на основе которой разработана схема компьютерного распознавания человека по его лицу. Выполнена программная реализация системы, которая осуществляет автоматическое ведение электронного журнала группы. Проведены эксперименты по видеорегистрации студентов при входе в аудиторию с помощью компьютерного зрения.

Введение

Искусственный интеллект находит все большее применение в различных сферах человеческой деятельности [1], в том числе и в образовании [2-5].

Актуальность данной задачи обусловлена, во-первых, потерями времени преподавателя, которые возникают при ручной регистрации присутствия большого количества учащихся в аудитории (в потоке из 70 человек потери на «переключку» составляют в среднем 85 мин в семестре и даже больше); во-вторых, «переключка» проходит шумно и не всегда достоверно [6].

Основные трудности компьютерного распознавания лиц в реальном времени связаны с быстрой изменчивостью изображений объектов в видеопотоке: положение, размер и ракурс лица в кадре, освещение и т. д. В настоящее время большие перспективы в преодолении перечисленных проблем связывают с применением глубоких нейронных сетей. К этому классу относится многослойная свёрточная нейронная сеть.

На данный момент свёрточная нейронная сеть и её модификации считаются лучшими по точности и скорости распознавания объектов на изображении.

Более того, распознавание лиц из видеопотока с камеры показывает, что нейронные сети с такой архитектурой способны работать в режиме реального времени даже на устройствах с ограниченными ресурсами [7].

Поэтому целью данной работы является оценка возможности реализации нейросетевого распознавания лиц из видеопотока на базе существующих инструментальных средств и создание системы оперативного визуального контроля присутствия студентов на учебных занятиях, что подчёркивает её уникальность.

Выделение признаков во входном изображении свёрточной нейронной сетью

Свёрточная нейронная сеть — это специализированный тип нейросетевой модели, предназначенной для работы с данными двумерных изображений, хотя они могут использоваться и с одномерными, и с трёхмерными данными.

Центральным элементом свёрточной нейронной сети является свёрточный слой, благодаря которому сеть и получила свое название. Этот слой выполняет операцию, называемую «свёрткой».

В контексте свёрточной нейронной сети свёртка — это линейная операция, которая включает в себя умножение набора весов на вход, как в традиционной нейронной сети. Учитывая, что техника была разработана для двумерного ввода, умножение выполняется между массивом входных данных и двумерным массивом весов, называемым фильтром или ядром.

Фильтр меньше входных данных, и тип умножения, применяемый между участком входных данных, представляет собой точечное произведение. Точечное произведение — это умножение по элементам между участками входных данных и фильтра, которые затем суммируются, в результате чего всегда получается одно значение. Поскольку в результате получается одно значение, эту операцию часто называют «скалярным произведением».

Использование фильтра меньшего размера, чем входные данные, является умышленным, поскольку это позволяет нам умножать один и тот же фильтр, набор весов, на входной массив несколько раз в разных точках входных данных. В частности, фильтр систематически применяется к каждой перекрывающейся части или участку входных

данных размером с фильтр, слева направо, сверху вниз.

Систематическое применение одного и того же фильтра ко всему изображению – это хорошая идея. Если фильтр предназначен для обнаружения определенного типа признаков на входных данных, то систематическое применение этого фильтра на всем входном изображении дает ему возможность обнаружить этот признак в любой точке изображения. Эту возможность принято называть трансляционной инвариантностью, то есть общий интерес заключается в том, присутствует ли признак, а не в том, где он был обнаружен.

Результатом однократного умножения фильтра на входной массив является одно значение. Поскольку фильтр многократно применяется к входному массиву, результатом является двумерный массив выходных значений, которые представляют собой фильтрацию входных данных. Таким образом, двумерный выходной массив, полученный в результате этой операции, называется «картой признаков» (рис. 1) [7].

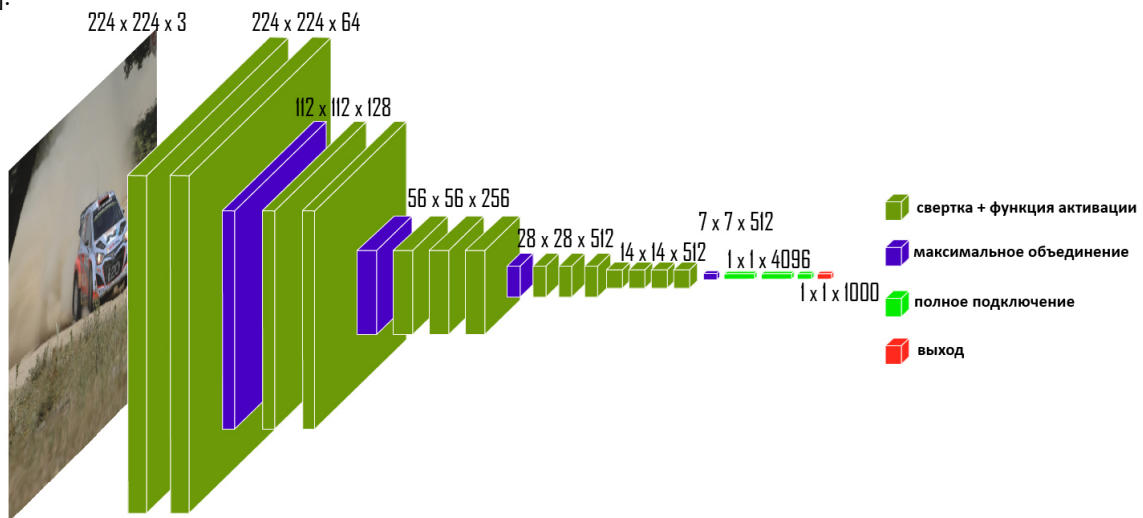


Рисунок 2 – Архитектура VGGFace

VGGFace известна своей простотой и эффективностью, а также способностью достигать высоких результатов в различных задачах компьютерного зрения, включая классификацию изображений и распознавание объектов. Архитектура модели представляет собой стек конволюционных слоев, за которыми следуют слои с максимальным объединением, с постепенным увеличением глубины. Такая конструкция позволяет модели обучаться сложным иерархическим представлениям визуальных признаков, что приводит к надежным и точным предсказаниям. Несмотря на свою простоту по сравнению с более современными архитектурами VGG остается популярным выбором для многих приложений глубокого обучения благодаря своей

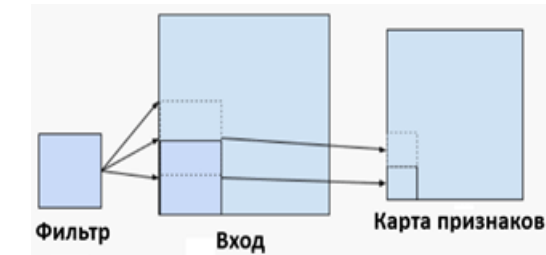


Рисунок 1 – Схема применения фильтра к двумерному входному изображению для создания карты признаков

Архитектура свёрточной нейронной сети модели VGGFace

Модель VGGFace – это архитектура свёрточной нейронной сети, предложенная группой Visual Geometry Group в Оксфордском университете [8]. Она отличается своей глубиной, состоящей из 16 слоев, включая 13 свёрточных и 3 полностью связанных слоя (рис. 2).

универсальности и отличной производительности.

Данная архитектура с функциями активации, приведенные на рисунке, обучалась и тестировалась на наборе данных ImageNet, который содержал 14 млн. изображений, относящихся к 1000 классов [8].

Видеоконтроль присутствия студентов на занятии с помощью компьютерного зрения

Схема видеорегистрации студентов при входе в аудиторию осуществляется в реальном времени. На рисунке 3 показана функциональная схема распознавания студентов по изображению лица из видеопотока.

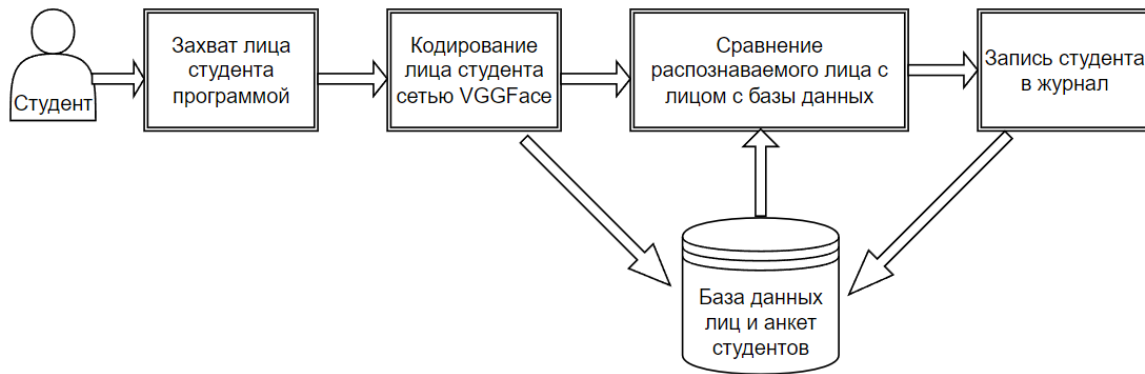


Рисунок 3 – Схема видеорегистрации студентов при входе в аудиторию

При входе в аудиторию, камера, подключенная к компьютеру преподавателя, локализует лицо студента, в результате чего получается изображение локализованного лица. Далее данное изображение передается в блок формирования вектора признаков лиц свёрточной нейронной сетью VGGFace, которая выделенному лицу ставит в соответствие вектор признаков, взятый с базы данных лиц и анкет студентов. Процедура сравнения основывается на методе вычисления косинусного сходства вектора распознаваемого лица с каждым вектором-эталонном из базы данных по формуле 1:

$$\text{Сходство} = \frac{Y \cdot \bar{Y}}{\|Y\| \cdot \|\bar{Y}\|} = \frac{\sum_{i=1}^n y_i \bar{y}_i}{\sqrt{\sum_{i=1}^n y_i^2} \cdot \sqrt{\sum_{i=1}^n \bar{y}_i^2}}, \quad (1)$$

где Y и $\bar{Y}$ – вектора признаков соответственно распознаваемого лица и лица-эталона из базы данных; $n = 2622$ (n – количество классов, т.е. распознаваемых лиц).

Для выработки признаков лица свёрточная нейронная сеть предварительно была обучена её создателями на примерах фотографий 2622-х человек (по 1000 фотографий на человека). Сеть настроена на классификацию распознаваемого лица, используя в качестве классов лица из обучающего множества. Поэтому результатом работы сети является 2622-мерный вектор, каждый элемент которого представляет собой вероятность сходства лица с одним из обучающего множества. Считается, что два изображения лица относятся к одному человеку, если они в одинаковой мере похожи на каждое лицо из обучающего множества. Для этого вектора признаков этих изображений в пространстве лиц из обучающего множества должны образовывать между собой достаточно острый угол [6, 9].

Распознаваемое лицо считается соответствующим эталону, если полученный коэффициент сходства выше определённого значения (в работе использовалось значение 0,7). После успешного сравнения лиц, происходит

запись в электронный журнал о явке студента на занятие.

Последовательная работа слоёв нейросети организована таким образом, чтобы осуществлялся переход от конкретных особенностей изображения к более абстрактным деталям, и далее к ещё более абстрактным. В этом процессе ключевую роль играет операция свёртки (см. форм. 2):

$$C_{i,j} = f\left(\sum_{k=1}^{m_y} \sum_{l=1}^{m_x} A_{i+k-1,j+l-1} B_{k,l}\right), \quad (2)$$

где A – матрица размером $(n_x \times n_y)$; B – матрица (ядро свёртки) размера $(m_x \times m_y)$;

C – результирующая матрица размера $[(n_x - m_x + 1) \times (n_y - m_y + 1)]$;
 $i=1,2,\dots, n_x - m_x + 1$; $j=1,2,\dots, n_y - m_y + 1$;
 $f()$ – функция активации.

Подвыборочный слой также как свёрточный имеет карты, но их количество совпадает с предыдущим (свёрточным) слоем. Его задача – уменьшить размерность карт предыдущего слоя. Для начала работы с системой требуется создать группу и анкету студента с фотографиями. Создание анкеты представлено на рисунке 4.

Рисунок 4 – Форма создания анкеты студента

Фотографию в анкету можно загрузить двумя способами: использовать подготовленное изображение, сделанное с фотоаппарата либо телефона; сделать фото с веб-камеры. Примеры загрузки изображений представлены на рисунках 5 и 6.

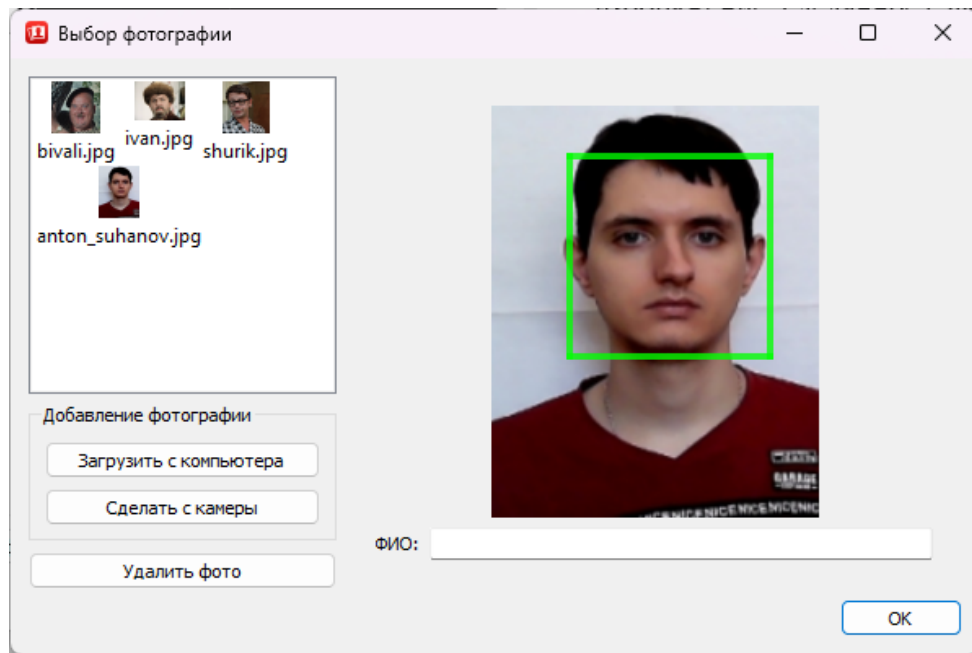


Рисунок 5 – Загрузка подготовленной фотографии

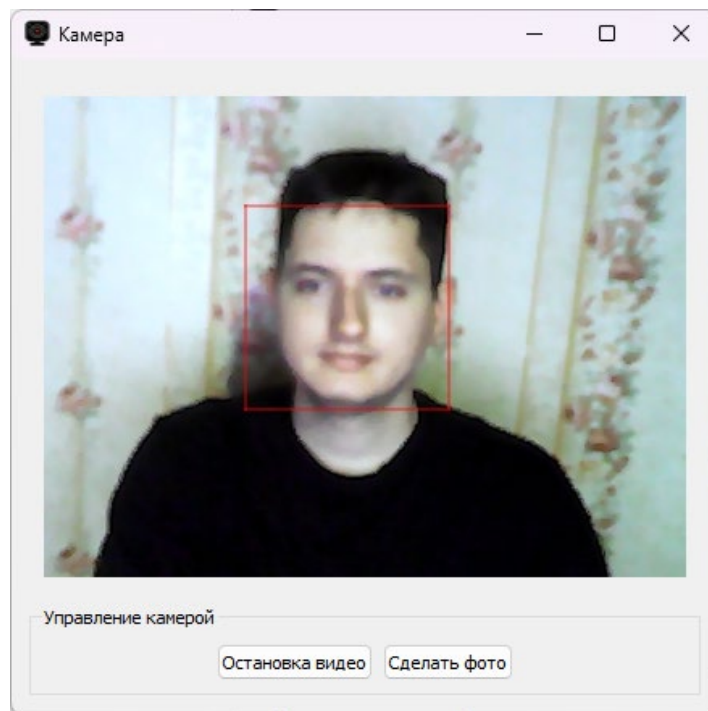


Рисунок 6 – Загрузка фото с камеры

После добавления анкет студентов и их фотографий в базу данных, на главной форме системы получим следующую таблицу, изображённую на рисунке 7.

Проведены эксперименты для оценки работы системы при предъявлении видеокамере фотографий лиц на бумажном носителе. Если выделенное лицо на изображении распознано, то в электронном журнале делается отметка о времени прибытия студента. Эксперименты

проводились на персональном компьютере со следующими техническими характеристиками:

- процессор: Intel Core i5-9400F CPU 2.90GHz;
- оперативная память: 16 ГБ;
- тип системы: 64-разрядная операционная система, процессор x64;
- видеокарта: NVIDIA GeForce RTX 4060 TI;
- система: Windows.

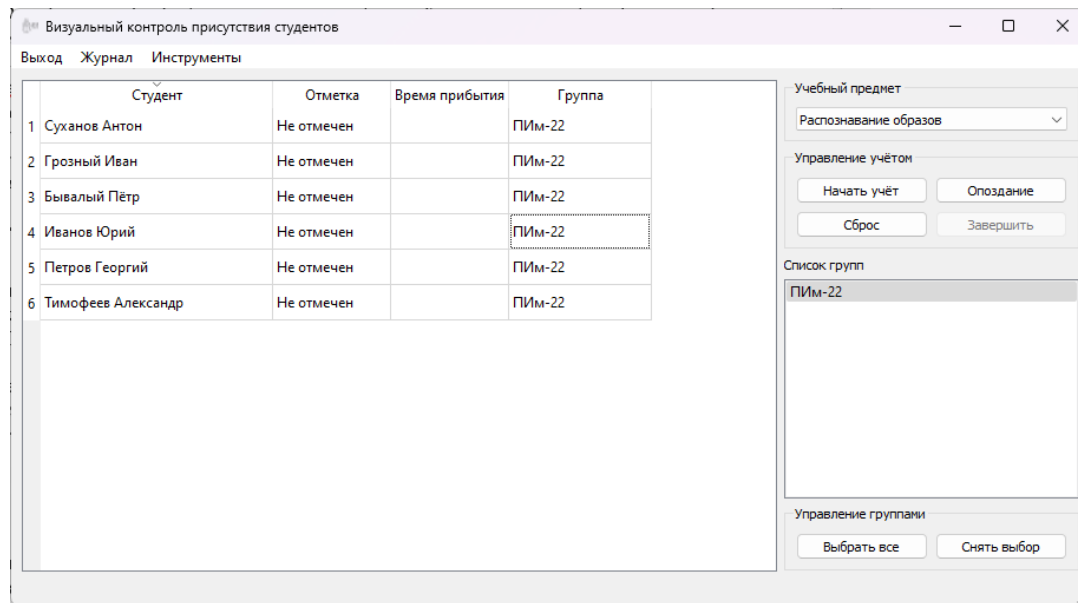


Рисунок 7 – Главная форма системы

Технические характеристики веб-камеры:

- разрешение в pix: 640*480;
- тип матрицы: CMOS;
- фокусное расстояние: 2,8 мм.;
- частота кадров: до 30 кадров/секунду

при разрешении в 640*480.

Выбрав группу для контроля и нажав на кнопку «Начать учёт», система начнёт видеорегистрацию студентов. Результат распознавания студентов с отметкой времени прибытия представлен на рисунке 8.

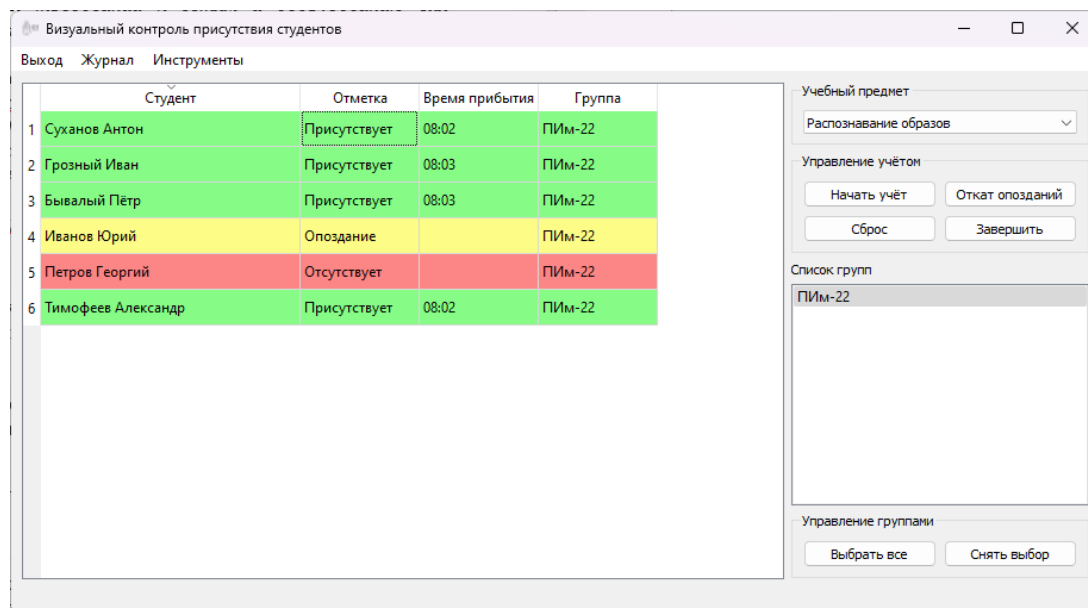


Рисунок 8 – Результат учета студентов

Как видно из результатов работы программы, изображенных на рисунке 8, программа успешно справилась с экспериментами.

Выводы

В данной статье представлена и описана архитектура свёрточной нейронной сети VGGFace. Разработана схема компьютерного

распознавания человека по его лицу. Расширен функционал системы регистрации студентов, позволяющий формировать видеобазу студентов по фотографиям, вводимых из фалов и видеокамеры.

Проведены эксперименты по распознавания студентов с автоматической отметкой в электронном журнале на небольшом объёме видео данных. При переходе обучения в режим оффлайн система будет тестироваться в

реальных условиях присутствия студентов группы (потока) в аудитории. Внедрение системы позволит расширить возможности виртуальной кафедры [10].

Литература

1. Федяев О.И., Бакаленко В.С. Интеллектуальная система принятия решений в отделении медицинского учреждения на основе нейросетевых, продукционных и статистических моделей. *Статистика и Экономика*. 2019;16(3):70-77. <https://doi.org/10.21686/2500-3925-2019-3-70-77>
2. Дворяткина, С. Н. Интеграция фрактальных и нейросетевых технологий в педагогическом контроле и оценке знаний обучаемых / С. Н. Дворяткина // Вестник РУДН. Серия : Психология и педагогика. - 2016. - Том. 14. - № 4. - С. 451-465.
3. Федяев, О. И. Прогнозирование остаточных знаний студентов по отдельным дисциплинам с помощью нейронных сетей / О. И. Федяев // Известия ЮФУ. Технические науки. - 2016. - С. 122-136.
4. AI в обучении: на что способны технологии уже сейчас? // EduTech. - 2022. - № 4 [49]. - 60 с.
5. Струнин, Д. А. Искусственный интеллект в сфере образования / Д. А. Струнин. — Текст : непосредственный // Молодой ученый. — 2023. — № 6 (453). — С. 15-16. — URL: <https://moluch.ru/archive/453/99921/>

6. Федяев О.И., Коломойцева И.А. Автоматическая регистрация присутствия студентов на учебном занятии с помощью компьютерного зрения // XXI Национальная конференция по искусственному интеллекту с международным участием КИИ-2023 (Смоленск, 16-20 октября 2023 г.). Труды конференции. В 2-х томах. Т.1. – Смоленск: Принт-Экспресс, 2023. – С. 294-303.

7. Machine Learning Mastery. Making Developers Awesome at Machine Learning [Электронный ресурс] / Интернет-ресурс. – Режим доступа URL: <https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/>. – Загл. с экрана.

8. GeeksforGeeks | A computer science portal for geeks. VGG-16 CNN model [Электронный ресурс] / Интернет-ресурс. – Режим доступа URL: <https://www.geeksforgeeks.org/vgg-16-cnn-model/>. – Загл. с экрана.

9. Blog. Great Learning. Introduction to VGG16 | What is VGG16? model [Электронный ресурс] / Интернет-ресурс. – Режим доступа URL: <https://www.mygreatlearning.com/blog/introduction-to-vgg16/#VGG%20-%20The%20Idea>. – Загл. с экрана

10. Федяев О.И., Жабская Т.Е. Проектирование виртуальной кафедры университета на основе многомодельного агентно-ориентированного подхода // Искусственный интеллект. – 2010. – № 3. – С. 679-686.

Суханов А.А., Баев Д.Э., Федяев О.И. Видеорегистрация присутствия студентов в аудитории на основе нейросетевого распознавания лиц. В данной статье описана архитектура свёрточной нейронной сети VGGFace, на основе которой разработана схема компьютерного распознавания человека по его лицу. Выполнена программная реализация системы, которая осуществляет автоматическое ведение электронного журнала группы. Проведены эксперименты по видеорегистрации студентов при входе в аудиторию с помощью компьютерного зрения.

Ключевые слова: свёрточная нейронная сеть, распознавание лиц, нейронная сеть VGGFace, компьютерное зрение, регистрация студентов.

Sukhanov A.A., Baev D.E., Fedyaev O.I. Video recording of students' presence in the classroom based on neural network facial recognition. This paper describes the architecture of the convolutional neural network VGGFace, on the basis of which the scheme of computer recognition of a person by his face is developed. The program realization of the system, which performs automatic maintenance of the electronic group journal, is performed. Experiments on video registration of students at the entrance to the classroom with the help of computer vision are carried out.

Keywords: convolutional neural network, face recognition, VGGFace neural network, computer vision, student registration.

Статья поступила в редакцию 12.06.2024
Рекомендована к публикации профессором Григорьевым А. В.