

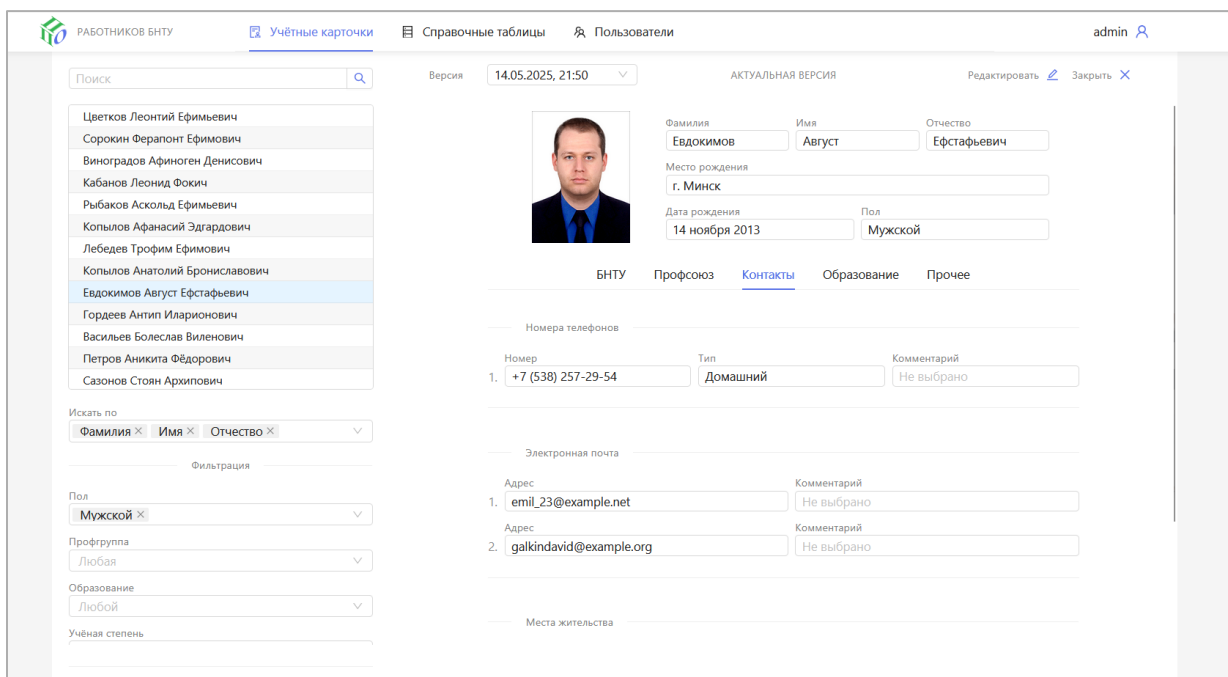


Рис.2. Веб-страница для работы с учётными карточками членов ППО работников БНТУ.

Литература

1. React Documentation [Электронный ресурс]. – Режим доступа: <https://react.dev/learn/> – Дата доступа: 14.05.2025.
2. Django Documentation [Электронный ресурс]. – Режим доступа: <https://docs.djangoproject.com/> – Дата доступа: 14.05.2025.
3. PostgreSQL Documentation [Электронный ресурс]. – Режим доступа: <https://www.postgresql.org/docs/> – Дата доступа: 14.05.2025.
4. TypeScript Documentation [Электронный ресурс]. – Режим доступа: <https://www.typescriptlang.org/docs/> – Дата доступа: 14.05.2025.
5. ППО работников БНТУ [Электронный ресурс]. – Режим доступа: <https://bntu.by/departments/ppo/> – Дата доступа: 14.05.2025.

УДК 004.942

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ЧТЕНИЯ ПО ГУБАМ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТИ

Кругликовский М.А.

Научный руководитель –Белова С.В., ст. преподаватель

Технология автоматического чтения по губам является одним из перспективных направлений на стыке компьютерного зрения и распознавания речи. Данная технология имеет широкий спектр

практического применения: от создания вспомогательных систем для людей с нарушениями слуха до обработки видеозаписей в условиях отсутствия звука или при наличии сильного шумового фона.

Традиционные методы чтения по губам основывались на детектировании контрольных точек губ и классификации их форм. Однако такой подход имеет существенные ограничения в точности распознавания, особенно при работе с непрерывной речью и в различных условиях освещения.

В основе разработанной системы лежит глубокая сверточная нейронная сеть, которая обрабатывает последовательность видеокадров и преобразует визуальную информацию о движении губ в текстовую транскрипцию произносимой речи. Архитектура нейросети включает несколько ключевых компонентов: модуль предварительной обработки видео, который извлекает и нормализует область губ на каждом кадре; сверточные слои для извлечения пространственных признаков из изображений губ; рекуррентный модуль на основе LSTM для моделирования временных последовательностей; CTC-декодер для преобразования выходных данных нейросети в текст.

При разработке системы был использован следующий технологический стек: Python как основной язык разработки, TensorFlow и Keras для построения и обучения нейросетевых моделей, OpenCV для обработки видеоданных и детектирования области губ, FastAPI для создания веб-интерфейса и API, HTML/CSS/JavaScript для пользовательского интерфейса.

В рамках дипломного проекта было разработано полнофункциональное веб-приложение, которое позволяет загружать пользовательские видеофайлы для анализа, использовать тестовые видео для демонстрации работы системы, отображать результаты распознавания в удобном формате, визуализировать работу нейросети в процессе анализа. Интерфейс приложения разработан с учетом принципов UX/UI дизайна и обеспечивает интуитивно понятное взаимодействие с системой для пользователей с любым уровнем технической подготовки.

Основные преимущества системы заключаются в работе с видеофайлами различных форматов, возможности обработки как отдельных слов, так и связной речи, визуализации работы нейросетевой модели для лучшего понимания процесса анализа, наличии веб-интерфейса, не требующего установки дополнительного ПО.

Разработанная система чтения по губам имеет широкие перспективы применения в качестве вспомогательной технологии для людей с нарушениями слуха, для улучшения систем распознавания речи в шумных условиях, при работе с видеоматериалами без звуковой дорожки, в системах повышения безопасности коммуникаций и в образовательных системах для изучения артикуляции.

В качестве перспектив дальнейшего развития системы можно выделить повышение точности распознавания за счет использования мультимодальных моделей, объединяющих аудио- и видеоинформацию, расширение языковой поддержки, улучшение устойчивости к различным условиям съемки и артикуляционным особенностям разных дикторов, оптимизацию для работы в режиме реального времени на мобильных устройствах.

Разработанное в рамках дипломного проекта веб-приложение для чтения по губам с использованием нейросетей демонстрирует эффективность применения современных методов искусственного интеллекта для решения сложных задач компьютерного зрения. Система обладает практической ценностью и имеет потенциал для дальнейшего совершенствования и применения в различных сферах.

Литература

1. TensorFlow Documentation [Электронный ресурс]. – 2024. – Режим доступа: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Conv3D – Дата доступа: 18.03.2025.
2. LipNet: End-to-End Sentence-level Lipreading [Электронный ресурс]. – 2024. – Режим доступа: <https://arxiv.org/abs/1611.01599> – Дата доступа: 18.03.2025.
3. OpenCV Documentation [Электронный ресурс]. – 2024. – Режим доступа: https://docs.opencv.org/4.x/d9/df8/tutorial_root.html – Дата доступа: 18.03.2025.
4. FastAPI Documentation [Электронный ресурс]. – 2024. – Режим доступа: <https://fastapi.tiangolo.com/> – Дата доступа: 18.03.2025.
5. Visual Speech Recognition [Электронный ресурс]. – 2024. – Режим доступа: <https://paperswithcode.com/task/lipreading> – Дата доступа: 18.03.2025.
6. Towards Practical Lipreading with Distilled and Efficient Models [Электронный ресурс]. – 2024. – Режим доступа: <https://arxiv.org/abs/2007.06504> – Дата доступа: 18.03.2025.
7. Keras Documentation [Электронный ресурс]. – 2024. – Режим доступа: https://keras.io/api/layers/recurrent_layers/lstm/ – Дата доступа: 18.03.2025.
8. CTC Decoder [Электронный ресурс]. – 2024. – Режим доступа: <https://distill.pub/2017/ctc/> – Дата доступа: 18.03.2025.
9. Deep Learning for Lip Reading [Электронный ресурс]. – 2024. – Режим доступа: <https://github.com/rizkiarm/LipNet> – Дата доступа: 18.03.2025.
10. Visual Speech Recognition: A Review [Электронный ресурс]. – 2024. – Режим доступа:

УДК 004.942

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ РАСПОЗНАВАНИЯ ЛИЦА В ВИДЕОПОТОКЕ С ИСПОЛЬЗОВАНИЕМ ИЗОБРАЖЕНИЯ

Нестерков Д.А.

Научный руководитель –Белова С.В., ст. преподаватель

Технология распознавания лиц в видеопотоке — одно из самых динамично развивающихся направлений компьютерного зрения. Это технология искусственного интеллекта, позволяющая идентифицировать личность по цифровому изображению или видеокадру. Сегодня данная технология активно применяется в системах общественной безопасности, поиске пропавших людей, идентификации лиц в местах массового скопления, а также в правоохранительной деятельности.

Современные системы распознавания лиц функционируют по определённому алгоритму. Сначала происходит захват изображения — получение исходного кадра с камеры или видеопотока. Затем следует детекция лица, когда система обнаруживает область изображения, содержащую лицо человека. После этого выполняется предобработка изображения — нормализация, корректировка освещения, выравнивание. На следующем этапе происходит выделение признаков, то есть извлечение ключевых характеристик лица. Далее система проводит сопоставление с эталоном, сравнивая полученные признаки с базой данных. И наконец, формируется вывод о соответствии обнаруженного лица эталону.

Традиционные методы распознавания лиц включают несколько основных подходов. Метод фишеровских лиц (Fisherfaces) использует линейный дискриминантный анализ для лучшего разделения изображений лиц разных людей. Локальные бинарные шаблоны (LBP) описывают текстурные особенности лица путем сравнения яркости центрального пикселя с соседними. Гистограммы ориентированных градиентов (HOG) основаны на подсчете частоты направлений градиента в локальных областях изображения.

Методы глубокого обучения значительно превосходят традиционные подходы. DeepFace использует глубокую сверточную нейронную сеть с 3D-выравниванием лица для компенсации изменений ракурса. RetinaFace представляет собой многозадачный детектор лиц, объединяющий обнаружение лица, локализацию ключевых точек и реконструкцию 3D-формы.