

УДК 004.932.2:687.01

РАЗРАБОТКА МОДЕЛИ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ ОДЕВАНИЯ ОДЕЖДЫ НА ОСНОВЕ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ

Жәліш Жасұлан Убайдуллаұлы

магистрант, Образовательная программа «Прикладной анализ данных»,
Astana IT University, г.Астана, Казахстан

Научный руководитель: Тұрар О.Н.

Данная статья посвящена разработке системы на основе глубокого обучения, способной генерировать реалистичные изображения людей в процессе надевания одежды, используя методы семантической сегментации. Интеграция современных технологий компьютерного зрения, таких как U-Net, условные генеративно-сопоставительные сети (Conditional GANs) и модели оценки позы, позволяет решить проблему виртуальной примерки одежды в сфере онлайн-торговли. Предлагаемое решение обеспечивает фотореалистичную симуляцию, улучшает пользовательский опыт и открывает новые возможности для цифрового дизайна одежды. В статье описываются этапы подготовки данных, архитектура модели, методы аугментации, стратегия обучения, а также обсуждаются результаты и перспективы применения.

Ключевые слова: виртуальная примерка, семантическая сегментация, генерация одежды, conditional GAN, мода и ИИ, OpenPose, U-Net.

Введение

Модная индустрия переживает цифровую трансформацию: дополненная реальность, 3D-моделирование и ИИ-решения меняют способы взаимодействия потребителей с одеждой. Виртуальная примерка стала ключевым инструментом, сокращающим количество возвратов и повышающим уверенность покупателей.

Однако большинство существующих систем примерки отображают одежду в статичном виде, не отражая процесс надевания и естественного движения ткани. В данной работе предлагается использовать семантическую сегментацию в сочетании с генеративными моделями для создания динамичных изображений примерки.

Семантическая сегментация позволяет классифицировать каждый пиксель изображения (одежда, части тела, фон), а генеративные модели формируют

новые изображения на её основе. Это дает возможность имитировать взаимодействие одежды с телом в различных позах.

Обзор литературы

Технологии виртуальной примерки опираются на достижения в области сегментации изображений и генеративных моделей. Модель U-Net, предложенная Роннебергером и др., успешно применяется в медицине и моде благодаря своей архитектуре «кодировщик-декодировщик» и соединениям, сохраняющим пространственную информацию.

OpenPose — система для оценки позы человека — позволяет точно определить ключевые точки (суставы), что критически важно для корректного наложения одежды на тело.

Генеративные состязательные сети (GANs) радикально изменили подходы к синтезу изображений. Условные GANs (cGANs) дают возможность управлять результатом генерации, используя такие параметры, как сегментационные маски. В моде активно используется Pix2Pix — фреймворк для преобразования масок в реалистичные изображения.

Системы VITON и CP-VTON предложили двухэтапную архитектуру: деформация одежды под тело пользователя и её визуальное улучшение. Новые подходы, например DiOr, позволяют моделировать последовательность надевания одежды. FashionGAN и SwapNet расширяют возможности редактирования, позволяя менять цвет, текстуру и стиль одежды.

Сбор и разметка данных

Для обучения модели использовались как открытые датасеты (DeepFashion2, VITON, COCO, MPII), так и собранные вручную изображения с модных сайтов и соцсетей. Последние обеспечили культурное и стилистическое разнообразие.

Аннотация проводилась вручную с использованием LabelMe и Supervisely. Сегментировались верхняя, нижняя и полная одежда, а также аксессуары. Позы людей были размечены для последующей синхронизации одежды и тела.

Предобработка и аугментация

Для повышения устойчивости модели применялись методы аугментации:

Повороты, отражения, масштабирование

Изменение освещенности и цвета

Случайное обрезание и шум

Оригинальный набор из 15 000 изображений был уменьшен до 5 000 для обучения и расширен обратно до 15 000 с помощью аугментации. Это позволило избежать переобучения и повысить универсальность модели.

Архитектура модели

Рассматривались три архитектуры:

U-Net: Использовался для точной сегментации одежды.

Conditional GAN (cGAN): Генерирует реалистичные изображения на основе масок и поз. Генератор формирует изображение, дискриминатор оценивает его правдоподобность.

Вариационные автокодировщики (VAE): Кодировать изображения в латентное пространство и обратно. Обеспечивают разнообразие, но менее фотореалистичны.

Модель состоит из:

U-Net для сегментации

OpenPose для поз

cGAN с механизмами внимания для генерации

Использовались функции потерь: перцептуальная, L1 и состязательная.

Обучение проводилось в среде PyTorch с использованием CUDA.

Стратегия обучения и оценки

Разделение данных: 70% — обучение, 15% — валидация, 15% — тест.

Метрики:

mIoU: точность сегментации

FID: сходство с реальными изображениями

SSIM: визуальное сходство

Для предотвращения переобучения использовались ранняя остановка и сохранение лучших моделей.

Результаты и анализ

Лучшие показатели модели:

75% точности наложения одежды

SSIM > 0.88

Улучшенные FID при обучении на аугментированных данных

Визуальный анализ показал реалистичные складки, драпировку и соответствие телу. Основные проблемы: артефакты при сложных текстурах и редких позах.

Пользовательское тестирование (20 участников) показало среднюю оценку реалистичности — 8.1 из 10. Предпочтение отдавалось изображениям, сгенерированным cGAN.

Обсуждение и заключение

Предложенная модель сочетает сегментацию и генерацию, обеспечивая реалистичную симуляцию процесса примерки. Гибкая архитектура позволяет включать новые компоненты — например, трансформеры или стилизацию одежды.

Возможные применения:

Виртуальные примерочные

Цифровое проектирование одежды

Игры и анимация

Визуализация в интернет-магазинах

Исследование доказало эффективность генерации изображений примерки одежды с использованием семантической сегментации и генеративных моделей. В будущем планируется:

Внедрение трансформеров

Улучшение генерации сложных текстур

Оптимизация работы в реальном времени

Персонализация под параметры пользователя

Работа закладывает основу для нового уровня взаимодействия пользователей с цифровой модой.

Список использованной литературы

1. Lassner C., Pons-Moll G., Gehler P. V. Генеративная модель людей в одежде (A generative model of people in clothing) // Труды Международной конференции IEEE по компьютерному зрению (ICCV). – 2017. – С. 853–865. – DOI: <https://doi.org/10.1109/iccv.2017.98>

2. Lateef F., Ruichek Y. Обзор методов семантической сегментации с использованием технологий глубокого обучения (Survey on semantic segmentation using Deep Learning Techniques) // Neurocomputing. – 2019. – Т. 338. – С. 321–348. – DOI: <https://doi.org/10.1016/j.neucom.2019.02.003>

3. Sultana F., Sufian A., Dutta P. Эволюция сегментации изображений с использованием глубоких сверточных нейронных сетей: обзор (Evolution of image segmentation using deep convolutional neural network: A survey) // Knowledge-Based Systems. – 2020. – Т. 201–202. – Ст. 106062. – DOI: <https://doi.org/10.1016/j.knosys.2020.106062>

4. Ronneberger O., Fischer P., Brox T. U-Net: сверточные сети для сегментации биомедицинских изображений (U-Net: Convolutional Networks for Biomedical Image Segmentation) // Медицинская обработка изображений и компьютерная интервенция (MICCAI). – 2015. – arXiv:1505.04597. – DOI: <https://doi.org/10.48550/arXiv.1505.04597>

5. Long J., Shelhamer E., Darrell T. Полносвязные сверточные сети для семантической сегментации (Fully Convolutional Networks for Semantic Segmentation) // Труды конференции IEEE по компьютерному зрению и распознаванию образов (CVPR). – 2015. – DOI: <https://doi.org/10.1109/CVPR.2015.7298965>

6. Cao Z., Simon T., Wei S. E., Sheikh Y. Оценка позы нескольких людей в реальном времени с использованием полей аффинности частей (Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields) // Труды конференции IEEE по компьютерному зрению и распознаванию образов (CVPR). – 2017. – DOI: <https://doi.org/10.1109/CVPR.2017.143>

СЕМАНТИКАЛЫҚ СЕГМЕНТАЦИЯ НЕГІЗІНДЕ КИІМ КИЮ БЕЙНЕЛЕРІН ҚАЛЫПТАСТЫРУ ҮЛГІСІН ЖАСАУ

Жәліш Жасұлан Убайдуллаұлы

Ғылыми жетекші: Тұрар О.Н.

Бұл мақала киімді кию процесінде адамдардың шынайы бейнелерін генерациялай алатын терең оқытуға негізделген жүйені әзірлеуге арналған. Семантикалық сегментация әдістерін қолдану арқылы жүзеге асырылатын бұл жүйе U-Net, шартты генеративті қарсылас желілер (Conditional GANs) және позаны бағалау модельдері сияқты заманауи компьютерлік көру технологияларын біріктіреді. Ұсынылған шешім онлайн сауда саласындағы виртуалды киім өлшеу мәселесін шешуге мүмкіндік береді. Бұл жүйе фотореалистік симуляцияны қамтамасыз етіп, пайдаланушы тәжірибесін жақсартады және цифрлық киім дизайны үшін жаңа мүмкіндіктер ашады. Мақалада деректерді дайындау кезеңдері, модель архитектурасы, аугментация әдістері, оқыту стратегиясы сипатталады және нәтижелер мен қолдану перспективалары талқыланады.

Кілт сөздері: виртуалды киім өлшеу, семантикалық сегментация, киім генерациясы, Conditional GAN, мода және ЖИ (Жасанды интеллект), OpenPose, U-Net.

DEVELOPMENT OF A MODEL FOR GENERATING IMAGES OF PUTTING ON CLOTHES BASED ON SEMANTIC SEGMENTATION

Zhalish Zhassulan

Scientific Supervisor: Turar O.N.

This article is dedicated to the development of a deep learning-based system capable of generating realistic images of people in the process of putting on clothes, using semantic segmentation methods. The integration of modern computer vision technologies, such as U-Net, conditional generative adversarial networks (Conditional GANs), and pose estimation models, makes it possible to address the problem of virtual clothing try-on in the field of online retail. The proposed solution provides photorealistic simulation, enhances user experience, and opens up new opportunities for digital clothing design. The article describes data preparation stages, model architecture, augmentation methods, training strategy, and discusses the results and potential applications.

Keywords: Virtual try-on, semantic segmentation, clothing generation, conditional GAN, fashion and AI, OpenPose, U-Net.

REFERENCES

1. Lassner C., Pons-Moll G., Gehler P. V. A Generative Model of People in Clothing // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – 2017. – pp. 853–865. – DOI: <https://doi.org/10.1109/iccv.2017.98>
2. Lateef F., Ruichek Y. Survey on Semantic Segmentation Using Deep Learning Techniques // Neurocomputing. – 2019. – Vol. 338. – pp. 321–348. – DOI: <https://doi.org/10.1016/j.neucom.2019.02.003>
3. Sultana F., Sufian A., Dutta P. Evolution of Image Segmentation Using Deep Convolutional Neural Network: A Survey // Knowledge-Based Systems. – 2020. – Vols. 201–202. – Article 106062. – DOI: <https://doi.org/10.1016/j.knosys.2020.106062>
4. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). – 2015. – arXiv:1505.04597. – DOI: <https://doi.org/10.48550/arXiv.1505.04597>
5. Long J., Shelhamer E., Darrell T. Fully Convolutional Networks for Semantic Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – DOI: <https://doi.org/10.1109/CVPR.2015.7298965>
6. Cao Z., Simon T., Wei S. E., Sheikh Y. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – DOI: <https://doi.org/10.1109/CVPR.2017.143>