

SMART CROPPING: ТЕХНОЛОГИЯ УМНОЙ ОБРЕЗКИ ИЗОБРАЖЕНИЙ

Сорокина В. В.

*Белорусский государственный университет, Минск, Беларусь,
e-mail: Viktoria.sorokina.96@gmail.com*

В данной работе рассматривается задача генерации изображений, представляющих одежду и обувь на модели, в каталог товаров электронной коммерции. Генерация изображений подразумевает нарезку входного изображения человека в полный рост на серию снимков, представляющих определенные товары. Нарезка изображений производится за счет решения задачи определения частей тела человека – задачи компьютерного зрения, которая вычисляет ключевые точки человеческого тела или объекта на изображении или видео. Также можно рассматривать задачу определения частей тела человека как проблему определения положения и ориентации камеры относительно данного человека или объекта.

Smart Cropping – полностью автоматическая система, позволяющая подготавливать изображения, формирующие каталог товаров, для классов плечевой, поясной и верхней одежды, а также обуви и головных уборов.

Архитектура системы представлена блоком сверточной нейронной сети на основе архитектуры OpenPose с VGG-19 для выделения признаков, а также алгоритмами для вычисления позиционного отношения между ключевыми точками человеческого тела, которые используются для обрезки по заданным правилам.

OpenPose – это глубокая нейронная сеть с прямой связью. Данный метод позволяет эффективно обнаруживать 2D-позиции частей тела человека на изображении RGB в реальном времени. Это делается путем обнаружения и ассоциации частей тела с использованием полей сходства частей (PAFs) и карт достоверности. Карта достоверности представляет собой функцию плотности вероятности для нового изображения, присваивающую каждому пикселю нового изображения вероятность, которая представляет собой вероятность принадлежности пикселя части тела в объекте на предыдущем изображении. Обнаружение частей тела происходит в последовательном стиле, выполняя прогнозирование снизу-вверх с использованием пространственного контекста. Извлеченная информация впоследствии используется для создания начальной структуры человеческого скелета в текущем кадре. Архитектура OpenPose представлена на рисунке 1.

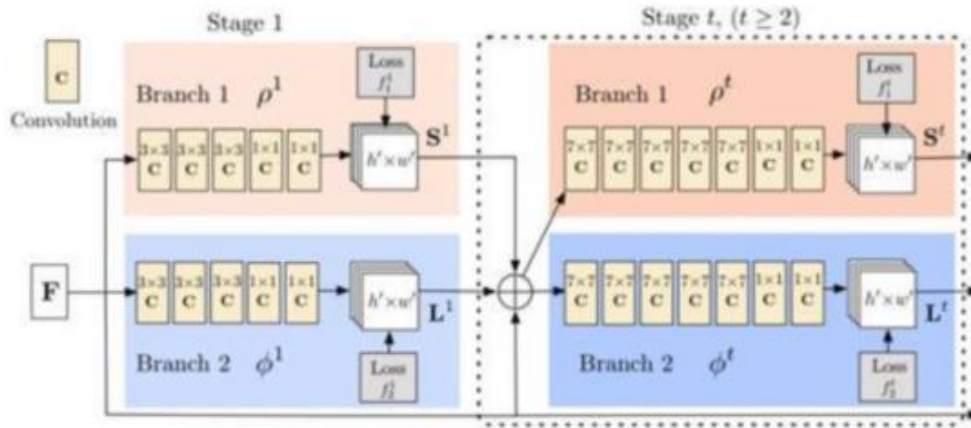


Рис. 1. Архитектура OpenPose

Smart Cropping – это метод, позволяющий производить обрезку автоматически без привлечения человеческих ресурсов. Чтобы реализовать такую обрезку исходного снимка и создание набора изображений, представляющих товары, нужно получить координаты ключевых точек человеческого тела, а затем вычислить позиционные отношения между ними через координаты каждой ключевой точки. Необходимо знать координаты трех точек, чтобы узнать угол, образованный тремя сторонами, а затем использовать диапазон значений этих углов, чтобы оценить позу человека на изображении и произвести корректную обрезку. Формулы для расчета этих углов показаны ниже.

Пусть известны 3 точки $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$ и соответствующие векторы:

$$\begin{aligned}\overrightarrow{AB} &: (x_2 - x_1, y_2 - y_1), \\ \overrightarrow{AC} &: (x_3 - x_1, y_3 - y_1), \\ \overrightarrow{BC} &: (x_3 - x_2, y_3 - y_2).\end{aligned}\tag{1}$$

Тогда

$$\begin{aligned}|AB| &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \\ |AC| &= \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}, \\ \cos \angle A &= \frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)}{|AB||AC|}.\end{aligned}\tag{2}$$

В ходе выполнения работы на наборе данных СОСО была обучена нейронная сеть на основе архитектуры OpenPose, способная распознавать 23 ключевые точки человеческого тела.

Полученные ключевые точки формируют прямые, составляющие позу человека, например, левое и правое плечи, левое и правое бедра и т.д.

Взаимное расположение полученных прямых является инструкцией к обрезке изображения. При этом все прямые классифицируются следующим образом – ключевые точки, составляющие прямые, относятся к классу «верх», если они расположены выше запястий, и к классу «низ», если ниже. Данная классификация необходима, т.к. в случае не нахождения одной или всех ключевых точек в паре

система автоматически спускается вверх/вниз по иерархии прямых в зависимости от типа обрезки.

Сведения о позиционных отношениях между данными прямыми позволили создать систему для автоматической обрезки изображений и сократить время подготовки изображений для электронного каталога товаров. Данная система была протестирована на реальных изображениях товаров одного интернет-магазина. Пример работы системы представлен на рисунке 2.



Рис. 2. Пример работы Smart Cropping системы

Результаты, полученные в разработанной программе, могут быть использованы для обрезки изображений других категорий товаров.

Литература

1. Chen, Y. Monocular human pose estimation: A survey of deep learning-based methods / Y. Chen, Y. Tian, and M. He // Computer Vision and Image Understanding – 2020. – Vol. 192
2. Liu, Z. A survey of human pose estimation: the body parts parsing based methods / Z. Liu [et al.] // Journal of Visual Communication and Image Representation – 2015. – Vol. 32, P. 10–19.
3. 2d/3d pose estimation and action recognition using multitask deep learning / Luvizon, D.C., Picard, D., Tabia, H. // IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, June 18-23 2018. – Salt Lake City, 2017. – P.5137–5146.
4. Bi-Manual Articulated Robot Teleoperation using an External RGB-D Range Sensor / Emily-Jane Rolley-Parnell [et al.] // International Conference on Control, Automation, Robotics and Vision, Singapore, September 2018. – Singapore, 2018.
5. eCommerce Product Image Guide [Electronic resource] – Mode of access: <https://www.threekit.com/blog/ecommerce-product-image-guide-2020>. – Date of access: 25.03.2021.
6. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields / Zhe Cao [et al.] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 21-26 July 2017. – Honolulu, 2017. – P. 1302-1310.
7. Very Deep Convolutional Networks for Large-Scale Image Recognition / Karen Simonyan and Andrew Zisserman // International Conference on Learning Representations, San Diego, May 7-9, 2015. – San Diego, 2015. – P. 1137–1149
8. He, K. Deep Residual Learning for Image Recognition / K. He [et al.] // 2016 IEEE Conference on Computer Vision and Pattern Recognition – Las Vegas, 2016. – P. 770-778
9. Gesture Recognition Based on Deep Learning / Tingxuan Zhang // Journal of Physics Conference Series – 2020. – P. 1449-1457
10. COCO dataset // COCO 2018 Keypoint Detection Task [Electronic resource] – Mode of access: <http://cocodataset.org/#overview> – Date of access: 05.04.2019.