

АВТОМАТИЧЕСКОЕ ВЫДЕЛЕНИЕ ГРАНИЦ ЧЕКА НА ИЗОБРАЖЕНИИ

Янкова А. Д., Пастухов К. Д.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Петюкевич Н.С. – ст. преподаватель, м.т.н.

В данном докладе будет рассмотрена задача автоматического выделения границ чека на изображении и её решение на основе дообучения существующей нейронной сети. Будет описан следующий перечень вопросов: поиск объектов на изображении, структура выбранной модели нейронной сети и её дообучение. Полученная реализация выделения границ чека может послужить составной частью любой системы автоматической обработки и анализа финансовых данных.

Выделение границ чека на изображении является первым шагом в процессе автоматического извлечения данных из него.

Существует два способа практической реализации данной задачи. Первый основан на использовании различных алгоритмов обработки изображений, второй – на обучении или дообучении нейронной сети.

Первый способ, хоть и намного легче в реализации, сильно зависит от начальных условий и не способен достичь приемлемого результата во всех возможных случаях. Обрабатываемые изображения могут содержать посторонние объекты или фоновый шум, иметь неоднородное освещение и качество, сами чеки могут быть в разной степени деформированы. Также для достижения сколько-нибудь адекватного распознавания границ необходимо комбинировать существующие алгоритмы, что вычислительно недопустимо, либо прибегать к дополнительной предварительной обработке изображений, не связанной с непосредственным выделением границ.

Таким образом, не существует единого алгоритма, способного одновременно учесть все перечисленные нюансы.

Все перечисленные проблемы актуальны и для второго способа, однако, в отличие от любых алгоритмов, нейронные сети обладают гораздо большей гибкостью за счёт своей способности к обучению и выявлению сложных признаков и свойств объектов, сущностей и понятий. По этой причине они могут показывать качественные результаты в общем случае, без привязки к конкретным параметрам обрабатываемых изображений.

В зависимости от своей структуры нейронные сети могут обнаруживать сущности в двух формах [1]. В первой форме объект обрамляется прямоугольной рамкой, что иногда не совсем точно показывает действительное положение объекта на фотографии. Во второй форме на весь объект накладывается попиксельная маска. Преимуществом маски является то, что на её основе возможно точно вырезать чек из изображения с учётом перспективы, в итоге получая намного более удобные входные данные для дальнейшей обработки, нежели при обрамлении прямоугольной рамкой.

Отличные результаты в области обработки изображений в наше время показывают нейронные сети на основе трансформенной архитектуры несмотря на то, что первоначально они были разработаны для других задач. Суть трансформенной архитектуры в общем виде заключается в разделении обрабатываемого объекта на составные части с последующим расчётом взаимосвязей между этими частями. В частности, данной архитектуре следует выбранная модель Swin Transformer [2], которая также включает в себя обработку изображения на различных масштабах и ряд улучшений классической структуры. Общая схема структуры модели приведена на рисунке 1.

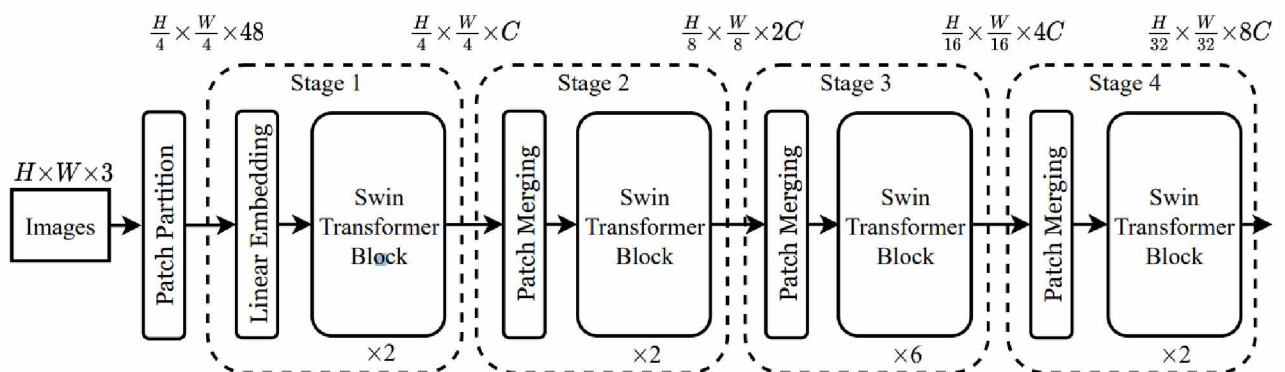


Рисунок 1 – Схематическое изображение структуры модели архитектуры Swin Transformer

Как видно из рисунка 1, после разбиения изображения на части в блоке Patch Partition процесс обработки состоит из четырёх стадий. Каждая стадия работает с изображением на своём уровне

масштаба, пропуская его через последовательность нескольких Swin Transformer Block, которые рассчитывают взаимосвязи между составными частями изображения. Между стадиями общая размерность фотографии уменьшается в два раза как по высоте, так и по ширине. В самом конце, что не отображено на общей схеме, данные попадают в классификатор, который превращает результирующие численные данные в осмысленные параметры, задающие границы маски или обводящего прямоугольника, класс объекта или объектов на изображении и иную информацию.

Создание собственной нейронной сети в соответствии с этой архитектурой оказалось бы слишком неэффективной и трудоёмкой задачей, поскольку её обучение с нуля потребовало бы огромных вычислительных и временных ресурсов. В связи с этим лучшим решением оказывается применение метода дообучения [3].

Дообучение нейронной сети – процесс адаптации предварительно обученной модели к новой задаче, связанной с предыдущей, не начиная обучение с нуля. Данный подход основан на принципе переноса обучения, где знания, представления и понятия, полученные моделью в процессе тренировки на большом наборе данных, используются для решения более специфической проблемы. В процессе дообучения модель может быть модифицирована путём изменения архитектуры, например, замены последних слоёв для соответствия новой задаче, а также могут быть применены различные стратегии, такие как замораживание слоёв модели для контроля того, веса каких составных частей нейронной сети останутся в неизменном положении, а какие будут адаптированы к новым данным.

Следовательно, дообучение позволяет эффективно использовать уже существующие знания, полученные в процессе обучения на обширных наборах данных, что существенно сокращает время и ресурсы, необходимые для решения новых задач.

Для проведения дообучения под задачу распознавания границ чеков не существовало готового набора данных, поэтому он был создан вручную путём объединения и разметки более полутора тысяч изображений чеков, найденных как в сети Интернет, так и сфотографированных самостоятельно. Метка представляла собой полигон, описывающий контур истинной маски чека на изображении. Также у исходной модели был изменён классификатор и заморожена ветвь, отвечающая за формирование прямоугольной границы объекта.

Таким образом, выбранная модель, способная распознать восемьдесят типов объектов на изображении, была натренирована для задачи определения единственного класса – чеков. На рисунке 2 проиллюстрирован пример предсказания модели. В данном примере чек выделен не полной попиксельной маской, а обведён только лишь её контуром.



Рисунок 2 – Контур чека, предсказанный дообученной моделью

Несмотря на тот факт, что при тестировании нейронной сети было обнаружено, что она определяет границы чеков с незначительными неточностями, практически во всех случаях полученная маска полностью покрывала содержащийся на нём текст. В следствие этого поставленную задачу можно считать выполненной.

Список использованных источников:

1. CS231n: Deep Learning for Computer Vision [Электронный ресурс]. – Режим доступа: <http://cs231n.stanford.edu/> – Дата доступа: 15.03.2024.
2. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2103.14030.pdf> – Дата доступа: 15.03.2024.
3. Fine-tune a pretrained model [Электронный ресурс]. – Режим доступа: <https://huggingface.co/docs/transformers/training> – Дата доступа: 15.03.2024.