

Отчёт по курсу: «Современное компьютерное зрение»

Выполнила: Зайченкова Екатерина, М05-302а.

Общее задание: собрать и разметить изображения баркодов.

Результат выполнения: собраны и размечены 87 изображений, доступны по ссылке:
<https://github.com/Zayrina/mipt2024s-5-zaychenkova-e-e/tree/main/data>

Индивидуальное задание: реализовать алгоритм для расправления баркодов.

Содержание

Анализ задачи.....	1
Обзор литературы.....	2
Проведённые эксперименты.....	5
Планы по улучшению работы.....	11

Анализ задачи

Постановка задачи: по изображению с искаженным (смятым, изогнутым) баркодом восстанавливать изображение с распрямленным.

Возможные типы искажений:

- смятие (без перекрытия и с перекрытием баркода другими объектами);
- загибы части баркода, при которых эта часть остается видимой;
- отсутствие части баркода (она может быть оторвана, закрашена, закрыта другими объектами, загнута под остальную часть баркода);
- цилиндрические искажения, вызванные тем, что баркод расположен на цилиндрическом объекте (например, бутылке) или на скрученном листе бумаги и подобных материалов;

Ситуации, когда баркод частично не виден, принципиально делятся на случаи, когда отсутствующую часть можно восстановить по видимой без потери смысла (например, если частично не видны полосы в одномерном баркоде, но для каждой полосы есть сохранившийся участок, который можно просто продолжить) и когда для восстановления баркода имеющегося контекста недостаточно.

Входные данные для алгоритма расправления: изображение вырезанного по bounding box-у баркода, получаемое от детектора баркодов (алгоритм Корчагина Сергея).

Выходные данные: расправляемое изображение баркода.

Обзор литературы

Ключевые слова для поиска существующих решений: barcode, document / paper unwarping / dewarping / dewrapping / unrolling / unfolding / flattening / unfurling.

Задачу расправления баркодов было решено обобщить до расправления документов. Обзор показал, что существующие для документов методы глобально делятся на нейросетевые (SotA) и “классические” (более старые, без использования нейронных сетей).

Классические подходы

[Stamatopoulos](#) и др. [1] предложили один из первых подходов, основанный на нахождении слов и строк текста, по результатам которого осуществлялось грубое улучшение изображения, а после этого отдельно улучшалось отображение отдельных слов.

[Tian](#) и др. [2] тоже детектировали строки текста, восстанавливали по ним двумерную сетку искажения, а затем по ней оценивали трехмерную деформацию.

[Wada](#) и др. [3] и [Courteille](#) и др. [4] восстанавливали трехмерную форму искаженного документа на основе анализа теней (техника SfS), а затем проецировали трехмерную поверхность на плоскость.

Аналогично [Zhang](#) и др. [5] предложили общий метод SfS, учитывающий модель перспективной проекции и различные условия освещения.

[Brown](#) и др. [6] опирались на двумерные границы изображенного материала (бумаги и т.п.).

[Tsoi](#) и др. [7] для восстановления геометрических искажений извлекали информацию о границах на основе снимков документа с разных ракурсов.

Ограничениями применимости подобных подходов является то, что они значительно опираются на предположения о свойствах документа: что он имеет цилиндрическую форму, или содержит текст, или имеет упорядоченное содержание (разбиение текста на строки, параллельные края листа границы изображений). Кроме того, такие методы в большинстве своем направлены на исправление искажений определенного типа (например, загибов), поэтому для полноценного выпрямления документа придется использовать комбинацию таких методов.

Нейросетевые подходы

Первый нейросетевой алгоритм был предложен в 2018 году [Ma](#) и др. [8]. Они использовали U-Net для предсказания выпрямляющего преобразования по деформированному изображению и для его обучения и тестирования сгенерировали синтетический набор данных из приблизительно 100.000 изображений.

[Ramanna](#) и др. [9] в 2019 году предложили использовать нейронную сеть, переводящую изображение в изображение (pix2pixhd) и основанную на Conditional

Generative Adversarial Network (CGAN). Обучение производилось на синтетических данных, тестирование — и на синтетических, и на реальных. Для проверки качества работы алгоритма использовались метрики, основанные на пиксельном сравнении исходного и восстановленного изображений: Pixel Accuracy, Mean Accuracy, Frequency Weighted IU, а также индекс структурного сходства SSIM и Haar Perceptual Similarity Index (HaarPSI), который отличается высокой коррелированностью с MOS (восприятием человека).

Liu и др. [10] использовали адверсиальную нейросеть для предсказания плотной сетки деформации изображения.

В 2020 году Xie и др. [11] обучили полносвязную нейросеть на синтезированных данных, где изображения смятого документа сопровождались разметкой “потоков смещений” (displacement flow). Эта двухголовая сеть предсказывает для каждого пикселя изображения смещение (регрессия) и является ли этот пиксель фоном (бинарная классификация). Авторы также предложили метод регуляризации для контроля гладкости потока смещений, то есть того, что выровненный документ имеет непрерывную форму, и того, что выравнивание сохраняет локальные детали. Метод проверялся на реальных и симулированных изображениях. В качестве метрики для оценки алгоритма использовался Multi-Scale Structural Similarity (MS-SSIM) — индекс структурного сходства на разных масштабах изображений, а также Local Distortion (LD), который оценивает локальные признаки путем подсчета dense SIFT flow — плотного потока на основе нахождения особых точек.

В 2021 году Xie и др. предложили усовершенствование своего метода [12]. Новый подход был основан на нахождении контрольных точек на изображении с помощью нейросетевого энкодера. Архитектура энкодера строилась на основе полносвязной сети, использованной в работе [11]. Энкодер использовался для нахождения признаков, по которым определялись контрольные и референсные точки. Затем с помощью интерполяционного подхода (TPS, Linear или Cubic) находилось попиксельное преобразование, которое переводило контрольные точки в референсные. Как и в предыдущей статье, для оценки качества алгоритма использовались MS-SSIM и Local Distortion (LD).

Источники

- [1] Stamatopoulos, N., Gatos, B., Pratikakis, I., Perantonis, S.J.: Goal-oriented rectification of camera-based document images. *IEEE Transactions on Image Processing* 20(4), 910–920 (2010)
- [2] Tian, Y., Narasimhan, S.G.: Rectification and 3d reconstruction of curved document images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 377–384. IEEE (2011)
- [3] Wada, T., Ukida, H., Matsuyama, T.: Shape from shading with interreflections under a proximal light source: Distortion-free copying of an unfolded book. *International Journal of Computer Vision* 24(2), 125–135 (1997)

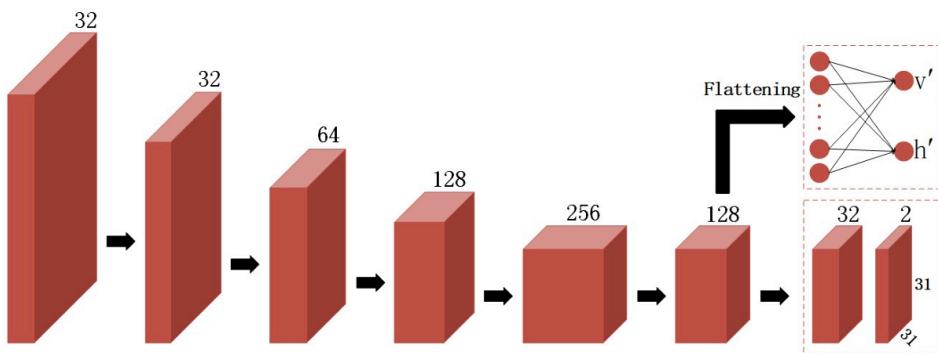
- [4] Courteille, F., Crouzil, A., Durou, J.D., Gurdjos, P.: Shape from shading for the digitization of curved documents. *Machine Vision and Applications* 18(5), 301–316 (2007)
- [5] Zhang, L., Yip, A.M., Brown, M.S., Tan, C.L.: A unified framework for document restoration using inpainting and shape-from-shading. *Pattern Recognition* 42(11), 2961–2978 (2009)
- [6] Brown, M.S., Tsoi, Y.C.: Geometric and shading correction for images of printed materials using boundary. *IEEE Transactions on Image Processing* 15(6), 1544–1554 (2006)
- [7] Tsoi, Y.C., Brown, M.S.: Multi-view document rectification using boundary. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1–8. IEEE (2007)
- [8] Ma, K., Shu, Z., Bai, X., Wang, J., Samaras, D.: Docunet: document image unwarping via a stacked u-net. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4700–4709 (2018)
- [9] Ramanna, V., Bukhari, S.S., Dengel, A.: Document image dewarping using deep learning. In: *International Conference on Pattern Recognition Applications and Methods* (2019)
- [10] Liu, X., Meng, G., Fan, B., Xiang, S., Pan, C.: Geometric rectification of document images using adversarial gated unwarping network. *Pattern Recognition* 108, 107576 (2020)
- [11] Xie G. W. et al: Dewarping document image by displacement flow estimation with fully convolutional network //*Document Analysis Systems: 14th IAPR International Workshop, DAS 2020, Wuhan, China, July 26–29, 2020, Proceedings* 14. – Springer International Publishing (2020)
- [12] Xie G. W. et al: Document dewarping with control points //*Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I* 16. – Springer International Publishing (2021).

Проведённые эксперименты

Ссылка на гитхаб: <https://github.com/Zayrina/mipt2024s-5-zaychenkova-e-e/tree/main>

1. Начала с алгоритма [12] как с самого нового из найденных.

Метод основан на обучении нейросетевого энкодера вот такой архитектуры:



Верхняя голова нейросети считает расстояние между референсными точками, нижня — ищет контрольные точки.

Для начала попробовала с помощью опубликованной авторами нейронной сети, обученной на 30.000 симулированных изображениях деформированных документов, выровнять собранные реальные баркоды.

Результат ожидаемо оказался слишком плохой, поэтому проверять на симулированных не стала. Проблема возникла с тем, что маска контрольных точек захватывала много фона, объединяла несколько баркодов в один и в результате распознавала неправильную форму.

Примеры работы:



2. Чтобы решить проблему с фоном и объединением баркодов, вырезала каждый баркод с помощью разметки. В дальнейшем решила работать именно с вырезанными отдельными баркодами, потому что распрямитель баркодов логично в общей системе встроить после детектора.

Оказалось, что теперь нейросеть наоборот почти на всех изображениях обрезает значительную часть, что тоже критично плохо.

Примеры работы (сверху маска контрольных точек, снизу выровненное изображение):



3. Следующий шаг — дообучение рассматриваемой нейросети на симулированных данных. Для этого с помощью симулятора Плохотнюка Всеволода и Белкова Алексея сгенерировала аугментированные баркоды, а с помощью кода из статьи [12] — соответствующие маски контрольных точек.

При симуляции подобрала для баркодов более-менее реалистичные на мой взгляд фоны:

картон	крафтовая бумага	полиэтиленовый пакет	плёнка
бумага	упаковка / обои	ткань типа 1	ткань типа 2

ткань типа 3	дерево	абстрактный фон	
			

Примеры симулированных изображений:



Размер полученного набора данных составил 200 изображений.

4. На симулированных изображениях дообучила нейросеть со 143 эпох до 500 эпох, размер батча взяла 4, оптимизатор Adam, learning rate 5e-5 (каждые 40 эпох уменьшается в 2 раза). Данные поделила на трейн/валидацию/тест в отношении 8:1:1 с соблюдением баланса классов баркодов и фонов.

В качестве метрик для оценки качества выбрала Structural Similarity Index(SSIM) — классическую метрику для сравнения структуры, Haar Perceptual Similarity Index (HaarPSI) — использовалась в статьях по выпрямлению документов, считается хорошо отражающей человеческой восприятие, а также FSIM — одна из SotA метрик структурного сходства изображений.

Для подсчёта этих метрик сравниваемые изображения должны быть одного размера, поэтому истинное и тестируемое изображение перед подсчетом метрик приводила к минимальному из их размеров по каждой оси. Для этого использовала метод resize из opencv, попробовала разные интерполяции: кубическую (cv2.INTER_CUBIC), линейную (cv2.INTER_LINEAR), а также cv2.INTER_AREA.

На симулированных данных получились такие результаты:

интерполяция	SSIM, %	HaarPSI, %	FSIM, %
cv2.INTER_CUBIC	41.4 ± 27.7	63.0 ± 30.5	48.7 ± 12.1
cv2.INTER_LINEAR	42.1 ± 27.4	63.2 ± 30.4	49.1 ± 12.2
cv2.INTER_AREA	39.8 ± 27.0	62.8 ± 30.7	47.8 ± 12.1

Значения близкие, от способа интерполяция зависят не сильно. Лучше всего получилось с линейной.

Примеры изображений (сверху с найденными контрольными точками, под ним — распрямленное, в самом низу — оригинал):





Невысокие значения метрик, скорее всего, обусловлены:

- относительными сдвигами исходного и найденного баркодов, вызванными тем, что маска находит границы исходного баркода с небольшой погрешностью;
 - искажением формы баркода в результате растяжения или сжатия при смятии и закруглении;
 - изменениями цвета, которые происходили при аугментациях смятия и сгиба.
- Визуально с точки зрения структуры все полученные изображения похожи на исходные.

На реальных данных:

Качество визуально всё ещё заметно хуже, чем на симулированных данных.

На слабо искаженных изображениях нейросеть в большинстве случаев адекватно работает:



Полученный результат уже лучше, чем в экспериментах без дообучения: визуально больше IoU для маски контрольных и истинного баркода.

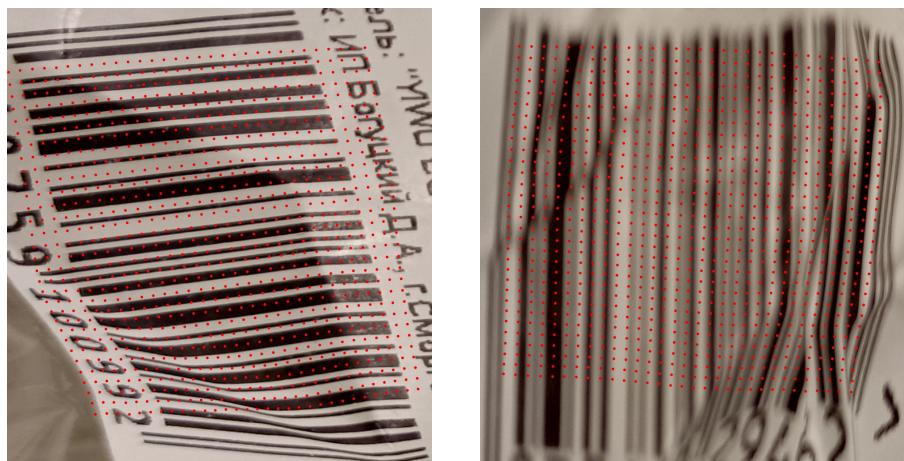
Анализ ошибок

Очень распространена ситуация, когда нейросеть предлагает квадратную маску контрольных точек вместо “прямоугольной” для одномерного баркода (22 случая), причем часто эта маска имеет неправильный наклон (12 случаев из этих 22).



В трех случаях для двумерных баркодов распознался значительно неверный (больше 10 градусов) наклон маски (сама маска при этом квадратная).

Также пока не распознается, что баркод сильно помят и находится под заметным углом к наблюдателю (то есть плотность маски слишком низкая там, где должна быть высокой).



Сейчас у всех масок плотность примерно равномерная (хотя в целом алгоритм допускает, что маски могут быть сильно неравномерными — мне попадались такие ситуации, правда, при неудачных гиперпараметрах обучения):



— для неудачных гиперпараметров (другая сетка)

В целом на двумерных баркодах нейросеть отработала лучше, чем на одномерных. Возможно, это связано с тем, что для одномерные она путает с текстом. Также похоже, что не хватило аугментаций, при которых баркод доходил бы верхнего и нижнего краев изображения — возможно, из-за этого для одномерных баркодов находятся квадратные (обрезанные) маски.

При этом неожиданно есть сильно мятые или поврежденные баркоды, где нейросеть работает относительно хорошо:



Планы по улучшению работы

Пока не успела реализовать, но осознаю, что стоит:

1. Оценить, насколько полученный алгоритм улучшает распознавание баркодов и качество системы в целом.
2. Для сравнения запустить другие нейросетевые методы.
3. Реализовать аугментации, моделирующие более сложное смятие и скручивание баркодов.
4. Добавить еще метрик для оценки качества результата, например, Local Distortion (LD), попиксельное сравнение: Pixel Accuracy, Mean Accuracy, Frequency Weighted IU.

5. Приближать симулированный датасет к реальному, чтобы повысить качество на реальных данных.