



Активное обучение и перенос знаний в задаче сегментации изображений документов

^{1,2} Д.М. Киранов, ORCID:0000-0002-3507-3803 <kiranov.dm@ispras.ru>

¹ М.А. Рындин, ORCID:0000-0002-7504-3975 <mxrynd@ispras.ru>

¹ И.С. Козлов, ORCID:0000-0002-0145-1159 <kozlov-ilya@ispras.ru>

¹ Институт системного программирования им. В.П. Иванникова РАН,
109004, Россия, г. Москва, ул. А. Солженицына, д. 25

² Московский физико-технический институт,
141701, Россия, Московская область, г. Долгопрудный, Институтский пер., 9

Аннотация. В данной работе исследуется эффективность классических подходов активного обучения в задаче сегментации изображений документов с целью уменьшения обучающей выборки. Приводится свой модифицированный подход выбора изображений для разметки и последующего обучения. Результаты, полученные с помощью активного обучения, сравниваются с переносом знаний, использующим полностью размеченные данные. Также исследуется, как предметная область обучающего набора, на котором инициализируется модель для переноса знаний, влияет на последующее дообучение модели.

Ключевые слова: активное обучение; перенос знаний; сегментация изображений

Для цитирования: Киранов Д.М., Рындин М.А., Козлов И.С. Активное обучение и перенос знаний в задаче сегментации изображений документов. Труды ИСП РАН, том 33, вып. 6, 2021 г., стр. 205-216. DOI: 10.15514/ISPRAS-2021-33(6)-14

Active learning and transfer learning for document segmentation

^{1,2} D.M. Kiranov, ORCID:0000-0002-3507-3803 <kiranov.dm@ispras.ru>

¹ M.A. Ryndin, ORCID:0000-0002-7504-3975 <mxrynd@ispras.ru>

¹ I.S. Kozlov, ORCID:0000-0002-0145-1159 <kozlov-ilya@ispras.ru>

¹ Ivannikov Institute for System Programming of the Russian Academy of Sciences,
25, Alexander Solzhenitsyn st., Moscow, 109004, Russia

² Moscow Institute of Physics and Technology,
9, Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russia

Abstract. In this paper, we investigate the effectiveness of classical approaches of active learning in the problem of segmentation of document images in order to reduce the training sample. A modified approach to the selection of images for marking and subsequent training is presented. The results obtained through active learning are compared to transfer learning using fully labeled data. It also investigates how the subject area of the training set, on which the model is initialized for transfer learning, affects the subsequent additional training of the model.

Keywords: active learning; transfer learning; image segmentation

For citation: Kiranov D.M., Ryndin M.A., Kozlov I.S. Active learning and transfer learning for document segmentation. Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 6, 2021, pp. 205-216 (in Russian). DOI: 10.15514/ISPRAS-2020-33(6)-14

1. Введение

При анализе структурированных документов, таких как научные статьи, правовые акты, полезным первым шагом для последующего анализа является выделение сложных структурных элементов: текстовых блоков, таблиц, списков, графиков и названий. Разметка изображений для обучения – времяемкое занятие, кроме того задачи сегментирования и обнаружения объектов в некоторых узкоспециализированных предметных областях могут требовать специфических знаний для корректной аннотации изображений, что ограничивает круг лиц, способных осуществить разметку для обучения. Таким образом, ограниченный человеческий ресурс требует от нас минимизации числа обучающих примеров.

Кроме того, не все изображения являются одинаково полезными для обучения; к примеру, в отсканированных документах объект-текст присутствует почти на каждом изображении, в то время как объект-список встречается не так часто, и, стало быть, обучаясь на нем, модель может извлечь больше полезной информации. Отсюда, в условии минимизации числа обучающих примеров естественным образом возникает вопрос о получении тех примеров, обучившись на которых, модель могла бы лучше всего справляться с поставленной задачей.

В данной работе рассматривается подход, в котором примеры для обучения выбираются с помощью активного обучения. Активное обучение (Active Learning) – это раздел машинного обучения, в котором модель может обращаться к оракулу для разметки выбранных неразмеченных данных [1]. Основной целью активного обучения является получение высокой точности модели на новых примерах с минимальным возможным количеством обращений к оракулу (подробнее см. разд. 2). Главной трудностью при этом является выбор наиболее информативных примеров для обучения.

Классическим подходом отбора примеров для разметки является «выбор по степени неуверенности». К основным мерам неуверенности относятся:

- максимальный отступ [2]

$$\varphi(x)_{margin} = \left[1 - \left(\max_{c_1 \in K} p(c_1|x) - \max_{c_2 \in K \setminus c_1} p(c_2|x) \right) \right]^2;$$

- максимальная энтропия [3]

$$\varphi(x)_{entropy} = - \sum_{c \in K} p(c|x) \log p(c|x);$$

- минимальная уверенность [4]

$$\varphi(x)_{min_conf} = 1 - \max_{c_1 \in K} p(c_1|x).$$

1.1 Активное обучение в задачах обнаружения объектов и сегментации

На сегодняшний день существует немало статей [5-9], посвященных исследованию подходов активного обучения в задачах обнаружения объектов и сегментации изображений, которые показали свою эффективность. Так, в статье [5] авторы рассматривали максимальный отступ (*margin*), как способ определения меры неуверенности отдельного предсказания на изображении, а также исследовали способы их агрегации по всему изображению для получения меры неуверенности всего изображения. Рассмотренные способы агрегации неуверенности:

- сумма

$$\Phi_{Sum} = \sum_{i \in D} \varphi_{1vs2}(x_i);$$

- среднее

$$\Phi_{Avg} = \frac{1}{|D|} \sum_{i \in D} \varphi_{1vs2}(x_i);$$

- максимум

$$\Phi_{Max} = \max_{i \in S} \varphi_{1vs2}(x_i).$$

Отдельно исследовался вопрос о несбалансированности классов в обучающей выборке, для чего был рассмотрен подход с применением весовой схемы.

В статье [6] рассматривались подходы активного обучения, основанные не только на работе классификатора, но и более совершенные стратегии, учитывающие локализацию объектов, а также ее устойчивость к гауссовским шумам.

1.2 Активное обучение в задаче сегментации документов

Задача сегментации документов имеет ряд важных особенностей по сравнению с сегментацией естественных изображений, что усложняет задачу и требует еще большего числа обучающих примеров. Во-первых, число объектов для обнаружения в документе превосходит число объектов для обнаружения на изображении из естественного обучающего набора (например, MS-COCO [10]). Во-вторых, имеется значительная несбалансированность предсказываемых классов; так, например, в документах можно встретить сноски, но их количество значительно меньше заголовков. Поэтому обучение с нуля модели, способной достаточно качественно сегментировать изображения документов, очень трудоемкая задача, из-за чего эта предметная область долгое время оставалась мало исследуемой.

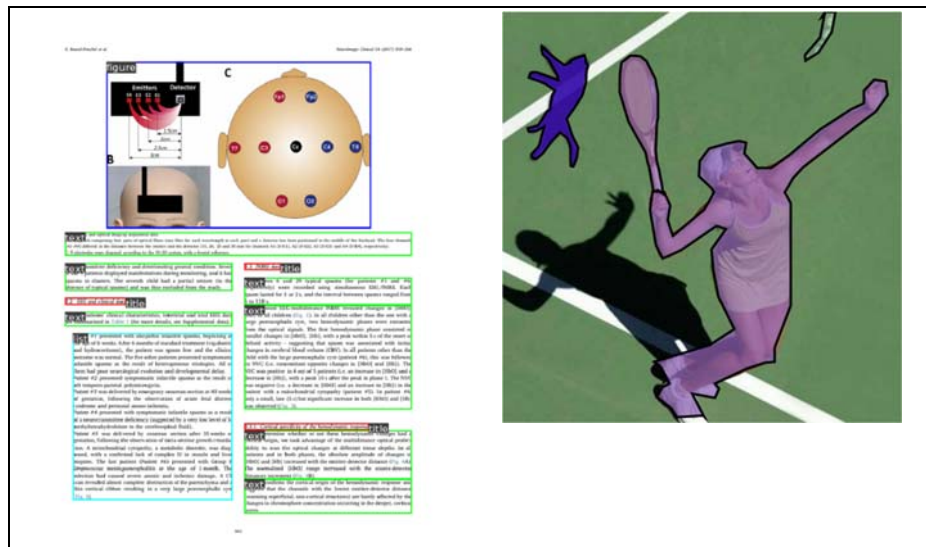


Рис. 1. Типичные примеры изображений из PubLayNet и MS-COCO
Fig. 1. Typical examples of images from PubLayNet and MS-COCO

Существенным прорывом в этой области стало появление в 2019 году обучающего набора *PubLayNet* [11]. Это крупнейший набор размеченных структурированных документов.

В 2020 году была опубликована статья [13], посвященная сегментации документов. Авторы полагают, что стандартный подход активного обучения к выбору изображений для обучения содержит существенный недостаток: методы основываются на агрегированной информативности по всему изображению, в то время как конечная цель – это определение правильной ограничивающей рамки с предсказанием для конкретного объекта на изображении, так для задач со значительной несбалансированностью классов метод может быть не оптимальным из-за чрезмерной распространенности одного из классов. Таким

образом, авторы предлагают подход, в котором оракулу нужно размечать не все изображение, а только объекты, в которых модель наименее уверена.

Однако до сих пор не существует исследований, посвященных использованию классических подходов активного обучения к задаче сегментации изображений документов. Выбирая примеры для обучения только на основании классификации, мы предположительно могли бы получать приемлемое качество, используя меньше ресурсов.

2. Общий вид алгоритма активного обучения

Введём следующие обозначения:

- множество неразмеченных объектов: $X = \{x_1, x_2, \dots, x_n\}$;
- множество меток: $Y = \{y_1, y_2, \dots, y_n\}$;
- хотим восстановить зависимость $a : X \rightarrow Y$.

На каждом шаге активного обучения имеем следующее:

- множество размеченных объектов: X_l ;
- множество неразмеченных объектов: $X \setminus X_l$;
- множество меток для X_l : Y_l ;
- функция выбора самых информативных примеров: $v : X \rightarrow X_l \subset X$;
- оракул, число обращений к которому минимизируется: $f : X_l \rightarrow Y_l$;
- $X_l = X_l \cup X_i, Y_l = Y_l \cup Y_i$.

Тогда общий вид алгоритма активного обучения может быть описан псевдокодом, показанным на Листинге 1.

- 1: инициализация весов на k произвольных изображениях:
($(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$)
- 2: for $i = 1$ to $count$ do
- 3: выбор n изображений по предсказаниям модели, в соответствии с выбранной стратегией:
 $\Phi : X \setminus X_l \rightarrow X_n \subset X$
- 4: Разметка соответствующих изображений
 $f : X_n \rightarrow Y_n$
- 5: обучение модели по выбранным экземплярам
- 6: $X_l = X_l \cup X_n, Y_l = Y_l \cup Y_n$
- 7: end for

Листинг 1. Псевдокод модели активного обучения

Listing 1. Pseudocode for an active learning model

3. Исследуемые стратегии разметки

Так как нами не было найдено исследований классических подходов активного обучения в исследуемой предметной области, в данной работе мы исследовали следующие классические способы подсчета неуверенности отдельного предсказания:

- максимальный отступ [2],
- максимальная энтропия [3],

а также сумму и максимум в качестве функции агрегации:

- $\Phi_{Sum} = \sum_{x_i \in X} \varphi(x_i)$;
- $\Phi_{Max} = \max_{x_i \in X} \varphi(x_i)$.

Классификатор используемой модели помимо вероятностей структурных элементов документа выдаёт также вероятность фона изображения, из чего появилось предположение о том, что использование вероятности фона может быть несодержательным и искажает оценку

неуверенности, что может приводить к неоптимальному решению. Для проверки этого предположения был рассмотрен новый способ подсчета информативности изображения, не учитывающий вероятности предсказания фона изображения.

Пусть p_1, p_2, \dots, p_n – исходные вероятности, вычисленные классификатором модели, где p_1 – вероятность фона, а p_2, \dots, p_n – вероятности структурных элементов документа. Тогда переопределим вероятности следующим образом:

$$p_i^* = \frac{p_{i+1}}{\sum_{i \in \{2, \dots, n\}} p_i}, i = 1, n-1.$$

Откуда неуверенность отдельного предсказания определяется следующим образом:

$$\varphi(x)_{margin_norm} = \left[1 - \left(\max_{c_1 \in K^*} p^*(c_1|x) - \max_{c_2 \in K^* \setminus c_1} p^*(c_2|x) \right) \right]^2,$$

$$\varphi(x)_{entropy_norm} = - \sum_{c \in K^*} p^*(c|x) \log p(c|x),$$

где $K^* = K \setminus c_{фон}$.

4. Наборы данных

MS COCO [10] – популярный крупный (порядка 330 тысяч изображений) набор данных для задач обнаружения объектов и сегментации. *MS COCO* содержит распространенные объекты в их естественном окружении.

PubLayNet [11] – крупнейший обучающий набор, посвященный анализу структуры документов. *PubLayNet* содержит изображения статей и исследовательских работ; в нем представлены категории, приведенные в табл. 1.

Табл. 1. Список категорий в *PubLayNet*

Категории согласно статье [11]	Категория
Text	Текст
Title	Заголовок
List	Список
Table	Таблица
Figure	Изображение

DLA [12] – маленький (порядка 500 документов для обучения и 300 документов для валидации) обучающий набор ручной разметки, состоящий из правовых актов на русском языке. В нем представлены категории, приведенные в табл. 2.

Табл. 2. Список категорий в *DLA*

Категории в <i>DLA</i>	Категория
Text	Текст
Table	Таблица
Figure	Изображение

5. Перенос знаний

Одним из способов сокращения обучающей выборки является перенос знаний (Transfer Learning). Суть метода заключается в том, чтобы использовать знания, полученные моделью из решения одной задачи (для которой имеется большой размеченный обучающий набор), к задаче, в которой получение большого числа обучающих примеров невозможно, либо очень затратно.

Для получения базовой модели мы выполнили перенос знаний с модели предобученной на *MS COCO* на *PubLayNet*, так как используемый фреймворк не имел модели, предобученной на документах. Далее был выполнен перенос знаний на обучающий набор *DLA*, с целью получения верхней оценки на будущее решение.

Кроме того, нас интересовало, как предметная область, на которой обучалась модель изначально, влияет на последующее дообучение. Формально, пусть в нашем распоряжении имеется 2 обучающих набора *A* и *B* с классами $\{a_1, a_2, \dots, a_n\}$ и $\{b_1, b_2, \dots, b_m\}$ соответственно. Даст ли модель M_B , предобученная на обучающем наборе *B*, преимущество при активном обучении на обучающем наборе *C* с классами $\{b_1, b_2, \dots, b_p\}$, $p' < m$, по сравнению с моделью M_A , предобученной на обучающем наборе *A*?

Изначально мы попробовали применить обученную *PubLayNet*-модель для сегментации документов из *DLA*, так как она умела определять все интересующие нас классы. Однако показатель качества такой модели на новых данных был равен нулю, что свидетельствовало о том, что модели для качественного предсказания требуется дообучение даже в рамках одной предметной области. Тогда мы решили проверить, как изменится качество предсказаний выбранной модели, если исходная *MS COCO*-модель не будет предобучаться на *PubLayNet*, а сразу будет активно обучаться на небольшом наборе *DLA* из другой предметной области.

6. Эксперименты

6.1 Показатель качества

Популярный и часто используемый показатель качества в задаче сегментации – модифицированный *Mean Average Precision*. Это обычный *mAP*, усредненный для 10 пороговых значений *IoU* с шагом 0.05: [*IoU*=0.50:0.95]. Авторы метрики [10] заявляют, что усреднение по *IoU* позволяет лучше локализовать объекты.

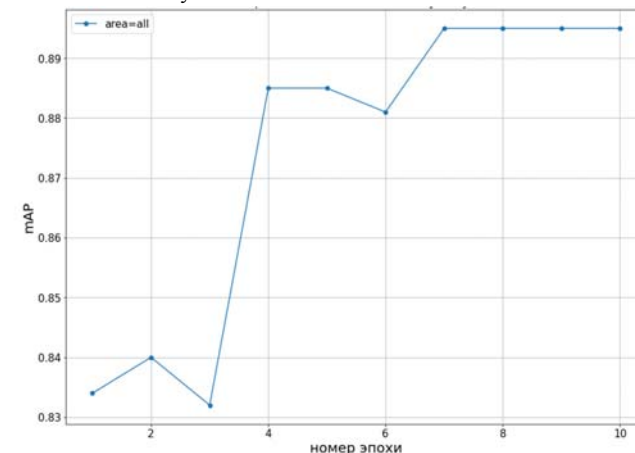


Рис. 2. Перенос знаний с *MS COCO* на *PubLayNet*
Fig. 2. Transfer of knowledge from *MS COCO* to *PubLayNet*

6.2 Модель

В силу структуры документов задачи семантической сегментации, сегментации экземпляров и обнаружения объектов являются практически эквивалентными в этой предметной области. В частности, это подтверждают результаты, полученные авторами статьи [11], они обучали

модель сегментации экземпляров (*Mask RCNN* [15]), а также модель обнаружения объектов (*Faster RCNN* [14]). Существенных отличий в качестве предсказаний моделей не было, а так как наш метод для определения степени неуверенности изображения использует только результаты работы классификатора, то использование более тяжелой модели мы посчитали неразумным, поэтому нами была выбрана модель *Faster RCNN*. Плюсом является то, что в *PyTorch*¹ имеется реализация *Faster RCNN*, предобученная на *MS COCO*.

Так как у нас не было модели *Faster RCNN*, предобученной на *PubLayNet*, а обучение было очень затратным, мы выполнили перенос знаний с *MS COCO* на 1/7 часть *PubLayNet* с результатами на отложенной выборке *PubLayNet*-а, показанными на рис. 2.

Полученная модель имитировала модель, предобученную на *PubLayNet*.

6.3 Дизайн экспериментов

6.3.1 Выбор стратегии сэмпинга

Модель, предобученная на *PubLayNet*, инициализировала веса выходного слоя на 10 произвольных изображениях из *DLA* в течение 10 эпох обучения. Эта модель сохранялась и была начальной для всех последующих экспериментов с различными стратегиями активного обучения. Далее для каждой стратегии выбора экземпляров по степени неуверенности следовало 49 итераций активного обучения, на каждой итерации модель получала 8 новых примеров с наибольшей мерой неуверенности и дообучалась на всех размеченных на текущий момент данных ($8 \times \text{номер итерации}$) в течение 10 эпох.

Для получения более объективных результатов был реализован скользящий контроль с разбиением на 5 групп. После инициализации выходного слоя, в нашем распоряжении было 490 изображений для обучения, то есть одна группа состояла из 98 изображений. Таким образом, на первой итерации обучение происходило на 8 изображениях, а на 49-й итерации – на $490 \times \frac{1}{5} \times 4 = 392$ изображениях; валидация все время осуществлялась на оставшихся 98 изображениях (как это и устроено в классическом скользящем контроле).

6.3.2 Критерий выбора стратегии

Получение показателя качества, соответствующего верхней границе, на меньшем числе обучающих примеров – малореальная задача. Тем не менее, нам нужно формализовать то, что мы будем считать достаточно качественным обучением в условиях минимизации обращений к оракулу для разметки.

Мы предполагаем, что на ранних итерациях активный выбор обучающих примеров приведет к более быстрому обучению (по сравнению с произвольным выбором примеров для обучения) до некоторого порогового значения, которое впоследствии будет не сильно меняться при добавлении новых размеченных данных. Под достаточно качественной моделью будем понимать такую, что ее относительное качество не будет меняться на 1 процент и более в течение трех итераций дообучения на новых данных. Таким образом, мы получим модель, показатель качества которой относительно быстро выходит на некоторое плато, а затем мало изменяется при добавлении новых обучающих примеров.

6.4 Результаты

6.4.1 Выбор стратегии сэмпинга

На рис. 3 приведен график зависимости среднего показателя качества модели в зависимости от числа итераций активного обучения (фактически, числа обучающих примеров). Из-за

количества рассмотренных стратегий выбора примеров для обучения, а также схожести многих из них на некоторых участках, по данному графику трудно делать выводы о качестве какой-то конкретной модели, тем не менее можно отметить, что на начальных этапах обучения (≤ 20 итераций) *random* и *margin_max* отстают от остальных стратегий, а к концу обучения (≥ 30 итераций) все стратегии выходят на плато.

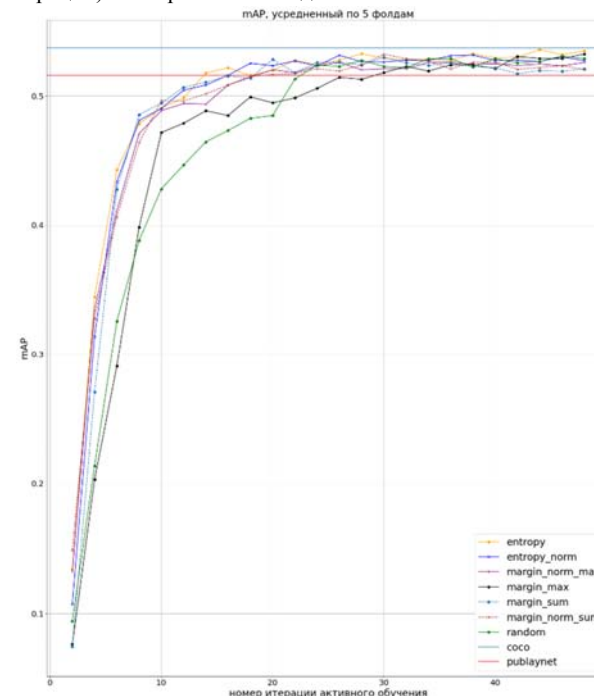


Рис. 3. Усредненный по 5-ти фолдам mAP
Fig. 3. mAP averaged over 5 folds

Также на рис. 3 приведен результат моделей, для которых была выполнена перекрестная проверка переноса знаний с *MS COCO* и *PubLayNet*.

Для сравнения разных стратегий активного обучения приводим диаграммы размаха (рис. 4) показателя качества для разных стратегий активного обучения в зависимости от номера итерации. Также на графиках отображены средние значения показателя качества модели, выбиравшей примеры для обучения произвольно.

margin_sum, *entropy*, *entropy_norm* имеют наибольший прирост по сравнению с *random* стратегией; кроме того, их отличает наименьшая величина межквартильного размаха, что говорит о небольшом разбросе рассматриваемых значений, то есть качество модели не сильно зависит от того, на каких данных обучались модели.

Худший результат демонстрирует *margin_max* с минимальными значениями медианы, а также наибольшим значением межквартильного размаха. При этом использование нормировки позволяет быстрее выйти на плато в области $mAP \geq 0.5$, а также уменьшить разброс значений показателя качества модели.

¹ <https://pytorch.org/>

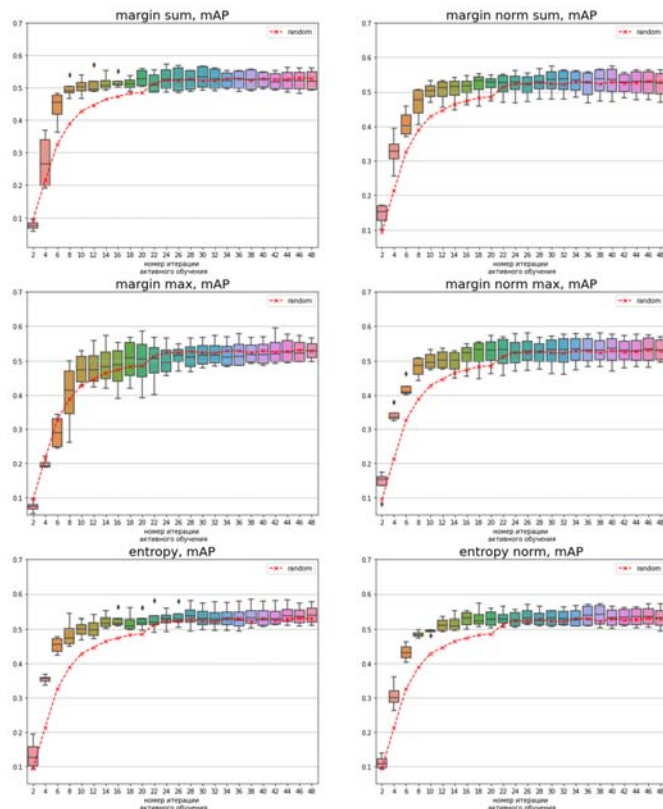


Рис. 4. Сравнение разных стратегий активного обучения
Fig. 4. Comparison of different active learning strategies

В соответствии с критерием отбора (п. 6.3.2) приведем также табл. 3, демонстрирующую, на какой итерации разные стратегии стали удовлетворять критерию.

Табл. 3. Сравнение разных стратегий активного обучения
Table 3. Comparison of different active learning strategies

Стратегия	№ итерации	mAP_{cv}	Transfer ₁	Transfer ₂
margin sum	12	0.507	0.98	0.94
entropy	14	0.518	1.00	0.96
entropy norm	18	0.525	1.01	0.98
margin norm sum	18	0.515	0.99	0.96
random	24	0.524	1.02	0.98
margin norm max	24	0.523	0.99	0.97
margin max	30	0.518	1.00	0.96
Transfer _{publaynet}	-	0.516	1.00	0.96
Transfer _{coco}	-	0.537	1.04	1.00

Как видно из табл. 3, стратегии выходят на плато добиваясь 98%-го (и более) показателя качества от переноса знаний с *PubLayNet* и 94%-го (и более) показателя качества от переноса знаний с *MS COCO*. Кроме того, перенос знаний с *PubLayNet* показал худшее качество по сравнению с переносом знаний с *MS COCO*, несмотря на то, что классы обучающего набора, на котором проходили эксперименты, содержатся в *PubLayNet* и отсутствуют в *MS COCO*.

Для полноты картины приведем также табл. 4, которая показывает, какого показателя качества удалось добиться каждой модели к итерации, когда некоторая стратегия вышла на плато согласно критерию.

Табл. 4. Зависимость mAP_{cv} от номера итерации
Table 4. Dependence of mAP_{cv} on the iteration number

Стратегия/итерация	12	14	18	24	30
margin sum	0.507	0.511	0.514	0.526	0.529
entropy	0.498	0.518	0.516	0.524	0.529
entropy norm	0.504	0.508	0.525	0.523	0.526
margin norm sum	0.496	0.502	0.515	0.521	0.532
random	0.446	0.464	0.482	0.524	0.523
margin norm max	0.494	0.4936	0.514	0.523	0.521
margin max	0.479	0.488	0.499	0.506	0.518

Таким образом, использование в качестве меры не уверенности *margin* (или *entropy*) позволяет к 12-й (14-й) итерации получить 94% (96%) качества от переноса знаний с *MS COCO*, используя лишь $10 + 12 \times 8 = 106$ ($10 + 14 \times 8 = 122$) изображений из 392. В то же время произвольный выбор изображений для обучения показывает лишь 83% (86%) качества модели, обучающейся на всех данных.

6.4.2 Перенос знаний

Также были проведены дополнительные эксперименты по исследованию зависимости качества активного обучения от инициализации начальной модели.

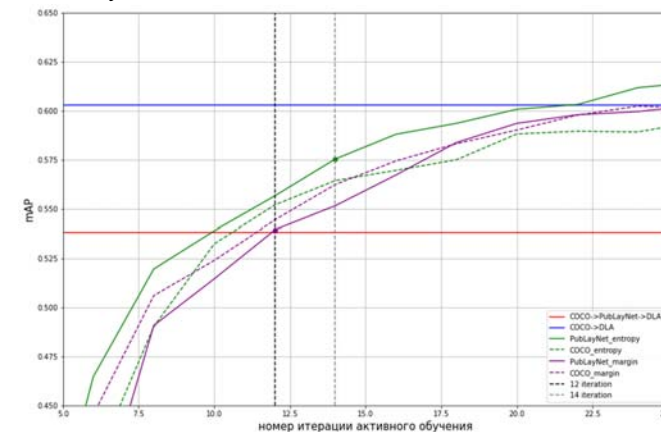


Рис. 5. Качество активного обучения на отложенной выборке в зависимости от инициализации
Fig. 5. The quality of active learning on deferred sampling depending on initialization

На рис. 5 приведены зависимости двух лучших стратегии активного обучения из предыдущего пункта для разных инициализаций на отложенной выборке: инициализация моделями, предобученными на *PubLayNet* и *MS COCO*.

Лучшее качество демонстрирует модель, предобученная на *PubLayNet* с использованием *entropy* в качестве меры не уверенности. Однако сказать, что модель, предобученная на смежном обучающем наборе, даёт преимущество при активном обучении, нельзя, так как *margin* с инициализацией на *MS COCO* превосходит аналогичную модель с инициализацией на *PubLayNet*.

7. Заключение

Использование классических подходов активного обучения в задаче сегментации изображений документов можно считать эффективным, так они позволяют получать высокое качество предсказания, используя лишь около четверти данных, требуемых для переноса знаний.

Модифицированные подходы активного обучения на основе перенормировки исходных вероятностей не показали преимуществ по сравнению с классическими стратегиями выбора по степени неуверенности.

Перенос знаний на смежный обучающий набор не показал преимущества по сравнению с переносом знаний с обучающего набора с произвольными классами.

Список литературы / References

- [1] Settles B. Active learning literature survey. Technical Report #1648, University of Wisconsin-Madison, Department of Computer Sciences, 2009, 47 p.
- [2] Scheffer T., Decomain C., Wrobel S. Active hidden markov models for information extraction. In Proc. of the International Symposium on Intelligent Data Analysis, 2001, pp. 309-318.
- [3] Dagan I., Engelson S. Committee-based sampling for training probabilistic classifiers. In Proc. of the Twelfth International Conference on Machine Learning, 1995, pp. 150-157.
- [4] Culotta A., McCallum A. Reducing labeling effort for structured prediction tasks. In Proc. of the 20th National Conference on Artificial Intelligence, 2005, pp. 746-751.
- [5] Brust C., Käding C., Denzler J. Active Learning for Deep Object Detection. In Proc. of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2019, pp. 181-190.
- [6] Kao C., Lee T. et al. Localization-Aware Active Learning for Object Detection. In Proc. of the 14th Asian Conference on Computer Vision, 2018, pp. 506-522.
- [7] Roy S., Unmesh A., Namboodiri V. Deep active learning for object detection. In Proc. of the 29th British Machine Vision Conference, 2018, 12 p.
- [8] Aghdam H., Gonzalez-Garcia A. et al. Active Learning for Deep Detection Neural Networks. In Proc. of the 17th IEEE/CVF International Conference On Computer Vision, 2019, pp. 3671-3679.
- [9] Lv X., Duan F. et al. Deep active learning for surface defect detection. *Sensors*, vol. 20, no. 6, 2020, article no. 1650.
- [10] Lin T., Maire M. et al. Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science*, vol. 8693, 2014, pp. 740-755.
- [11] Zhong X., Tang J., Yepes A. PubLayNet: largest dataset ever for document layout analysis. In Proc. of the International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 1015-1022.
- [12] Беляева О.В., Перминов А.И., Козлов И.С. Использование синтетических данных для тонкой настройки моделей сегментации документов. *Труды ИСП РАН*, том 32, вып. 4, 2020 г., стр. 189-202 / Belyaeva O.V., Perminov A.I., Kozlov I.S. Synthetic data usage for document segmentation models fine-tuning. *Trudy ISP RAN/Proc. ISP RAS*, vol. 32, issue 4, 2020. pp. 189-202 (in Russian). DOI: 10.15514/ISPRAS-2020-32(4)-14
- [13] Shen Z., Zhao J. et al. OLALA: Object-Level Active Learning for Efficient Document Layout Annotation. *arXiv:2010.01762*, 2021, 12 p.
- [14] Ren S., He K. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proc. of the 28th International Conference on Neural Information Processing Systems, 2015, pp. 91-99.
- [15] He K., Gkioxari G. et al. Mask R-CNN. In Proc. of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988.

Информация об авторах / Information about authors

Дмитрий Маратович КИРАНОВ – студент магистратуры МФТИ, лаборант в ИСП РАН. Научные интересы: активное обучение, непрерывное обучение.

Dmitry Maratovich KIRANOV – MIPT master's student, laboratory assistant at ISP RAS. Research interests: active learning, lifelong learning.

Максим Алексеевич РЫНДИН – аспирант. Научные интересы: методы адаптации к домену и переноса знаний, онлайн обучение, обработка текстов на естественном языке, векторное представление слов/предложений/текстов, генеративные модели, активное и проактивное обучение, анализ социальных сетей.

Maxim Alexeevitch RYNDIN – PhD Student. Research interests: domain adaptation, transfer learning, online learning, natural language processing, embeddings, generative models, active and proactive learning, social media analysis.

Илья Сергеевич КОЗЛОВ является стажером-исследователем. Научные интересы: распознавание структуры документов, цифровая обработка изображений, нейросетевая обработка данных, распознавание образов.

Ilya Sergeevich KOZLOV – researcher at ISP RAN. Research interests: document layout analysis, digital image processing, neural network data processing, image pattern recognition.