# Использование синтетических данных для тонкой настройки моделей сегментации документов



ИНСТИТУТ СИСТЕМНОГО ПРОГРАММИРОВАНИЯ ИМ. В.П. ИВАННИКОВА РОССИЙСКОЙ АКАДЕМИИ НАУК

Беляева Оксана Владимировна Перминов Андрей Игоревич Козлов Илья Сергеевич

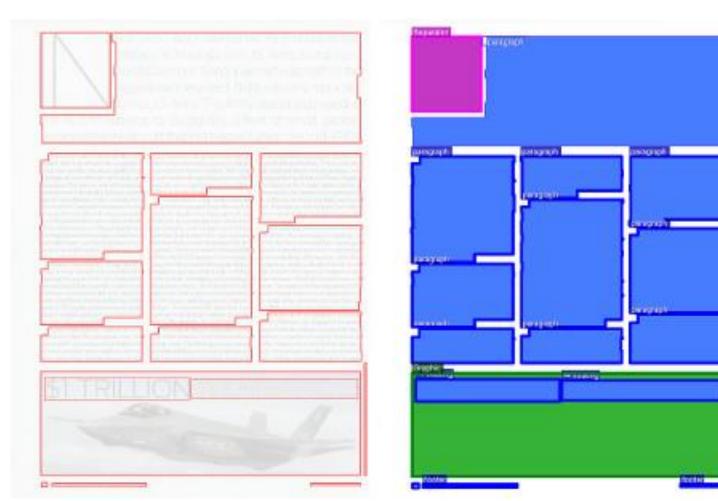
### Актуальность

- документы в PDF, зачастую с плохим текстовым слоем или даже сканы
- необходимо автоматизировать обработку документов
- в документах есть блоки, в блоках таблицы, заголовки, текст, списки, ...
- нужно извлекать текстовую составляющую, а также дополнительную информацию
- необходима сегментация документа

### Сегментация документа

Сегментация документа – это процесс разделения изображения

документа на несколько сегментов



[TABLE III]
RESULTS FOR THE ICDAR2017 COMPETITION ON LAYOUT ANALYSIS
FOR CALLENGING MEDIEVAL MANUSCRIPTS [7] - TASK-1 (IOU)

Method	CB55	CSG18	CSG863	Overall
System-1 (KFUPM)	.7150	.6469	.5988	.6535
System-6 (IAIS)	.7178	.7496	.7546	.7407
System-4.2 (MindGarage-2)	.9366	.8837	.8670	.8958
System-2 (BYU)	.9639	.8772	.8642	.9018
System-3 (Demokritos)	.9675	.9069	.8936	.9227
System-4.1 (MindGarage-1)	.9864	.9357	.8963	.9395
dhSegment	.9757	.9322	.9130	.9403
dhSegment + Page	.9783	.9317	.9205	.9435
System-5 (NLPR)	.9835	.9365	.9271	.9490









Fig. 5. Example of layout analysis on the DIVA-HisDB test test. On the left the original manuscript image, in the middle the classes pixel-wise labelled by the dhSegment and on the right the comparison with the ground-truth (refer to the evaluation tool\*for the signification of colors)

The private dataset [23] used for this task is composed of several printed books. The selected pages were manually annotated and each ornament was marked by a bounding rectangle. A total of 912 annotated pages were produced, with 612 containing one or several ornaments. The dataset is split in the following way [

- 610 pages for training (427 with ornaments)
- 92 pages for evaluation (62 with ornaments)
- 183 pages for testing (123 with ornaments)

The original images are resized to  $8\cdot 10^5$  and the model is trained for 30 epochs with batch size of 16. The training takes less than two hours.

To obtain the binary masks, a threshold of t=0.6 is applied to the output of the network. Opening and closing operations are performed and a bounding rectangle is fitted to each detected ornament. Finally, very small boxes (those with areas less than 0.5% of the image size) are removed.

The evaluation of ornaments detection task is measured using the standard Intersection over Union (IoU) metric, which calculates how well the predicted and the correct boxes overlap. Table IV lists the precision, recall and f-measure for three IoU thresholds as well as the mean IoU (mIoU) measure Our results are compared to the method implemented in [23] which uses a region proposal technique coupled with a CNN classifier to filter false positives

### E. Photo-collection extraction

A very practical case comes from the processing of the scans of an old photo-collection. The inputs are high resolution scans of pieces of cardboard with an old photograph stuck in the

Fig. 6. The left image illustrates the case of a partially detected ornamer the middle one shows the detection of an illustration but also a false positive detection of the banner and the right image is a correct example of multiple.

TABLE IV

ORNAMENTS DETECTION TASK. EVALUATION AT DIFFERENT IOU

TUBERROL DE ON TEST SET

Method	IoU	F-val	P-val	R-val	mIoU
(22)	0.5	- 7	0.800	0.430	- 12
[23]	0.5		0.470	0.600	
dhSegment	0.7	0.941	0.969	0.914	
	0.8	0.874	0.847	0.902	0.870
	0.9	0.510	0.374	0.803	

middle, and the task is to properly extract the part of the scan containing the cardboard and the image respectively.

Annotation was done very quickly by directly drawing on the scans the part to be extracted in different colors (background, cardboard, photograph). Leveraging standard image diting software, 60 scans per hour can be annotated. Data was split in 100 scans for training, 20 for validation and 150 for testing. Training for 40 epochs took only 20 minutes. The network is trained to predict for each pixel its belonging to one of the classes.

The predicted classes are then cleaned with a simple morphological opening, and the smallest enclosing rectangle of the corresponding region is extracted. Additionally, one can use the layout constraint that the area of the photograph has to be enclosed in the area of the piece of cardboard. We compare the extracted rectangle with the smallest rectangle coming from the groundfurth mask and display relevant metrics in Table V.

TABLE V
PHOTO-COLLECTION EXTRACTION TASK, EVALUATION OF MIOU ON TEST
SET, AND SOME RECALL AT IOU THRESHOLDS OF 0.85 AND 0.95

Method	Cardboard	25	Photo	
	mIoU	mIoU	R@0.85	R@0.95
Predictions-only	0.992	0.982	0.980	0.967
+ layout constraint	0.992	0.988	1.000	0.993

### V. DISCUSSIO

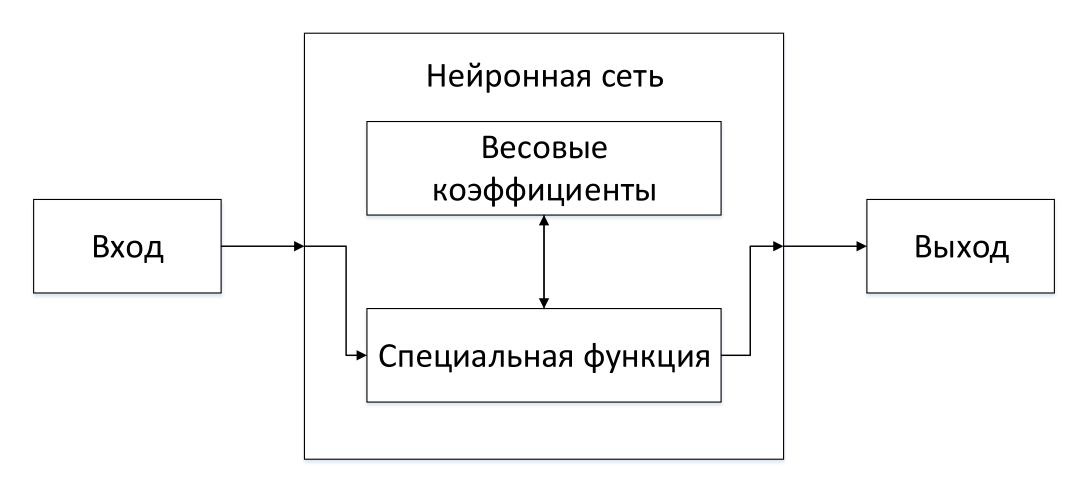
While using the same network and almost the same training configurations, the results we obtained on the five different

### Алгоритмы сегментации документа

	Методы, основанные на правилах	Методы, использующие машинное обучение
Требуют обучения	-	+
Высокое качество	-	+
Большой набор классов	-	+
Малое время работы	+	-
Нетрудоёмкая адаптация к новым данным	-	+

### Нейронные сети

**Искусственная нейронная сеть** — отображение входных данных в выходные, используя сравнительно большой набор весовых коэффициентов



### Обучение и дообучение

**Обучение нейронной сети** — процесс настройки весовых коэффициентов выбранной модели для наилучшего отображения входных данных в выходные

### Обучение с нуля

- Коэффициенты нейронной сети инициализируются случайным образом
- Требуется большой объём обучающих данных
- Требуется много времени на обучение

### Добучение

- Используются коэффициенты нейронной сети, обученной на схожих данных
- Обычно требуется небольшой объём обучающих данных
- Требуется сравнительном меньше времени на обучение

### Формат входных данных

- документы технических заданий и нормативно-правовых актов;
- тёмный текстом на светлом однородном фоне;
- документы манхэттенского типа;

### 12. Этапы выполнения ОКР

Νe	Наименование и краткое	Cpox
π/π	содержание этапа	выполнения
1	Разработка эскизного проекта	
2	Разработка технического проекта	
3	Разработка рабочей конструкторской документации	
4	Изготовление ОО и проведение предварительных испытавий. Доработка опытного обраща, корректировка рабочей конструкторской документации по результатам предварительных испытавий	
5	Государственные испытания	
ő	Доработка опытного обраща по результатам государственных испытаний. Утверждение рабочей конструкторской документации для организации промышленного производства	
~	Сертификация опытного образца изделия	

- 1. ОКР начинается с даты подписания государственного контракта.
- 2. Сроки выполнения этапов ОКР определяются ведомостыс исполнения государственного контракта по результатам конкурсных
- 13. Порядок выполнения и приемки этапов ОКР
- 13.1 Порядок выполнения и приёмки этапов ОКР должен осуществляться в соответствие с требованиями ГОСТ РВ 15.203-2001.
- 13.2 Отчетная документация должна удовлетворять требованиям ГОСТ РВ 15.110-2003, ГОСТ РВ 15.203-2001. Созданная в ходе выполнения ОКР учтенная копия отчетной документации поставляется Заказчику в электронной форме (в соответствии с ГОСТ Р 2.903-96, ГОСТ РВ 2.902-2005 и ΓΟCT 2.102-2013).
- 13.3 Настоящее ТТЗ может быть изменено в порядке, предусмотренном ГОСТ РВ 15.201-2003 и ГОСТ РВ 15.203-2001 по результатам разработки эскизного (технического) проекта, а также этапа разработки рабочей конструкторской документации
- 13.4 Соисполнители ОКР определяются Исполнителем по

- 14. Требования по учету результатов ОКР и результатов интеллектуальной деятельности
- 14.1 Создаваемая при выполнении ОКР научно-техническая продукция является собственностью Заказчика ОКР.
- 14.2 Право интеллектуальной собственности, а также патентное и авторское права при выполнении ОКР устанавливаются действующим
- 14.3 Головной исполнитель ОКР должен выполнить ОКР и передать Заказчику результаты, в том числе разработанные согласно настоящему ТТЗ отчеты, научно-техническую, техническую документацию (в терминологии пункта 3 ГОСТ Р 1.15-2009), результаты интеллектуальной деятельности которым предоставлена или предоставляется правовая охрана, а также изготовленные макеты, стенды, экспериментальные образцы, в предусмотренный настоящим ТТЗ и государственным контрактом срок.
- 14.4 Права на результаты ОКР, включая права на результаты интеллектуальной деятельности, полученные (созданные) и (или) использованные при выполнении ОКР, принадлежат Российской Федерации, от имени которой выступает Заказчик.
- 14.5 В месячный срок после завершения ОКР головной исполнитель ОКР должен представить Заказчику оформленные в установленном в Российской Федерации порядке (по форме 1, утвержденный приказом Минюста Российской Федерации и Минпромнауки от 17.07.2003 № 173/178 «Об утверждении форм документов, необходимых для ведения единого реестра результатов научно-исследовательских, опытно-конструкторских и технологических работ военного, специального и двойного назначения, права на которые принадлежат Российской Федерации») документы, необродимые для государственного учета результатов интеллектуальной деятельности, полученных в ходе выполнении ОКР в соответствии с требованиями постановления Правительства Российской Федерации 2002 г., № 131.
- 14.6 Головной исполнитель ОКР должен провести:
- при завершении (прекращении) выполнения ОКР инвентаризацию результатов интеллектуальной деятельности, созданных (использованных) в данной работе и представить Заказчику документы, для осуществления бюджетного (стоимостного) учета таких результатов в качестве нематериальных активов по их фактической стоимости
- патентные исследования в соответствии с ГОСТ Р 15.001-96 и на дату завершения ОКР обеспечить защиту используемых при выполнении ОКР авторских прав и патентообладателей.

- 14.7 Головной исполнитель ОКР должен уведомлять Заказчика в 30-лиевный срок о каждом полученном при выполнении Контракта результате ОКР, способного к правовой охране в качестве объекта интеллектуальной собственности с обоснованием предлагаемого порядка его использования и формы правовой охраны.
- 14.8 Головной исполнитель ОКР должен согласовать с Заказчиком необходимость использования при выполнении ОКР результатов интеллектуальной деятельности, исключительные права на которые принадлежат Исполнителю или третьим лицам.
- 14.9 Головной исполнитель ОКР должен гарантировать Заказчику передачу полученных результатов, в том числе результатов интеллектуальной деятельности, не нарушающих исключительных прав других лиц.
- 14.10 Распоряжение от имени Российской Федерации правами на результат ОКР, включая права на результаты интеллектуальной деятельности, полученные (созданные) и (или) использованные при выполнении ОКР, осуществляются Заказчиком в порядке, установленном законодательством Российской Федерации.
- 14.11 Техническая или иная покументация, перепледемая Заказчиком головному исполнителю, должна использоваться головным исполнителем исключительно для выполнения работ по Контракту. По завершению работ головной исполнитель обязан возвратить Заказчику всю переданную техническую или иную документацию.
- 14.12 Головной исполнитель ОКР должен заключить лицензионные договора на используемые результаты интеллектуальной деятельности, права на которые принадлежат Российской Фелерации
- 14.13 Головной исполнитель ОКР должен гарантировать установлени режима конфиденциальности в отношении передаваемой технической документации и постановку окраняемых результатов интеплектуальной установленном законом порядке.
- 14.14 В случае нарушения по вине головного исполнителя ОКР исключительного права авторов и (или) иных правообладателей объектов интеллектуальной собственности он несет за это единоличную ответственность в соответствии с законодательством Российской Федерации.

- программный комплекс выявления изменений ведомственных

пк ои программный комплекс мониторинга данных открытых

- программный комплекс мониторинга изменений по данным дистанционного зондирования Земли

- программный комплекс предварительной оценки достоверности данных

ПК БДИМ - программный комплекс ведения базы данных изменений

TIK OOBT - плогозминый комплекс обновления объектноориентированной базы данных

подвижный многоцелевой топографо-геодезический комплекс

- программное обеспечение

персонатьная эпектронная вычистительная машина

тактико-техническое залание

Топографическая служба Вооруженных Сил Российской

### 1. Наименование, шифр ОКР, основание, исполнитель и сроки выполнения ОКР

- 1.1 Наименование ОКР: «Разработка автоматизированной системы топографического мониторинга в целях обеспечения военного картографирования в масштабе времени близком к реальному».
- 1.2 Шифр: ОКР «Приоритет»
- 1.3 Основание для выполнения ОКР: постановление Правительства Российской Федерации от «\_\_\_» \_\_\_\_\_ 201\_\_г. № \_\_\_ «О государственном оборонном заказе на 2020 год и плановый период 2021 и 2022 годов».
  - 1.4 Головной исполнитель ОКР определяется по результатам конкурса.
  - 1.5 Споки выполнения работы:
- начало работы 2020 г. (с момента заключения государственного контракта):
  - окончание работы 2022 г.
  - 2. Цель выполнения ОКР, наименование и индекс изделия
- 2.1 Целью выполнения ОКР является повышение эффективности системы выявления изменений местности и автоматизации процессов их учета базе геопространственных данных на основе непрерывного
- топографического мониторинга. 2.2 Полное наименование изделия: «Автоматизированная система топографического мониторинга».
  - Сокращенное наименование изделия: АС ТМ.
  - 2.3 Инлекс создаваемого изделия: АС ТМ 001.
  - 2.4 Назначение изделия
- 2.4.1 AC ТМ предназначена для автоматизированного получения и обработки данных изменений местности, получаемых из различных источников, с целью обновления баз реограстранственных данных в автоматизированных аппаратно-программных комплексах создания ЕС ЭК
- Источниками геопространственных данных изменений местности для АС ТМ должны быть:
- базы данных открытого контура НЦУО РФ, формируемые по информации АИС ФОИВ, в том числе: АИС Минтранса, ГИС Росполресурсов, АИС Минприроды, АИС Роспесхоза, Геопортал Роспестра, АИС Росседькоза и другие. С этой целью АС ТМ должна иметь техническое (с инфолексимуникационными средствами ЕАСО ГПИ) и информационное

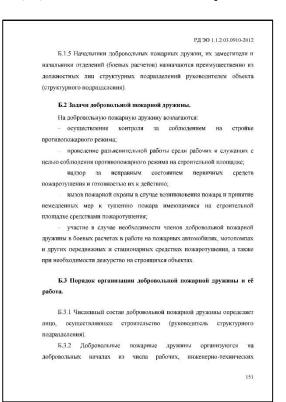
### Классы сегментирования

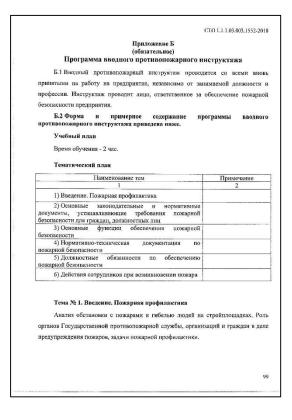
**Text** — блоки, содержащие однородный текст с единым форматированием; **Table** — таблицы с границами, которые могут содержать объединенные ячейки по вертикали или по горизонтали;

**Picture** — класс, содержащий печати, подписи, изображения в документах;

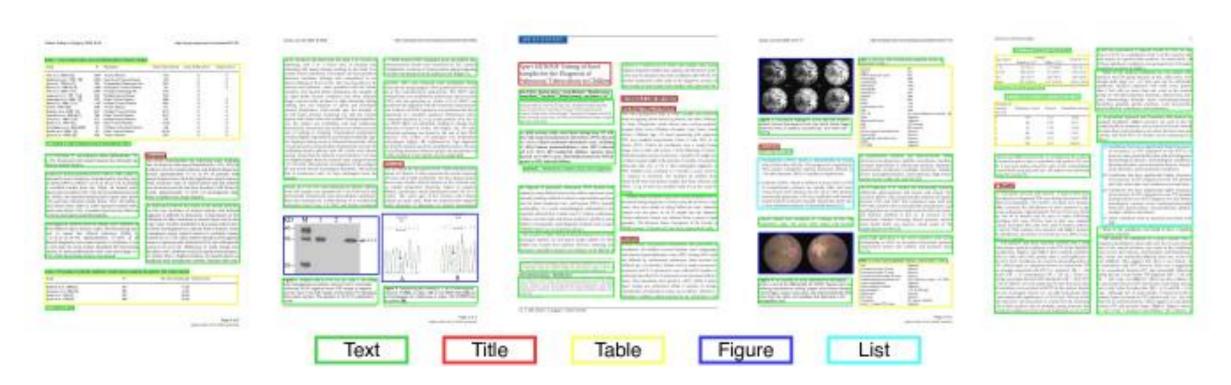








# Готовые наборы данных

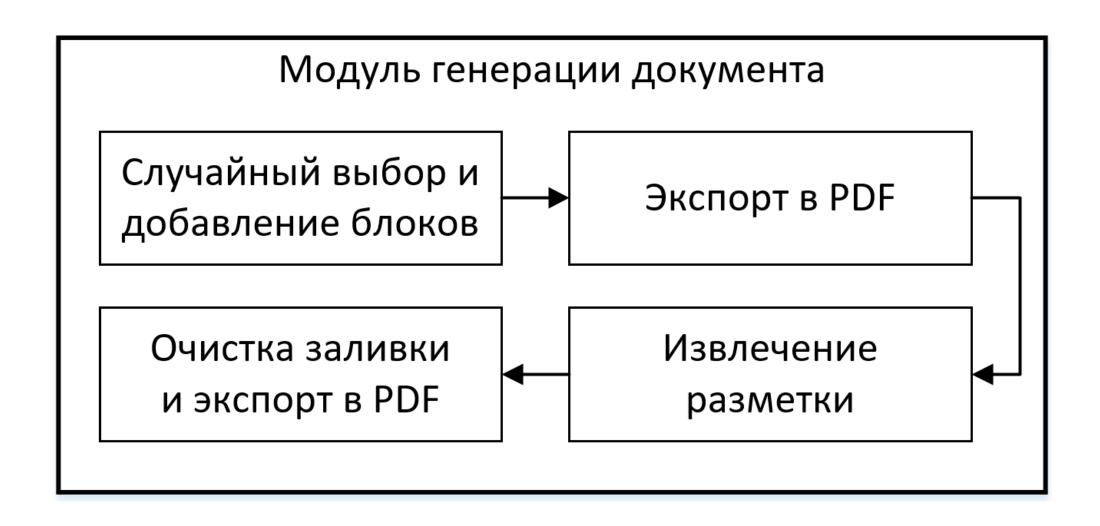


Набор данных PubLayNet

### Зачем генерировать документы

- для обучения модели необходимо иметь большой набор реальных данных;
- существующие размеченные наборы зачастую не подходят;
- разметка данных процесс дорогостоящий;
- решение создать набор из искусственных данных путём самостоятельной генерации;

### Генерация документов



### Извлечение сегментационной разметки

8

Свыше 60 до 80 включительно	100	0,0077	1,1	27
Ke 32-150	£0,09	(63)	30	196
Предварительная наплавка кромок высокохромистыми или перлитными присадочными материалами	28		0,050	0,200

- 8.1.5. При определении границ регистрации оборудования и трубопроводов необходимо руководствоваться следующими требованиями:
- Измерительная линия горизонтальная линия на измерительной поверхности, вдоль которой расположены контрольные точки для установки микрофона.

### приложение в

- выгорание полупроводящего покрытия по ширине хотя бы одного пакета стали;
- ж) проводят проверку исправности действия механизма выключателя по перечислению в):
- б) для внутренней изоляции трансформаторов напряжения и тока, аппаратов и конденсаторов связи - импульсы положительной и отрицательной полярностей;
- в) осуществлять контроль за подготовкой и своевременным предъявлением трубопроводов для технического освидетельствования;
  - б) насос с арматурой или без нее (см. 5.3.4);
  - м) требования к грузозахватным органам и приспособлениям;

### Форма удостоверения контролера

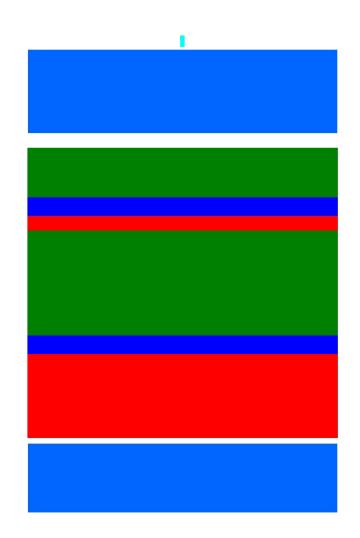
состав имитирующих и моделирующих средств с указанием наименований, шифров и числа:

ГОСТ 15150-69 Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и транспортирования в части воздействия климатических факторов внешней среды

г) зольные помещения:

ГОСТ 12971-67 Таблички прямоугольные для машин и приборов. Размеры

- комплекс взаимоувязанных	П6.7	Показания приборов, регистрирующих
диагностических параметров,		биение и подсчет отклонений в точках
методов, правил, СТД,		по окружности
указанных в технической		
документации, необходимых		



### text table 0,0077 включительно Ke 32-150 £0.09 (63) Предварительная 0.200 наплавка кромок высокохромистыми или перлитными присапочными материалами 1.1.5. При определении границ регистрации оборудования и трубопроводов необходим руководствоваться следующими требованиями: 4. Измерительная линия - горизонтальная линия на измерительной поверхности, вдоль оторой расположены контрольные точки для установки микрофона. приложение в ие полупроводящего покрытия по ширине хотя бы одного пакета стали; перечислению в); б) для внутренней изоляции трансформаторов напряжения и тока, аппаратов онденсаторов связи - импульсы положительной и отрицательной полярностей; в) осуществлять контроль за подготовкой и своевременным предъявлением грубопроводов для технического освидетельствования; б) насос с арматурой или без нее (см. 5.3.4); м) требования к грузозахватным органам и приспособлениям; Рорма удостоверения контролера состав имитирующих и моделирующих средств с указанием наименований, шифров ГОСТ 15150-69 Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и гранспортирования в части воздействия климатических факторов внешней среды г) зольные помещения; ГОСТ 12971-67 Таблички прямоугольные для машин и приборов. Размеры - комплекс взаимоувязанных Показания приборов, регистрирующих биение и подсчет отклонений в точках диагностических параметров, методов, правил, СТД, по окружности указанных в технической

документации, необходимых

### Особенности создаваемых документов

- документы одноколоночны;
- различные семейства шрифтов;
- жирные шрифты и шрифты большего размера для заголовков;
- таблицы имеют все границы, а также объединённые ячейки;
- создаются многостраничные таблицы;
- случайное выравнивание колонтитулов и нумерации страниц;
- высокая вариативность межстрочных интервалы и отступов;

### Примеры сгенерированных документов

.

### МЕХАНИЧЕСКИЕ СВОЙСТВА МЕТАЛЛА ШВА И НАПЛАВЛЕННОГО МЕТАЛЛА И ХАРАКТЕРИСТИКИ СВАРНЫХ СОЕДИНЕНИЙ

 b) для подачи жидкостей, которые не входят в технологический процесс перекачивания насосом с целью;

Vупр. = Vcпр. - Voпр..

(Ш.7)

(нижняя граница)

 - определение основных параметров и характеристик системы возбуждения и отдельных устройств согласно технической документации на системы возбуждения конкретных типов или отдельные устройства;

применения новых прогрессивных технологических процессов восстановления (текущего ремонта).

### 3. МАТЕРИАЛЫ И ПОЛУФАБРИКАТЫ ДЛЯ РЕМОНТА И РЕКОНСТРУКЦИИ

2.4. Сварные швы и их расположение

7) периодичность каждого из видов контроля;

4.3.2.3 Вертикальные насосы

Скоро сть подач и прово локи, мм/с	ЭЛЕКТРООБО РУДОВАНИЕ ПЕРЕМЕННОГ О ТОКА НА НАПРЯЖЕНИЯ ОТ 1 ДО 7Класс напряжения электробору дования	Уровни звуковой мощности, дБ, в октавных полосах со среднегеомет рическими частотами, Гц	Классифик ационное обозначен ие продукции НП-001-97	Трубопр овод зарегист рирован за №В 	Москва Стандарт информ 2016
Первы й замест итель директ ора по произв одству и эксплу атации АЭС			Кинематиче ская вязкость vж		0,0428
A00	74	RTm, MΠa	33,83	570 78	1

II. Условия применения импортного оборудования, изделий, материалов и комплектующих

РД ЭО 4.5 3-6 8

### - обозначение ТЗ/ТУ на оборудование;

ГОСТ 11929-87 Машины электрические вращающиеся. Общие методы испытаний. Определение уровня шума

г) сведения о группе, классе безопасности и категории сейсмостойкости трубопровода;

безопасной эксплуатации оборудования

отм - средние напряжения растяжения по сечению болта или шпильки, МПа (кгс/мм2)

Установочные размеры двигателей

### Приложение 2 (справочное)

В целях подтверждения заявленной вибрационной характеристики для партии машин для испытаний предъявляют случайную выборку машин (не менее трех) из этой партии.

 - НП-046-03 «Правила устройства и безопасной эксплуатации паровых и водогрейных котлов для объектов использования атомной энергии»;

### Приложение N (справочное)

### - без резисторов.

В случае известного вида закона распределения случайной величины планирование наблюдений в условиях эксплуатации проводится по планам [NUN], [NMR], [NMT] в соответствии с РД 50-690 при однородной доверительной вероятности 0,8 или 0,9 и предельной относительной опиобке  $\delta$ , выбираемой по согласованию с заказчиком из ряда: 0,05, 0,10; 0,15; 0,20.

 а) тепловыделяющие элементы и сборки, рабочие органы системы управления и защиты (далее - СУЗ) и другие коиструкции внутри корпусов реакторов, технологических и иных каналов, содержащие делящиеся, поглощающие или замедляющие материалы;

При местном «термическом отдыхе» ширина зоны нагрева должна быть такой же, как при предварительном и сопутствующем подогреве согласно пп. 6.2.14 и 6.2.15.

### Приложение 2

- его наибольший пик ів вплоть до равного нормированному значению ів. н, которое должно быть не менее 2,5 Ів. н.
- контроль герметичности (статус «НР»);

Москва 2013

Чувствительность регулятора

Добавочные потери в тарированном двигателе с фазным ротором могут быть определены при испытании данным методом двух одинаковых машин с фазным ротором в асинхронном или синхронном режиме.

при  $a \le r \le 1$ 

Обкатку следует проводить на номинальном или близком к номинальному режиме работы насоса, если иное не предусмотрено в ПМ или не оговорено в договоре.

TO 3 8.3.6 529-77

Контрольным сварным соединением является соединение, вырезанное из числа производственных сварных соединений или сваренное отдельно, но являющееся идентичным либо однотипным по отношению к производственным сварным соединениям и предназначенное для проведения разрушающего контроля при аттестации технологий сварки или проверки качества и свойств производственных сварных соединений.

- б) после гибки без нагрева труб:
- б) 5 мм при применении слюдяных пластин.
- б) переменного синусоидального напряжения;
- в) приводит к выходу из строя одного из каналов какой-либо системы безопасности:
  - а) способ сварки:
  - в) способы проверки приборов безопасности;
- в) моделирование схемы со снятием осциллограммы процесса восстановления напряжения:

Характеристики собственного переходного напряжения	Минимальный стаж работы по способу сварки, мес.		Тип расчетной модели	X -
Цикл технологической операции		(Фамилия, инициалы)	1800	1,8
3,5		0 3 10	11	
Контролируется:	Без термообработки	392	0,046	2,0
35	54	Тахометр	28	2

### Приложение Q (справочное)

В случае снижения рабочего давления по результатам технического освидетельствования пробное давление при гидравлическом испытании определяется исходя из разрешенного давления.

Допускается располагать в барабане приварные элементы для крепления внутренних устройств. Организация-изготовитель обязана в инструкции по монтажу и эксплуатации указать порядок съема и установки этих устройств.

Испытания проводят до пересечения одной из линий несоответствия или соответствия, причем возможна постановка на испытания дополнительных образцов для достижения суммарной наработки.

Уровень 2. Предотвращение проектных аварий системами нормальной эксплуатации:

При приварке накладок под опоры и подвески к трубопроводам из труб со спиральными швами минимальное расстояние между краем углового шва приварки накладки и краем

10

### Обучение модели

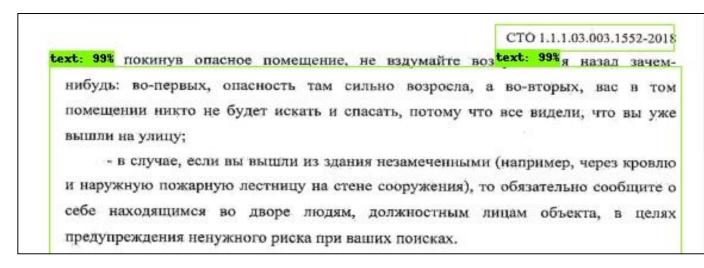
- модель Faster RCNN с претренированными весами на наборе COCO;
- 20 эпох обучения на изображениях PubLayNet (125000);
- 4 эпохи дообучения на сгенерированных изображениях (18000);
- дообучение на малом (100) и большом (500) наборах реальных данных;



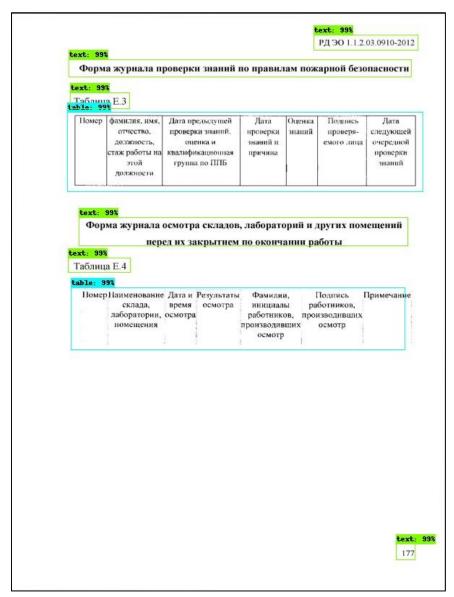
### Возникающие ошибки модели

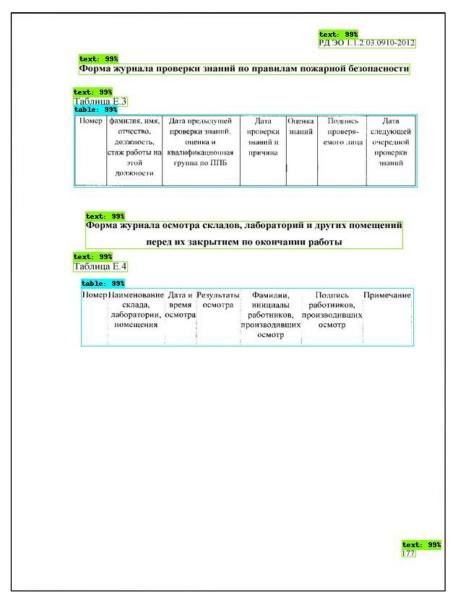
В результате предсказания модели могут образовываться артефакты, негативно влияющие на качество сегментации:

- избыточный захват фоновых пикселей по границам выделенного блока;
- наложение блоков друг на друга;
- обрезка текстового блока;
- обрезка таблиц;



- изменённый алгоритм NoneMaximum-Supression (NMS) для максимального сохранения текстовой информации на изображениях
- бинаризация и инвертирование;
- расширение границ сегментируемых объектов;
- сужение границ сегментируемых объектов;
- устранение пересекающихся текстовые объекты;





CTO 1.1.1.03.003.1552-2018

text: 99% покинув опасное помещение, не вздумайте воз text: 99% я назал зачемнибудь: во-первых, опасность там сильно возросла, а во-вторых, вас в том помещении никто не будет искать и спасать, потому что все видели, что вы уже вышли на улицу;

 в случае, если вы вышли из здания незамеченными (например, через кровлю и наружную пожарную лестницу на стене сооружения), то обязательно сообщите о себе находящимся во дворе людям, должностным лицам объекта, в целях предупреждения ненужного риска при ваших поисках.

Х.3.4 Если дым и пламя в соседних помещениях не позволяют выйти наружу;

- не поддавайтесь панике; помните, что современные железобетонные конструкции в состоянии выдержать высокую температуру;
- если вы отрезаны огнем и дымом от основных путей эвакуации в многоэтажном здании, проверьте, существует ли возможность выйти на крышу или спуститься по незадымляемой пожарной лестнице, или пройти через соседние лоджии;
- если возможности эвакуироваться нет, то для защиты от тепла и дыма постарайтесь надежно загерметизировать свое помещение. Для этого плотно закройте входную дверь, намочите водой любую ткань, обрывки одежды или штор и плотно закройте (заткните) ими щели двери изнутри помещения. Во избежание тяги из коридора и проникновения дыма с улицы – закройте окна, форточки, заткните вентиляционные отверстия, закройте фрамуги вентиляционных решеток;
  - если есть вода, постоянно смачивайте двери, пол, тряпки;
- если в помещении есть телефон, звоните по «01», даже если вы уже звонили
   туда до этого, и даже если вы видите подъехавшие пожарные автомобили.

Объясните диспетчеру, где именно вы находитесь, и что вы отрезаны огнем от выхода;

 если комната наполнилась дымом, передвигайтесь ползком, так будет легче дышать (около пола температура ниже и кислорода больше). Оберните лицо повязкой из влажной ткани, наденьте защитные очки; продвигайтесь в сторону окна, находитесь возле окна и привлекайте к себе внимание людей на улице;

201

toxt. 999

CTO 1.1.1.03.003.1552-2018

- техт: 39%

   покинув опасное помещение, не вздумайте возвращаться назад зачемнибудь: во-первых, опасность там сильно возросла, а во-вторых, вас в том помещении никто не будет искать и спасать, потому что все видели, что вы уже вышли на улицу;
- в случае, если вы вышли из здания незамеченными (например, через кровлю и наружную пожарную лестницу на стене сооружения), то обязательно сообщите о себе находящимся во дворе людям, должностным лицам объекта, в целях предупреждения ненужного риска при ваших поисках.

Х.3.4 Если дым и пламя в соседних помещениях не позволяют выйти наружу;

- не поддавайтесь панике; помните, что современные железобетонные конструкции в состоянии выдержать высокую температуру;
- если вы отрезаны огнем и дымом от основных путей эвакуации в многоэтажном здании, проверьте, существует ли возможность выйти на крышу или спуститься по незадымляемой пожарной лестнице, или пройти через соседние лоджии;
- если возможности эвакуироваться нет, то для защиты от тепла и дыма постарайтесь надежно загерметизировать свое помещение. Для этого плотно закройте входную дверь, намочите водой любую ткань, обрывки одежды или штор и плотно закройте (заткните) ими щели двери изнутри помещения. Во избежание тяги из коридора и проникновения дыма с улицы – закройте окна, форточки, заткните вентиляционные отверстия, закройте фрамуги вентиляционных решеток;
  - если есть вода, постоянно смачивайте двери, пол, тряпки;
- если в помещении есть телефон, звоните по «01», даже если вы уже звонили туда до этого, и даже если вы видите подъехавшие пожарные автомобили.

Объясните диспетчеру, где именно вы находитесь, и что вы отрезаны огнем от выхода;

 если комната наполнилась дымом, передвигайтесь ползком, так будет легче дышать (около пола температура ниже и кислорода больше). Оберните лицо повязкой из влажной ткани, наденьте защитные очки; продвигайтесь в сторону окна, находитесь возле окна и привлекайте к себе внимание людей на улице;

анные доверенности должны быть представлены «Предприятием» в течение пяти календарных дней с момента заключения настоящего договора, а если срок предоставленной доверенности на уполномоченное лицо составляет менее срока действия настоящего договора, «Предприятие» предоставляет новую доверенность за три календарных дня до истечения срока ранее выданной доверенности. В случае не предоставления указанных доверенностей, «Общество» вправе не исполнять принятые на себя 43% зательства по настоящему договору.

- 7.4. «Общество» имеет право на односторонний отказ от исполнения договора полностью или в части в связи с несогласованием «Предприятием» цен на услуги или неоплатой услуг, выполненных «Обществом», и иных платежей причитающихся «Обществу». Отказ от исполнения договора (полностью или частично) по инициативе «Общества» происходит путём направления «Предприятию» письменного уведомления. Договор считается соответственно расторгнутым или измененным по истечении 15 дней с момента получения уведомления, если иное не указано в уведомлении.
- 7.5. По всем вопросам, не предусмотренным настоящим договором, стороны руководствуются Федеральными законами «Уставом железнодорожного транспорта Российской Федерации», «О железнодорожном транспорте в Российской Федерации», Правилами перевозок грузов железнодорожным транспортом, Правилами эксплуатации и обслуживания железнодорожных путей необщего пользования и другими нормативными актами Российской Федерации.
- 7.6. В случае издания правомочными органами нормативных актов, изменяющих условия транспортного обслуживания, тарифы на перевозку грузов, размеры плат и сборов, порядок их оплаты и т.п., стороны руководствуются названными нормативными актами без внесения дополнений и изменений в текст договора.
- 7.7. Стороны обязуются информировать друг друга в письменной форме об изменении существенных фактов деятельности организации: изменение адресов, платежных реквизитов, реорганизации, возбуждении процедуры банкротства и т п. в течение 10
- 7.8. Неотъемлемой частью договора являются: Приложение № 1 «Протокол согласования договорных тарифов на оказание услуг АО «Уралпромжелдортранс» для филиала АО «Концерн Росэнергоатом» «Белоярская атомная станция», Приложение № 2 «Акт обследования грузовых фронтов».
- Настоящий Договор составлен в двух экземплярах, имеющих одинаковую юридическую силу: один экземпляр для «Предприятия», один экземпляр для «Общества».
- 7.10. Период оказания услуг по настоящему договору 36 месяцев с даты подписания договора, в части оплаты - до полного исполнения обязательств.
- 7.11. В случае расторжения договора по решению суда в связи с существенным нарушением «Обществом» условий договора, информация об «Обществе» заносится в реестр недобросовестных поставщиков, предусмотренный Федеральным законом от 18 июля 2011г. № 223-ФЗ «О закупках товаров, работ, услуг отдельными видами юридических лиц» сроком на 2 (два) года».
- 7.12. «Стороны» обязаны обеспечить конфиденциальность сведений, касающихся предмета договора, хода его исполнения и полученных результатов.
- 7.13. «Общество» уведомлено, что в случае нарушения условий настоящего договора в информационную систему «Расчет рейтинга деловой репутации поставщиков», ведение которой осуществляется на официальном сайте по закупкам атомной отрасли www.rdr.rosatom.ru в соответствии с утвержденными Госкорпорацией «Росатом» Едиными отраслевыми методическими указаниями по оценке деловой репутации, могут быть внесены сведения и документы о таких нарушениях. Основанием для внесения сведений в информационную систему «Расчет рейтинга деловой репутации поставшиков» могут являться:
  - выставленные «Предприятием» и принятые «Обществом» неустойки за нарушение

text: 99%



### text: 71%

занные доверенности должны быть представлены «Предприятием» в течение пяти календарных дией с момента заключения настоящего договора, а если срок предоставленной доверенности на уполномоченное лицо составляет менее срока действия настоящего договора, «Предприятие» предоставляет новую доверенность за три календарных дня до истечения срока ранее выданной доверенности. В случае не предоставления указанных доверенностей, «Общество» вправе не исполнять принятые на себя обязательства по настоящему договору.

- «Общество» имеет право на односторонний отказ от исполнения договора полностью или в части в связи с несогласованием «Предприятием» цен на услуги или неоплатой услуг, выполненных «Обществом», и иных платежей причитающихся «Обществу». Отказ от исполнения договора (полностью или частично) по инициативе «Общества» происходит путём направления «Предприятию» письменного уведомления. Договор считается соответственно расторгнутым или измененным по истечении 15 дней с момента получения уведомления, если иное не указано в уведомлении.
- По всем вопросам, не предусмотренным настоящим договором, стороны руководствуются Федеральными законами «Уставом железнодорожного транспорта Российской Федерации», «О железнодорожном транспорте в Российской Федерации», Правилами перевозок грузов железнодорожным транспортом, Правилами эксплуатации и обслуживания железнодорожных путей необщего пользования и другими нормативными актами Российской Федерации.
- В случае издания правомочными органами нормативных актов, изменяющих условия транспортного обслуживания, тарифы на перевозку грузов, размеры плат и сборов, порядок их оплаты и т.п., стороны руководствуются названными нормативными актами без внесения дополнений и изменений в текст договора.
- Стороны обязуются информировать друг друга в письменной форме об изменении существенных фактов деятельности организации: изменение адресов, платежных реквизитов, реорганизации, возбуждении процедуры банкротства и т п. в течение 10
- Неотъемлемой частью договора являются: Приложение № 1 «Протокол согласования договорных тарифов на оказание услуг АО «Уралпромжелдортранс» для филиала АО «Концерн Росэнергоатом» «Белоярская атомная станция», Приложение № 2 «Акт обследования грузовых фронтов».
- Настоящий Договор составлен в двух экземплярах, имеющих одинаковую юридическую силу: один экземпляр для «Предприятия», один экземпляр для «Общества».
- Период оказания услуг по настоящему договору 36 месяцев с даты подписания договора, в части оплаты - до полного исполнения обязательств.
- 7.11. В случае расторжения договора по решению суда в связи с существенным нарушением «Обществом» условий договора, информация об «Обществе» заносится в реестр недобросовестных поставщиков, предусмотренный Федеральным законом от 18 июля 2011г. № 223-ФЗ «О закупках товаров, работ, услуг отдельными видами юридических лиц» сроком на 2 (два) года».
- 7.12. «Стороны» обязаны обеспечить конфиденциальность сведений, касающихся предмета договора, хода его исполнения и полученных результатов.
- «Общество» уведомлено, что в случае нарушения условий настоящего договора в информационную систему «Расчет рейтинга деловой репутации поставщиков», ведение которой осуществляется на официальном сайте по закупкам атомной отрасли www.rdr.rosatom.ru в соответствии с утвержденными Госкорпорацией «Росатом» Едиными отраслевыми методическими указаниями по оценке деловой репутации, могут быть внесены сведения и документы о таких нарушениях. Основанием для внесения сведений в информационную систему «Расчет рейтинга деловой репутации поставшиков» могут являться:
- выставленные «Предприятием» и принятые «Обществом» неустойки за нарушение



# Результаты

Набор	Изображений
NPA-test – тестовый набор реальных данных	278
PLN — набор научных статей PupLayNet	125000
<b>GEN</b> – набор сгенерированных данных	18000
NPA-small – набор реальных данных	100
NPA-big — набор реальных данных	500

Количество изображений в разных наборах данных

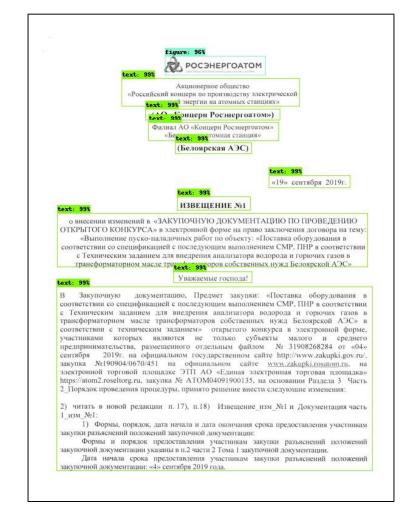
# Результаты

	Текст	Таблица	Картинка
GEN	104514	12413	0
NPA-small	474	53	31
NPA-big	2146	196	96
NPA test	1379	74	6

Количество объектов каждого класса в разных наборах данных

# Результаты сегментации (реальные данные)







# Результаты

	Текст	Таблица	Картинка	Итого
PLN+GEN+NPA-big	0.810	0.937	0.696	0.820
PLN+GEN+NPA-small	0.502	0.824	0.336	0.559
PLN+NPA-small	0.489	0.846	0.346	0.565
PLN	0.045	0.065	0.004	0.039

Результаты сегментации ( $F_1$ -мера)

### Результаты

	Текст	Таблица	Картинка	Итого
PLN+GEN+NPA-big	0.810	0.937	0.696	0.820
PLN+GEN+NPA-big+Post	0.840	0.968	0.755	0.855
PLN+GEN+NPA-small	0.502	0.824	0.336	0.559
PLN+GEN+NPA-small+Post	0.652	0.888	0.448	0.663

Результаты сегментации ( $F_1$ -мера) с постобработкой

### Заключение

Выбрана модель сегментации и проведена ее тонкая настройка в условиях ограниченного объёма реальных данных.

Сгенерировали искусственные данные

Дообучили модель на искусственных и малом количестве реальных данных

Удалось существенно повысить качество сегментации