

**УТВЕРЖДЕНА**

Приказом Ректора АНО ВО  
«Центральный университет»  
Ивашкевич Е.В.  
от «19» января 2024 г. № 0119.37

**Рабочая программа дисциплины (модуля)  
«Computer vision (Компьютерное зрение)»  
дополнительной профессиональной программы – программы  
профессиональной переподготовки «Академия data science»**

**Траектория: Backend-разработка**

**Москва  
2024**

## Содержание

<b>1. Краткая характеристика дисциплины (модуля) .....</b>	<b>3</b>
<b>2. Тематический план .....</b>	<b>4</b>
<b>3. Содержание дисциплины (модуля) .....</b>	<b>4</b>
<b>4. Учебно-методическое обеспечение .....</b>	<b>5</b>
<b>5. Материально-техническое обеспечение .....</b>	<b>5</b>
<b>6. Методические и оценочные материалы .....</b>	<b>7</b>

## 1. Краткая характеристика дисциплины (модуля)

Изучение дисциплины (модуля) «Computer vision (Компьютерное зрение)» позволяет понять, как машины могут интерпретировать и анализировать визуальную информацию, что открывает новые возможности в таких областях, как автономные транспортные средства, медицинская диагностика и безопасность. Эта дисциплина (модуль) сочетает в себе элементы машинного обучения, обработки изображений и искусственного интеллекта, что делает её ключевой для развития современных технологий и инновационных решений.

**Цель изучения дисциплины (модуля):** освоение методов и алгоритмов для автоматического анализа, обработки и понимания визуальной информации с помощью компьютерных систем.

### **Задачи изучения дисциплины (модуля):**

— формирование знаний по темам: принципы получения и хранения изображения в компьютере, классические необучаемые методы обработки изображений, архитектуры бэкбонов нейронных сетей для задач классификации, сегментации, детекции, адаптация трансформерных архитектур для области CV, устройство и принципы работы архитектур нейросетей для задач сегментации, pose estimation и распознавания действий, актуальные подходы в задаче детектирования объектов: two-stage/anchor-based, anchor-free;

— приобретение знаний и развитие понимания по темам: устройство и принципы работы фундаментальных моделей в CV, таких как CLIP, SAM, Grounding DINO, подходы для обучения моделей эффективного извлечения признаков, в частности для задач FaceRec/ReID/Retrieval, основные методы для трекинга одного/множества объектов, пайплайн оптического распознавания символов, современные методы обучения сетей без учителя, подходы к генерации изображений, в том числе генеративные и диффузионные модели;

— освоение умений применять классические методы обработки изображений, обучать нейронные сети для задач классификации/ детекции/ сегментации/ нахождения ключевых точек/распознавания действий, решать задачи эффективно за счет построения лучшего пайплайна обучения и использования современных библиотек, применять фундаментальные модели для решения задач, решать задачи не только на статичных кадрах, но и на видеопотоке;

— формирование навыков проведения и оценки экспериментов по обучению сетей сообразных поставленной задаче, декомпозирования сложносоставных задач на более мелкие и создания пайплайна, решающего целевую задачу.

## 2. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		<i>Очная форма</i>				
		Аудиторная работа		Контроль	Самостояте льная работа	
Лекции	Семинары (практичес кие занятия)					
1	Базовые задачи и методы CV	16	17		42	Домашние задания, Тесты
2	Продвинутые задачи и методы CV	17	17		42	Домашние задания, Тесты
	<i>Зачет с оценкой</i>			4		
	<b>Итого:</b>	<b>33</b>	<b>34</b>	<b>4</b>	<b>94</b>	
	<b>Объем дисциплины (модуля) (в ак. ч.)</b>	<b>165</b>				

## 3. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Базовые задачи и методы CV	Изображение. Классические методы в CV. Современные архитектуры CNN. Построение эффективного пайплайна обучения сети. Vision Transformers, CLIP. Сегментация изображений, SAM. Современные детекторы объектов. Key-Point/Pose Estimation.
2	Продвинутые задачи и методы CV	Face Recognition/Re-Identification/Image Retrieval. Генерация. Multi-Object Tracking (MOT). Optical Character Recognition (OCR). SOTA Self-Supervised Learning. Ускорение и инференс моделей. Vision Language Models (VLMs). Action Recognition.

#### 4. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый слушатель в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Слушателям обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

##### *Основная литература:*

1. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. — СПб.: Питер, 2018. — 480 с.: ил. — (Серия «Библиотека программиста»). — ISBN 978-5-496-02536-2.

2. Richard Szeliski. Computer Vision: Algorithms and Applications. – Springer Science & Business Media, 2010.

##### *Дополнительная литература:*

1. Траск Э. Грокаем глубокое обучение. — СПб.: Питер, 2019. — 352 с.: ил. — (Серия «Библиотека программиста»). — ISBN 978-5-4461-1334-7.

2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил. – ISBN 978-5-97060-618-6.

#### 5. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том

числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	<a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a>
2.	База данных для IT-специалистов	<a href="https://habr.com">https://habr.com</a>
3.	База данных ScienceDirect	<a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	<a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>
5.	Федеральный портал «Российское образование»	<a href="https://www.edu.ru/">https://www.edu.ru/</a>
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	<a href="http://window.edu.ru/">http://window.edu.ru/</a>
7.	Единая коллекция цифровых образовательных ресурсов	<a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a>
8.	Федеральный центр информационно - образовательных ресурсов	<a href="http://fcior.edu.ru/">http://fcior.edu.ru/</a>

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
<b>Операционные системы:</b>		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
<b>Браузеры:</b>		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
<b>Офисные приложения:</b>		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
<b>Программное обеспечение для планирования и учета времени:</b>		
Toggle app	зарубежное	свободно распространяемое
<b>Системы управления проектами:</b>		
Microsoft Imagine (Project)	зарубежное	лицензионное
<b>Системы управления базами данных:</b>		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
<b>Системы резервного копирования (backup):</b>		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
<b>Справочно-правовые системы:</b>		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
<b>Средства антивирусной защиты:</b>		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
<b>Среды разработки:</b>		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое

Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
<b>Пакеты программных средств и библиотек:</b>		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
<b>Системы управления библиографической информацией:</b>		
Zotero	зарубежное	свободно распространяемое
<b>Сервисы и службы:</b>		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

## 6. Методические и оценочные материалы

### Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Computer vision (Компьютерное зрение)» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, практические занятия, домашние задания, тесты, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

*Лекция* – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

*Участие в семинаре (практическом занятии)* – активная работа слушателя на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре слушателям рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

*Домашнее задание* – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

*Тест* – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины (модуля).

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины (модуля), понимать смысл научных категорий и уметь их использовать в профессиональной

лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания слушателями важных методологических категорий.

*Самостоятельная работа* – работа слушателей, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы слушатели взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи слушателя включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

### **Система оценивания результатов обучения по дисциплине (модулю)**

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

**Промежуточная аттестация** по дисциплине (модулю) осуществляется в форме **зачета с оценкой**.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Слушатель полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	Зачтено	
8	Отлично	Зачтено	
7	Хорошо	Зачтено	Слушатель обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Слушатель обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Слушатель способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Слушатель не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Computer vision (Компьютерное зрение)» оценивается следующим образом:

Активность	Вес	Количество	Описание
<b>Накопительная оценка</b>			
Домашние задания	60%	5	Набор задач по темам недели
Тесты		5	Письменная работа с набором задач, которые нужно решить за ограниченное время
<b>Промежуточная аттестация</b>			
Зачет с оценкой	40%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

**Формула расчёта итоговой оценки по дисциплине (модулю) «Computer vision (Компьютерное зрение)»:** « $0,6 \times$  накопительная оценка ( $0,85 \times$  среднее за домашние задания +  $0,15 \times$  среднее за тесты) +  $0,4 \times$  зачет с оценкой».

## Текущий контроль успеваемости обучающихся по дисциплине (модулю)

### Примерные домашние задания

#### Домашнее задание 1.

##### Задание 1.

1. Объясните, что такое цифровое изображение, и опишите основные параметры, характеризующие изображение (разрешение, глубина цвета, цветовое пространство).
2. Реализуйте на Python классический метод обнаружения границ на изображении (например, оператор Собеля или Кэнни) с использованием библиотеки OpenCV.
3. Проведите эксперимент: примените выбранный метод к нескольким изображениям с разной степенью шумов и прокомментируйте полученные результаты.

##### Задание 2.

1. Изучите архитектуру ResNet и опишите ключевую идею остаточных блоков (residual blocks) и их влияние на обучение глубоких сетей.
2. Постройте пайплайн обучения простой сверточной нейросети для задачи классификации изображений (например, CIFAR-10) с использованием PyTorch или TensorFlow. Включите этапы подготовки данных, настройки оптимизатора, функции потерь и метрик.
3. Проведите эксперимент с различными гиперпараметрами (например, размером батча, скоростью обучения) и опишите, как они влияют на обучение и качество модели.

##### Задание 3.

1. Опишите принцип работы Vision Transformer (ViT) и объясните, как происходит разбиение изображения на патчи и их обработка.
2. Ознакомьтесь с моделью CLIP (Contrastive Language-Image Pre-training) и кратко изложите, как она объединяет визуальные и текстовые представления.
3. Реализуйте на практике простой пример использования предобученной модели CLIP для поиска изображений по текстовому запросу с помощью библиотеки OpenAI CLIP или аналогичной.

#### Домашнее задание 2.

##### Задание 1.

1. Изучите основные методы сегментации изображений, такие как сегментация на основе порогов, метод активных контуров (snakes) и сегментация с использованием нейронных сетей (например, U-Net). Напишите краткое описание каждого метода.
2. Реализуйте алгоритм сегментации изображения с использованием модели SAM (Segment Anything Model) на наборе данных (например, COCO или ваш собственный набор изображений). Оцените качество сегментации с помощью метрик, таких как IoU (Intersection over Union).
3. Проведите эксперимент, изменяя параметры модели SAM, и проанализируйте, как это влияет на качество сегментации.

##### Задание 2.

1. Изучите архитектуры современных детекторов объектов, такие как YOLO, SSD и Faster R-CNN. Напишите краткое сравнение их особенностей и областей применения.
2. Реализуйте детектор объектов (например, YOLOv5) на наборе данных (например, Pascal VOC или COCO). Оцените производительность модели с помощью метрик, таких как mAP (mean Average Precision).

3. Изучите методы оценки позы (Key-Point Estimation), такие как OpenPose или PoseNet. Реализуйте один из этих методов на наборе данных и визуализируйте результаты, отметив ключевые точки на изображениях.

### **Задание 3.**

1. Изучите методы распознавания лиц (Face Recognition) и переидентификации (Re-Identification). Напишите краткое описание подходов, таких как FaceNet и DeepFace, а также их применения.

2. Реализуйте систему распознавания лиц с использованием предобученной модели (например, FaceNet) на наборе изображений. Оцените точность распознавания на тестовом наборе данных.

3. Изучите методы многократного отслеживания объектов (MOT), такие как SORT (Simple Online and Realtime Tracking) и Deep SORT. Реализуйте один из методов на видео с несколькими движущимися объектами и визуализируйте результаты отслеживания.

## **Домашнее задание 3.**

### **Задание 1.**

1. Изучите основные этапы OCR: предобработка изображения, сегментация символов, распознавание и постобработка. Опишите типичные проблемы OCR и методы их решения.

2. Реализуйте простую систему OCR на базе открытой библиотеки (например, Tesseract) и протестируйте её на наборе сканированных документов или фотографий с текстом.

3. Изучите методы распознавания действий (Action Recognition) в видео, например, 3D-CNN или модели на основе трансформеров. Опишите ключевые особенности и задачи, которые решают эти методы.

### **Задание 2.**

1. Ознакомьтесь с принципами самоконтролируемого обучения (Self-Supervised Learning) и опишите, как оно применяется в компьютерном зрении (например, методы contrastive learning: SimCLR, MoCo).

2. Реализуйте тренировку простой модели самоконтролируемого обучения на наборе изображений (например, CIFAR-10) с использованием одного из популярных методов.

3. Изучите архитектуру и применение Vision Language Models (VLMs), таких как CLIP или Flamingo. Опишите, как они связывают визуальные и текстовые данные.

### **Задание 3.**

1. Изучите методы оптимизации моделей для ускорения инференса: квантование, праунинг, knowledge distillation. Опишите принципы каждого метода.

2. Выберите предобученную модель (например, ResNet) и проведите эксперимент по её оптимизации с помощью одного из методов ускорения. Сравните скорость и точность до и после оптимизации.

3. Исследуйте библиотеки и инструменты для ускорения инференса (TensorRT, ONNX Runtime, OpenVINO) и опишите их основные возможности.

## **Примерные задания по тестам**

### **Тест 1.**

**Вопрос 1.**

Какова основная функция фотодиодов в камере?

- A. Увеличение яркости изображения
- B. Преобразование света в электрический сигнал
- C. Сохранение изображения на носителе
- D. Управление экспозицией камеры

Ответ: B.

**Вопрос 2.**

Что такое дебайеризация?

- A. Процесс преобразования черно-белого изображения в цветное
- B. Устранение цветных артефактов на изображениях
- C. Преобразование данных с сенсора в полноцветное изображение
- D. Процесс выравнивания изображений

Ответ: C.

**Вопрос 3.**

Каково основное назначение цветовой модели HSV?

- A. Отображение цвета в терминах яркости
- B. Создание черно-белых изображений
- C. Облегчение работы с цветами, учитывая восприятие человека
- D. Хранение изображения в трех других первичных цветах: циан, маджента, желтый

Ответ: D.

**Вопрос 4.**

Какой цвет соответствует комбинации всех трех первичных цветов в модели RGB?

- A. Черный
- B. Белый
- C. Серый
- D. Красный

Ответ: B.

**Вопрос 5.**

Что такое попиксельная операция?

- A. Операция, применяющаяся ко всему изображению целиком
- B. Операция, применяемая к отдельным пикселям изображения
- C. Процесс пересчета значения пикселя в новой цветовой модели
- D. Улучшение контраста изображения

Ответ: B.

**Тест 2.****Вопрос 1.**

Какого типа детектирования не существует?

- A. Two-stage
- B. Three-stage
- C. Anchor-based
- D. Anchor-free

Ответ: B.

**Вопрос 2.**

Какая из этих метрик/операций не используется при подсчете AP?

- A. IoU
- B. Precision
- C. Recall
- D. NMS

Ответ: D.

**Вопрос 3.**

Что такое Average Precision?

- A. Это площадь под Precision-Recall кривой
- B. Это Precision при score=0.5
- C. Это средний precision при скорях 0.5..0.95
- D. Это среднее геометрическое между Precision и Recall для всех возможных скоров детектора

Ответ: A.

**Вопрос 4.**

Как изменятся показатели метрик при увеличении  $th\_iou$  с 0.5 до 0.95

- A. AP уменьшится
- B. AP увеличится
- C. Максимальный Recall увеличится
- D. Максимальный Precision увеличится

Ответ: A.

**Вопрос 5.**

Какой из этих лоссов обычно не используется в задачах детектирования?

- A. GIoU Loss
- B. CIoU Loss
- C. BIoU Loss
- D. DIoU Loss

Ответ: C.

**Тест 3.**

**Вопрос 1.**

Выберите **ложные** утверждения о задаче OCR:

- A. Задача OCR предназначена для преобразования текстовой информации в изображения.
- B. OCR может работать как с печатным, так и с рукописным текстом.
- C. Scene Text Recognition - это подзадача OCR, связанная с распознаванием текста на документах
- D. Все датасеты OCR содержат расшифрованные тексты.
- E. Метрики OCR базируются на расстоянии Левенштейна.
- F. Расстояние Левенштейна можно использовать только для сравнения слов одинаковой длины.
- G. В задаче OCR обычно рассматриваются две метрики - Character Error Rate и Word Error Rate.

Ответ: A, C, D, F.

**Вопрос 2.**

Выберите **ложное** утверждение о пайплайне OCR:

- A. Извлечение ключевой информации зависит от бизнес-задачи.
- B. Выравнивание текста перед детекцией необходимо только при работе с печатными документами.
- C. Детекция текста — это процесс нахождения и локализации текстовых областей на изображении.
- D. Распознавание текста может использовать как CTC-based, так и Attention-based подходы.
- E. Для извлечения ключевой информации можно использовать графовые сети и языковые модели.
- F. Подходы локализации текста делятся на regression-based и segmentation-based.

Ответ: B.

**Вопрос 3.**

Электронный документ

Выберите верные утверждения о детекции текста:

- A. В OCR можно детектировать как отдельные символы, так и слова или строки.
- B. Регрессионные методы лучше подходят для текста произвольной формы.
- C. Сегментационные методы детекции текста определяют текстовые области, предсказывая маски для отдельных букв или слов.
- D. Регрессионные методы детекции текста используют детекторы объектов (anchor-based/anchor-free).
- E. Сложность детекции текста обусловлена только низким качеством изображений.
- F. Сегментационные методы требуют сложной постобработки для извлечения текстовых областей.

Ответ: A, C, D, F.

#### Вопрос 4.

Выберите **ложные** утверждения про DBNet:

- A. DBNet использует карту бинаризации для предсказанных хитмапов чтобы выделить текстовые области.
- B. DBNet — это объектный детектор, который использует anchor-based подход.
- C. Бинаризация в DBNet осуществляется с фиксированным порогом, заданным вручную.
- D. DBNet аппроксимирует бинаризацию с помощью непрерывной функции.
- E. В основе DBNet лежит архитектура FPN, которая помогает учитывать контекст на разных уровнях.
- F. DBNet лучше всего подходит для детекции текста на документах, но неэффективна для детекции текста “in the wild”.

Ответ: B, C, F.

#### Вопрос 5.

Выберите верные утверждение о модели CRNN:

- A. CRNN объединяет сверточные и рекуррентные нейросети для распознавания текста.
- B. CRNN использует объектный детектор для локализации текста перед распознаванием.
- C. CRNN использует механизм внимания (attention) при предсказании текста.
- D. Выходные векторы CNN в CRNN имеют фиксированную длину, которая соответствует количеству символов в тексте.
- E. CRNN использует сверточные слои для классификации символов.
- F. CRNN является end-to-end решением.
- G. CRNN предсказывает сразу готовую строку текста, для которой не требуется этапа декодирования.
- H. CRNN предсказывает дублирующиеся символы из-за того, что receptive field эмбедингов накладываются друг на друга.

Ответ: A, H.

### Примерный перечень тем к зачету с оценкой

1. Как устроена камера. Дебайеризация. Цветовые модели. Попиксельные операции. Выравнивание гистограммы. Интерполяция. Проблема алиасинга.
2. Бинаризация. Морфологические операции. Градиент для изображения. Оператор Собеля. Фильтр Кэнни. Контуры и примитивы. Гомография.
3. Постановка задачи классификации (one-label, multi-label). ImageNet, Метрики. Операция свёртки, её параметры и разновидности: 2D Conv, DWS Conv, Grouped Conv. Первые CNN-ки (LeNet, AlexNet, VGG).
4. Семейство ResNet и его конструкция (ResNet, ResNeXt). Residual Block, Bottleneck Block, Grouped Conv.
5. MobileNet v1-v3, NAS, MBConv, SE block.

6. EfficientNet v1, v2, NAS, MBConv, Fused-MBConv, особенности обучения.
7. Гиперпараметры обучения: инициализация, претрейны, Loss, Optimizer, Batch, Epoch, LR, LR Scheduling. Техники семплирования данных, mixed precision.
8. Техники регуляризации: Weight Decay, Label Smoothing, EMA, Random Erasing, GridMask, Миксинг (CutMix, MixUp). Пайплайны аугментаций: Auto, Random, AugMix, Trivial. Test Time Augmentations. Progressive training.
9. Vision Transformer (ViT): устройство, основные блоки. Hybrid ViT. Особенности обучения и сравнение с CNN.
10. Постановка задачи детектирования. Loss-ы. Метрики: AP, Precision/Recall, IoU. Two-stage детекторы: R-CNN family.
11. Anchor-based детекторы: YOLO, SSD. Пирамиды признаков, FPN, RetinaNet.
12. Anchor-free детекторы: CenterNet, FCOS, DETR
13. Постановка задачи сегментации. Метрики и лоссы. UNet
14. Архитектуры для сегментации: LinkNet, UNet++, DeepLabV3, FPN, HRNet
15. Segment Anything Model, Segment Anything Model 2.
16. Постановка задач Key Point Detection и Pose Estimation. Метрики. Top-down и bottom-up подходы. DeepPose.
17. CPMs, Stacked Hourglass, HRNet, RSN.
18. ViTPose. OpenPose.
19. Системы распознавания лиц. Open-set и Closed-set задачи распознавания. Верификация и идентификация. Связь с ReID, Image Retrieval. Метрики для задачи распознавания лиц. Knowledge Distillation.
20. Metric Learning Losses: Contrastive Loss, Triplet Loss, Softmax Loss, Softmax margin-based losses, ArcFace.
21. Постановка задачи MOT. Метрики MOT: MOTA, MOTP, IDF1, HOTA, IDsw. SOT: постановка задачи, GOTURN. MCMOT: постановка задачи, идеи реализации.
22. SORT, Фильтр Калмана.
23. DeepSORT, FairMOT, ByteTrack, BOT-SORT.
24. SSL: постановка задачи, оценка качества. Классические pretext tasks. SimCLR, BYOL.
25. DINO, MAE, DINO v2.
26. VLM, Foundation Model - определения. VLM: подходы к обучению, ключевые аспекты обучения, метрики. CLIP. Zero-shot classification с CLIP. Florence.
27. BLIP, BLIP-2, Frozen.
28. LLaVa, Florence-2.
29. Grounded Models: GLIP, Grounding DINO, Grounded SAM, etc
30. OCR: постановка задачи, датасеты, метрики. Пайплайн OCR. Text Detection: CRAFT, DBNet.
31. Text Recognition: CRNN, CTC, SAR.
32. GAN: vanilla, cGAN, pix2pix, cycleGAN, StyleGAN
33. Diffusion: forward and backward processes. Latent Diffusion, Stable Diffusion.