

Аннотация рабочей программы дисциплины

Б1.В.06 «Современные методы компьютерного зрения»

Направление

подготовки/специальность 02.03.02 Фундаментальная информатика и информационные технологии

(код и наименование направления подготовки/специальности)

Курс 3 Семестр 6 Количество з.е. 4

Объем трудоемкости: 4 зачетных единиц (144 ч., из них – 89,3 час. аудиторной нагрузки: лекционных 32 ч., лабораторных работ - 48 ч., 19 часов самостоятельной работы, 4 часов КСР, 0,3 часа ИКР, курсовой проект – 5 часов, подготовка к экзамену 35,7), форма контроля – экзамен.

Цель дисциплины: Сформировать у студентов систематизированное представление о современных архитектурах нейронных сетей для компьютерного зрения и выработать практические навыки их применения для решения реальных задач анализа изображений и видео.

Задачи дисциплины:

1. Изучить архитектуры нейронных сетей для компьютерного зрения: от классических методов (Виолы-Джонс) до современных сверточных сетей, трансформеров и генеративных моделей.
2. Освоить математические основы работы нейросетевых моделей и методы их оптимизации и регуляризации.
3. Сформировать практические навыки реализации и обучения моделей для детекции, сегментации, генерации и мультимодального анализа изображений.
4. Научить работать с инструментами и библиотеками для обработки данных, обучения моделей и оценки их качества.
5. Развить умение адаптировать и применять модели для решения прикладных задач с учетом ограничений.

Место дисциплины в структуре ООП ВО:

Дисциплина «Современные методы компьютерного зрения» относится к «Часть, формируемая участниками образовательных отношений» Блока 1 «Дисциплины (модули)» учебного плана.

Дисциплина использует знания и умения студентов, полученные в рамках изучения дисциплин: Технологии компьютерного зрения, Современные методы машинного обучения, Нейросетевые технологии, Интеллектуальные методы оптимизации.

Результаты освоения дисциплины будут использоваться при прохождении следующих дисциплин/практик: технологическая (проектно-технологическая) практика, Генеративные нейронные сети, Системы искусственного интеллекта, Проектная работа.

Результаты обучения (знания, умения, опыт, компетенции):

Содержание и структура дисциплины:

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

DL-1.3; DL-1.4; DL-1.9; DL-1.10; DL-1.12; DL-3.1; DL-3.2; DL-3.3; DL-3.4; PL-1.1; FC-1.2

DL-1 Способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей

DL-1.3 Применяет современные архитектуры глубоких сетей для решения различных задач, понимая их внутреннюю структуру и особенности обучения

Применяет основные архитектуры глубокого обучения (VGG, ResNet)

Применяет принцип построения вычислительного блока Google Inception; Применяет принцип работы блока остатка в ResNet; Разрабатывает решения с применением backbone сетей; Знает отличия и способен применять нейронные сети для отслеживания объектов (семейство R-CNN, YOLO)

Регулирует поток вычисления градиента в глубоких нейронных сетях

DL-1.4 Разрабатывает и оптимизирует специализированные архитектуры для работы с изображениями, учитывая их уникальные свойства

Выбирает значения гиперпараметров нейронов свертки и пулинга (размер ядра свертки, шаг и отступ); применяет готовые архитектуры сверточных нейронных сетей;

Применяет принцип работы обратного распространения ошибки на слоях свертки и слоях пулинга; разрабатывает собственные архитектуры сверточных нейронных сетей; разрабатывает решения на основе сложных конфигураций сверточных нейронных сетей (EfficientNet, RetinaNet); понимает принцип работы блока единичной задержки в рекуррентных нейронных сетях.

Создает принципиально новые, эффективные архитектурные решения для CNN (новые типы слоев, схемы соединений, механизмы взаимодействия между признаками), основанные на глубоком понимании теории CNN и свойств данных

DL-1.9 Разрабатывает, адаптирует и внедряет трансформерные архитектуры для решения задач обработки последовательностей, включая создание новых моделей, оптимизацию обучения и промышленное развертывание

Понимает принципы работы трансформеров (самовнимание, позиционное кодирование). Умеет применять готовые модели (BERT, GPT, ViT) через Hugging Face/PyTorch. Настраивает базовые параметры обучения (скорость обучения, размер пакета)

Понимает и применяет специфические архитектурные варианты для задач (например, вариант "энкодер-декодер" для перевода/суммаризации, вариант "только энкодер" для генерации); использует для работы с длинными последовательностями эффективные схемы внимания (Linformer, Performer, Longformer, Sparse Transformers). Проводит детальный анализ ошибок модели, понимает типичные проблемы (смещение в данных/модели, переобучение) и применяет методы борьбы с ними (регуляризация, аугментация данных)

Разрабатывает разреженные трансформеры. Понимает принцип работы многоголовочного внимания

DL-1.10 Проектирует, разрабатывает и внедряет мультимодальные модели глубокого обучения, эффективно комбинируя различные типы данных.

Понимает принципы мультимодального обучения (раннее/позднее слияние, кросс-модальное внимание). Работает с готовыми мультимодальными моделями (CLIP, Flamingo, BEiT-3). Умеет подготавливать и согласовывать разнородные данные (текст+изображения)

Разрабатывает кастомные архитектуры для специфических комбинаций модальностей. Оптимизирует многозадачное обучение. Применяет продвинутое техники (кросс-модальная дистилляция, контрастивное обучение). Создает эффективные пайплайны обработки мультимодальных данных

Разрабатывает универсальные архитектуры для произвольных комбинаций модальностей. Решает фундаментальные проблемы (дисбаланс модальностей, отсутствующие модальности). Строит масштабируемые системы для промышленного применения. Оптимизирует модели для работы в условиях ограниченных ресурсов

DL-1.12 Применяет, адаптирует и разрабатывает методы дообучения нейронных сетей для эффективной адаптации моделей к новым задачам и доменам.

Понимает базовые техники дообучения (перенос обучения, обучение по нескольким примерам). Умеет работать с популярными предобученными моделями (ResNet, BERT, GPT). Применяет стандартные подходы к адаптации последних слоев. Использует базовые стратегии заморозки слоев.

Владеет продвинутыми техниками (слои адаптера, LoRA, настройка префикса). Комбинирует различные стратегии адаптации. Работает с малыми датасетами (обучение по нескольким примерам)

Разрабатывает новые методы параметрически-эффективного обучения.

Создает универсальные фреймворки для адаптации моделей. Решает фундаментальные проблемы (катастрофическое забывание, доменный разрыв)

DL-3 Способен применять и (или) разрабатывать алгоритмы, методы и технологии компьютерного зрения

DL-3.1 Применяет (проводя выбор и эксперименты) известные алгоритмы и библиотеки компьютерного зрения, предобученные глубокие нейросетевые модели для прикладных задач анализа изображений и видеопотока, при необходимости дообучая и валидируя на собственных наборах данных

Использует готовые модели из библиотеки OpenCV и популярных фреймворков (TensorFlow Hub, TorchVision). Умеет применять стандартные архитектуры (ResNet, YOLO) для базовых задач (классификация, детекция объектов). Запускает инференс на изображениях и простых видеопотоках. Использует базовые функции библиотек OpenCV и PIL для обработки. Умеет сохранять и загружать модели..

Сравнивает разные предобученные модели под конкретную задачу. Проводит перенос обучения на своих данных. Оптимизирует гиперпараметры для улучшения качества. Создает сложные пайплайны аугментации (с использованием библиотеки Albumentations). Умеет работать с видео: извлечение кадров, обработка временных последовательностей путём применения CNN+RNN, 3D CNN..

Разрабатывает стратегии применения моделей компьютерного зрения под бизнес-задачи. Комбинирует несколько моделей в комплексные решения. Строит сквозные пайплайны обработки видеопотоков.

DL-3.2 Определяет стек технологий, методов и алгоритмов для построения продуктов с компьютерным зрением (системы видеоаналитики, поисковые системы по изображениям и т.д.)

Понимает принципы представления изображений и кодирования цвета; применяет фильтрацию изображений, включая частотную; Владеет аппаратом математической морфологии; Понимает базовые задачи анализа изображений - классификация, детекция,

сегментация. Применяет известные архитектуры нейронных сетей (CNN) в решении простых задач распознавания изображений

Разрабатывает алгоритмы сегментации изображений (разделение-слияние регионов, нормализованный разрез графа, mean shift), включая семантическую сегментацию; применяет преобразование Хафа и RANSAC; применяет алгоритмы детекции характеристических точек (детектор Харриса, детектор Фестнера, SUSAN, бобы, DoG); применяет дескрипторы изображений, например, SIFT Нейросетевые архитектуры для анализа изображений VGG, Inception, ResNet, EfficientNet и т.д. особенности обучения и дообучения. Архитектуры FCN и Unet в задачах сегментации, функции потерь для задачи сегментации. Одностадийные (SSD, YOLO) и двухстадийные (FASTER R-CNN, Mask R-CNN) детекторы в задачах детекции, функций потерь в задаче детекции.

Применяет марковские случайные поля и условные случайные поля Исследование и имплементация SOTA-моделей в задачах анализа изображений. Vision Transformer (ViT) - глубокое понимание подходов и методов дообучения. Metric Learning для задач поиска и распознавания на изображениях, функции потерь contrastive loss, triplet loss, arcface. Распознавание текста на изображении (CRNN, Attention OCR и Transformer OCR)

DL-3.3 Имплементирует известные алгоритмы, архитектуры и модели компьютерного зрения на реальных данных, строит пайплайны обучения моделей и развертывания сервисов компьютерного зрения в продуктивной среде

Умеет применять стандартные архитектуры CNN (ResNet, EfficientNet, YOLO) для базовых задач компьютерного зрения (классификация, детекция, сегментация). Использует готовые реализации из библиотек (OpenCV, TensorFlow/PyTorch, Hugging Face). Проводит базовую аугментацию и нормализацию изображений. Умеет оценивать качество по стандартным метрикам (точность, mAP, IoU).

Кастомизирует архитектуры под задачу (изменение слоев, замена опорной сети).

Применяет методы ускорения инференса (квантизация, прунинг, TensorRT). Строит сложные стратегии аугментации (augmentations, кастомные трансформеры). Настраивает распределённое обучение (DDP, Horovod). Создает пайплайны CI/CD для моделей компьютерного зрения.

Умеет разрабатывать продвинутое архитектуры в задачах анализа видеопотока. Разрабатывает высоконагруженные CV-сервисы.

Строит A/B-тестирование для CV-решений. Оптимизирует модели анализа видеопотока

DL-3.4 Разрабатывает новые алгоритмы и библиотеки компьютерного зрения, новые архитектуры глубоких нейронных сетей и методы их обучения для задач анализа изображений и видео

Умеет адаптировать существующие архитектуры (ResNet, ViT, ConvNext) под специфические задачи. Проводит исследования абляций для анализа компонентов моделей. Умеет реализовывать идеи из продвинутых статей (на уровне воспроизведения результатов)

Владеет аппаратом эвклидовой геометрии. Способен применять алгоритмы стереозрения. Способен применять алгоритмы фотограмметрии. Использует классические подходы GAN в задачах генерации изображений (перенос стиля, замена лиц, улучшение качества фотографий высокого разрешения, реконструкция фотографий)

Знает математические основы представления поверхностей.

Строит диффузионные модели для генерации изображений и видео. Имплементирует генеративные модели для преобразования текста в изображение.

PL-1 Способен применять язык программирования Python для решения задач в области ИИ

PL-1.1 Разрабатывает и отлаживает прикладные решения разной сложности и для разного круга конечных пользователей с использованием языка программирования Python, тестирует, испытывает и оценивает качество таких решений

Знает основы синтаксиса языка, пишет небольшие скрипты для автоматизации ручной работы по обработке небольших объемов данных с помощью встроенных модулей и внешних библиотек (csv, json, requests).

Владеет основными библиотеками для выполнения большинства рутинных задач в крупных проектах: ввод-вывод, серверное программирование (FastAPI, Flask, Django REST Framework), применение многопоточности (модуль threading). Самостоятельно участвует в разработке серверных приложений и их поддержке

Использует особенности виртуальной машины Python (например, GIL), разрабатывает библиотечный код общего пользования, а также документацию к нему. Профилирует и оптимизирует приложения на Python, используя встроенные инструменты (например, cPython).

FC-1 Способен проводить передовые исследования в области архитектур, алгоритмов МО, оптимизации и математики

FC-1.2 Разрабатывает новые архитектуры глубоких нейросетей

Знает основные соответствия в триаде: архитектура-данные-задача, способен по описанию данных и задачи подобрать архитектуру-бейзлайн. Активно пользуется алгоритмами автоматизации подбора архитектур.

Знает передовые архитектуры в основных триадах: архитектура-данные-задача, принципы их построения, сильные и слабые стороны. Знает особенности наиболее часто встречающихся вычислителей, умеет подбирать архитектуры, адекватные особенностям вычислительных устройств.

Способен по описанию создать архитектуры нейросетей, которые могут подстраиваться под конкретные задачи и типы данных с использованием минимального количества сторонних библиотек. Использует методы автоматизации машинного обучения для создания новых архитектур и приспособления архитектур под специализированные ПАКи (программно-аппаратные комплексы) и вычислительные устройства

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 6 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа СРС
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Фундаментальные архитектуры и методы глубокого обучения для зрения	20	8		8	4
2.	Основные задачи компьютерного зрения: детекция, сегментация, генерация	28	10		14	4
3.	Современные архитектуры: внимание, трансформеры и метрическое обучение	20	6		10	4
4.	Продвинутые и гибридные методы	31	8		16	7
ИТОГО по разделам дисциплины		99	32		48	19
Контроль самостоятельной работы (КСР)		4				

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
	Промежуточная аттестация (ИКР)	0,3				
	Курсовой проект	5				
	Подготовка к текущему контролю	35,7				
	Общая трудоемкость по дисциплине	144				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Курсовые проекты или работы.

Примерные темы курсовых проектов

1. Разработка системы детекции и трекинга транспортных средств на видео с городских камер
2. Создание системы семантической сегментации медицинских изображений для диагностики
3. Разработка генеративной модели для синтеза фотореалистичных изображений интерьеров
4. Система мультимодального поиска товаров по изображению и текстовому описанию
5. Детекция аномалий на производственных линиях с помощью видеоаналитики
6. Создание системы оценки позы человека в реальном времени для фитнес-трекера
7. Разработка 3D реконструкции объектов по нескольким 2D изображениям
8. Система распознавания эмоций по видео лица с учетом временной последовательности
9. Автоматическая колоризация черно-белых исторических фотографий
10. Детекция и распознавание дорожных знаков для систем ADAS
11. Создание системы подсчета людей в crowded-сценах
12. Разработка системы повторной идентификации человека (ReID) по видео с нескольких камер
13. Генерация синтетических данных для дообучения моделей компьютерного зрения
14. Система анализа спортивных матчей: детекция игроков и мяча, тактический анализ
15. 3D детекция объектов для автономного вождения по данным LiDAR
16. Мультимодальная система описания изображений для помощи слабовидящим
17. Детекция deepfake видео с использованием временных и пространственных признаков
18. Система автоматической разметки датасетов с активным обучением
19. Оптимизация нейросетевых моделей для embedded-устройств
20. Создание системы оценки качества продукции по изображениям с производственной линии

Вид аттестации: экзамен.

Автор А.С.Жук, доцент КВТ