

# Компьютерное зрение – CV

01	Кому подойдёт курс	<p>Курс подойдёт тем, у кого есть опыт работы с данными, Machine Learning и Deep Learning</p> <p>Специалистам в области Data Science</p> <ul style="list-style-type: none"><li>Научитесь строить модели на основе информации из изображений, освоите SOTA-архитектуры и сможете применять CV для решения бизнес-задач</li></ul> <p>DL- и ML-инженерам</p> <ul style="list-style-type: none"><li>Разберёте способы работы с визуальными данными, изучите продвинутые архитектуры и узнаете на практике, как применять CV в ваших проектах</li></ul> <p>Разработчикам</p> <ul style="list-style-type: none"><li>Поймёте принципы обработки визуальных данных, научитесь применять OpenCV и TensorFlow, сможете интегрировать CV в разработку</li></ul>
02	Чему научитесь на курсе	<ul style="list-style-type: none"><li>Создавать решения для сегментации и детекции объектов</li><li>Обучать и адаптировать генеративные модели под собственные задачи</li><li>Работать с мультимодальными данными и архитектурами на их основе</li><li>Разбираться в трансформерах для изображений — ViT</li><li>Оценивать и интерпретировать работу CV-моделей</li></ul>
03	Как проходит курс	<ul style="list-style-type: none"><li>Сопровождение кураторами</li><li>Обратная связь от опытных наставников</li><li>Воркшопы с экспертами</li><li>Теория на платформе Практикума</li><li>Практические задания с рецензией на готовой инфраструктуре в облаке</li></ul>

## Что вас ждёт на курсе

Детекция, сегментация, генерация изображений и мультимодальные модели

Архитектуры, loss-функции, метрики, принципы повышения качества

Удостоверение о повышении квалификации

# Компьютерное зрение – CV

**2 месяца**

продолжительность курса

**4 проекта**

в портфолио

2 ЧАСА

**00**

**Бесплатная часть.  
Нейросетевые  
решения на практике**

- Знакомство с курсом
- Введение в нейросети и их роль в решении сложных задач
- Работа с текстами
- Работа с изображениями
- Нейросети на практике

2 НЕДЕЛИ

**01**

**Детекция объектов**

- Введение и быстрый старт с YOLOv8
- YOLO-модель: подготовка данных и обучение
- Постобработка и визуализация результатов
- Альтернативные архитектуры: SSD и Faster R-CNN
- Решение задачи на альтернативных архитектурах

2 НЕДЕЛИ

**02**

**Сегментация  
изображений**

- Введение и быстрый старт с сегментацией
- Подготовка данных и обучение модели
- Постобработка и улучшение результатов
- Альтернативные архитектуры
- Решение задачи на альтернативных архитектурах

2 НЕДЕЛИ

**03**

**Генерация  
изображений**

- Введение в генерацию изображений
- Autoencoder и VAE
- GAN
- Диффузионные модели
- Файн-тюнинг и кастомизация diffusion-моделей

2 ЧАСА

**04**

**Трансформеры  
и мультимодальные  
модели**

- Введение в Vision Transformers
- Файн-тюнинг ViT и гибридные трансформеры
- Мультимодальные модели
- Текстово-визуальный поиск и описание изображений
- Визуальные вопросы и ответы

# Бесплатная часть. Нейросетевые решения на практике

00

2 часа

1 практическая работа

## Практическая работа

Научитесь использовать предобученные модели для анализа текста и изображений и интерпретировать результаты их работы

## Инструменты и технологии

- PyTorch
- CV
- NLP

## Содержание

01. Знакомство с курсом	Узнаете, как организован курс: из чего состоит программа, как устроена платформа, какие активности вас ждут
02. Введение в нейросети и их роль в решении сложных задач	Поймёте, в каких областях применяются нейросети и какие задачи они решают в реальной жизни
03. Работа с текстами	Классифицируете тексты с использованием предобученной модели
04. Работа с изображениями	Классифицируете изображения с помощью предобученной модели
05. Нейросети на практике	Научитесь использовать нейросети в повседневной практике и для решения профессиональных задач

2 недели

1 проект

## Проект

Создадите высокоточный детектор на кастомном наборе данных

## Инструменты и технологии

- PyTorch
- YOLO
- Faster R-CNN
- SSD
- TorchVision

## Содержание

01. Введение и быстрый старт с YOLOv8	Задача детекции объектов. Предсказания модели: координаты, класс, уверенность. Эволюция от YOLO до YOLOv8 (Backbone, Neck, Detection Head). Запуск предобученной YOLO-модели на изображениях. Интерпретация результатов: рамки, метки, уверенность.
02. YOLO-модель: подготовка данных и обучение	Форматы данных и их конвертация. Аугментация данных и её влияние на разметку. Loss-функции и метрики для детекции. Разметка данных, конфигурации обучения и управление им по логам от модели.
03. Постобработка и визуализация результатов	Фильтрация по порогу уверенности. Non-Maximum Suppression (NMS). Визуализация предсказаний (Matplotlib/OpenCV). Написание визуализатора результатов. Вывод top-N-предсказаний. Визуальное сравнение с ground truth.
04. Альтернативные архитектуры: SSD и Faster R-CNN	SSD. Faster R-CNN: двухэтапная архитектура. Ключевые отличия от YOLO: области применения, время инференса, сложность настройки и обучения. Различия в loss-функциях.
05. Решение задачи на альтернативных архитектурах	Использование SSD или Faster R-CNN. Подготовка данных. Обучение модели на том же наборе данных. Сравнение результатов: по метрикам (mAP, IoU), по скорости, по визуальному качеству.

2 недели

1 проект

## Проект

Разработаете модели сегментации для решения бизнес-задачи

## Инструменты и технологии

- PyTorch
- U-Net
- DeepLab
- TorchVision
- MM Segmentation

## Содержание

01. Введение и быстрый старт с сегментацией	Задача семантической сегментации: отличие от детекции. Карта классов (mask), формат выхода. Быстрый запуск предобученной модели. Визуализация результатов и интерпретация масок. Запуск инференса на изображениях. Отображение маски и соответствующих цветов для классов.
02. Подготовка данных и обучение модели	Разметка для задачи сегментации. Форматы масок: индексные, PNG, One Hot Encoding. Связь масок и аугментации. Выбор loss-функции и метрик для модели. Разбор популярных архитектур и конфигурация обучения для них.
03. Постобработка и улучшение результатов	Softmax/Argmax и получение финальной маски. Morphological operations: opening, closing. CRF как опция. Визуальное сравнение до и после. Постобработка результата модели и эксперименты. Сравнение метрик до и после.
04. Альтернативные архитектуры	Обзор Encoder-Decoder-архитектур. U-Net. DeepLabv3+: dilated convolutions, ASPP. SegFormer на основе трансформеров. Ключевые различия: строение, подход к масштабам, скорость и качество, области применения. Сравнение U-Net, DeepLab, SegFormer.
05. Решение задачи на альтернативных архитектурах	Обучение U-Net, DeepLabv3, SegFormer. Сравнение: по точности, скорости, качеству масок. Обучение альтернативной модели. Визуальное сравнение и по метрикам. Применимость моделей.

2 недели

1 проект

## Проект

Кастомизируете генеративную сеть под конкретный запрос

## Инструменты и технологии

- PyTorch
- GAN
- VAE
- Diffusion

## Содержание

01. Введение в генерацию изображений	Зачем нужна генерация изображений. Типы генеративных моделей: Autoencoders, GANs, Diffusion Models. Подходы к обучению и генерации. Предобученный VAE и визуализация латентного пространства.
02. Autoencoder и VAE	Архитектура: Encoder + Decoder. Ограничения обычных AE (неявное моделирование вероятностей). Variational Autoencoder: обучение распределения + sampling. Обучение VAE, генерация и интерполяция, латентное пространство в 2D.
03. GAN	Архитектура: Generator + Discriminator. Типичные проблемы: mode collapse, нестабильность. Варианты: DCGAN, Conditional GAN, StyleGAN. Обучение DCGAN на простом датасете, визуализация результатов по эпохам.
04. Диффузионные модели	Denoising Diffusion Probabilistic Models, Stable Diffusion, Imagen, DALL·E. Преимущества и ограничения diffusion-архитектур. Использование diffusers от Hugging Face. Генерация по текстовому промпту. Замена параметров: guidance scale, steps, seed.
05. Файн-тюнинг и кастомизация diffusion-моделей	Принципы дообучения: Transfer Learning. Технология LoRA. Дообучение Stable Diffusion, сравнение оригинальных и кастомных промптов, визуализация влияния кастомного обучения.

# Трансформеры и мультимодальные модели 04

2 недели

1 проект

## Проект

Обучите классификационную модель на трансформерной архитектуре

## Инструменты и технологии

- PyTorch
- ViT
- CLIP
- VQA

## Содержание

01. Введение в Vision Transformers	Ограничения классических CNN. Архитектура ViT: патчи, позиционные эмбеддинги, self-attention. Преимущества ViT и примеры использования. Загрузка и применение предобученного ViT. Классификация изображений. Визуализация attention maps.
02. Файн-тюнинг ViT и гибридные трансформеры	Обучение и файн-тюнинг Vision Transformer. Гибридные модели: архитектурные особенности и преимущества. Сравнение ViT, гибридных и CNN. Советы по обучению и выбору моделей. Эксперименты.
03. Мультимодальные модели	Модели Vision + Language. Задачи: image captioning, VQA, image-text retrieval. Архитектуры Encoder-Decoder, Dual Encoder, Fusion. Обзор датасетов: COCO Captions, VQA, Flickr30k. Модели CLIP для простых визуально-текстовых задач.
04. Текстово-визуальный поиск и описание изображений	Архитектура моделей для генерации описания: BLIP, GIT. Метрики качества captioning. Модель CLIP для текстово-визуального поиска: принцип работы и применение. Генерация описаний и поиск изображений по тексту с помощью BLIP или GIT.
05. Визуальные вопросы и ответы	Архитектуры для VQA: multi-modal fusion, cross-attention. Задача grounding: локализация объектов по тексту. Обзор моделей: OFA, LXMERT. Flamingo (DeepMind). Использование LXMERT или OFA для VQA-задач.