

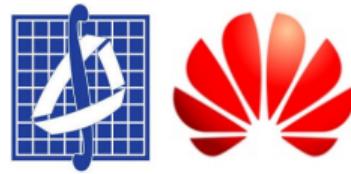
Введение в искусственный интеллект. Современное компьютерное зрение

Лекция 1. Вводная

Бабин Д.Н., Иванов И.Е., Петюшко А.А.

кафедра Математической Теории Интеллектуальных Систем

01 октября 2019 г.



План лекции

- ① Организационные вопросы
- ② Историческая справка
- ③ Нейросетевой подход
- ④ Постановка основных задач компьютерного зрения
- ⑤ Примеры прикладных задач



Авторы курса



Руководитель курса: д.ф.-м.н. Бабин Дмитрий Николаевич



Лектор: к.ф.-м.н. Иванов Илья Евгеньевич



Лектор: к.ф.-м.н. Петюшко Александр Александрович

Сотрудничество

- Авторы имеют более 15 лет опыта участия в проектах, связанных с машинным обучением и компьютерным зрением
- Являются постоянными участниками группы распознавания образов кафедры МаТИС
- В качестве научных консультантов работают или работали с такими крупнейшими российскими и международными компаниями как Нейроком, LSI Research, Fotonation, Huawei и др.



- В данный момент времени авторы ведут исследования в области компьютерного зрения в московском научно-исследовательском центре Хуавэй
- Данный курс является частью программы **SHARE**
 - **SHARE** = School of Huawei Advanced Research Education
 - **SHARE** = Школа опережающего научного образования Хуавэй



Зачем посещать этот курс

- 1 Компьютерное зрение — самое бурно развивающееся направление в области искусственного интеллекта



Зачем посещать этот курс

- ➊ Компьютерное зрение — самое бурно развивающееся направление в области искусственного интеллекта
- ➋ Возможность получить state-of-the-art знания, которые пригодятся в работе



Зачем посещать этот курс

- ➊ Компьютерное зрение — самое бурно развивающееся направление в области искусственного интеллекта
- ➋ Возможность получить state-of-the-art знания, которые пригодятся в работе
- ➌ Шанс максимально использовать своё образование



Зачем посещать этот курс

- ① Компьютерное зрение — самое бурно развивающееся направление в области искусственного интеллекта
- ② Возможность получить state-of-the-art знания, которые пригодятся в работе
- ③ Шанс максимально использовать своё образование
- ④ Для лучших студентов возможны стажировки и бонусы от партнеров



Что же такое искусственный интеллект?

Естественный интеллект (человек)

- Может воспринимать информацию, ее анализировать, принимать решения на основе анализа



Что же такое искусственный интеллект?

Естественный интеллект (человек)

- Может воспринимать информацию, ее анализировать, принимать решения на основе анализа

Искусственный интеллект

- (Сильный) то же самое, что и естественный, только на месте человека — компьютер



Что же такое искусственный интеллект?

Естественный интеллект (человек)

- Может воспринимать информацию, ее анализировать, принимать решения на основе анализа

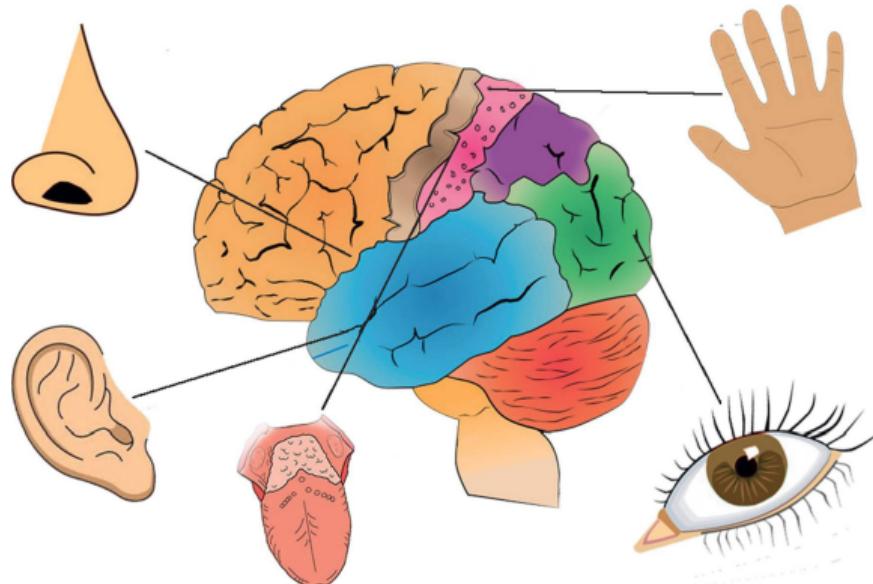
Искусственный интеллект

- (**Сильный**) то же самое, что и естественный, только на месте человека — компьютер
- (**Слабый**) алгоритм, способный обучиться на основе массива входных данных, чтобы затем выполнять задачу вместо человека



Взаимодействие со средой

- Около 90 % информации поступает через **зрение**¹
- Около 9 % информации поступает через слух



¹https://www.rlsnet.ru/books_book_id_2_page_40.htm

Общая структура курса «Введение в компьютерный интеллект»

① Машинное обучение²

- Необходимые основы для всего курса

²<https://github.com/mlcourse-mm/mlcourse-mm2019spring>

Общая структура курса

«Введение в компьютерный интеллект»

- ① Машинное обучение²
 - Необходимые основы для всего курса
- ② Компьютерное зрение
 - Извлечение информации из визуальных образов (изображений и видео)

²<https://github.com/mlcourse-mm/mlcourse-mm2019spring>



Общая структура курса

«Введение в компьютерный интеллект»

- ① **Машинное обучение²**
 - Необходимые основы для всего курса
- ② **Компьютерное зрение**
 - Извлечение информации из визуальных образов (изображений и видео)
- ③ **Обработка естественного языка**
 - Извлечение информации из речи и текста

²<https://github.com/mlcourse-mm/mlcourse-mm2019spring>

Общая структура курса

«Введение в компьютерный интеллект»

- ① Машинное обучение²
 - Необходимые основы для всего курса
- ② Компьютерное зрение
 - Извлечение информации из визуальных образов (изображений и видео)
- ③ Обработка естественного языка
 - Извлечение информации из речи и текста
- ④ Обучение с подкреплением
 - Интерактивное взаимодействие со средой

²<https://github.com/mlcourse-mm/mlcourse-mm2019spring>

- Видеонаблюдение: нахождение и слежка за объектами

Компьютерное зрение в жизни

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей
- Улучшение качества фотографий

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей
- Улучшение качества фотографий
- Зрение роботов

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей
- Улучшение качества фотографий
- Зрение роботов
- Поиск по фотографиям

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей
- Улучшение качества фотографий
- Зрение роботов
- Поиск по фотографиям
- Дополненная реальность

Компьютерное зрение в жизни

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей
- Улучшение качества фотографий
- Зрение роботов
- Поиск по фотографиям
- Дополненная реальность
- Автомобили без водителей

Компьютерное зрение в жизни

- Видеонаблюдение: нахождение и слежка за объектами
- Медицина: обнаружение злокачественных опухолей
- Улучшение качества фотографий
- Зрение роботов
- Поиск по фотографиям
- Дополненная реальность
- Автомобили без водителей
- Магазины без продавцов

Что будет в этом курсе

Теоретическая часть

- Задачи компьютерного зрения
 - Классификация, детекция, сегментация
- Известные сверточные нейросети
 - AlexNet, VGG, ResNet, R-CNN, Mask R-CNN
- Генеративные модели
 - Генеративные состязательные сети, (вариационный) автоэнкодер
- Обучение нейросетей
 - Градиентный спуск, обратное распространение ошибки, инициализация



Что будет в этом курсе

Теоретическая часть

- Задачи компьютерного зрения
 - Классификация, детекция, сегментация
- Известные сверточные нейросети
 - AlexNet, VGG, ResNet, R-CNN, Mask R-CNN
- Генеративные модели
 - Генеративные состязательные сети, (вариационный) автоэнкодер
- Обучение нейросетей
 - Градиентный спуск, обратное распространение ошибки, инициализация

Практическая часть

- Обработка изображений и нейросетевые фреймворки
 - Scikit-Image, Tensorflow, Keras
- Соревнования по компьютерному зрению

Чего не будет в этом курсе

- Теория обработки изображений³
 - Преобразование Фурье, свертки

³См. “Математические основы обработки изображений”, Мазуренко И.Л.



Чего не будет в этом курсе

- Теория обработки изображений³
 - Преобразование Фурье, свертки
- Классическое компьютерное зрение
 - Теория фильтрации

³См. “Математические основы обработки изображений”, Мазуренко И.Л.



Оценки за курс

- Оценки за курс будут выставляться в соответствии с набранными баллами за выполнение домашних заданий.
- По курсу будут предложены домашние задания трёх видов:
 - теоретические
 - практические
 - соревнования
- В конце семестра состоится экзамен, на котором при желании можно будет повысить свою оценку
- Предварительная шкала оценок:

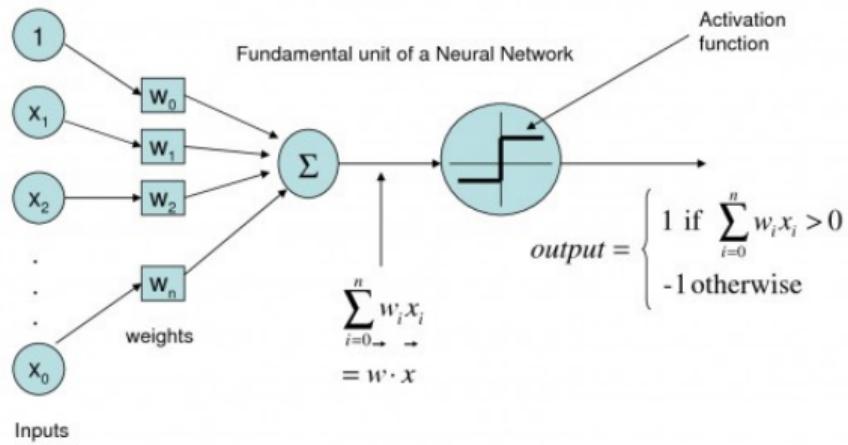
Оценка	Процент выполненных заданий
Отлично	80 %
Хорошо	60 %
Зачет	40 %

Полезные ресурсы

- Страница курса: <https://github.com/mlcoursemm/cvcoursemm2019autumn>
- Телеграмм-канал: <https://t.me/joinchat/AAAAAEUmx5cJL0dLXs0t8g>
- Группа обсуждения: <https://t.me/joinchat/B2iAGkx8IrWpQdM3LL-H-w>
- Почта курса: mlcoursemm@gmail.com

Нейросетевой подход

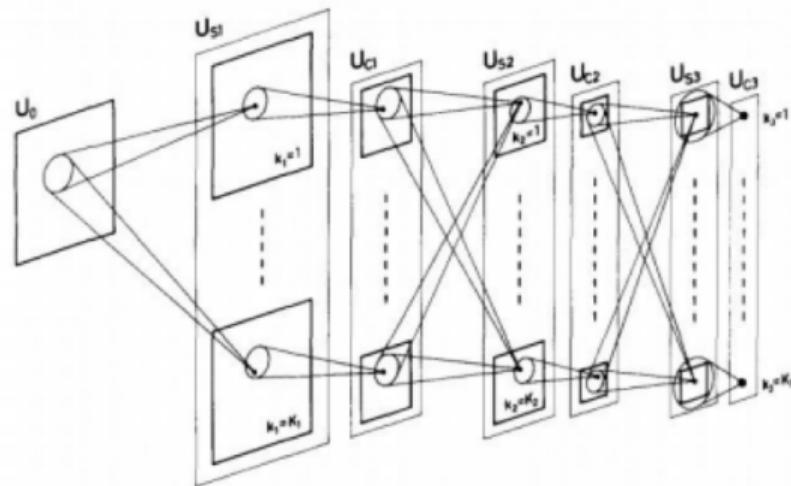
- На данный момент (2019 г.) подавляющее большинство алгоритмов при работе с компьютерным зрением использует нейронные сети
- Простейшая нейронная сеть — однослойный перцептрон — состоит из линейного слоя и следующей за ним нелинейной функции
- В 1943 г. Маккалок и Питтс формализуют понятие нейрона



- 1949 — Хебб предлагает первый алгоритм обучения перцептрана
- 1958 — Розенблattt изобретает однослойный персептрон и демонстрирует его способность решать задачу классификации
- 1974 — Галушкин и Вербос предлагают метод обратного распространения ошибки (впоследствии улучшен Хинтоном в 1986)
- 1975–1980 — Фукусима предлагает мощную модель когнитрана и неокогнитрана (прототип сверточных сетей)
- 1989 — Лекун представляет первую работу по объединению свёрточных нейросетей и метода обратного распространения ошибки
- 2012 — Крижевский и Хинтон публикуют первую успешную глубокую классификационную нейросеть AlexNet

Неокогнитрон

- Фукусима предложил практически современный метод построения архитектуры нейросетей, заимствованный им из модели первичной зрительной коры
- Два вида нейронов:
 - Простые (Simple), отвечающие за локальные признаки
 - Сложные (Complex), отвечающие за компенсацию искажения
 - Организованы в каскадную структуру SCSCSC...
 - В сверточной сети S=свертка, C=субдискретизация



- Главный минус: не было предложено метода обратного распространения ошибки для обучения

Нейронные сети — почему сейчас?

- Несмотря на то, что весь математический аппарат был готов уже к середине 80-х годов прошлого века, революция в компьютерном зрении произошла только в 2012 году

⁴<https://arxiv.org>

⁵<https://github.com>

Нейронные сети — почему сейчас?

- Несмотря на то, что весь математический аппарат был готов уже к середине 80-х годов прошлого века, революция в компьютерном зрении произошла только в 2012 году
- Основные причины начала новой эры в нейронных сетях:

⁴<https://arxiv.org>

⁵<https://github.com>



Нейронные сети — почему сейчас?

- Несмотря на то, что весь математический аппарат был готов уже к середине 80-х годов прошлого века, революция в компьютерном зрении произошла только в 2012 году
- Основные причины начала новой эры в нейронных сетях:
 - Появились большие наборы данных для обучения ($\sim 10^6$ изображений)

⁴<https://arxiv.org>

⁵<https://github.com>

Нейронные сети — почему сейчас?

- Несмотря на то, что весь математический аппарат был готов уже к середине 80-х годов прошлого века, революция в компьютерном зрении произошла только в 2012 году
- Основные причины начала новой эры в нейронных сетях:
 - Появились большие наборы данных для обучения ($\sim 10^6$ изображений)
 - Появились мощные видеокарты для обучения нейронных сетей

⁴<https://arxiv.org>

⁵<https://github.com>

Нейронные сети — почему сейчас?

- Несмотря на то, что весь математический аппарат был готов уже к середине 80-х годов прошлого века, революция в компьютерном зрении произошла только в 2012 году
- Основные причины начала новой эры в нейронных сетях:
 - Появились большие наборы данных для обучения ($\sim 10^6$ изображений)
 - Появились мощные видеокарты для обучения нейронных сетей
 - Открытость исследований (arxiv⁴, github⁵)

⁴<https://arxiv.org>

⁵<https://github.com>



Ресурсы для работы с нейросетями

- На данный момент обучение нейросетей немыслимо без использования **GPU** (Graphical Processing Unit) от NVIDIA либо **TPU** (Tensor Processing Unit) от Google
- Для обучения промышленных нейросетей (например, Mask R-CNN) необходим объем памяти GPU от 12-16 ГБ
 - В силу этого подходящие карты на данный момент: Titan X, GTX 1080 Ti, RTX 2080 Ti, P100, V100
 - Для любого проекта необходимо обеспечить доступ к данному оборудованию (например, удаленно через ssh)
 - Либо воспользоваться облачными сервисами (например, через Google Colaboratory⁶ либо Amazon Web Services⁷)

⁶<https://colab.research.google.com>

⁷<https://aws.amazon.com/ru/ec2/>

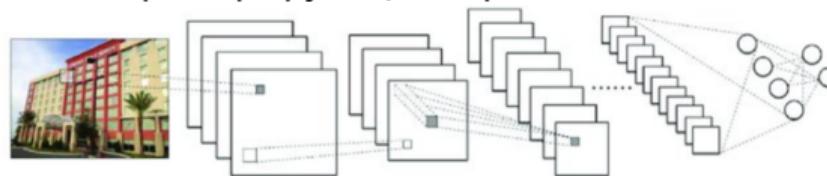
Задача классификации

- **Задача:** отнести входное изображение к одному из классов
 - По фотографии человека определить пол, возраст и т. д.
 - По фотографии автомобиля определить марку и модель
- Самая простая задача компьютерного зрения
 - Решается с помощью сверточных нейросетей (СНС)
- **Проблема:** двусмысленность изображений
 - Не любое изображение можно описать одним классом

Задача классификации

- **Задача:** отнести входное изображение к одному из классов
 - По фотографии человека определить пол, возраст и т. д.
 - По фотографии автомобиля определить марку и модель
- Самая простая задача компьютерного зрения
 - Решается с помощью сверточных нейросетей (СНС)
- **Проблема:** двусмысленность изображений
 - Не любое изображение можно описать одним классом

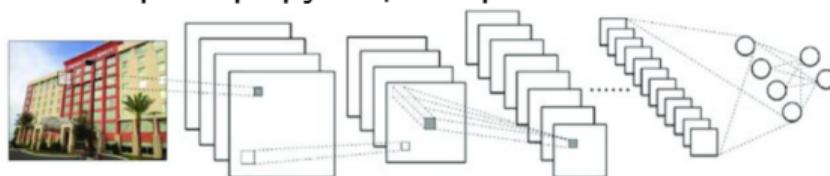
Пример функционирования СНС



Задача классификации

- **Задача:** отнести входное изображение к одному из классов
 - По фотографии человека определить пол, возраст и т. д.
 - По фотографии автомобиля определить марку и модель
- Самая простая задача компьютерного зрения
 - Решается с помощью сверточных нейросетей (СНС)
- **Проблема:** двусмысленность изображений
 - Не любое изображение можно описать одним классом

Пример функционирования СНС



Основные представители:

- 2012 — AlexNet
- 2014 — VGGNet
- 2014 — Inception
- 2015 — ResNet
- 2016 — ResNeXt
- 2017 — MobileNet
- 2017 — (P)NASNet
- 2018 — GPipe
- 2019 — EfficientNet

Проблемная классификация

Сено, лошадь или женщина?



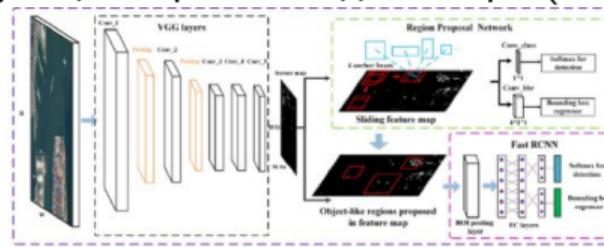
Задача детекции

- **Задача:** по входному изображению требуется найти объекты заранее заданных классов и обвести их прямоугольником
- Сложнее, чем классификация, так как необходимо не только распознать класс, но и локализовать объект
- Существуют алгоритмы, которые решают эту задачу в т. ч. в реальном времени
 - Faster R-CNN, YOLO, SSD

Задача детекции

- **Задача:** по входному изображению требуется найти объекты заранее заданных классов и обвести их прямоугольником
- Сложнее, чем классификация, так как необходимо не только распознать класс, но и локализовать объект
- Существуют алгоритмы, которые решают эту задачу в т. ч. в реальном времени
 - Faster R-CNN, YOLO, SSD

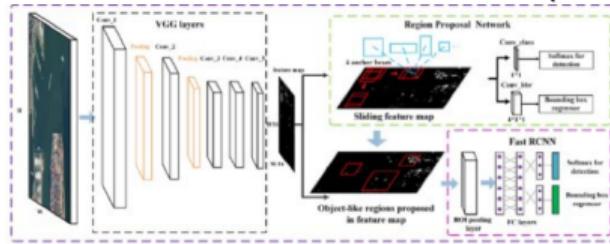
Пример функционирования детектора (Faster R-CNN)



Задача детекции

- **Задача:** по входному изображению требуется найти объекты заранее заданных классов и обвести их прямоугольником
- Сложнее, чем классификация, так как необходимо не только распознать класс, но и локализовать объект
- Существуют алгоритмы, которые решают эту задачу в т. ч. в реальном времени
 - Faster R-CNN, YOLO, SSD

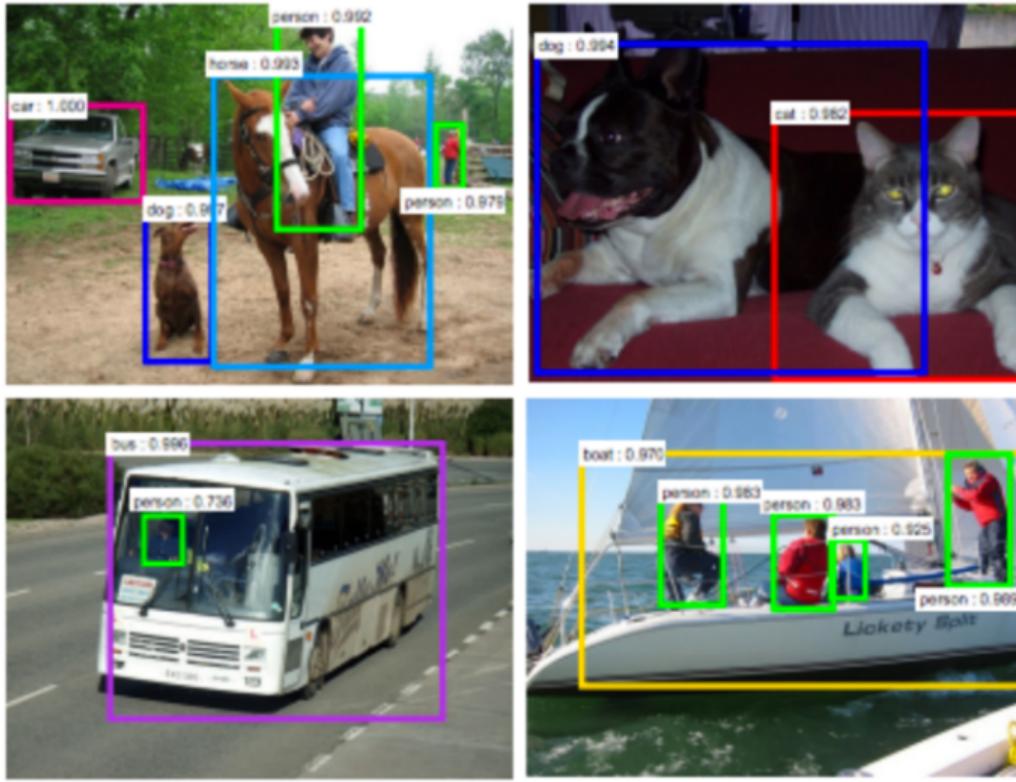
Пример функционирования детектора (Faster R-CNN)



Основные представители:

- 2013 — Overfeat
- 2014 — R-CNN
- 2015 — Faster R-CNN
- 2015 — YOLO
- 2015 — SSD
- 2016 — R-FCN
- 2017 — FPN
- 2017 — RetinaNet
- 2018 — SNIPER
- 2019 — CenterNet

Пример работы детектора



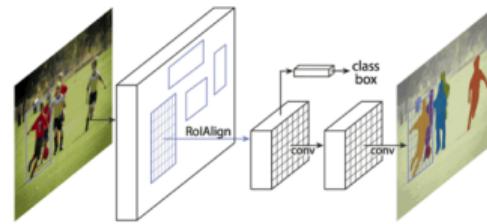
Задача сегментации

- **Задача:** по входному изображению найти объекты заранее заданных классов и выдать пиксельную маску объекта
- Сложнее, чем детекция — нужно точно определять границы объекта, а не просто заключить в рамку
- На данный момент существуют алгоритмы-комбайны (решающие одновременно задачи классификации, детекции и сегментации)

Задача сегментации

- **Задача:** по входному изображению найти объекты заранее заданных классов и выдать пиксельную маску объекта
- Сложнее, чем детекция — нужно точно определять границы объекта, а не просто заключить в рамку
- На данный момент существуют алгоритмы-комбайны (решающие одновременно задачи классификации, детекции и сегментации)

Пример функционирования решения-комбайна (Mask R-CNN)



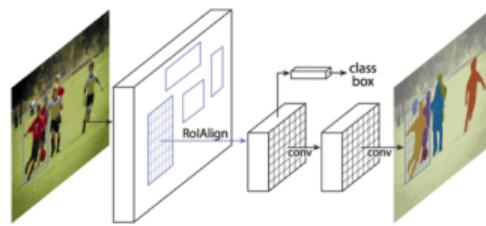
Задача сегментации

- **Задача:** по входному изображению найти объекты заранее заданных классов и выдать пиксельную маску объекта
- Сложнее, чем детекция — нужно точно определять границы объекта, а не просто заключить в рамку
- На данный момент существуют алгоритмы-комбайны (решающие одновременно задачи классификации, детекции и сегментации)

Основные представители:

- 2014 — FCN
- 2015 — U-Net
- 2015 — SegNet
- 2016 — RefineNet
- 2016 — PSPNet
- 2017 — Mask R-CNN
- 2018 — DeepLabv3+

Пример функционирования решения-комбайна (Mask R-CNN)



Пример работы решения-комбайна



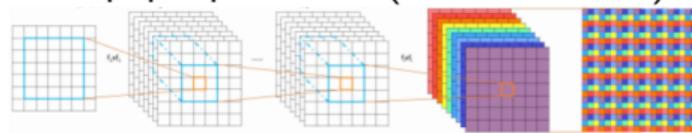
Задачи улучшения изображений

- **Задача сверхразрешения:** увеличить пространственное разрешение фотографии / видео-кадра
- **Метод решения:** применить нейронную сеть, заранее обученную на множестве картинок / видео
- **Цель:** улучшить классические алгоритмы интерполяции

Задачи улучшения изображений

- **Задача сверхразрешения:** увеличить пространственное разрешение фотографии / видео-кадра
- **Метод решения:** применить нейронную сеть, заранее обученную на множестве картинок / видео
- **Цель:** улучшить классические алгоритмы интерполяции

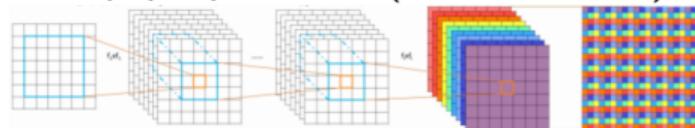
Пример функционирования алгоритма сверхразрешения (ESPCN, 2016)



Задачи улучшения изображений

- **Задача сверхразрешения:** увеличить пространственное разрешение фотографии / видео-кадра
- **Метод решения:** применить нейронную сеть, заранее обученную на множестве картинок / видео
- **Цель:** улучшить классические алгоритмы интерполяции

Пример функционирования алгоритма сверхразрешения (ESPCN, 2016)

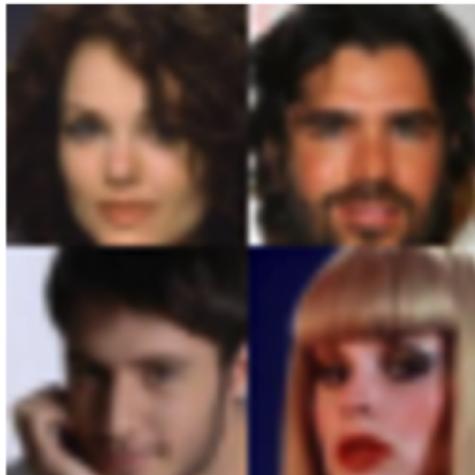


Метод сверхразрешения	PSNR
Бикубическая интерполяция	24.32 dB
Бикубическая интерполяция + GIMP	24.69 dB
Бикубическая интерполяция + Photoshop	24.72 dB
<i>Нейро сверхразрешение</i>	<i>27.20 dB</i>

Задачи улучшения изображений

Сверхразрешение для лиц — пример

Бикубическая интерполяция: PSNR = 24.32



Нейросверх-разрешение: PSNR = 27.20



Малоразмерный вход

Задачи улучшения изображений

- **Задача:** улучшить качество дефектных изображений

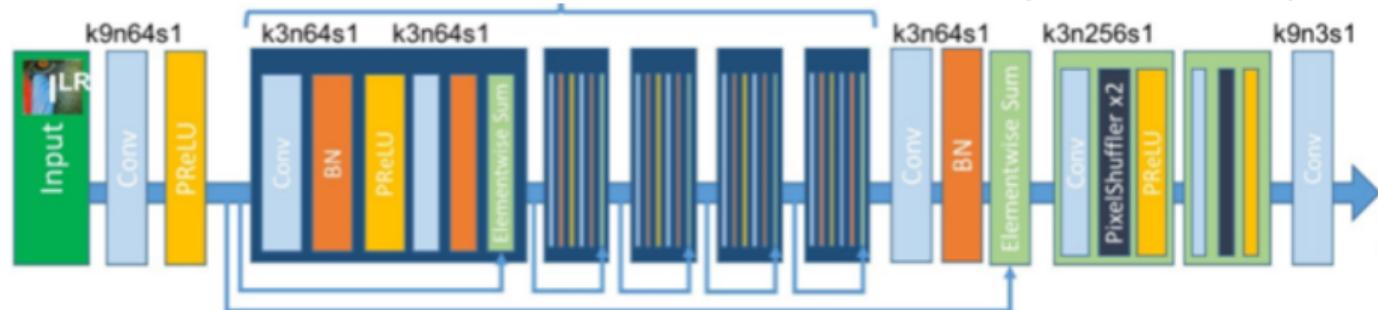
- Устранить смаз (motion blur)
- Устранить размытие (defocus blur)
- Удалить артефакты, являющиеся результатом попытки устранить дефекты зашитым в устройстве алгоритмом (image reconstruction)

Задачи улучшения изображений

- **Задача:** улучшить качество дефектных изображений

- Устранить смаз (motion blur)
- Устранить размытие (defocus blur)
- Удалить артефакты, являющиеся результатом попытки устранить дефекты зашитым в устройстве алгоритмом (image reconstruction)

Пример функционирования алгоритма деблюра (SRGAN, 2016)



Задачи улучшения изображений

Исходное
размытие



Деблюр для лиц — пример
Классический
Wiener



Коммерческое
решение



Нейросетевое
решение



Распознавание автомобильных номеров

- **Задача:** по видеопотоку детектировать транспортное средство и распознать госномер
- Зачастую итоговое решение состоит из следующих модулей:
 - Детектирование автомобиля
 - Распознавание марки и модели (опционально)
 - Трекинг автомобиля
 - Детектирование номера
 - Распознавание номера
 - Объединение результатов с разных кадров / поиск ключевых кадров

Распознавание автомобильных номеров

- **Задача:** по видеопотоку детектировать транспортное средство и распознать госномер
- Зачастую итоговое решение состоит из следующих модулей:
 - Детектирование автомобиля
 - Распознавание марки и модели (опционально)
 - Трекинг автомобиля
 - Детектирование номера
 - Распознавание номера
 - Объединение результатов с разных кадров / поиск ключевых кадров



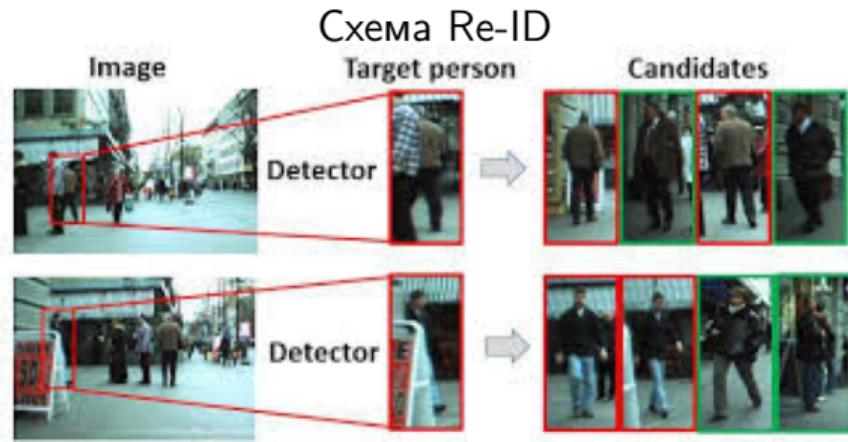
Ре-идентификация пешеходов

- **Задача:** идентифицировать пешеходов на камерах наблюдения
- **Метод решения:**
 - Детектирование пешеходов
 - Трекинг пешеходов
 - Извлечение признаков
 - Поиск в пространстве признаков

Примеры прикладных задач

Ре-идентификация пешеходов

- Задача: идентифицировать пешеходов на камерах наблюдения
- Метод решения:
 - Детектирование пешеходов
 - Трекинг пешеходов
 - Извлечение признаков
 - Поиск в пространстве признаков



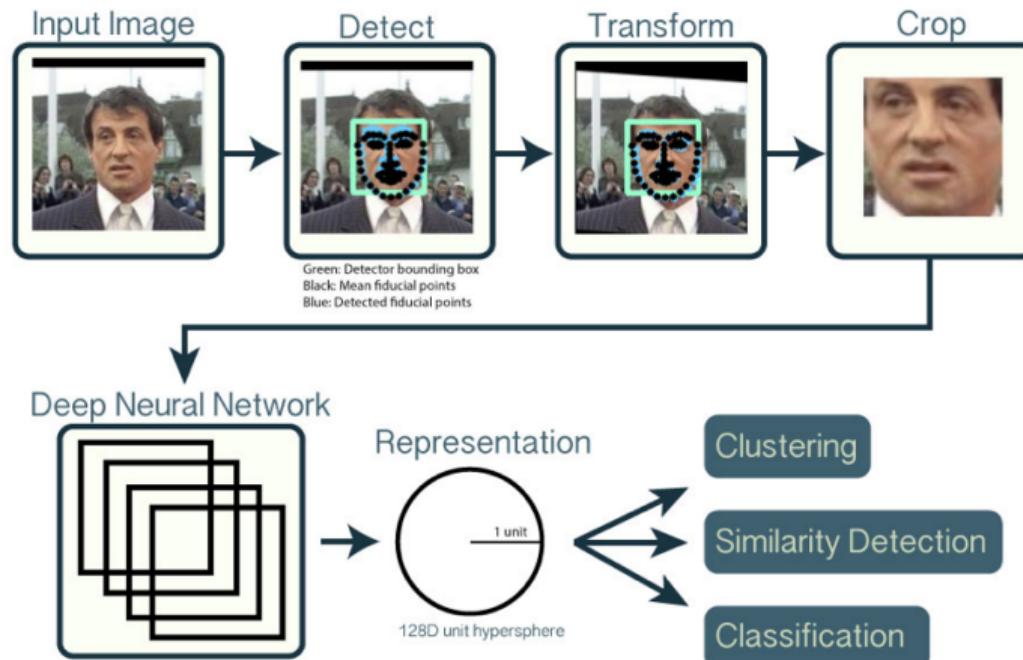
Распознавание лиц

- **Задача:** по входному изображению лица распознать человека. Есть два подтипа задач:
 - **Верификация** — принадлежат ли две входные картинки одному человеку?
 - **Идентификация** — нахождение в базе данных человека, фотография которого подана на вход
- **Метод решения:** с помощью каскада СНС
- Главная особенность — нахождение N -мерного вектора признаков, соответствующего данному человеку



Примеры прикладных задач

Типичная система распознавания лиц:



Синтез (создание) лиц

- **Задача:** сгенерировать набор синтетических лиц, неотличимых от реальных фотографий
 - Зачастую нужно для расширения обучающей выборки для задачи распознавания лиц
 - Необходимо генерировать как можно больше разных личностей
- **Метод решения:** с помощью специальных **генеративных состязательных нейросетей (GAN)**

Примеры прикладных задач

Какое из этих лиц — синтетическое⁸?



⁸<https://thispersondoesnotexist.com>

Спасибо за внимание!

