МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«СЕВАСТОПОЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт Информационных технологий   
и управления в технических системах

Кафедра Информационных систем и технологий

|  |  |  |  |
| --- | --- | --- | --- |
| № | Дата поступления на кафедру | Подпись отв. за регистрацию | Подпись преподавателя |
|  |  |  |  |
|  |  |  |  |

**ОТЧЕТ**

о \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) практике

(вид практики) (тип практики)

в \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(наименование организации)

Выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия И.О. обучающегося)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(шифр группы)

Направление / специальность \_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(код, наименование)

Руководитель практики от Университета

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия И.О. руководителя)

Севастополь

2022 г.

Индивидуальное задание на практику

Выполнить обзор литературных источников по методам машинного обучения, применяемым для поиска объектов на изображении.

Содержание

[Введение 3](#_Toc106403820)

[Основная часть 4](#_Toc106403821)

[1 Машинное зрение 4](#_Toc106403822)

[2 Задача нахождения объектов на изображении 5](#_Toc106403823)

[3 Базовые методы поиска объектов на изображении 7](#_Toc106403824)

[4 Методы машинного обучения 7](#_Toc106403825)

[5 Свёрточные нейронные сети 9](#_Toc106403826)

[Заключение 11](#_Toc106403827)

[Список литературы 12](#_Toc106403828)

# Введение

Машинное (компьютерное) зрение используется для решения многих прикладных задач, возникающих в современном мире, таких как автономное управление беспилотным транспортом, автоматизированные системы безопасности, поддержка работы медицинских сотрудников. Технологии машинного зрения основаны на методах машинного обучения. Классическая задача в компьютерном зрении – это определение наличия в видеоданных некоторого характерного объекта, особенности или активности. Эта задача может быть достоверно и легко решена человеком, существуют также алгоритмы машинного зрения, позволяющие распознавать один или несколько предварительно заданных объектов или классов объектов. Однако до сих пор данная задача не решена удовлетворительно в общем случае: распознавание случайных объектов в случайных ситуациях.

Объектом исследования являются методы машинного обучения, используемые для поиска объектов на изображении.

Предмет исследования – это характеристики, характерные свойства, особенности и возможности методов машинного обучения, используемых для поиска объектов на изображении.

Целью практики является выполнение обзора литературных источников и электронных ресурсов, посвященных теме компьютерного зрения, а именно задач обнаружения и распознавания объектов, и их решения при помощи классических методов и методов машинного обучения.

# Основная часть

## 1 Машинное зрение

Машинное зрение — это научное направление в области искусственного интеллекта и связанные с ним технологии получения изображений объектов реального мира, их обработки и использования полученных данных для решения разного рода прикладных задач без участия (полного или частичного) человека.

Машинное зрение сосредотачивается на применении, в основном промышленном, например, автономные роботы и системы визуальной проверки и измерений. Это значит, что технологии датчиков изображения и теории управления связаны с обработкой видеоданных для управления роботом и обработка полученных данных в реальном времени осуществляется программно или аппаратно.

Обработка изображений и анализ изображений в основном сосредоточены на работе с 2D изображениями, т.е. как преобразовать одно изображение в другое. Например, попиксельные операции увеличения контрастности, операции по выделению краёв, устранению шумов или геометрические преобразования, такие как вращение изображения. Данные операции предполагают, что обработка/анализ изображения действуют независимо от содержания самих изображений.

Компьютерное зрение сосредотачивается на обработке трехмерных сцен, спроектированных на одно или несколько изображений. Например, восстановлением структуры или другой информации о 3D сцене по одному или нескольким изображениям. Компьютерное зрение часто зависит от более или менее сложных допущений относительно того, что представлено на изображениях.

Также существует область названная визуализация, которая первоначально была связана с процессом создания изображений, но иногда имела дело с обработкой и анализом. Например, рентгенография работает с анализом видеоданных медицинского применения.

Наконец, распознавание образов является областью, которая использует различные методы для получения информации из видеоданных, в основном, основанные на статистическом подходе. Значительная часть этой области посвящена практическому применению этих методов.

Таким образом, можно сделать вывод, что понятие «машинное зрение» на сегодняшний день включает в себя: компьютерное зрение, распознавание зрительных образов, анализ и обработка изображений и т.д.

## 2 Задача нахождения объектов на изображении

Задача нахождения объектов на изображении может быть поставлена различным образом и включает в себя класс других задач, помогающих определить, какие объекты находятся на изображении и где они расположены в сетке пикселей исходного изображения.

Задача семантической сегментации (англ. semantic segmentation) — задача, в которой на вход модели подаётся изображение, а на выходе для каждого пикселя является метка принадлежности этого пикселя к определённой категории. Например, если в исходном изображении человек переходит дорогу, то для каждого пикселя необходимо вывести, является ли этот пиксель частью человеческого тела, профиля дороги, знака дорожного движения, неба, или какого-то другого типа. Существенный недостаток применения одной лишь семантической сегментации относительно задач, связанных с распознаванием объектов — маркировка пикселей по принадлежности только к типу объекта, что не создаёт различия между объектами как таковыми. Например, если назвать "объектом" связную область пикселей, характеризующих одинаковый тип, то два объекта, перегораживающих друг друга на исходном изображении, будут определены как один объект, что в корне неверно. Задача семантической сегментации изображения с дифференцированием объектов называется задачей сегментации экземпляров (англ. instance segmentation). Модели, решающие задачу сегментации экземпляров, применяются, в том числе, для подсчёта людей в массовых скоплениях, для автомобилей с автоматическим управлением.

Задача классификации с локализацией (англ. classification and localization) — задача, в которой в дополнение к предсказанию метки категории класса определяется рамка, ограничивающая местоположение экземпляра одиночного объекта на картинке. Как правило, рамка имеет прямоугольную форму, её стороны ориентированы параллельно осям исходного изображения, а площадь является минимальной при условии полного нахождения экземпляра объекта внутри этой рамки. Такую прямоугольную рамку называют термином "ограничивающая рамка" (англ. bounding box). Ограничивающую рамку можно задать как при помощи центра, ширины и высоты, так и при помощи четырёх сторон. Модель в данном случается одновременно обучается как верной классификации, так и максимально точному определению границ рамки.

Задача детекции объектов (англ. object detection) — задача, в рамках которой необходимо выделить несколько объектов на изображении посредством нахождения координат их ограничивающих рамок и классификации этих ограничивающих рамок из множества заранее известных классов. В отличие от классификации с локализацией, число объектов, которые находятся на изображении, заведомо неизвестно.

На сегодняшний день задача поиска объектов на изображении может быть решена при помощи следующих подходов: цветовые фильтры, выделение и анализ контуров, сопоставление с шаблоном, работа с особыми точками, методы машинного обучения.

## 3 Базовые методы поиска объектов на изображении

К базовым методам поиска объектов на изображении относятся следующие методы: цветовые фильтры, выделение и анализ контуров, сопоставление с шаблоном, работа с особыми точками.

Цветовые фильтры используются в том случае, если объект существенно выделяется на фоне по цвету, в таком случае для его обнаружения достаточно правильно подобрать соответствующий фильтр.

Если заранее известна форма объекта, то можно применять метод выделения и анализа контуров. Таким образом, в случае поиска круглого объекта анализируются окружности на изображении.

Сопоставление с шаблоном может быть использовано, если имеется изображение искомого объекта. Данный метод заключается в поиске на изображении областей, совпадающих с шаблонным изображением объекта.

При работе с особыми точками на изображении объекта выделяются особые точки, например углы, и сопоставляются с такими же особенностями на анализируемом изображении.

## 4 Методы машинного обучения

Методы машинного обучения основываются на получении классификатора путём обучения на тренировочной выборке объектов и дальнейшем его применении на всех частях некоторым образом разделенного изображения. Главным преимуществом этих методов является высокая обобщающая способность: в то время, как подходы сопоставления шаблонов или особых точек распознают только одно конкретное изображение, методы машинного обучения способны определять объект, класс или группу объектов.

Задачу поиска объектов на видео можно разделить на три подзадачи или этапа: обнаружение, распознавание и сопровождение. На этапе обнаружения необходимо на изображении выделить области поиска интересующих нас объектов. Затем, на этапе распознавания, для каждого из найденных объектов уточняется его тип. Сопровождение заключается в локализации распознанных объектов на следующих кадрах.

Машинное обучение можно применять, по крайней мере, для решения задач обнаружения и распознавания, сводимых к построению классификатора изображений. Классификатор изображений состоит из двух частей: метод извлечения признаков и, собственно, сам классификатор. Среди методов извлечения признаков, используемых для построения классификаторов изображений, наиболее популярны методы HOG, BoW, Виолы-Джонса.

Гистограммы направленных градиентов (HOG) – это метод извлечения признаков из изображений, который напоминает метод вычисления дескрипторов SIFT для особых точек, но в этом методе вычисление происходит не для окрестности, а для всего изображения. Общий алгоритм вычисления HOG состоит в следующем: изображение разделяется на части – ячейки, затем для каждой ячейки осуществляется построение гистограммы направлений яркости градиента, после чего гистограммы ячеек нормируются по контрасту и объединяются в классы.

Мешок слов (BoW)– это метод извлечения признаков из изображений, который является адаптацией метода частотного анализа текстов. Применение методов машинного обучения для распознавания объектов на изображении предполагает наличие учебного набора, состоящего как минимум из двух частей: набор изображений, содержащих объект, и набор изображений, его не содержащих. Аналогично методу частотного анализа текстов, для использования мешка слов необходимо для начала собрать словарь. Алгоритм построения словаря состоит в следующем. Сначала выбирается метод выявления особых точек и метод вычисления их дескрипторов. Затем для каждого изображения, входящего в учебный набор, определяются особые точки и вычисляются дескрипторы для них. После этого похожие дескрипторы объединяются в группы, то есть выполняется кластеризация множества полученных дескрипторов. Полученные кластеры будут играть роль визуальных слов и составлять словарь для частотного анализа изображения. Признаки BoW извлекаются из изображения следующим образом: методом, который был выбран на этапе составления словаря, определяются особые точки на изображении и вычисляются для них дескрипторы, найденные дескрипторы разбираются по кластерам словаря и затем для каждого кластера посчитывается количество найденных дескрипторов. Таким образом, вектор BoW-признаков – это количество найденных дескрипторов в каждом кластере словаря.

Метод Виолы-Джонса, который часто используется для поиска лиц на изображении, построен на основе признаков Хаара. Признак Хаара вычисляется следующим образом: выбирается прямоугольная область на изображении, эта область разбивается на несколько смежных прямоугольных частей, в каждой из них суммируется яркость точек, после чего вычисляется разность между полученными суммами. Итоговые разность и являются значением признака Хаара. Для формирования характеристики изображения используют несколько разных признаков Хаара, каждый из которых имеет свои параметры: размер области для вычисления признака, количество частей, позиция на изображении.

Существует также модель классификатора изображений, обладающая встроенным механизмом извлечения признаков из изображения – свёрточная нейронная сеть. Механизм извлечения признаков является частью сети и «самонастраивается» в процессе обучения.

## 5 Свёрточные нейронные сети

При распознавании изображения, если речь идет о машинном обучении, как правило используют сверточные сети. Эти сети, как и обычные, состоят из нейронов с обучаемыми весами и смещениями. Каждый нейрон имеет несколько входов, генерирует числовое значение и передает его другим нейронам. Как и обычная нейронная сеть, сверточная сеть представляет собой функцию, принимающую на вход вектор со значениями в неограниченном интервале, и отдающую на выходе вектор со значениями в ограниченном интервале. Точно так же она имеет функцию потерь на последнем слое, необходимую для обучения. Различие в том, что сверточная сеть напрямую учитывает то, что на вход ей подается изображение, поэтому возможно сделать оптимизации, облегчающие процесс обучения и распознавания.

В обычной нейронной сети, имеющей последовательность скрытых слоев, нейроны между слоями полностью связаны между собой, и полностью независимы между собой внутри одного слоя. Из-за того, каждый нейрон связан со всеми нейронами из соседних слоев, такие сети плохо масштабируемы для обучения на больших векторах, к которым относятся, в частности, большие изображения.

Сверточные сети берут в преимущество тот факт, что на вход им даются именно изображения, и их архитектура приспособлена под такие данные. В отличии от обычных нейросетей, нейроны в сверточной сети располагаются в трех измерениях: ширине, высоте и глубине. Входные изображения по размеру должны совпадать с входным слоем. Нейроны в следующих слоях имеют связи только с небольшой областью нейронов предыдущего слоя вместо того, чтобы быть соединенными со всеми предыдущими нейронами. Слой на выходе может быть одномерным, если необходимо классифицировать изображение по конечному числу классов.

# Заключение

В результате практики был проведен сбор и анализ источников информации о методах машинного обучения, которые используются для распознавания объектов на изображении. Были рассмотрены задачи компьютерного зрения, типы задач нахождения объекта на изображении, базовые методы и методы машинного обучения распознавания объектов на изображении.

# Список литературы

1. Richard Szeliski Computer Vision: Algorithms and Applications – Springer, 2010.
2. Введение в компьютерное зрение // Лекториум [Электронный ресурс] – URL: https://www.lektorium.tv/course/22847.
3. Задача нахождения объектов на изображении // Университет ИТМО. Викиконспекты [Электронный ресурс] – URL: https://neerc.ifmo.ru/wiki/index.php?title=Задача\_нахождения\_объектов\_на\_изображении.
4. Машинное зрение. Что это и как им пользоваться? Обработка изображений оптического источника // Хабр [Электронный ресурс] – URL: https://habr.com/ru/post/350918/.
5. О задаче поиска объекта на изображении. Часть 1: Базовые методы // Дом-страница Евгения Сергеевича Борисова. [Электронный ресурс] – URL: http://mechanoid.su/cv-image-detector.html.
6. О задаче поиска объекта на изображении. Часть 2: Применение методов машинного обучения // Дом-страница Евгения Сергеевича Борисова. [Электронный ресурс] – URL: http://mechanoid.su/cv-image-detector2.html.
7. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход.: Пер. с англ. – М.: Издательский дом «Вильямс», 2004. – 928с. ISBN 5-8459-0542-7.