

# Донейросетевые подходы к работе с изображениями

Морозов Дмитрий, БПМИ172

# Компьютерное зрение

- **Компьютерное зрение (Computer vision)** — теория и технология создания машин, которые могут производить обнаружение, отслеживание и классификацию объектов.

# Примеры применения компьютерного зрения

- Обработка изображений в медицине. Эта область характеризуется получением информации из видеоданных для постановки медицинского диагноза пациентам.
- Промышленность. Примером может служить контроль качества, когда детали или конечный продукт автоматически проверяются на наличие дефектов.
- Военное применение. Примерами являются обнаружение вражеских солдат и транспортных средств и управление ракетами.
- Автономные транспортные средства, включая подводные, наземные (роботы, машины), воздушные.

# Типичные задачи компьютерного зрения

- **Распознавание.** Классическая задача в компьютерном зрении, обработке изображений и машинном зрении это определение содержат ли видеоданные некоторый характерный объект, особенность или активность.
- **Движение.** Несколько задач, связанных с оценкой движения, в которых последовательность изображений (видеоданные) обрабатываются для нахождения оценки скорости каждой точки изображения или 3D сцены. Примерами таких задач являются: определение трехмерного движения камеры; слежение, то есть следование за перемещениями.
- **Восстановление сцены.** Даны два или больше изображения сцены, или видеоданные. Восстановление сцены имеет задачей воссоздать трехмерную модель сцены.
- **Восстановление изображений.** Задача восстановления изображений — это удаление шума (шум датчика, размытость движущегося объекта и т. д.).

# Типы признаков

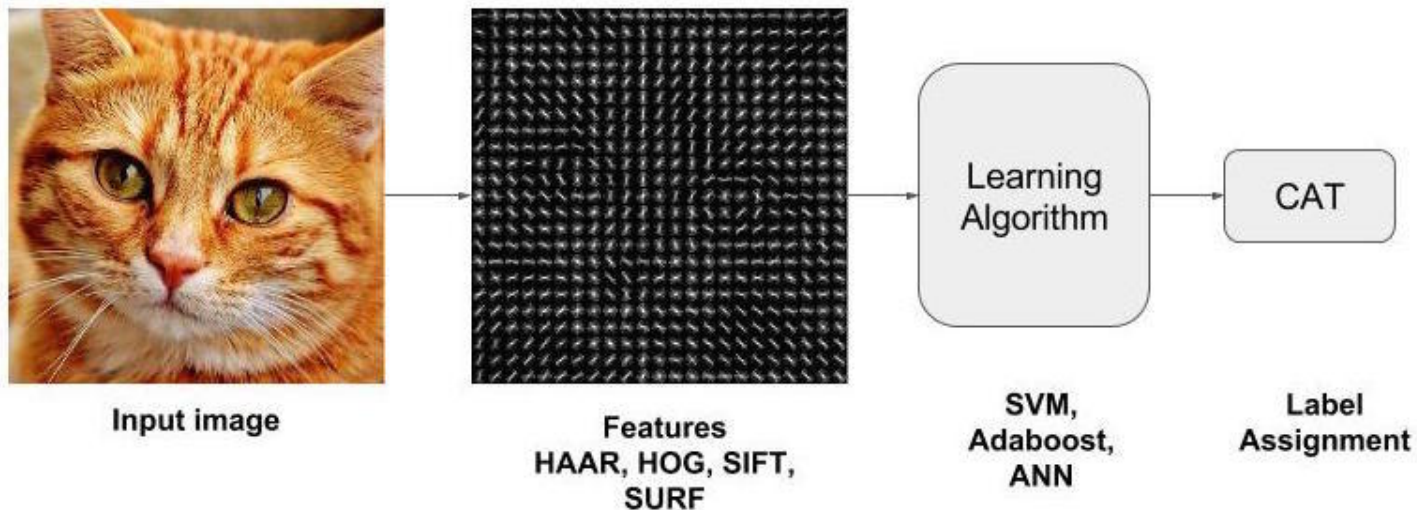
- **Границы / края (Edges).** Границы — это точки, обозначающие разделение двух областей изображения. В общем случае граница может иметь почти произвольную форму. На практике границы обычно определяются как наборы точек на изображении, которые имеют значительную градиентную величину.
- **Углы / точки интереса (Corners).** Углы относятся к точечным элементам изображения, которые имеют локальную двумерную структуру. Название «угол» возникло, так как ранние алгоритмы сначала выполняли обнаружение границ, а затем анализировали их, чтобы найти быстрые изменения в направлении (углы). Затем эти алгоритмы были доработаны, чтобы явное обнаружение границ больше не требовалось.
- **Пятна / области точек интереса (Blobs).** Пятна предоставляют дополнительное описание структур изображения в терминах областей, в отличие от углов, которые являются более точечными. Детекторы пятен могут обнаруживать области на изображении, которые являются слишком гладкими, чтобы быть обнаруженными детектором углов.
- **Хребты / ребра (Ridges).** Ребро можно рассматривать как одномерную кривую, которая представляет ось симметрии, и, кроме того, имеет атрибут локальной ширины ребра, связанный с каждой его точкой.

# Распознавание изображений (Классификация изображений)

- Алгоритм распознавания изображений принимает изображение в качестве входных данных и выводит то, что оно содержит. Другими словами, выводом является принадлежность к определенному классу (например, «кошка», «собака»).

# Распознавание изображений

- Следующая диаграмма иллюстрирует шаги, включенные в традиционный классификатор изображений.



# Предварительная обработка

- Часто входное изображение предварительно обрабатывается для нормализации эффектов контрастности и яркости. Распространенным этапом предварительной обработки является вычитание среднего значения интенсивности изображения и деление на стандартное отклонение. При работе с цветными изображениями преобразование цветового пространства может помочь получить лучшие результаты.



# Извлечение признаков

- Входное изображение содержит слишком много дополнительной информации, которая не нужна для классификации. Поэтому первым шагом в классификации является упрощение изображения путем извлечения важной информации и исключения остальной.

# Извлечение признаков

- Известными алгоритмами обнаружения признаков являются **Гистограмма направленных градиентов (Histogram of Oriented Gradients, HOG)** и **Масштабно-инвариантная трансформация признаков (Scale-Invariant Feature Transform, SIFT)**.

# Histogram of Oriented Gradients

- **Гистограмма направленных градиентов** — дескриптор признаков, основанный на подсчете количества направлений градиента в локальных областях изображения.
- **Дескриптор функции** — это представление изображения или части изображения, которое упрощает его, отбрасывая постороннюю информацию.
- Как правило, дескриптор функции преобразует изображение размером ширина × высота × 3 (каналы) в вектор длины  $n$ . В случае дескриптора функции HOG входное изображение имеет размер  $64 \times 128 \times 3$ , а вектор соответственно имеет длину 3780.
- В дескрипторе функции HOG в качестве функции используется распределение (гистограмма) направлений градиентов. Идея заключается в том, что величина градиентов велика вокруг краев и углов объектов (областей с резким изменением интенсивности).

# Histogram of Oriented Gradients

- **Расчет градиента**
- Вычисляются величины  $g_x$  и  $g_y$  градиента из оригинального изображения. Это можно сделать фильтрацией этого изображения со следующими ядрами:

$$[-1 \ 0 \ 1] \text{ и } [-1 \ 0 \ 1]^T$$

- Используя  $g_x$  и  $g_y$  вычисляется величина и направление градиента:

$$g = \sqrt{g_x^2 + g_y^2}, \theta = \arctan \frac{g_y}{g_x}$$

# Histogram of Oriented Gradients

- **Блоки.** Изображение разделяется на блоки  $8 \times 8$ .
- **Расчет гистограммы градиентов в блоках.** Для каждого пикселя в блоке  $8 \times 8$  известен градиент (его величина и направление). Таким образом, имеется 64 значения величины и 64 направления – 128 чисел. Гистограмма этих градиентов обеспечивает более полезное и компактное представление. Градиенты преобразуются в 9-интервальную гистограмму. Интервал гистограммы соответствует направлению градиента в 0, 20, 40 ... 160 градусов. Каждый пиксель распределяет значение градиента в один или два интервала, основываясь на принадлежности угла градиента к тому или иному интервалу.

# Histogram of Oriented Gradients

- **Нормализация блоков.** Гистограмма, рассчитанная в предыдущем шаге не очень устойчива к изменению освещения. Чтобы побороться с данным эффектом гистограмма нормализуется, т.е. гистограмму представляют как вектор из 9 элементов и каждый элемент вектора делят на величину вектора. В оригинальном описании метода HOG эта нормализация выполняется не для ячейки  $8 \times 8$ , которая вырабатывала гистограмму, а для блоков  $16 \times 16$ . Идея точно такая же, но вместо вектора длины в 9 элементов получается вектор в 36 элементов.
- **Характеристический вектор.** Для подсчета конечного вектора из исходного изображения блок  $16 \times 16$  двигается с шагом 8 и 36 значений, вычисленных на каждом шаге, объединяется для получения окончательного вектора признаков. Таким образом, с учетом размера входного изображения  $64 \times 128$  пикселей, производится 7 шагов по горизонтали и 15 шагов по вертикали. На каждом шаге мы рассчитывали 36 чисел, а значит, длина конечного вектора равна  $105 \times 36 = 3780$ .

# Обучающий алгоритм для классификации

- Перед стабильной работой, классифицирующий алгоритм нужно натренировать, показывая тысячи изображений на которых есть интересующий нас предмет и на которых его нет. Каждый алгоритм обучения учится по-разному, но главный принцип заключается в том, что они рассматривают характеристические векторы как точки в пространстве большей размерности и пытаются найти плоскости/поверхности, которые разбивают это пространство таким образом, что все объекты, принадлежащие, а одному классу, находятся на одной стороне плоскости/поверхности.
- В качестве такого алгоритма можно использовать **Метод опорных векторов (Support Vector Machine, SVM)**.

# Scale-Invariant Feature Transform

- **Масштабно-инвариантная трансформация признаков** — алгоритм для выявления и описания локальных признаков в изображениях.
- Сначала в SIFT извлекаются ключевые точки объектов из набора контрольных изображений и запоминаются в базе данных. Объект распознаётся в новом изображении путём сравнения каждого признака из нового изображения с признаками из базы данных и нахождения признаков-кандидатов на основе евклидова расстояния между векторами признаков.

Из полного набора соответствий в новом изображении отбираются поднаборы ключевых точек, которые наиболее хорошо согласуются с объектом по его местоположению, масштабу и ориентации. Определение подходящих блоков признаков осуществляется быстро с помощью эффективной реализации хеш-таблицы обобщённого преобразования Хафа. Каждый блок из 3 или более признаков, согласующийся с объектом и его положением, подлежит дальнейшей подробной проверке соответствия модели, и резко отклоняющиеся блоки отбрасываются.

Наконец, вычисляется вероятность, что определённый набор признаков говорит о присутствии объекта, что даёт информацию о точности совпадения и числе возможных промахов. Объекты, которые проходят все эти тесты, могут считаться правильными с высокой степенью уверенности.



# Scale-Invariant Feature Transform

- **Нахождение особых точек.** Основным моментом в детектировании особых точек является построение пирамиды гауссианов и разностей гауссианов. Гауссианом (или изображением, размытым гауссовым фильтром) является изображение

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

*Здесь  $L$  — значение гауссиана в точке с координатами  $(x, y)$ , а  $\sigma$  — радиус размытия.  $G$  — размытие по Гауссу,  $I$  — значение исходного изображения,  $*$  — операция свертки.*

Разностью гауссианов называют изображение, полученное путем попиксельного вычитания одного гауссиана исходного изображения из гауссиана с другим радиусом размытия.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Ключевые точки затем отбираются как максимум/минимум разности гауссианов, которые встречаются в различных масштабах.

# Scale-Invariant Feature Transform

- **Локализация особых точек.** Проводится проверка пригодности точки экстремума на роль ключевой. Первым делом определяются координаты особой точки с субпиксельной точностью. Это достигается с помощью аппроксимирования функции разности многочленом Тейлора второго порядка, взятого в точке вычисленного экстремума.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x$$

*Здесь  $D$  — функция разности,  $X = (x, y, \sigma)$  — вектор смещения относительно точки разложения, первая производная разности — градиент, вторая производная разности — матрица Гессе.*

- Экстремум многочлена Тейлора находится путем вычисления производной и приравнивания ее к нулю. В итоге получим смещение точки вычисленного экстремума, относительно точного

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}$$

# Scale-Invariant Feature Transform

- Если смещение  $\hat{x}$  больше 0.5 в любом направлении, то это означает, что на самом деле точка экстремума была вычислена неверно и нужно сдвинуться к соседней точке в направлении указанных компонент. Для соседней точки все повторяется заново. Если таким образом мы вышли за пределы октавы, то следует исключить данную точку из рассмотрения.

# Scale-Invariant Feature Transform

- **Отбрасывание ключевых точек низкого контраста.** Когда положение точки экстремума вычислено, проверяется на малость само значение разности в этой точке по формуле

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x}$$

Если эта проверка не проходит, то точка исключается, как точка с малым контрастом.

# Scale-Invariant Feature Transform

- **Исключение вклада рёбер.** Если особая точка лежит на границе какого-то объекта или плохо освещена, то такую точку можно исключить из рассмотрения. Эти точки имеют большой изгиб вдоль границы и малый в перпендикулярном направлении. Этот большой изгиб определяется матрицей Гессе  $H$ . Для проверки подойдет  $H$  размера  $2 \times 2$ .

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Пусть  $\text{Tr}(H)$  — след матрицы, а  $\text{Det}(H)$  — её определитель.

$$\begin{aligned} \text{Tr}(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ \text{Det}(H) &= D_{xx} D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned}$$

Пусть  $r$  — отношение большего изгиба к меньшему,  $\alpha = r\beta$ . Тогда

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}$$

и точка рассматривается дальше, если

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r + 1)^2}{r}$$

# Scale-Invariant Feature Transform

- **Назначение ориентации.** Каждой ключевой точке назначается одна или более ориентация, основываясь на направлениях градиентов в локальном изображении. Для изображения  $L(x, y)$  с масштабом  $\sigma$  предварительно вычисляются на основе разности пикселей величина градиента  $m(x, y)$  и ориентация  $\theta(x, y)$

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y)))$$

Вычисление величины и направления для градиента делается для каждого пикселя в окрестности ключевой точки в размытом по Гауссу изображении  $L$ . Формируется гистограмма направлений с 36 областями, каждая из которых покрывает 10 градусов. Каждая точка в окружающем окне добавляется в область гистограммы, взвешенная по величине градиента. Пики в этой гистограмме соответствуют доминирующим направлениям.

Как только гистограмма заполнена, направления, соответствующие самым высоким пикам и локальным пикам, которые в пределах 80 % от самых высоких пиков, назначаются ключевой точке. В случае назначения нескольких направлений создаётся дополнительная ключевая точка, имеющая то же местоположение и масштаб, что и оригинальная точка для каждого дополнительного направления.

# Scale-Invariant Feature Transform

- **Построение дескрипторов.** Предыдущие шаги находят местоположения ключевых точек в конкретных масштабах и назначают им ориентацию. Теперь мы хотим вычислить вектор дескрипторов для каждой ключевой точки.

Создаётся набор гистограмм направлений на  $4 \times 4$  соседних пикселях с 8 областями в каждой. Эти гистограммы вычисляются из значений величины и ориентации элементов в области  $16 \times 16$  вокруг ключевой точки, так что каждая гистограмма содержит элементы из  $4 \times 4$  подобласти исходной области соседства. Величины далее взвешиваются функцией Гаусса. Дескриптор затем становится вектором всех значений этих гистограмм. Поскольку имеется  $4 \times 4 = 16$  гистограмм с 8 областями в каждой, вектор имеет 128 элементов.

Полученный вектор нормализуется до единичной длины, после чего все его компоненты, значение которых больше 0.2, урезаются до значения 0.2 и затем дескриптор нормализуется ещё раз. В таком виде дескриптор готов к использованию.

# СПИСОК ИСТОЧНИКОВ

- [https://ru.wikipedia.org/wiki/Компьютерное зрение](https://ru.wikipedia.org/wiki/Компьютерное_зрение)
- [https://en.wikipedia.org/wiki/Feature detection \(computer vision\)](https://en.wikipedia.org/wiki/Feature_detection_(computer_vision))
- <https://www.learnopencv.com/image-recognition-and-object-detection-part1/>
- <https://www.learnopencv.com/histogram-of-oriented-gradients/>
- <https://medium.com/the-downlinq/histogram-of-oriented-gradients-hog-heading-classification-a92d1cf5b3cc>
- <https://towardsdatascience.com/sift-scale-invariant-feature-transform-c7233dc60f37>
- [https://iupr.ru/domains\\_data/files/zurnal\\_43/Ledyaev%20V.P.%20-1%20%20\(Informacionnye%20i%20kommunikativnye%20tehnologii\).pdf](https://iupr.ru/domains_data/files/zurnal_43/Ledyaev%20V.P.%20-1%20%20(Informacionnye%20i%20kommunikativnye%20tehnologii).pdf)



# Список вопросов

- Приведите примеры алгоритмов получения признаков из картинок.
- Дайте краткое описание **HOG**.
- Как можно классифицировать признаки на изображении?