

УДК 004

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ СЕГМЕНТАЦИИ ВИДЕОПОТОКА В КВАЗИРЕАЛЬНОМ ВРЕМЕНИ

Власов Глеб Геннадьевич

Городничев Михаил Геннадьевич

МТУСИ, к.т.н., доцент кафедры МКиИТ

Аннотация: В статье рассматривается механизм реализации интеллектуальной системы сегментации видеопотока в квазиреальном времени. Описывается процесс первичной обработки собранного датасета и его дальнейший анализ, структура свёрточной нейронной сети, используемая для решения данной проблемы.

Ключевые слова: Искусственный интеллект, нейронные сети, компьютерное зрение, сегментация видеопотока.

DEVELOPMENT OF AN INTELLIGENT SYSTEM FOR SEGMENTING A VIDEO FLOW IN QUASI-REAL TIME

Vlasov Gleb Gennadyevich

Gorodnichev Mikhail Gennadyevich

Abstract: The article discusses the realization mechanism of intelligent system used for segmentation of video stream in quasi-real time. It describes the process of initial processing of the collected dataset and its further analysis, the structure of the convoluted neural network used to solve this problem.

Key words: Artificial intelligence, neural networks, computer vision, video-flow segmentation.

С начала 20-го века идёт бурный процесс автомобилизации, в связи чем математики, физики и инженеры пытаются разработать модели, которые смогут обеспечить безопасность и оптимальность передвижения транспортных средств. Однако теоретические модели в реальной жизни работают достаточно плохо, так как за рулём находится человек, чьё поведение непредсказуемо и зависит от многих факторов, поэтому в

последнее десятилетие многие компании разрабатывают беспилотный транспортные средства на основе методов глубокого обучения. Данная работа посвящена одному из важных модулей БТС такому как сегментация удержания полосы и детектированию нарушения динамического габарита, которое обеспечивает безопасность передвижения на дорогах общего пользования.

1. Выбор данных для обучения

Для реализации данного проекта использовалась база данных Udacity Self-Driving car Nanodegree Program, которая содержит 16 048 RGB изображений с разрешением 64×64 пикселей.



Рис. 1. «Пример негативных изображений»



Рис. 2. «Пример позитивных изображений»

В изображениях класса «транспортные средства» содержится 8048 фотографий различных автомобилей с разных ракурсов. Эти изображения содержат множество шумов, таких как тени, световые блики, присутствуют просто размытые изображения. Во втором классе изображений «части дорожного полотна» находятся фотографии различных объектов дорожного полотна. Для задачи классификации несбалансированная выборка как положительных, так и отрицательных обучающих примеров оказывает огромное влияние на производительность сети.

2. Расширение данных (аугментация)

Свободно распространяемые, обучающие базы изображений отлично подходят для учебных работ, но почти всегда их нельзя применять в реальных

задачах. Приём расширения данных позволяет создавать дополнительные обучающие данные из уже доступных путём изменения изображений различными случайными преобразованиями. В идеале модель на этапе обучения не должна видеть одно и то же изображение дважды. Существует несколько способов подготовки выборки изображений для создания систем распознавания.

Обучающие наборы данных из естественных изображений, которые создаются на основе реальных данных. Создание таких датасетов состоит из следующих этапов:

1. Сбор данных (например снятие видеопотока с видеорегистратора машины);
2. Фильтрация - проверка изображений на ряд необходимых требований;
3. Написание разметчика датасета или настройка готового.
4. Разметка (выделение интересующих объектов);
5. Присвоение меток;

Этот алгоритм требует значительных временных затрат, но с другой стороны по результатам обучения системы на таких данных можно узнать об её эффективности в реальных условиях.

Обучающие наборы данных из искусственных изображений. Достаточно взять несколько реальных изображений и с помощью разных манипуляций сгенерировать необходимое число примеров для обучения. Существуют следующие виды изменений:

1. Изменение яркости/цвета;
2. Замена фона (или других частей изображения);
3. Геометрические (аффинные, проективные и др.);
4. Уникальные преобразования, применяемые для определённой задачи (световые блики, размытия, шумы, добавление капель дождя, снега)

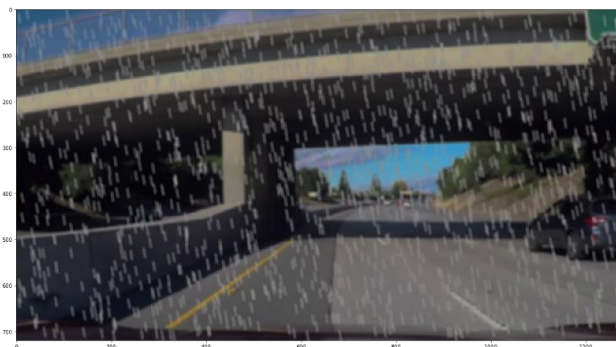


Рис. 3. «Добавление эффекта искусственного дождя»



Рис. 4. «Добавление эффекта искусственного снега»

Этот подход помогает значительно экономить время, но достаточно сложно предсказать эффективность данного метода в реальных условиях.

3. Описание нейронной сети

Свёрточный слой

Главной составляющей свёрточной нейронной сети является свёртка. Эти слои содержат фильтры для каждого канала изображения и ядро свёртки, которое является матрицей $N \times N$.

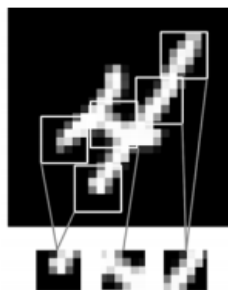


Рис. 5. «Примеры локальных шаблонов»

Свёрточные нейронные сети обладают несколькими важными свойствами:

1) Свёрточные ядра, которые они изучают являются инвариантными в отношении переноса. После изучения какого-либо ядра в центре изображения свёрточная нейронная сеть сможет найти его в любом другом месте. Эта способность повышает эффективность свёрточных сетей в задачах компьютерного зрения (так как видимый мир по сути является инвариантным в плане переноса): этим сетям требуется намного меньше обучающих образцов для получения результата.

2) Способны изучать пространственные иерархии ядер. Первый свёрточный слой будет изучать небольшие локальные признаки, второй - более крупные, состоящие из признаков, возвращаемых первым слоем, и т.д.

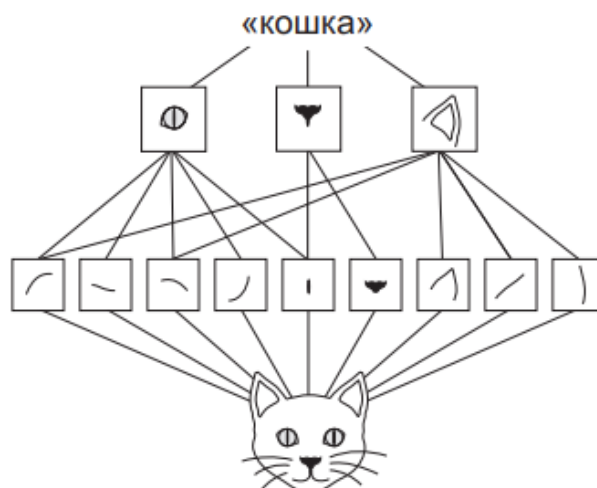


Рис. 6. «Пример второго свойства»

Свёртка двигает окно размером 3 x 3 или 5 трёхмерной входной карте признаков, останавливается в каждой возможной позиции и извлекает трёхмерный шаблон окружающих признаков. Затем каждый из полученных шаблонов преобразуется в одномерный вектор. Затем все эти векторы формируют трёхмерную выходную карту признаков.

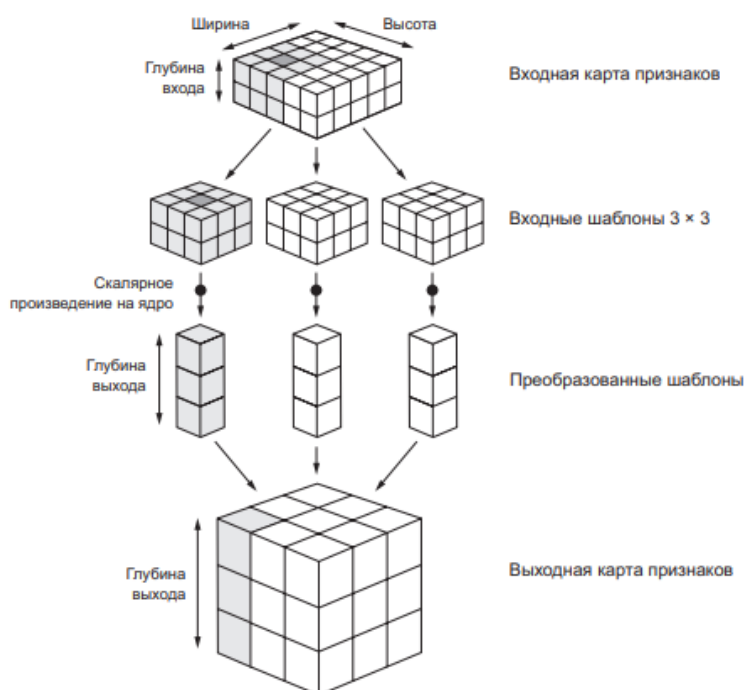


Рис. 7. «Принцип работы свёрточного слоя»

Слой пулинга

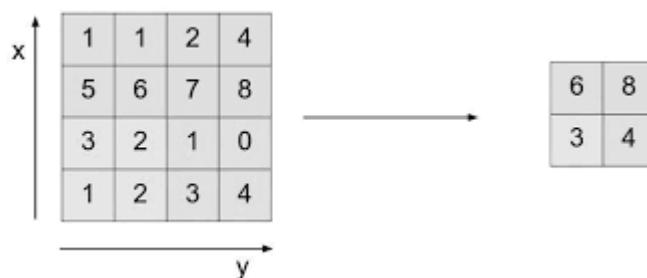


Рис. 8. «Пример пулинга»

Операция пулинга является нелинейным уплотнением карты признаков, из входной карты признаков извлекается ядро, и из него выбирается максимальное значение для каждого канала. Уменьшение разрешения используется для уменьшения количества коэффициентов в карте признаков для обработки. Пулинг даёт возможность сделать инвариантным представление относительно малых переносов входа, выделяя основные признаки. Инвариантность переноса говорит о том, что при сдвиге входа на некоторую величину, значение большинства задействованных пулингом выходов не изменится. Обычно слой пулинга размещают после слоя свёртки, но перед новой свёрткой.

Сеть с прямым распространением

После свёрточной нейронной сети идёт полносвязная сеть, которая выполняет роль связи для абстрактных признаков. Как правило такая сеть представляет собой многослойный персептрон. Каждый нейрон предыдущего слоя этой сети связан со всеми нейронами следующего слоя.

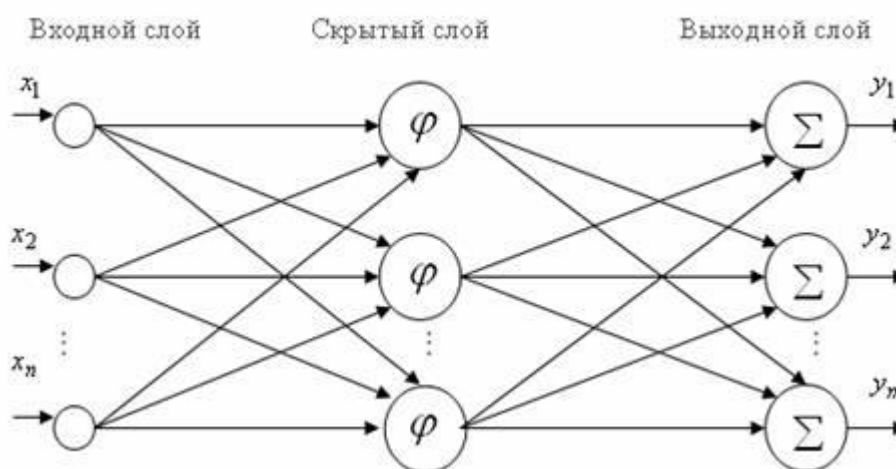


Рис. 9. «нейронная сеть прямого распространения»

Функция активации

Огромную роль в свёрточных нейронных сетях играет функция активации. В моём проекте используется функция активации ReLU. Она возвращает значение x , если x положительно, и 0 в других случаях.

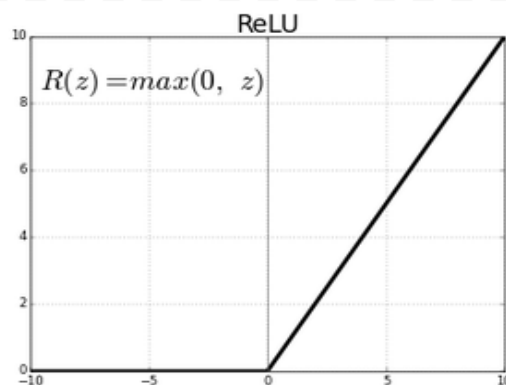


Рис. 10. «Рисунок функции активации ReLu»

Преимущества ReLU:

1) Для вычисления гиперболического тангенса или сигмоиды требуются ресурсоёмкие операции, такие как возведение в степень, а ReLu можно реализовать с помощью простого порогового преобразования матрицы.

2) ReLu значительно повышает скорость сходимости стохастического градиентного спуска (этот вид градиентного спуска часто используется в задачах компьютерного зрения).

Одним из недостатков функции ретификации является то, что она может привести к такому обновлению весов, что данный нейрон никогда больше не активируется. В таком случае, градиент, проходящий через этот нейрон всегда будет равен нулю. Обычно эта проблема решается настройкой скорости обучения.

Существует несколько разных блоков линейно ретификации:

Leaky ReLU (с «утечкой») - является одной из попыток решить указанный выше недостаток выхода из строя обычных ReLu. Функция LReLU имеет вид $f(x) = ax$ при $x < 0$ и $f(x) = x$ при $x \geq 0$;

Parametric ReLU - авторы утверждают, что использование этой функции позволило превзойти уровень человека при распознавании изображений базы данных ImageNet;

Randomize ReLU - позволяет уменьшить переобучение благодаря характерному элементу случайности;

Способы регуляризации

Регуляризация - это методы контроля ёмкости нейронной сети, которые позволяют предотвратить переобучение, математическая операция, ограничивающая запоминание, которая способствует обобщённому обучению.

В своём проекте я использовал Dropout это простой, но очень эффективный метод. Его суть заключается в том, что во время обучения некоторые узлы, которые попадают в полносвязный слой случайным образом отбрасываются. Удаление узла означает, что значение его веса устанавливается на 0. Такие узлы выбираются случайно на каждом этапе градиентного спуска.

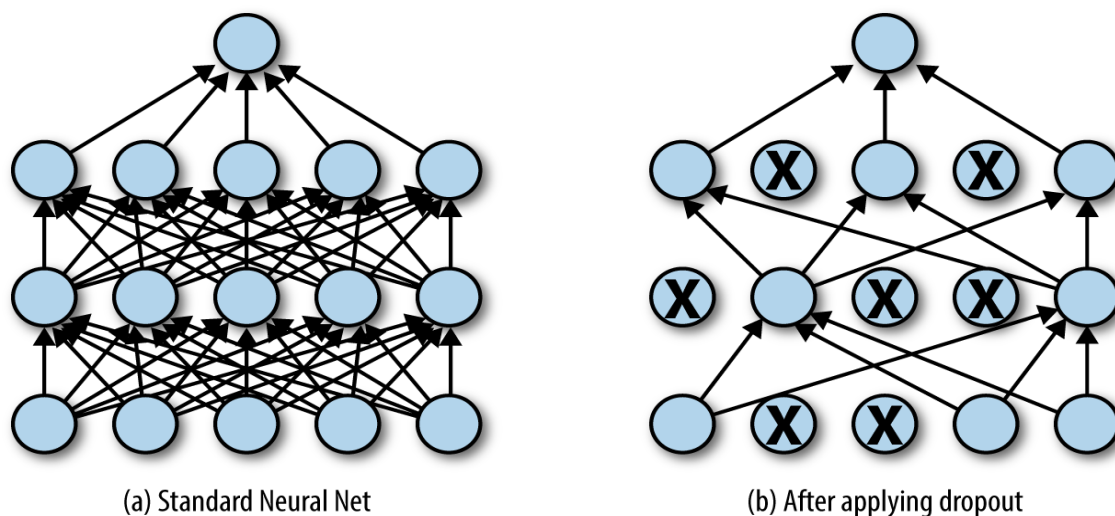


Рис. 11. «Отображение принципа работы Dropout»

4. Заключение

В результате работы реализована программа по сегментированию видеопотока, детектированию транспортных средств и определению дорожной полосы. Изучены методы первичной обработки датасета, расширения данных. Построена структура свёрточной нейронной сети для поставленной задачи.

Список литературы

1. Deep Learning with Python François Chollet - by Manning Publications .
2. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification - Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun.
3. A Simplified Computer Vision System for Road Surface Inspection and Maintenance Marcos Quintana, Juan Torres, and José Manuel Menéndez.
4. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In BMVC, 2014.