

Разработка оптико-электронной системы видеонаблюдения с возможностью обнаружения движения и последующей классификацией объектов и стала основной темой этого доклада.

Оптико-электронная система видеонаблюдения в общем случае представляет собой ip-камеру, работающую на базе связки FPGA и основного процессора с ARM архитектурой. Предобработка и фильтрация входных данных будет выполняться на FPGA, сокращая время последующего анализа и обработки видеопотока на ARM. Перед системой видео наблюдения стоит проблема решения следующих задач видеоаналитики:

1. Детектирование движения в кадре

Это несложная с точки зрения вычислений задача, заключающаяся в анализе двух соседних кадров, полученных в разное время, и нахождении между ними разницы свыше заданных параметров. Такими параметрами в простейших случаях являются:

- размер движущегося объекта.

Который задается в процентах от размера всего кадра и позволяет обнаружить тот размер объекта, при движении которого необходимо начинать запись.

- чувствительность.

Контраст движущегося объекта. То как движущийся объект выделяется на общем фоне. Фильтрация неконтрастных объектов позволяет исключить ложные срабатывания детектора, таких как при раскачивании на ветру деревьев, веток, травы.

2. Классификация и распознавание движущихся объектов

Является более сложной задачей, требующей значительных затрат вычислительной мощности. Для оптимизации этих процессов в настоящее время все чаще используются различные нейросети, которые обучаются для распознавания конкретных объектов.

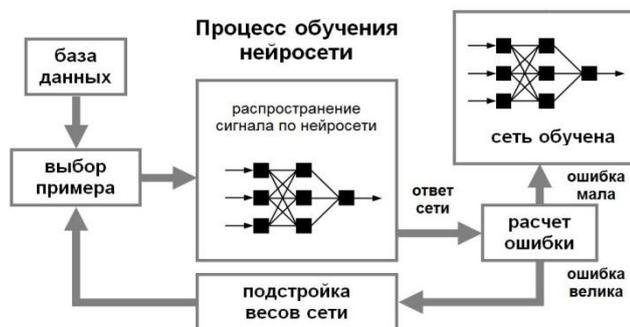


Рис. 1 – Обучение нейронной сети

Для реализации этого проекта была выбрана нейронная сеть GoogleNet. Она обучалась на заранее сформированной выборке, подобранной специально чтобы обучить нейросеть на классификацию объектов на людей и транспортные средства. Общий принцип обучения нейронной сети представлен на рисунке 1

Использование таких систем видеонаблюдения позволит с применением ограниченных ресурсов основного процессора и FPGA получать требуемый результат и обеспечивать безопасность и наблюдение за объектом. Использование нейронных сетей делает эту систему весьма гибкой и настраиваемой, т.к. переобучение нейросети для обнаружения объектов другого типа не займет много сил и времени.

Список использованных источников:

1. OpenCV [Электронный ресурс]. – Электронные данные – Режим доступа: <https://opencv.org>
2. Медведев В. С., Потемкин В. Г. Нейронные сети. MATLAB 6. М.: ДИАЛОГ-МИФИ, 2001. 630 с.
3. Demuth H., Beale M. Neural Network Toolbox. For Use with MATLAB. The MathWorks Inc. 1992-2000.

КЛАССИФИКАЦИЯ ЗВУКОВ С ПОМОЩЬЮ SEMI-SUPERVISED ПОДХОДА

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Жук И.Н.

Петровский А.А. – д.т.н., профессор

Данные-это ключ для успешного обучения глубоких нейронных сетей, однако для классического обучения с учителем данные должны быть размечены. Для того, чтобы использовать неразмеченные данные были разработаны различные методы semi-supervised обучения. Один из таких методов называется лестничная сеть [2]. Если применить данный подход к задаче классификации звуков окружающей среды, можно добиться значительного повышения результатов классификации.

Лестничная архитектура состоит из трёх нейронных сетей: “чистого” кодировщика, “зашумленного” кодировщика и декодировщика. При этом оба кодировщика по сути являются одной сетью, т.к. они разделяют весовые коэффициенты, но в “зашумленной” версии кодировщика после каждого слоя добавляется нормально-распределенный шум $N(0, \sigma^2)$. Декодировщик преобразует информацию таким же образом, как и кодировщик, но в другую сторону для того, чтобы на выходе получились предсказания данных, которые были поданы на кодировщик.

Лестничная архитектура одновременно обучается двум задачам. Первая задача – это обучение классификации размеченного набора данных. В ней принимает участие только “зашумленный” кодировщик и обучение происходит по классической схеме обучения классификатора. Вторая задача – это подавление шума. Именно благодаря обучению этой задачи выполняется semi-supervised подход, потому что нам не нужны метки для процесса обучения. Соответственно для этой задачи используются неразмеченные данные. В этой задаче используются все три сети, «зашумленный» кодировщик и декодировщик для расчета предсказания входных данных, а «чистый» кодировщик для расчета функции потерь при подавлении шума.

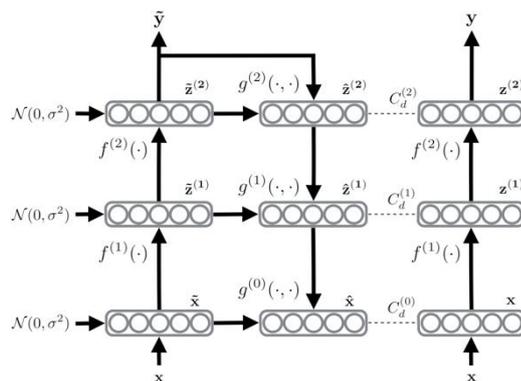


Рис. 1 – Архитектура лестничной сети для классификации

Где x – входные данные, y – предсказания классов, \tilde{x} – зашумленные входные данные, \tilde{y} – зашумленные предсказания классов, $C_d^{(n)}$ – ошибка подавления шума, $f^{(n)}(\cdot)$ – батч-нормализация и функция активации (ReLU), $g^{(n)}(\cdot, \cdot)$ – функция шумодав с обучаемыми параметрами.

Исследования строились на основе двух наборов данных. Набор данных ESC 10 [1] использовался как основной, размеченный набор для классификации. Набор данных UrbanSound8K использовался как неразмеченный, второстепенный набор. Для унификации входных данных все записи были приведены к длине 5с и затем были рассчитаны логарифмические мел-частотные спектрограммы с перекрытием 50%. Итоговое разрешение входных данных 128x431.

Экспериментальные результаты:

Характеристический вектор	Классификатор	Доля правильных ответов
Мел-кепстральные коэффициенты	“случайный лес”	0.73 ± 0.078
спектрограмма	Сверточная нейронная сеть	0.81 ± 0.044
спектрограмма	лестничная архитектура	0.91 ± 0.025
	человек	0.96 ± 0.014

Список использованных источников:

1. K. J. Piczak. ESC: Dataset for Environmental Sound Classification // In Proceedings of the 23rd ACM international conference on Multimedia, pp. 1015-1018, ACM, 2015.
2. Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. Semi-supervised learning with ladder networks // NIPS, 2015.