

Нейросетевой детектор госномера и его портирование на ПЛИС и GPU для устройств Элвис-НеоТек

Куратор модельно-ориентированного проектирования от ЦИТМ Экспонента подключился к проекту заказчика на этапе модернизации существующего изделия. Заказчик искал возможность программно повысить точность распознавания госномеров и перенести больше функций по определению госномера непосредственно на камеру, снизив тем самым ошибки и требования к каналам передачи данных и зависимость от серверных мощностей, используемых ранее для части алгоритма распознавания. Это позволило бы провести относительно дешевый апгрейд существующих изделий уже работающих на автомагистралях, повысив их производительность и точность. Была известна архитектура устройства и его ЭКБ, существовали некоторые базы размеченных и неразмеченных изображений. Ключевой нейросетевой алгоритм заказчика, разработанный на языке C, выдавал недостаточную точность, был крайне сложен в модификации и поддержке. Команда разработчиков алгоритма уже не была связана с командой, выполняющей модернизацию изделия.

Ключевые результаты:

1. Вместо переноса существующей сети, написанной на си, был проведен быстрый реинжиниринг алгоритма с помощью построения нейросети в матлаб и ее обучения на существующих данных. Это сэкономило время проекта оценочно в 100 раз, т.к. нейросеть на си создавалась заказчиком в течение 2-х лет и на нее было потрачено так много усилий, что инженеры заказчика не верили в возможность быстрого реинжиниринга этого алгоритма. Реинжиниринг же сети занял считанные дни. Это характерно для проектов, в которых большое количество ресурсов тратится на разработку - конечные инженеры часто встают в защитную позицию по отношению к затраченным ранее усилиям, что иногда снижает скорость развития предприятия.
2. Реинжиниринг нейросетевого алгоритма средствами матлаб открыл упрощенный способ тестирования и отладки качества алгоритма. Был создан фреймворк тестирования и отладки алгоритма по живым данным, с автоматизацией подсчета точности и с возможностью отладки как эталонного алгоритма в моделях, так и портированного алгоритма в железе (Processor-in-the-Loop). Сама автоматизация тестирования алгоритма с помощью симулинка снизила временные затраты на верификацию оценочно на порядок.
3. В рекордные сроки (2 дня) удалось создать и обучить алгоритм более высокой точности 98%, что на 3% пункта выше по сравнению с алгоритмом разработанным заказчиком ранее в течение 2х лет.
4. Часть алгоритма в камере работала на ПЛИС, а часть на GPU. С помощью системной модели всех алгоритмов появилась возможность принять верные технические решения по поводу того какая часть алгоритма на какой из ЭКБ должна работать. Поскольку при Модельно-Ориентированном проектировании разработчик не сильно ограничен в вопросе переноса алгоритма между ЭКБ с помощью автоматических генераторов кода, появилась возможность алгоритмической оптимизации проекта на системном уровне.

5. Старая ЭКБ как ПЛИС, так и GPU была одним из препятствий проекта, поскольку генераторы кода не поддерживали “из коробки” данную ЭКБ. Эту проблему удалось разрешить благодаря глубокой компетенции куратора МОП от ЦИТМ Экспонента по глубокой настройке генераторов кода. В проекте удалось настроить сквозную верификацию как на моделях, так и на аппаратуре (FIL - FPGA-in-the-Loop, PIL-Processor-In-the-Loop)
6. Проект длился с перерывами несколько месяцев. Всего на проект было затрачено 10 рабочих дней куратора МОП от Экспоненты и порядка 20 рабочих дней специалиста Заказчика. За это время специалисту заказчика были переданы процессные компетенции по МОП, а также настроены необходимые фреймворки для самостоятельного масштабирования полученного опыта силами заказчика на подобных проектах.

Ключевые проблемы и особенности проекта по курированию МОП для Элвис-НеоТек:

1. Для идентификации положения номера в MATLAB была построена модель, эквивалентная эталонной с последующим получением HDL-кода, работающего с частотой, даже превышающей требуемую. После подтверждения качества модели (для чего использовалось модельно-ориентированное проектирование) было получено IP-ядро. Загруженное в эталонный проект оно подтвердило свою работоспособность, показав результаты «на железе» такие же, какие были получены при проверке модели, не требуя дополнительных доработок и отладок. Подтверждена ключевая особенность Модельно-Ориентированного проектирования - возможность проведения основной итеративной отладки на моделях без железа, с последующим практически автоматическим подтверждением характеристик в аппаратуре.
2. В MATLAB была написана сверточная нейронная сеть. Результаты тестирования показали, что, после обучения на наборе данных, уже размеченных заказчиком, точность распознавания ею номеров составила 98%.
3. Сгенерирован CUDA код для последующего портирования на платформу Tegra. Была разработана операция оптимизации кода, которая позволила обеспечить обработку информации со скоростью до 500 символов в секунду на достаточно старой и малоресурсной ЭКБ.
4. Проведена и настроена автоматизированная верификация CUDA кода на устаревшей ЭКБ Tegra. С помощью тестовой обвязки процессора и Simulink моделей был обеспечен процесс сквозной верификации.
5. Посредством изменения параметров слоев оптимизирована архитектура нейронной сети, которая изначально оказалась достаточно требовательна к ресурсам. В результате удалось упростить сложность архитектуры нейронной сети в 10 раз, потеряв всего лишь один процент в точности распознавания.
6. В качестве изначальной точки сравнения был взят си код проекта подлежащий модернизации.

Выводы и рекомендации от Эксперта по МОП:

Приглашение эксперта МОП в команду позволяет сразу эффективно применить к текущему проекту подходящие инструменты, опыт использования которыми формируется у инженеров в течение года или более. Использование этих инструментов позволяет значительно сократить время разработки, а в командной работе с экспертом можно ожидать предсказуемый положительный результат, что очень полезно для компании заказчика в плане бизнеса.

Используемые инструменты:

[MATLAB](#)

[Simulink](#)

[Stateflow](#)

[GPU Coder](#)

[HDL Coder](#)

[Embedded Coder](#)

[Deep Learning Toolbox](#)

[DSP System Toolbox](#)

[Image Processing Toolbox](#)

[Computer Vision Toolbox](#)

[Vision HDL Toolbox](#)



Контакты

exponenta.ru

E-mail: info@exponenta.ru

Тел.: +7 (495) 009 65 85

Адрес: 115088 г. Москва,

2-й Южнопортовый проезд, д. 31, стр. 4

