

# Использование OpenVINO Toolkit для распределенной обработки информации

Выполнил: Киреев Константин Александрович, гр. 8383

Руководитель: Черниченко Дмитрий Александрович, к.т.н., доцент

# Цель и задачи

Актуальность исследования обусловлена:

- спросом на системы глубокого обучения
- высоким потреблением ресурсов при использовании

**Цель:** проанализировать эффективность оптимизации OpenVINO в случае распределенной обработки данных

**Задачи:**

1. Изучить функционал OpenVINO
2. Разработать демо-приложение
3. Проанализировать эффективность Multi-device плагина
4. Проанализировать эффективность гетерогенного плагина

# Задача 1. Функционал OpenVINO (1)

Основной функционал OpenVINO:

1. Model Optimizer – оптимизация моделей и конвертация в IR-формат
  2. Inference Engine – запуск модели на целевом устройстве/устройствах
- Дополнительные утилиты для оптимизации и работы с данными

## Задача 1. Функционал OpenVINO (2)

Основные оптимизации с помощью которых OpenVINO ускоряет работу нейронной сети:

- Оптимизации топологии сети
- Понижение точности
- Распределенная обработка данных

## Задача 2. Разработка демо-приложения (1)

Цели написания приложения:

1. Оценка сложности интеграции OpenVINO
2. Оценка производительности при использовании нескольких устройств

Детали эксперимента:

- Модель НС: взятая из открытого источника обученная модель MobileNet
- Метрики: задержка инференса и FPS
- Входные данные: видео поток с камеры

## Задача 2. Разработка демо-приложения (2)

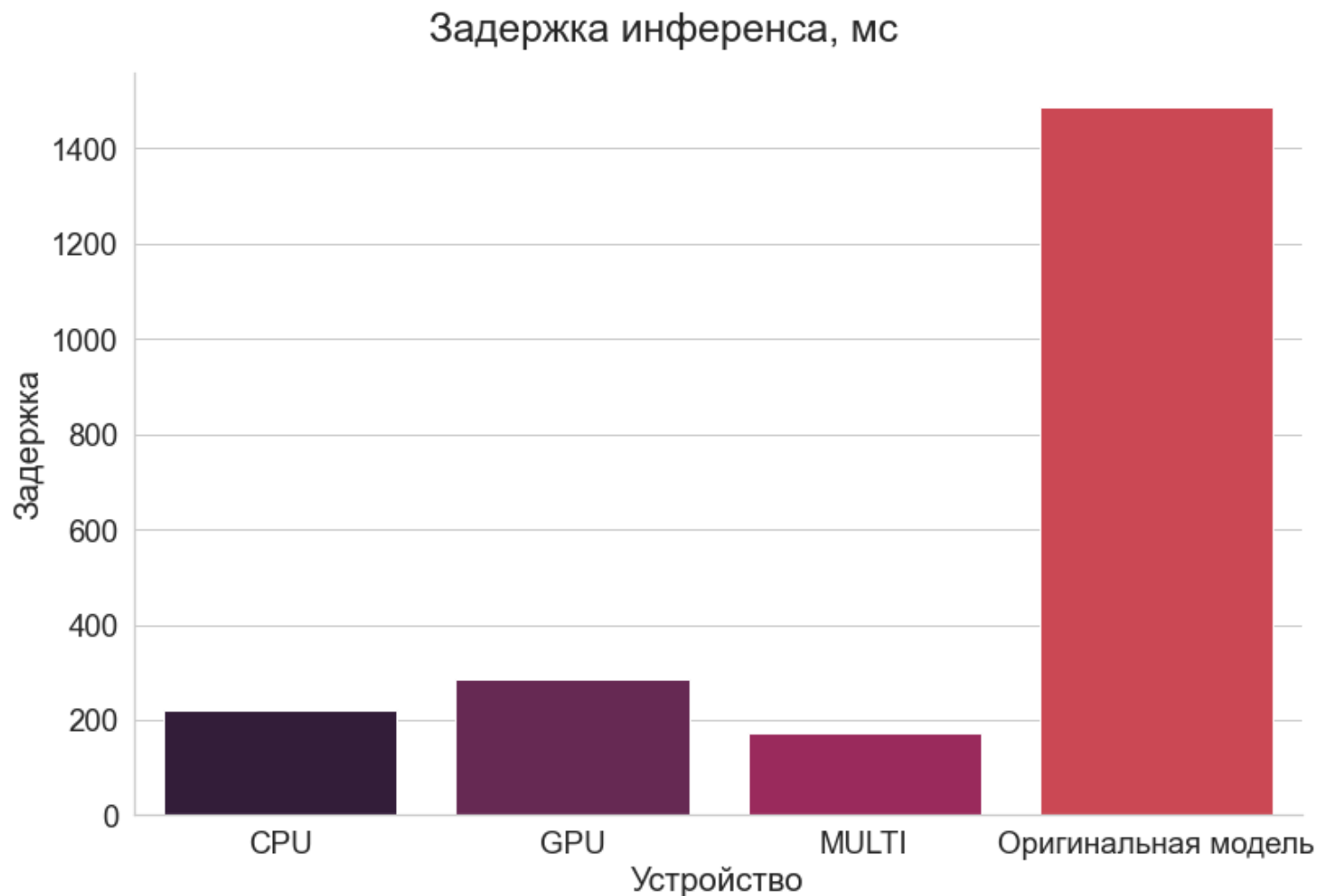
Основной функционал приложения:

- Распознавание поз на видео
- Изменение устройств исполнения

Этапы работы приложения:

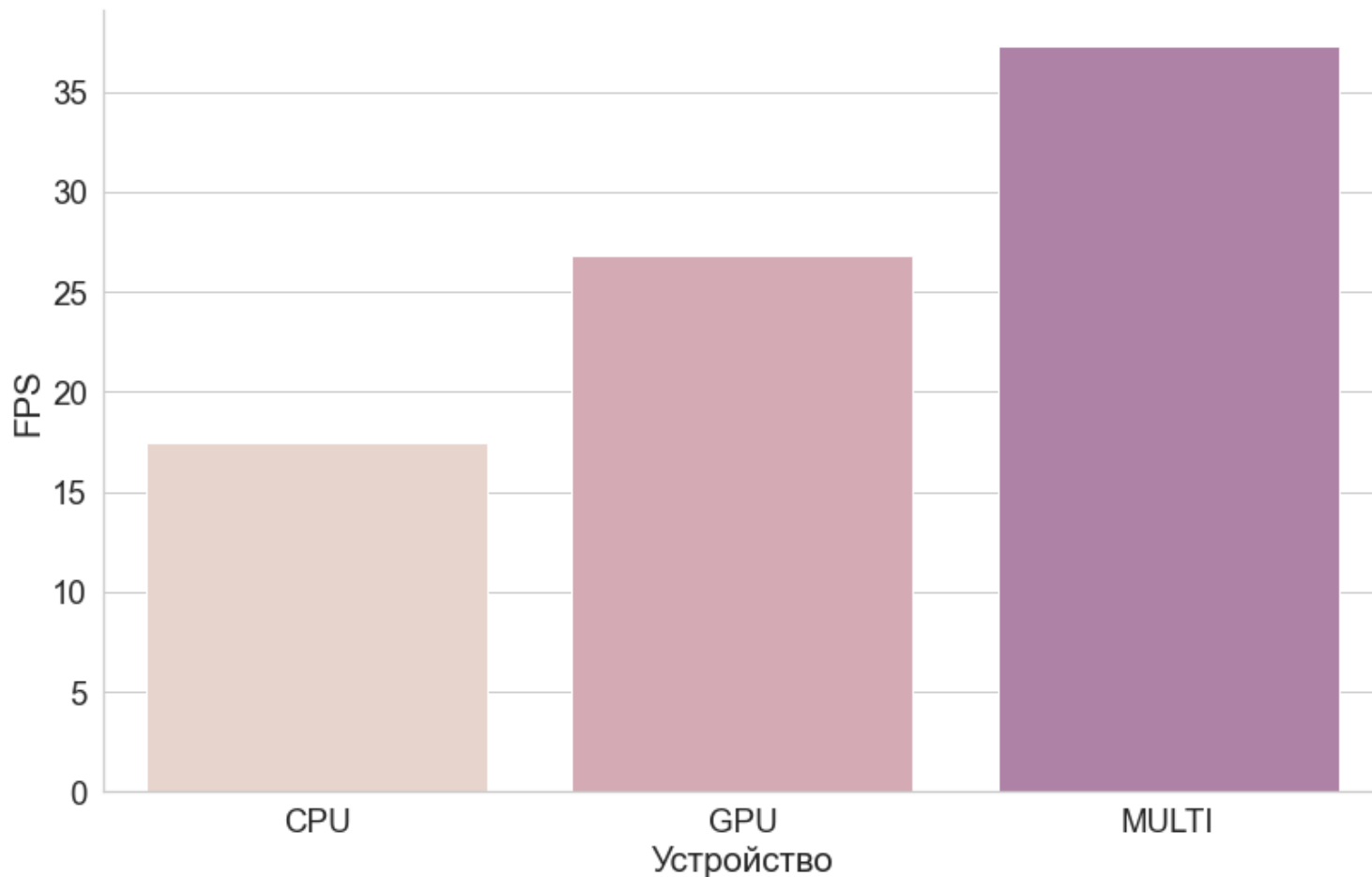
1. Загрузка модели
2. Загрузка входного видео
3. Предобработка
4. Запуск НС
5. Отрисовка поз и вывод конечного видео

# Задача 3. Анализ эффективности Multi-device плагина (1)



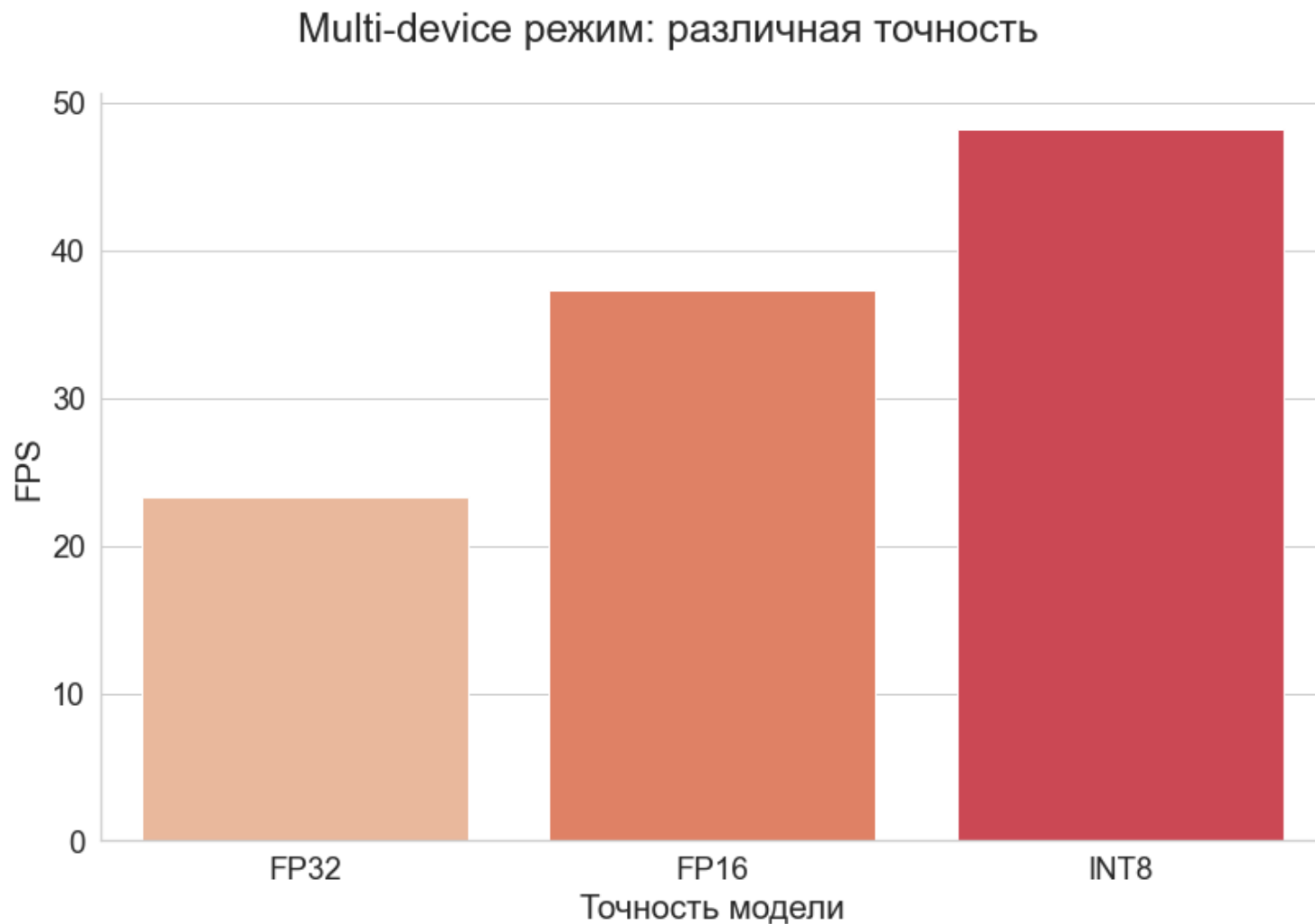
## Задача 3. Анализ эффективности Multi-device плагина (2)

FP16: Пропускная способность (FPS)

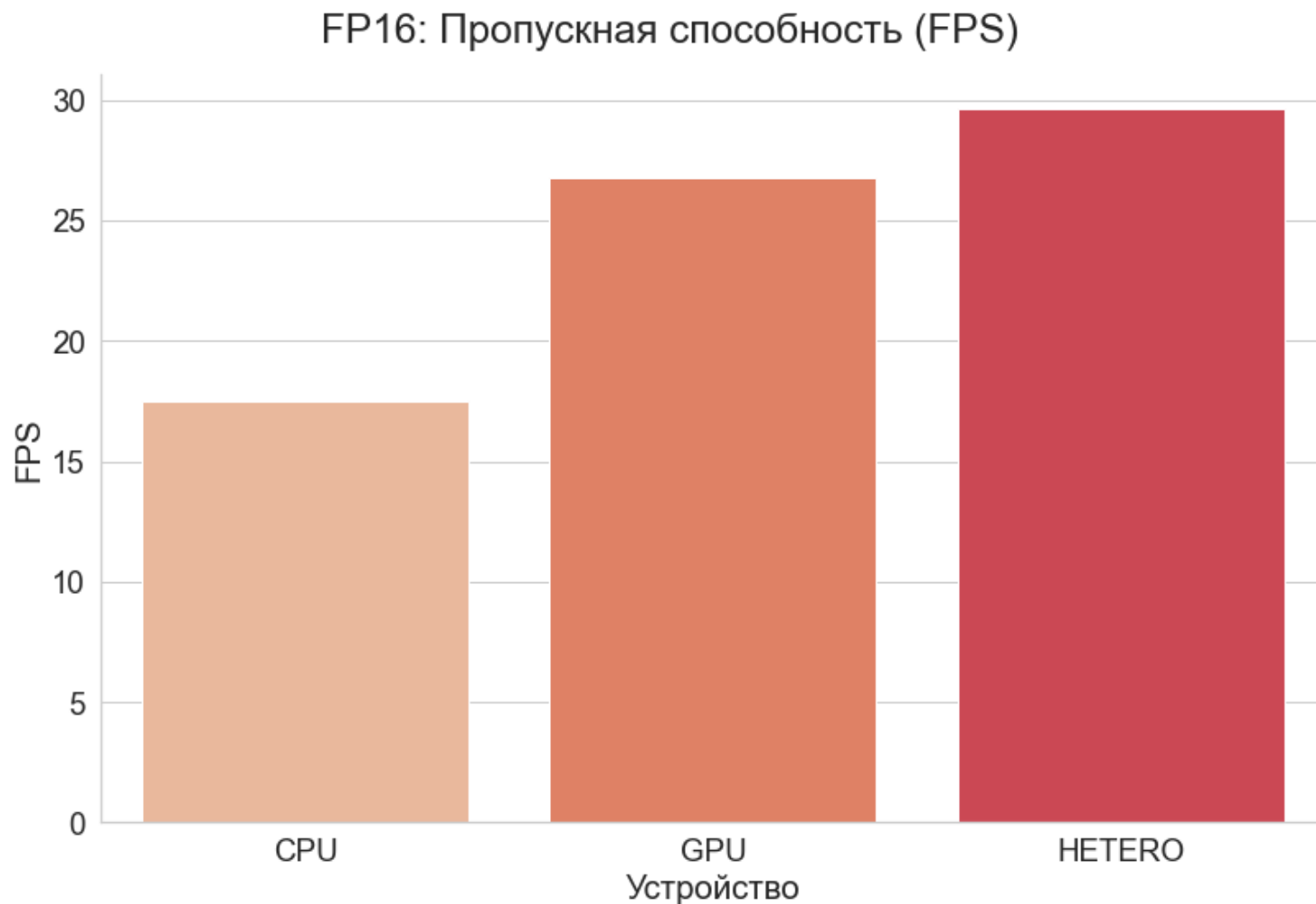




## Задача 3. Анализ эффективности Multi-device плагина (3)



## Задача 4. Анализ эффективности гетерогенного плагина



# Апробация работы

Код приложения размещен в интернет ресурсе github:

- [https://github.com/GandhiKK/OpenVINO\\_hetero](https://github.com/GandhiKK/OpenVINO_hetero)

# Заключение

- Изучен функционал OpenVINO для распределенной обработки данных
- Разработано приложение для демонстрации функционала OpenVINO и оценки производительности
- Исследование показало, что использование Multi-device плагина дает значительный прирост производительности в два раза
- Использование гетерогенного плагина дало прирост 18%

## Дальнейшие направления исследований:

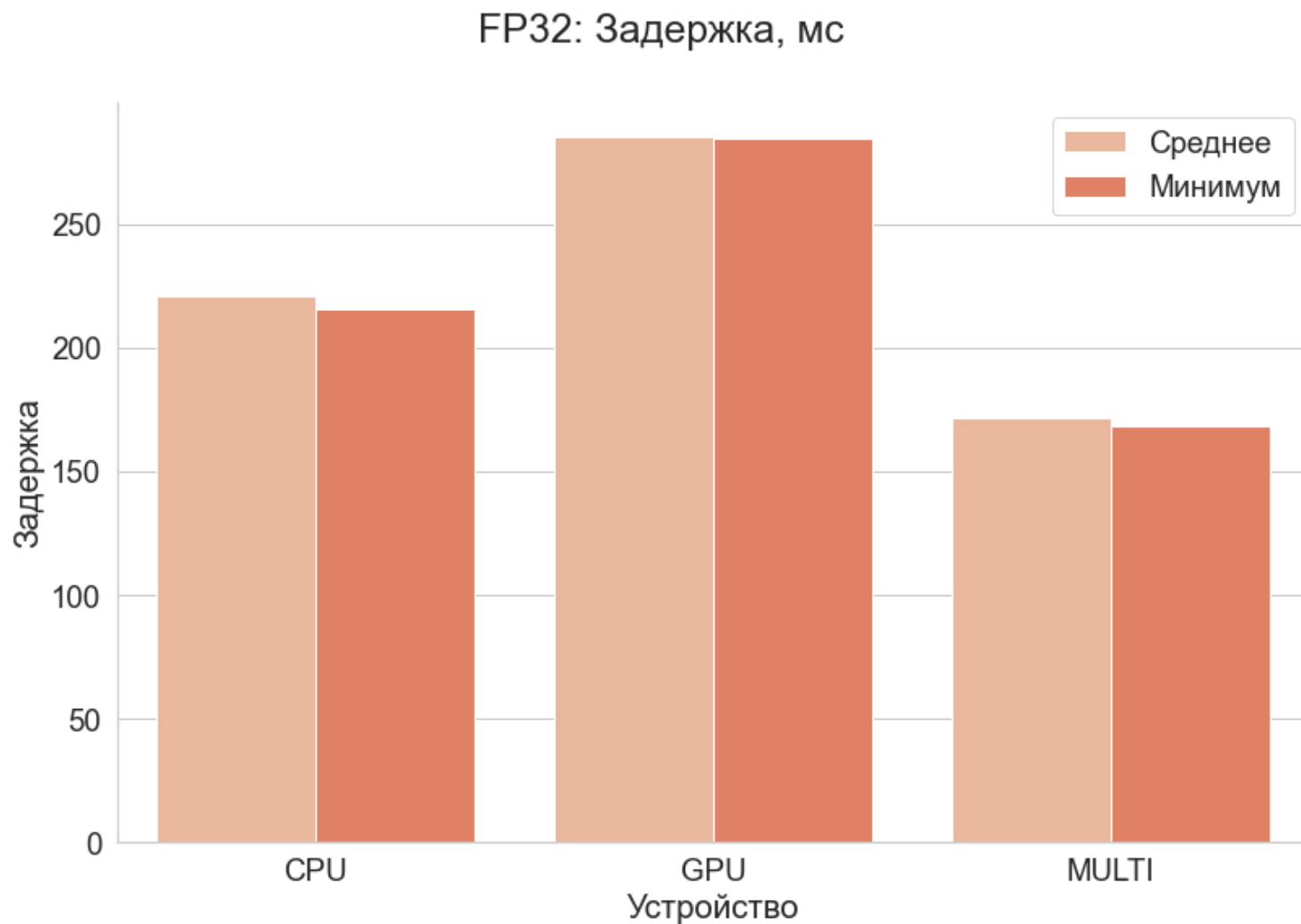
- Эксперименты на других устройствах компании Intel
- Улучшение производительности гетерогенного режима

Запасные слайды

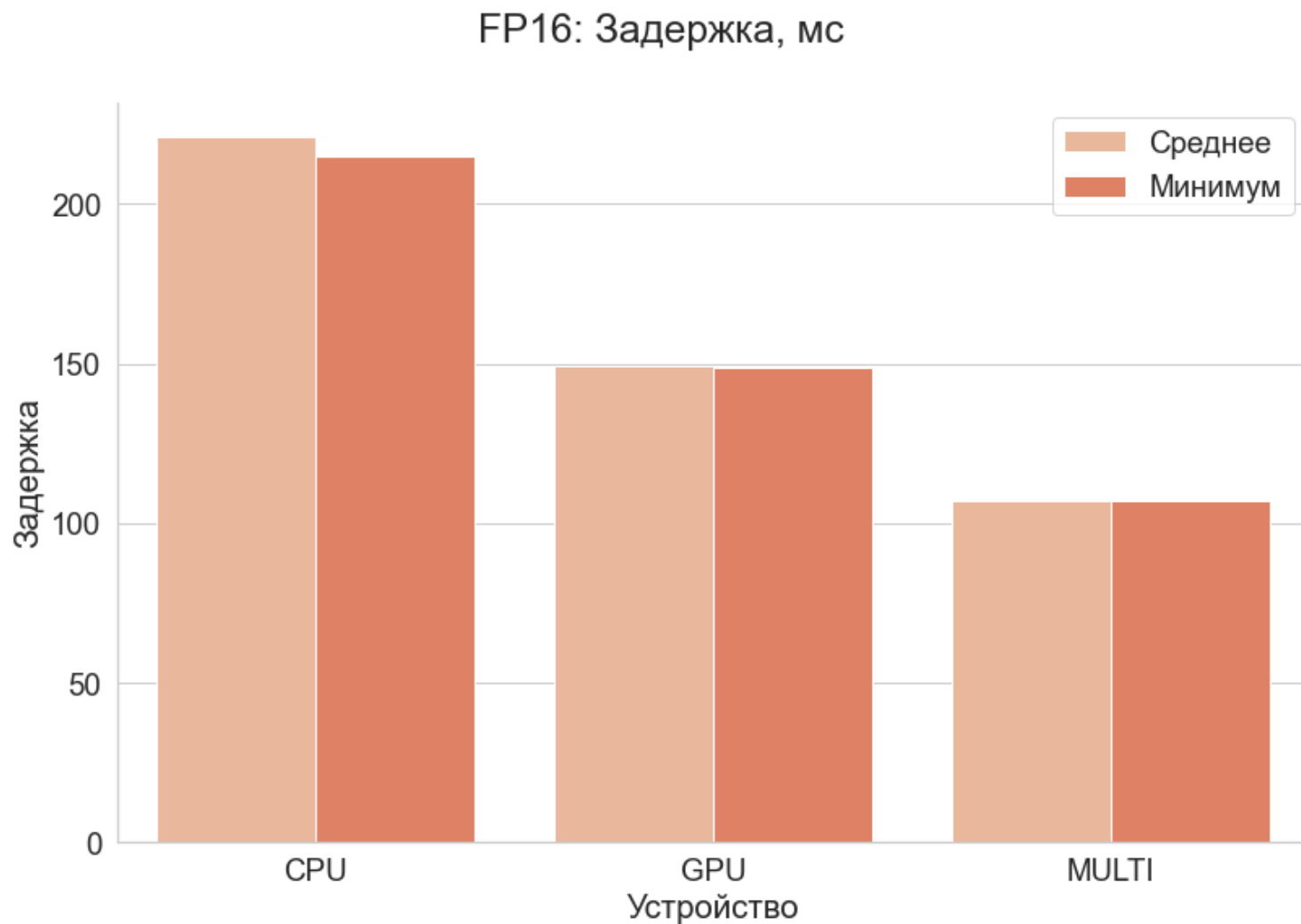
# Характеристики лабораторной машины

CPU	Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz 2.40 GHz
Модель основной платы	LNVNB161216
DIMM1	Kingston MSI16D3LS1KBG/8G 8 ГБ DDR3 1600 DDR3 SDRAM
GPU	Intel(R) UHD Graphics 630
GPU	NVIDIA GeForce GTX 1650

# Точность FP32. Задержка

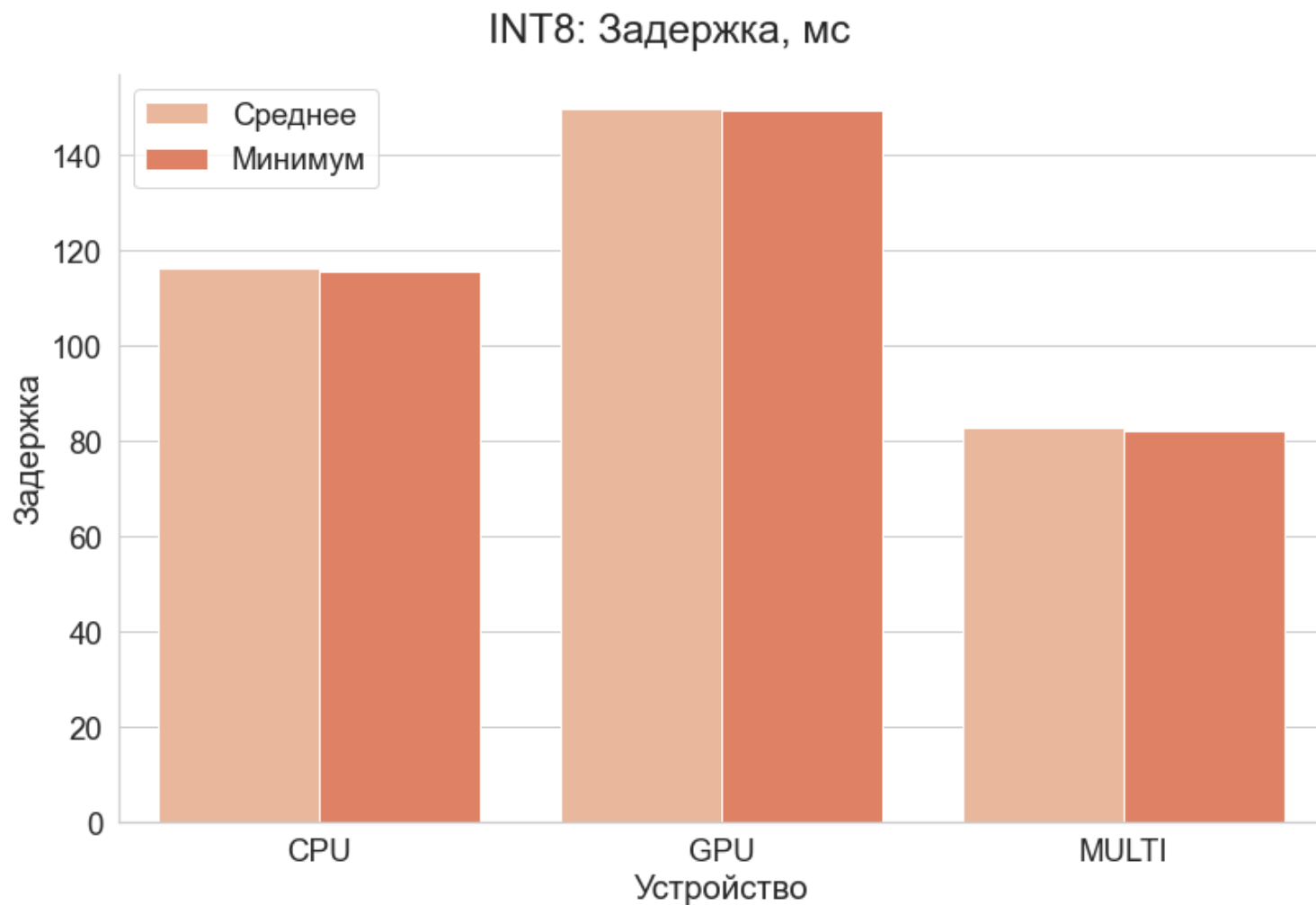


# Точность FP16. Задержка





# Точность INT8. Задержка



# Демо-приложение

