



Национальный исследовательский  
Нижегородский государственный университет им. Н.И. Лобачевского  
Институт информационных технологий, математики и механики

Образовательный курс  
«Введение в глубокое обучение с использованием  
Intel® neon™ Framework»

# Обзор возможностей инструмента Intel® nGraph™

*При поддержке компании Intel*

Кустикова Валентина,  
к.т.н., ст.преп. каф. МОСТ ИИТММ,  
ННГУ им. Н.И. Лобачевского

# Содержание

---

- ❑ Общая информация об инструменте Intel® nGraph™
- ❑ Основные концепции Intel® nGraph™
- ❑ Сборка и установка Intel® nGraph™
- ❑ Сборка и установка Intel® neon™ Framework для работы с Intel® nGraph™
- ❑ Пример обучения и тестирования модели





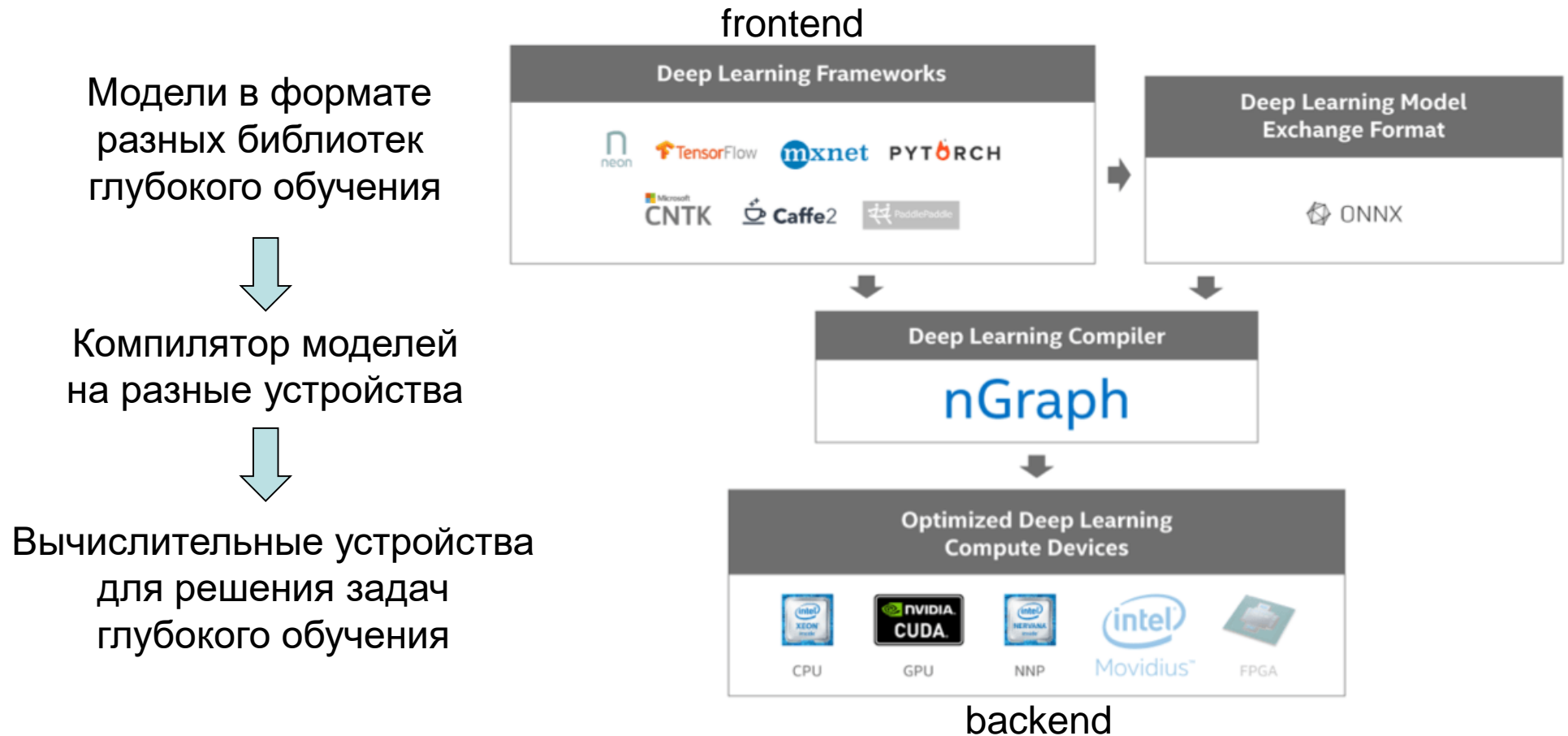
# ОБЩАЯ ИНФОРМАЦИЯ ОБ ИНСТРУМЕНТЕ INTEL® NGRAPH™

# Инструмент Intel® nGraph™ (1)

- **Цель разработки инструмента** – сделать работу аналитиков данных, разработчиков алгоритмов и программистов **портируемой, адаптивной и развертываемой** на самом современном аппаратном обеспечении, предназначенном для задач глубокого обучения
  - Intel® Architecture Processors (CPUs)
  - Intel® Integrated Graphics (GPUs)
  - Intel® Nervana™ Neural Network Processor™ (NNPs)
  - NVIDIA CUDA (GPUs)
  - Field Programmable Gate Arrays (FPGAs)
  - Movidius
- Инструмент Intel® nGraph™ (nGraph) включает библиотеку C++, среду исполнения и компилятор для глубокого обучения



# Инструмент Intel® nGraph™ (2)



\* nGraph [<http://ngraph.nervanasys.com/docs/latest>].

# Схемы работы с инструментом nGraph

- ❑ Схемы работы с инструментом nGraph:
  - Сборка стороннего инструмента глубокого обучения (TensorFlow, MXNet, Intel® neon™ Framework) для работы с nGraph и последующее обучение глубокой модели
  - Конвертация обученной модели в формат [Open Neural Network Exchange Format \(ONNX\)](#) для запуска на аппаратной инфраструктуре, поддерживаемой nGraph
  
- ❑ Примечание: далее рассматривается процедура установки Neon для работы с nGraph и пример обучения/тестирования модели



# Портабельность модели

- ❑ На данный момент поддерживаются три инструмента глубокого обучения с предварительно оптимизированной средой для обучения глубоких моделей:
  - TensorFlow
  - MXNet
  - Intel® neon™ Framework
  
- ❑ ***Портабельность модели***
  - Модель может быть размещена там, где лежат данные, а не ядро инструмента глубокого обучения
  - Нет необходимости размещать данные на вычислительных ресурсах, используемых для обучения



# Адаптивность модели

## □ *Адаптивность модели*

- Обученная модель может быть импортирована для тестирования на целевых серверах с использованием средств nGraph
- Импорт модели осуществляется в открытый формат представления моделей глубокого обучения ONNX

[<http://onnx.ai>]

- Тестирование моделей в формате ONNX обеспечивается вспомогательным инструментом, имеющим программный интерфейс на Python

[<https://github.com/NervanaSystems/ngraph-onnx>]

- Инструменты, поддерживающие формат ONNX:

- TensorFlow, MXNet, neon, CNTK, PyTorch, Caffe2





# Развертываемость

## □ *Развертываемость*

- Любая современная архитектура CPU может быть использована для обучения и тестирования глубоких нейросетевых моделей
- Модель может быть адаптирована для запуска на разнообразном аппаратном обеспечении

## □ Поддерживаемое аппаратное обеспечение:

- Intel® Architecture Processors (CPUs)
- Intel® Nervana™ Neural Network Processor™ (NNPs)
- NVIDIA GPUs

## □ Планируется поддержка:

- Field Programmable Gate Arrays (FPGAs)
- Movidius [<https://www.movidius.com>]



# С чего начать?

## □ Основные источники:

- Документация Intel® nGraph™ [<http://ngraph.nervanasys.com/docs/latest>]
- Репозиторий исходных кодов Intel® nGraph™ [<https://github.com/NervanaSystems/ngraph>]

## □ Дополнительные источники:

- Статья на странице Intel AI Academy [<https://ai.intel.com/ngraph-a-new-open-source-compiler-for-deep-learning-systems>]
- Intel® nGraph™. An Intermediate Representation, Compiler, and Executor for Deep Learning [<https://arxiv.org/pdf/1801.08058.pdf>]
- Другие статьи и видео-туториалы по данной тематике



---

# ОСНОВНЫЕ КОНЦЕПЦИИ БИБЛИОТЕКИ NGRAPH



# Представление глубокой модели в nGraph

- Глубокая модель в nGraph представляется **графом потоков данных** (data-flow graph)
  - **Узлы** этого графа соответствуют математическим операциям
  - **Ребра** соответствуют многомерным массивам данных (тензорам)
  
- Примечание: такое представление моделей поддерживается в TensorFlow, MXNet и Neon, поэтому разработчики nGraph обеспечивают прямое обучение моделей в формате указанных инструментов через программный интерфейс языка Python



# Состав библиотеки nGraph (1)

- **Мосты с инструментами** (framework bridges)
  - **Мост** – слой программного обеспечения, который транслирует наукоемкий язык данных в язык вычислений, представленный в виде **графа потока данных** (data-flow graph)
  - Мост предоставляет задачу библиотеке nGraph в виде **абстрактного слоя** (abstract layer), который отвечает за оптимальное исполнение
  - Оптимизация исполнения осуществляется посредством замены некоторых подграфов вычислений на более оптимальные
  - **Мост представляет собой плагин или расширение инструмента глубокого обучения, которые позволяют конвертировать модель в формат nGraph**



## Состав библиотеки nGraph (2)

---

- Преобразователи операций (transformer ops)
  - Мосты могут определять собственные специфические операции, поддерживаемые конкретным инструментом глубокого обучения
  - Преобразователи позволяют конвертировать специфичные операции инструмента в последовательность базовых операций библиотеки nGraph



# Состав библиотеки nGraph (3)

- **Элементы для построения графа** (graph shaping)
  - **Тензоры** (tensors) – отображения из множества координат на множество скалярных значений одного типа данных
    - Реализуются в виде многомерных массивов
    - Соответствуют данным в графе потоков данных
  - **Операции** (ops) – композиция тензорных преобразований
    - Операции являются узлами графа потока данных
    - Выход каждой операции имеет фиксированный тип данных элементов и размерность, которая отвечает размерности тензоров в графе
    - Входов и выходов может быть несколько
    - Каждая операция является узлом графа потока данных, но не все узлы являются операциями
    - Некоторые узлы графа описывают соответствие подграфов в процессе оптимизации исходного графа вычислений



# Состав библиотеки nGraph (4)

- **Функции построения графа** (graph shaping)
  - **Функции** (functions) – сгруппированный набор операций
    - Функция описывает последовательность операций на входным тензором
    - Операции могут выполняться на одном участке памяти
    - Определение функции включает следующие этапы:
      - Определение параметров – тензоров. При создании тензора указывается тип его элементов и размер
      - Определение узлов – операций, выполняемых над входными тензорами





# Заключение

- ❑ nGraph нацелен на программирование и развертывание приложений глубокого обучения на самых современных вычислительных устройствах
- ❑ Обеспечивается свобода разработки пользовательских API-интерфейсов для различных аппаратных решений непосредственно в рамках инструмента
- ❑ nGraph содержит мосты для некоторых из наиболее широко известных инструментов глубокого обучения (мост действует как посредник между ядром nGraph и инструментом глубокого обучения, предоставляя средства для использования различных платформ исполнения)
- ❑ Инструмент находится на стадии разработки, актуальная документация доступна на официальной странице [<http://ngraph.nervanasys.com>]

# Основная литература

---

- ❑ Хайкин С. Нейронные сети. Полный курс. – М.: Издательский дом «Вильямс». – 2006. – 1104 с.
- ❑ Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика. – 2002. – 344 с.
- ❑ Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press. – 2016. – [<http://www.deeplearningbook.org>].



# Авторский коллектив

---

- ❑ **Кустикова Валентина Дмитриевна**  
к.т.н., ст.преп. каф. МОСТ ИИТММ,  
ННГУ им. Н.И. Лобачевского  
[valentina.kustikova@itmm.unn.ru](mailto:valentina.kustikova@itmm.unn.ru)
- ❑ **Жильцов Максим Сергеевич**  
магистрант каф. МОСТ ИИТММ,  
ННГУ им. Н.И. Лобачевского  
[zhiltsov.max35@gmail.com](mailto:zhiltsov.max35@gmail.com)
- ❑ **Золотых Николай Юрьевич**  
д.ф.-м.н., проф. каф. АГДМ ИИТММ,  
ННГУ им. Н.И. Лобачевского  
[nikolai.zolotykh@itmm.unn.ru](mailto:nikolai.zolotykh@itmm.unn.ru)

