

新闻稿

Khronos Group 发布神经网络交换标准NNEF 1.0

2017年12月20日 - 比佛顿, 俄勒冈州 - [The Khronos™ Group](#), 一个由领先硬件和软件公司组成、致力于创建高级加速标准的开源组织, 今天发布神经网络交换格式([NNEF™](#)) 1.0 预测版规范, 作为在培训框架和界面引擎之间的经培训的神经网络通用格式。通过各种神经网络培训工具和界面引擎的丰富组合, NNEF 降低了 machine learning 部署碎片问题, 实现了在各种设备和平台上应用中的使用。NNEF 1.0 测试版规范的发布, 有助于在未来公开发布最终标准前获取业界的反馈, 欢迎访问 [NNEF GitHub repository](#) 分享评论和反馈。

NNEF 的目标是让数据科学家和工程师可以简便地从其所选的培训框架传输经培训网络到各种界面引擎。设备制造商可以依靠的一个稳定、灵活、可扩展的标准, 对于神经网络在各种高端设备的广泛采用非常重要。NNEF 囊括了对经培训神经网络架构、处理、参数的完整描述, 使用独立的培训工具完成这些描述, 并使用界面引擎执行。

“machine learning 业界从该领域的各工作组中大大获益, 但是缺少一个通用标准, 特别是研究如何让各种部署系统彼此贴合,” **NNEF 工作组主席 Peter McGuinness**。“Khronos 看到了业界的这个需求, 并在过去的一年中开发 NNEF 作为神经网络交换通用标准, 其意义就相当于神经网络的 PDF”。

NNEF 的设计致力于实现在各种工具和引擎直接的可靠的导出和导入, 例如 **Torch、Caffe、TensorFlow、Theano、Chainer、Caffe2、PyTorch 和 MXNet**。NNEF 1.0 预测版规范包括丰富的使用案例和网络类别, 以及各种功能和可扩展设计, 借用 Python 的 syntactical 元素但是添加了正式元素以用于更正。NNEF 包括对自定义复合操作处理的定义, 者带来了复杂网络优化的可能性。未来的工作将以一种可预测的方式建立在这一架构上, NNEF 便可以与时俱进, 随着 machine learning 的发展, 不断提供稳定的部署平台。

Khronos 发起了一系列开源项目, 包括 NNEF **syntax** 语法分析器/验证器以及从各种框架导出器的实例, 例如 **TensorFlow**, 我们欢迎 machine learning 社群的参与, 让 NNEF 对他们的工作流更有帮助。除此之外, NNEF 还与 Khronos **OpenVX™** 工作组紧密合作, 实现其在 NNEF 文档中的整合。**OpenVX** 神经网络扩展实现了 **OpenVX 1.2** 作为一个跨平台界面引擎, 结合了计算机视觉和 deep learning 在单一图形中的操作。

业界支持:

“**AIotive** 非常骄傲可以作为 NNEF 的开发和早期部署的主要参与者, 不仅仅作为一个新标准的推动者, 还委托 AI 研究者 **Viktor Gyenes** 作为标准编辑”, **AIotive 公司硬件工程部负责人 Marton Feher** 讲到, “我们的自动驾驶硬件和软件技术自很久以来作为早期采用者, 我们完全认可神经网络交换格式的重要性。随着开发框架扩展的数量不断增加、执行平台不断的发展和多样化, 可以自由移动网络拓扑并在不同环境间转换对于供应商选择自由化和创新是至关重要的”。

“随着我们将 **Deep learning** 从实验室转移到丰富的响应客户需求的各种应用中, 我们业界需要推动 **Deep Learning** 的交互转换解决方案,” **AMD Radeon 技术部 系统工程 CTO Greg Stoner**。“我们非常高兴看到 **Khronos Group** 推出 NNEF 规范, 支持神经网络在培训框架和建立在 **OpenVX** 的界面引擎之间简便的移动”。

“标准化格式以互换神经网络模型, 这对于网络和处理器在不同框架、工具和界面引擎之间的兼容性和优化具有重大意义。” **ARM 公司 Machine Learning 软件技术总监 Robert Elliott**, “**ARM** 支持 NNEF 的开发, 进一步使框架和工具开发者可以开发更多的模型, 以运行和验证更多的 **Arm®** 生态系统中的处理器和加速器”。

“随着神经网络处理从云转移到移动和高端设备, 对于这些模型的统一表示法的需求不断增长, 这对于 **Qualcomm Technologies** 关注在开发最好的平台以在设备上执行这些模型的公司也有很大帮助。”

Qualcomm Technologies, Inc. 公司技术 VP **Jeff Gehlhaar**。“作为 Khronos 会员公司, **Qualcomm Technologies** 公司相信合并将有助于该领域的发展。这也是对神经网络模型表示法标准的支持, 例如 Khronos 神经网络交换, 就是流线化从云到设备的转移”。

“NNEF 标准和 OpenVX 整合实现了神经网络在各种再我们 openVX GPU 的培训框架、视觉和神经网络处理 IP 的部署的捷径” 芯原公司执行副总裁兼 IP 部总经理戴伟进讲到。“通过使用通用转换格式和我们跨各种性能层级的专为 VeriSilicon IP 优化的 runtime 界面，我们的客户可以在其平台上立刻实现优化性能，无论他们选择哪种培训框架”。

新的 NNEF 1.0 文件项目和规范已上线：[Khronos Registry](#)。NNEF 开源工具和项目请浏览 Khronos [NNEF GitHub repository](#)。更多关于 *OpenVX 用于视觉加速和界面 run-time API* 的信息，请浏览：[here](#)。

更多关于 NNEF 任何带来处理平坦和房子操作灵活性的信息，请浏览 <https://www.khronos.org/blog/nnef-design-philosophy-network-structure-and-target-use-cases>

更多关于 The Khronos Group 的信息，请浏览 [Khronos.org](#)。

关于 The Khronos Group

The Khronos Group 是一个行业组织，创建开放标准，以实现并行计算、图形、视觉和神经网络在各种平台和设备上的编写和加速。Khronos 标准包括 Vulkan®, OpenGL®, OpenGL® ES, OpenGL® SC, WebGL™, SPIR-V™, OpenCL™, SYCL™, OpenVX™, NNEF™, COLLADA™, OpenXR™ 和 glTF™。Khronos 会员可以为 Khronos 规范的开发做贡献，并在标准公开发布前的不同阶段拥有投票权，通过提前获得规范草本和一致性测试，加速其高端加速平台和应用产品的推出。

###

Vulkan 是 The Khronos Group 的注册商标。Khronos, OpenXR, DevU, SPIR, SPIR-V, SYCL, WebGL, WebCL, COLLADA, OpenKODE, OpenVG, OpenVX, EGL, glTF, OpenKCAM, StreamInput, OpenWF, OpenSL ES, NNEF 和 OpenMAX 是 Khronos Group Inc. 的商标。OpenCL 是 Apple Inc. 的商标；OpenGL 是注册商标；OpenGL ES 和 OpenGL SC 是 Silicon Graphics International 的标识和商标，Khronos 在授权下使用。所有其他的产品名称、商标和/或公司名称仅供识别，并归属各自所有者。

媒体联系人

Caster Communications, Inc. 401-792-7080

Alex Crabb alex@castercomm.com 手机: +1-401-318-2229

Laura Shubel laura@castercomm.com