



Introduction to deep learning for the modeling of complex systems with applications

Alexandre Gonçalves Evsukoff

This seminar presents an introduction to machine learning methods for pattern recognition and presents an overview of the breakthrough models that are now being called deep learning methods. The seminars also present some application human like intelligence with deep learning in problems like natural language processing, image recognition and self-driving cars.

Alexandre G. Evsukoff received the M.Sc. degree in Mechanical Engineering in 1990 and the Ph.D. degree in Automatic Control in 1998 from Institut National Polytechnique de Grenoble, Grenoble, France. He is currently Associate Professor of Computational Systems at the Engineering Graduated Center COPPE/UFRJ. His research interests is data science focusing on machine learning and computational intelligence methods for the modeling of complex systems.



Introdução ao Deep Learning com o TensorFlow

Mauro Carlos Pichiliani

Esta palestra vai apresentar uma introdução ao framework de Deep Learning do Google chamado TensorFlow, incluindo suas principais características e usos. Também será comentado como utilizá-la junto com uma GPU da NVIDIA para acelerar o treinamento da rede neural. Por fim, alguns exemplos de uso vão ser abordados destacando as principais aplicações e usos práticos desta tecnologia.

Mauro Pichiliani é bacharel em Ciência da Computação, mestre e doutor pelo ITA (Instituto Tecnológico de Aeronáutica). Trabalha há mais de 15 anos utilizando diversas tecnologias de programação e já participou de projetos envolvendo aplicações móveis, realidade virtual, bancos de dados e machine learning. É pesquisador do lab de research da IBM Brasil, escritor da revista SQL Magazine e mantenedor do DatabaseCast (@databasecast), o podcast brasileiro sobre banco de dados.



HPC containers for CUDA Apps with Singularity

Paulo Souza

O uso de containers habilita um usuário sem privilégios de root executar um sistema operacional diferente na máquina, com isso o usuário tem o poder de configurar toda pilha de software e.g. S.O. e bibliotecas. Esta abordagem é conhecida como BYOE(Bring-Your- Own-Environment). Além disso o uso de containers facilita a distribuição de software e a reprodutibilidade de resultados. As tecnologias atuais de containers como o Docker, não foram desenhadas para o uso em ambientes de HPC. O projeto Singularity nasceu no Lawrence Berkeley National Laboratory e foi projetado desde o início para implementar containers em ambientes de HPC. Neste trabalho vamos apresentar o Singularity e um caso de uso em um cluster de GPUs utilizando uma aplicação CUDA com MPI+Infiniband.

Introdução ao Microsoft Cognitive Toolkit (CNTK)



Aurélio Moraes Figueiredo

Nesta apresentação, vamos demonstrar as principais características do CNTK, o toolkit de Deep Learning da Microsoft. Também vamos expor as vantagens da integração do CNTK ao hardware de GPUs da NVIDIA. Por fim, apresentaremos alguns exemplos de uso para essa ferramenta.

Aurélio Figueiredo é Engenheiro de Computação, com Mestrado em Computação Gráfica e Doutorado em Aprendizado de Máquina pelo Departamento de Informática da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Atuou por quase 10 anos como pesquisador do Instituto Tecgraf da PUC-Rio, desenvolvendo sistemas de Data Mining para Interpretação Automática de Dados Sísmicos. Desenvolveu várias aplicações de Machine Learning para a área de Óleo & Gás, publicando artigos nas maiores conferências mundiais dessa área. Atualmente é Senior Data Scientist da Metta Innovations, atuando em projetos ligados a Big Data, Machine Learning e Visão Computacional.



Como a tecnologia & GPUs podem nos empoderar a ler imagens & vídeos e aprender muito além do nosso alcance

Leonardo Bombonato Simões Coelho

Um papo recheado de cases reais e práticos da área da inteligência artificial que lida com a visão. Demonstrando como a tecnologia pode nos fortalecer em um mundo hiper-lotado de informação. O papo será conduzido por Leonardo Bombonato, sócio a Stilingue Inteligência Artificial, líder do núcleo de Visão Computacional e cientista voltado ao assunto.

Sócio da Stilingue Inteligência Artificial e líder do núcleo de Visão Computacional da empresa. Formado em Ciência da computação e mestrando em visão computacional pela Universidade Federal de Ouro Preto (UFOP). Cientificamente, seu atual desafio está no reconhecimento de contexto em larga escala baseado em análise visual, combinando técnicas de aprendizado de máquina, mineração de dados e visão computacional.

Metodologias Fundamentais para Otimizações em Cuda

Ricardo Farias

Vários fatores podem degradar a performance de kernels GPU. Analisaremos alguns detalhes das arquiteturas das GPUs, indicando boas práticas para se evitar degradações de performance. Abordaremos ainda algumas ideias de como se medir o grau de eficiência de códigos GPU.



Co-Design 101: Qual a Melhor Infra para Hospedar sua GPU?

Renato Miceli

Que GPUs são potentes aceleradores computacionais, todo mundo já sabe! Mas esse poder só vem com certo amparo: para um bom desempenho e escalabilidade das aplicações GPU-enabled, a escolha dos demais componentes da infra computacional faz toda a diferença. Qual a CPU mais adequada? Quanto necessito de RAM? Qual GPU escolher? E o número de nós? Este storage é suficiente? Nesta palestra vamos falar tntim por tntim sobre o time de apoio ao trabalho das GPUs e solucionar alguns dos deslizes mais comuns de iniciantes no processo de montagem de infraestruturas eficientes para GPU Computing.

Renato Miceli é Consultor de Negócios em Computação Avançada, especializado no planejamento estratégico e desenvolvimento de projetos de inovação científica e tecnológica para/atraves do uso de Supercomputação e Big Data. Renato tem Bacharelado em Ciência da Computação pela Universidade Federal de Campina Grande e é Doutorando em Ciência da Computação pela Université de Rennes I. Já atuou como Coordenador do Centro de Supercomputação do SENAI CIMATEC em Salvador, dirigindo ações de prospecção de negócios e transferência de conhecimento e tecnologia acadêmico-industriais, incluindo a gestão do supercomputador mais potente da América Latina; e fomentou colaborações de P&D na Europa como Cientista Computacional no Centro de HPC irlandês, onde seu trabalho em Fintech recebeu o prêmio HPCwire por "Melhor uso de HPC para serviços financeiros". A missão de Renato é apoiar seus parceiros a superar desafios de negócio e ciência através de inovação técnico-estratégica, trabalhando junto aos parceiros na concepção e operação de infraestruturas de HPC e HPDA co-projetadas sob medida e custo benefício para aprimorar modelos, garantir sucesso dos projetos, mitigar riscos e fomentar oportunidades



Do Stanley ao Waymo: as tecnologias que possibilitaram a direção autônoma.

Carlos Vido

Breve panorama histórico da direção autônoma, seguido de detalhes sobre como funcionam os veículos, do que são capazes e quais são as tecnologias que vêm permitindo sua disseminação.

Breve panorama histórico da direção autônoma, seguido de detalhes sobre como funcionam os veículos, do que são capazes e quais são as tecnologias que vêm permitindo sua disseminação.

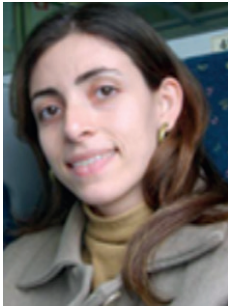


What comes next in Gpu computing

Esteban Clua

Esta palestra irá apresentar as novas arquiteturas de GPUs, bem como novidades nas novas versões de CUDA.

Professor da Universidade Federal Fluminense e coordenador geral do UFF Medialab, jovem cientista do nosso Estado pela FAPERJ em 2009 e 2013. Possui graduação em Computação pela Universidade de São Paulo, mestrado e doutorado em Informática pela PUC-Rio. Sua área de atuação está especialmente focada na área de Video Games, Realidade Virtual, GPUs e visualização. É um dos fundadores do SBGames (Simpósio Brasileiro de Games e Entretenimento Digital), tendo sido presidente da Comissão Especial de Jogos da SBC entre 2010 e 2014. Atualmente é o representante para o Brasil do Technical Committee de Entretenimento Digital da International Federation of Information Processing (IFIP) e membro honorário do conselho diretivo da ABrGames (Associação Brasileira de Desenvolvimento de Games). Em 2015 foi nomeado NVIDIA Fellow. Atualmente é coordenador do Centro de Excelência da NVIDIA no Brasil, que funciona no instituto de computação da Universidade Federal Fluminense.



Um convite ao aprendizado profundo

Cristina Nader Vasconcelos

Neste Workshop serão apresentados um breve resumo sobre a área de aprendizado profundo e suas inúmeras aplicações. Como convite para iniciantes na área, serão apresentados os conceitos básicos de redes neurais convolucionais (CNNs). Com esse objetivo, será apresentado um paralelo entre processamento de sinais por abordagens tradicionais além de uma comparação entre as redes neurais clássicas versus CNNs. Por fim, serão mencionados frameworks com diferentes características e funcionalidades.

Possui graduação no curso de Bacharel em Informática pela Universidade Federal do Rio de Janeiro (2003), mestrado em Informática pela Pontifícia Universidade Católica do Rio de Janeiro (2005) e doutorado (2009) pela mesma instituição. Tem experiência na área de Ciência da Computação, com ênfase em Computação Gráfica, atuando principalmente nos seguintes temas: visão computacional e processamento de imagens, reconhecimento de padrões e computação gráfica. Suas contribuições principais incluem temas de processamento genérico paralelo em hardware gráfico aplicados a tarefas de visão computacional, métodos de otimização discreta em grafos, processamento, gerenciamento e formatos de imagens e vídeo, estruturas de dados espaciais. Desde 2014, tem atuado na divulgação e formação de pesquisadores na área de aprendizado profundo. Hoje é professora adjunta da Universidade Federal Fluminense no Instituto de Computação.



High-Level and Productive Parallel Programming on GPUs

Dalvan Griebler

In the last decade, GPU accelerators have dictated a new trend in computer science and the industry. Also they have contributed to significant innovations in several areas. Artificial intelligence, autonomous vehicles, image processing, and big data are among examples of domains of applications that may take advantage of many-core parallelism. However, implementation has been not a trivial task for application programmers due to the low-level and architecture-oriented programming. Thus, a set of new tools and frameworks have emerged. This presentation will provide a discussion of preferred alternatives to develop efficient and productive parallel code for GPUs and also present the main initiatives from the scientific community to provide higher-level parallelism abstractions in the future.

Currently, he is a postdoc and computer science researcher at PUCRS. He received his Ph.D. from both the Pontifical Catholic University of Rio Grande do Sul (PUCRS) and University of Pisa (UNIP) in 2016. His main research is on high-level and productive parallelism abstractions, parallel programming, domain-specific languages, source-to-source compilers, cloud computing, and high-performance computing.



Uso de GPU em aplicações científicas.

Roberto Pinto Souto (LNCC).

Nesta palestra iremos abordar o uso de GPU em pesquisas realizadas no LNCC, para diferentes áreas do conhecimento. Em algumas aplicações foram utilizadas bibliotecas com soluções otimizadas para esta arquitetura. Em outros casos, foram desenvolvidas implementações próprias, por meio de técnicas de programação para GPU. Serão apresentadas as estratégias de otimização de código empregadas, bem como a comparação de desempenho com a versão original das aplicações.

Possui graduação em Matemática Aplicada Computacional pela Universidade Federal do Rio Grande do Sul (1997). Mestrado em Sensoriamento Remoto (2000) e doutorado em Computação Aplicada, ambos pelo Instituto Nacional de Pesquisas Espaciais (2006). É tecnólogo do LNCC desde 2013, atuando principalmente em processamento de alto desempenho aplicado a computação científica. Ministra curso de programação em CUDA no LNCC desde 2011.



Filtragem recursiva em paralelo para processamento de imagens

Diego Nehab

Essa palestra cobrirá o conteúdo de dois artigos publicados no SIGGRAPH Asia. No artigo de 2011, demonstramos como quebrar a cadeia de dependências que existe em filtros recursivos para aumentar a quantidade de paralelismo da computação. A situação é semelhante à existente no cálculo de somas cumulativas, exceto que cada saída do filtro recursivo é realimentada por um número arbitrário de saídas anteriores e os pesos são também arbitrários. Para resolver o problema, quebramos a entrada em blocos e, em uma primeira etapa, realizamos todos os cálculos independentemente por bloco, salvando apenas as condições finais das fronteiras de cada bloco. Uma etapa intermediária processa essas fronteiras transformando-as nos valores necessários para a realimentação inicial de cada bloco. Na etapa final, essas realimentações viabilizam o cálculo do resultado independentemente por bloco. Essa nova forma de realizar a computação permite uma grande economia na quantidade de acessos à memória global e, em consequência, levou a um ganho de eficiência considerável em relação ao estado da arte (12x). Em 2016, resolvemos atacar o problema do que fazer perto das fronteiras da entrada. A prática tradicional consiste em expandir às dimensões da entrada em todas as direções, preenchendo os novos valores de acordo com uma dentre várias alternativas: repetição periódica, reflexão periódica, ou expansão por constante. O resultado é uma aproximação que pode ser muito custosa para computar, já que filtros recursivos tipicamente têm uma resposta infinita ao impulso com suporte efetivo muito largo. Descobrimos uma forma de calcular, de forma fechada e exata, as realimentações iniciais ao redor da imagem. Além disso, conseguimos fazê-lo sem impactar a eficiência do algoritmo de 2011. O resultado é que agora é possível calcular o resultado exato de filtros recursivos em extensões infinitas da entrada e fazer isso em paralelo e eficientemente.

Diego Nehab is an associate professor at the National Institute for Pure and Applied Mathematics (IMPA) in Rio de Janeiro, Brazil. Before joining IMPA in 2010, Diego worked as a postdoctoral researcher at Microsoft Research in Redmond, WA. He is interested in most topics related to computer graphics, but focuses on parallelism, real-time rendering, and image processing. He received B.Eng. and M.S. degrees in computer science from PUC-Rio in 2000 and 2002, respectively, and a Ph.D. in computer science from Princeton University in 2007.



Virtual Reality applied to train electric substation professionals

Edgar Vilela Gadbem

In this talk, we will see an applied use case of virtual reality in a professional training. We will analyze the problem of training professionals to work on electric substations and how we could leverage the benefits of Virtual Reality to enable the professionals to practice what they learn without risks. We will also analyze the weakness of the technology and how we hem used the feedbacks of the final users to improve the experience.

Project Manager of innovative software projects in the areas of Virtual Reality, Image Processing, Computer Vision and Machine Learning. Previous background in Computer Graphics for games and high performance GPGPU programming using CUDA. Former professor of Computer Graphics, Artificial Intelligence and Game Engines at PUC – Campinas.



A deep learning strategy for the classification of constructed area on Google Earth images

Hélio Lopes

In this talk we will present an accurate deep neural network model for the classification of constructed area on Google Earth images. Moreover, we will show its applications to the analysis of census data. This work is in collaboration with Cassio Almeida and William Fernandes.

Hélio Lopes is an Associate Professor in the Department of Informatics at PUC-Rio. He has a Doctoral degree in Mathematics (1996), a Master's degree in Informatics (1992) and a Bachelor degree in Computer Engineering (1990). All of these degrees were obtained at PUC-Rio. He likes to develop and to apply advanced mathematical and computational techniques to solve real-world problems that have engineering, scientific and industrial relevance. In the present, he is particularly interested in Data Science on a Cloud Computing setting. He has also interest in Uncertainty Quantification and Visualization and their applications to industrial problems.



Graphics: A real-time future with the Unreal Engine

Paulo Souza

Paulo Souza irá contar como engines de jogos serão o motor que moverá TV, Filmes e Arquitetura ainda nesta década, e fará uma breve análise das tendências em computação gráfica para simulações, jogos e realidade virtual.

Paulo Souza ajuda os desenvolvedores brasileiros a criar experiências incríveis com a Unreal Engine. Com 15 anos de experiência na indústria de jogos, Paulo tem mais de 10 jogos lançados globalmente, de títulos AAA a jogos mobile.



New NVIDIA platform for Artificial Intelligence, Big Data & Analytics, and HPC.

Pedro Mário Cruz e Silva (NVIDIA)

In this talk, some of key applications of Artificial Intelligence (AI) applications to the Oil & Gas industry will be presented. Modern AI is the 4th industrial revolution. Deep Learning (DL) is the Machine Learning (ML) technique enabling breakthroughs in several industrial, business, and scientific workflows. The new NVIDIA's Deep Learning platform is providing the computational power demanded by the recent advances in AI. The new Pascal family of GPUs was specially designed for the High-Performance Computing workloads necessary to train a Deep Neural Network with a huge amount of data. The latest version of the CUDA language (version 8), and NVIDIA SDKs were improved to include specialized and highly optimized algorithms to extract GPUs full potential in DNN training and inference tasks. Large variety of training data can be efficiently used for training including text, audio, images, and video. This new computing model is delivering outstanding results in Computer Vision, Natural Language Processing, Language Translation, Speech Recognition, Recommendation Systems, Logistics, and Autonomous Cars and other machines.

Pedro Mário Cruz e Silva did his Bachelor in Mathematics (1995), and Master in Science (MSc) in Combinatorial Optimization (1998) at Federal University of Pernambuco (UFPE), he also did his Doctoral (DSc) in Computer Graphics (2004) at Pontifical University of Rio de Janeiro (PUC-Rio). Worked for 15 years at Tecgraf Institute/PUC-Rio as a Researcher, Project Manager and for 8 years as Manager of Computational Geophysics Group, during this period was responsible for several Software Development and R&D projects for Geophysics, and other fields of the Oil&Gas industry, with strong focus on innovation. Has more than 10 years experience with Agile Methodologies for Software Development, including Certified Scrum Master (CSM) and Certified Scrum Product Owner (CSPO). He also finished a Master's in Business Administration (MBA) in 2015 at Getúlio Vargas Foundation (FGV/RJ). As a volunteer project he is member of the main board of The Brazilian Geophysical Society (SBGf) as the Secretary of Publications. Recently joined the NVIDIA team as a Solution Architect for Latin America with special interest for Deep Learning applications for Oil&Gas industry.