NuGraph2
A Graph Neural Network for 3D Reconstruction in Liquid Argon Time Projection Chambers

V Hewes
9th May 2022
26th International Conference on Computing in High Energy & Nuclear Physics (CHEP23)
Exa.TrkX

- Exa.TrkX is a collaboration developing next-generation Graph Neural Network (GNN) reconstruction for HEP:
  - Energy Frontier
    - Expand on HEP TrkX's prototype GNN for HL-LHC.
    - Incorporate into ATLAS's simulation and validation chain.
  - Intensity Frontier
    - Explore viability of HEP TrkX network for neutrino physics.
    - Develop GNN-based reconstruction for Liquid Argon TPCs.

See Paulo Calafiura's overview talk!
Exa.TrkX

- Exa.TrkX is a collaboration developing next-generation Graph Neural Network (GNN) reconstruction for HEP:
  - **Energy Frontier**
    - Expand on HEP.TrkX's prototype GNN for HL-LHC.
    - Incorporate into ATLAS's simulation and validation chain.
Exa.TrkX

Exa.TrkX is a collaboration developing next-generation Graph Neural Network (GNN) reconstruction for HEP:

- **Energy Frontier**
  - Expand on HEP.TrkX’s prototype GNN for HL-LHC.
  - Incorporate into ATLAS’s simulation and validation chain.

- **Intensity Frontier**
  - Explore viability of HEP.TrkX network for neutrino physics.
  - Develop GNN-based reconstruction for Liquid Argon TPCs.
Exa.TrkX

- Exa.TrkX is a collaboration developing next-generation Graph Neural Network (GNN) reconstruction for HEP:
  
  **Energy Frontier**
  - Expand on HEP.TrkX's prototype GNN for HL-LHC.
  - Incorporate into ATLAS's simulation and validation chain.

  **Intensity Frontier**
  - Explore viability of HEP.TrkX network for neutrino physics.
  - Develop GNN-based reconstruction for Liquid Argon TPCs.

- See Paulo Calafiura's overview talk!
Liquid Argon TPCs

- Liquid Argon Time Projection Chambers (LArTPCs) currently a heavily utilised detector technology in neutrino physics.
  - At FNAL: MicroBooNE, Icarus, SBND.
  - Future: DUNE (70kT LArTPC deep underground, plus near detector).

- Charged particles ionize liquid argon as they travel.
- Ionisation electrons drift due to HV electrode field, and are collected by anode wires.
- Wire spacing ~3mm – high-resolution detector.
NuGraph2

• NuGraph2 is Exa.TrkX's second-generation GNN architecture for **semantically labelling LArTPC detector hits according to particle type.**
  
  • Utilise a **multi-head attention message-passing mechanism** within each detector plane.
  
  • Incorporate a number of improvements over first-generation proof-of-concept model (arxiv:2103.06233).
  
  • Incorporate **nexus connections** allowing information to pass between planes.
NuGraph2

- NuGraph2 is Exa.TrkX's second-generation GNN architecture for **semantically labelling LArTPC detector hits according to particle type**.
  - Utilise a **multi-head attention message-passing mechanism** within each detector plane.
  - Incorporate a number of improvements over first-generation proof-of-concept model (arxiv:2103.06233).
  - Incorporate **nexus connections** allowing information to pass between planes.
- Network trained on **simulated neutrinos** from MicroBooNE's Open Data Release.
  - See talk by Giuseppe Cerati this afternoon!
3D Nexus convolutions

- Perform message-passing independently in each detector view.
3D Nexus convolutions

- Perform message-passing independently in each detector view.
3D Nexus convolutions

• Perform message-passing independently in each detector view.
3D Nexus convolutions

- Perform message-passing independently in each detector view.
3D Nexus convolutions

- Add additional 3D step to the standard message-passing loop.
3D Nexus convolutions

• Add additional 3D step to the standard message-passing loop.

Pass features from 2D nodes to shared 3D nodes generated from simple spacepoint reconstruction
3D Nexus convolutions

• Add additional 3D step to the standard message-passing loop.

Pass features from 2D nodes to shared 3D nodes generated from simple spacepoint reconstruction

Convolve each 3D node to mix together features from all views
3D Nexus convolutions

• Add additional 3D step to the standard message-passing loop.

Pass features from 2D nodes to shared 3D nodes generated from simple spacepoint reconstruction

Convolve each 3D node to mix together features from all views

Propagate 3D features back down to 2D nodes and concatenate with 2D features to provide additional context.
• Network achieves \( \approx 86\% \) overall hit classification accuracy.

• With 3D connections, consistency of representations between views is now around 98\%, compared to \( \approx 70\% \) without.

Confusion matrix weighted by true semantic label to show efficiency.
NuGraph2

- Network achieves \(~86\)% overall hit classification accuracy.

- With 3D connections, consistency of representations between views is now around \(98\)% compared to \(~70\)% without.

Confusion matrix weighted by **predicted semantic label** to show **purity**.
NuGraph2

• Network achieves \( \sim 86\% \) overall hit classification accuracy.

• With 3D connections, consistency of representations between views is now around 98%, compared to \( \sim 70\% \) without.

• Inference takes \( 0.12 \text{ s / event on CPU} \).

GPU inference time as a function of batch size
Network performance on the Michel class primarily driven by **class imbalance in training dataset**.

- **EM showers** have large hit multiplicity, and make up **> 50% of hits** in training dataset.
- By contrast, **Michel electrons** are low multiplicity, and make up **< 1% of hits** in training dataset.

**Discussion**

- Multiple decoders
  - Binary classifier for background hit rejection (i.e., cosmic hit rejection).
  - Event classifier for event identification (i.e., neutrino interaction flavour).
  - Object condensation for instance segmentation (i.e., particle clustering).
  - Combine instance and semantic segmentation for panoptic segmentation, providing full particle reconstruction.
Discussion

- Network performance on the Michel class primarily driven by **class imbalance in training dataset**.
  - **EM showers** have large hit multiplicity, and make up > 50% of hits in training dataset.
  - By contrast, **Michel electrons** are low multiplicity, and make up < 1% of hits in training dataset.
- Looking forward: **multiple decoders**.
Discussion

- Network performance on the Michel class primarily driven by class imbalance in training dataset.
  - EM showers have large hit multiplicity, and make up > 50% of hits in training dataset.
  - By contrast, Michel electrons are low multiplicity, and make up < 1% of hits in training dataset.

- Looking forward: multiple decoders.
  - Binary classifier for background hit rejection (ie. cosmic hit rejection).
Discussion

• Network performance on the Michel class primarily driven by **class imbalance in training dataset**.
  
  • **EM showers** have large hit multiplicity, and make up **> 50% of hits** in training dataset.
  
  • By contrast, **Michel electrons** are low multiplicity, and make up **< 1% of hits** in training dataset.

• Looking forward: **multiple decoders**.
  
  • Binary classifier for **background hit rejection** (ie. cosmic hit rejection).
  
  • Event classifier for **event identification** (ie. neutrino interaction flavour).
Discussion

• Network performance on the Michel class primarily driven by **class imbalance in training dataset**.
  • **EM showers** have large hit multiplicity, and make up > 50% of hits in training dataset.
  • By contrast, **Michel electrons** are low multiplicity, and make up < 1% of hits in training dataset.

• Looking forward: **multiple decoders**.
  • Binary classifier for **background hit rejection** (ie. cosmic hit rejection).
  • Event classifier for **event identification** (ie. neutrino interaction flavour).
  • Object condensation for **instance segmentation** (ie. particle clustering).
Discussion

• Network performance on the Michel class primarily driven by **class imbalance in training dataset**.
  
  • **EM showers** have large hit multiplicity, and make up **> 50% of hits** in training dataset.
  
  • By contrast, **Michel electrons** are low multiplicity, and make up **< 1% of hits** in training dataset.

• Looking forward: **multiple decoders**.
  
  • Binary classifier for **background hit rejection** (ie. cosmic hit rejection).
  
  • Event classifier for **event identification** (ie. neutrino interaction flavour).
  
  • Object condensation for **instance segmentation** (ie. particle clustering).
  
  • Combine instance and semantic segmentation for **full particle reconstruction**.
Example $v_{\mu}$ interaction
Example $\nu_\mu$ interaction
Example $\nu_\mu$ interaction

Proton and muon tracks both efficiently reconstructed
Example $v_e$ interaction
Example $v_e$ interaction
Example $v_e$ interaction

Proton tracks and EM shower efficiently reconstructed
Example $v_e$ interaction

Minor confusion between shower and diffuse classes close to shower edges
Common abstraction for neutrino experiments

- Although the details of many neutrino physics experiments vary, the majority of them share a common paradigm at a high level.

**NOvA**
- Neutrino generator (GENIE)
- Particle simulation (Geant4)
- True light depositions
- Photoelectrons on APDs

**MicroBooNE**
- Neutrino generator (GENIE)
- Particle simulation (Geant4)
- True ionization electrons
- Pulses on TPC wires

**Shared structure**
- Event information
- True particles
- True energy deposits
- Detector hits
NuML & PyNuML

- The **NuML** package is a toolkit for writing **physics event records** to an **HDF5 file format**.
  - Hold low-level information such as **simulated particles**, **hits**, **true energy depositions** etc.
  - Generic data structure can be **shared across experiments**.
  - Common interface with **PandAna** analysis toolkit (see **CHEP 2021 talk**).
  - Available as **LArSoft package on GitHub**.

- The **PyNuML** package is designed to provide a **generic**, **accessible**, **efficient** and **flexible** solution for many of the necessary tasks in leveraging ML for particle physics.
  - Define particle ground truth labels for Geant4-simulated particles.
  - Arrange detector hits into ML objects, ie. graphs, CNN pixel maps, etc.
  - Efficiently preprocess ML inputs in parallel in HPC environments using MPI.
  - Available as **Python package on GitHub**, or install with **pip install pynuml**!
NuML & PyNuML

- The **NuML** package is a toolkit for writing **physics event records** to an **HDF5 file format**.
  - Hold low-level information such as **simulated particles, hits, true energy depositions** etc.
  - Generic data structure can be **shared across experiments**.
  - Common interface with **PandAna** analysis toolkit (see **CHEP 2021 talk**).
  - Available as **LArSoft package on GitHub**.

- The **PyNuML** package is designed to provide a **generic, accessible, efficient** and **flexible** solution for many of the necessary tasks in leveraging ML for particle physics.
  - Define **particle ground truth labels** for Geant4-simulated particles.
  - **Arrange detector hits into ML objects**, ie. graphs, CNN pixel maps, etc.
  - Efficiently **preprocess ML inputs in parallel in HPC environments** using MPI.
  - Available as **Python package on GitHub**, or install with **pip install pynuml**!
Summary

• **NuGraph2** is a state-of-the-art graph neural network for semantically labelling detector hits in neutrino physics experiments.
  • Model developed and tested in MicroBooNE and DUNE, and designed to be utilised across many neutrino physics detectors.
  • Targeting full particle reconstruction for next generation architecture.

• Standardised process of producing ML inputs from HEP data for general use with **NeutrinoML** toolkit.
  • Toolkit utilised for MicroBooNE's public data release.
  • Open-source, easy-to-install code packages.