Internship Miguel: LIDAR

Internship plan

Stef Vandermeeren

30/01/2025

Situation

Kapernikov is developing a strong expertise in point cloud analysis. Clients use this intelligence for efficient, centralized and reliable maintenance monitoring. A key step in this analysis is moving from point level representation to object level representation. For example, in the context of railways, those objects can be rails and cables or vegetation.

For this task, the current industry standard is feature engineering at the point level followed by a ml model to aggregate those. This approach isn't perfectly satisfying because of the amount of effort the feature engineering consumes.

For this reason, Kapernikov is developing a new deep learning based solution called Copernnet. Compared to existing research, this solution scales with the volume of the point cloud, rather than the number of points. This makes it suitable for real world large point clouds (e.g. the city of Brussels gets a new 100 billion + point cloud every year). The model is a transformer modified to accommodate point cloud data.

The internship

1ST OPTION: USE COPERNNET IN PRODUCTION

- Build a pipeline that will use Copernnet to predict alerts
- Handle the ETL of the produced data
- Deploy the changes to our Kubernetes cluster

2ND OPTION: COPERNET ECOSYSTEM IMPROVEMENT

- Be able to train and track progress with an experiment monitoring tool (eg. Determined)
  - This includes model versioning and keeping track of the data that was used for training/validation
- Visual debugging the network sampling / prediction

- Set up a training cluster with multiple gpu. Change the training script to split the work efficiently
- Find a more general learning procedure
- Automate hyper parameter testing.

What will you learn?

- Data pipeline architecture
- Point cloud and lidar processing tools
- MLops tools and practices
- ETL and relational db design