Grad-CAM feature for explaining areas of interest

Astraea Software Co., Ltd.

January 20, 2025

Overview

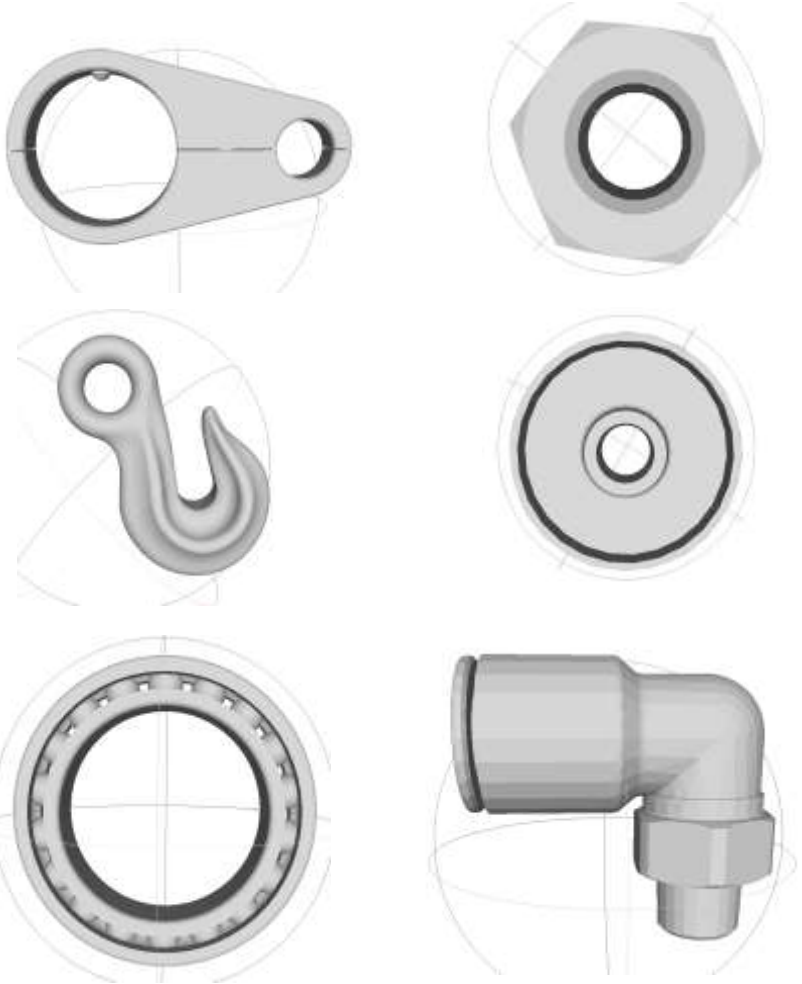
- Deep learning models automatically determine the criteria for judgment during training, making it challenging to explain the basis of their decisions. As a result, AI models are often regarded as "black boxes." However, there is a growing demand for methods to provide some level of explanation for these models' decisions.
- Grad-CAM (Gradient-weighted Class Activation Mapping) is one such method that addresses this issue. A brief explanation of Grad-CAM can be found on pages 6–7.
- To illustrate, we used PointNet to classify 3D shapes and applied Grad-CAM to generate a heat map. This heat map visualizes the parts of the 3D shape that the AI model focused on when making its classification. The visualization assigns colors to regions of the shape, where red indicates the highest attention, followed by orange, green, and blue in decreasing order of focus.

Overview

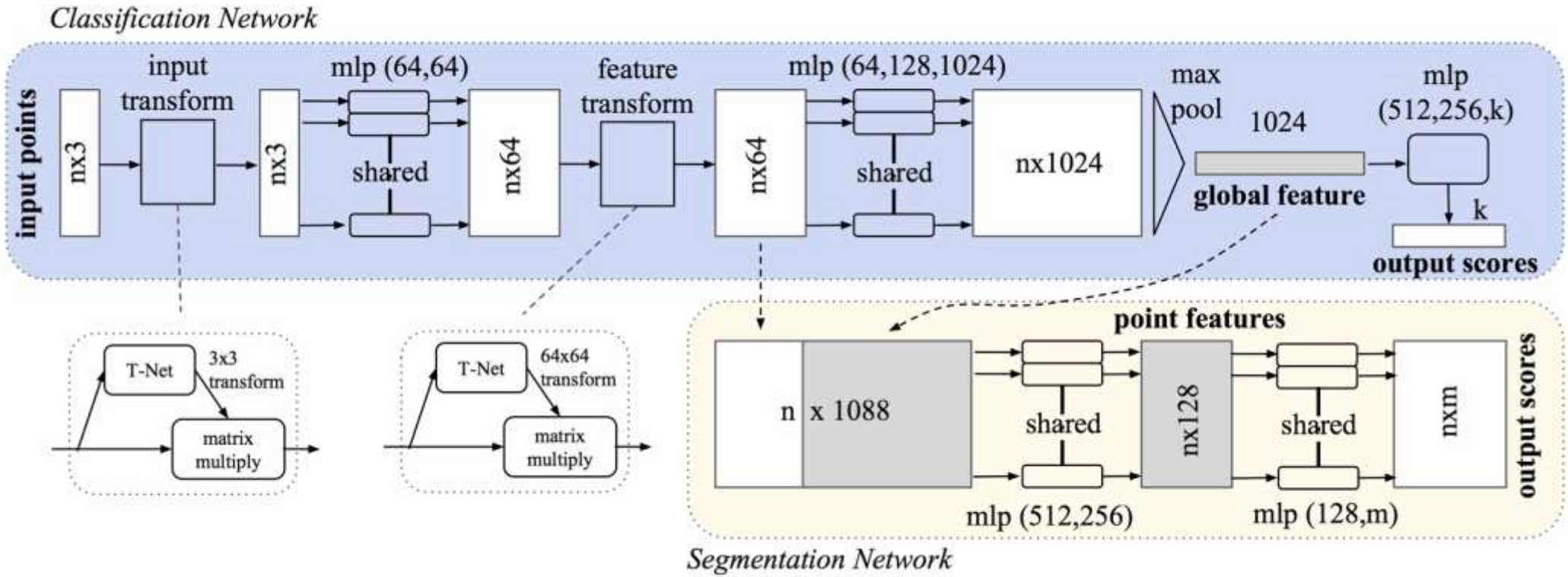
- Using PointNet, we developed an AI model to classify the mechanical parts in the training dataset shown on page 4. Additionally, we applied Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize the areas of focus that served as the basis for the model's decisions through a heat map.
- The classification results for the Hook part are summarized on page 9. The sample part was correctly classified into the Hook group, with attention concentrated on the hook's tip and the curved regions at the top and bottom of its shape. These areas are inferred to have significantly influenced the AI model's decision.
- Similarly, the classification results for the Wheel part are summarized on page 10. In this case, the model's attention was focused on the outer rim and central axis, suggesting these features played a key role in the classification.

Training Dataset

- 12 classes:
  - Articulations, eyelets and other articulated joints
  - Castor
  - Clamps
  - Cylindrical pins
  - Elbow fitting
  - Hexagonal nuts
  - Hook
  - Radial contact ball bearings
  - Tapping screws
  - Thrust washers
  - T-shape fitting
  - Wheel
- 96 per class, 1056 in total

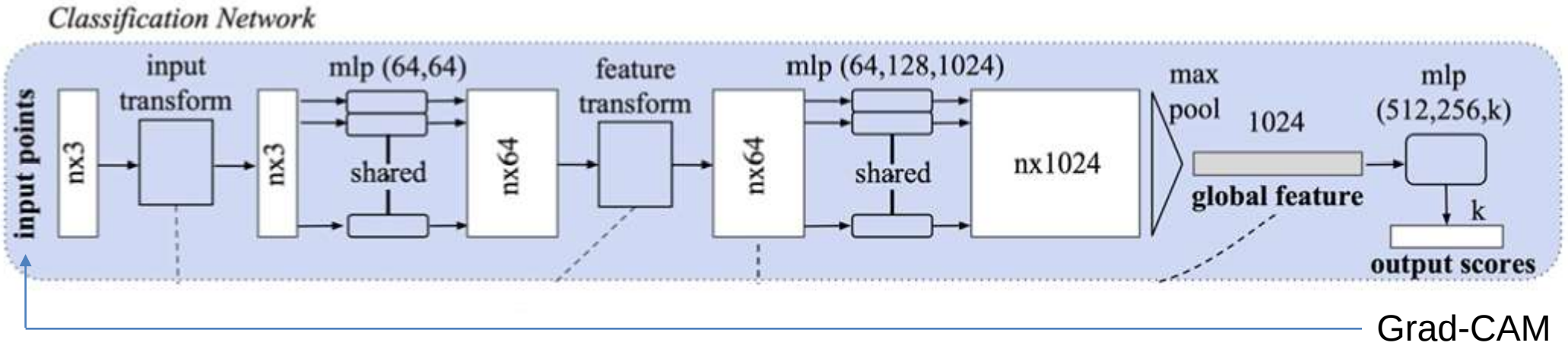


Classification NN using PointNet



Grad-CAM Overview

Reference from Grad-CAM (Gradient-weighted Class Activation Mapping), a technique used to visualize which parts of an input a deep learning model focuses on when making a decision.

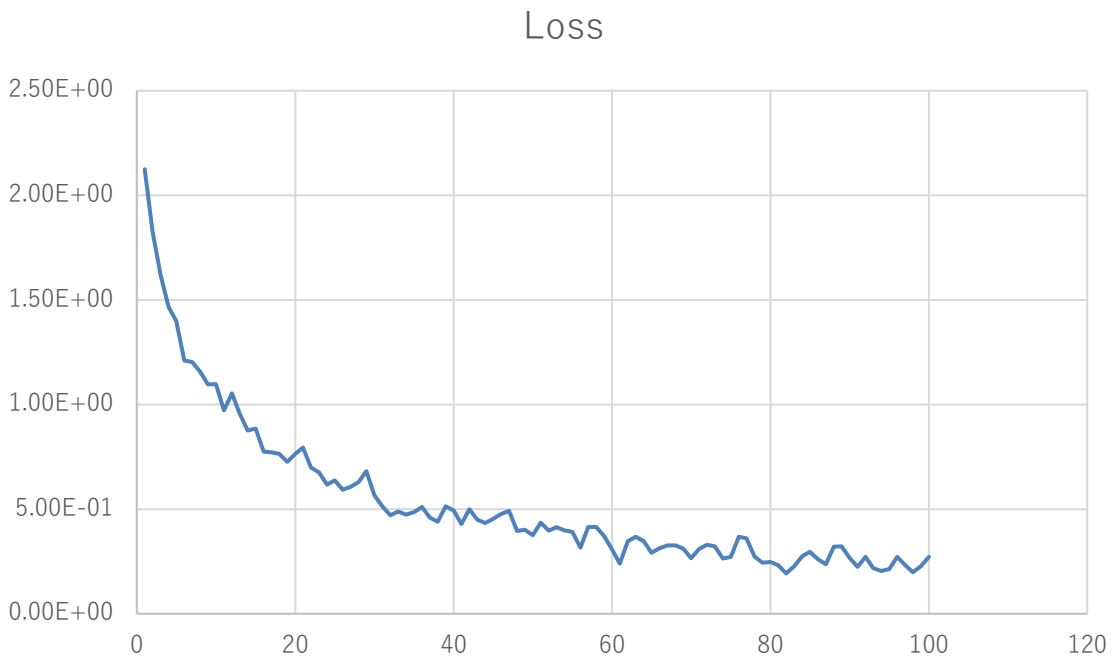


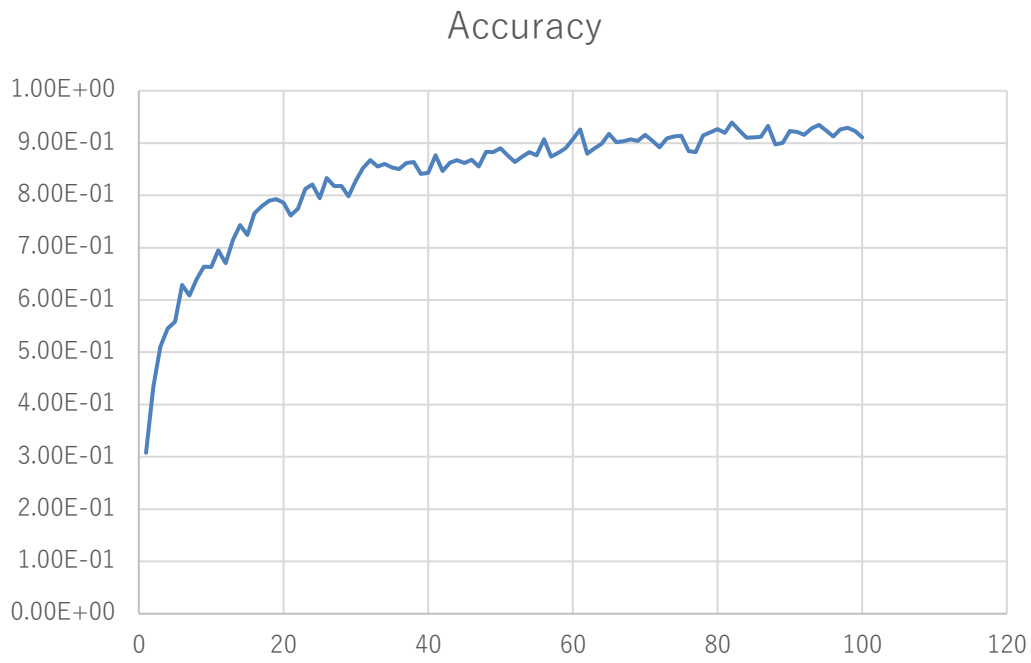
- Gradients of the model's predicted score (for a specific class) are calculated with respect to the feature maps which calculated from the last layer.
- The gradients are globally averaged to produce a set of weights.
- The feature maps are combined using these weights to generate a heatmap which is overlaid on the original input to create a visual explanation.

Grad-CAM function details

- Feature Map Extraction: Grad-CAM computes the feature maps from the last convolutional layer of a CNN. These feature maps contain spatial information about the image.
- Gradient Calculation: Weight Computation:. These weights represent how much each feature map contributes to the target class's prediction.
- Weighted Sum:. The heatmap highlights the most important regions of the input image that influenced the model's decision.
- Overlay: The heatmap is usually overlaid on the original image to create a visual explanation.

Loss function and accuracy convergence

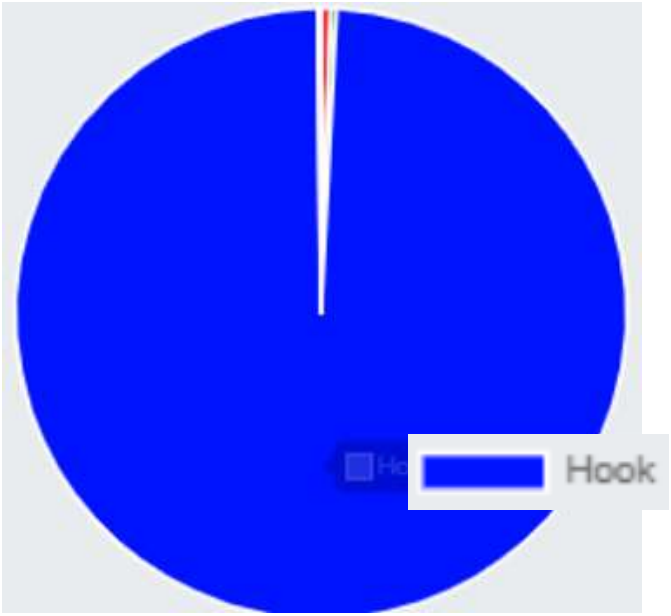




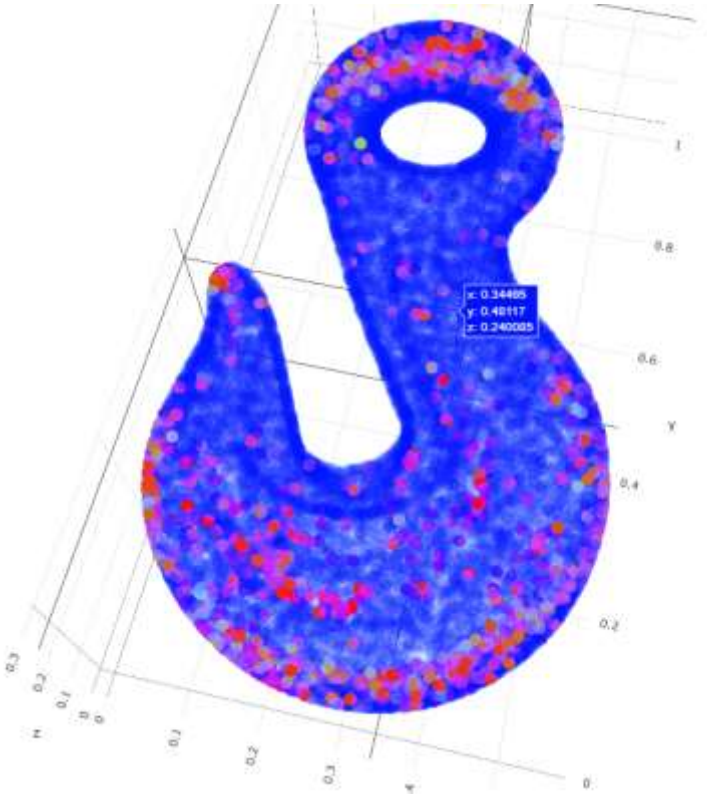
Example of Hook part judgment result

Input Geometry

Judgement：Hook



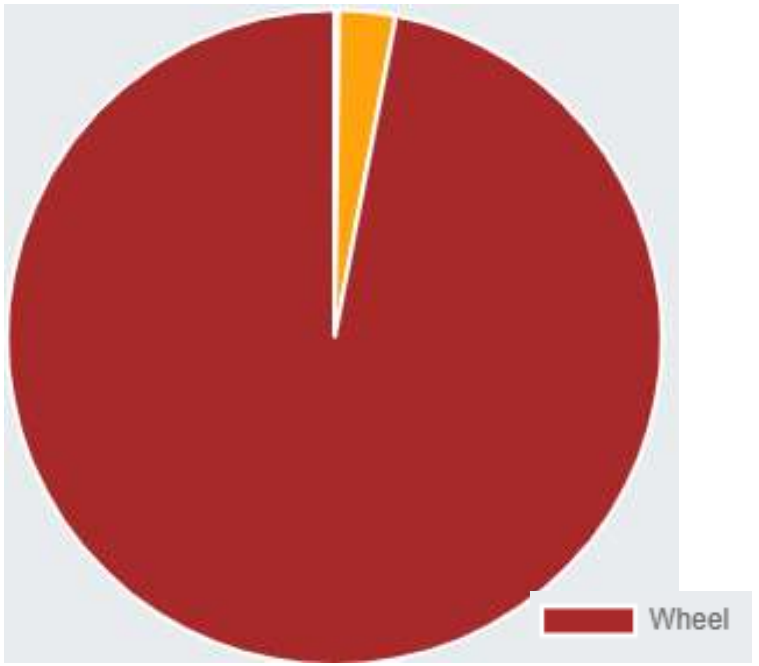
Heat map



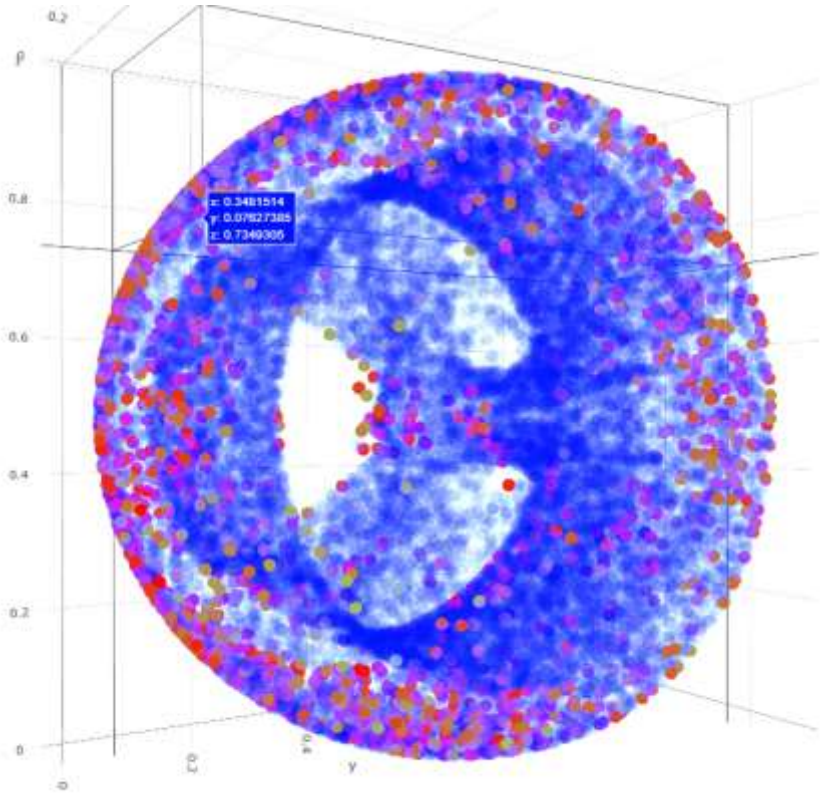
Example of wheel part judgment result

Input Geometry

Judgement：Wheel



Heat map



Demo site

- You can try out the Grad-CAM AI model we have created at the following URL:
- https://demo3.astraea-soft.com/test_pointcloud.html

End