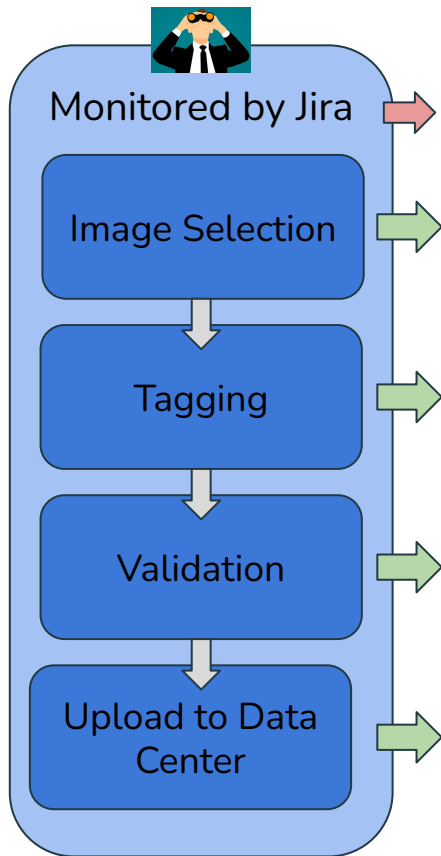# Train Data Tagging Pipeline

Monitored by Jira

Jira is used for managing the open tasks throughout the data tagging pipeline.

Image Selection

Images are selected using data from a high-end object detection network and results from our object detection model. Images are selected if they have edge case objects, or if they have a "ground truth" box that was not identified by the detection results and vice versa.

Tagging

The selected images are then sent to the tagging team that generates the ground truth data. This data includes: coordinates and size of the object in the image, classification of the object, 3D orientation. Metadata about the image such as time of day, weather conditions and location are added as well.

Validation

The generated data is then sent to be validated to ensure the integrity of the data. The data is checked for anomalies in various ways, such as: checking the coordinates and bounds of the box is within the bounds of the image, checking the width and height of the box are positive, checking the rider flag is only used in correctly, and checking the D3 separation value is used correctly and is indeed inside the box bounds.

Upload to Data Center

The images and validated data are then uploaded to Autobrains data center and are made accessible to the deep learning developers to be used to enhance the object detection model.