Validation Results Report

Converted from Validation Results Report.docx for Instructables supporting documentation.

Project: Fridge Helper Student: Hanna Koziun Model type: YOLO object detection Intermediate model: yolo26n Dataset: Fridge Helper Roboflow YOLO dataset

This report describes the first intermediate validation results for the Fridge Helper AI model. The goal of the mo del is to detect packaged fridge products or product label areas and classify theminto the correct product categ ory.

Dataset Preparation

The original Roboflow export was prepared before training. The exported dataset originally contained only a training split, so a preparation script created a train, validation, and test split. The prepared dataset contains:

Split	Images	Labels	Objects
Train	275	275	278
Validation	36	36	36
Test	36	36	36

The final model uses five classes:

Class
milk_carton_or_plastic_cup
yogurt_cup_or_lid
egg_carton_label
meat_package_sticker
cheese_package_sticker

The preparation script also checked the dataset for missing labels, labels without images, empty labels, invalid class IDs, malformed YOLO label rows, and incorrect bounding box values. After preparation, the dataset had no missing labels and no label errors. The original Roboflow export had inconsistent class names, such as several different milk classes.

Model Choice

The selected model for this is YOLO26n.

This model was chosen because it is:

suitable for object detection,

fast enough for a first prototype,

easier to train on a laptop,

more realistic for future Raspberry Pi use,

consistent with the YOLO example project used as reference.

For this project, the best model is not only the model with the highest accuracy. The final system should eventually work on a Raspberry Pi, so speed and latency also matter. A larger model may be more accurate,

but it may be too slow for the final embedded prototype.

Training Setup

Setting	Value
Epochs	30
Image size	640
Batch size	4
Patience	10
Model	YOLO26n

Mild data augmentation was handled during training by Ultralytics. This helps the model generalize better to new images by slightly changing colors, scale, translation, flipping, and mosaic combinations during training. The goal was to get an intermediate result, inspect the metrics, and decide what should be improved next.

Metric Of Interest

The main metric for this project is mAP@50.

mAP@50 is important because my project uses YOLO object detection. The model must not only choose the correct class but also place a bounding box around the correct product. mAP@50 checks whether the predicted bounding box overlaps enough with the real annotated object and whether the class is correct.

The secondary metric is recall.

Recall is important because missed products will not be saved into the fridge inventory. For example, if the model does not detect a milk carton, the system cannot add it to the product list. A model with high precision but low recall may look ok, but it will still fail the user if it misses products.

Precision is also useful because wrong detections could add incorrect products to the fridge list. However, for this first prototype, recall is especially important because missed detections directly affect the usefulness of the system.

Current Performance

The intermediate model performs well for a first validation run. The average mAP@50 is approximately 0.985, which means the model usually detects the correct class with a good enough bounding box. The average precision is approximately 0.939, and the average recall is approximately 0.926.

The mAP@50-95 is lower, approximately 0.749. This means that although the model usually finds the right object, the bounding box precision can still be improved. This is expected for an intermediate run with a relatively small dataset.

Classification Report

The table below shows the per-class validation metrics.

Class	Precision	Recall	mAP@50	mAP@50-95
milk_carton_or_plastic_cup	1.000	0.775	0.962	0.726
yogurt_cup_or_lid	0.886	1.000	0.995	0.765
egg_carton_label	0.933	1.000	0.995	0.789
meat_package_sticker	0.893	1.000	0.995	0.695
cheese_package_sticker	0.983	0.857	0.978	0.770

The strongest classes in this validation run are yogurt_cup_or_lid, egg_carton_label, and meat_package_sticker, because they reached perfect recall. This means the validation objects for those classes were detected.

The weaker classes are milk_carton_or_plastic_cup and cheese_package_sticker, because their recall is lower. This means that some validation objects from these classes were missed. For the next dataset iteration, these classes need more examples with different brands, angles, and lighting conditions.

The classification report shows that the model already separates most classes well, but some classes still need more data. The milk class is very broad because it includes cartons and plastic cups, so the model may need more variation to understand the class better. Cheese packaging can also be difficult because of reflections, transparent plastic, and different label positions.

Confusion Matrix

The confusion matrix shows that most predictions are on the correct diagonal. This means the model usually predicts the correct class. The main remaining errors are related to the background class. Some real objects are missed and counted as background. This is visible for classes such as milk_carton_or_plastic_cup, yogurt_cup_or_lid, meat_package_sticker, and cheese_package_sticker.

Training Results Graph

The training results graph shows that the model improved during training. The training and validation classification loss generally decreased. The mAP@50 and mAP@50-95 curves increased during training, which means the model learned useful patterns from the dataset. The precision and recall curves also improved during training. Precision reached a high level, meaning most predicted detections were correct. Recall also improved, but it is still not perfect. The mAP@50 curve reaches a high value near the end of training. This is a good sign for an intermediate result. The mAP@50-95 curve is lower, which means the model can still improve the exact bounding box quality.

Validation Prediction Example

The validation prediction image shows how the model performs on real validation examples. In the shown batch, the model correctly detects several cheese package stickers and draws bounding boxes around the package label area. One example appears to be missed or detected with lower confidence.

Current Limitations

The current validation result has potential, but there are still limitations:

The dataset is still small.

Some classes have fewer validation examples than others.

Reflections on plastic packaging can make labels harder to detect.

Some classes are broad, especially milk_carton_or_plastic_cup.

The model sometimes misses objects and predicts background instead.

Next Steps

Add more photos for milk_carton_or_plastic_cup.

Add more photos for cheese_package_sticker.

Add more examples for egg_carton_label, because this class has fewer validation samples.

Use more brands and package types.

Add more lighting conditions, including low light and shadows.

Add more angles and distances.

Improve bounding box consistency.

Add more validation images.

Retrain the model and compare the new metrics with this run.

Compare nano, small, and medium YOLO models if there is enough time.