



Detecting Windshield Damage Using Physically Realistic Synthetic Data

Authors: Jose Castro - Miguel Clavijo - Diego Torres

www.datadOO.com - hi@datadOO.ai

Motivation & Challenges

Motivation:

- To automate and reduce the costs associated with the 800 billion dollar [1] auto insurance industry who faces constant pressure to optimize claims efficiency. Windshield damage alone accounts for 30% [2] of all insurance claims. The TAM is \$37-41B [3][4].
- To reduce the process complexity of filing a claim for windshield related damages which causes frustration to customers resulting in churn.



Windshield cracking typical of laminated safety glass

Challenges:

- Real Data:** Very difficult to obtain assorted and high quality real data in this field. It's scarce and extremely difficult to label accurately.
- Synthetic Data Generation (SDG):** Transparent and reflective surfaces like windshields or any glass materials are very complex to simulate, especially damaged glass, due to the variability of lighting conditions (reflection and refraction). In order to assess and validate damage, there is a need for high fidelity data to produce a complete and balanced dataset.



Issues rendering glass material

Contributions

- End-to-end SDG pipeline for laminated glass damage with automatic pixel-wise labels.
- Domain control via Cosmos to cover weather, optics, curvature, and crack taxonomies.
- Segmentation output providing measured and qualified repair-vs-replace decisions for insurers.

Synthetic Data & Cosmos-Transfer

We author scenes in USD and programmatically vary illumination, weather, camera pose, lens model, shutter, sensor noise, glass curvature, lamination, and damage taxonomy. NVIDIA Replicator produces RGB, depth, normals, instance masks, and metadata for automatic labeling.

We add an extra step with NVIDIA Cosmos-Transfer [5] to control the generated domain. Cosmos acts as a physical AI foundation layer that allows us to augment the variability of our dataset in a consistent way. Like this, we fully control the scalability of our datasets covering specific environments and edge-cases leveraging natural language as an extra input.



Usage of Replicator for generating windshield synthetic data

Our data generation pipeline with NVIDIA Replicator and Cosmos-Transfer [5] allows us to obtain spatial-aware videos that detects recycled footage, timebase glitches or spatial-temporal inconsistencies. With this anti-fraud measure we authenticate the user submitted video. This is essential for insurance companies.



Depth and instance segmentation condition maps

Render Domain 1

Render Domain 2

Render Domain 3

Cosmos has helped us obtain variability in our dataset generation without compromising quality, obtaining better data diversity with Sim2Real transfer capabilities.

With the use of Cosmos-Transfer, we have kept the same 3D assets and condition maps; depth and instance segmentation, controlling the domain generation. The above shows variations in vehicle color, background of the picture and lighting conditions while maintaining the same geometry and location of the windshield crack.

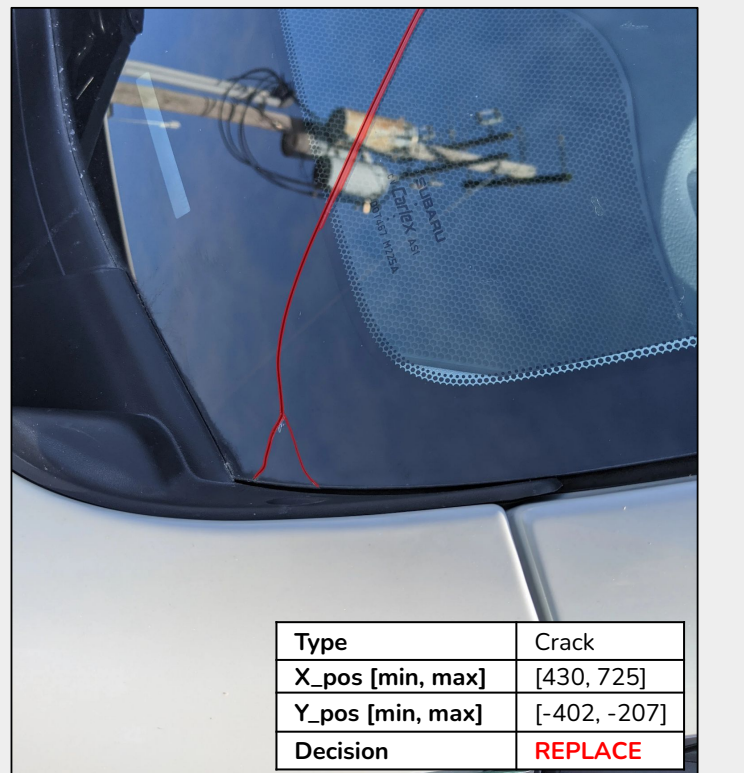
Results

We trained a semantic segmentation model with an Omniverse synthetically generated dataset, bridging the Sim2Real [6] gap without the need for real data.

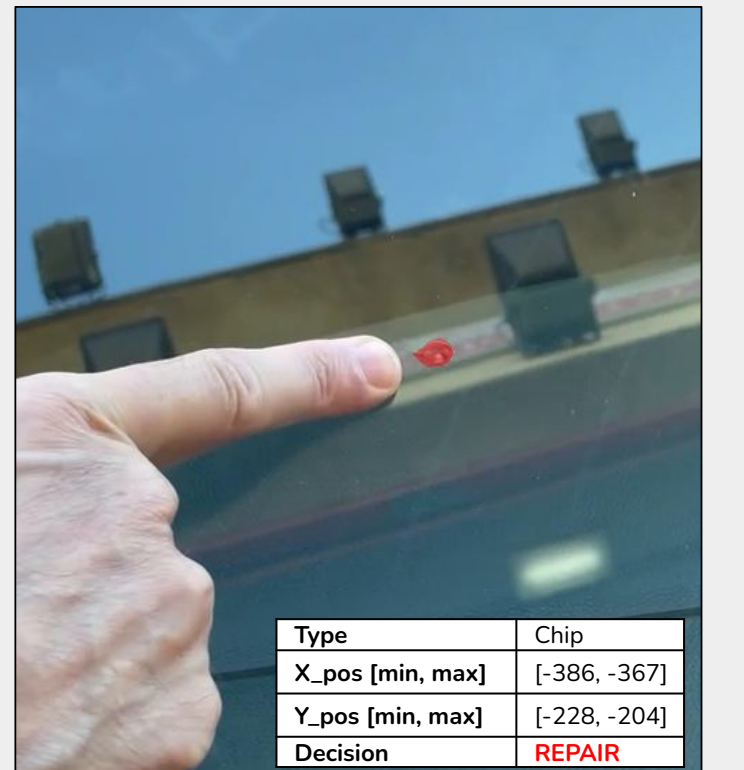
The iteration during development was accelerated by 1.4x speed (see table below) thanks to the quick SDG turnaround speeds together with curation possibilities of our pipeline (see pipeline below).

Our model provides powerful reporting capabilities with a segmentation mask. With this, we measure the damage and positioning within the windshield. This data is essential to deliver a decision on a "repair or replace" scenario.

The model helped achieve a reduction of human inspection providing a fully "notarized" output that serves the insurance companies as a validation comparable to the valuator output and help the customer to obtain a faster decision to repair or a replace without friction or churn.



Type	Crack
X_pos [min, max]	[430, 725]
Y_pos [min, max]	[-402, -207]
Decision	REPLACE



Type	Chip
X_pos [min, max]	[-386, -367]
Y_pos [min, max]	[-228, -204]
Decision	REPAIR

Repair or replace decision making

Segmented Results



Chip damage detection



Windshield cracks detection



Bullseye and crack damage detection

References

- [Verified Market Reports - Global Motor Insurance Market](#)
- [Insurance Journal - Facts and Findings about Windshields](#)
- [Automotive Glass Replacement Global Market Report 2025](#)
- [Automotive Glass Replacement Market Size & Forecast to 2029](#)
- [NVIDIA Cosmos-Transfer Github Repository](#)
- [NVIDIA - Closing the Sim2Real Gap](#)

SDG Performance Metrics

We reworked all stage randomization logic from Replicator achieving an increased generation speed of ≈ 40% ensuring quality and performance with RTX - Real-Time 2.0 rendering engine.

Setup	Render	Generation time [s/sample]
Standard Replicator	RTX - Interactive	53.36
Standard Replicator	RTX - Real-Time 2.0	6.38
DatadOO SDG + Replicator	RTX - Interactive	38.28
DatadOO SDG + Replicator	RTX - Real-Time 2.0	4.57

THE PIPELINE

