



SAGIVTECH

IMVC 2024

Can AI Train AI?



Spoiler: yes!

Ofer Lavi, CEO
dataspan.ai

“Obtaining the right data for training computer vision models is super easy.”

No one, ever.



Challenges in data acquisition hinder CV potential



Data volume:

Large-scale requirements for robust models



Acquisition hurdles:

Time-consuming and resource-intensive



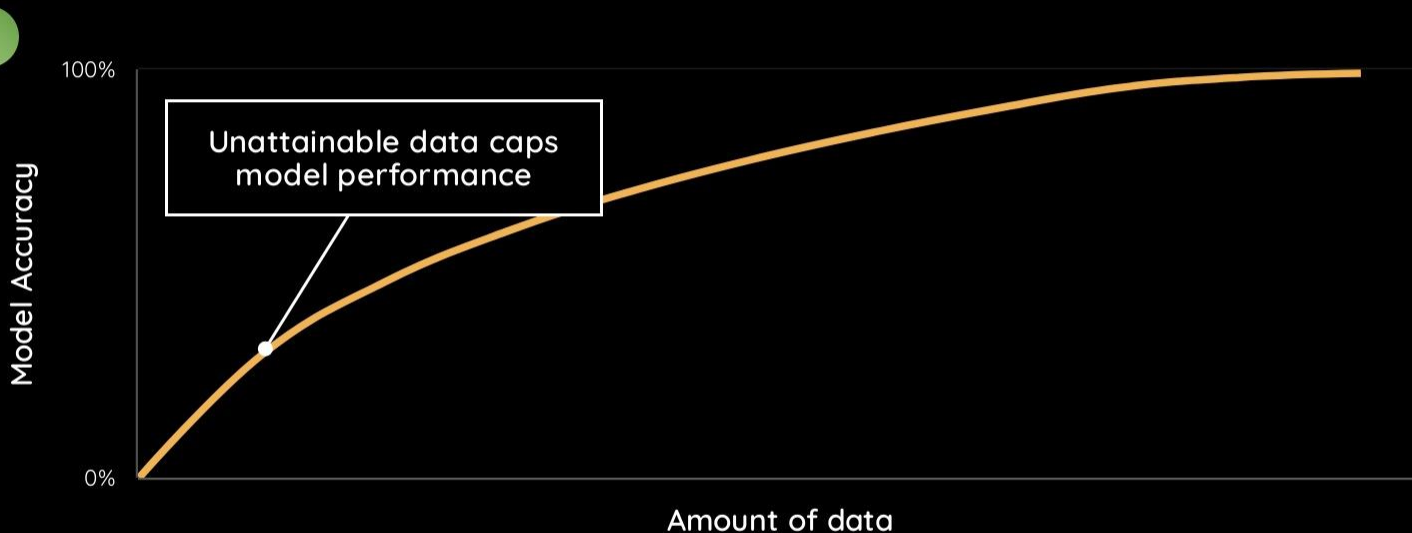
Annotation complexities:

Additional challenge in labeling data

80% of the time in a typical AI project spent on data

Bottom line: Traditional methods for data acquisition are insufficient to meet the needs of AI applications.

The "death valley" of unattainable data



Model performance is held back by unattainable data
Implications might be disastrous



GenAI
to the
rescue!



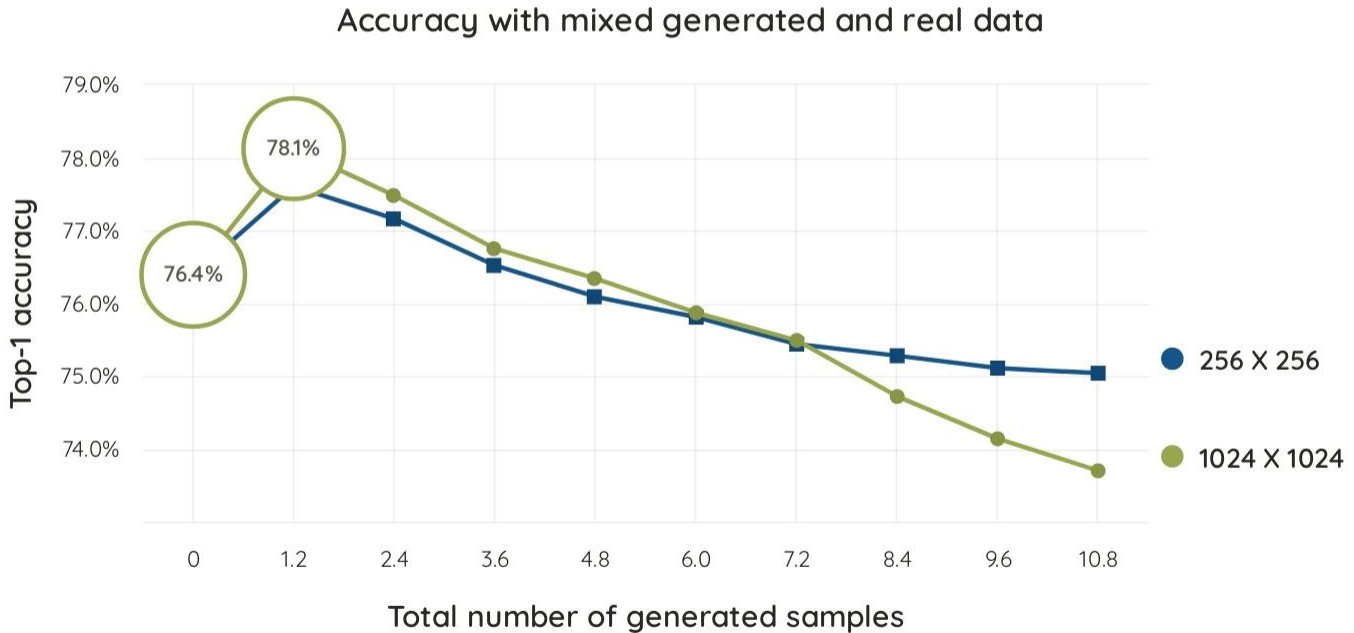
Diffusion models can improve ImageNet (Google Research, 2023)

+1.73%

HIGHEST ACCURACY IMPROVEMENT

1:1

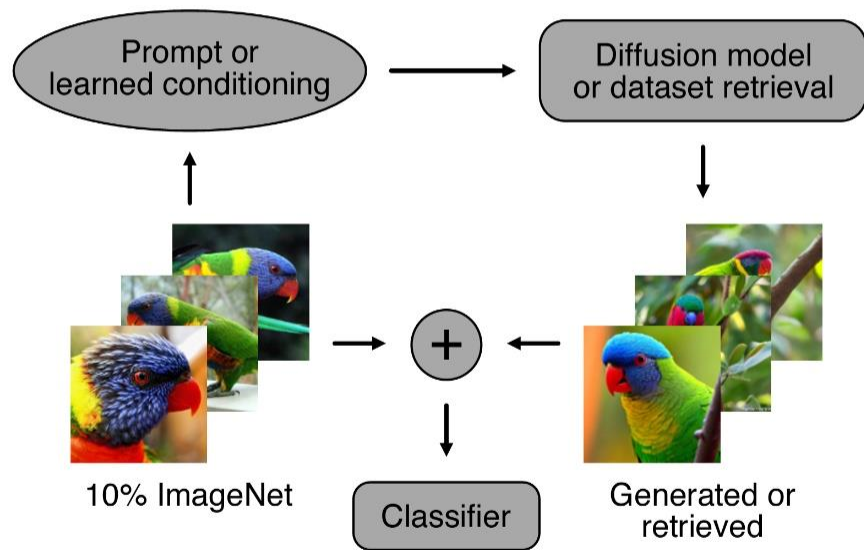
RATIO OF REAL:GENERATED DATA AT BEST ACCURACY



Azizi, Shekoofeh, et al. "Synthetic data from diffusion models improves imagenet classification." arXiv preprint arXiv:2304.08466 (2023).



Image retrieval outperforms diffusion models on data augmentation



57.2%

ImageNet accuracy on
10% of the data

60.9%

Accuracy
augmenting
with CLIP
prompts
generation

61.2%

Accuracy combining two
state of the art
generation methods

62.6%

Accuracy on **retrieved**
data from diffusion
model training set

NO GENERATION HERE!

Burg, Max F. et al. "Image retrieval outperforms diffusion models on data augmentation." (2023).





dataspan.ai

The **low-data** computer vision **rapid** development platform

dataspan.ai provides an intuitive visual platform with powerful API capabilities that leverage cutting-edge generative AI to automatically process data and train computer vision models.

Case study - predictive maintenance

Derailments per year

1000

Cost of a massive derailment occurred in 2023

\$803M

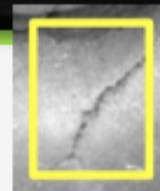
Cost of a small derailment

\$400,000

Images of damaged wheels in dataset

60

Collected in 6(!) years





Clear wheel

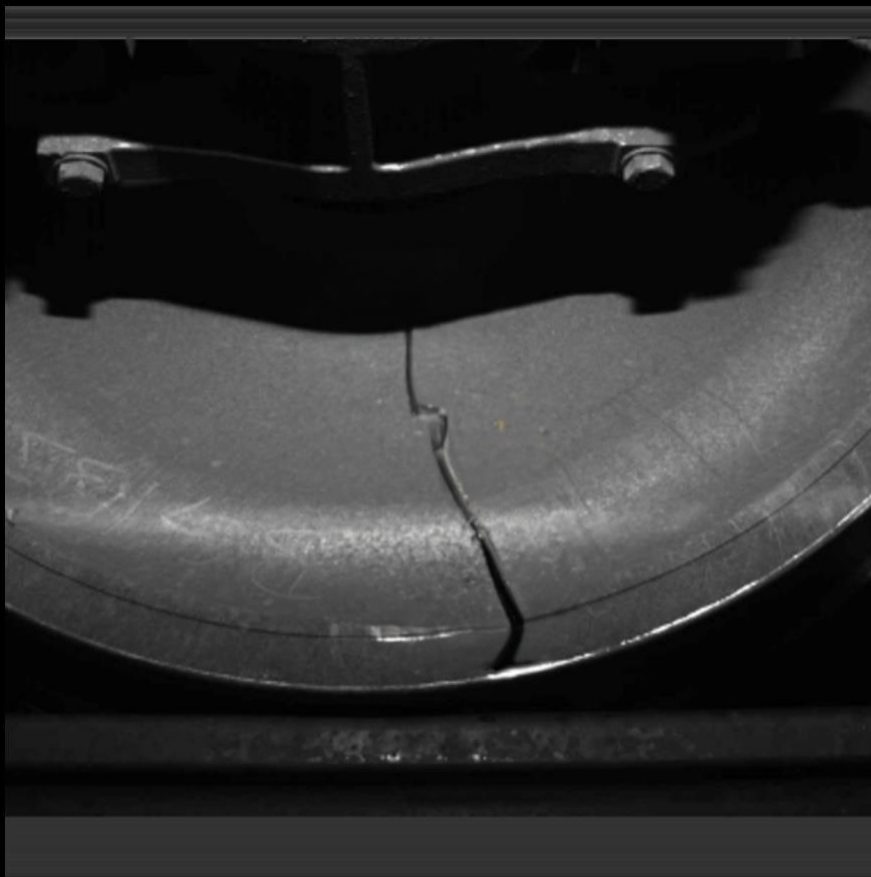




"a photo of a train wheel, taken from the level of the rail, where the wheel has a big crack that can cause a derailment"

Dall-E 2





Inpainting + prompt
*"a train wheel with a
big crack"*

Dall-E 2



No free lunch!



Keep original distribution



Add missing aspects



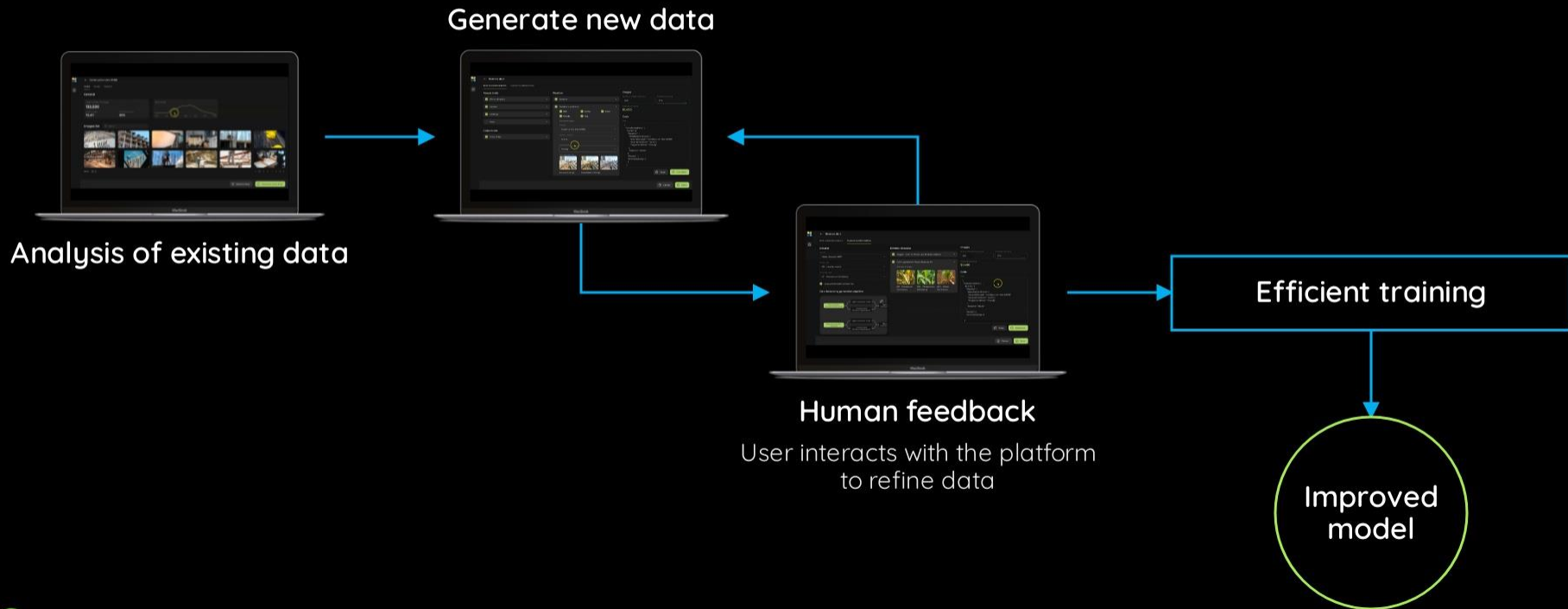
Make them as real as possible



Improve downstream model



How dataspan.ai works



The two questions to answer

Where?

dataspan.ai
Your data. Supercharged.

gmail Hubspot Calendar Zoom monday Dokka Drive Sales Navigator LinkedIn Notion Scholar Inbox buyme Invoice upload aut...

Datasets Concepts Generate Hello, John Smith JS

Add localization layer

Layer name

Choose localization type

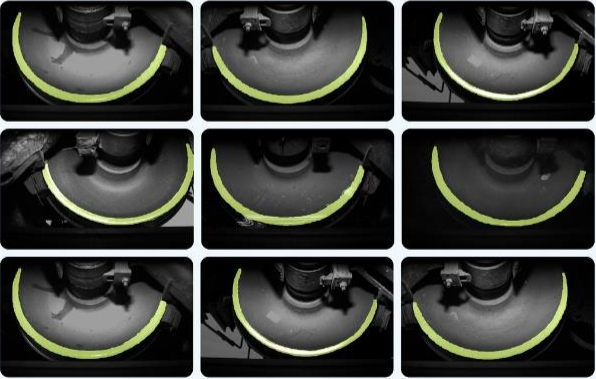
Choose localization algorithm

Positive prompt

Negative prompt

Preview

Show localization masks



1 2 3 4 5 6 >

SAVE LOCALIZATION LAYER

What?

dataspan.ai
Your data. Supercharged.

gmail Hubspot Calendar Zoom monday Dokka Drive Sales Navigator LinkedIn Notion Scholar Inbox buyme Invoice upload aut...

Datasets Concepts Generate Hello, John Smith JS

Creating concept

Concept name

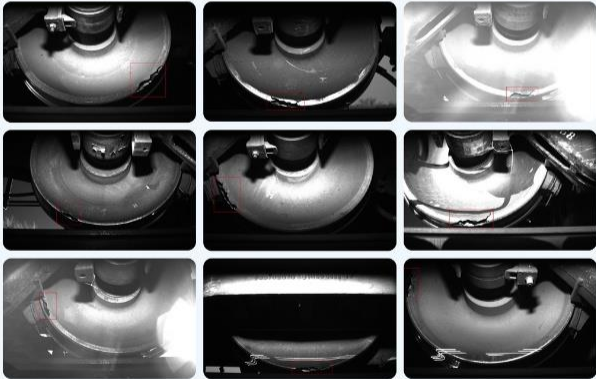
Concept prompt

Choose concept samples

Choose generation canvas

Preview

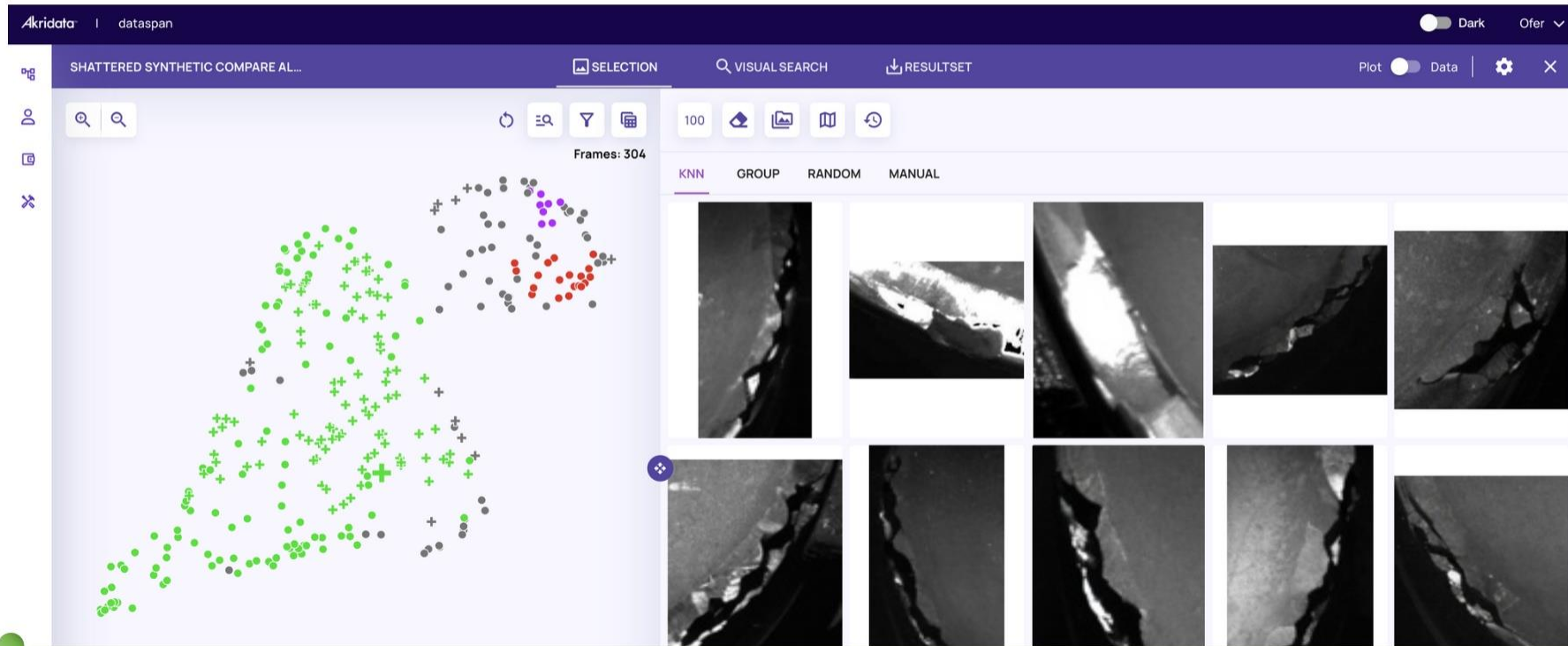
Show bounding boxes



CREATE CONCEPT



Generated data is similar to real data



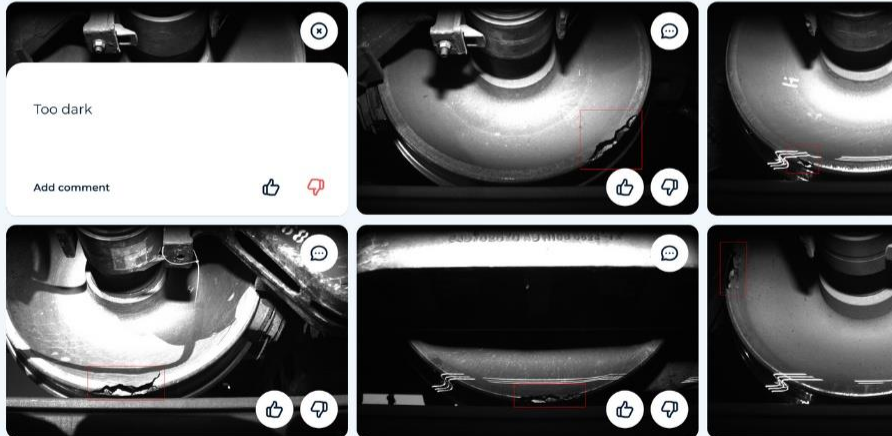
Iterative refinement



Datasets Concepts Generate

Refining concept

Please rate generated images



1 2 3 ... 7 8 9 >

REFINE PROGRESS

Subject matter expert provides feedback to improve the generation process
Visit our demo to learn more!

...cracks/2001770-4c6-4f92-a59d-aa121aac4100.jpg",49.876,"poor",
...cracks/2161190-a290-4074-d175-3a0390bfc456.jpg",84.242,"decent",
...cracks/29c21f52-4a1c-4100-4501-95bd640b6ace.jpg",10.853,"good",
...cracks/2a11c7e-4e7a-43fa-974d-8ad0635a8255.jpg",64.054,"decent",
...cracks/3012202-141e-49ec-a141-66beb0fa2c0a.jpg",8.47,"good",
...cracks/360604c-0485-4a47-0be5-96c92b6e341e.jpg",27.541,"decent",
...cracks/39ec075c-78ea-416d-000b-4728297eb4de.jpg",95.952,"poor",
...cracks/3c30231-41c3-4213-abc7-e3a08a43fac0.jpg",167.514,"poor",
...cracks/4050401-df50-4edc-06ae-0f0059ebb97f.jpg",77.641,"decent",
...cracks/4682e772-f952-4be2-0ecb-243cct408d73.jpg",29.943,"decent",
...cracks/4a00ae7-0001-4b46-805b-2c5bfa940d71.jpg",2040.092,"decent",
...cracks/525a0220-784c-49c3-0371-71744532763d.jpg",51.577,"good",
...cracks/520c9080-0935-4520-0f76-df10abda4232b.jpg",72.313,"decent",
...cracks/57403206-7102-4007-08c8-947ab55849cb.jpg",27.098,"decent",
...cracks/58437653-f910-45c3-0014-3d076423bca2.jpg",171.772,"poor",
...cracks/58a2730-7881-df10-0076-1819f30244e1.jpg",1029.694,"decent",
...cracks/5909538-377d-4b10-0076-1819f30244e1.jpg",22.000,"decent",



Clear wheel

Using a model built upon [dataspan.ai's](#) API 1.1M wheels examined weekly





dataspan.ai

Using a model built upon [dataspan.ai's](#) API 1.1M wheels examined weekly



Client case study - Defect detection in manufacturing

Challenge

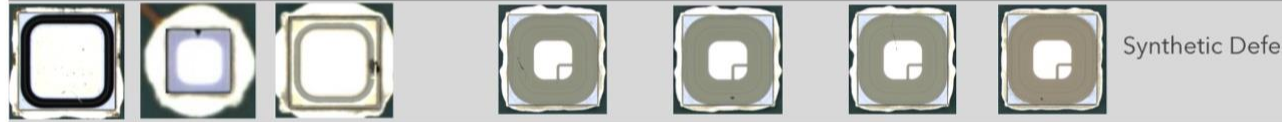
- Limited number of defect images in data, multiple components, multiple defect types

Result

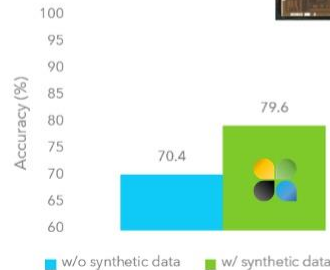
- PoC results - Reduction of up to 90% in error rate over a prominent competitor

Case 1b - Hybrid

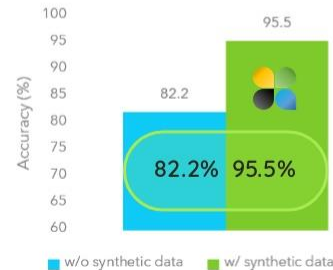
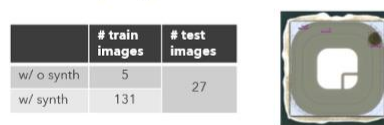
8-13% Performance Improvement when adding AI generated defect examples to training dataset



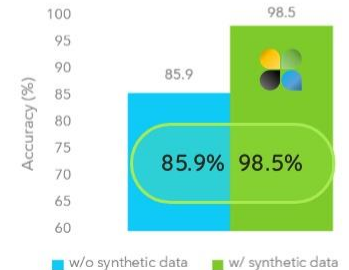
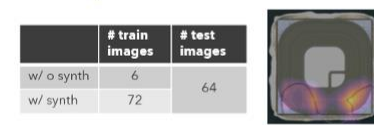
Anomaly Object Detection - All Die



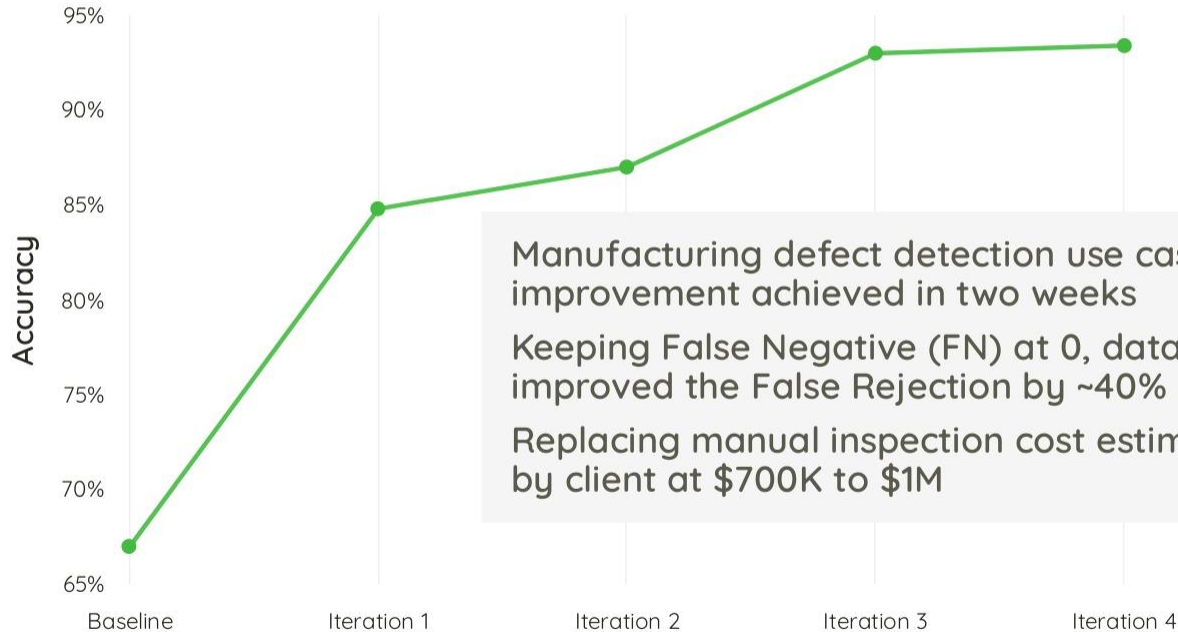
Anomaly Object Detection - One Die



Anomaly Classification - One Die



Defect detection accuracy



Manufacturing defect detection use case - improvement achieved in two weeks

Keeping False Negative (FN) at 0, dataspan improved the False Rejection by ~40%

Replacing manual inspection cost estimated by client at \$700K to \$1M



Thanks!

Scan the QR code for more info!

Joint work with:

Almog Elharar

Roy Toren

Elad Shaked

Sanjay Pichala and the Akridata team



Akridata[™]
AI Platform for Visual Data

PLUGANDPLAY





 <https://dataspan.ai>

 sales@dataspan.ai

 <https://www.linkedin.com/company/dataspan-ai>

730 Arizona Ave, Suite 206, Santa Monica,
California, 90401 US

