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.

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

Challenges in data acquisition hinder CV potential

Data volume:

Large-scale requirements for robust models

Acquisition hurdles:

Time-consuming and resource-intensive

Annotation complexities:

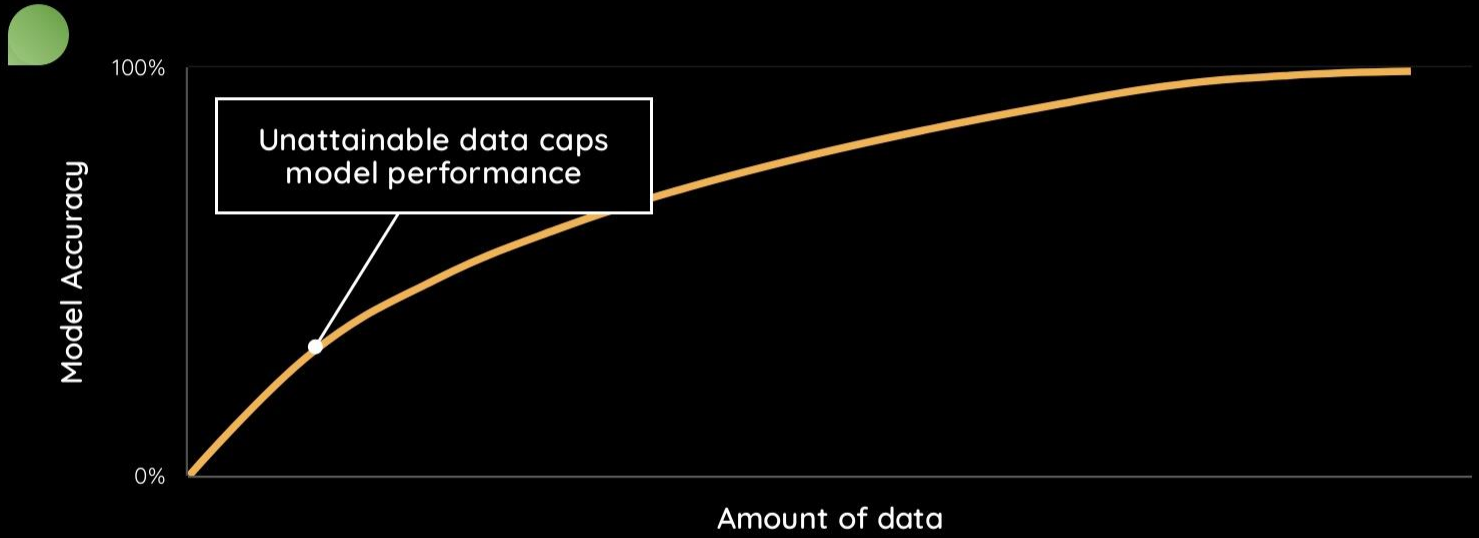
Additional challenge in labeling data

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.

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

The *"death valley"* of unattainable data



Model performance is held back by unattainable data

Implications might be disastrous

IMVC 2024

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

GenAI to the rescue!

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

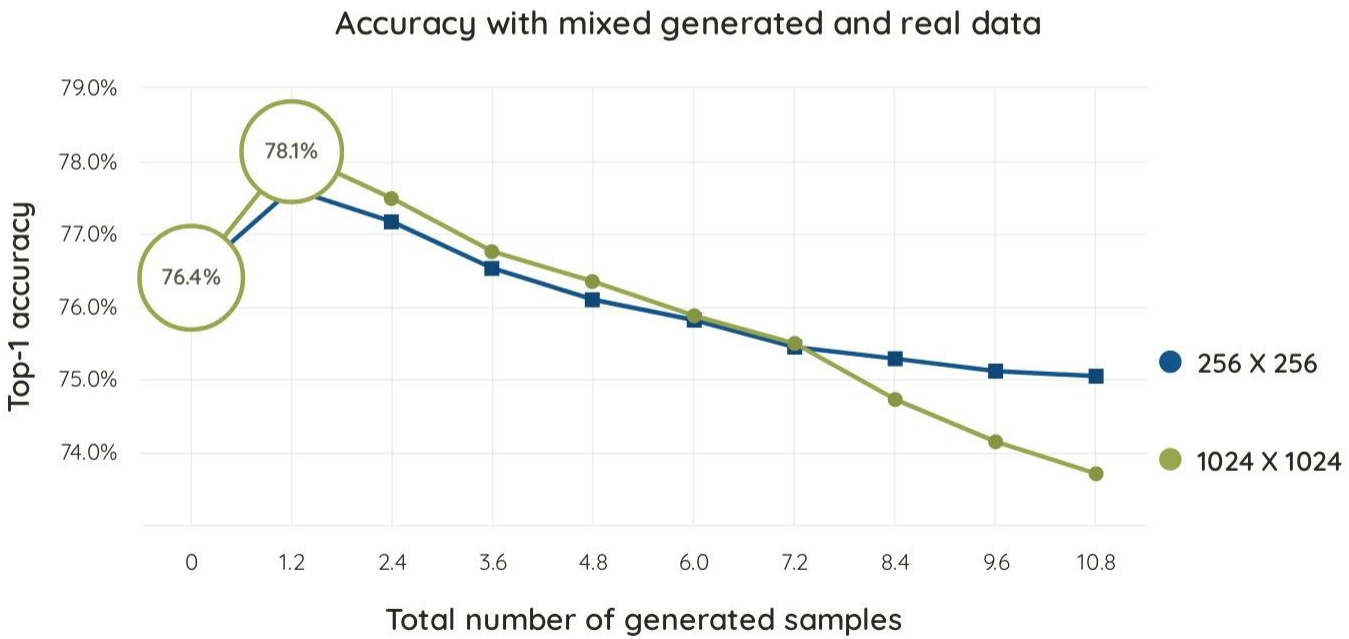
Diffusion models can improve ImageNet (Google Research, 2023)

+1.73%

HIGHEST ACCURACY IMPROVMENT

1:1

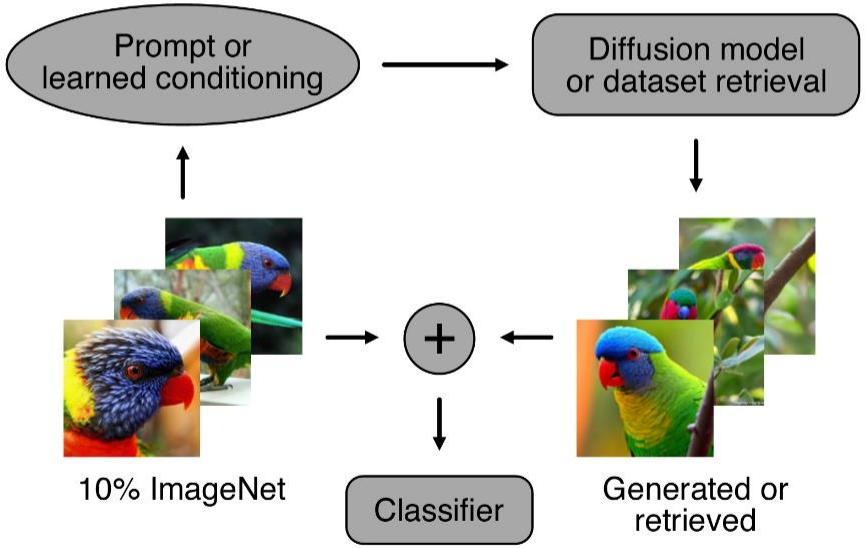
RATIO OF REAL:GENERATED DATA AT BEST ACCURACY



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

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

Image retrieval outperforms diffusion models on data augmentation



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

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!

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

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.

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

Case study - predictive maintenance

Derailments per year

1000

Cost of a massive derailement occured in 2023

$803M

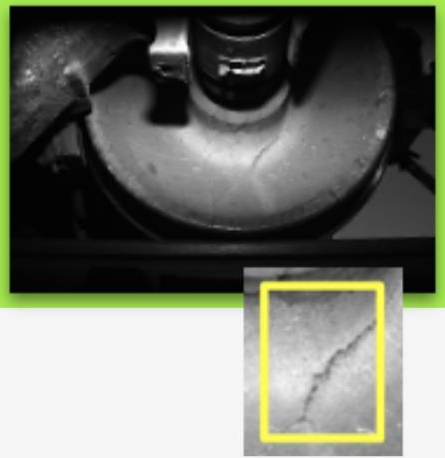
Cost of a small derailment

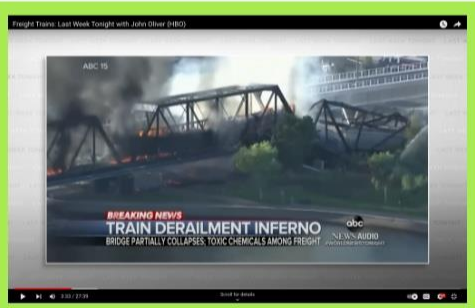
$400,000

Images of damaged wheels in dataset

60

Collected in **6(!)** years





©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

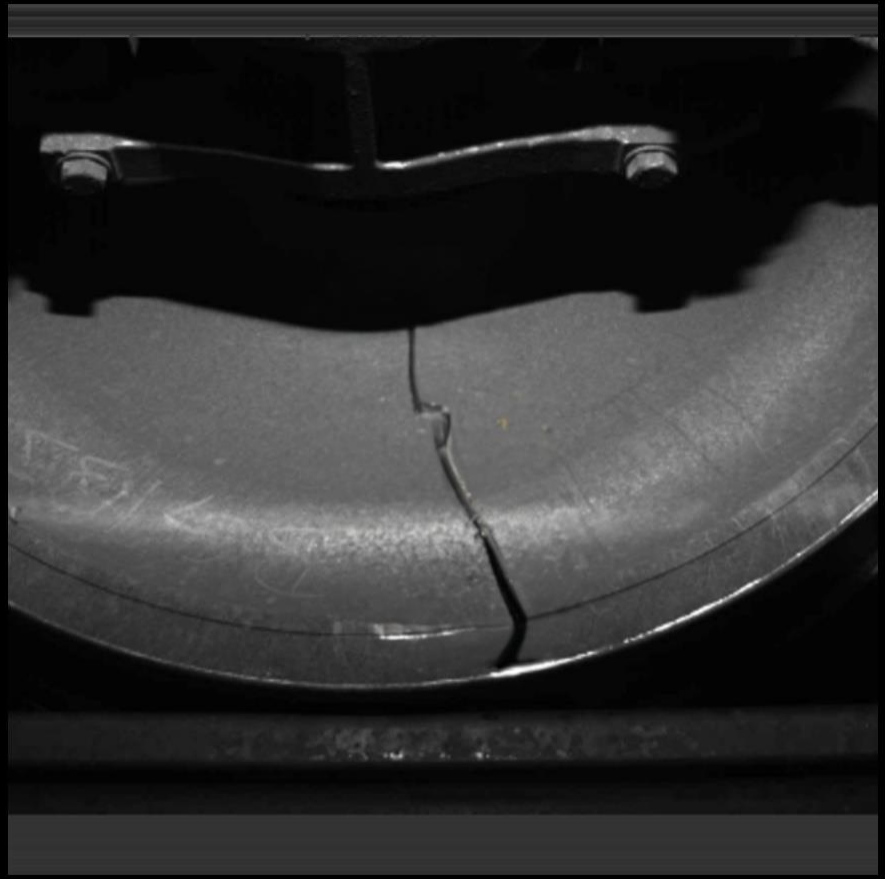
Clear wheel

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

"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

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.



Inpainting + prompt

"a train wheel with a big crack"

Dall-E 2

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

No free lunch!

Keep original distribution

Add missing aspects

Make them as real as possible

Improve downstream model

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

How dataspan.ai works

Analysis of existing data

Generate new data

Human feedback

User interacts with the platform to refine data

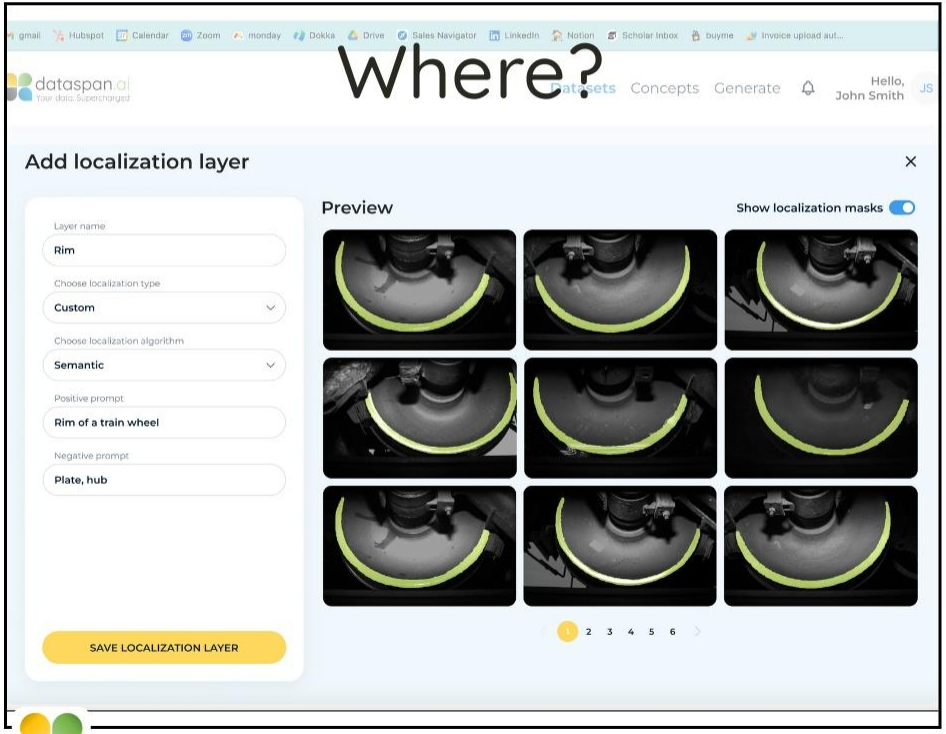
Efficient training

Improved model

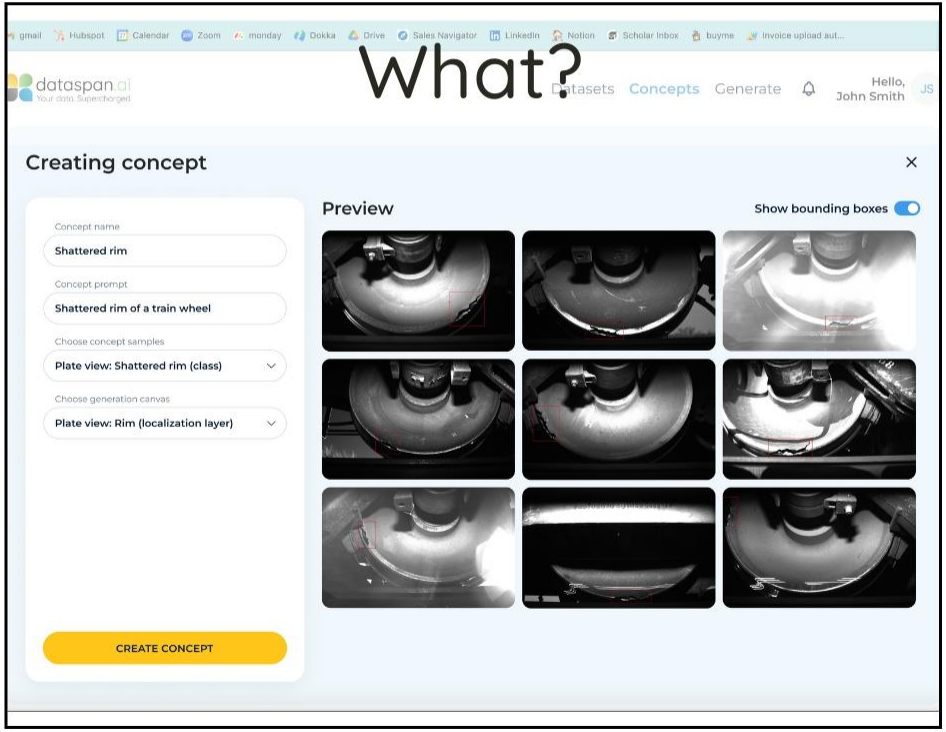
©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

The two questions to answer

Where?

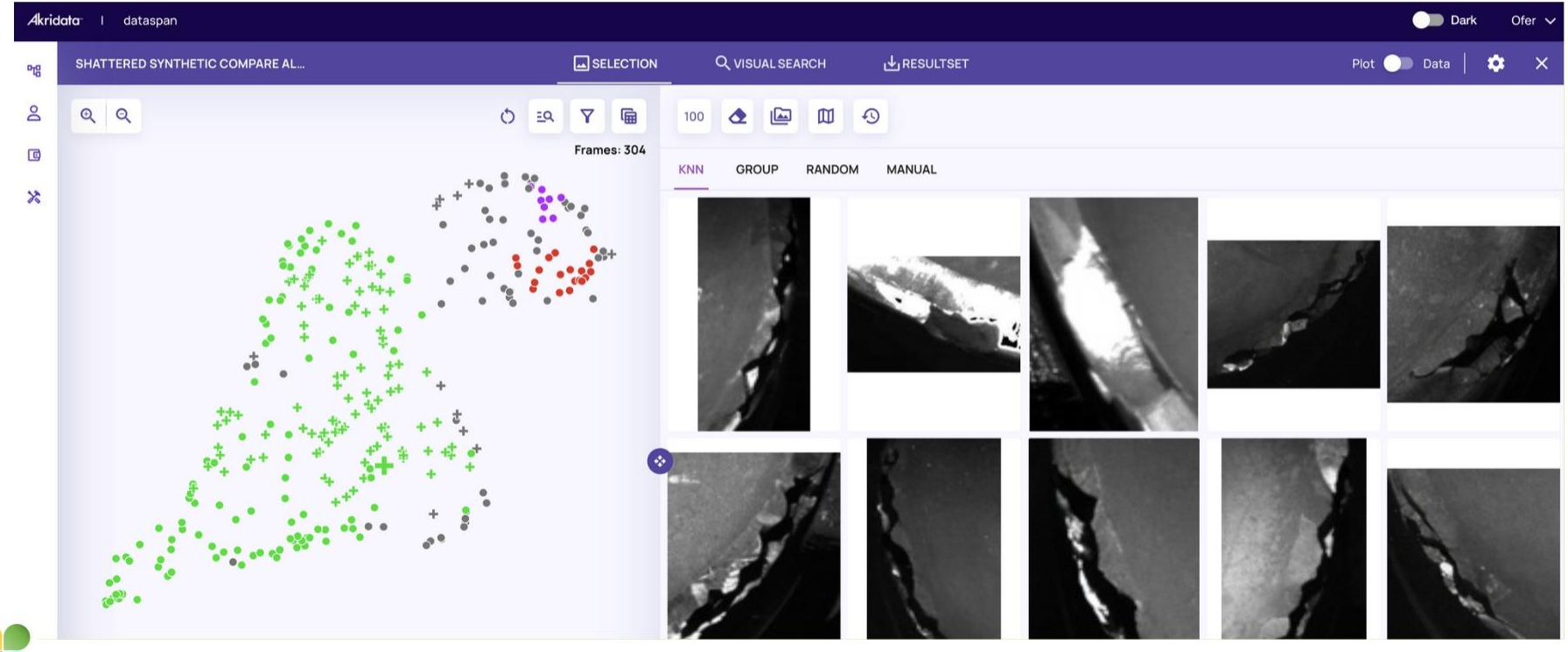


What?



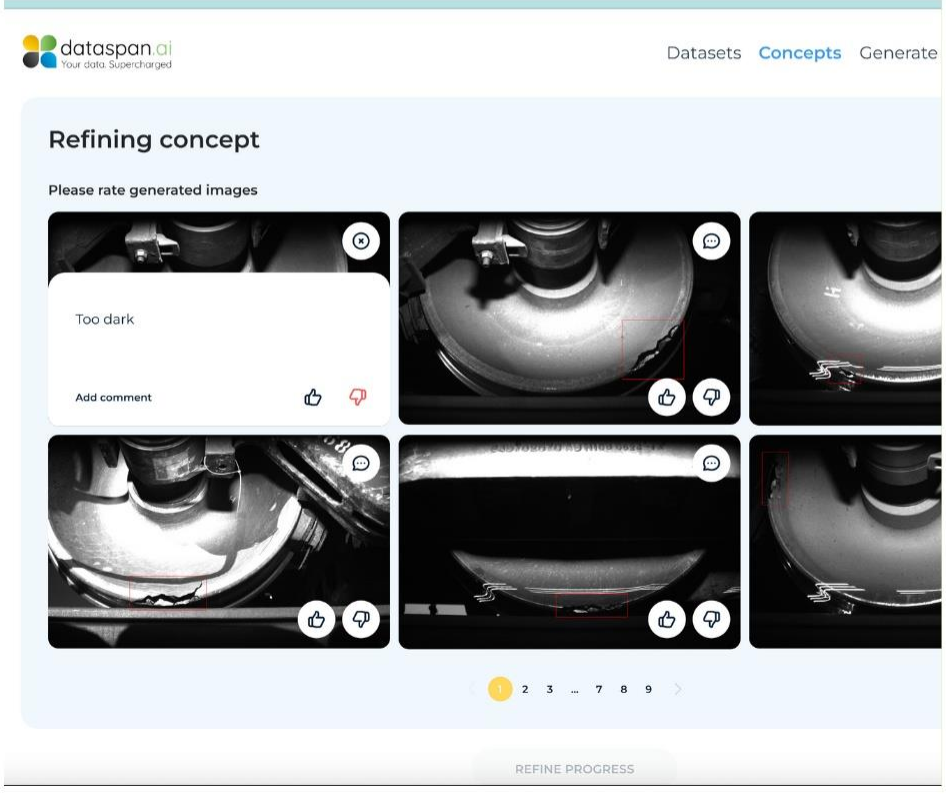
©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

Generated data is similar to real data



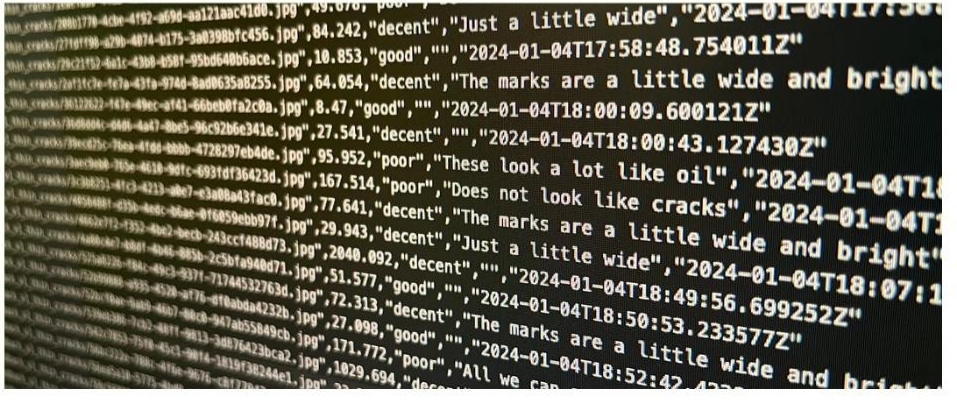
©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

Iterative refinement



Subject matter expert provides feedback to improve the generation process

Visit our demo to learn more!



Clear wheel

Using a model built upon dataspan.ai's API 1.1M wheels examined weekly

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

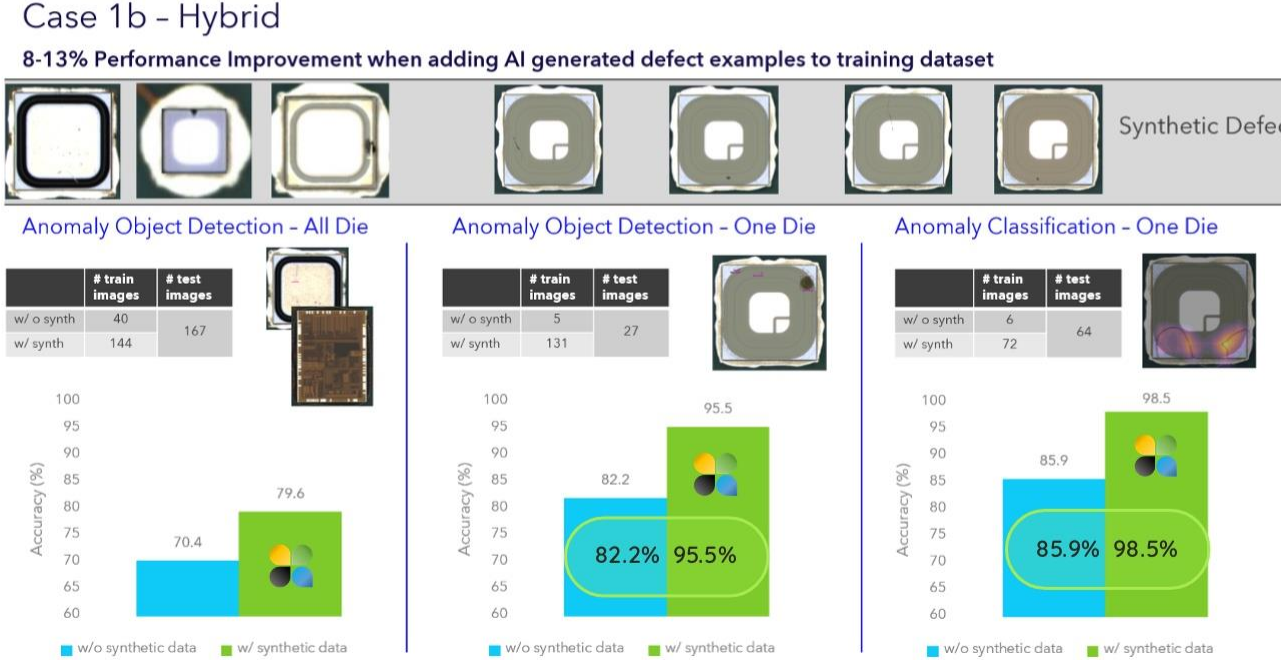
dataspan.ai

Using a model built upon dataspan.ai's API 1.1M wheels examined weekly

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

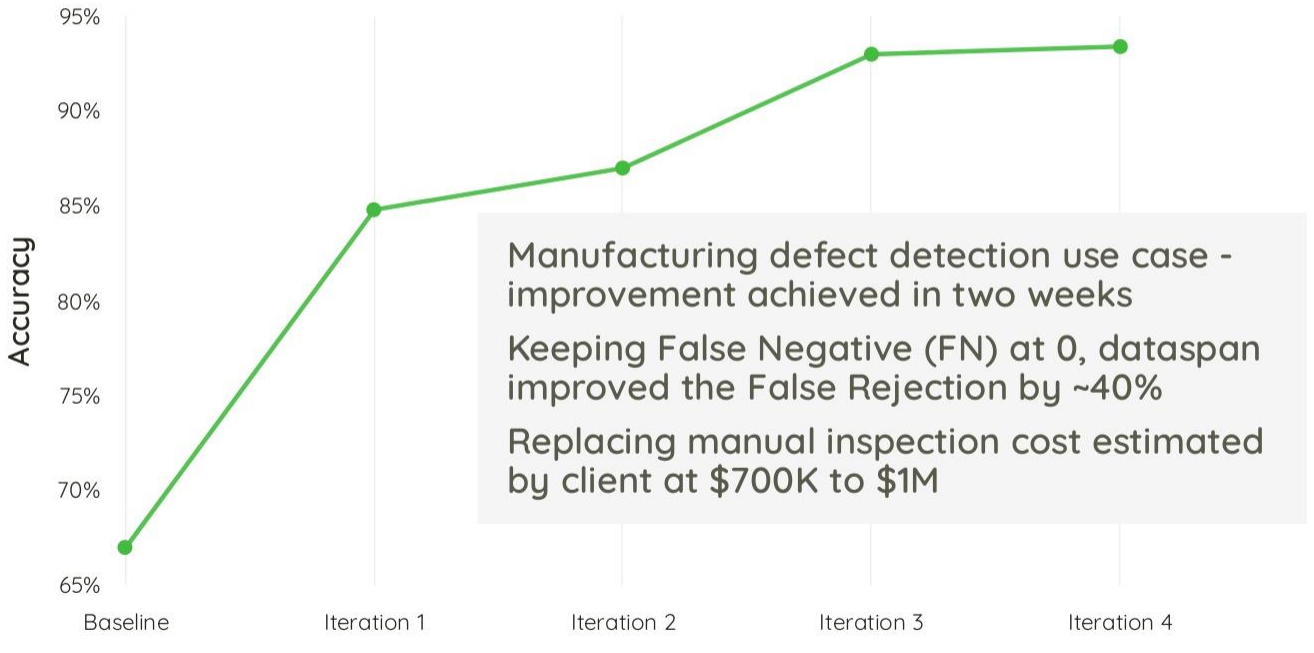
Client case study - Defect detection in manufacturing

- **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**



©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

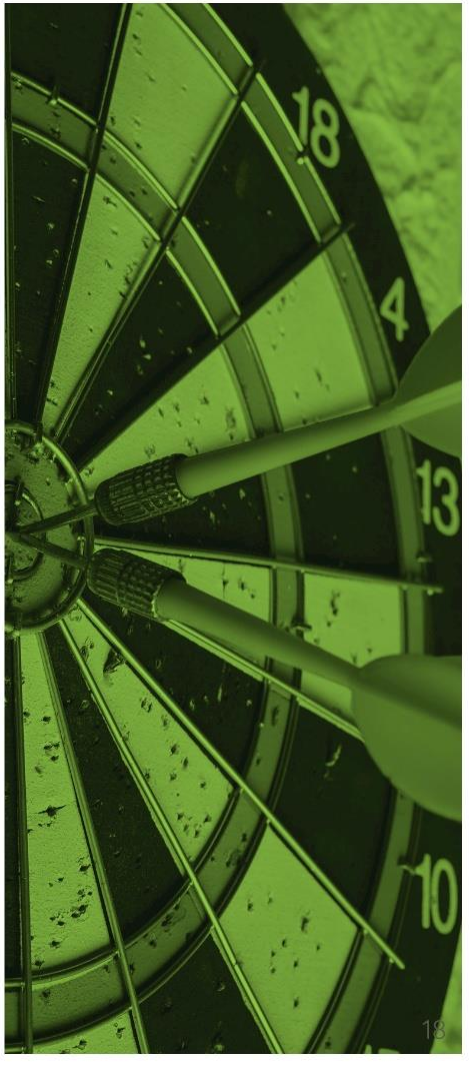
Defect detection accuracy



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



©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

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

NVIDIA INCEPTION PROGRAM

aws startup loft

FUSION

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.

dataspan.ai

https://dataspan.ai

sales@dataspan.ai

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

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

©2024 Dataspan AI Inc. Proprietary and Confidential. All Rights Reserved.