

Synthetic Image Data in Geopolitical Events and Armed Conflicts

SICSS-Istanbul

July 8, 2023

Boğaziçi University
Ergin Selim Gönen

Agenda

- What is synthetic data?
- Examples in geopolitical events and armed conflicts
- Tools and workflows

About me

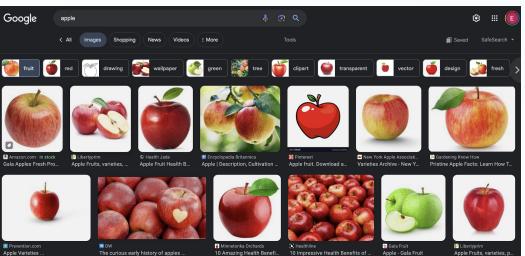
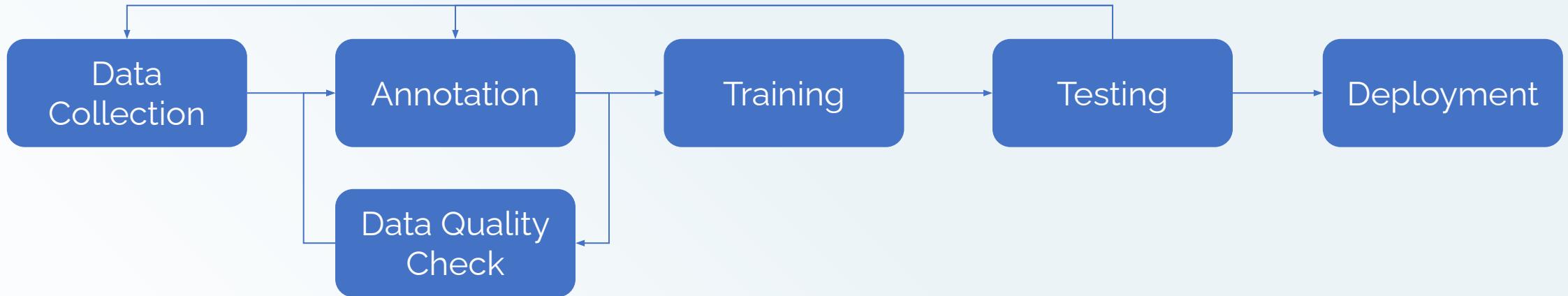
- B.S. Electrical and Electronics Engineering – Bilkent University
- M.B.A. – Sabanci University
- Strategy Analyst – Arama Consulting
- Strategy Analyst – Turk Telekom
- Freelance Strategy Consulting
- Strategy Consulting/Founder – Pretcog
- Synthetic Data Engineer/Founder – Mirage

What is synthetic data?

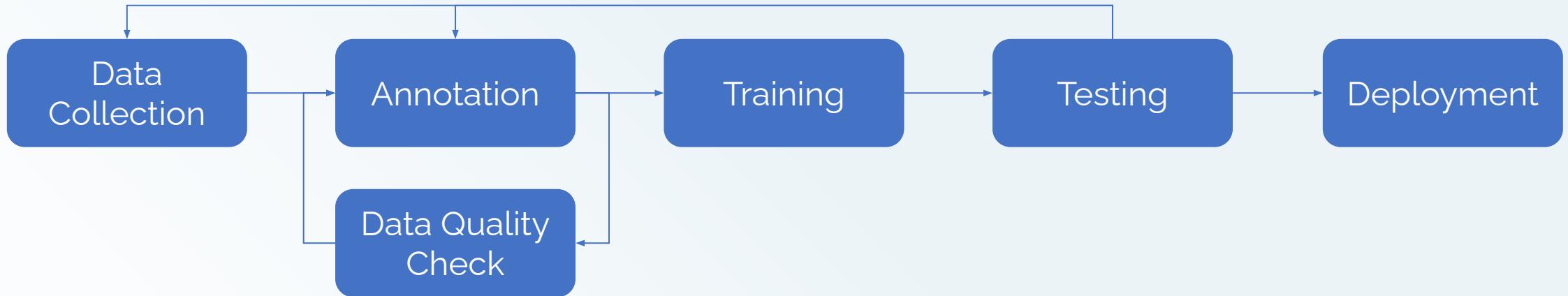
Synthetic data is annotated information that computer simulations or algorithms generate as an alternative to real-world data¹

¹: <https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/>

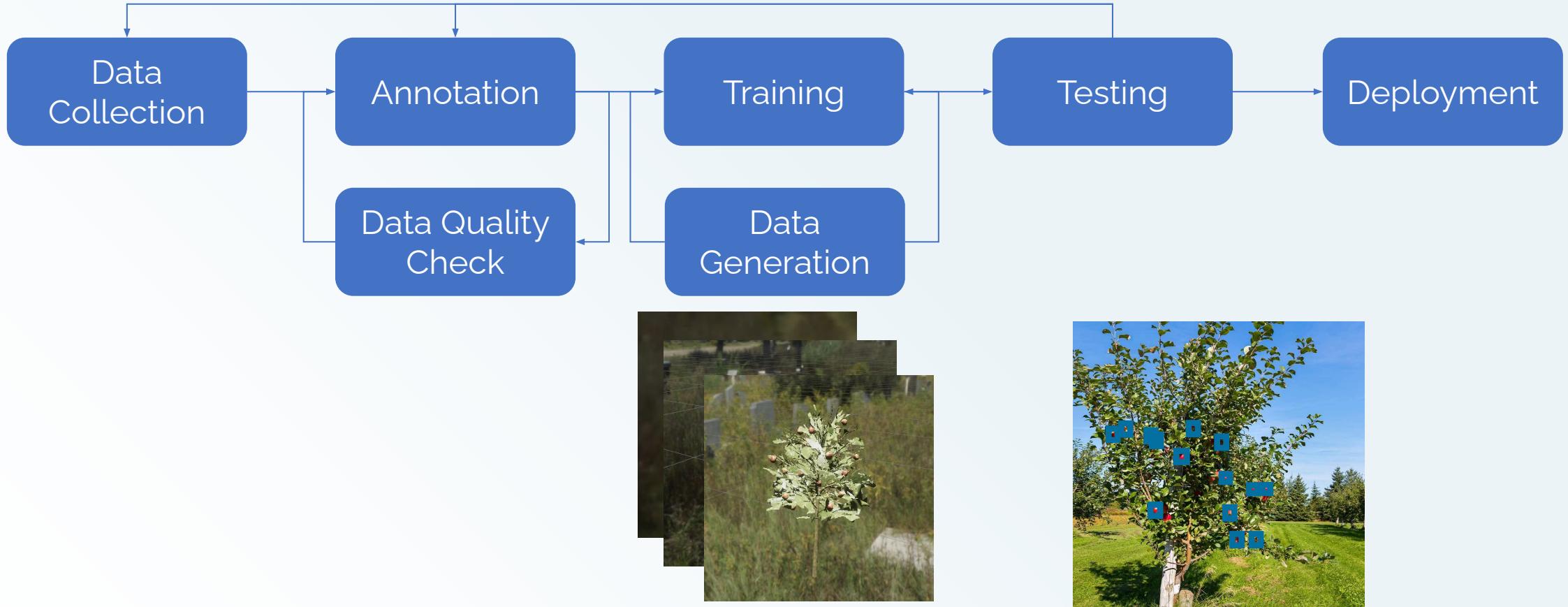
How a computer acquires sight?



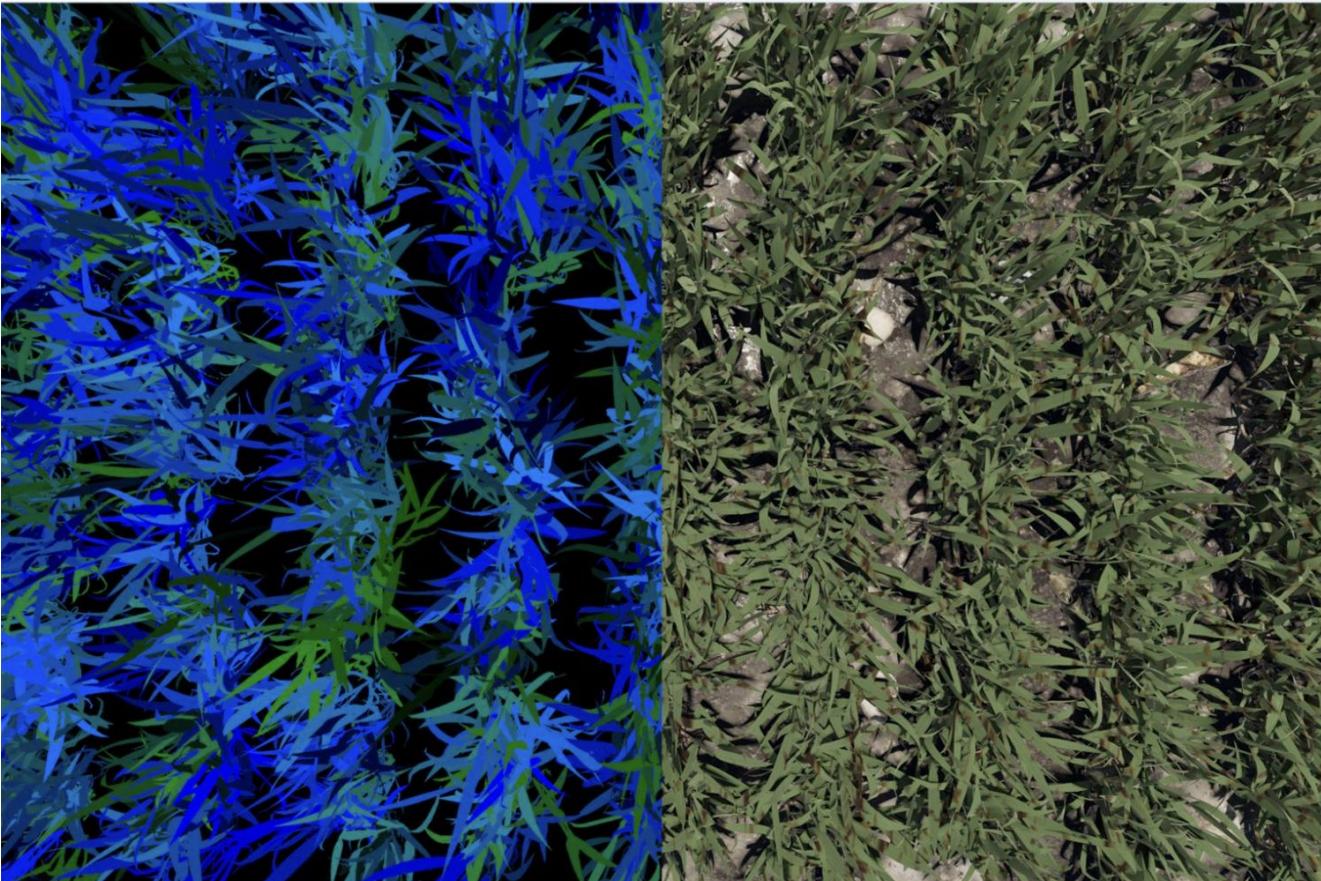
How a computer acquires sight?



How a computer acquires sight?



What is the point?



1. Procedurally generated; virtually unlimited variety - **no expensive data collection, no bias**
2. Pixel precise annotation - **no annotation errors, custom annotations are feasible & scalable**
3. No real-world connection - **no privacy concerns, regulation friendly**

How to create it?

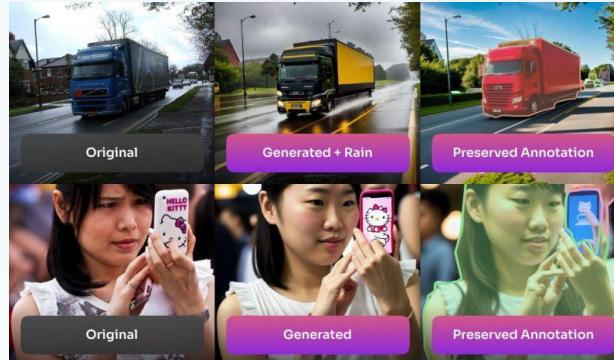
Tabular

Text

Audio

Sensor

Image



Using existing data

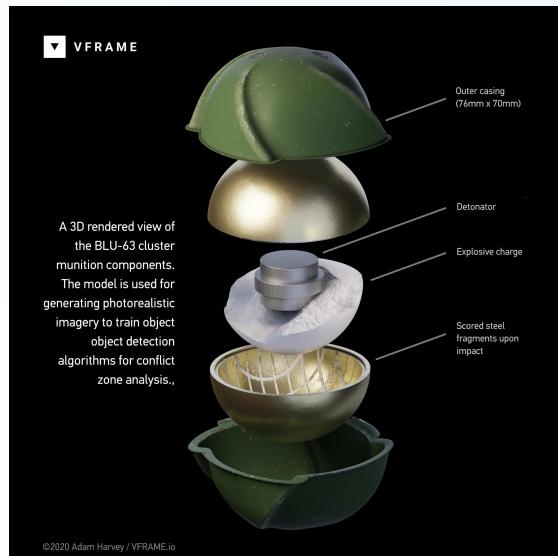


From scratch

1: Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, TsungYi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple 6 copy-paste is a strong data augmentation method for instance segmentation. In CVPR, 2021.
2: <https://www.kopikat.co>

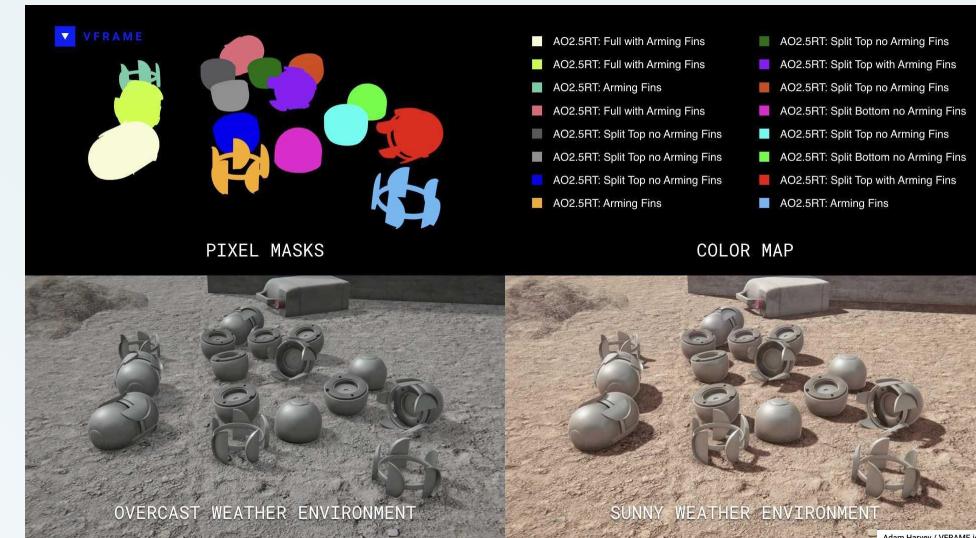
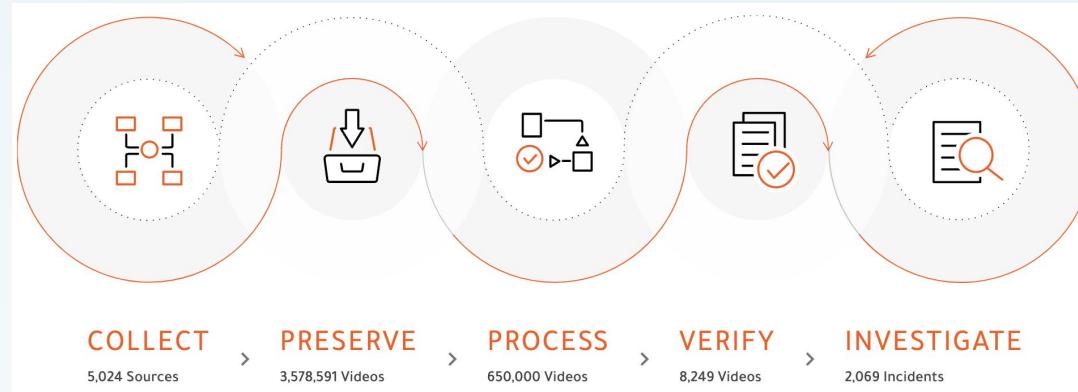
Questions?

Warcrimes in Yemen



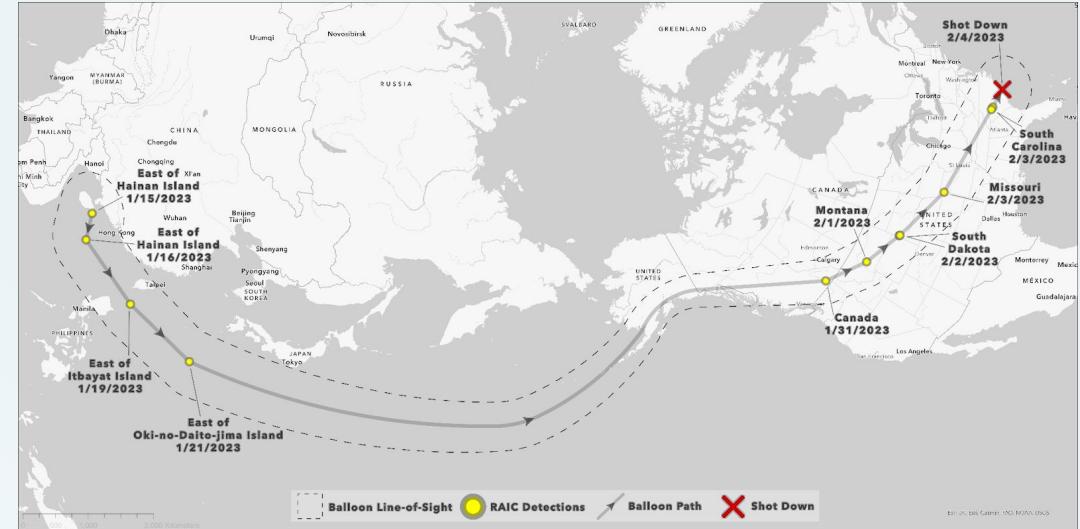
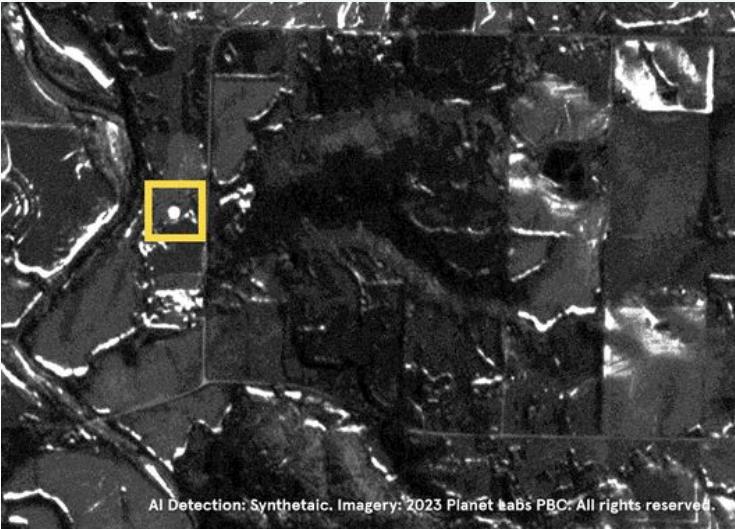
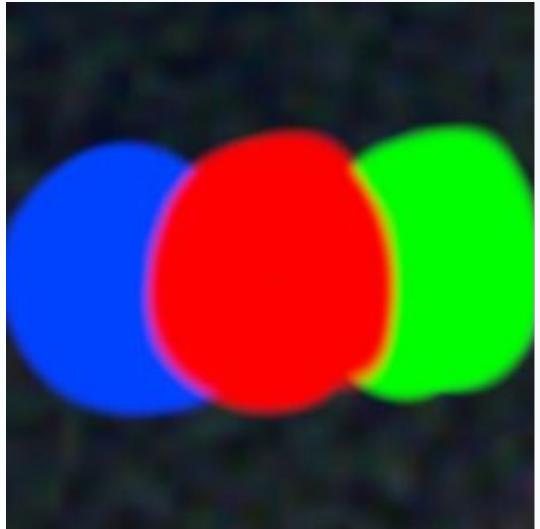
- 1: <https://vframe.io/>
- 2: <https://yemeniarchive.org/>
- 3: <https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/>

Warcrimes in Syria



- 1: <https://syrianarchive.org/>
- 2: <https://www.ft.com/content/8399873e-0dda-4c87-ba59-0e2678166fba>
- 3: <https://vframe.io/>

Chinese weather/spy balloon



- 1: <https://www.synthetaic.com/>
- 2: <https://www.bbc.com/news/world-65972168>
- 3: <https://www.wired.com/story/how-one-guys-ai-tracked-the-chinese-spy-balloon-across-the-us/>

SICKS

Demining Ukraine

 **BlogYourEarth**
@Ingmar_Stapel

In the late evening I was just playing with my stable diffusion neural net generating a few PFM-1 mine pictures like those sixteen... Not perfect but it is now going in a good direction to generate some synthetic data... But just test data... [@DemineTech](#) [@whatisthe_love_tv](#) [@adamhvr](#)



9:48 PM · Apr 11, 2023 · 3,519 Views

3 Retweets 1 Quote 38 Likes 1 Bookmark

 ergin selim gonen @selimgonen · Apr 12
Hey, looking awesome.
Would you mind if I ask couple of questions about your generation process and experience?

 BlogYourEarth @Ingmar_Stapel · Apr 12
No problem just ask.

 ergin selim gonen @selimgonen · Apr 15
How scalable do you think your workflow is in terms of prompt creation, image generation and computationally? How long did it take to generate the dataset and how many images did you generate? Have you considered using a game engine?

Possible use case examples – Crowd Counting



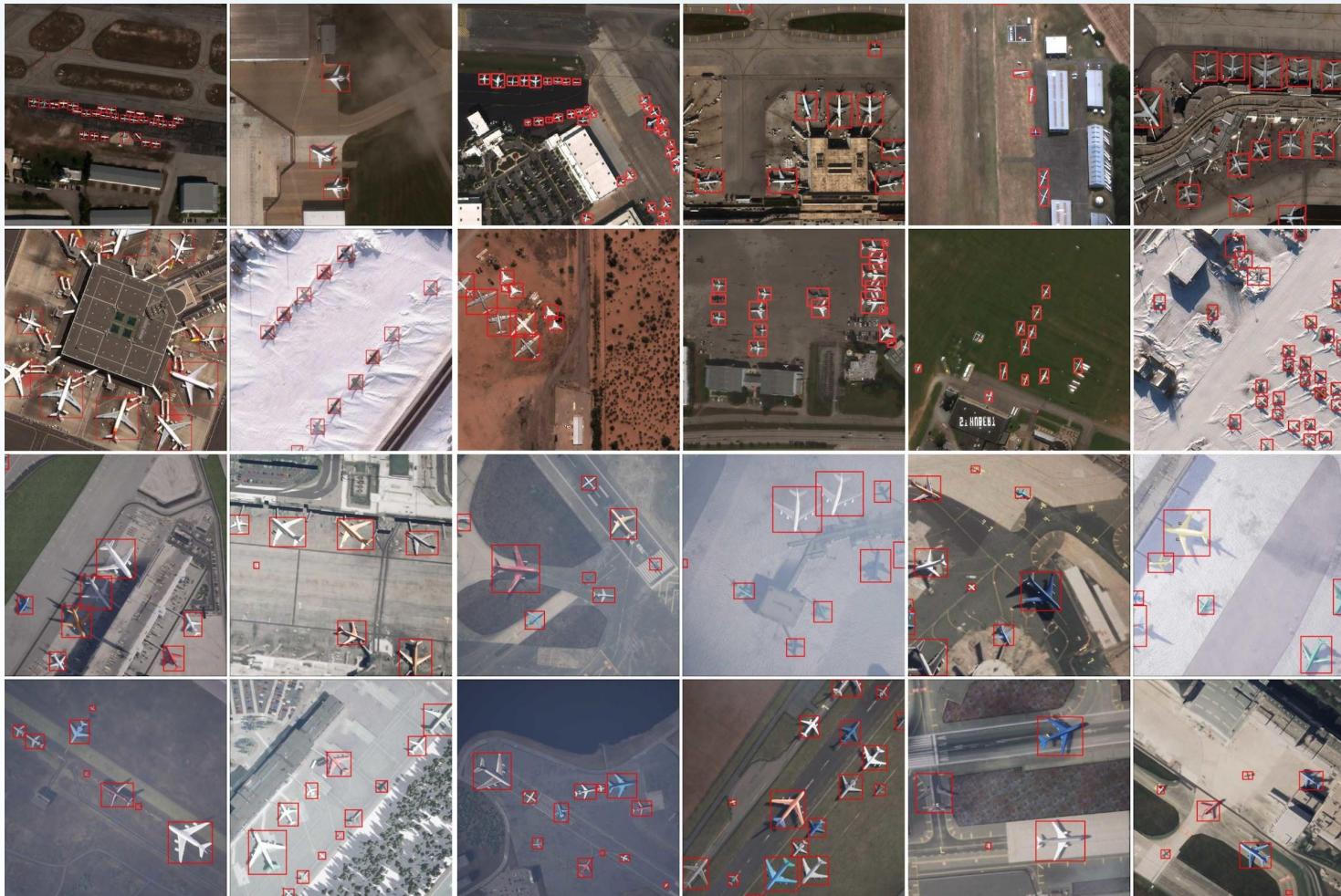
1: Qi Wang, Junyu Gao, Wei Lin, and Yuan Yuan. Pixel-wise crowd understanding via synthetic data. International Journal of Computer Vision, 129:225–245, 2021.

Possible use case examples – Emotion detection



1: Wood E, Baltrušaitis T, Hewitt C, et al., 2021. Fake it till you make it: Face analysis in the wild using synthetic data alone. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), p.3681-3691.

Possible use case examples – Geospatial datasets



1: J. Shermeyer, T. Hossler, A. Van Etten, D. Hogan, R. Lewis, and D. Kim, "Rareplanes: Synthetic data takes flight," arXiv:2006.02963, 2020

Possible use case examples – Rare object detection to enforce agreements



Figure 5: Sample synthetic manipulator arm image.



Figure 2: Real-world image of remote manipulator arm. Image credit: Savannah River National Laboratory via Twitter: <https://twitter.com/srnlab/status/983767392265306112>

1: Gastelum, Z. N., Shead, T., & Higgins, M. (2020). Synthetic Training Images for Real-World Object Detection. Proceedings of the Institute of Nuclear Materials Management. Virtual.

Smart war machines will proliferate

AI.Reverie Awarded \$950 Million IDIQ Contract for U.S. Air Force Advanced Battle Management System (ABMS)

Opportunity for AI.Reverie's Synthetic Data to Advance Computer Vision Capabilities of the U.S. Air Force

NEWS PROVIDED BY

[AI.Reverie](#) →

13 Jan, 2021, 08:01 ET

SHARE THIS ARTICLE



NEW YORK, Jan. 13, 2021 /PRNewswire/ -- AI.Reverie has been awarded a \$950,000,000 ceiling indefinite-delivery/indefinite- quantity contract for the maturation, demonstration and proliferation of capability across platforms and domains, leveraging open systems design, modern software and algorithm development in order to enable Joint All Domain Command and Control (JADC2). This contract is part of a multiple award multi-level security effort to provide development and operation of systems as a unified force across all domains (air, land, sea, space, cyber, and electromagnetic spectrum) in an open architecture family of systems that enables capabilities via multiple integrated platforms.

Today's contract follows collaboration between the Air Force and AI.Reverie in 2020 to improve AI-guided navigation and intelligence gathering. AI.Reverie generates fully annotated data at scale with an efficiency that makes AI training fast, flexible and productive. Gartner recently recognized the technology by naming AI.Reverie a "Cool Vendor" in AI Core Technologies.

AI.Reverie co-founder and CEO Daeil Kim said, "We are honored that the Air Force selected AI.Reverie to support its Advance Battle Management System. We believe that in partnership with AI.Reverie, the Air Force will have a significant opportunity to improve mission critical vision algorithms that ensure military advantage and keep our troops safe."

1: <https://www.prnewswire.com/news-releases/aireverie-awarded-1-5m-afwerx-phase-2-sbir-contract-from-us-air-force-301091960.html>

2: <https://www.armysbir.army.mil/topics/sensor-synthetic-data-generation/>

3: <https://www.businesswire.com/news/home/20210520005363/en/Synthetaic-Awarded-AFWERX-Contract-to-Leverage-Synthetic-Data-and-AI-Powered-Object-Detection>

SBIR STTR

HOME WHO WE ARE SCHEDULE ELIGIBILITY TOPICS OUR INNOVATORS NEWS FAQ & RESOURCES CONTACT US

Artificial Intelligence/Machine Learning, ASA(ALT), Direct to Phase II

Sensor Synthetic Data Generation

Release Date: 10/12/2021 Open Date: 10/26/2021 Application Due Date: 11/30/2021 Close Date: 11/30/2021

Solicitation: 21.4 Topic Number: A214-42 Duration: Up to 24 months Amount Up To: 1.7M

Objective

US Army requires large-scale, accurate and easily accessible training, test, and validation data to support AI model development for multiple security domains (e.g. SIPR, JWICS...). Sensor data is critical to develop AI/ML models. Unfortunately, there is not enough data yet to create highly performant models. Sensor Synthetic Data Generation will potentially serve to reduce bottleneck of training data supply that helps improve ML models by developing a synthetic data generation tool.



Synthetaic Awarded AFWERX Contract to Leverage Synthetic Data and AI-Powered Object Detection

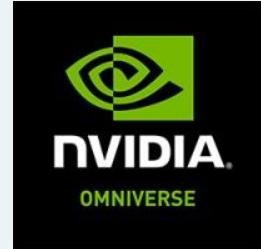
Synthetaic provides the US Air Force with faster ways to monitor change, extract key insights, and enable automation through AI models with minimal human labeling.

May 20, 2021 09:00 AM Eastern Daylight Time

DELAFIELD, Wis.--(BUSINESS WIRE)--Synthetaic, the leading synthetic data company, today announced it has won a contract with AFWERX, a program of the USAF focused on expanding technology, talent, and transition partnerships for rapid and affordable commercial and military capability. Synthetaic's Rapid Automatic Image Categorization technology (RAIC) will enable rapid object labeling, AI modeling, and Earth Observation object detection by analyzing large, multidimensional imagery datasets.



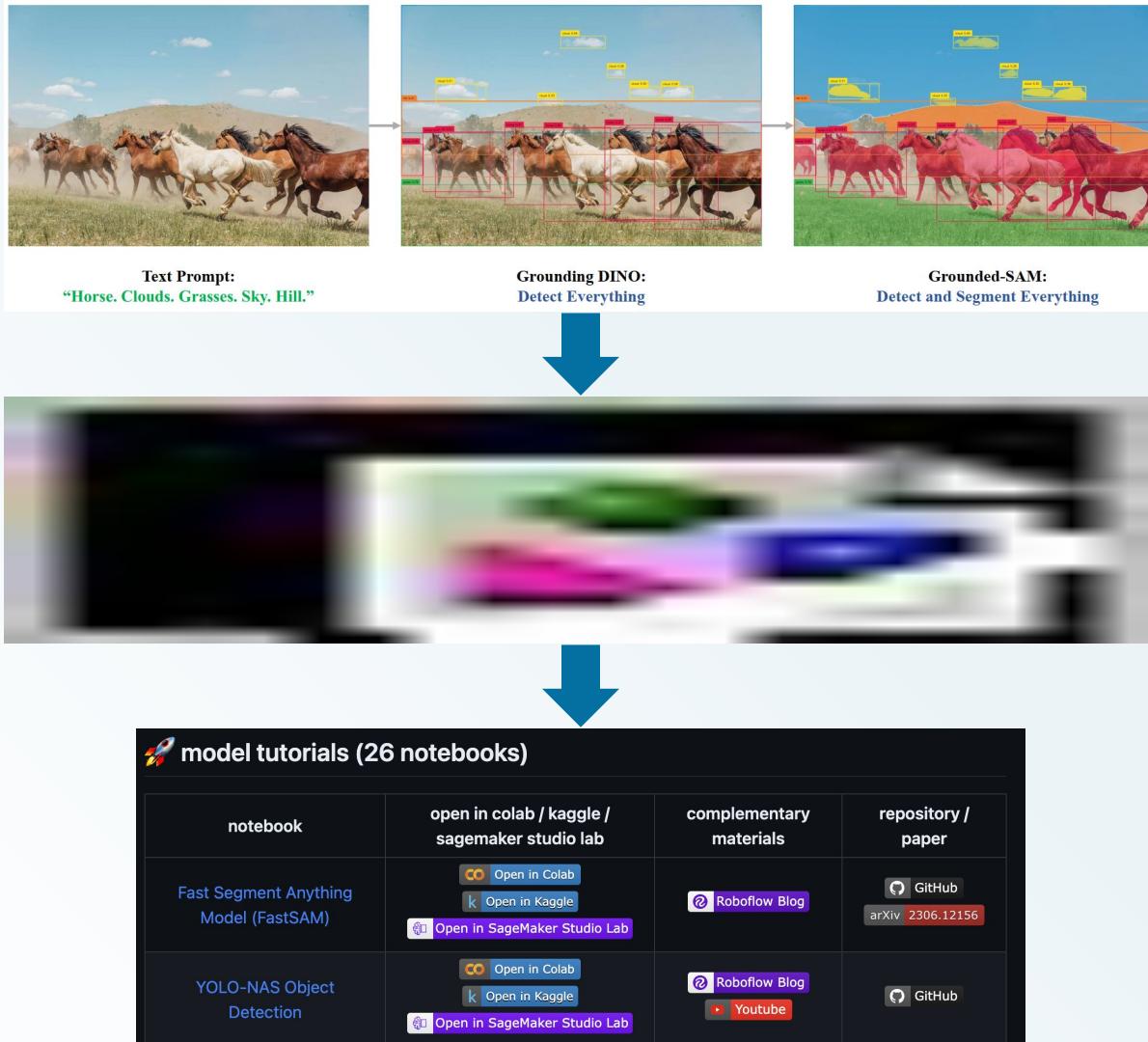
Synthetic data tools – 3D rendering based



- Open source
- Open source synthetic data libraries
- Huge community
- One tool for the whole workflow
- Python – API First Tool
- Kubric
- BlenderProc
- Infinigen
- Free to use¹
- Official, open source synthetic data library
- Real-time rendering
- Huge community
- Free to use²
- No official synthetic data library
- Realistic real-time rendering
- Huge community
- Free to use for individuals
- First class tool for synthetic data
- Realistic real-time rendering
- Node based system

1: <https://support.unity.com/hc/en-us/articles/208610336-What-subscription-tiers-are-available>
2: <https://www.unrealengine.com/en-US/faq>

An example workflow



- 1: <https://github.com/IDEA-Research/Grounded-Segment-Anything>
- 2: <https://github.com/andrewjouffray/Composite-Image-Generator>
- 3: <https://github.com/roboflow/notebooks>

Synthetic Image Data in Geopolitical Events and Armed Conflicts

esgnn@mirage.vision
@selimgonen

<https://www.linkedin.com/in/esgnn/>