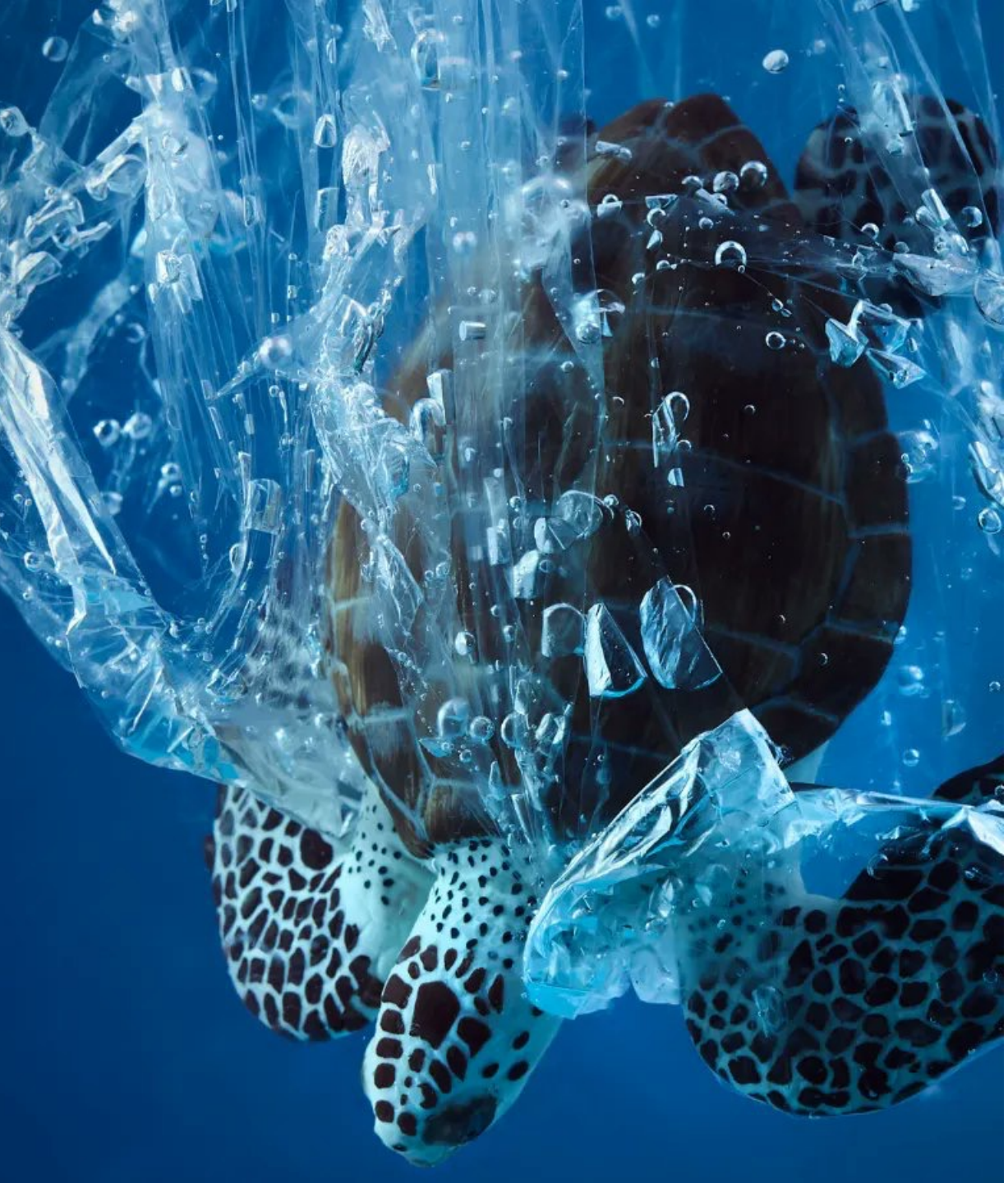
AI use cases 2024

Monica Rotulo,
Machine Learning Advisor
SURF, EUROCC Netherlands



Case study 1: THE OCEAN CLEANUP

“Using AI to monitor plastic density in the ocean”

https://theoceancleanup.com/

“Plastic soup”: Every year, 14 million tons of plastic in the ocean

Seascape: the state of our oceans

Plastic never dies: the museum of vintage waste – in pictures

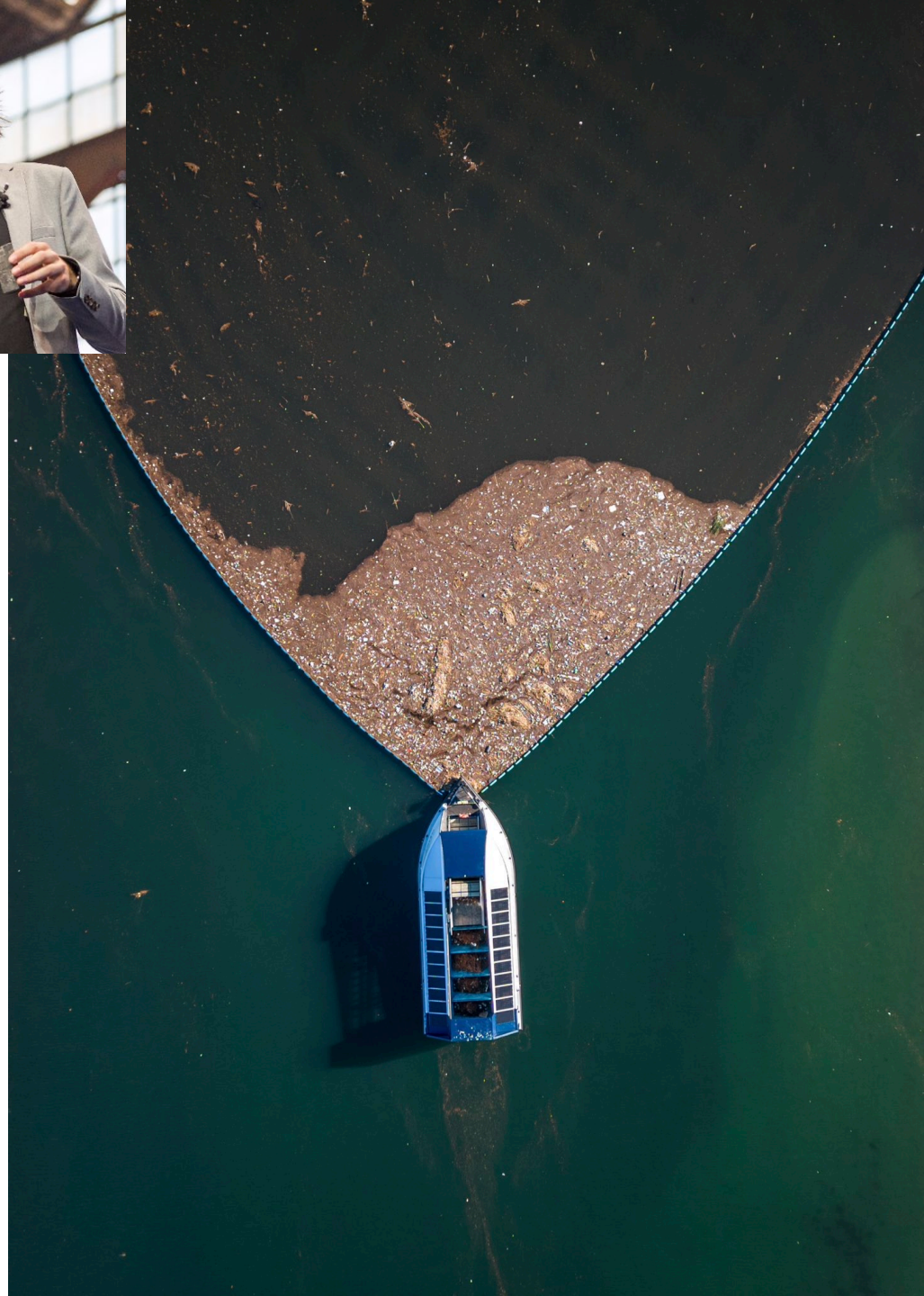
▲ A retro steering wheel found on the beach is among the items on display. Photograph: Archeoplastica

https://www.theguardian.com/

https://www.archeoplastica.it/

SURF

About The Ocean Cleanup



It all began when 16-year-old Boyan Slat (now CEO) was diving in Greece and saw more plastic bags than fish.

The Ocean Cleanup was founded to remove the plastic soup from the oceans.

How AI can come into play?

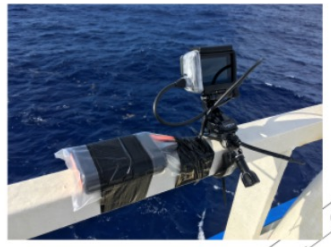
Boyan Slat, on Twitter

Idea: Using AI to monitor plastic density in the ocean

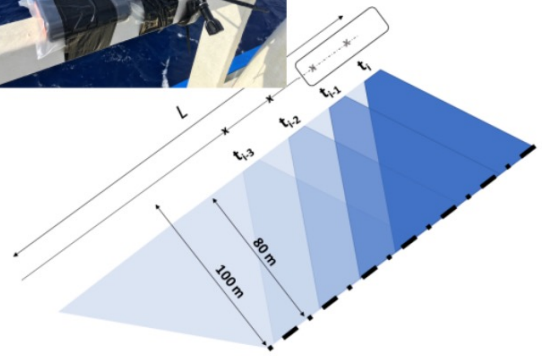
Goal:	Have insights to **determine optimal locations** for operations
Potential impact:	Knowing where the highest concentrations of plastic are located allows to **optimize resources** and maximize impact.
Conventional Method:	Airplanes and trawls: labor intensive, **costly, and complex**
Idea:	AI-based object detection software

Data collection & preparation

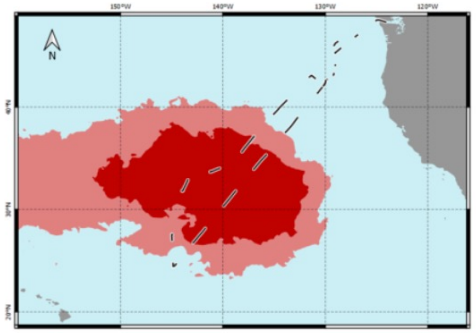
Data Collection:



GoPro
Photos with
GPS tag



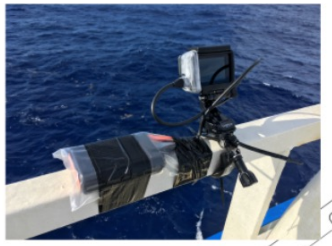
Over 100,000 photos collected



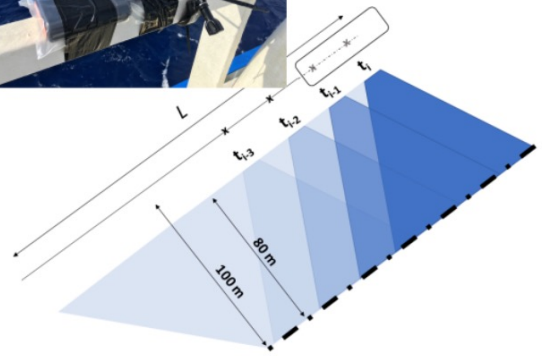
SURF

Data collection & preparation

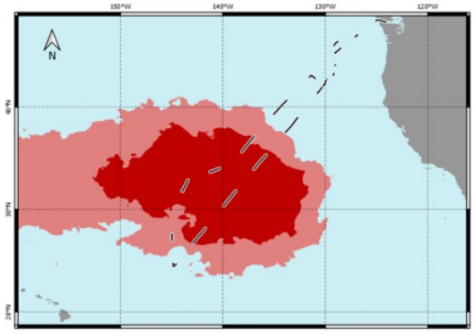
Data Collection:



GoPro
Photos with
GPS tag

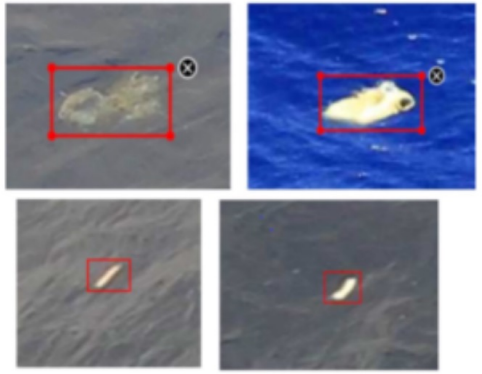


Over 100,000 photos collected

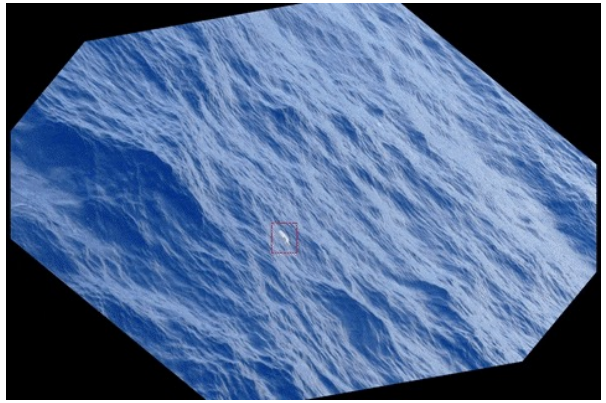


Data Preparation:

Labelling



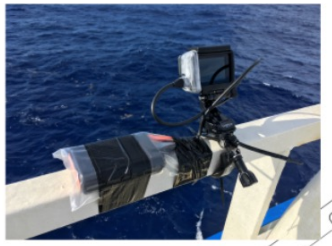
Augmentation



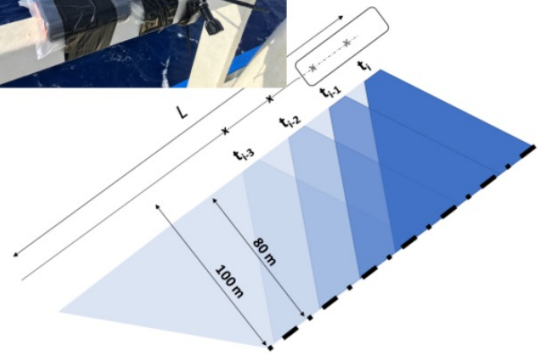
SURF

Data collection & preparation

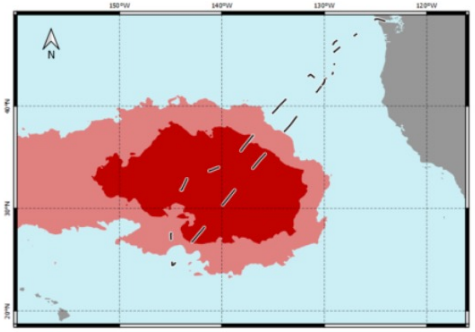
Data Collection:



GoPro
Photos with
GPS tag

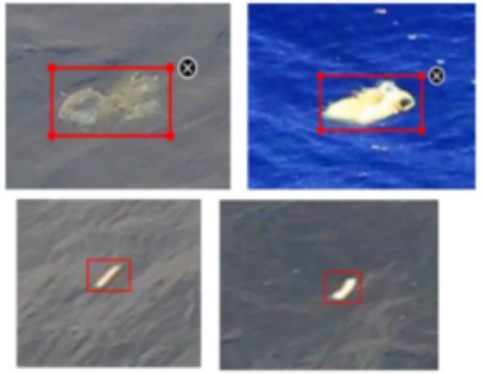


Over 100,000 photos collected

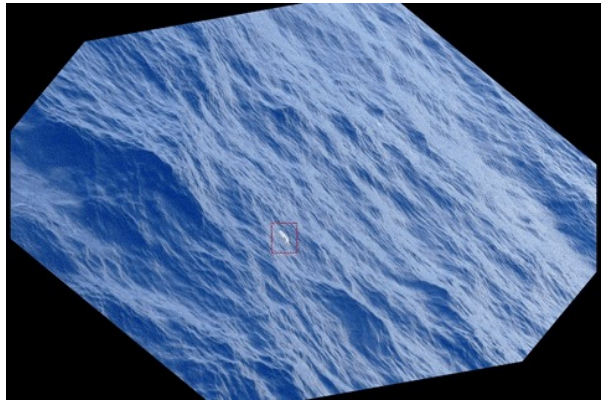


Data Preparation:

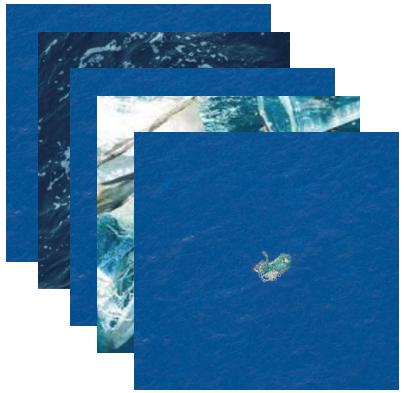
Labelling



Augmentation



Create Datasets:

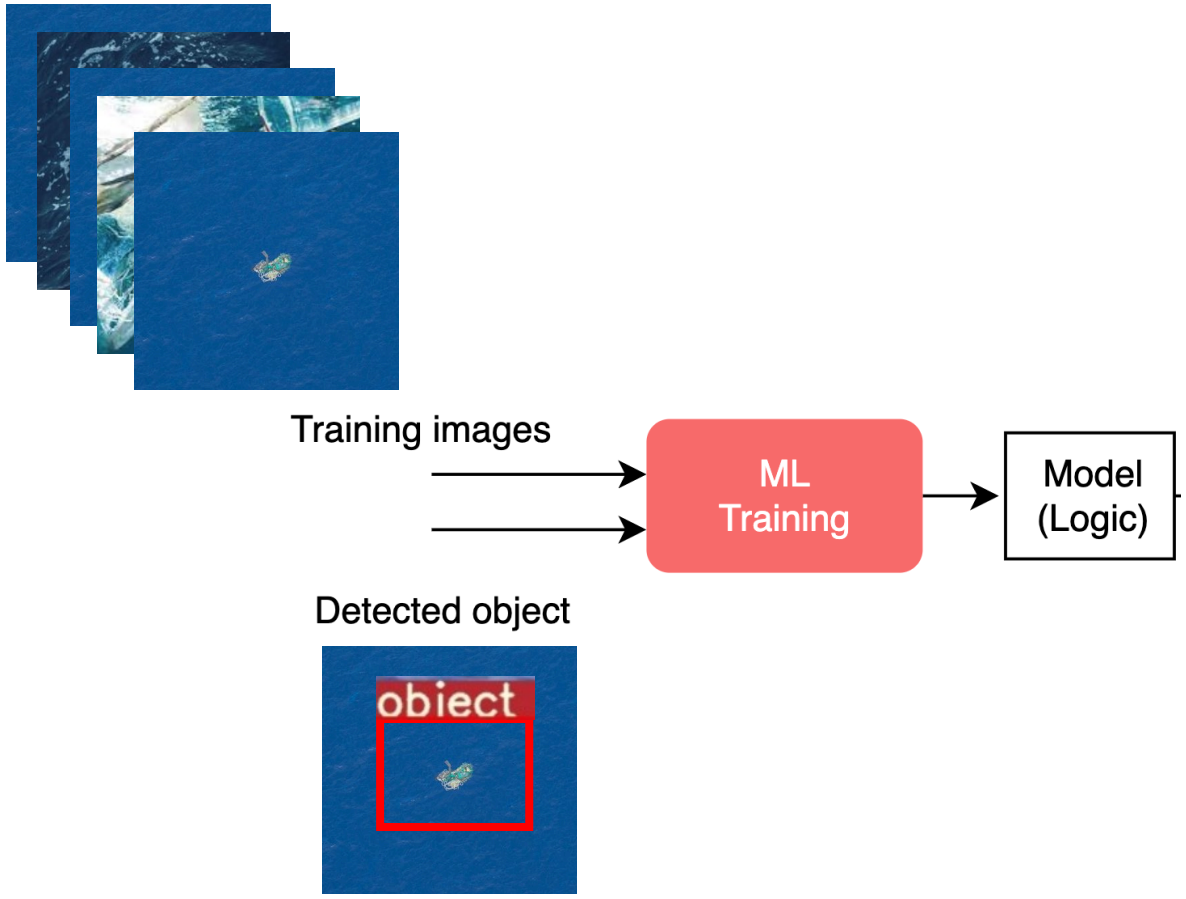


Training set:
18,589 images

Test set:
739 images

Total of 600 GB of images

Training the AI algorithm



Evaluating the AI algorithm

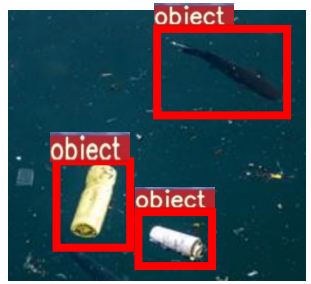
New deck of images
Never seen before

New image

Model (Logic)

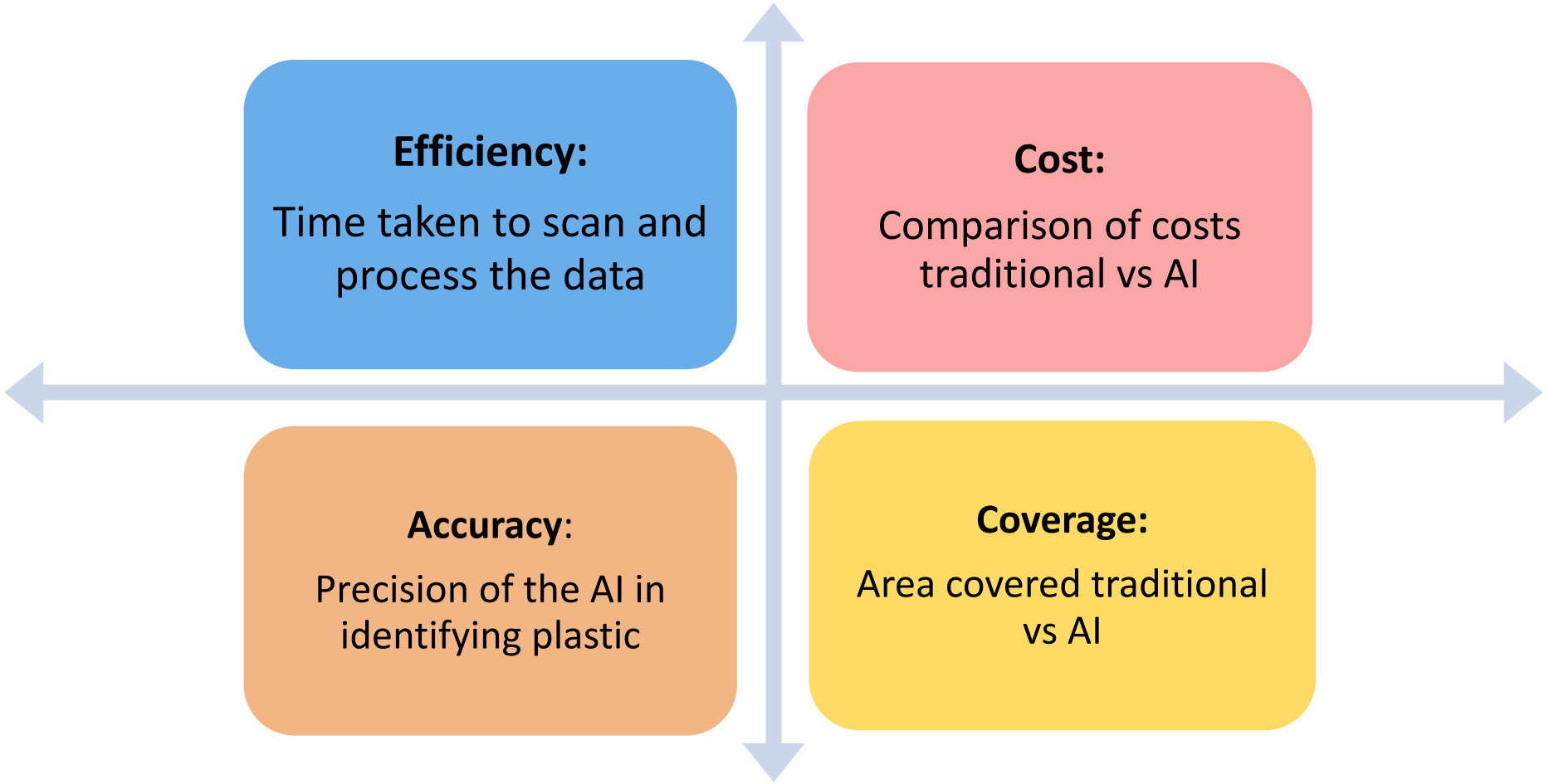
Model

Object detected!



Accuracy:
Precision of the AI in identifying plastic

Metrics to evaluate the AI vs Traditional approach?



From Prototype to Production (& Monitoring): What are the challenges?

Training data “created” in a controlled environment	Expect real-world data to be more messy (think of different light conditions)
The model is trained and tested in HPC clusters	Real-time object detection, integration with small devices and/or cloud
The prototype is handled by data scientist and engineers	For a production-ready product, a user-friendly interface is essential
Prototype is meant for a proof-of-concept, with iterative refinements based on feedbacks	Regular checks for anomalies. Hw might become outdated, model becomes obsolete

SURF

Case study 2:

“Pre-trained LLMs for caption generation of histopathology images”

Radboudumc
university medical center

AI advisor:
Bryan Cardenas Guevara
bryan.cardenasguevara@surf.nl

ExaMode

SURF

Digital pathology for cancer diagnosis:

Data collection

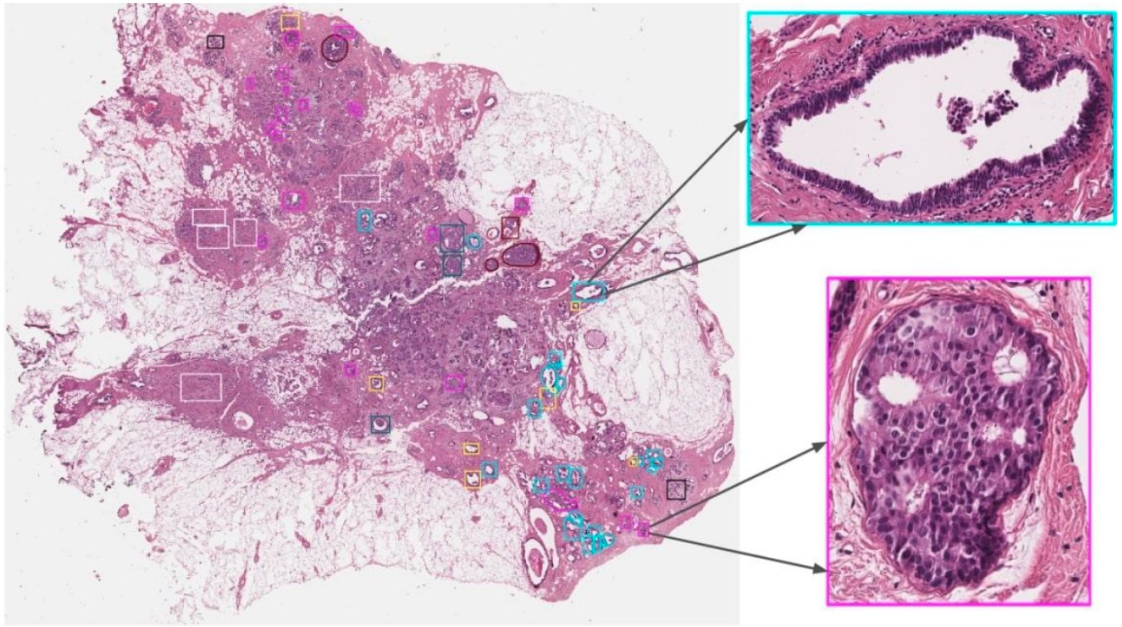
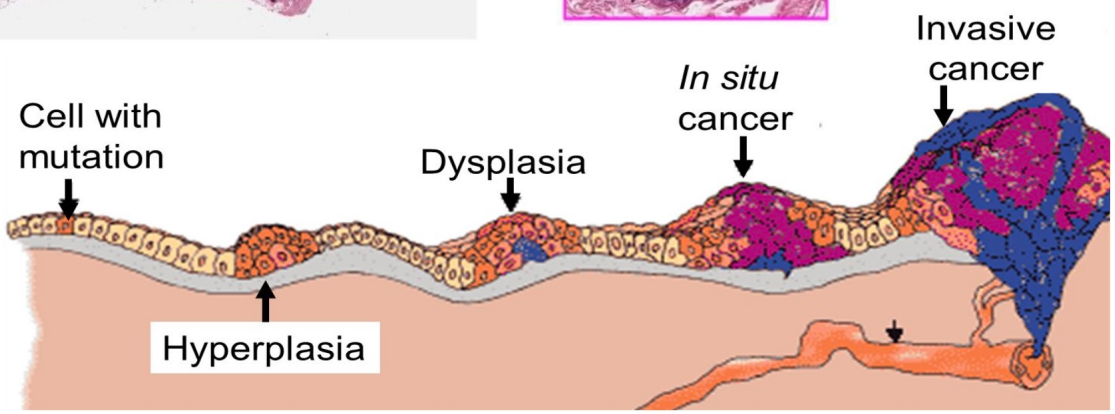


Image analysis And diagnosis



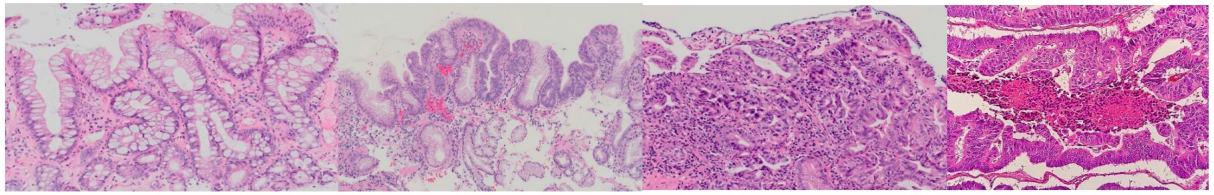


Image analysis: how ML is current used?

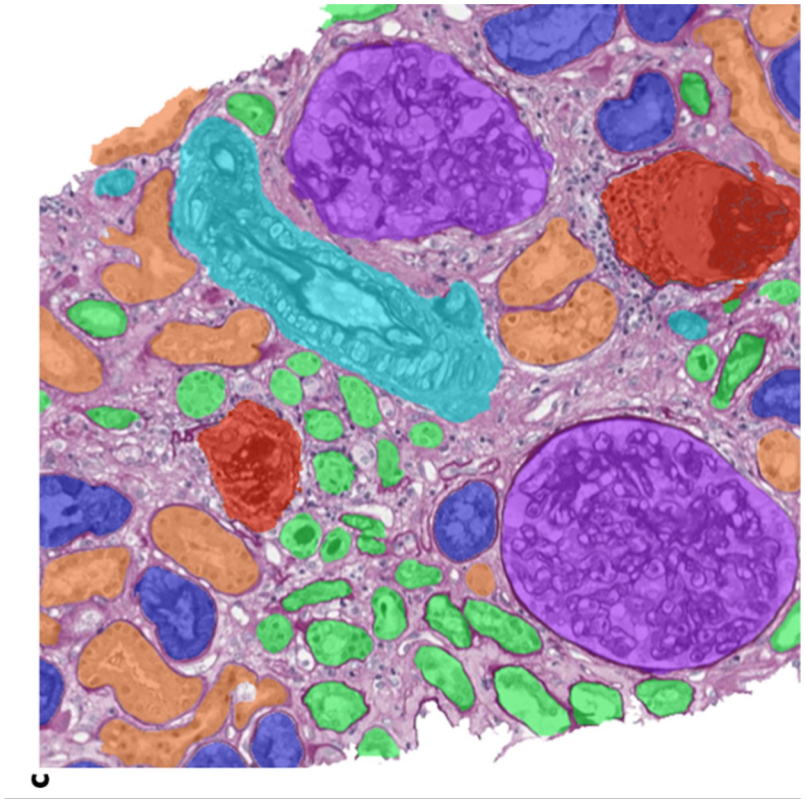
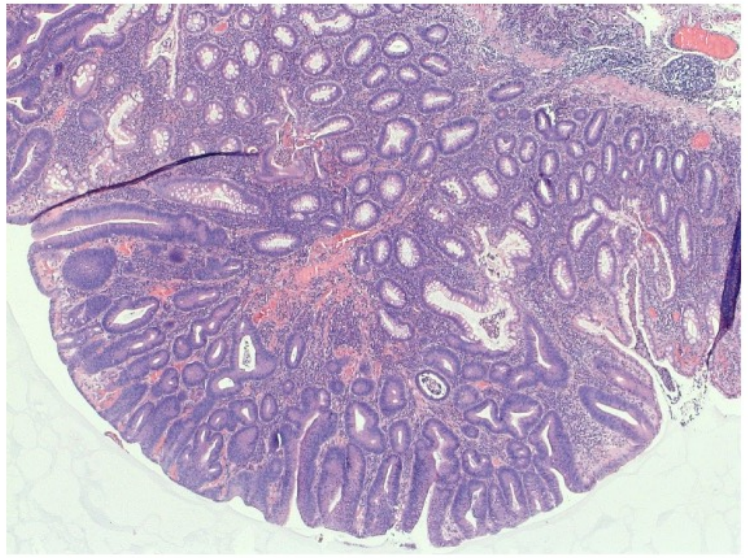
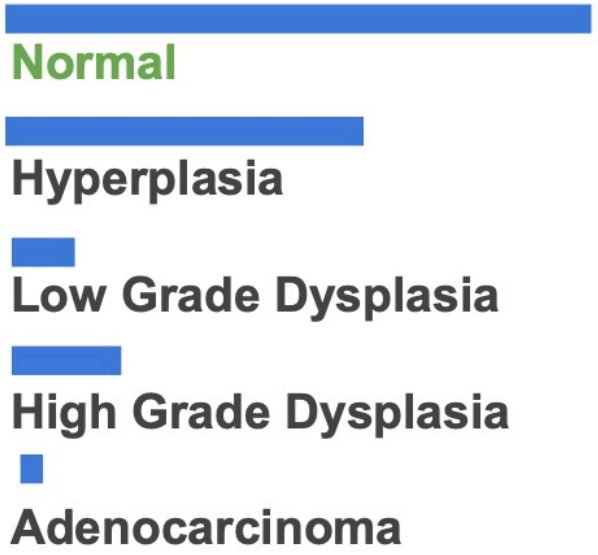


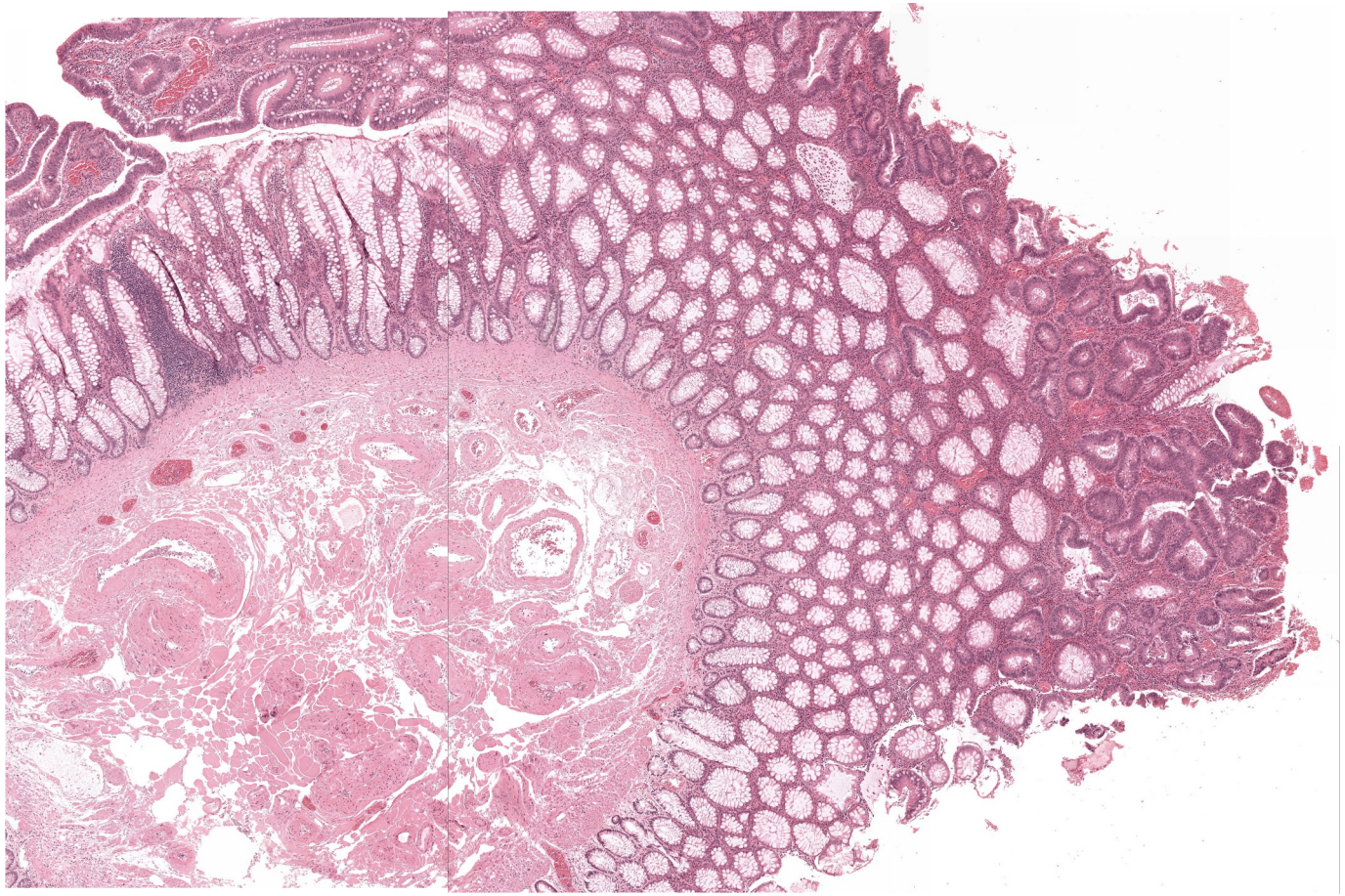
Image segmentation and multi-class classification





Can we improve the text output to further help doctors in the **diagnosis**?

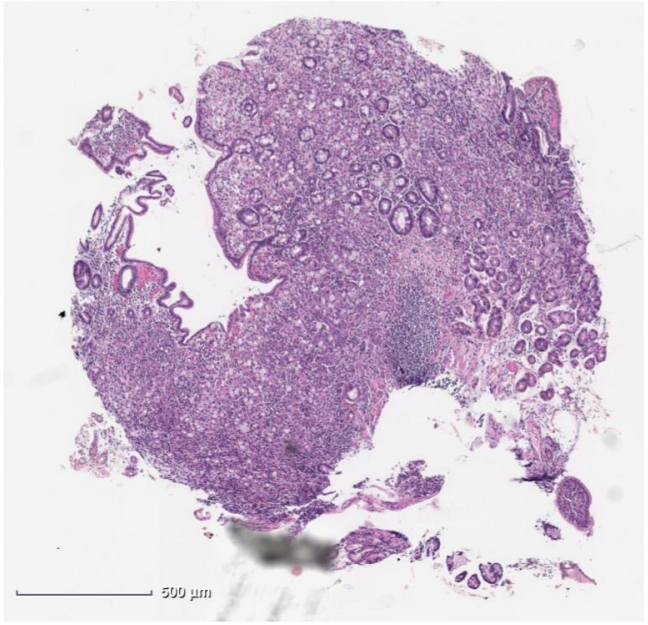
Histopathology data: Images and Text



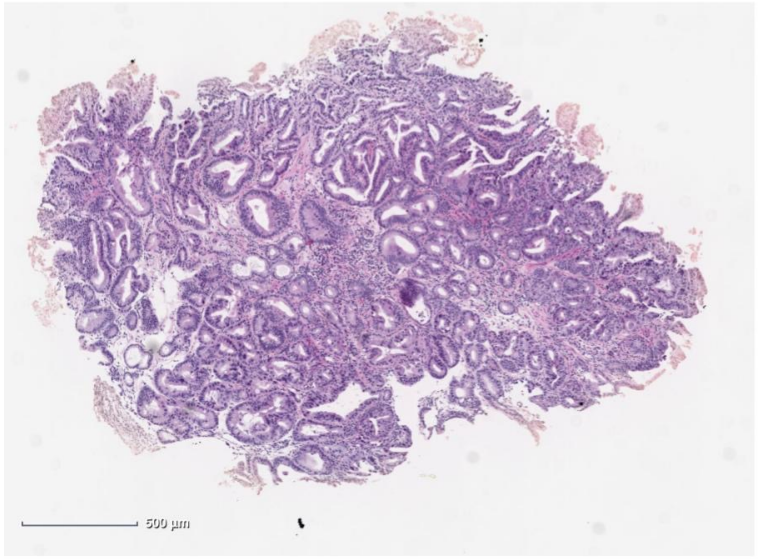
- “Colonbiopt 21-25 cm en 15 cm: in beide subrieken adenocarcinoma metastase, morfologisch en immuunhistochemisch geschikt voor het al bekend primaire ovariumcarcinoom van de patiënt.”
- 2 - adenoma tubulare con displasia bassa, focalmente, alto grado (moderato, focalmente grave)

Data collection: pair of whole slide images + captions

Pair **<WSI-Caption>** collected from two labs from two countries



A tumor tissue consisting of signet ring cell carcinoma cells is observed in the lamina propria. Signet ring cell carcinoma.



Tumor tissue consisting of medium-sized and irregular ducts is observed infiltrating in the superficial epithelium. Tumor cells are highly columnar, with nuclei aligned basolaterally and polarized. Well differentiated tubular adenocarcinoma.

- 5729 WSI-caption pair for training
- A subset of 569 WSI-caption for testing

Data preparation of the captions

Captions are:

- Machine translated from two languages to English
- Preprocessed with GPT-3.5-turbo to remove irrelevant or noisy information
- ChatGPT was prompted with few examples of how to restructure the captions

Original Caption	GPT-3.5 Cleaned	Generated Caption
biopsies distal colon: chronic inflammation, in partially active and slightly histiocytary. no specific characteristics. the microscopic preparations from elsewhere have been requested for revision.	chronic inflammation, no specific characteristics.	no abnormalities, no dysplasia or malignancy. cyclic inflammation.
biopt colon transversum: adenocarcinoma.	adenocarcinoma.	Metastasis of adenocarcinoma best suited to primary process.
1) fragments of tubular adenoma with high degree dysplasia.	tubular adenoma with high degree dysplasia.	adenocarcinoma on villous adenoma. no lymphovascular invasion is identified. enced enced enced ED ED ED ED

(failed caption generation)

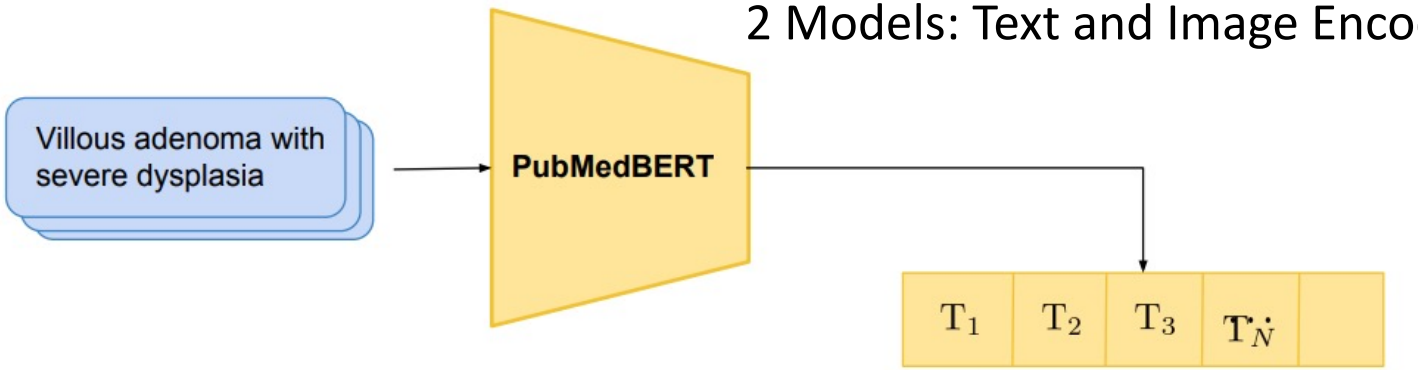
SURF

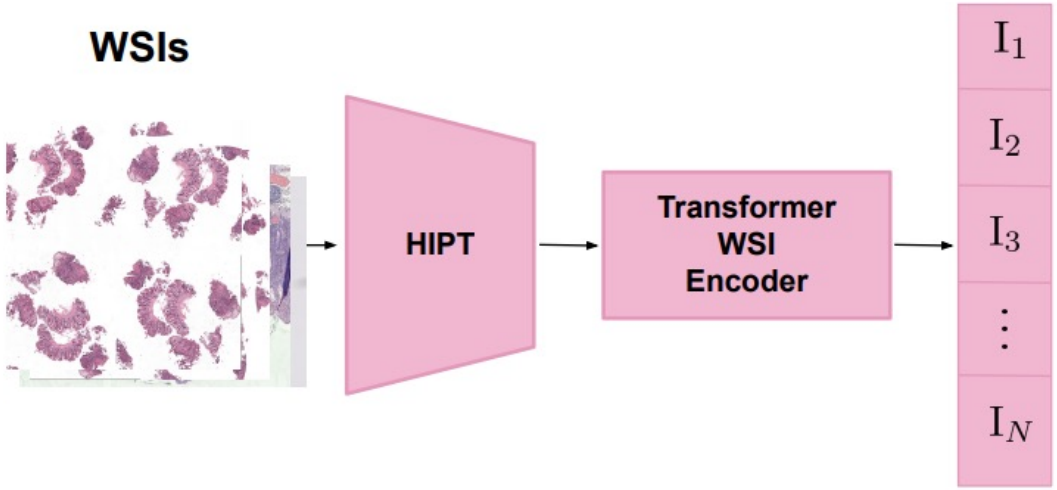
Guevara, Bryan Cardenas, et al. "Caption generation from histopathology whole-slide images using pre-trained transformers." Medical Imaging with Deep Learning, 2023

Model development: Pre-training phase

step 1: Encode information to numerical representation (vectors)

2 Models: Text and Image Encoders

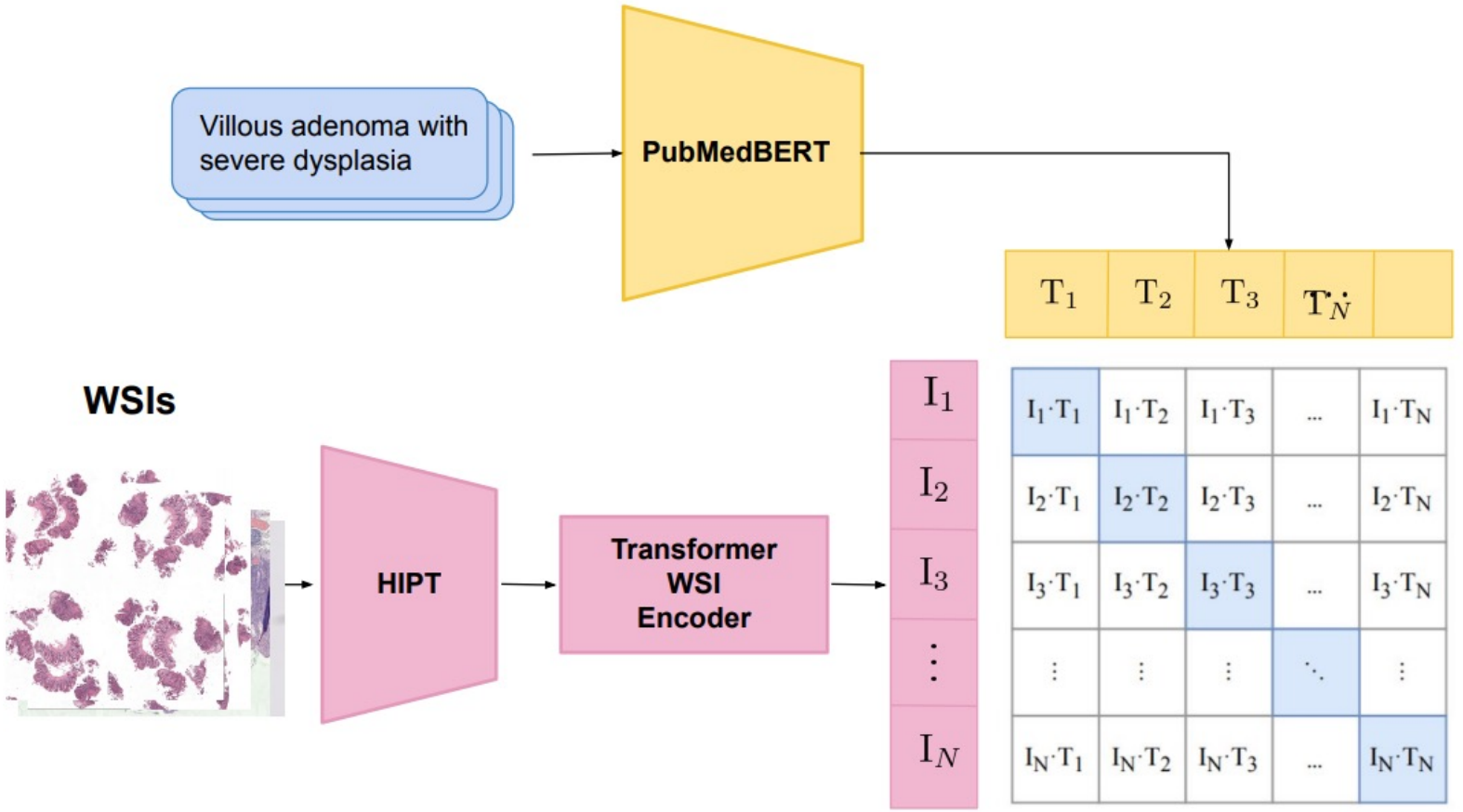




2 representations for text and images

Model development: Pre-training phase

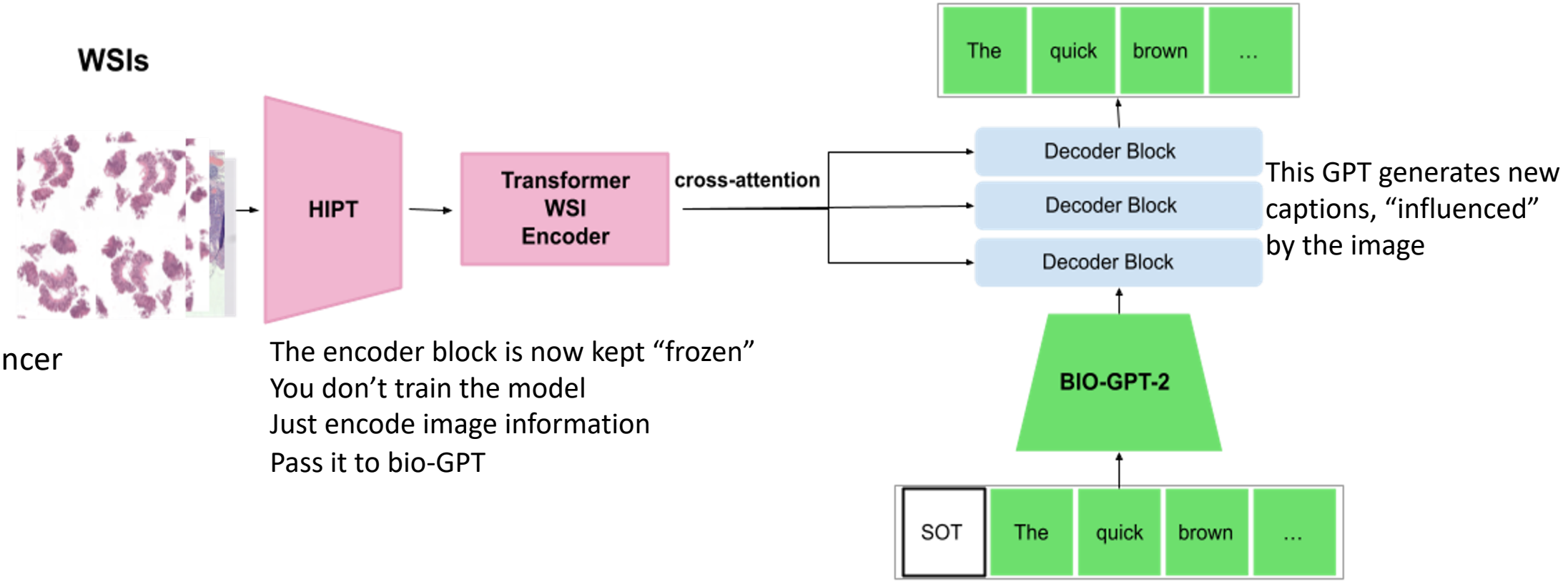
step 2: Train the model to maximize similarity



Vectors generated such that corresponding image-text receive a label 1 where non-related image-text receive a 0

Model development: Generation phase

"this image contains lesion of the form: cancer, high grade"



This GPT generates new captions, “influenced” by the image

Image
With cancer

The encoder block is now kept “frozen”
You don’t train the model
Just encode image information
Pass it to bio-GPT

"this image contains lesion of the form:"

Evaluating the captions generated

Real: "polyep coecum bottom : tubular adenoma with locally high - grade dysplasia.",

Generated: "[CLS] polyep rectosigmoid : tubular adenomas with low grade (moderate) dysplasia and shows shows minor pseudomel some serrated high grade dysplasia.

Real: Polipe rectum : moderately differentiated adenocarcinoma.

Generated: rectum bioptes: localization of moderately differentiated adenocarcinoma.

SURF

Some references and resources:

- SURF https://www.surf.nl/
- EuroCC Netherlands https://eurocc-netherlands.nl/nl/home-nederlands/

The Ocean Cleanup:

- https://www.surf.nl/en/case-computer-models-in-the-fight-against-plastic-soup
- https://theoceancleanup.com/updates/using-artificial-intelligence-to-monitor-plastic-density-in-the-ocean/

Histopathology:

- https://www.computationalpathologygroup.eu/projects/examode/
- https://github.com/sara-nl/WSI_Captioning

Thank you!

SURF, Amsterdam Science Park

monica.rotulo@surf.nl