Predicting the Needle in a Petabyte Scale Haystack



Satellites now image the Earth daily

Predicting the Needle in a Petabyte Scale Haystack

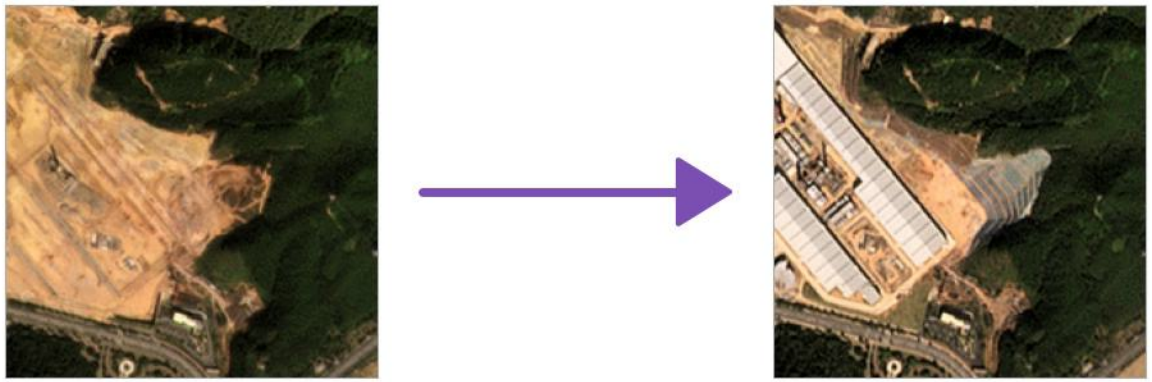


week 41

However, much of the Earth is not changing in interesting ways

Predicting the Needle in a Petabyte Scale Haystack

a needle: a data center going up



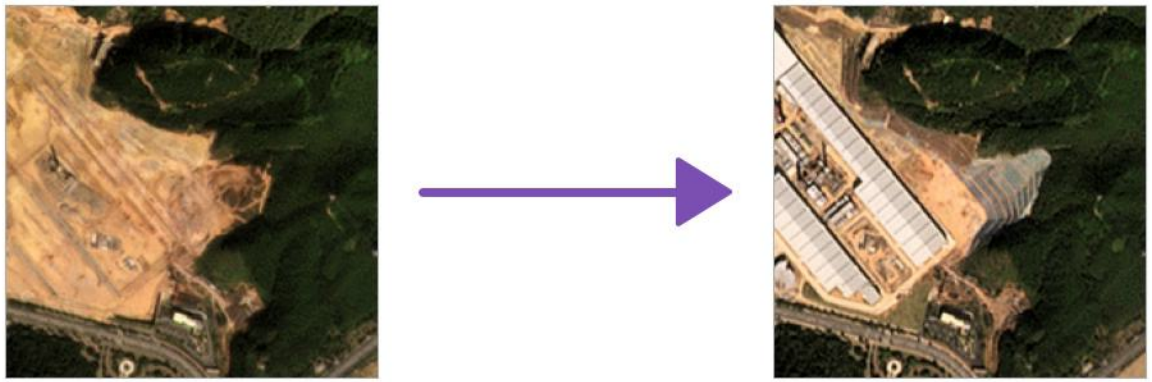
empty lot

data center

Making the events we care about relatively rare, like needles in a petabyte-scale haystack

Predicting the Needle in a Petabyte Scale Haystack

a needle: a data center going up



empty lot

data center

We ask: can we retrieve these needles, and maybe even anticipate events before they finish?

1 The problem

- To model event anticipation, we need lots of examples of events

To model event anticipation, we need lots of examples of events

1 The problem

- To model event anticipation, we need lots of examples of events
- Events are needles in a haystack: we need to retrieve them

But since events are needles in a haystack, we need to retrieve them somehow

1 The problem

- To model event anticipation, we need lots of examples of events
- Events are needles in a haystack: we need to retrieve them
- ~~No sufficient open-vocabulary models for event retrieval~~

We do not currently have sufficient open-vocabulary models for event retrieval

1 The problem

- To model event anticipation, we need lots of examples of events
- Events are needles in a haystack: we need to retrieve them
- ~~No sufficient open-vocabulary models for event retrieval~~
- **Good models exist for open-vocabulary static scene retrieval**

But we do have good models for open-vocabulary static scene retrieval

1 The problem

open-vocabulary static scene retrieval

"empty undeveloped rural land with natural vegetation"



This means we can use text descriptions to retrieve locations with relevant satellite imagery

2 Our insight

Our insight: while static models cannot reason about time, they can understand the before and after states of an event

2 Our insight

~~"data center construction"~~

static models cannot reason about time

For example, they cannot retrieve the event data center construction

2 Our insight

before

"empty lot"



after

"data center"



But they can retrieve images of empty lots, and data centers

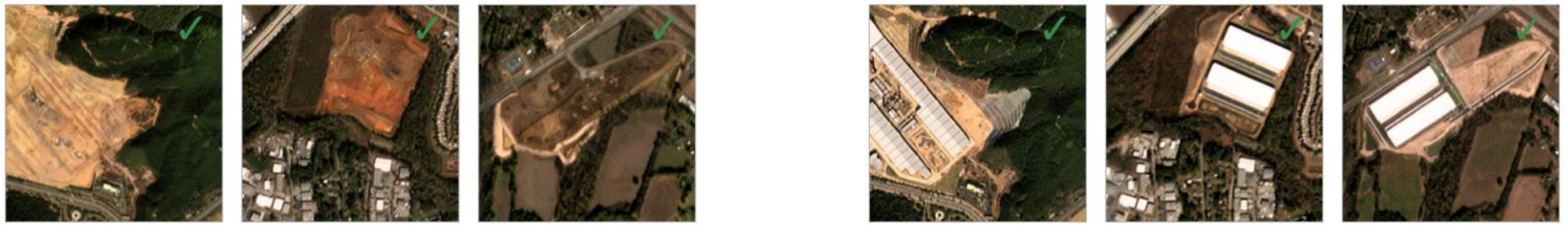
2 Our insight

before

"empty lot"

after

"data center"



We can leverage this knowledge to retrieve events

3 Strategy 1: intersection search

query @ time 1

"empty pier"



Before

The most naive way to leverage this: search for the before state at timepoint 1, the after state at timepoint 2, and intersect

3 Strategy 1: intersection search

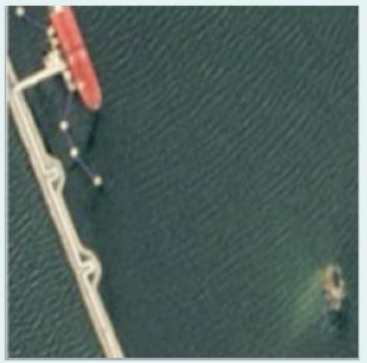
query @ time 1

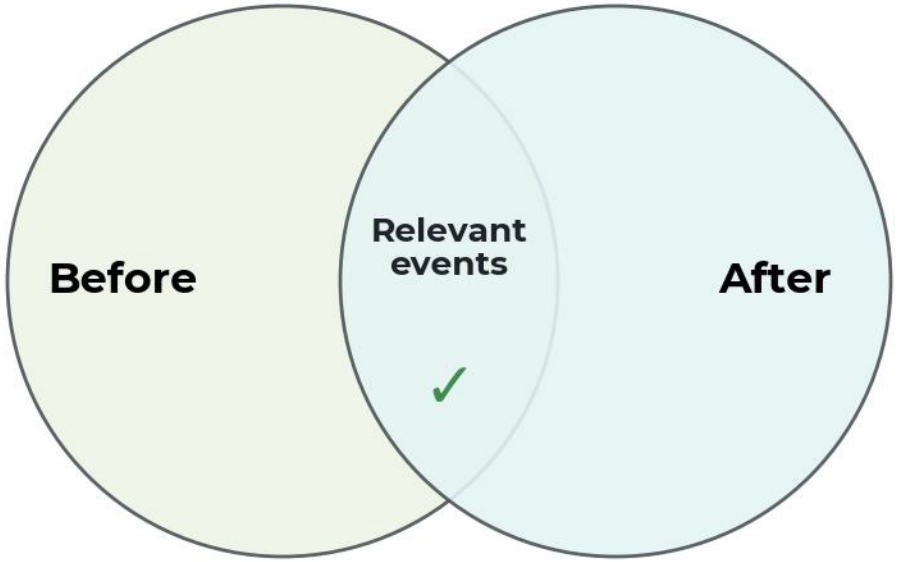
"empty pier"



query @ time 2

"pier with a ship"





Event: ship docking at pier

Say we are searching for ships docking at a pier: first we query for empty piers at timepoint 1

3 Strategy 1: intersection search

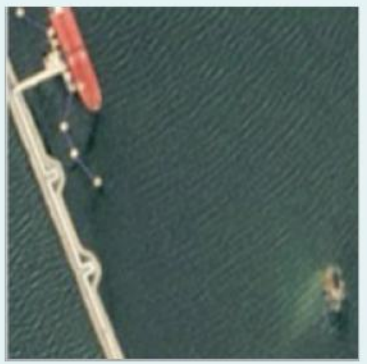
query @ time 1

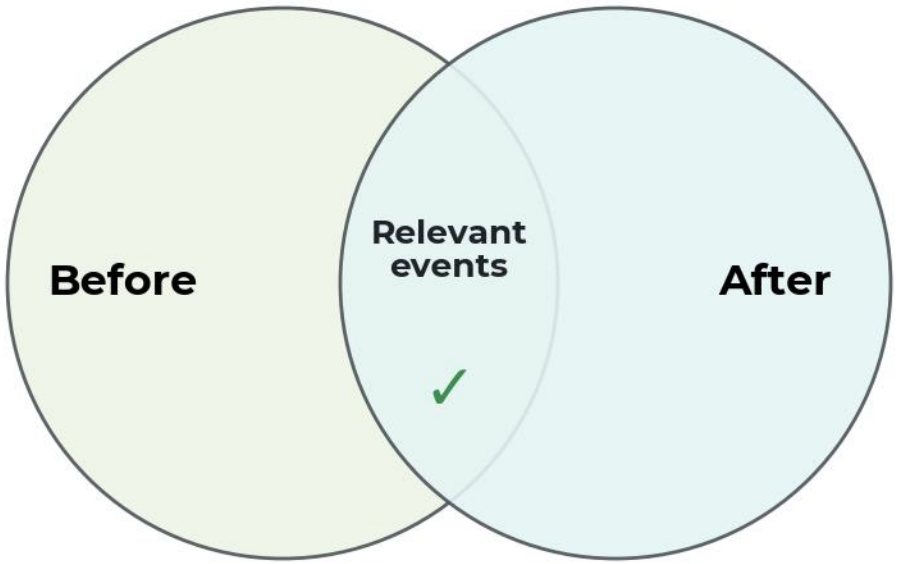
"empty pier"



query @ time 2

"pier with a ship"





Event: ship docking at pier

Then for piers with a ship at timepoint 2. Where the results overlap, we have found the event of a ship docking at a port

3 Strategy 1: intersection search

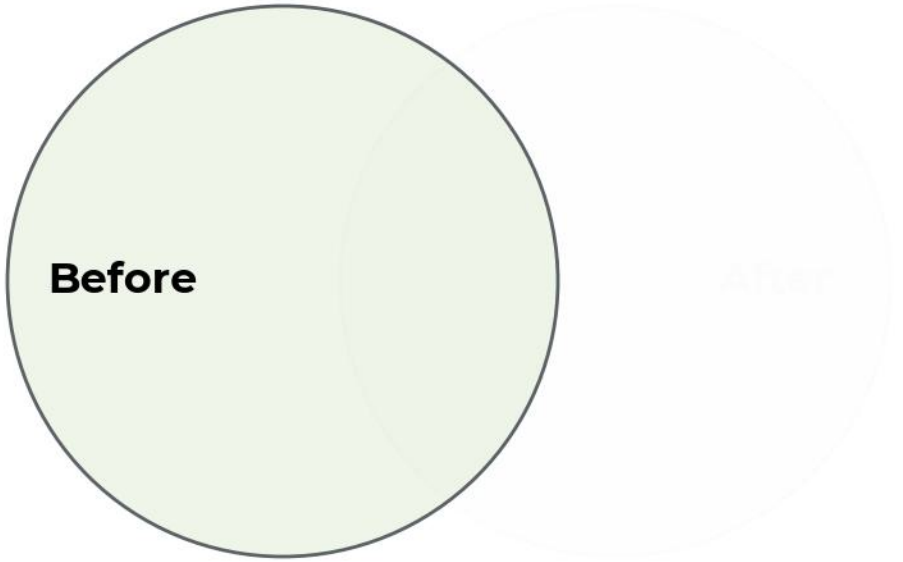
query @ time 1

"park with a baseball field"



query @ time 2

"park with two baseball fields"



But say we want to find parks where a second baseball field has been added on

3 Strategy 1: intersection search

query @ time 1

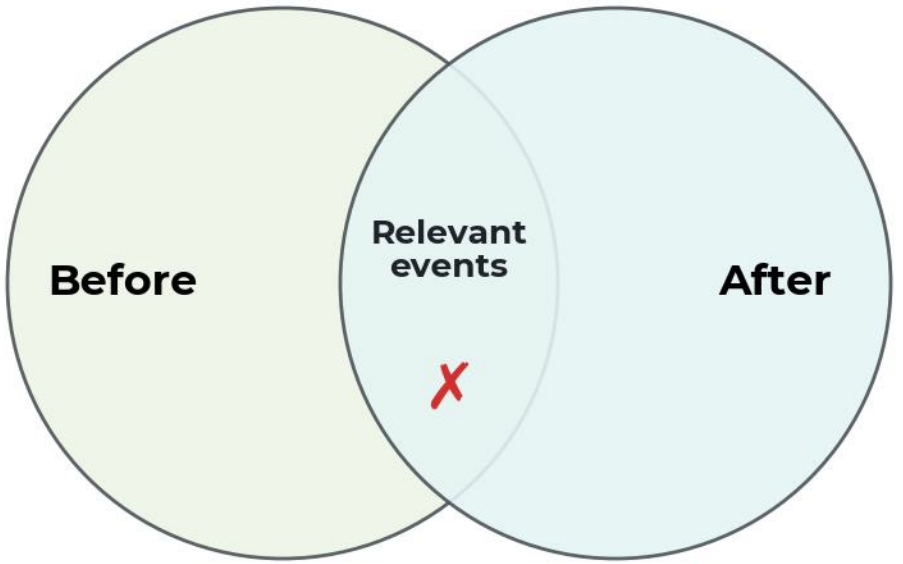
"park with a baseball field"



query @ time 2

"park with two baseball fields"





Event: new baseball field?

Parks that already have multiple baseball fields match both the before and after queries, without any event having happened

3 Strategy 1: intersection search

query @ time 1

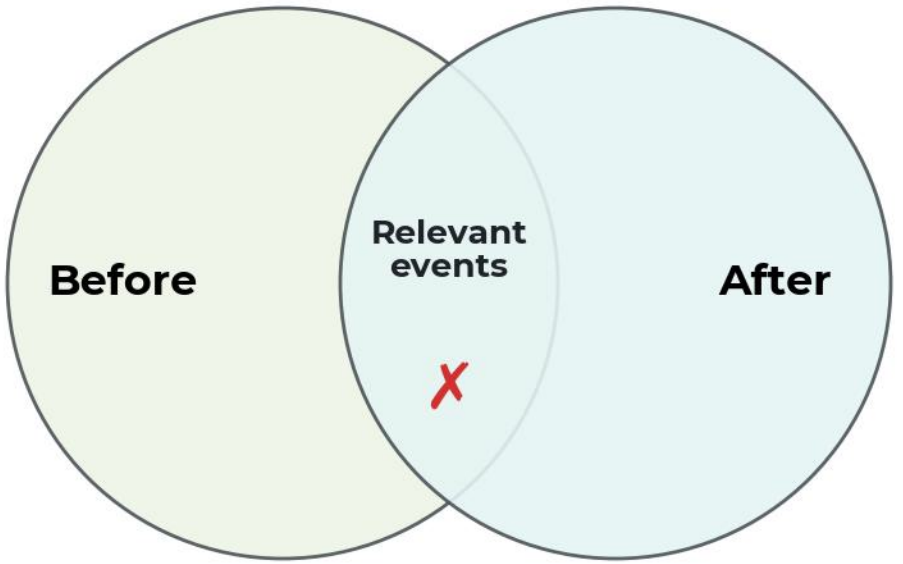
"park with a baseball field"



query @ time 2

"park with two baseball fields"





Event: new baseball field?

Here we see that intersection search fails to ensure that a *change* event has occurred

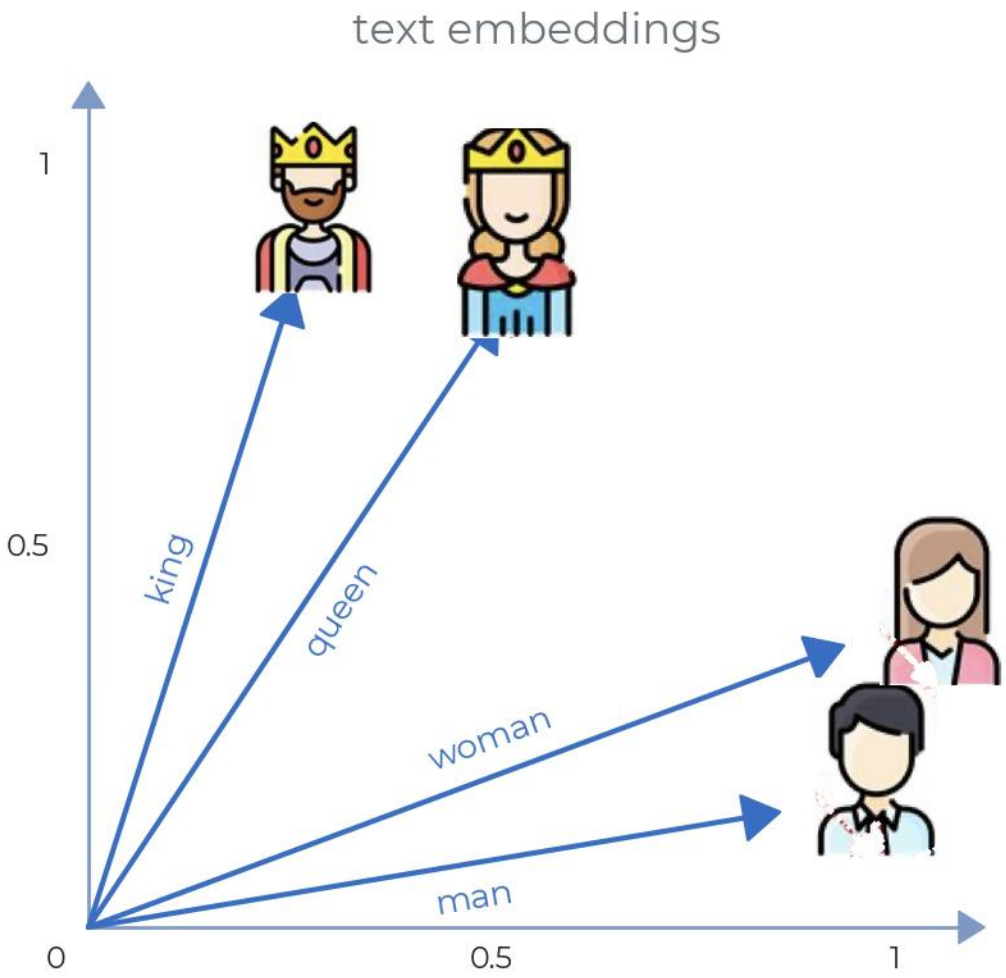
4 Strategy 2: change search

So how can we ensure

change?

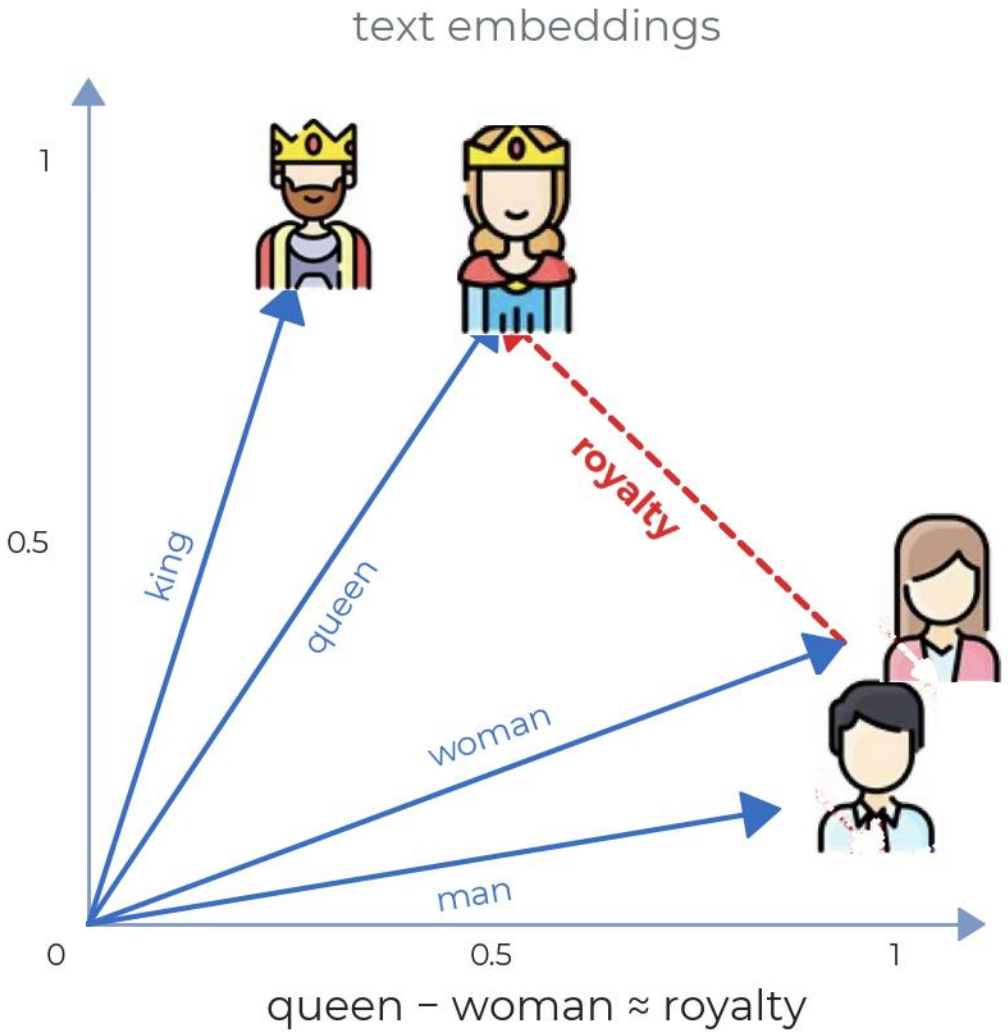
So how can we ensure change?

4 Strategy 2: change search



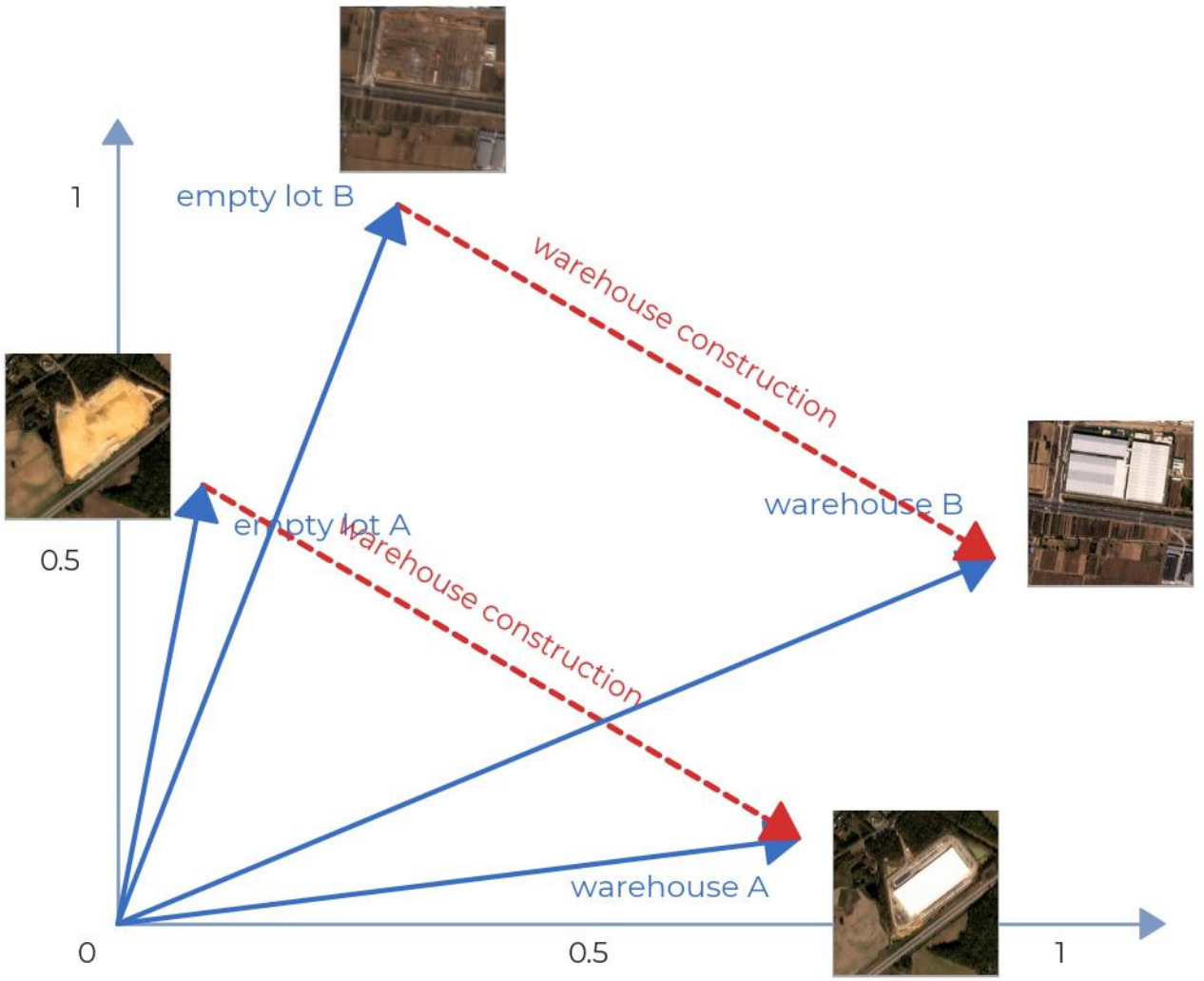
In language-aligned embedding spaces, the difference between two vectors can represent a semantic transformation

4 Strategy 2: change search



A famous example in text embeddings is the notion of a royalty direction, from woman to queen or man to king

4 Strategy 2: change search



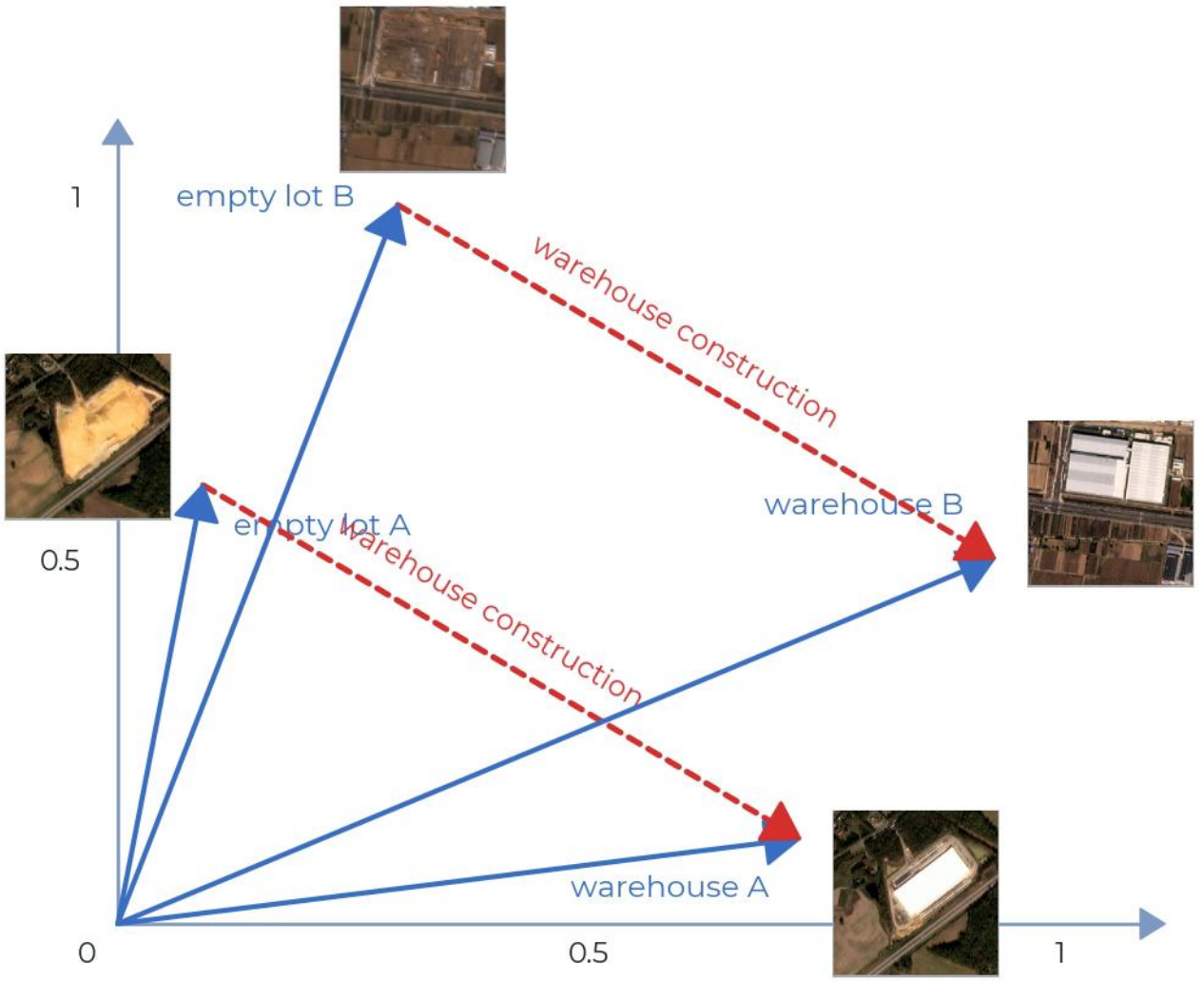
same location, two times

warehouse − empty lot

≈ warehouse construction

With satellite imagery, we can embed images of the same location over time

4 Strategy 2: change search



same location, two times

warehouse − empty lot

≈ warehouse construction

$z_t = f_\theta(I_t)$ (embedding of the pinned location at time t)

$\Delta z = z_T - z_0$ **(change vector)**

The transformation between two time points represents a physical change event occurring at that location

4 Strategy 2: change search



We can precompute this change, or delta vector, for all locations in our area of interest

4 Strategy 2: change search

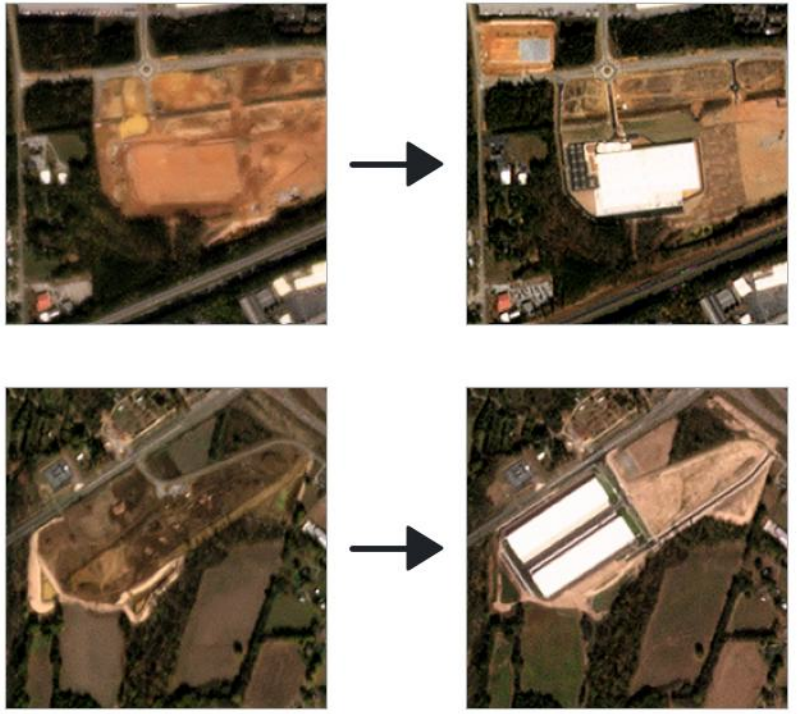
After "large warehouse building" − **Before** "empty undeveloped rural land with natural vegetation"

OR

After  − **Before** 

Retrieval on Δz

Retrieved change events



And perform retrieval on this delta, by creating delta queries as well

5 Two complementary strategies + automated relevance feedback



→ baseball field construction

Event: new baseball field in park? ✓

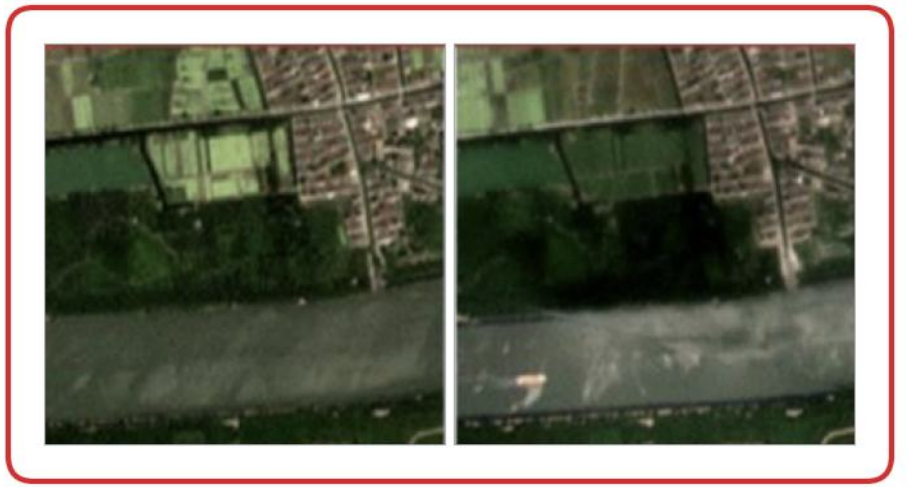
Change search ensures we capture a change: a new baseball field is found

5 Two complementary strategies + automated relevance feedback



→ baseball field construction

Event: new baseball field in park? ✓



→ ship appearing

Event: ship docking at pier? ✗

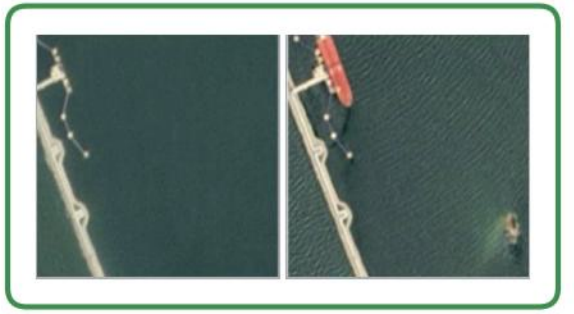
But searching only for the change loses context: a ship appearing isn't a ship docking at a pier

5 Two complementary strategies + automated relevance feedback

Intersection search



Event: new baseball field? ✗

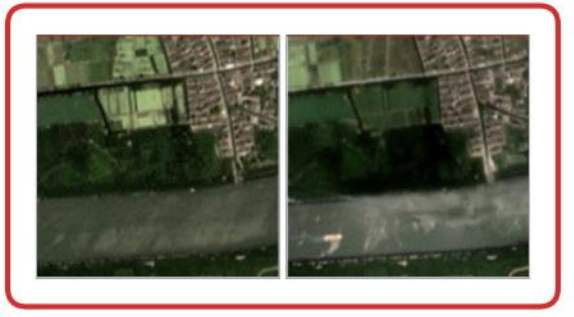


Event: ship docking at pier ✓

Change search



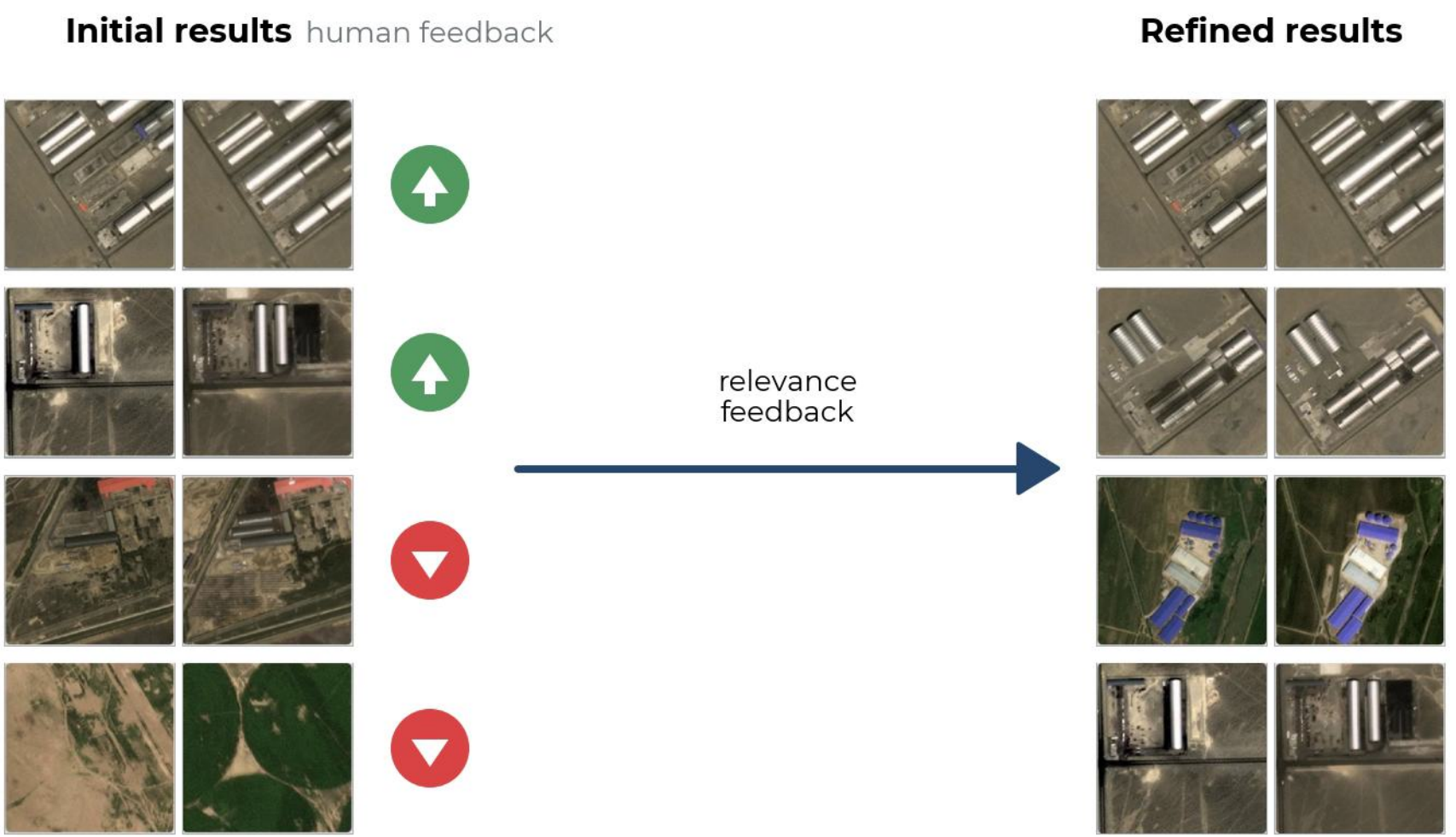
Event: new baseball field ✓



Event: ship docking at pier? ✗

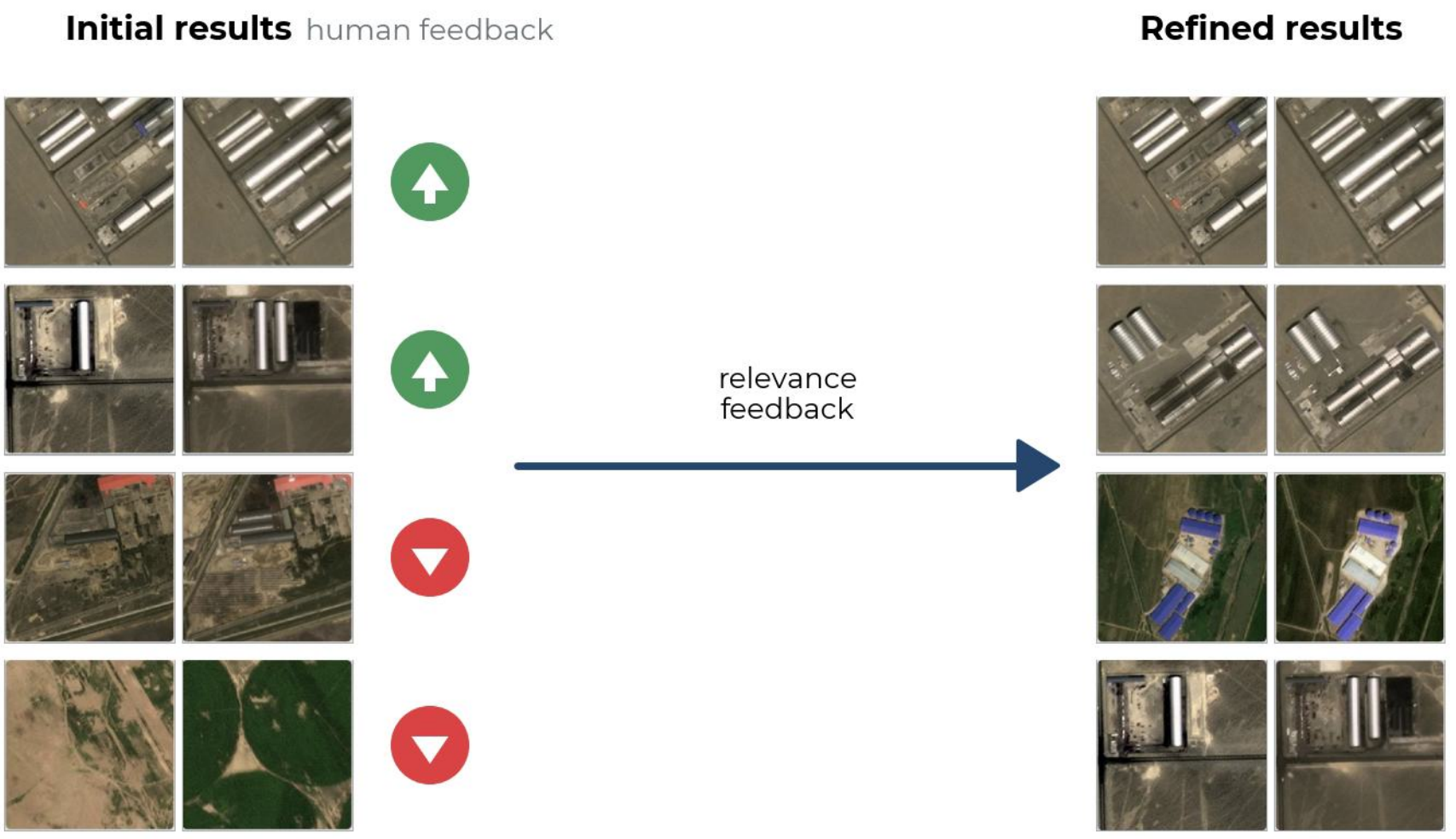
The two are complementary: context from one, specificity from the other

5 Two complementary strategies + automated relevance feedback



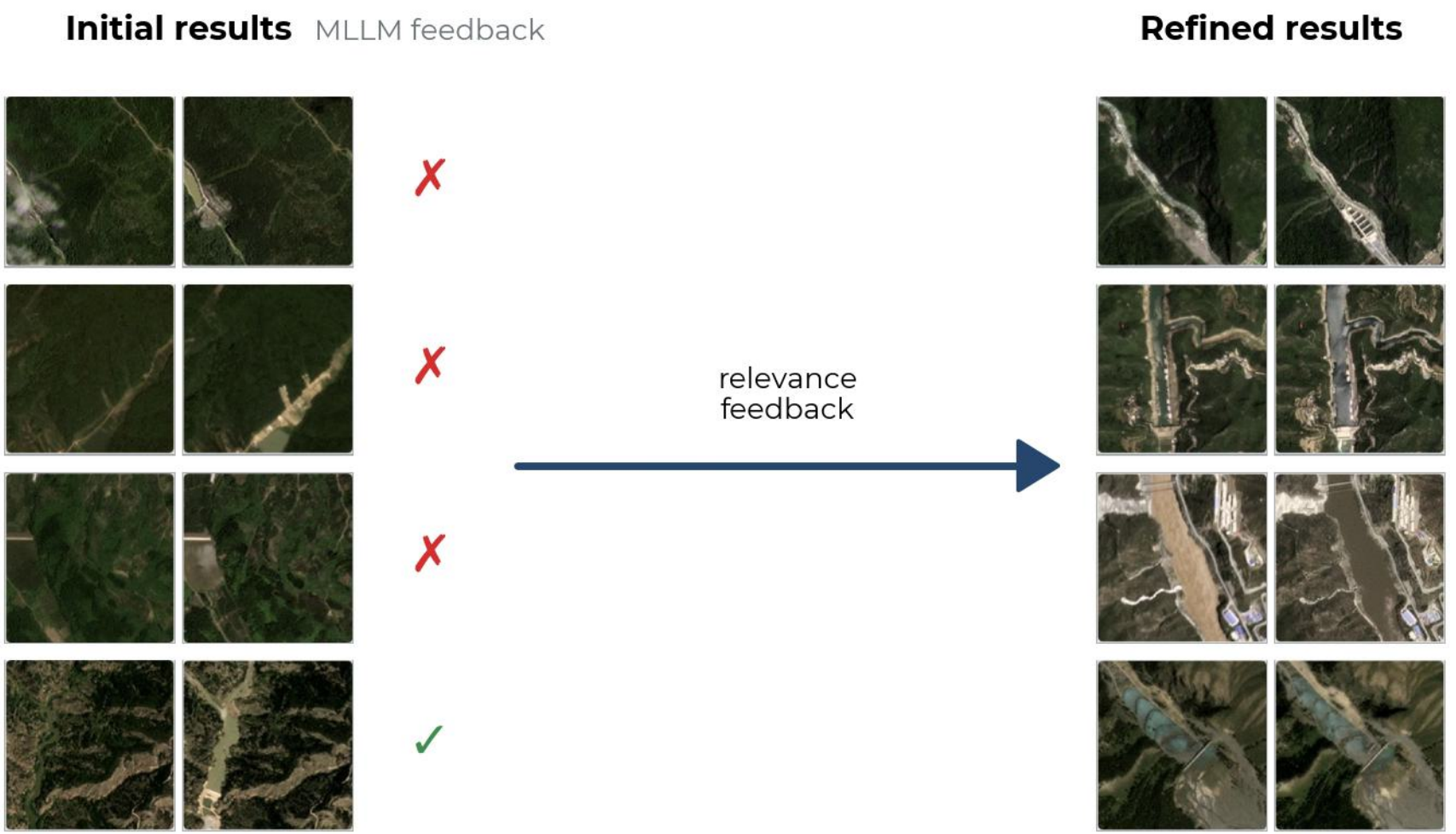
We tie them together with relevance feedback, a technique used to refine search probes

5 Two complementary strategies + automated relevance feedback



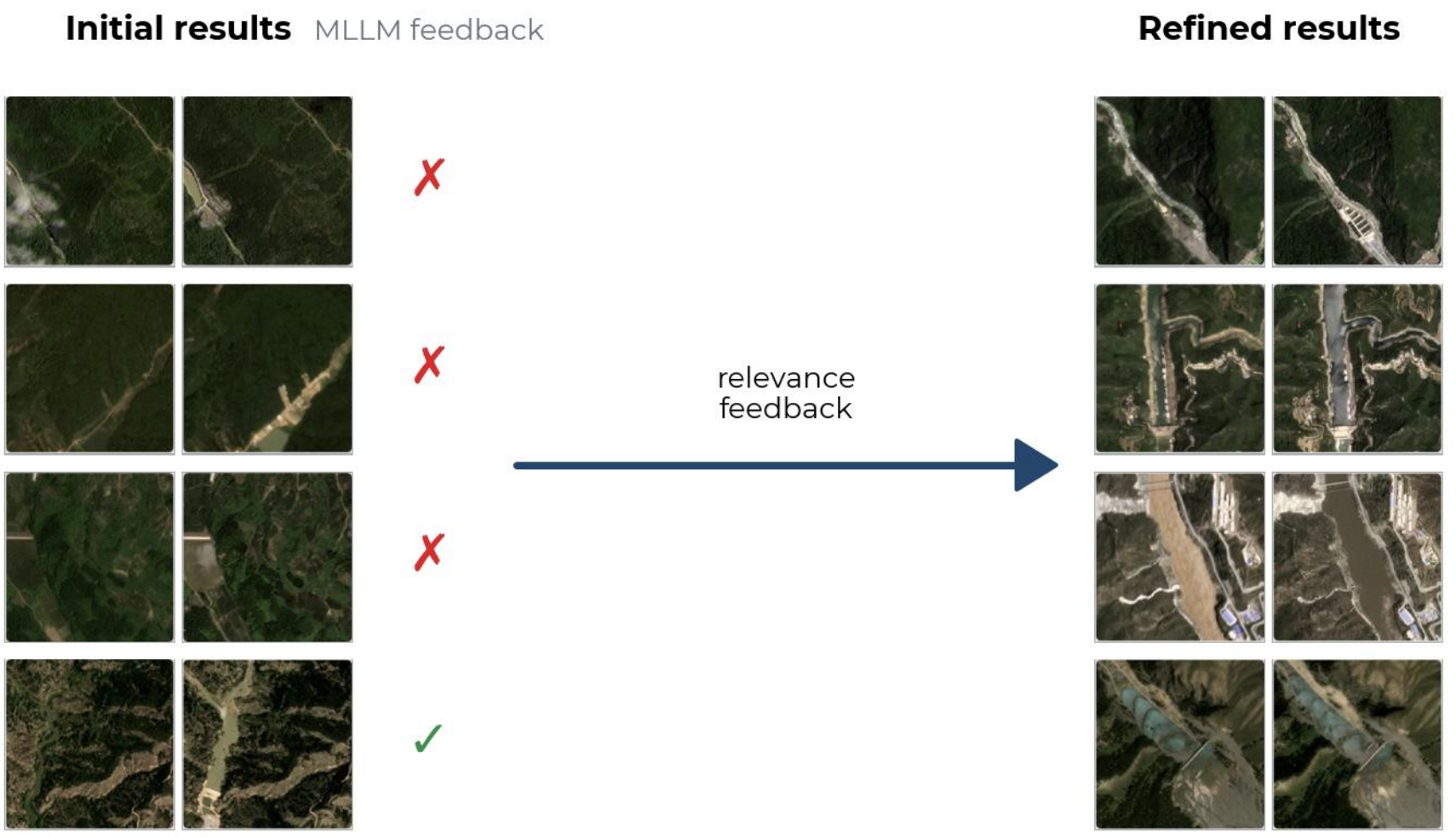
Traditionally, relevance feedback is done by humans

5 Two complementary strategies + automated relevance feedback



We accelerate it using a multimodal LLM to give the feedback: automated relevance feedback (ARF)

5 Two complementary strategies + automated relevance feedback



This rapidly scales feedback and dramatically improves search results

6 Collecting events at Earth scale

Δz for every location on Earth



Nov 2024

embed every location

−



Nov 2025

embed every location

To collect events at scale, we embed Earth's landmass at two time points

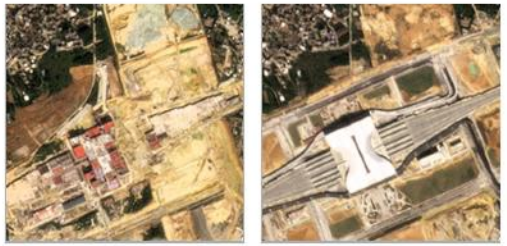
6 Collecting events at Earth scale

1. **Generate event descriptions**
   an MLLM writes before / after text for each event type
2. **Intersection retrieval**
   top-k candidates that match both descriptions
3. **ARF round 1**
   MLLM feedback refines the change probe
4. **Change-vector retrieval**
   retrieve with the refined probe
5. **ARF round 2**
   MLLM keeps the final positives, saved to the dataset

Then, for every event type, we combine both strategies with automated relevance feedback

6 Collecting events at Earth scale

Airport Construction



Bridge Construction



Coastal Industrial Facility Construction



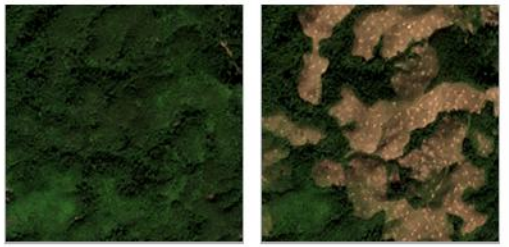
Dam Construction



Data Center Construction



Forest to Pasture Conversion



Greenhouse Complex Buildup



High-Voltage Transmission Line



Highway Construction



Hospital Wing Addition



Housing Development



Industrial Site Decommissioning

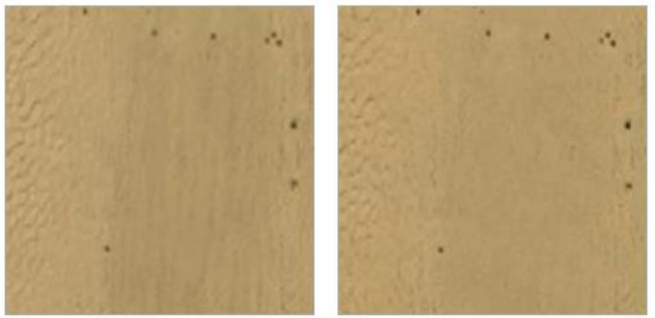


This yields a large collection of events across the world

6 Collecting events at Earth scale

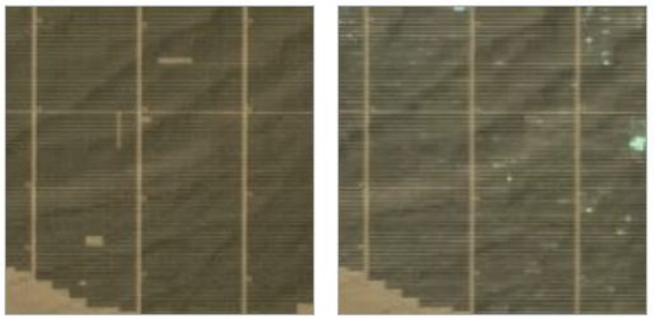
stable scenes: before description = after description

Stable desert



before after

Solar farm



before after

Urban center

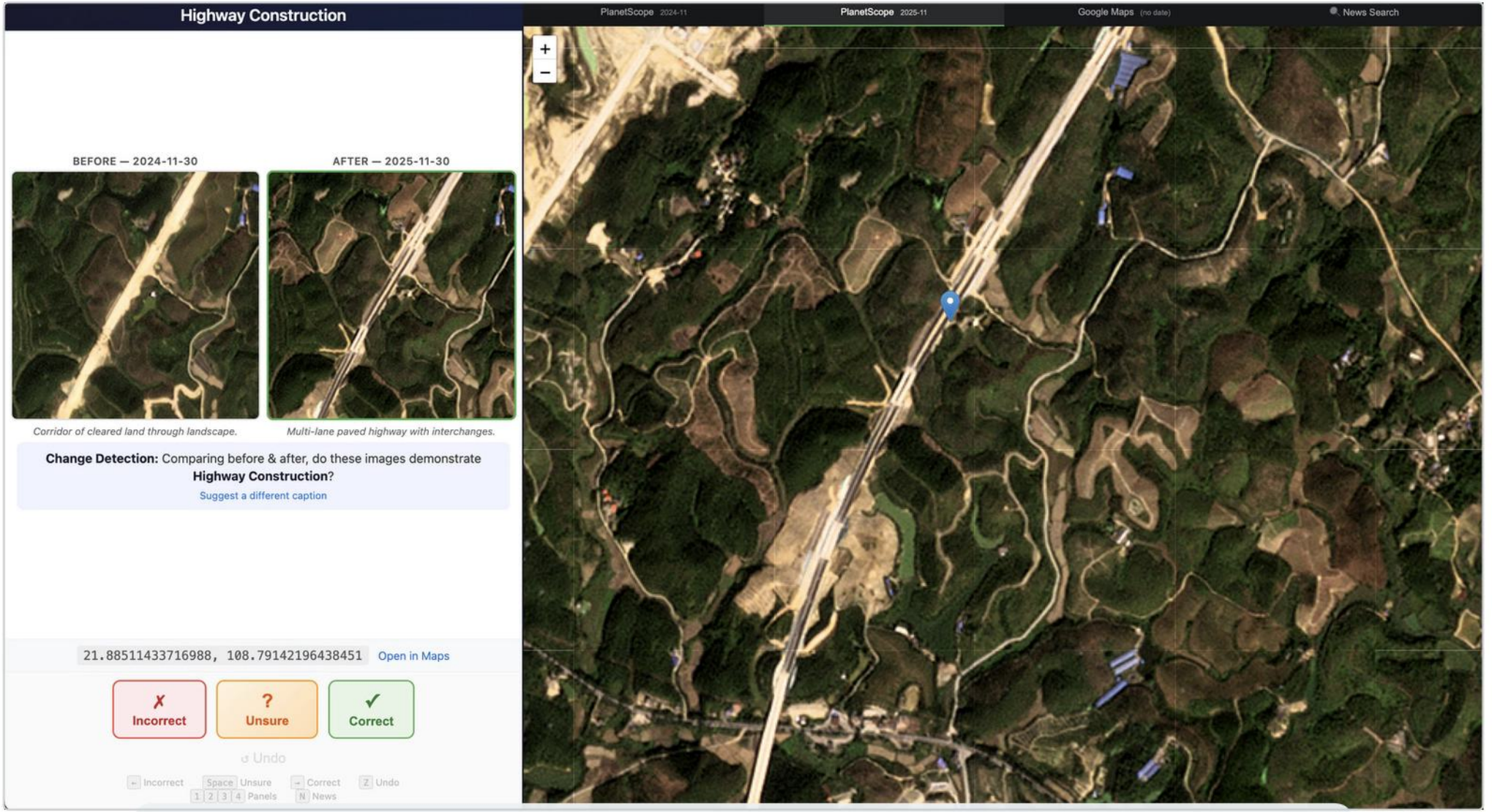


before after

$$||\Delta z|| \approx 0$$

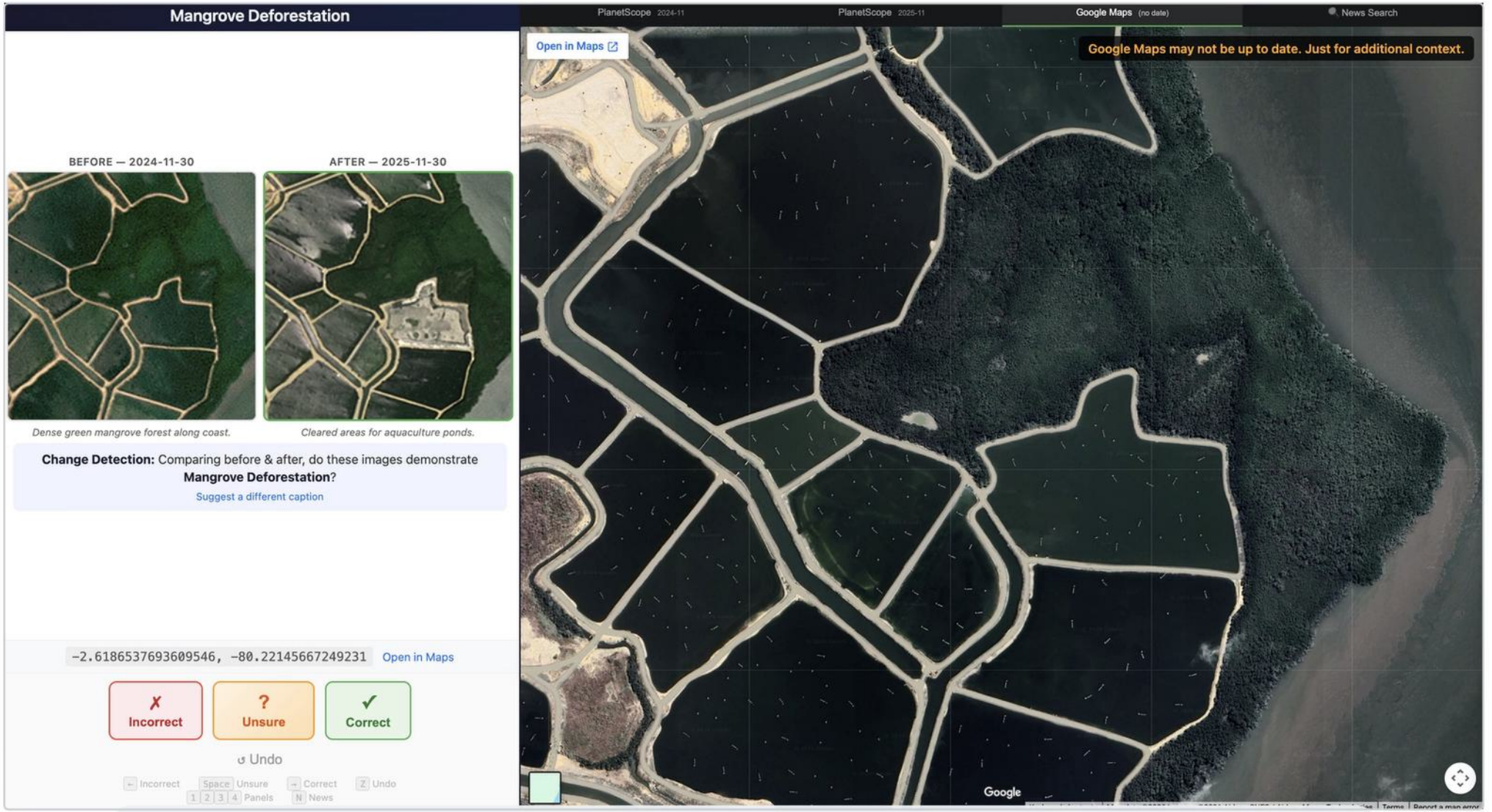
We can also extend this to get stable scenes: generate descriptions where the before equals the after

7 Human verification



To check this process and build a verified test set, we label samples by hand

7 Human verification



Annotators check each event using the before and after imagery, the full PlanetScope scene, a map, and news about the location

8 Event completion

before



Nov '24

after



Nov '25

But events are not just before and after

8 Event completion



They unfold over time: from bi-temporal pairs to event sequences

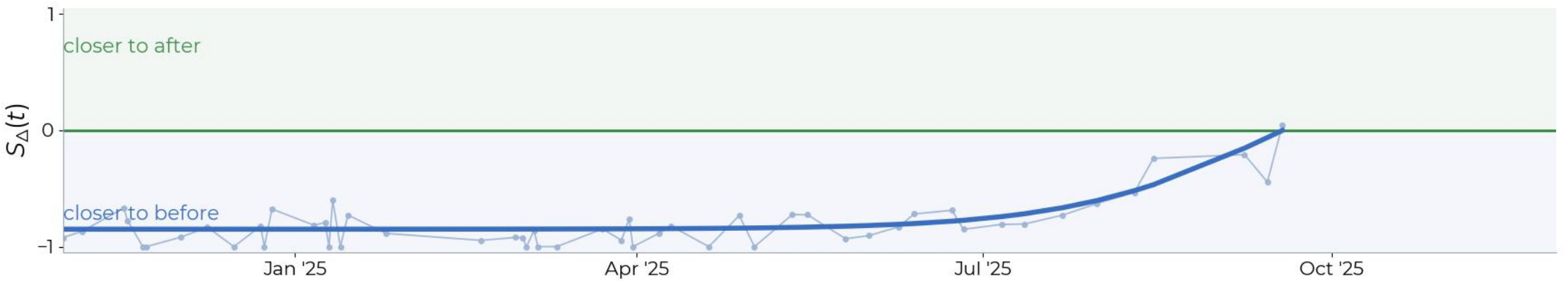
8 Event completion



And once completed, there is no longer an event occurring

8 Event completion



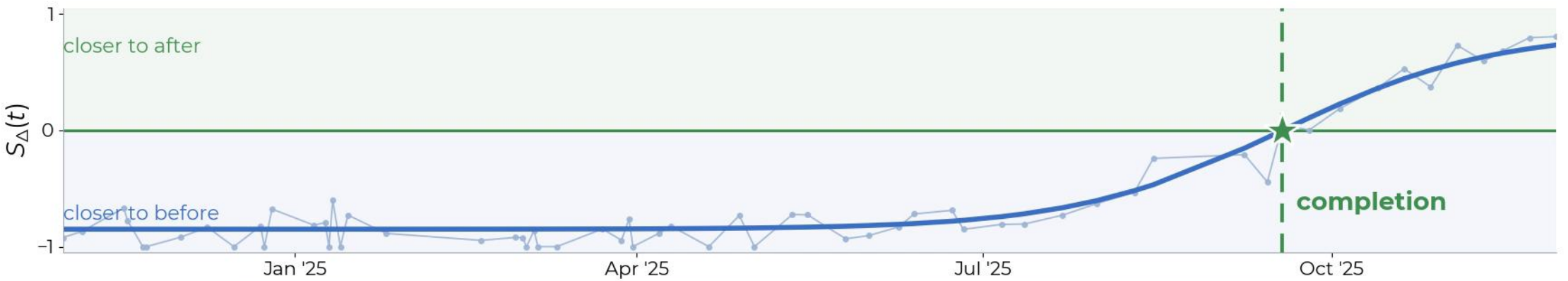


We define completion as the date beyond which anticipation provides no additional value

8 Event completion

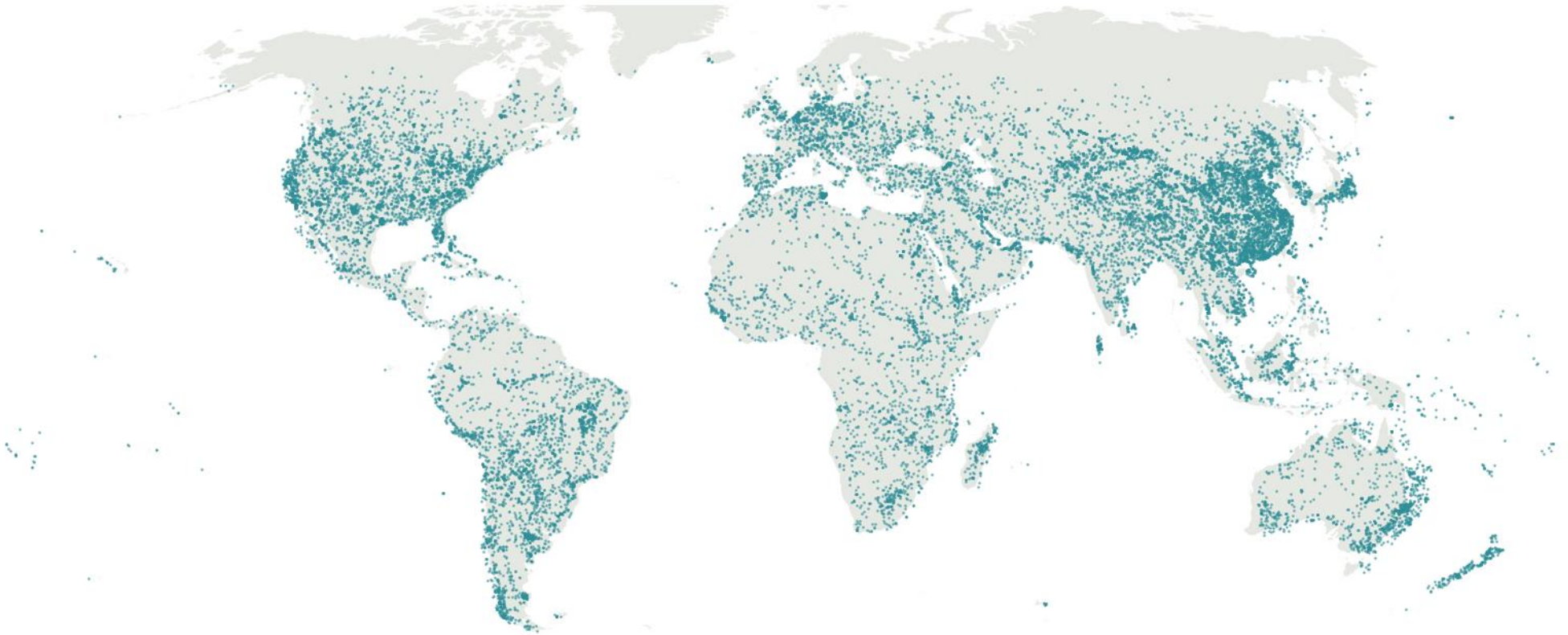


$$S_\Delta(t) = \mathrm{sim}(e_t,\ \mathrm{text}_{after}) - \mathrm{sim}(e_t,\ \mathrm{text}_{before})$$



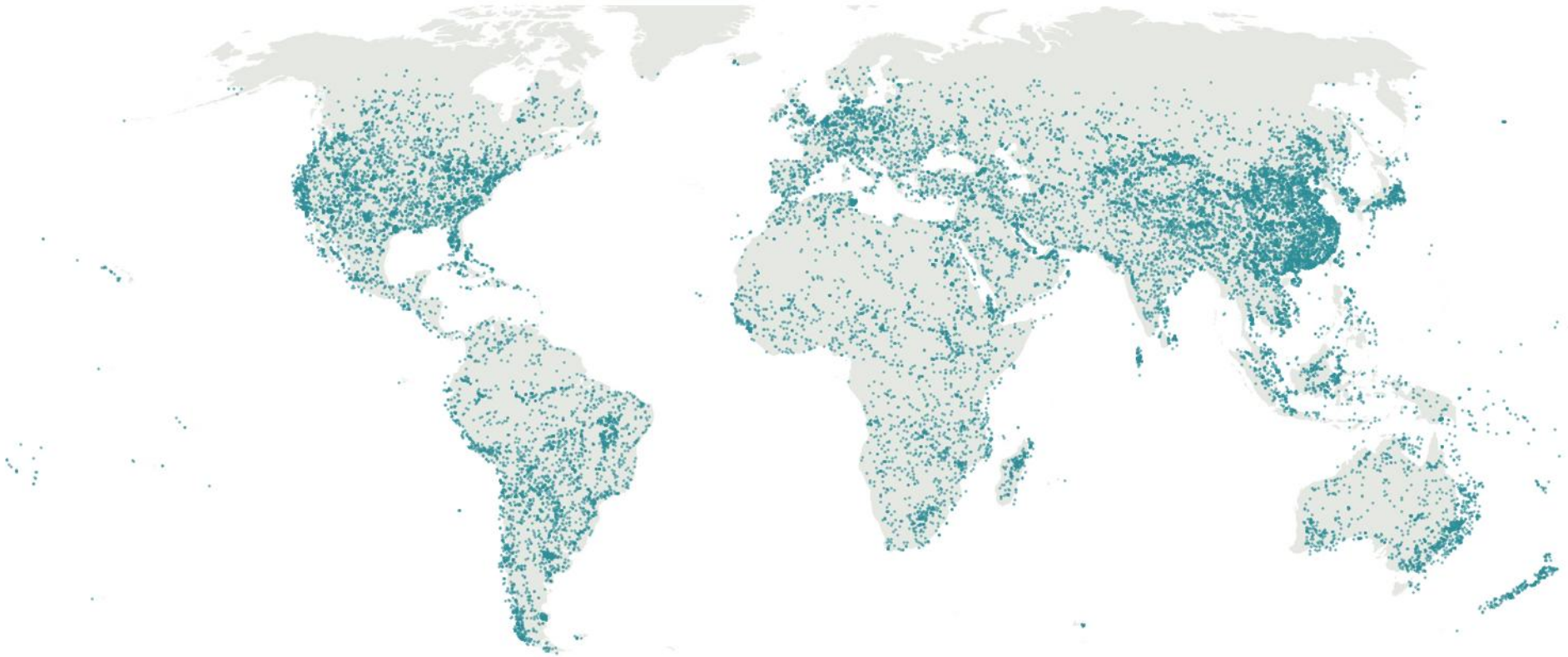
To find this change point, we compute S_Δ: each frame's cosine similarity to the after description minus the before

9 SatEvents-21k



Together, this builds SatEvents-21k: a large collection of events, with completion dates, and stable scenes

9 SatEvents-21k



21k+

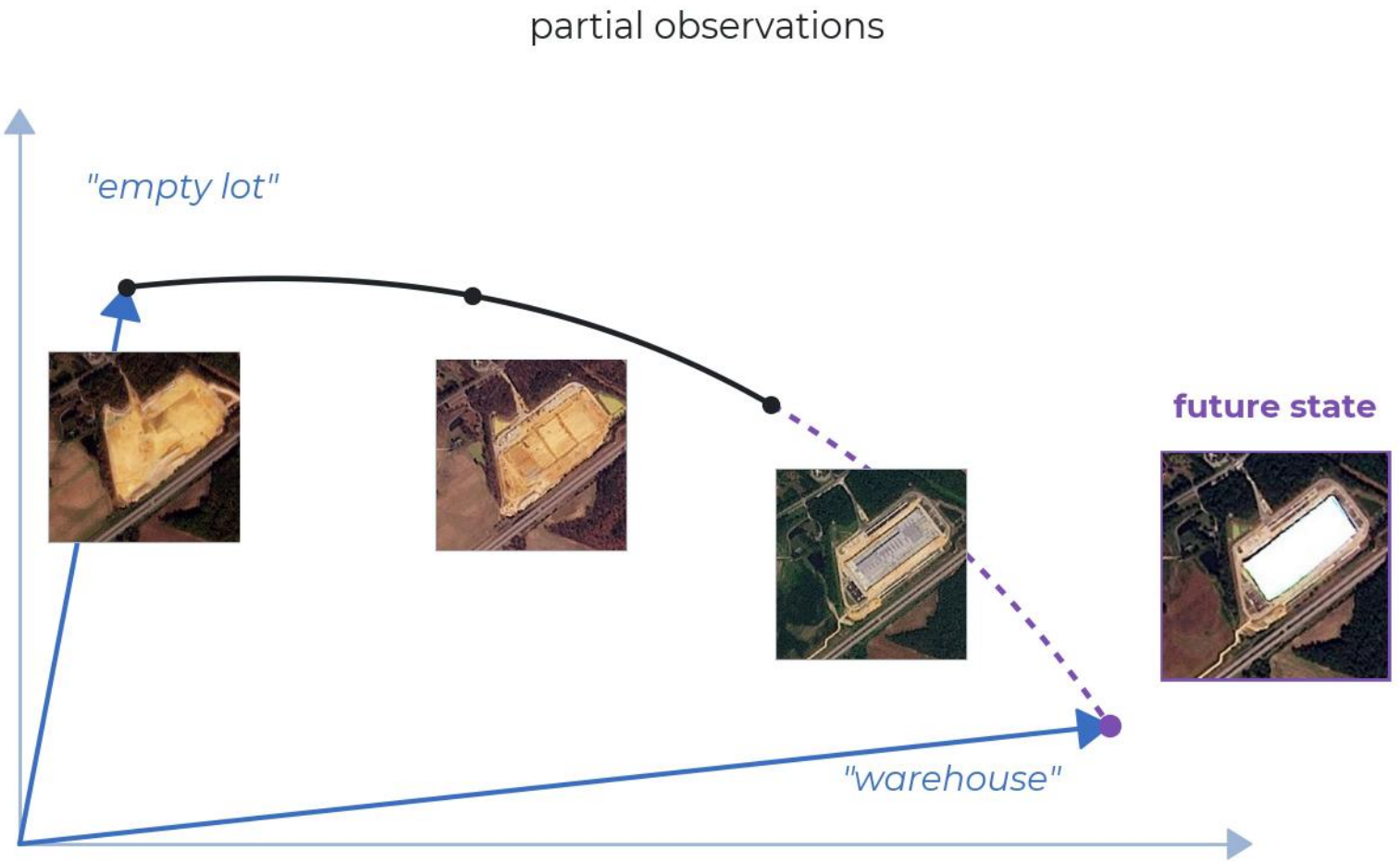
locations

~2M

PlanetScope + Sentinel-2 images

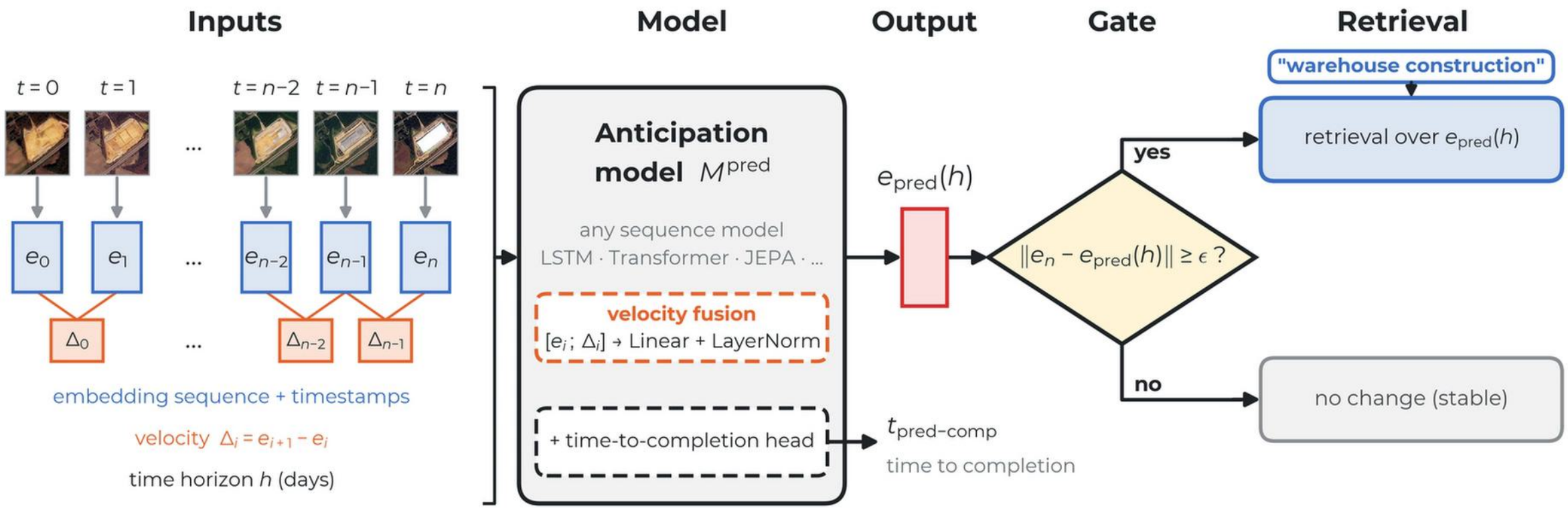
Over 21,000 locations across six continents, and about 2 million PlanetScope and Sentinel-2 images

11 Anticipating in embedding space



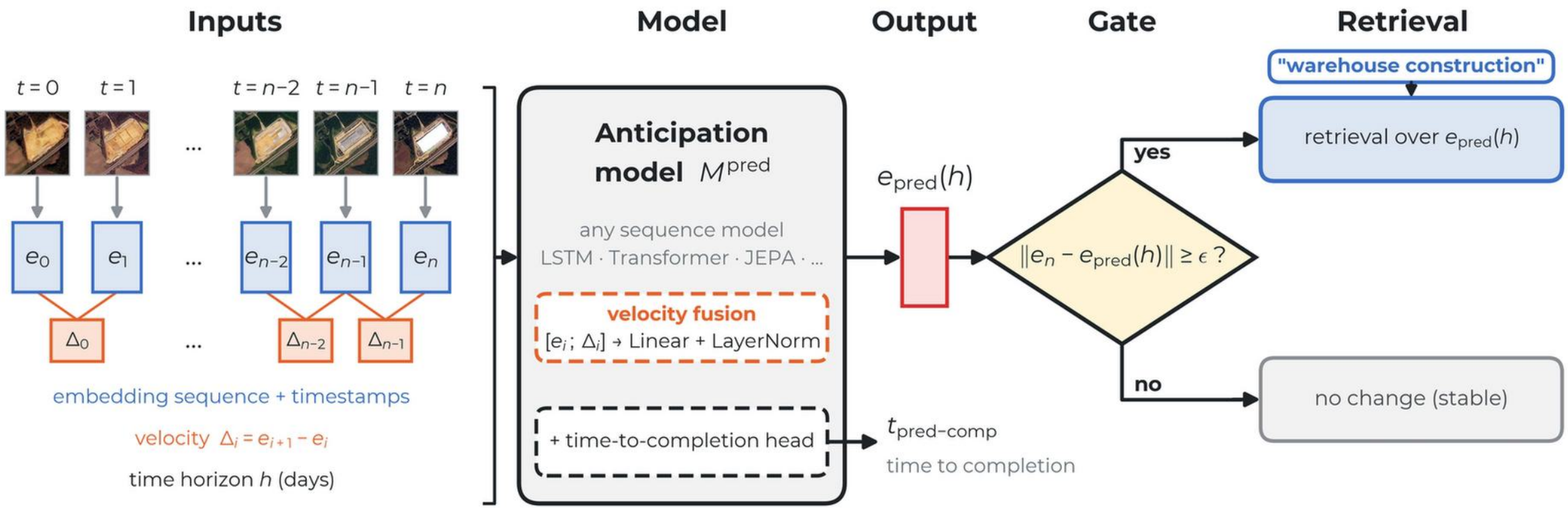
With our global dataset of events, we hypothesize that we can estimate the future state of an event given partial observations

11 Anticipating in embedding space



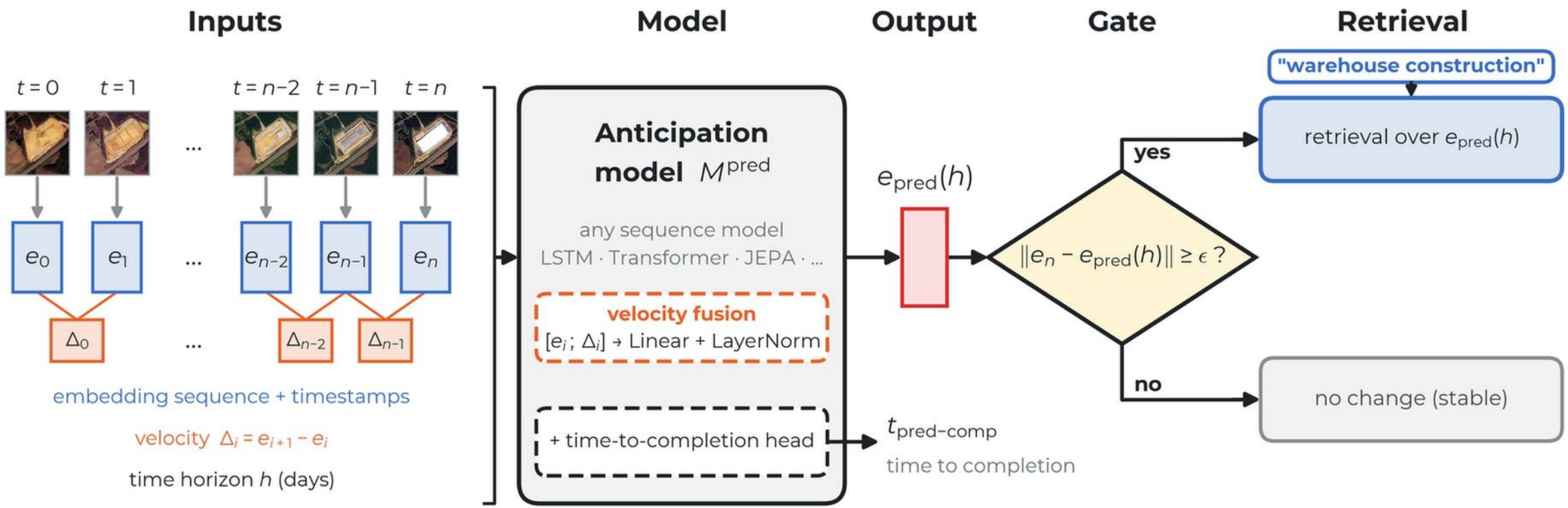
Our model accepts weekly observations of a location and a time horizon h

11 Anticipating in embedding space



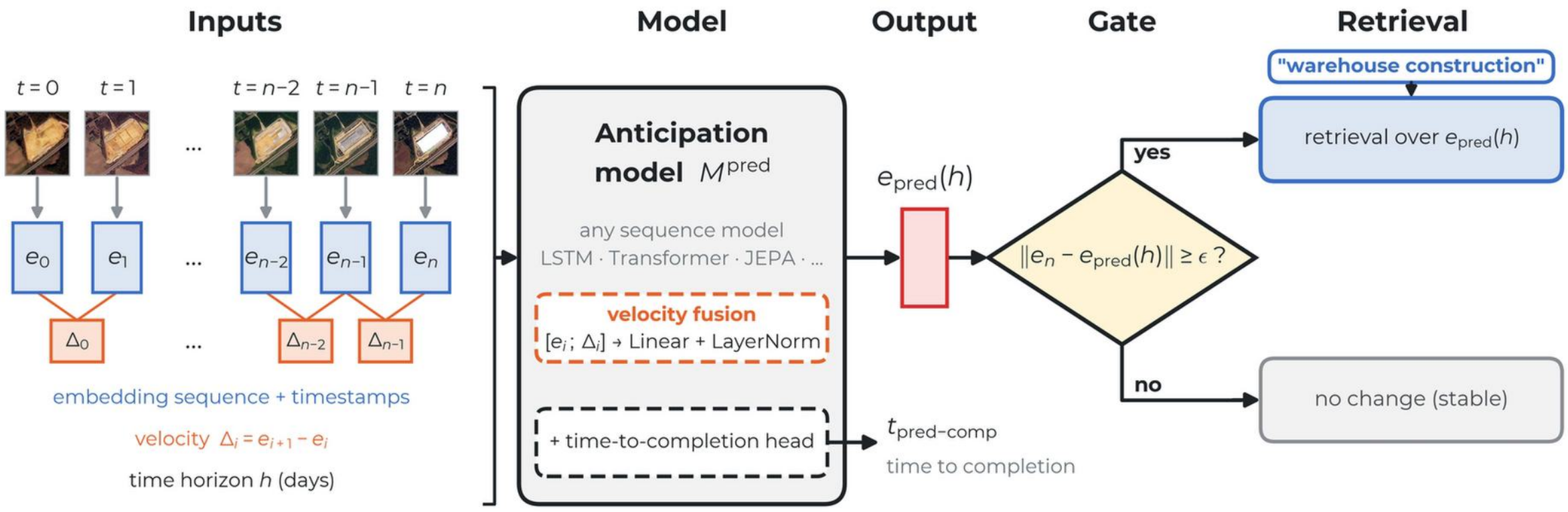
The model then predicts what the location will look like *h* days into the future

11 Anticipating in embedding space



Additionally, a time-to-completion head estimates the days until event completion

11 Anticipating in embedding space

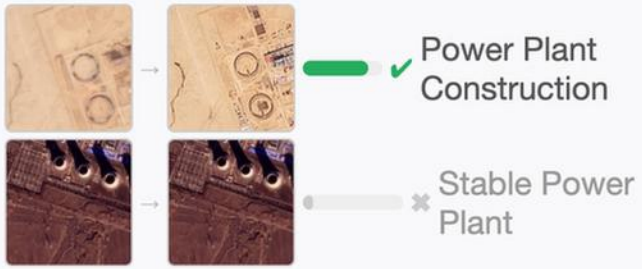


The magnitude of the predicted change acts as a detection gate

11 Anticipating in embedding space

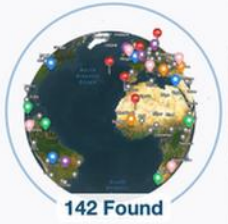
CAPABILITIES

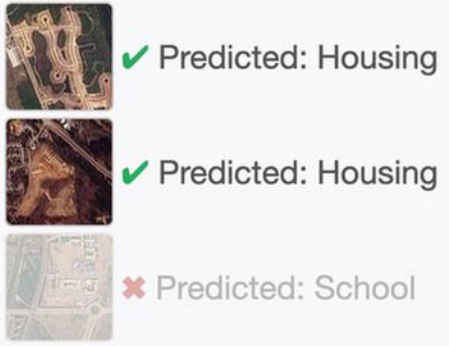
CHANGE DETECTION



OPEN-VOCABULARY PREDICTION

"Where is new housing development occurring?"





COMPLETION FORECAST

"When will this reservoir be filled?"





GROUNDTRUTH FUTURE
Power Plant
265 Days in Future



GROUNDTRUTH FUTURE
Housing
250 Days in Future



GROUNDTRUTH FUTURE
Reservoir
215 Days in Future

This gives us one model with three capabilities: change detection, open-vocabulary event anticipation, and time-to-completion prediction

12 Results

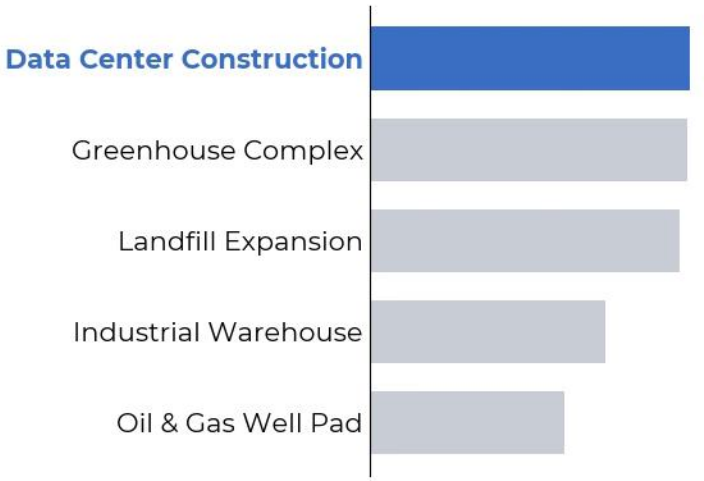
observed (model input)

$t = 0$ $t = 5$ $t = n$



10 weekly observations

model prediction



time to completion: 287 days

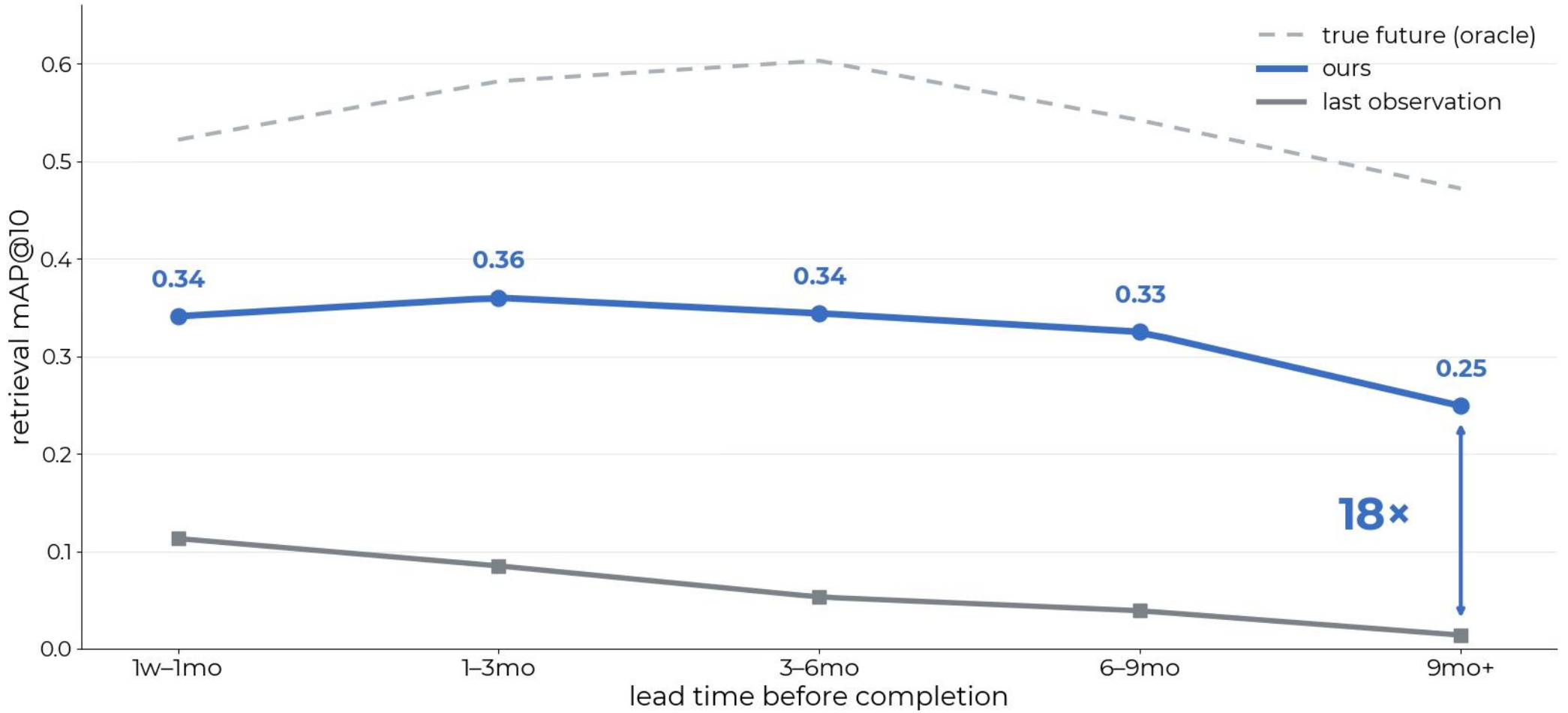
ground truth



actual: 317 days

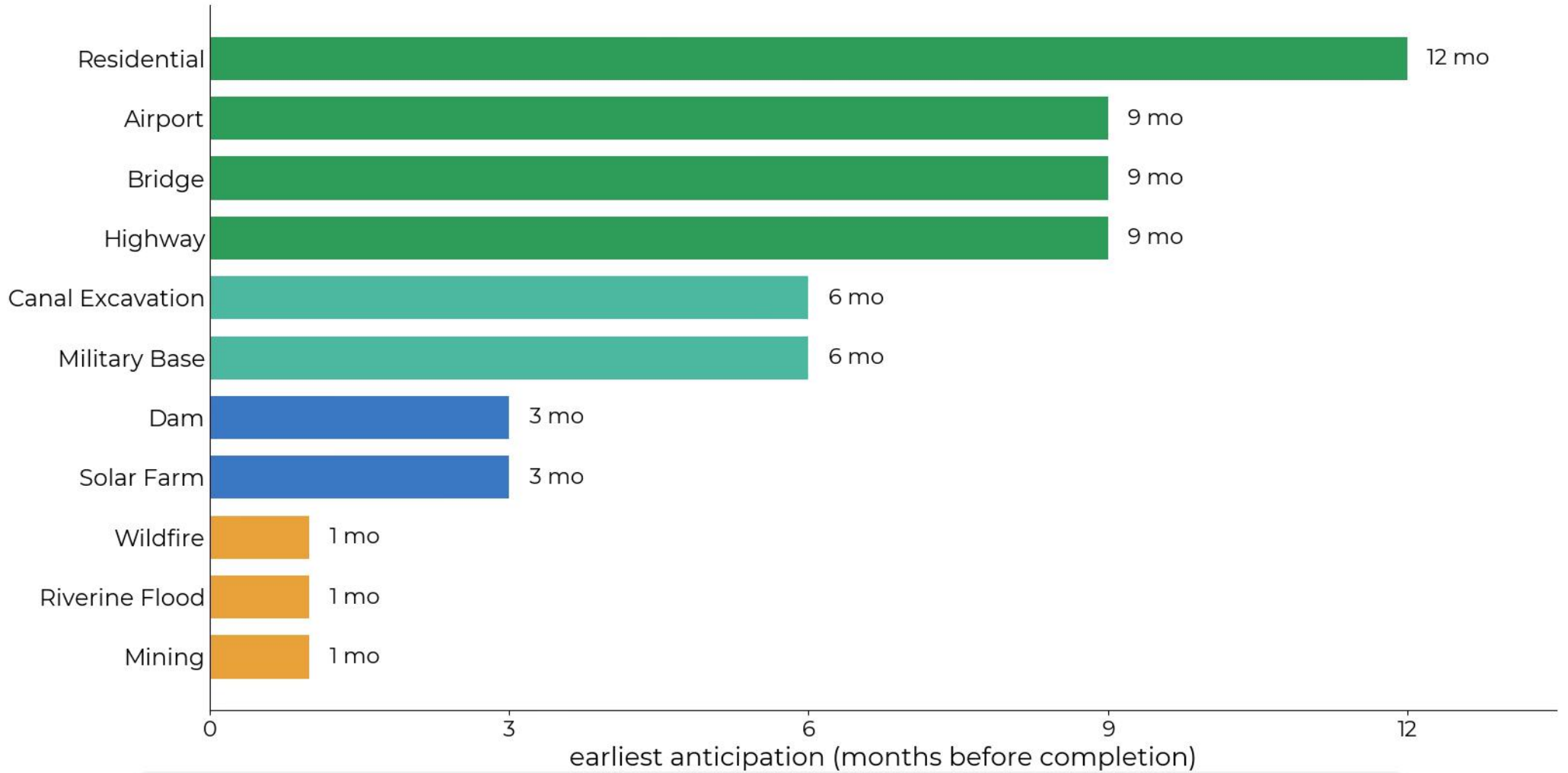
Given observations of any location, our model can anticipate the expected event months in advance

12 Results



Anticipation performance holds across time horizons, from a week to more than nine months into the future

12 Results



With larger infrastructure anticipated nine to twelve months before completion

Predicting the Needle in a Petabyte Scale Haystack

Open-Vocabulary Event Anticipation in Satellite Imagery

NEURAL INFORMATION PROCESSING SYSTEMS

website · code · data

Website

cs.cornell.edu/~revankar/predicting-the-needle

Code

github.com/ShreelekhaR/predicting-the-needle

Contact

revankar@cs.cornell.edu

Cornell University

planet.

Thanks for watching, and check out these resources!