Datamole

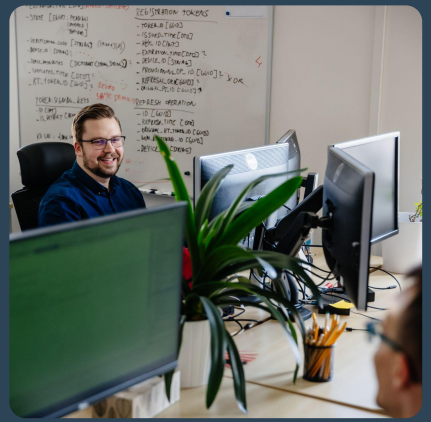
Animal 3D Reconstruction

Simon Pokorny & Tibor Szolar

2024-11-01

Since 2015

We’re helping companies worldwide become data & AI driven.



100+ projects

We have successfully delivered 100+ data, IoT, AI & UI projects.

80+ specialists

We’re a mix of data scientists, engineers, analysts, and full-stack developers based in Prague, Brno and Cesky Krumlov.

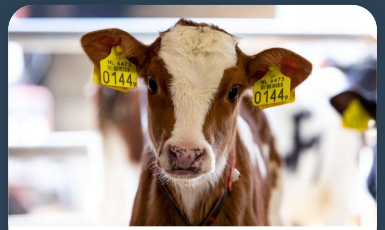
Industrial focus

We work with companies mostly from agritech, machinery, manufacturing, biotech & foodtech industries.

DATAMOLE

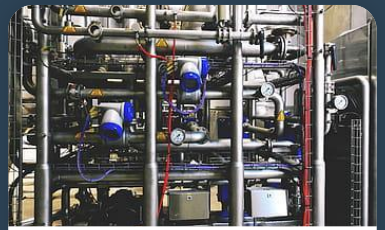
Industrial focus: For a sustainable future

At Datamole, we believe in responsible and ethical data applications that contribute positively to the physical world. This guides our choice of industries we serve:



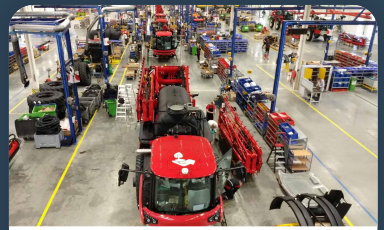
Agritech

Nurturing **sustainable** agriculture practices through data-driven solutions



Machinery

Empowering machinery with predictive maintenance, ensuring **uptime and machine efficiency**



Manufacturing

Streamlining processes, optimizing resources, and **minimizing waste** for a sustainable future

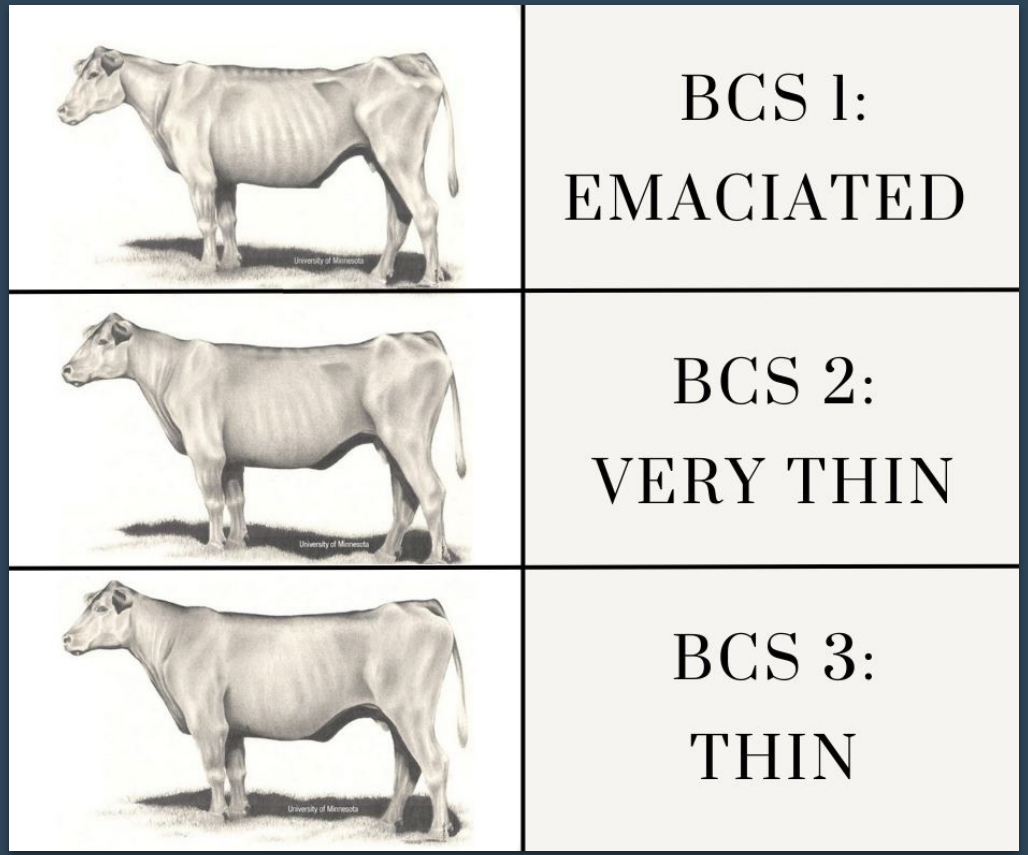
Biotech & foodtech

Revolutionizing the food industry with **responsible and efficient** data utilization

Introduction - problem definition

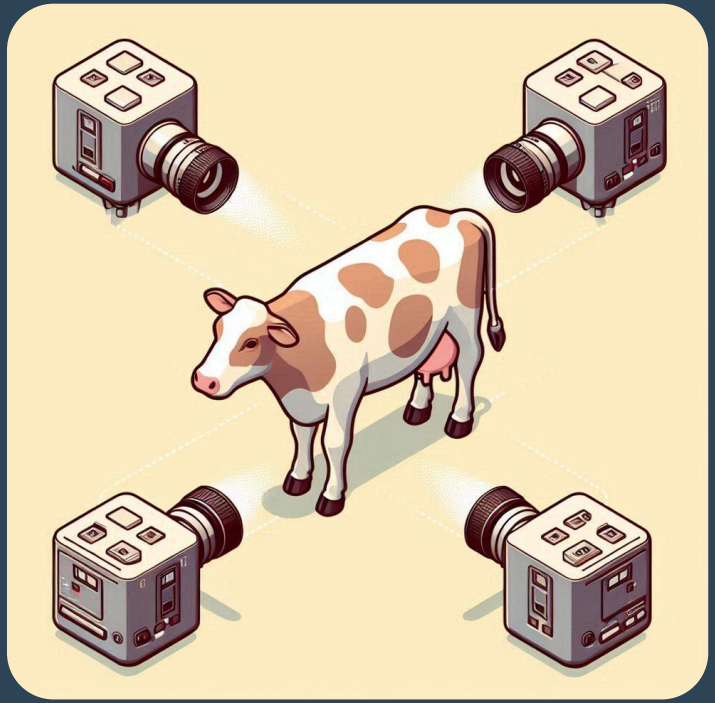
3D shape of a cow is very valuable:

- Automated determination of BCS
- Disease detection
- Feeding advice
- Reproduction advice
- Reproduction scoring
- Hypothetical:
  - Rumen movement evaluation
  - Cow identification



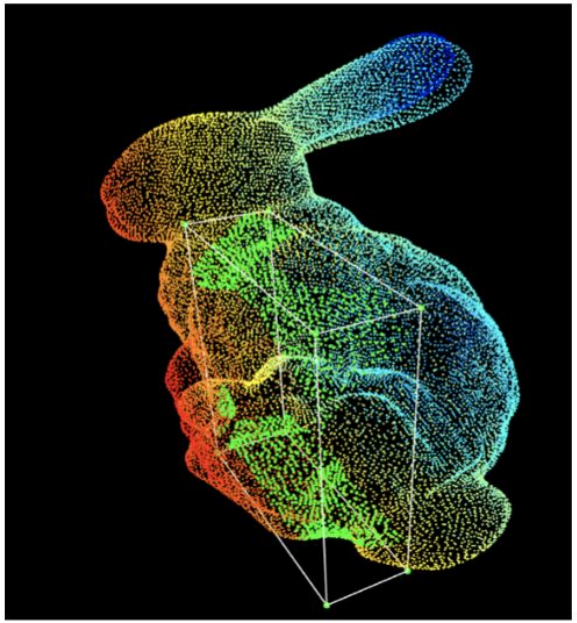
DATAMOLE

Cow 3D reconstruction approach

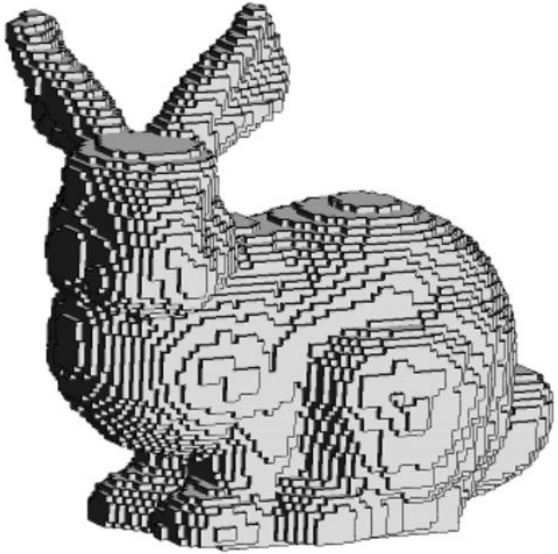


- Regular RGB cameras
- Lowest number that works
  ⇒ cost & maintenance
- Shape estimation is sufficient
  (we don’t need mm accuracy)

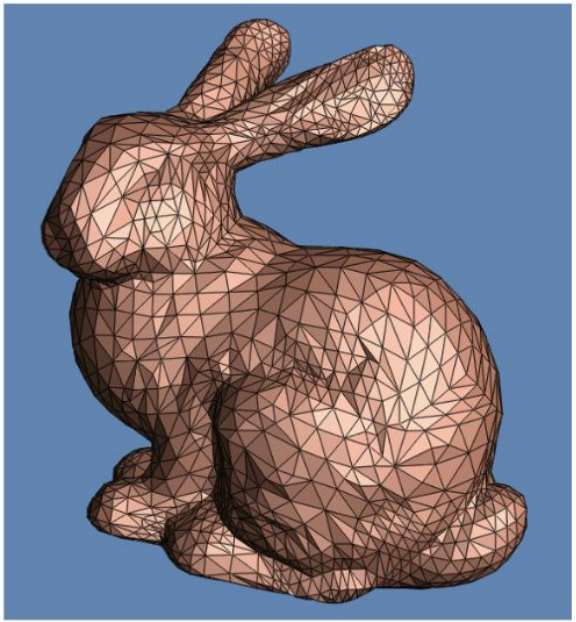
Description of 3D space



Point clouds

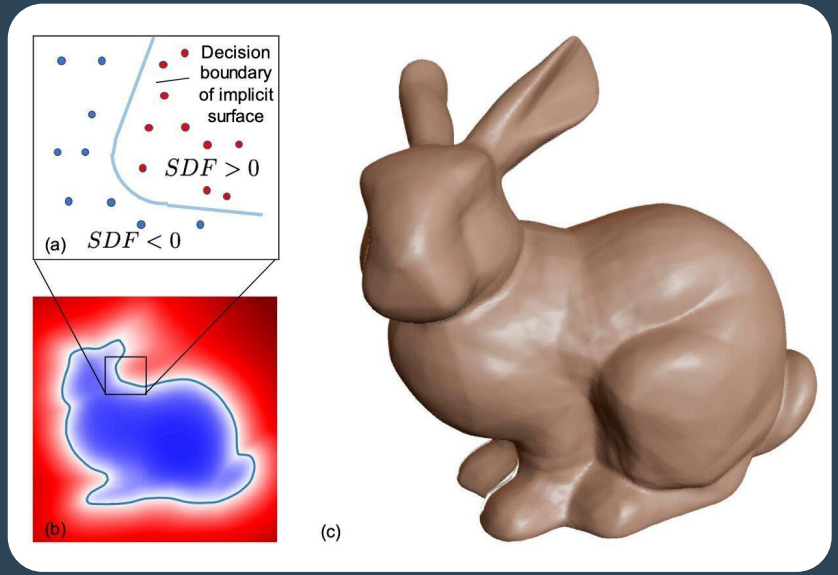


Voxels

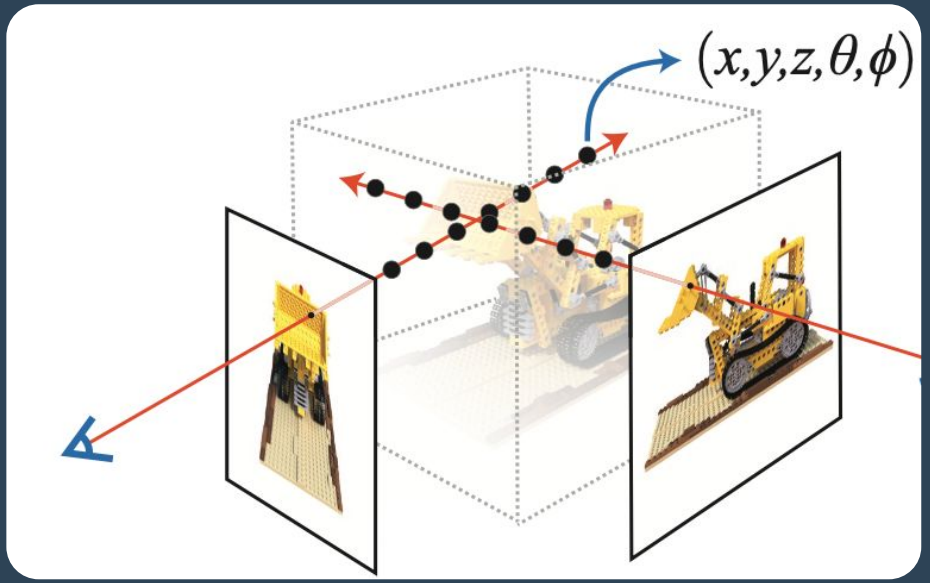


Meshes

Description of 3D space

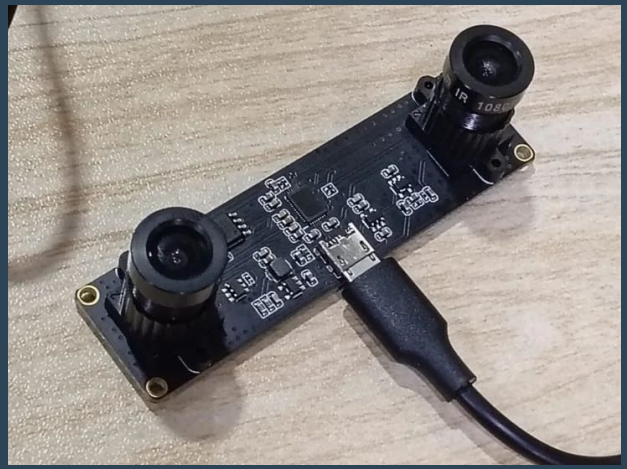


Signed distance function

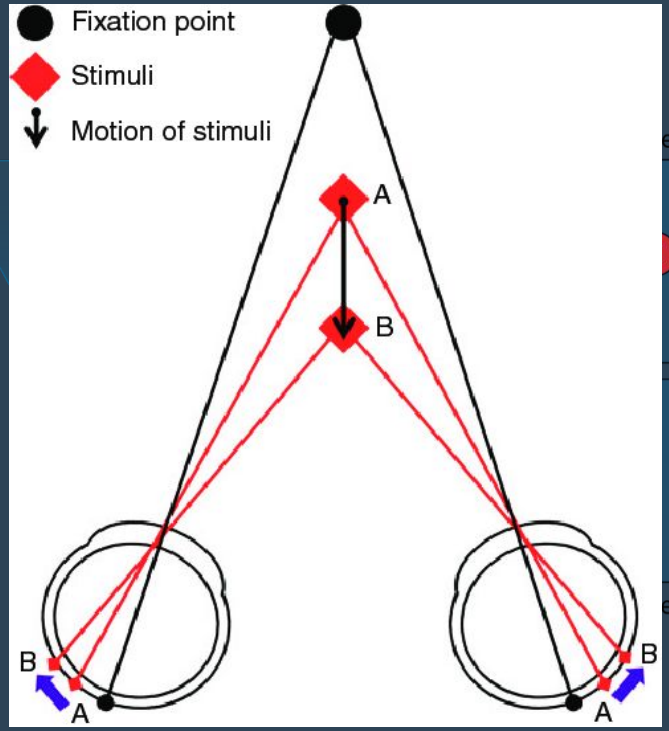


Implicit representations

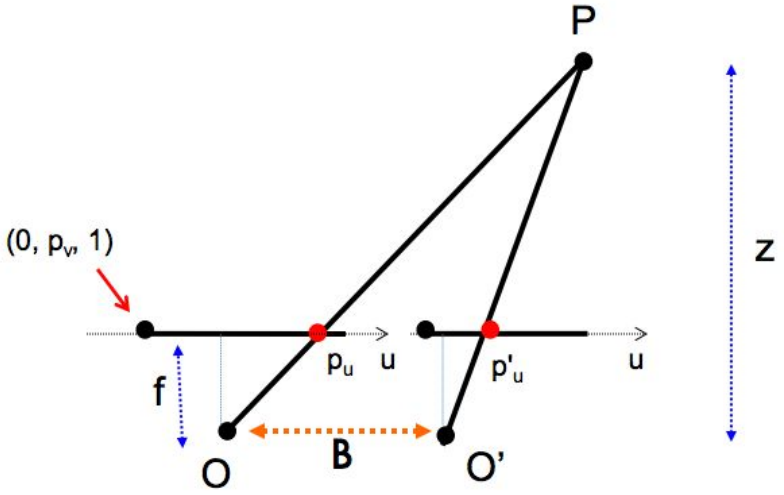
Calibrated stereo



Disparity



Calibrated stereo pair



z = depth

B = baseline
f = focal length

$$\text{disparity} = p_u - p'_u \propto \frac{B \cdot f}{z} \quad \text{[Eq. 1]}$$

Disparity is inversely proportional to depth z!

Output: depth map

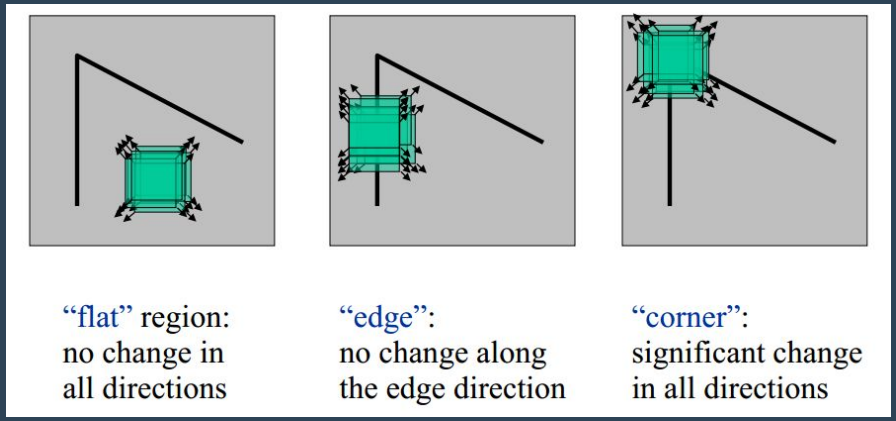
Photogrammetry

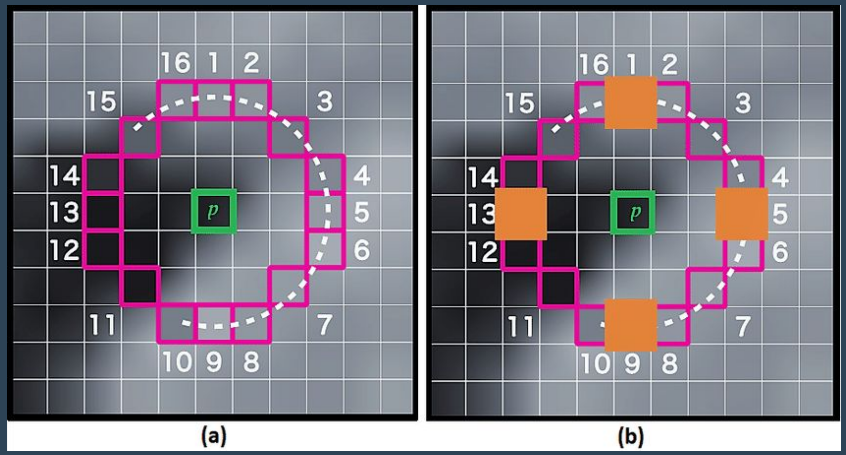
Our input to photogrammetry pipeline

Correspondence - Wide Baseline Stereo Pipeline

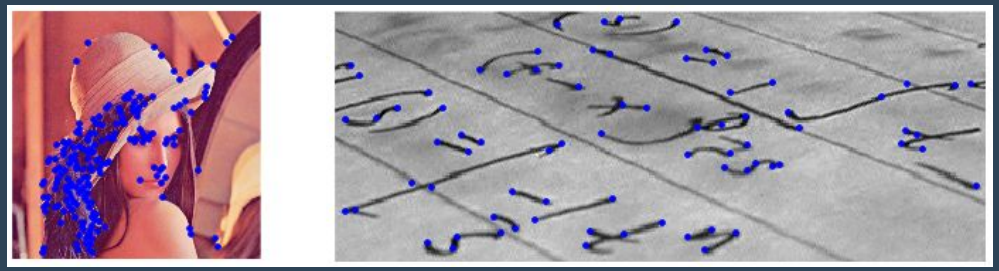
Feature Detection → Feature Description → Descriptors Matching → Geometrical Verification

Goal: find *some* points of interest.





FAST Corner detector

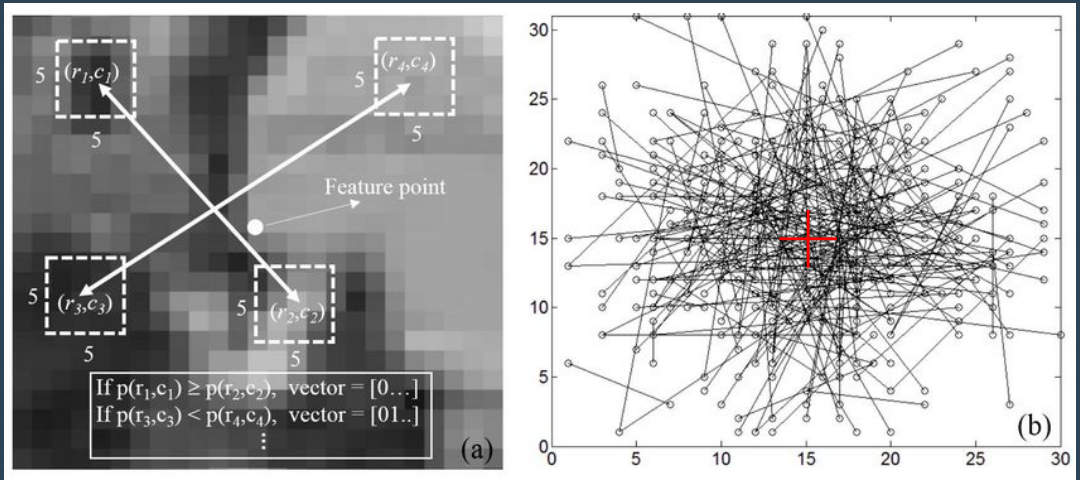


DATAMOLE

Correspondence - Wide Baseline Stereo Pipeline

Feature Detection → **Feature Description** → Descriptors Matching → Geometrical Verification

Goal: describe each point of interest with a feature vector.



Descriptor calculation: comparison of image intensities around detected feature.

Output: Binary vector of length 256 for each keypoint

ORB = Oriented FAST + Rotated BRIEF

DATAMOLE

Correspondence - Wide Baseline Stereo Pipeline

Feature Detection → Feature Description → **Descriptors Matching** → Geometrical Verification

Goal: match descriptors across two images.

Similarity metric

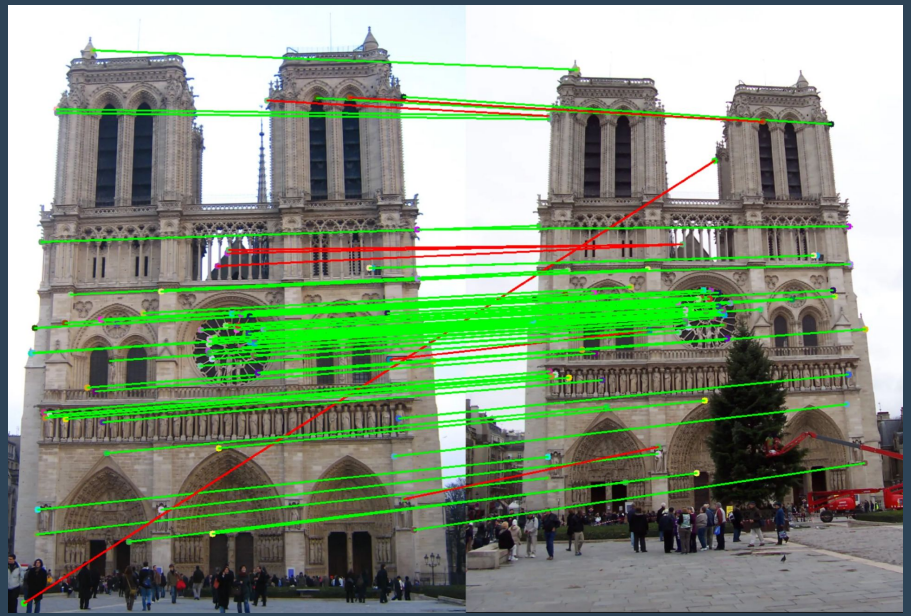
Binary descriptor:
- Hamming distance

Numerical descriptor:
- Euclidean dis.
- Cosine dist.
- ...

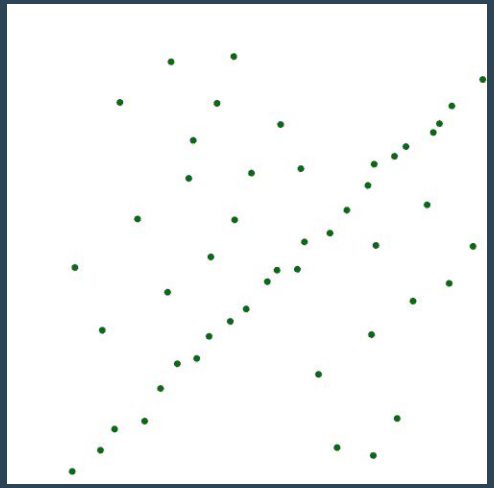
Correspondence - Wide Baseline Stereo Pipeline

Feature Detection → Feature Description → Descriptors Matching → Geometrical Verification

Goal: verify which matches are valid

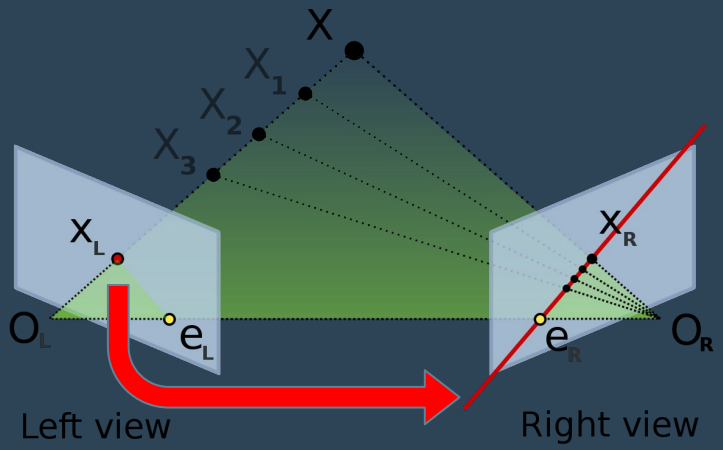


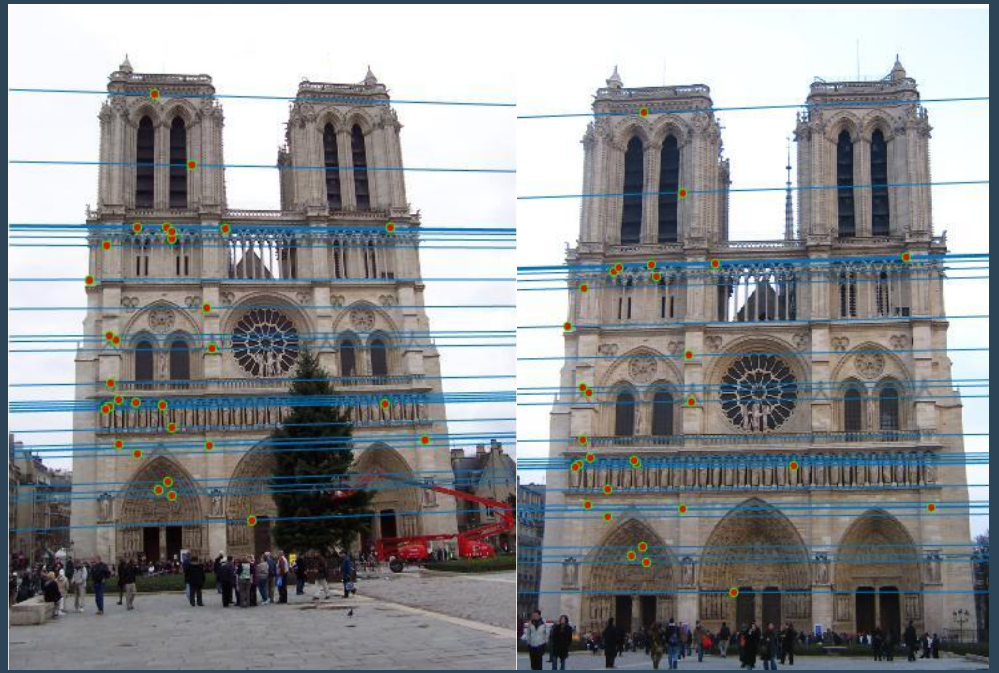
Geometric model



\+ RanSaC (Random Sample Consensus)

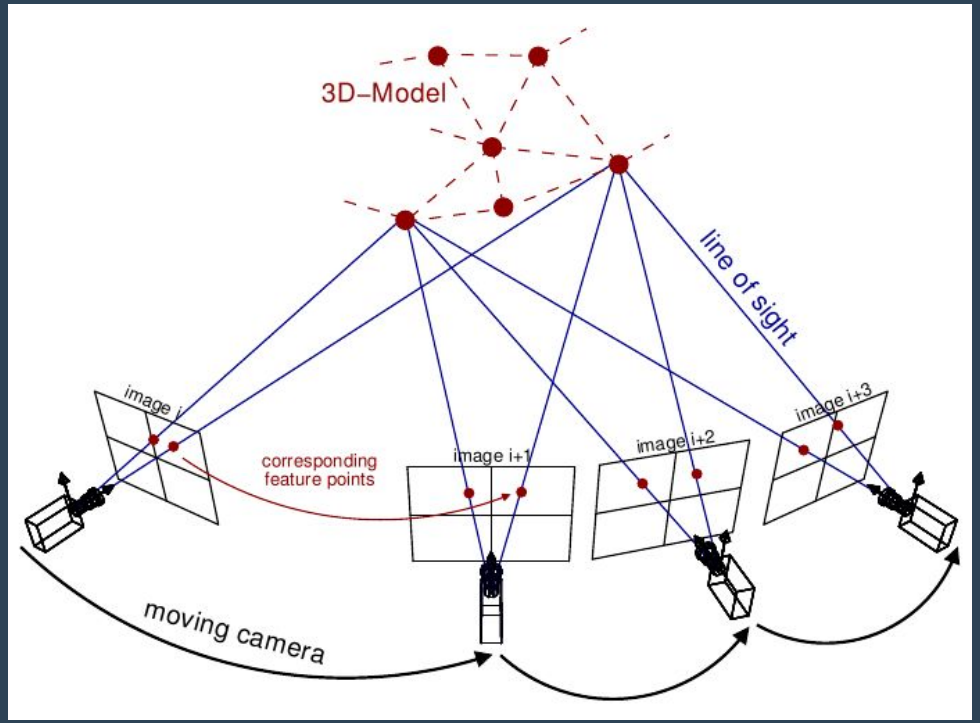
Epipolar constraint





Epipolar constraint: Point from one image must lie on an **epipolar line** in the other image.

Incremental Structure from Motion



Initialization

Image Registration

Triangulation

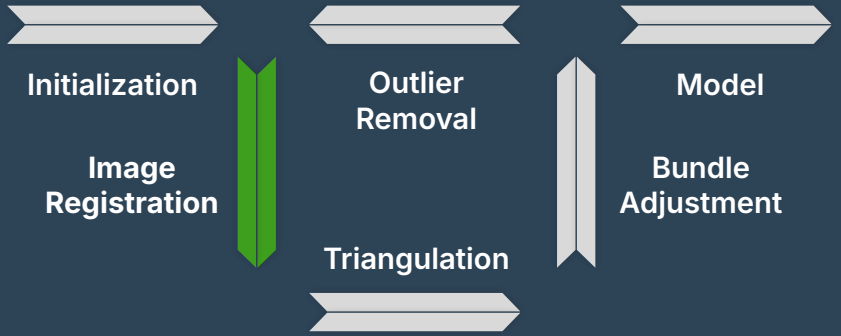
Outlier Removal

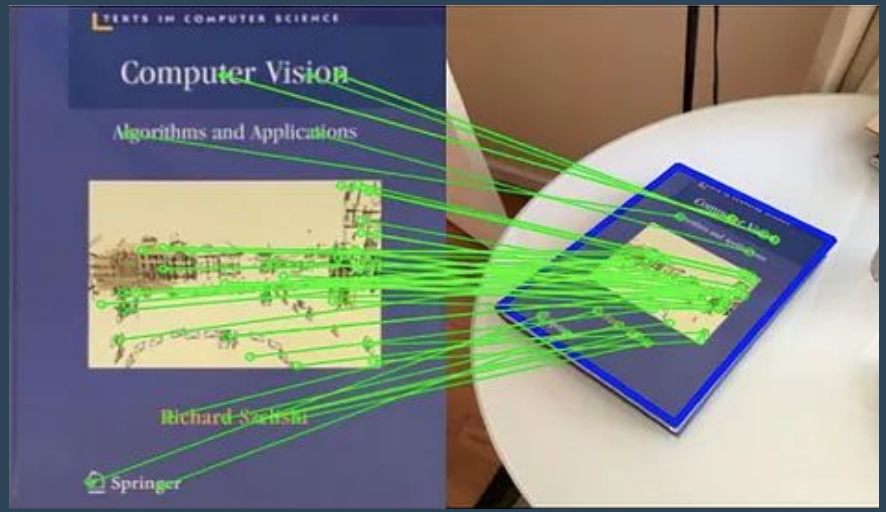
Model

Bundle Adjustment

DATAMOLE

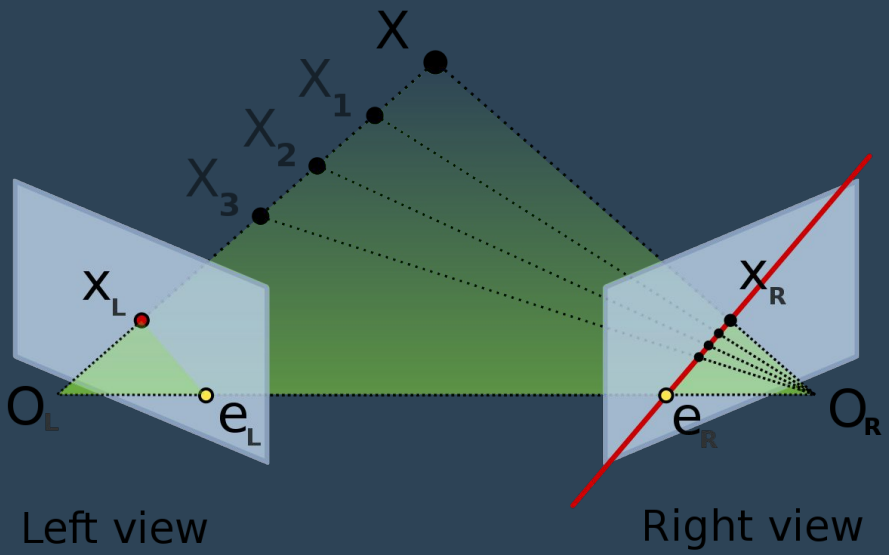
Incremental Structure from Motion





Feature detection and matching!

Incremental Structure from Motion



Epipolar geometry!

Initialization

Image Registration

Outlier Removal

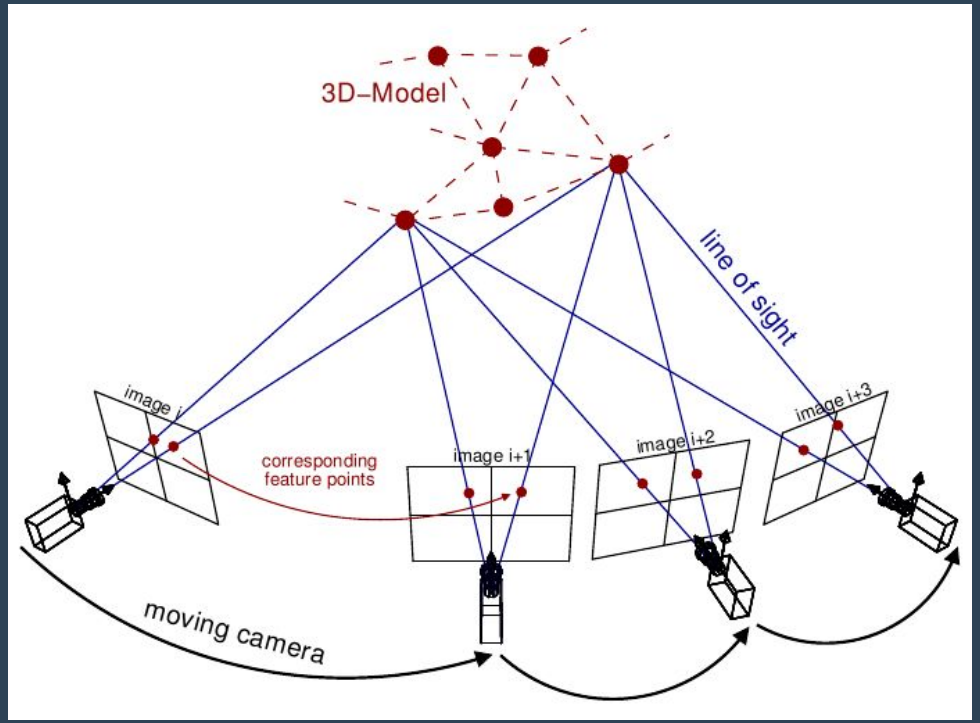
Triangulation

Model

Bundle Adjustment

DATAMOLE

Incremental Structure from Motion



Initialization

Image Registration

Outlier Removal

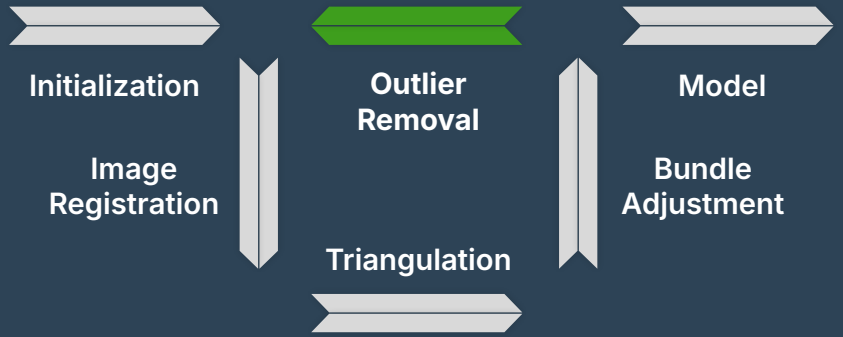
Triangulation

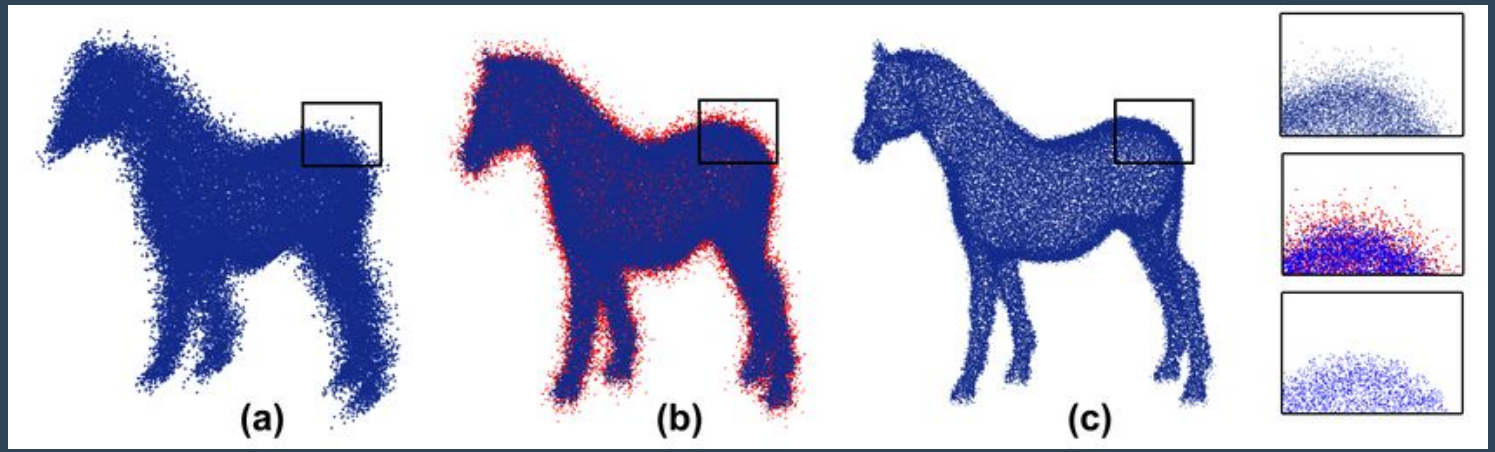
Model

Bundle Adjustment

DATAMOLE

Incremental Structure from Motion





Incremental Structure from Motion

Initialization

Image Registration

Triangulation

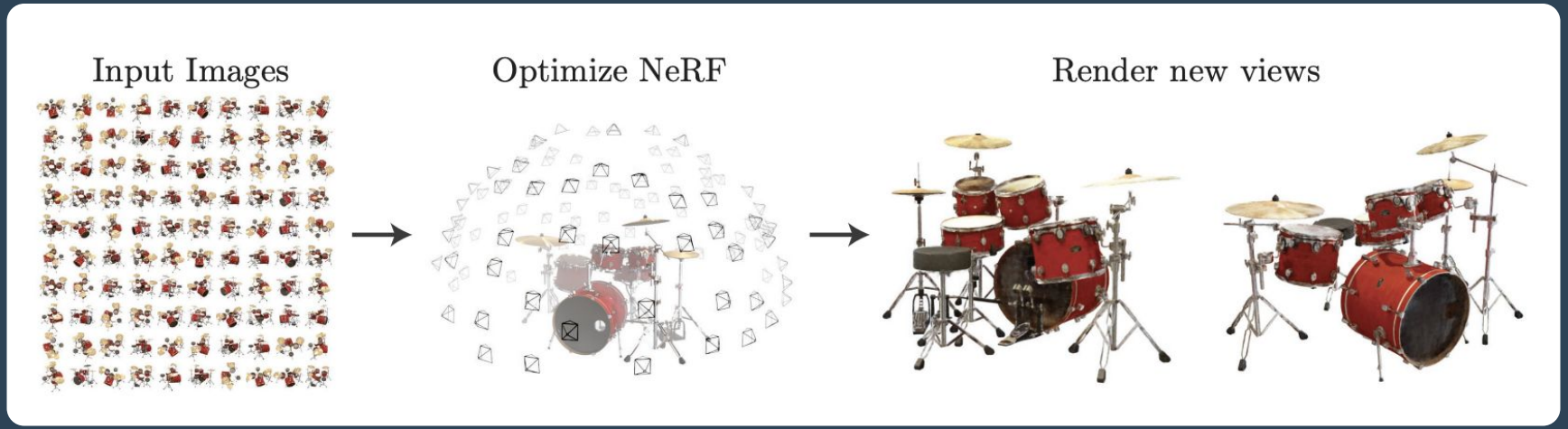
Outlier Removal

Bundle Adjustment

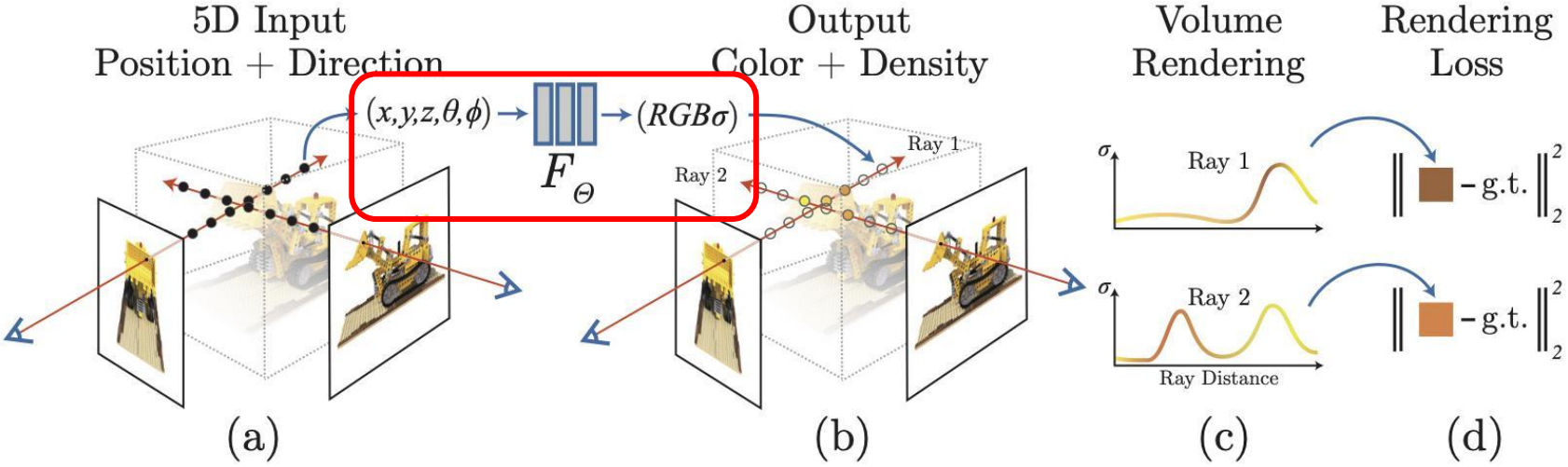
Model

Dense model (mesh) from SfM [COLMAP]

- Novel approach → fully-connected neural network to model a continuous 3D scene from a sparse set of 2D images [NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis]

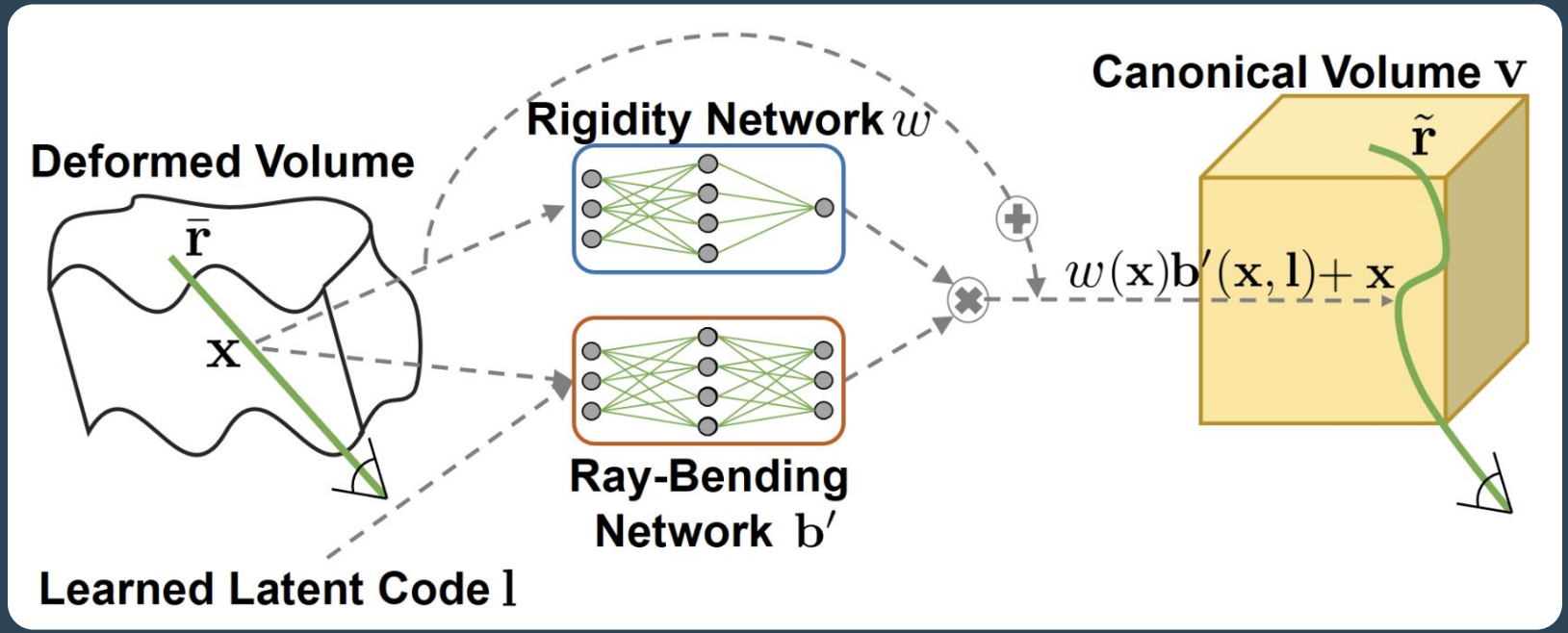


NERF → training regime



→ Dynamic scenes ?

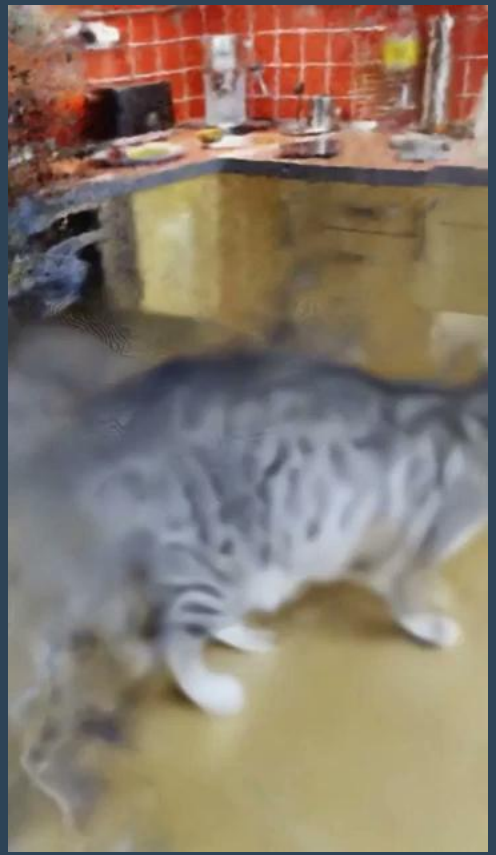
→ Dynamic scenes [Non-rigid Nerf]



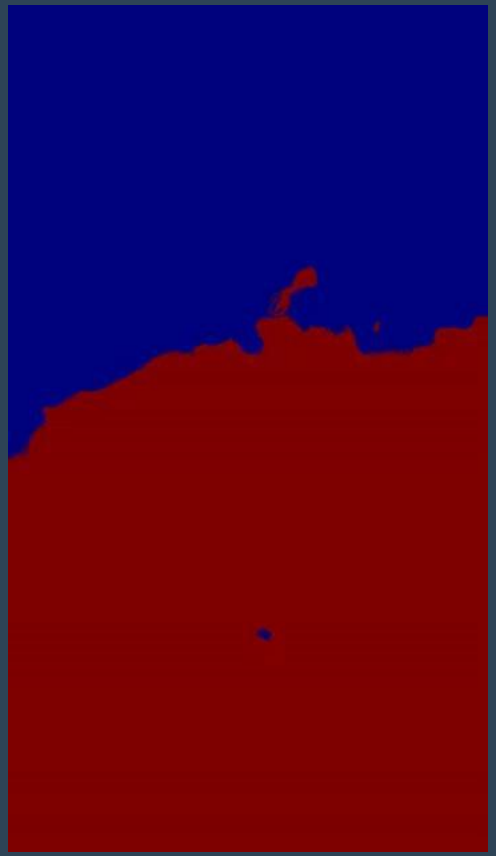
→ Results

→ Dynamic scene, input video

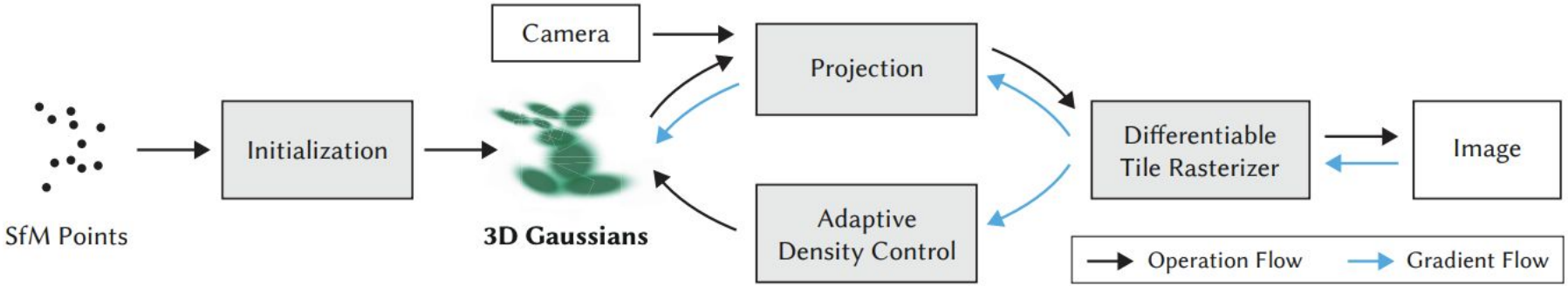
→ Dynamic scene, results → RGB



→ Dynamic scene, results → depth map



Gaussian Splatting



Gaussian Splatting

Gaussian Splatting

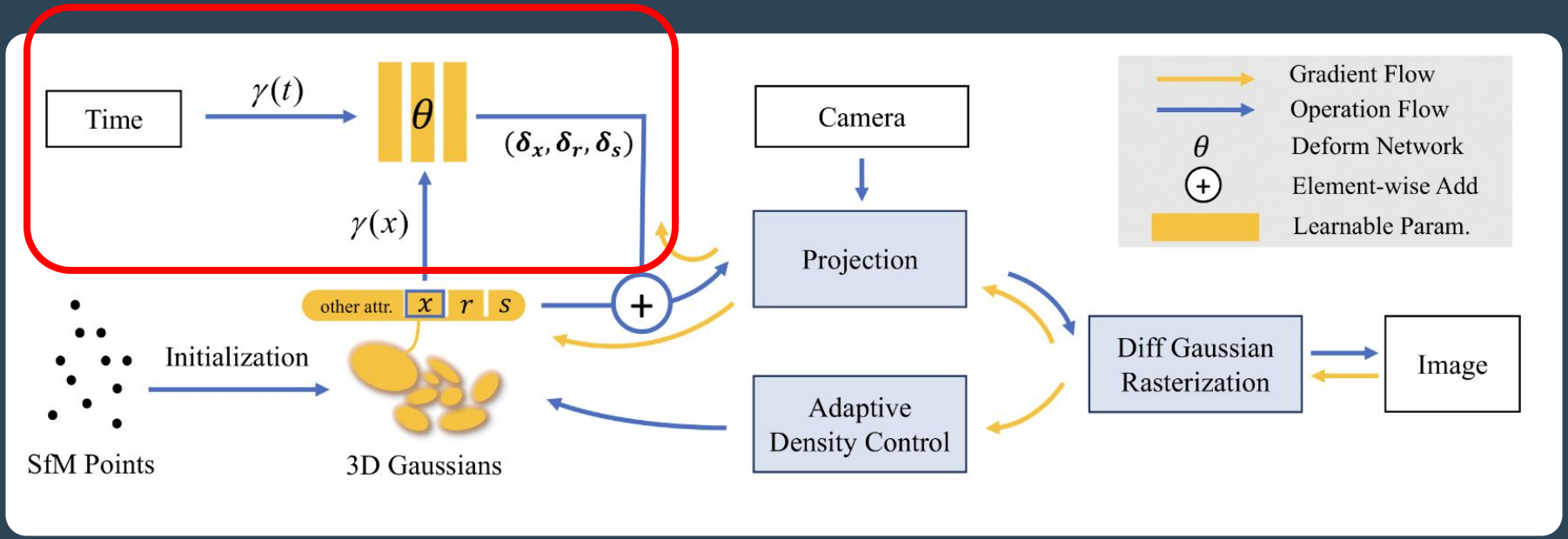
DATAMOLE

Gaussian Splatting

SfM
Initialisation

Random
Inititialisation

Gaussian Splatting → Dynamic scene ?



Gaussian Splatting → Dynamic scene, input video

Gaussian Splatting → Dynamic scene, result, continuous time

continuous time

Gaussian Splatting → Dynamic scene, result, fixed time

fixed time

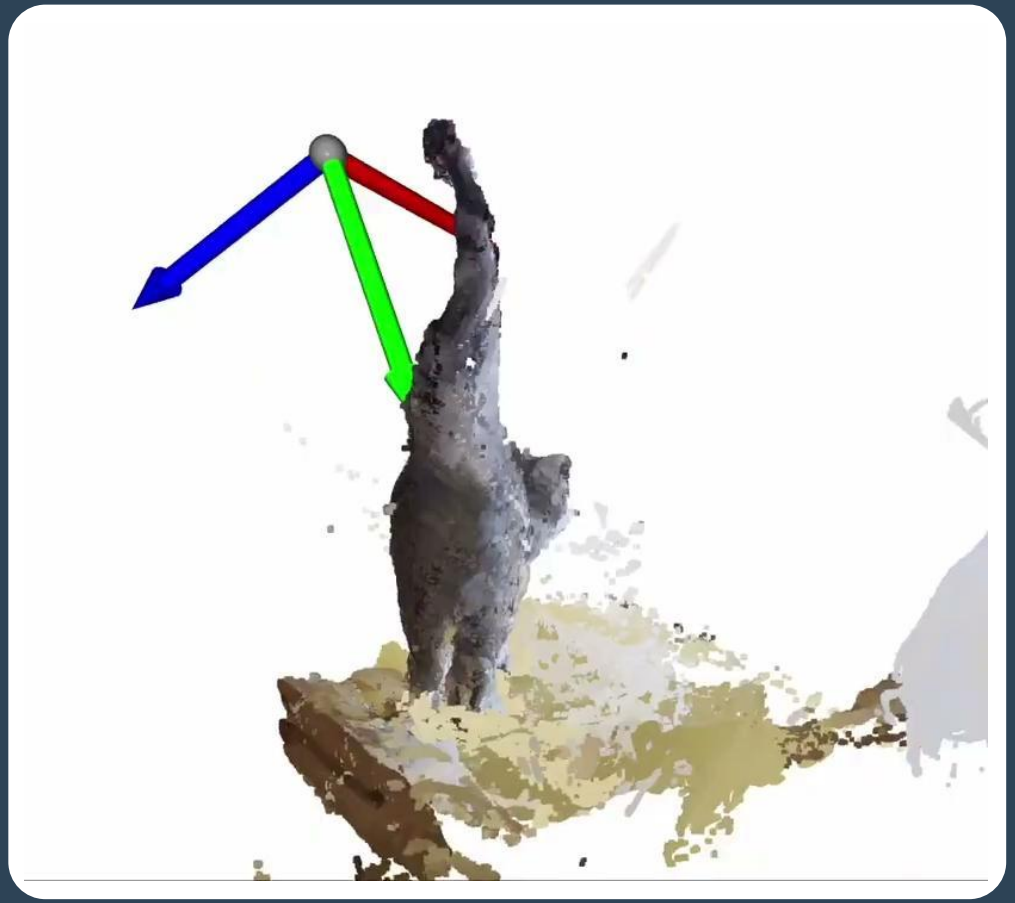
Gaussian Splatting

pointcloud

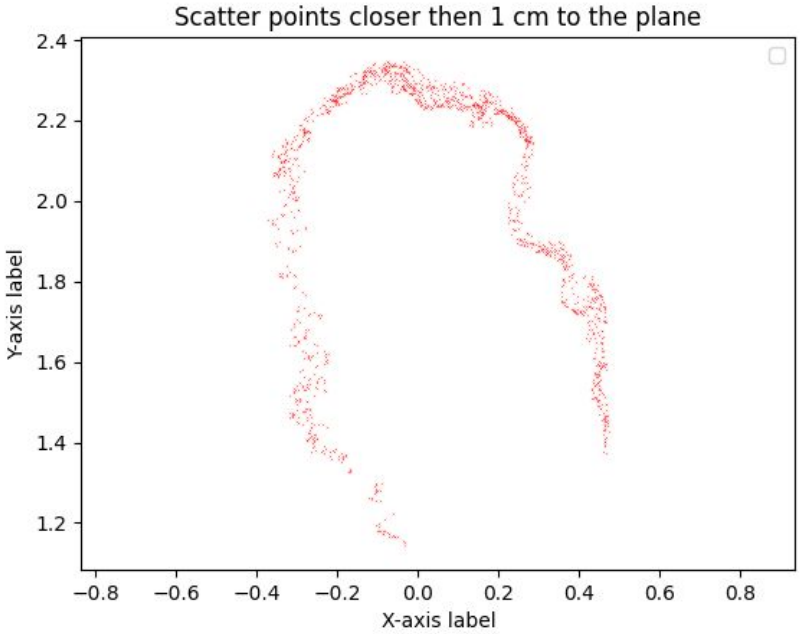
Gaussian Splatting

pointcloud

Gaussian Splatting



Gaussian Splatting



Gaussian Splatting → Static scene, few input images

DATAMOLE

Gaussian Splatting → Static scene, results on few images





DATAMOLE

Gaussian Splatting → comparison

Original

Depth

Prediction

Difference

Gaussian Splatting → depth

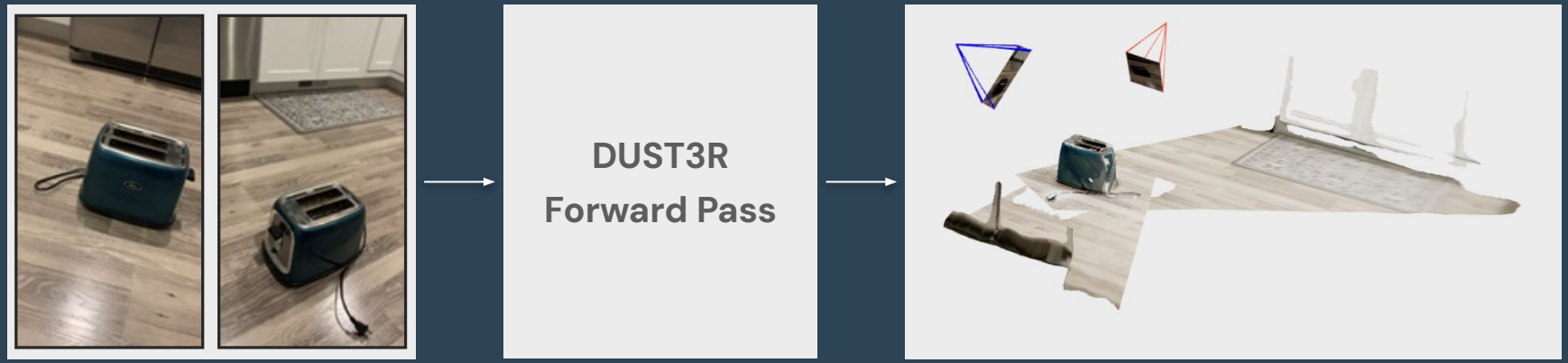
Depth

Gaussian Splatting

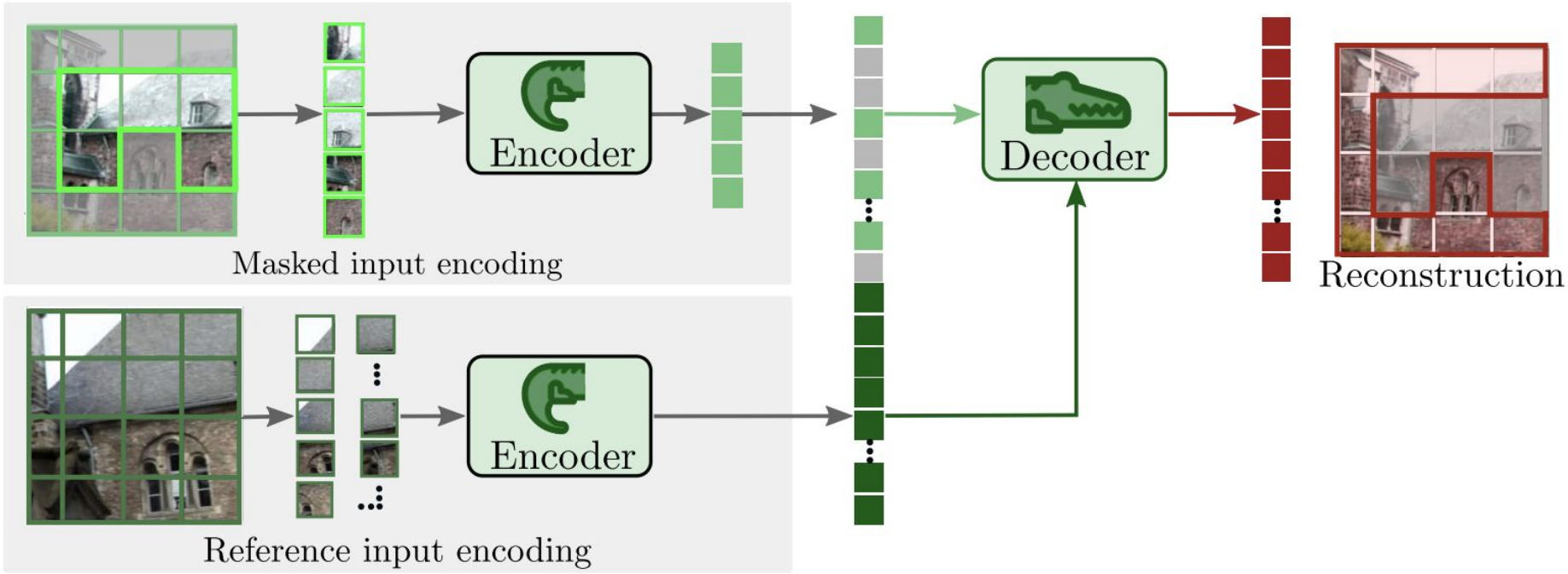
Problem with few images

Dense Unconstrained Stereo 3D Reconstruction (DUST3R)

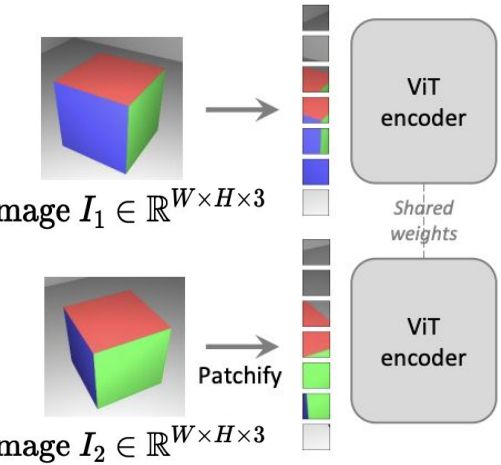
- unified foundational model for geometric 3D vision
- Direct RGB-to-3D Mapping
- Transformer-Based Architecture
- Network is trained in a fully supervised manner with large public datasets (8.5 M pairs)
- No prior knowledge about poses or intrinsics



DUST3R - pretraining

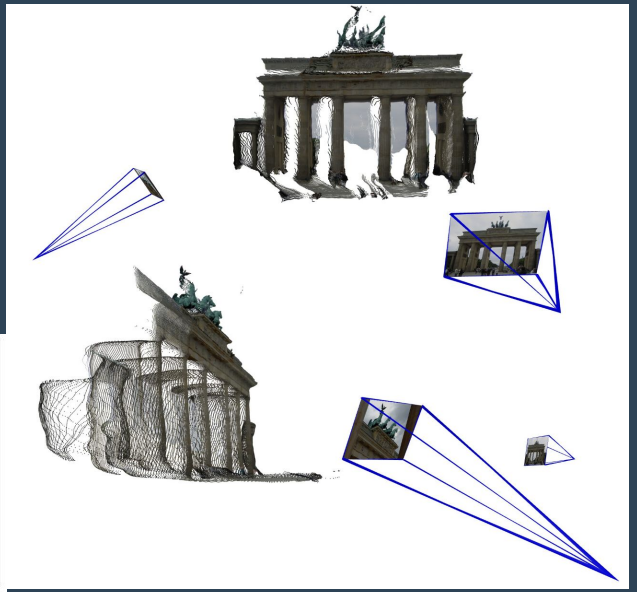


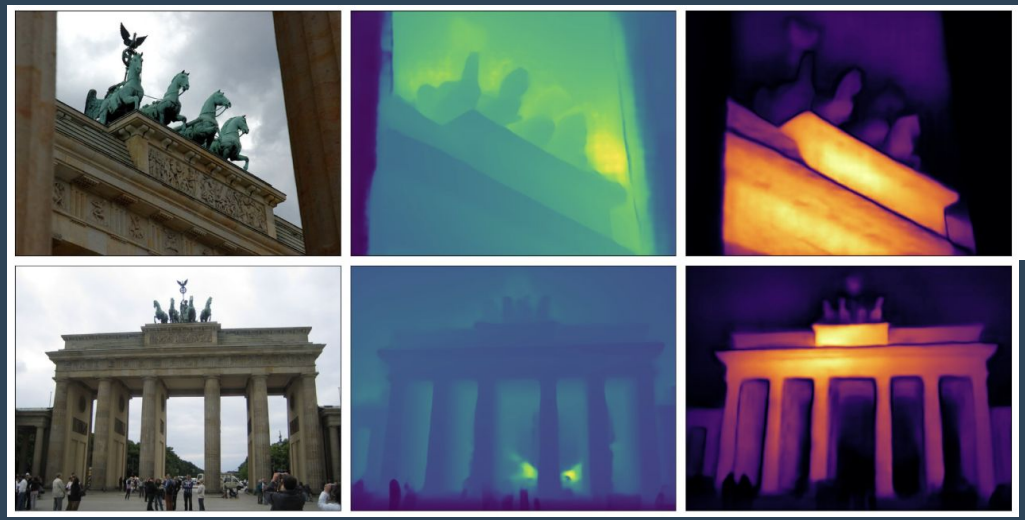
Dense Unconstrained Stereo 3D Reconstruction (DUST3R)



Dense Unconstrained Stereo 3D Reconstruction (DUST3R)

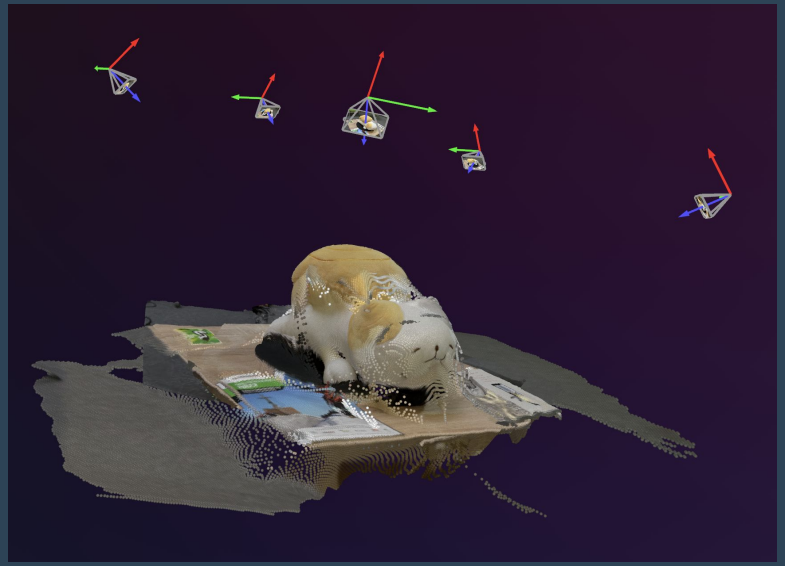
- Downstream applications
  - Camera calibration
  - Depth estimation
  - Pixel correspondences
  - Camera pose estimation
  - Dense 3D reconstruction

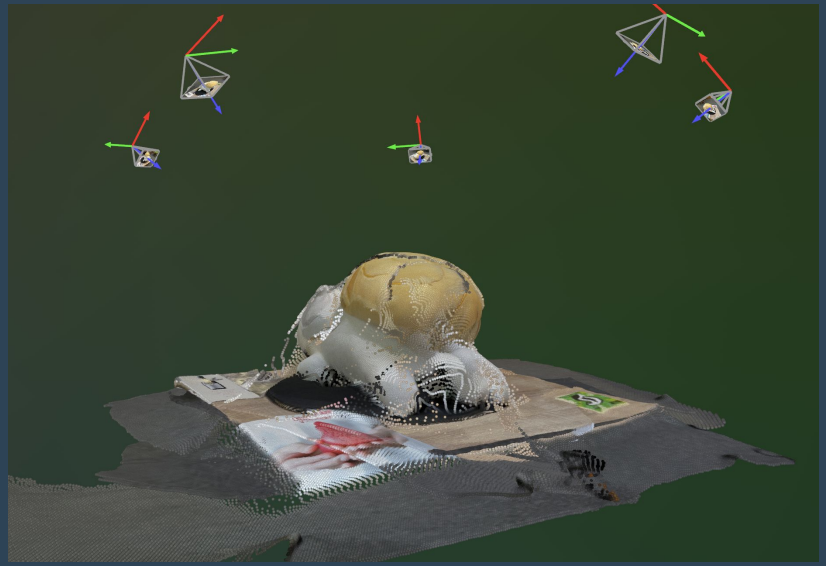




Dense Unconstrained Stereo 3D Reconstruction (DUST3R)

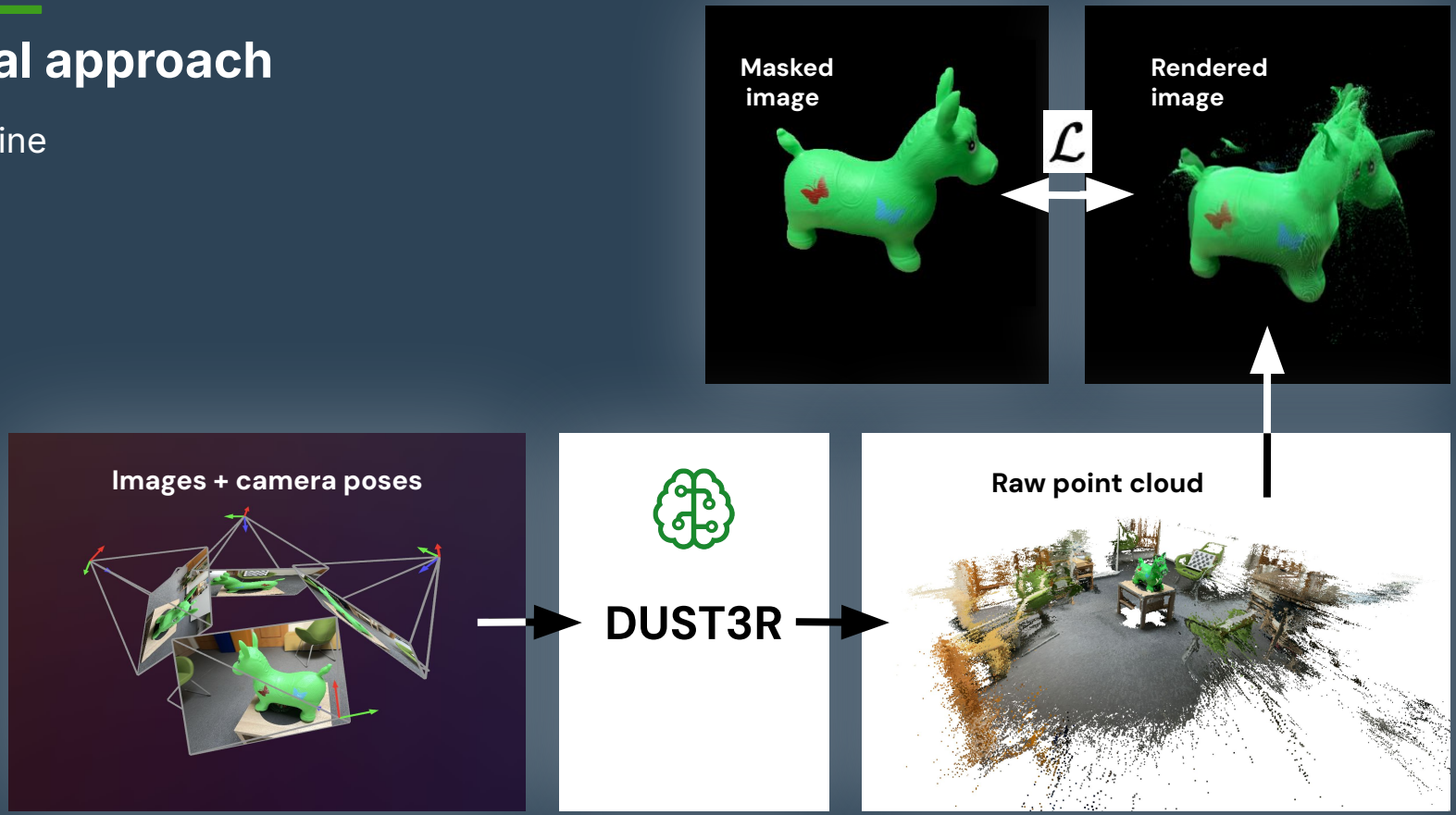
- Prediction on our 5 images of Bulbosaur (no prior about poses)





Final approach

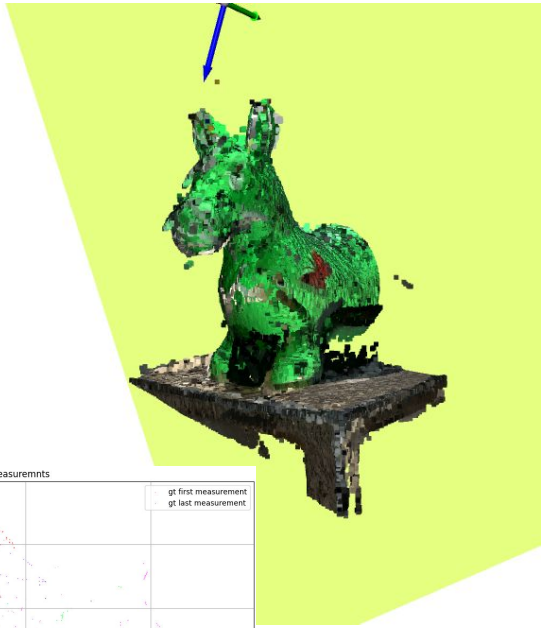
Pipeline

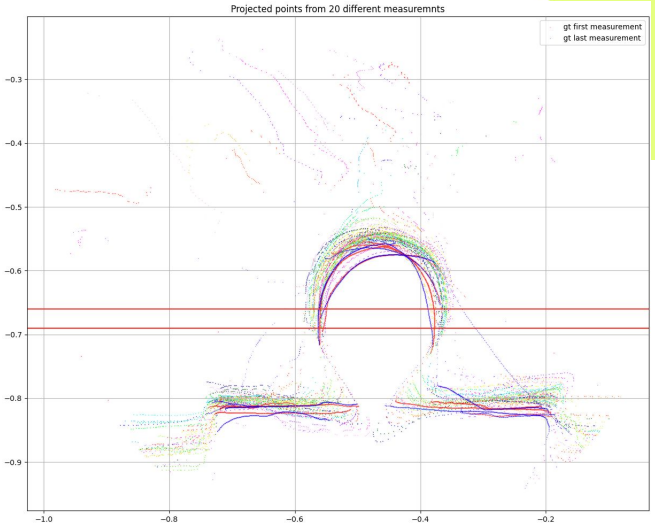


Final approach

Results

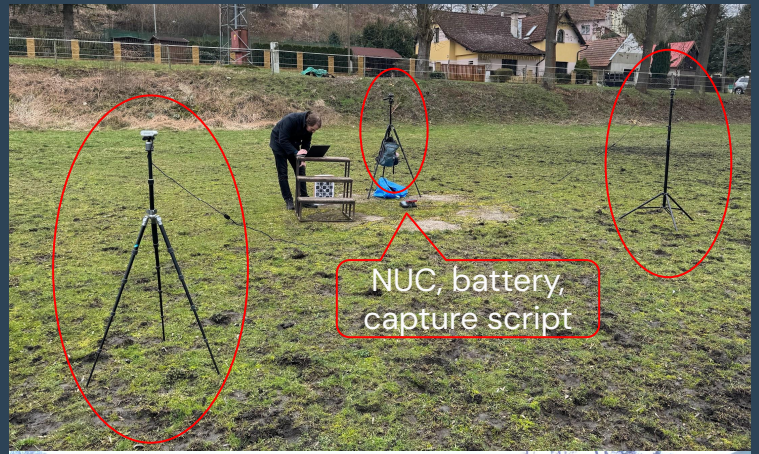
→We achieved error 0.7 cm





Horse 3D reconstruction experiment

Data collection setup



Encountered problems

- Depth quality issues
- Issues with camera pose calibration
- Different colours across cameras
- RGB camera quality and resolution
- Software/hardware stability problems

DATAMOLE

Summary

- Reasonable performance on the toy case
- Our naive ground truth experiment failed
- Currently on hold
- Future plans: find suitable ground truth experiment

DATAMOLE

Děkujeme za pozornost