ICLR

Cameras as Rays:
Pose Estimation via Ray Diffusion

Jason Y. Zhang* Amy Lin* Moneish Kumar

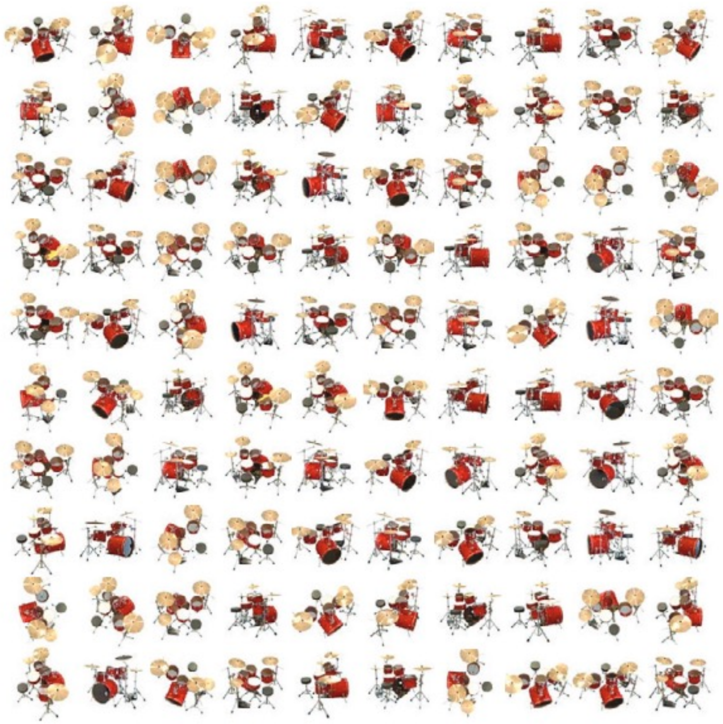
Tzu-Hsuan Yang Deva Ramanan Shubham Tulsiani

Carnegie Mellon University
Robotics Institute

(* denotes equal contribution)

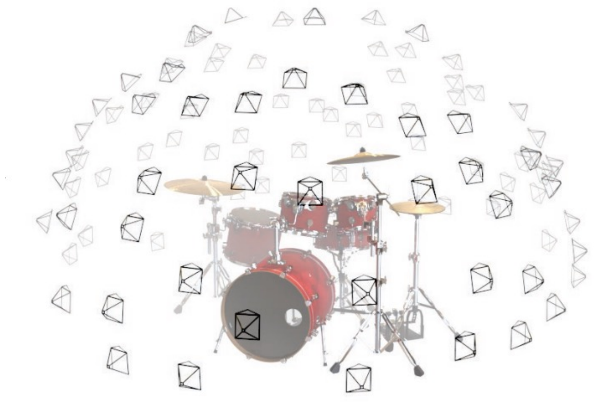
Lots of Progress in Novel-view Synthesis

Input Images



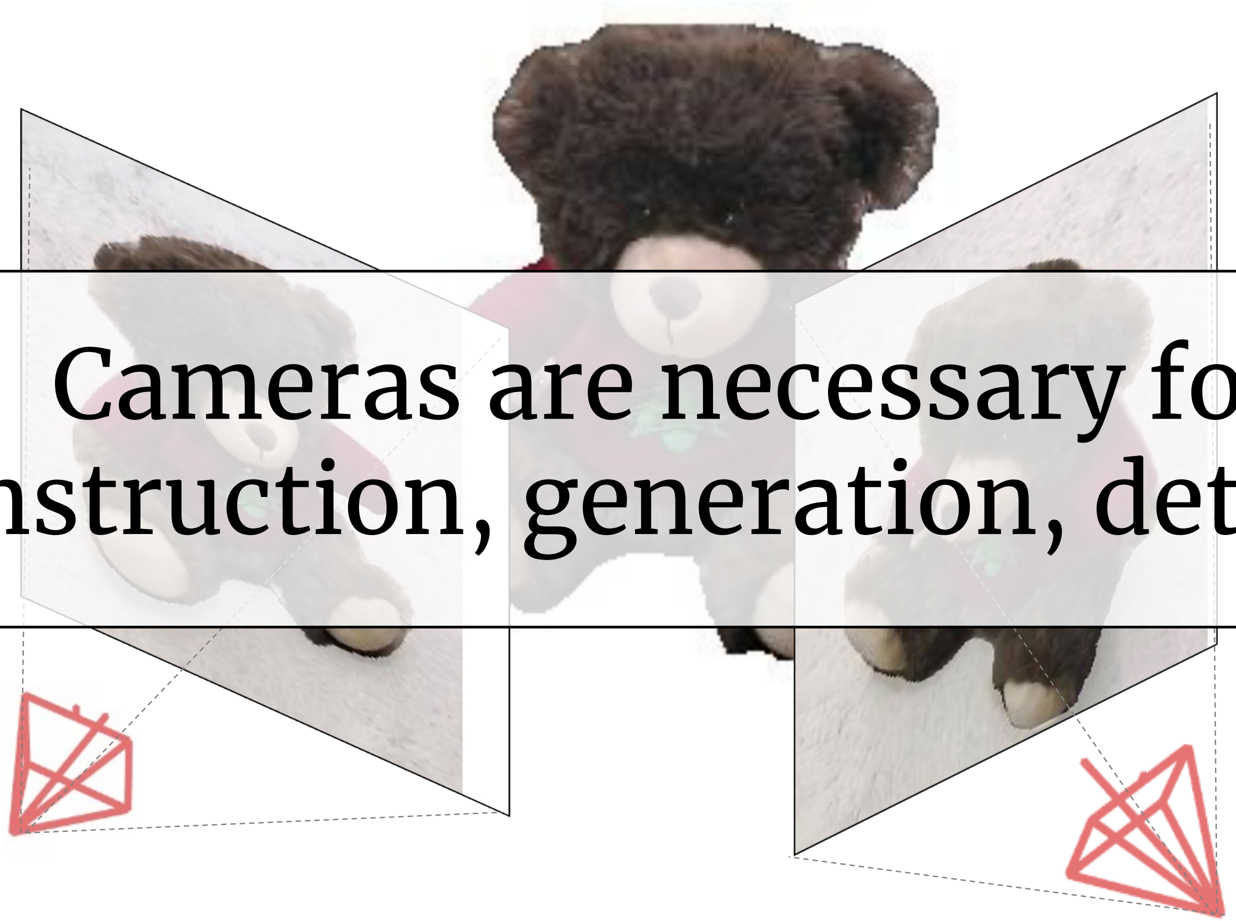
Rendered Views

Requirement:
Cameras



Mildenhall et al. *NeRF.* (ECCV 2020)

Recovering Cameras is a Pre-requisite for 3D Computer Vision



Cameras are necessary for:
3D reconstruction, generation, detection, etc.

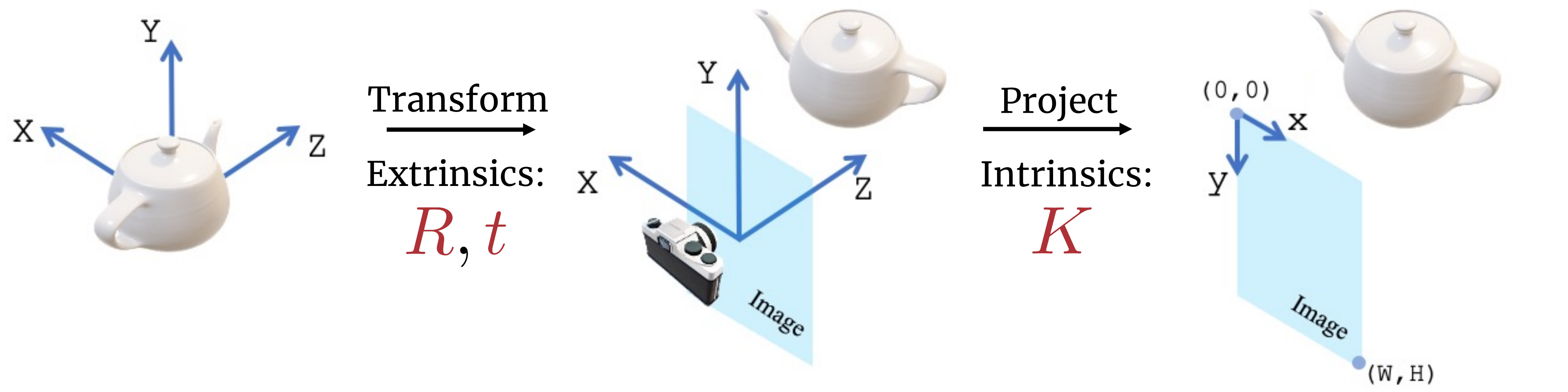
How are Cameras Typically Represented?

A camera describes how points in world coordinates project to pixel coordinates

World Coordinates

Camera Coordinates

Pixel Coordinates



Transform $\xrightarrow{\quad}$

Extrinsics:

R, t

Project $\xrightarrow{\quad}$

Intrinsics:

K

$$x$$

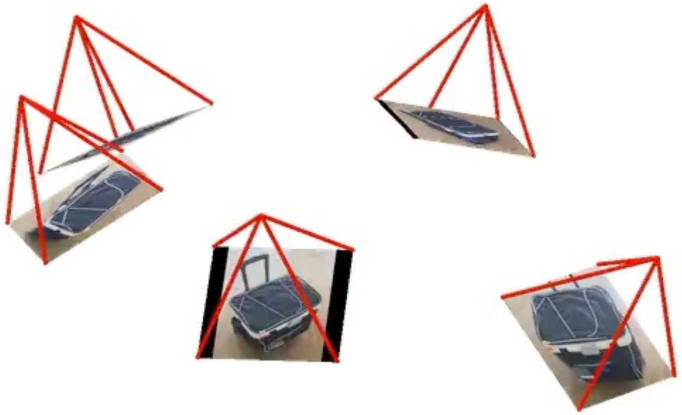
$$Rx + t$$

$$K(Rx + t)$$

Task: Sparse-view Camera Estimation

Input: Sparse Images (N≤8)

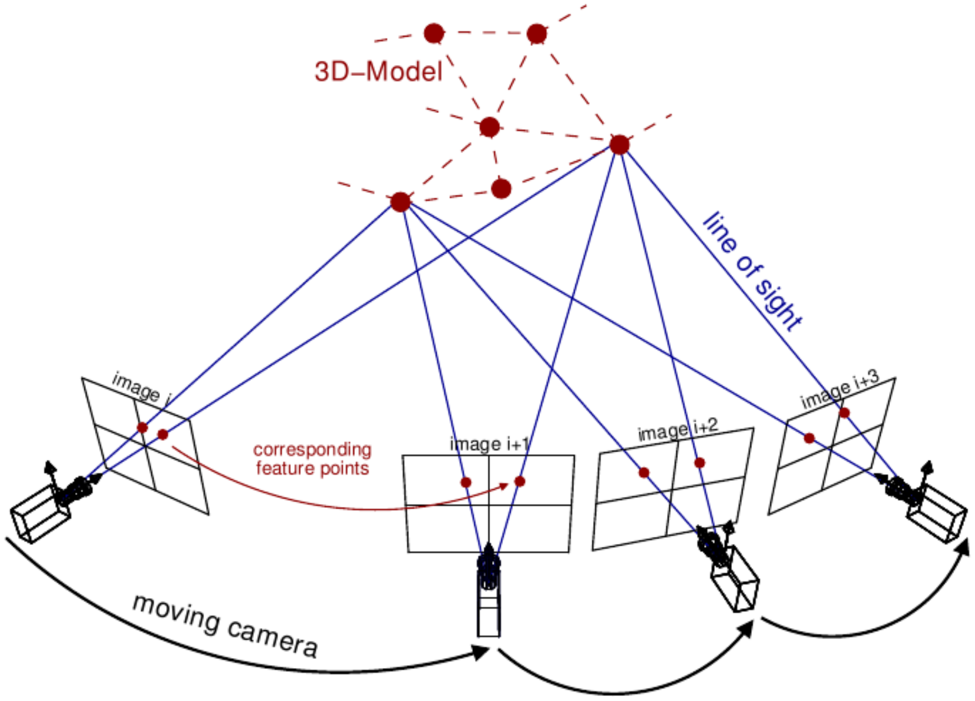
Output: Cameras



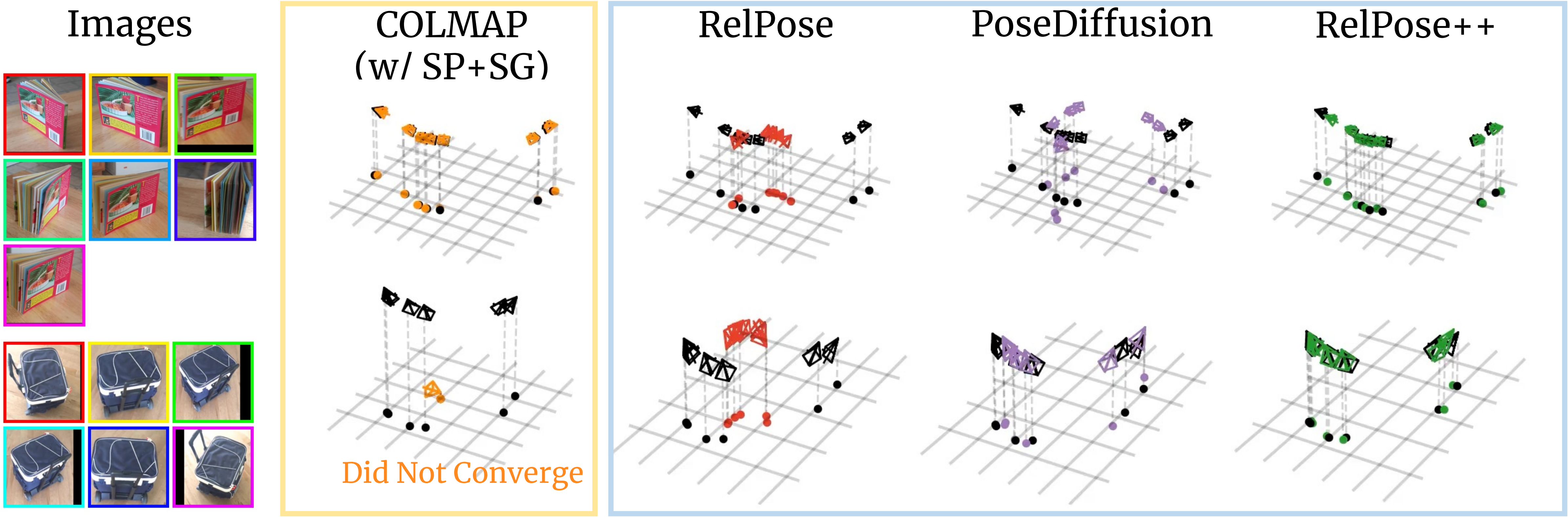
Structure-from-Motion:
Classical Pipeline for Recovering Cameras

Very challenging for sparse views!

Find point correspondences between images, triangulate them in 3D, solve for cameras parameters using Bundle Adjustment



Prior Work for Sparse-view Cameras

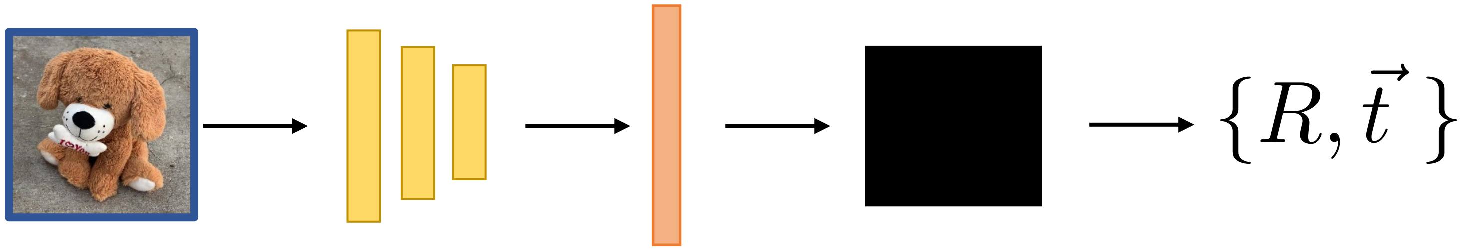


Classical Methods
+ Precise
- Lacks Robustness

Learning-based Methods
+ Robust
- Insufficient Precision

Schönberger et al. *COLMAP*. (CVPR 16, ECCV 16); Zhang et al. *RelPose*. (ECCV 22);
Wang et al. *PoseDiffusion*. (ICCV 23); Lin et al. *RelPose++*. (3DV 24)

Challenge: Global Features are a Bottleneck for Precision



Image

Encoder

(ResNet, DINO)

Feature

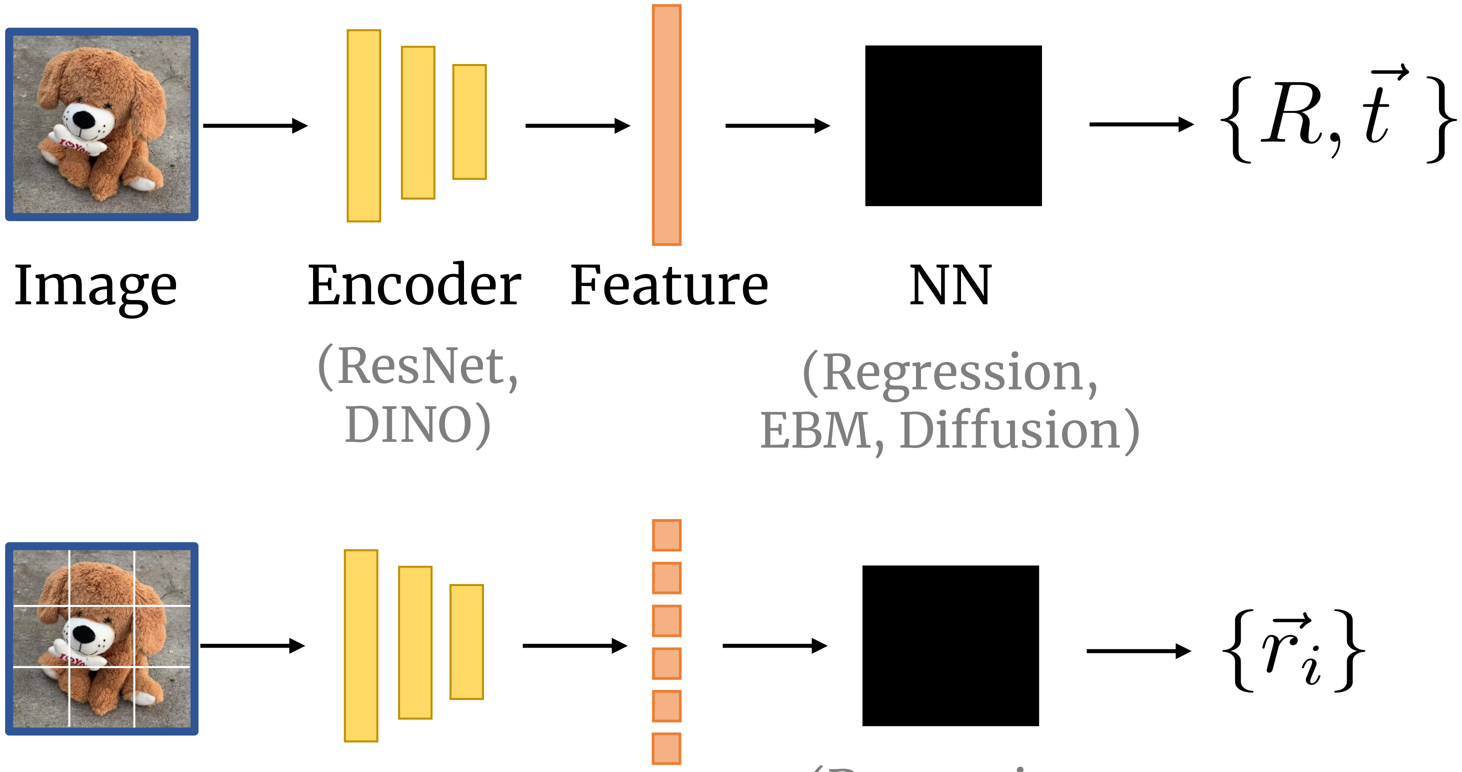
NN

(Regression, EBM, Diffusion)

$\{R, \vec{t}\}$

Cannot reason about low-level information (e.g., correspondences)

Challenge: Global Features are a Bottleneck for Precision



Image

Encoder

(ResNet, DINO)

Feature

NN

(Regression, EBM, Diffusion)

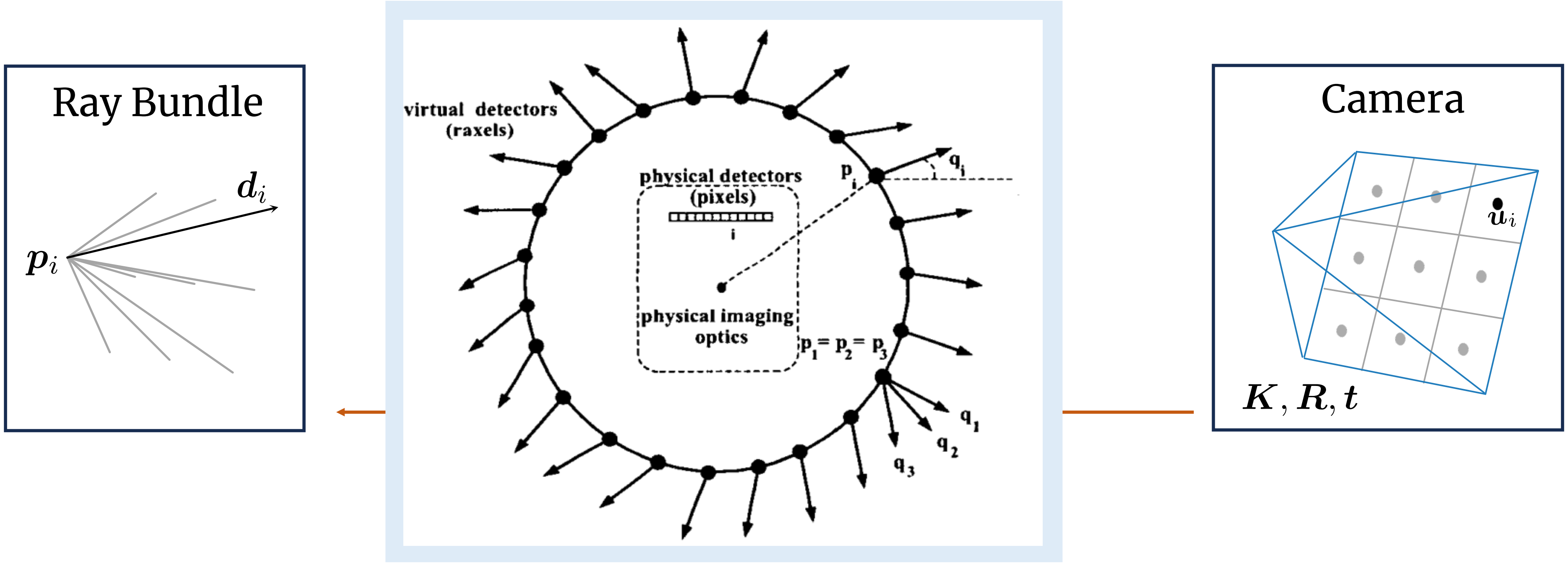
$\{R, \vec{t}\,\}$

Cannot reason about low-level information (e.g., correspondences)

$\{\vec{r}_i\}$

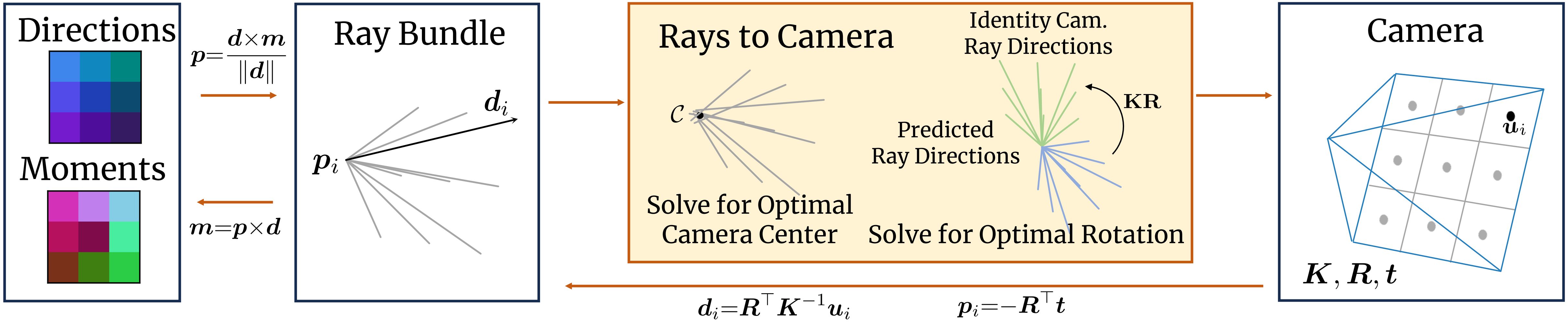
(Regression, Diffusion)

Representing Cameras via Ray Bundle



Grossberg & Nayar. *Raxel Imaging Model.* (IJCV 2005)

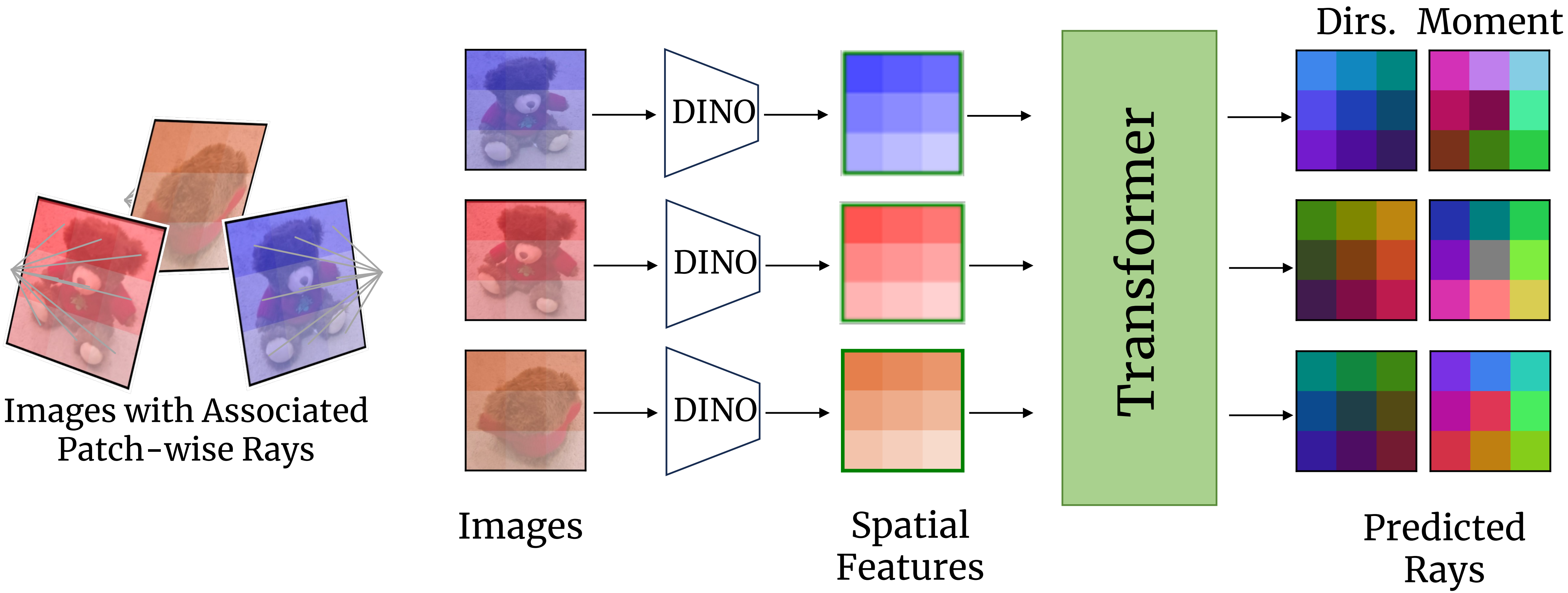
Representing Cameras via Ray Bundle


Directions
Moments
$p=\frac{\boldsymbol{d}\times\boldsymbol{m}}{\|\boldsymbol{d}\|}$
$\boldsymbol{m}=\boldsymbol{p}\times\boldsymbol{d}$
Ray Bundle
$\boldsymbol{d}_i$
$\boldsymbol{p}_i$
Rays to Camera
Identity Cam.
Ray Directions
$\mathcal{C}$
KR
Predicted
Ray Directions
Solve for Optimal
Camera Center
Solve for Optimal Rotation
Camera
$\boldsymbol{u}_i$
$\boldsymbol{K}, \boldsymbol{R}, \boldsymbol{t}$
$\boldsymbol{d}_i = \boldsymbol{R}^\top \boldsymbol{K}^{-1}\boldsymbol{u}_i$
$\boldsymbol{p}_i = -\boldsymbol{R}^\top \boldsymbol{t}$

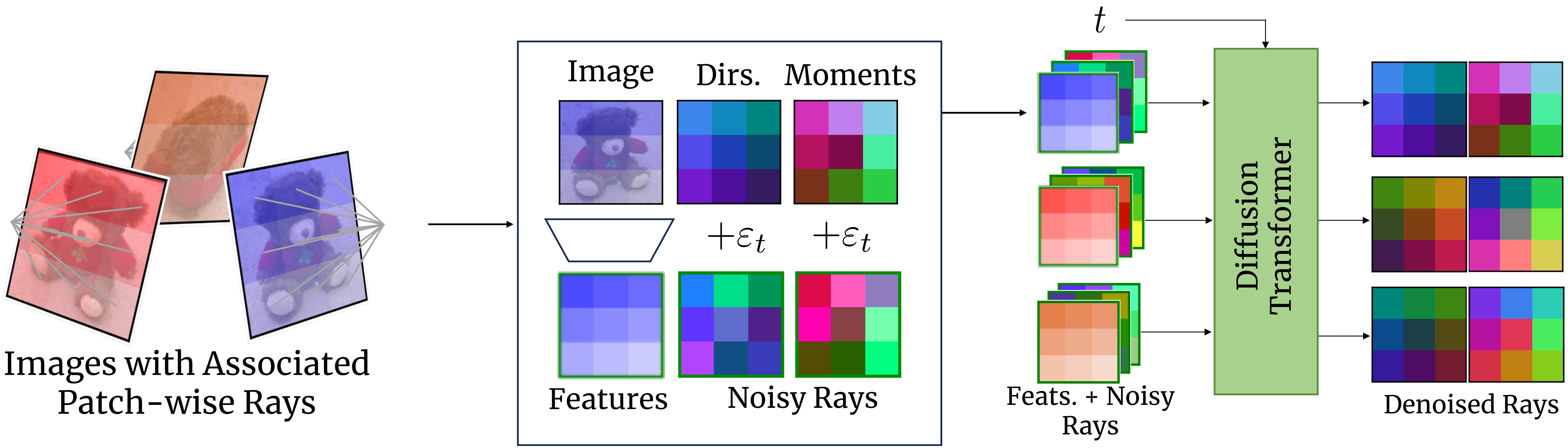
- Ray representation is distributed
- Ray representation is generic

Grossberg & Nayar. *Raxel Imaging Model.* (IJCV 2005)

Camera Estimation via Ray Regression



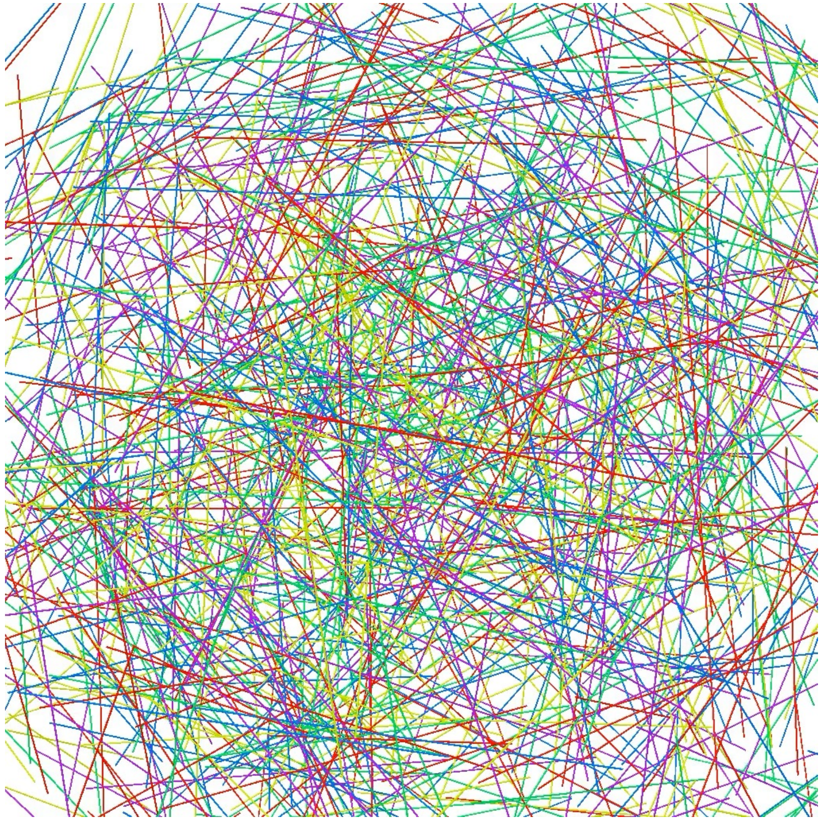
Camera Estimation via Ray Diffusion



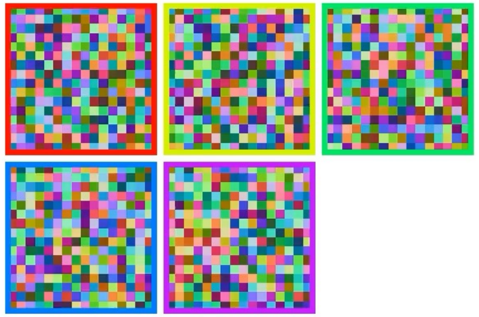
Backward Diffusion Process Visualization

Input Images

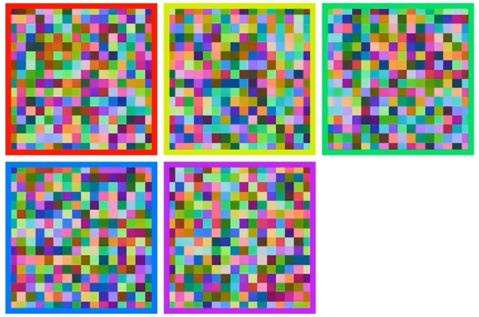
3D Rays



Directions



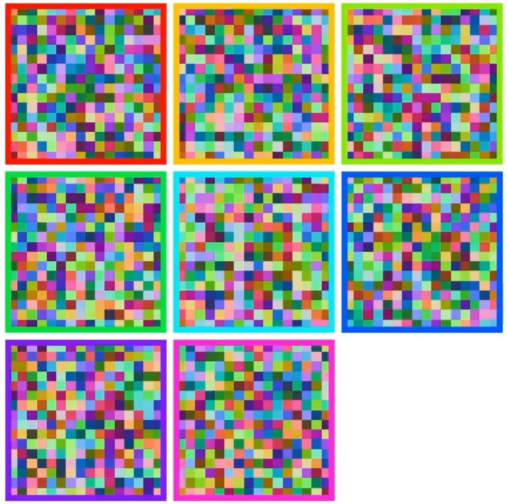
Moments



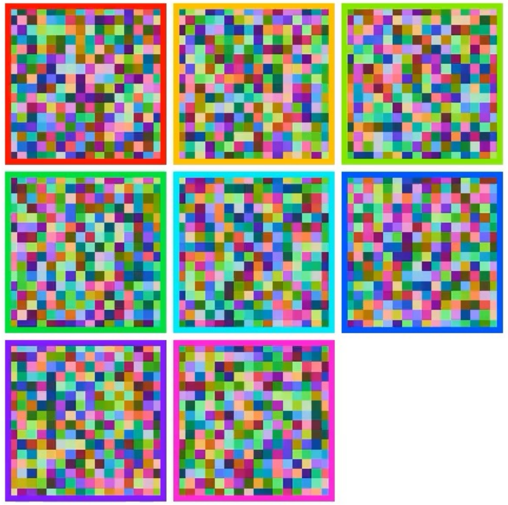
Backward Diffusion Process Visualization

Input Images

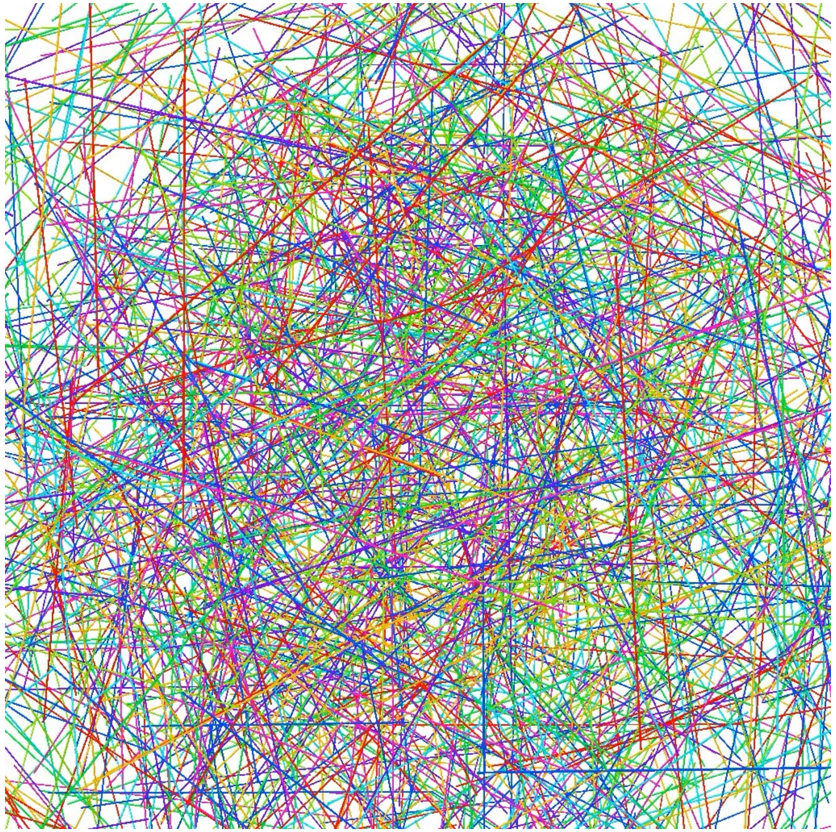
Directions



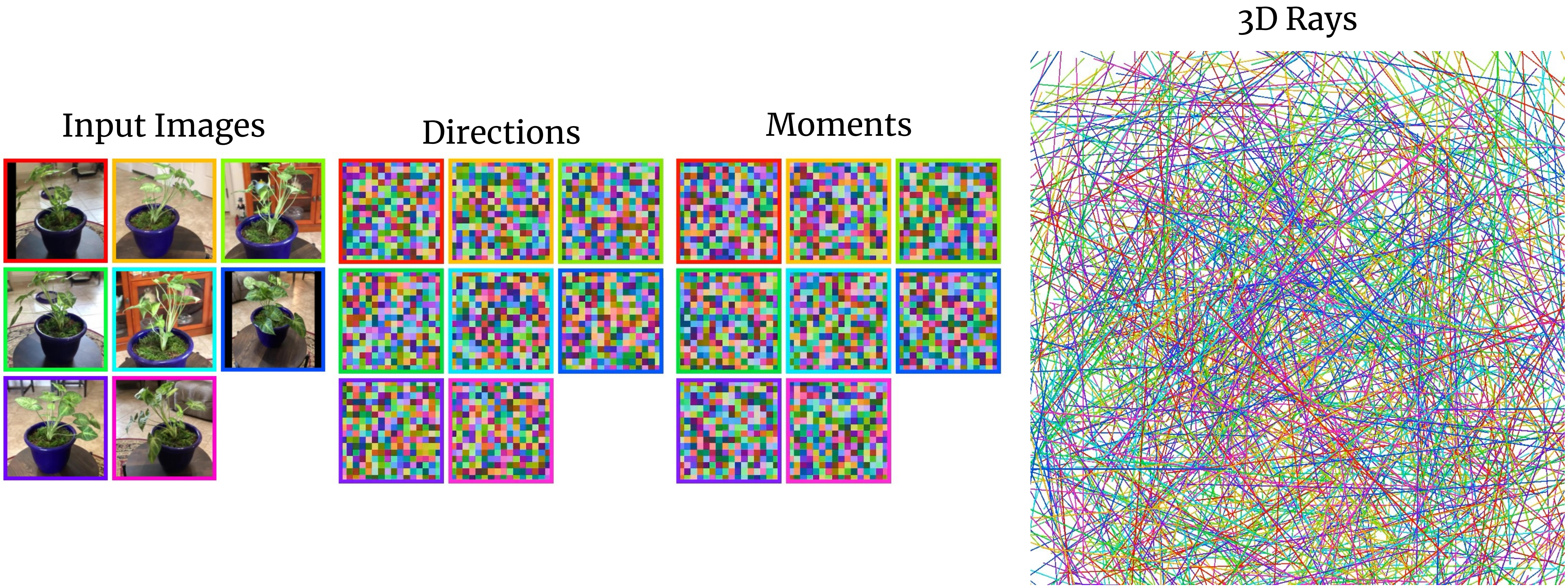
Moments



3D Rays



Backward Diffusion Process Visualization



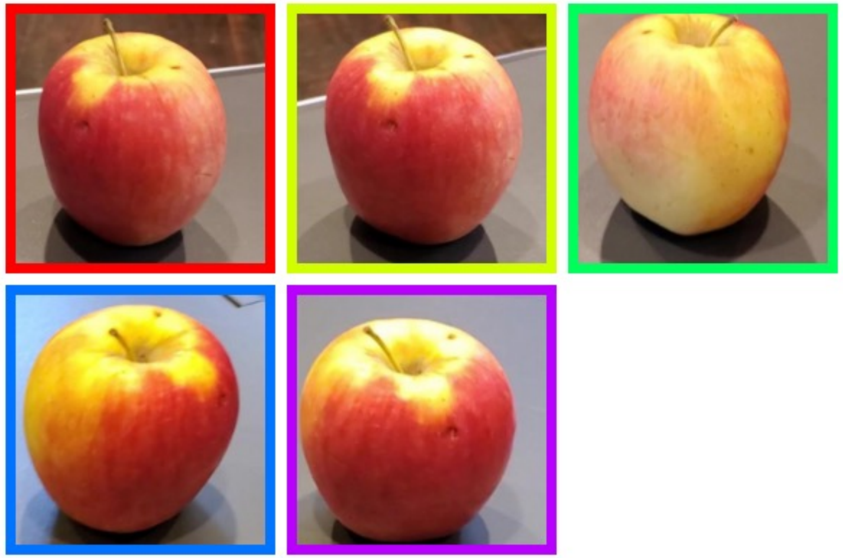
Input Images

Directions

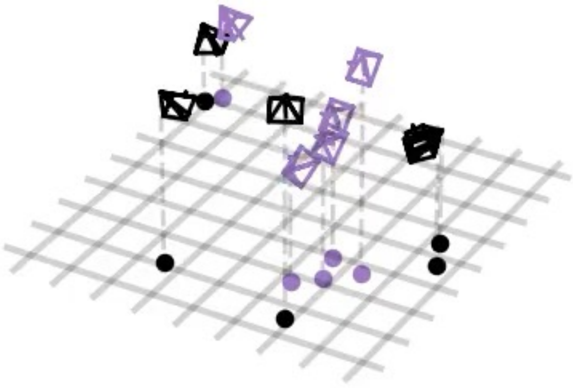
Moments

3D Rays

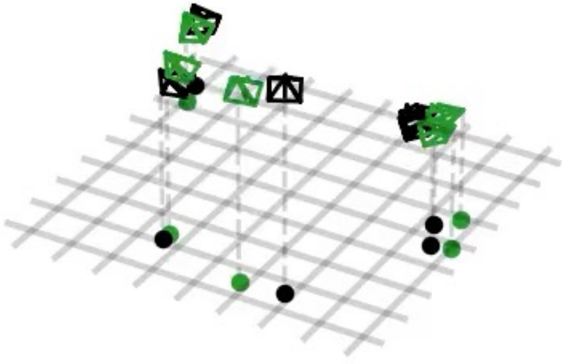
Qualitative Comparison



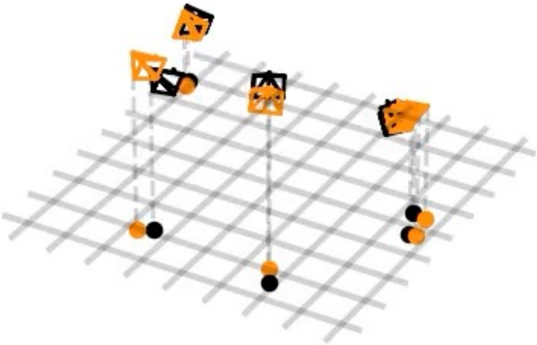
PoseDiffusion



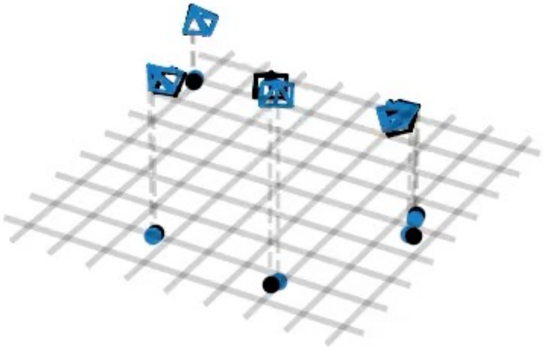
RelPose++



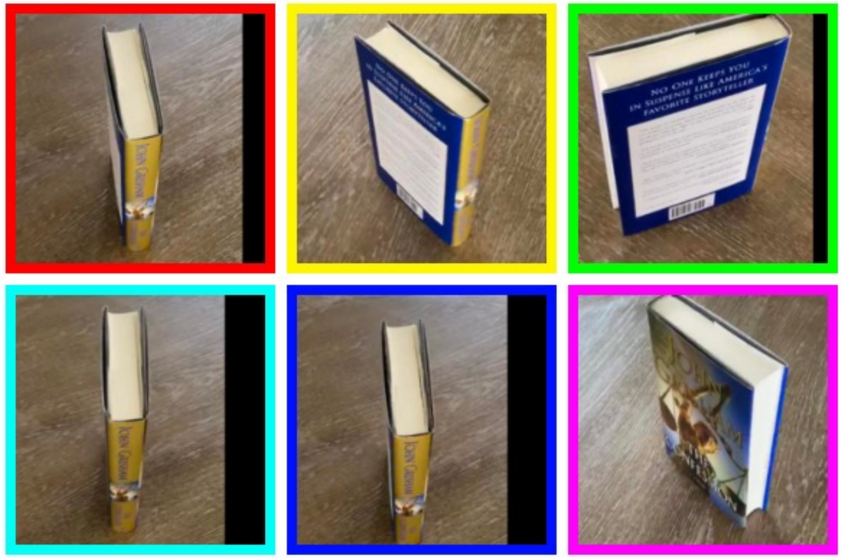
Ray Regression

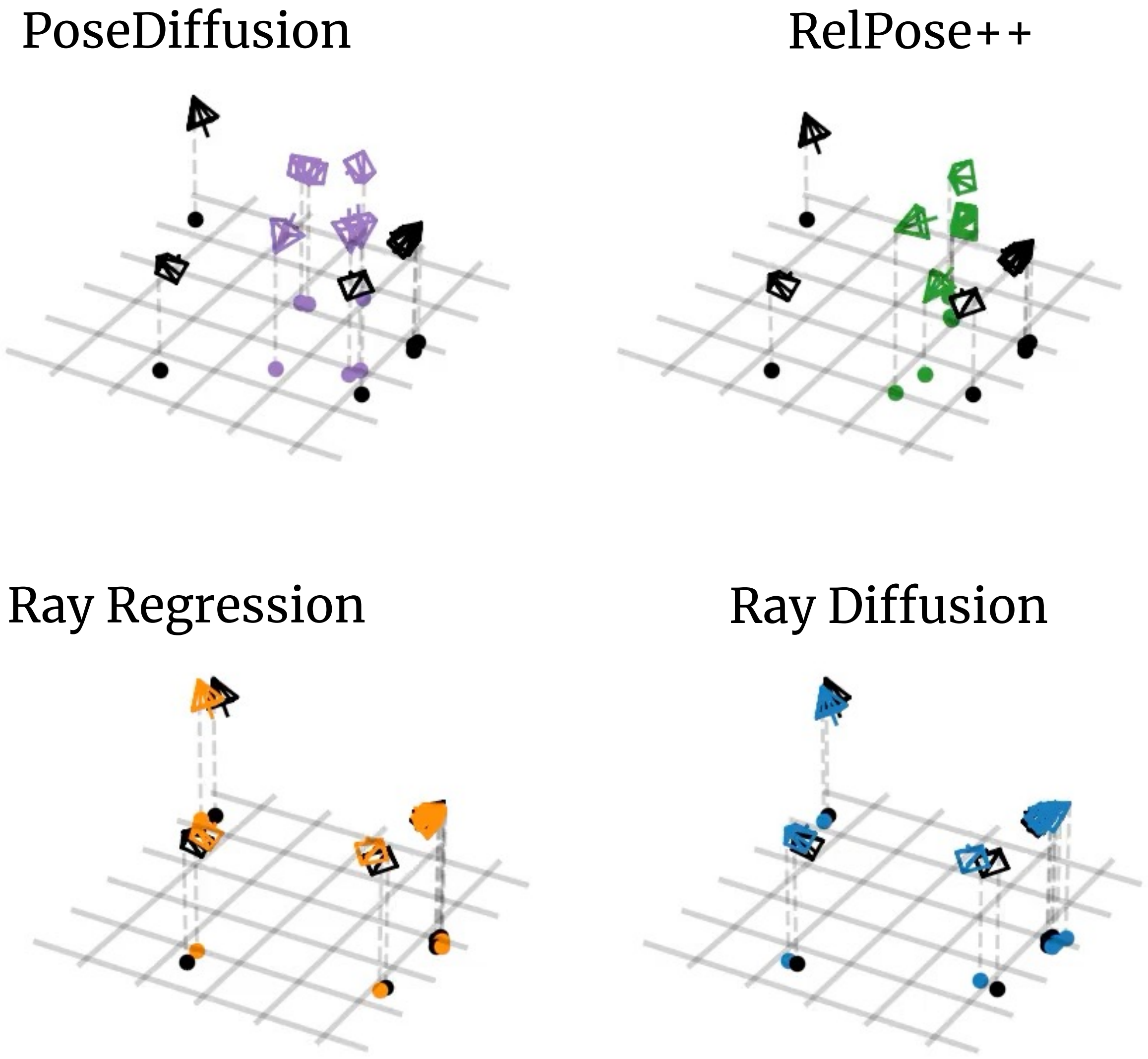


Ray Diffusion



Qualitative Comparison





PoseDiffusion

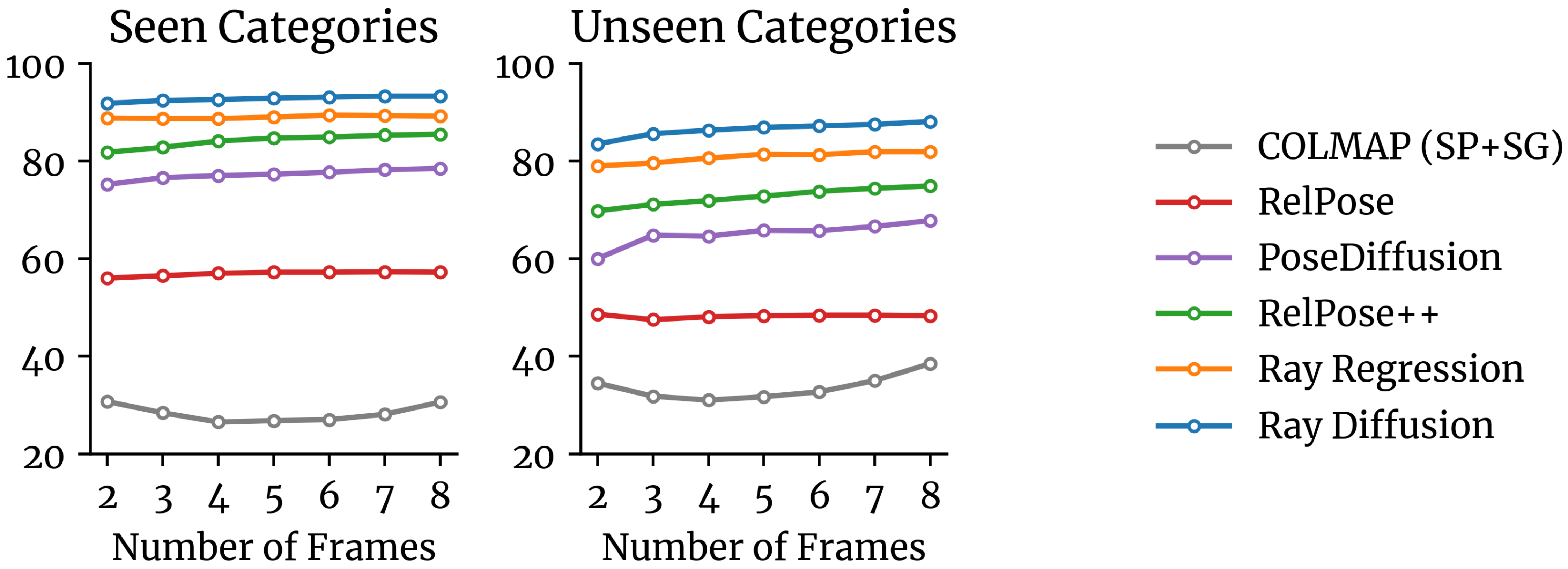
RelPose++

Ray Regression

Ray Diffusion

Quantitative Evaluation

Rotation Accuracy (% @ 15°)



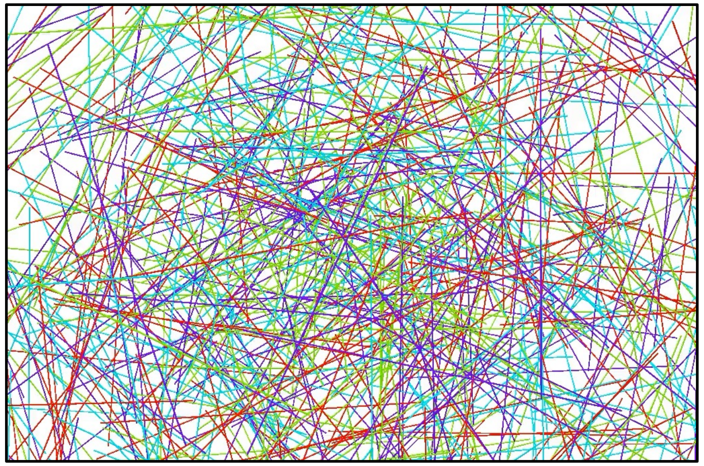
Takeaways

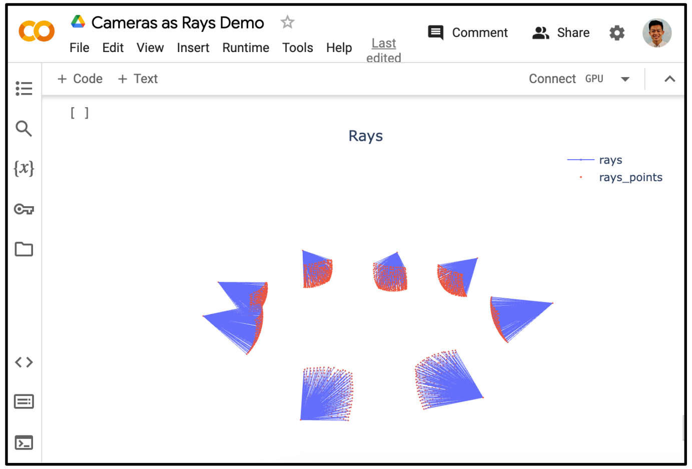
- We revisit the classical ray representation of cameras for learning-based pose estimation

- Present a diffusion-based model to predict the ray representation probabilistically

- **Future direction**: Train on all camera models jointly

Thank You for Listening!





Project Page (w/ Paper, Code, & More Results):

jasonyzhang.com/RayDiffusion

Join us at our poster: **Halle B #14**