Universität Stuttgart

Investigating Challenges in Generalizing Neural Radiance Fields with Learned Scene Priors

Master Thesis @ VISUS

Supervisor: Shohei Mori
Examiner: Dieter Schmalstieg

Jonas Geiselhart

What are Neural Radiance *F*ields ?

A Short Introduction

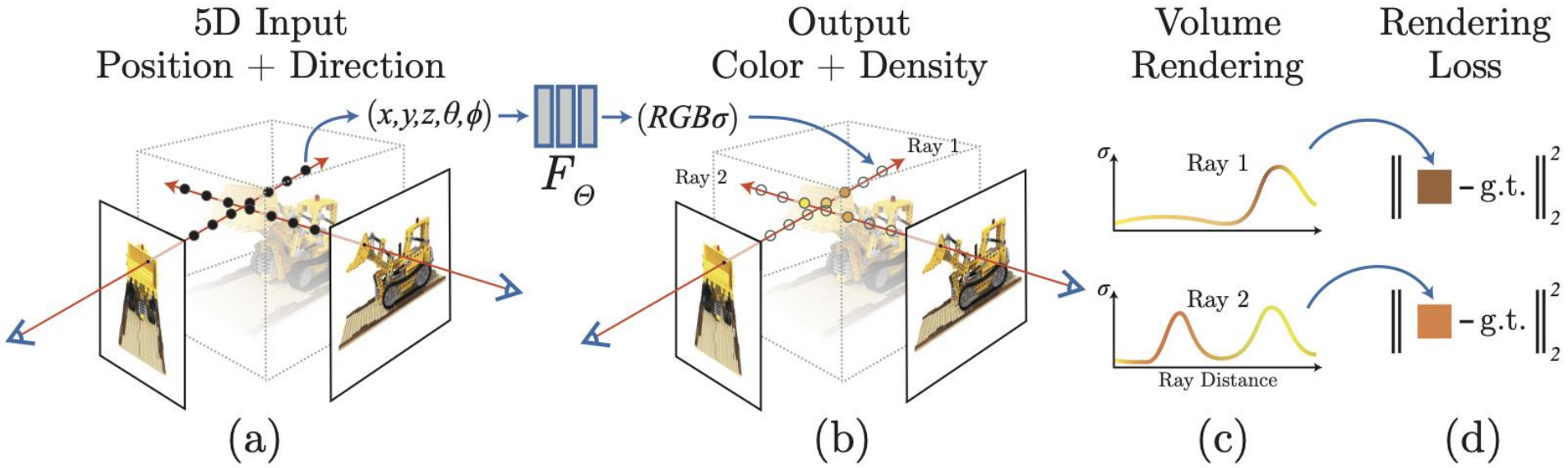


Image: Mildenhall et al. 2020

Problem

Motivation

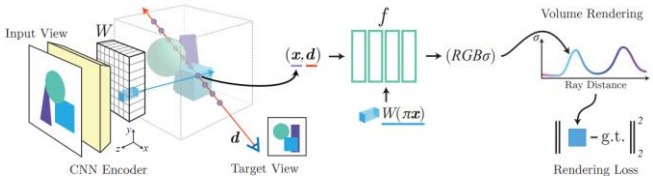
- Retraining takes a lot of time and computational resources (Minutes to Hours)
- One main objective in NeRF research:

Mitigate the per-scene retraining overhead to a tolarable amount !

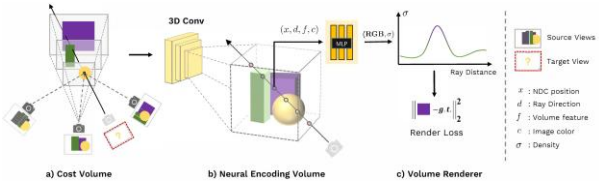
Problem

Related Works

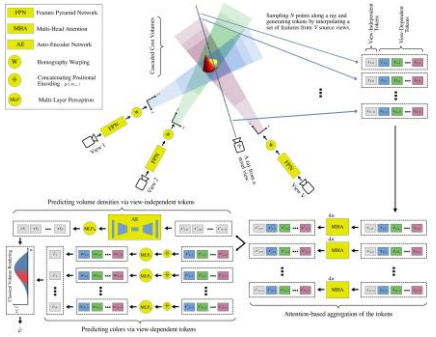
- How is this done in related works ?
  - Typically the priors get more fine grained and specific to the queried locations
  - Enforce consistency in priors through explicit algorithm, e.g.:
    - Contrastive Learning using point correspondences
    - Cost Volume Computation
    - Attentive prior selection



pixelNeRF (Yu et al. 2021). Pixelwise Prior



MVSNeRF (Chen et al. 2021) Neural Encoding Volumes



GeoNeRF (Johari et al. 2022). Attentive Priors

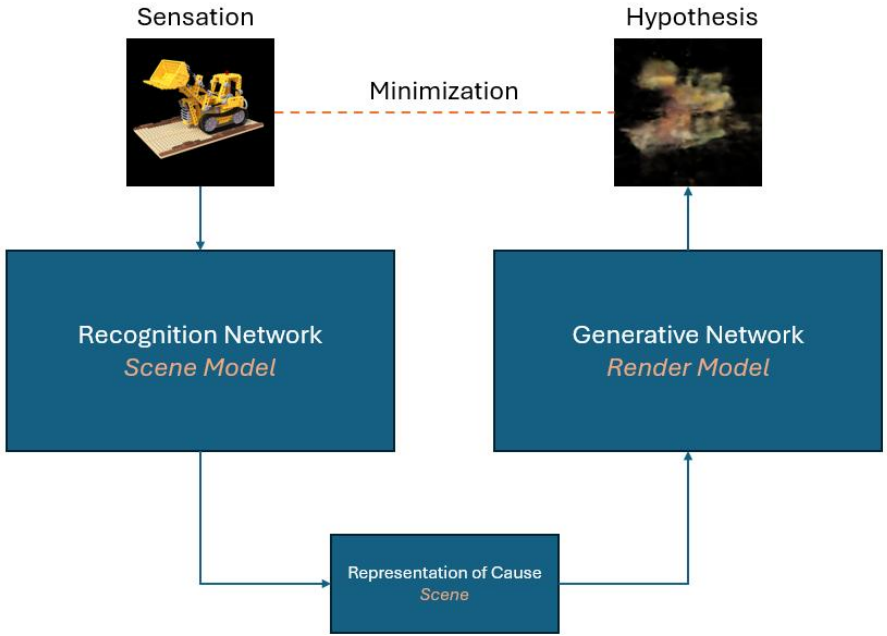
Proposed Approach

- Expore what happens when we focus on capturing the whole scene instead of a fine grained position- / ray-prior

- Why is this useful?
  - Only compute one prior in total → Eliminate computational overhead during Inference
  - Create a holistic prior → Benefit from non-local structures
  - Create a deep scene capture → Descriptive scene embedding for further neural processing

- **Relaxation:** *Learn, distinguish and render several pretrained scenes*
  - Enable gradual development towards generalizing network

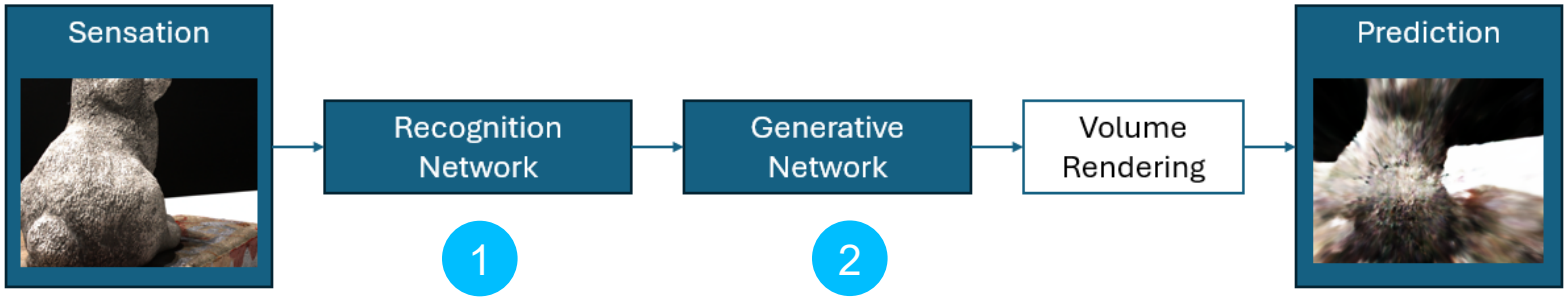
Theory
Free Energy Principle



- By Karl Friston *(The free-energy principle: a unified brain theory?, 2010)*
- Two complementary networks:
  - *Recognition Network* – reacts to sensation and represents the cause.
  - *Generative Network* – generates hypothesis from cause
- Move from learning geometries to perceiving geometries

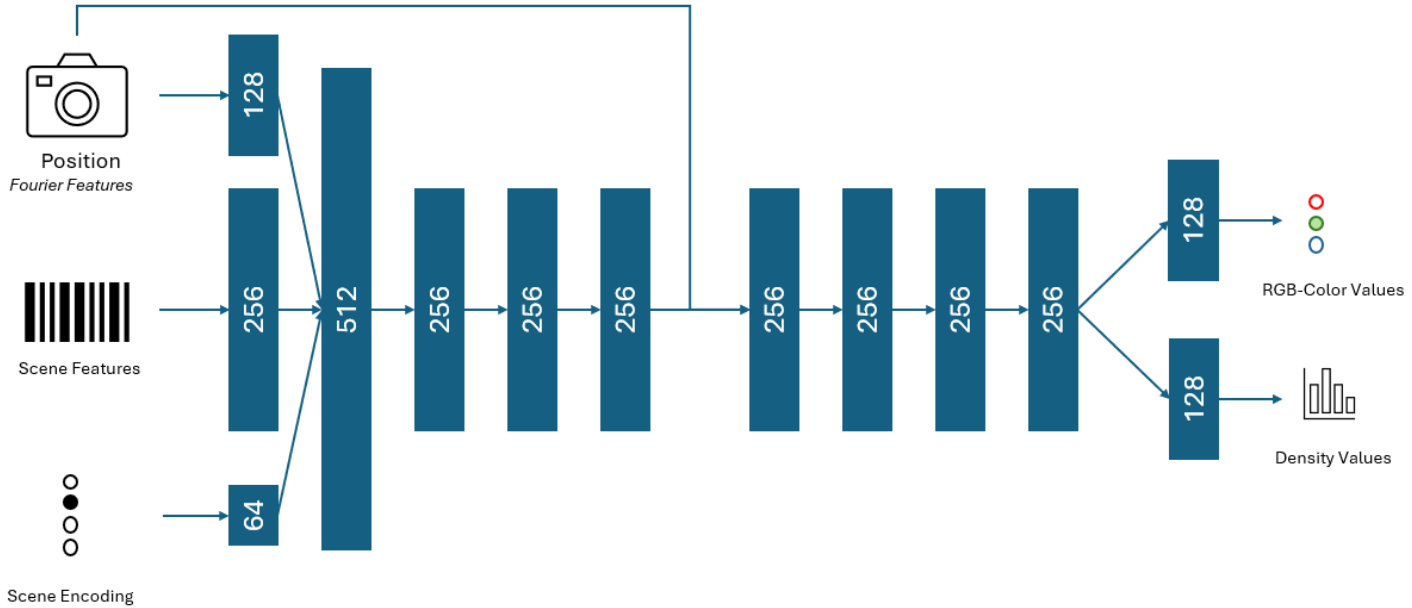
Framework

- Two networks:
    1. Receives images as sensation outputs a latent feature representation
    2. Receives latent feature representation outputs position prediction



Training & Architectures (1)
Generative Network

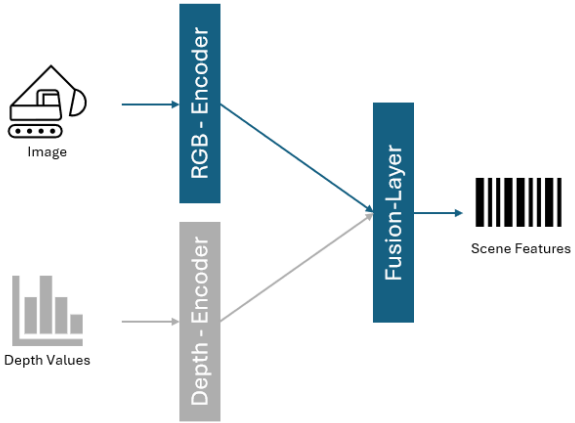
- One Version:

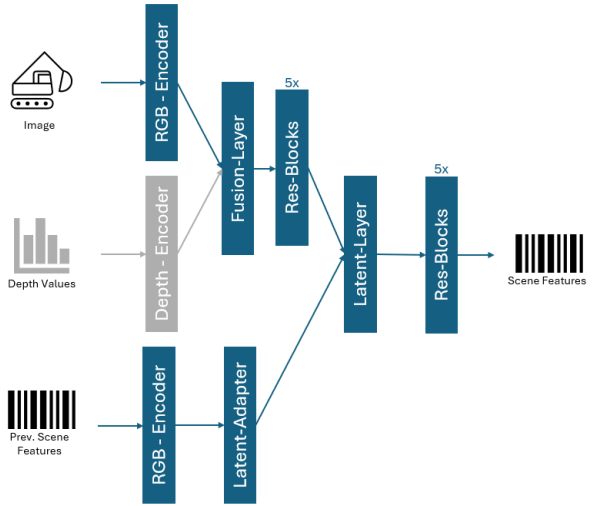


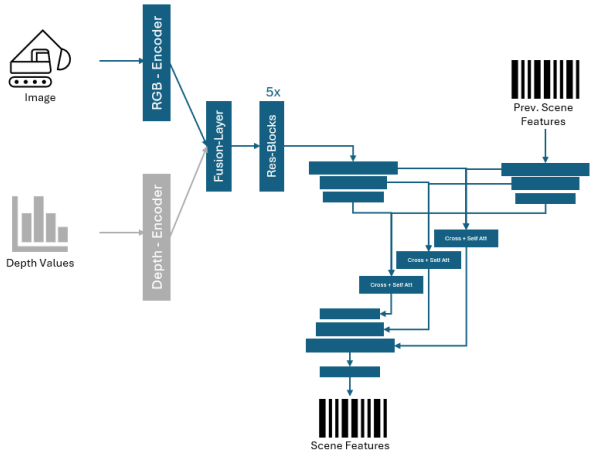
Training & Architectures (2)

Recognition Models

- Three variations to try:







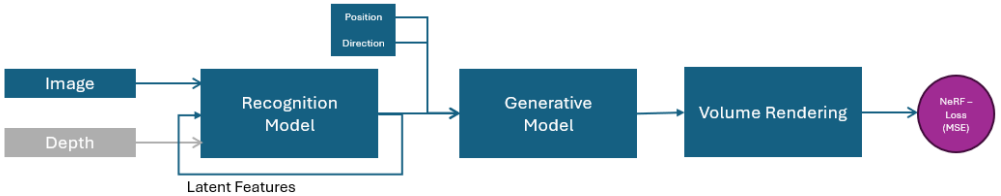
Training & Architectures (3)

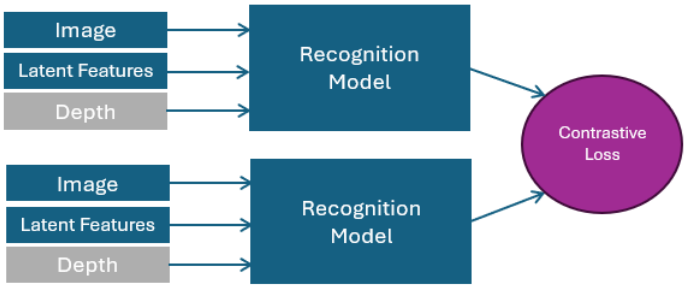
Training Procedure

LOSSES

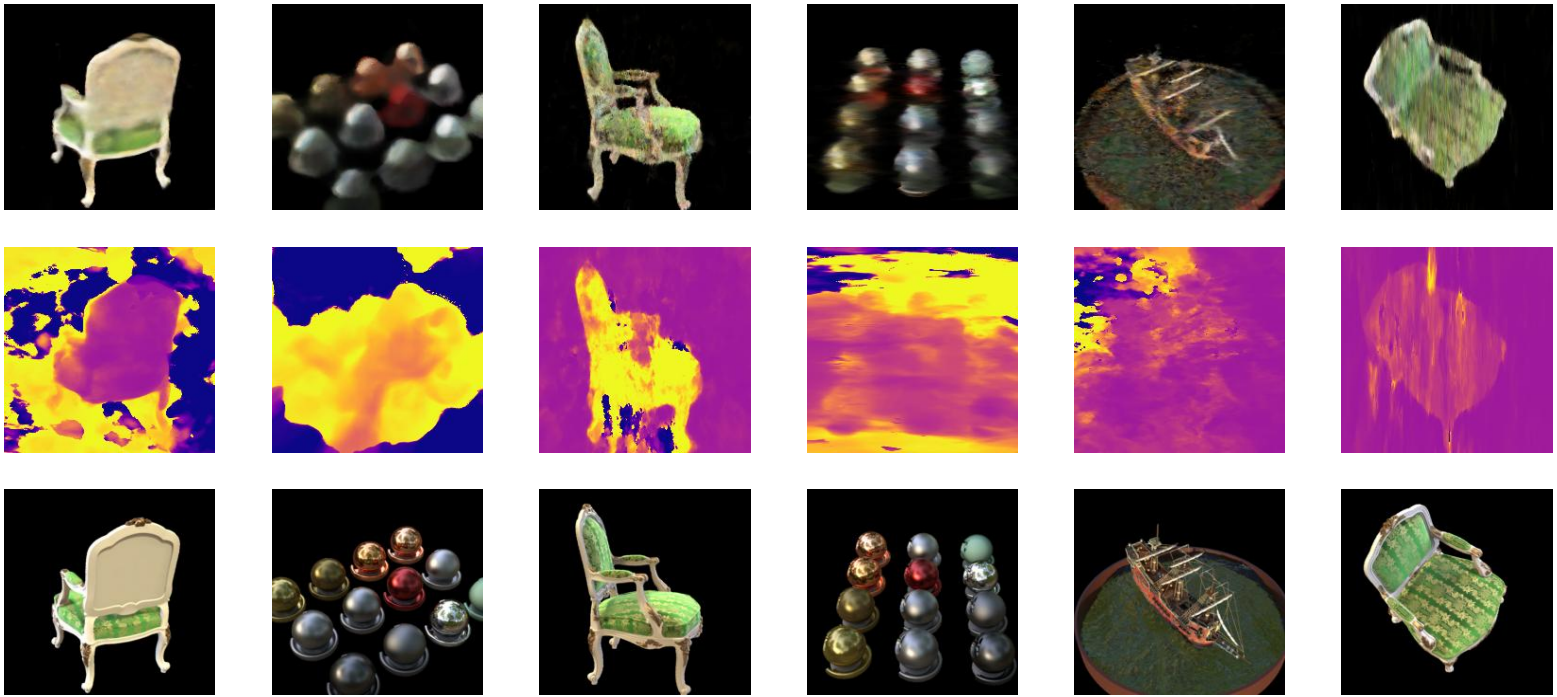
- RGB-MSE Loss (Standard)
- Depth Loss
  - Depends on Dataset
- Occlusion Loss
  - Penalize near camera floaters, no difference
- Consistency Loss
  - Very Expensive
- Contrastive Loss
  - See other side

TRAINING

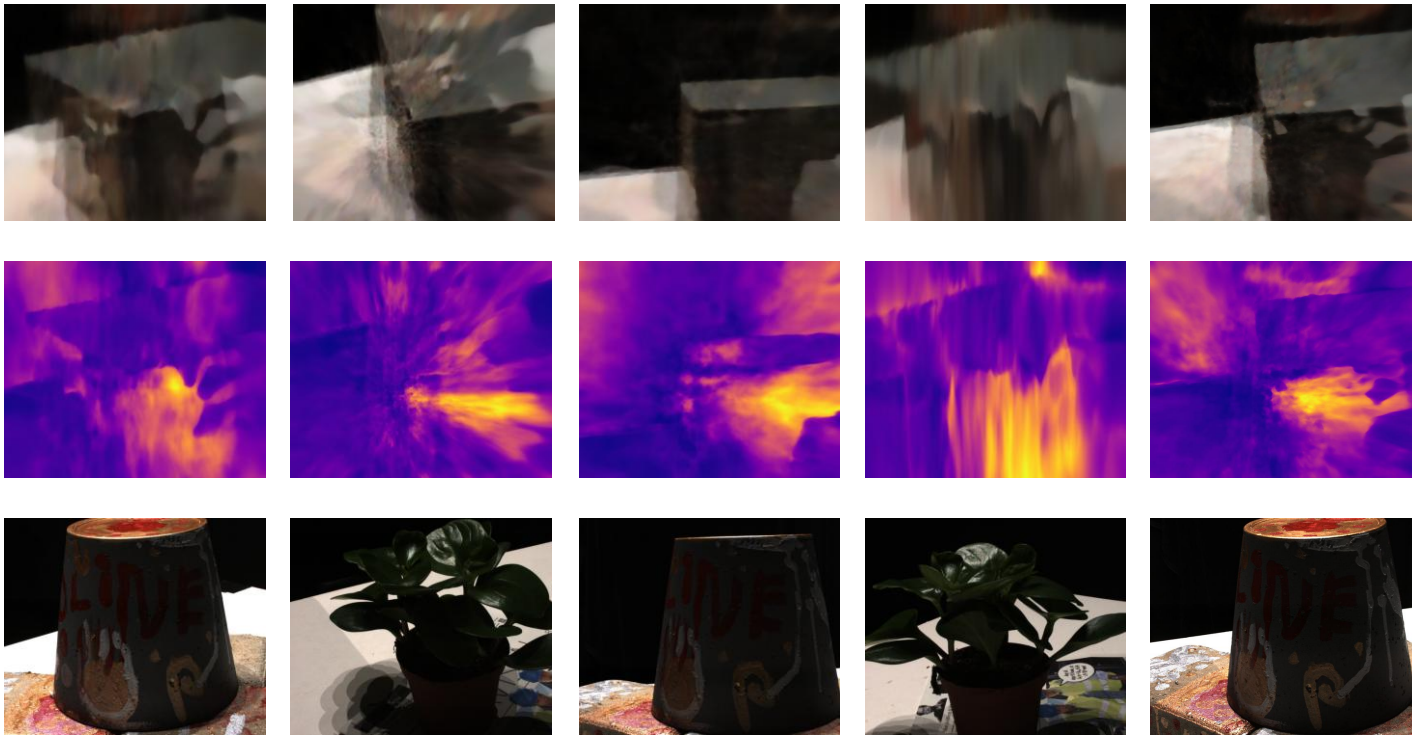




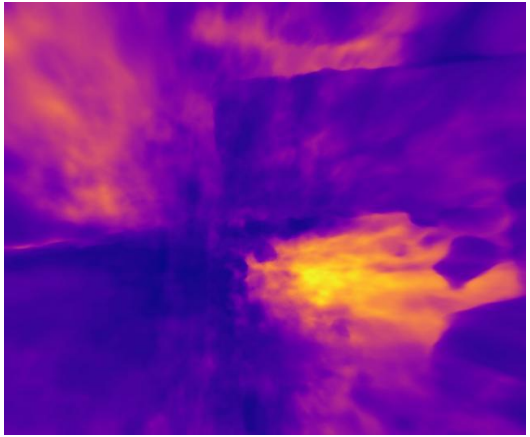
Challenges (1)
Distinguishability



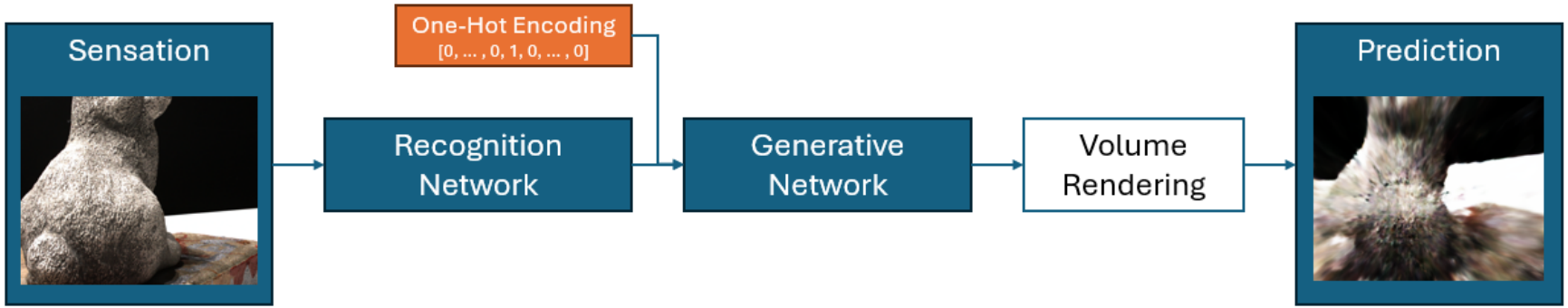
Challenges (1)
Distinguishability



Challenges (1)
Distinguishability

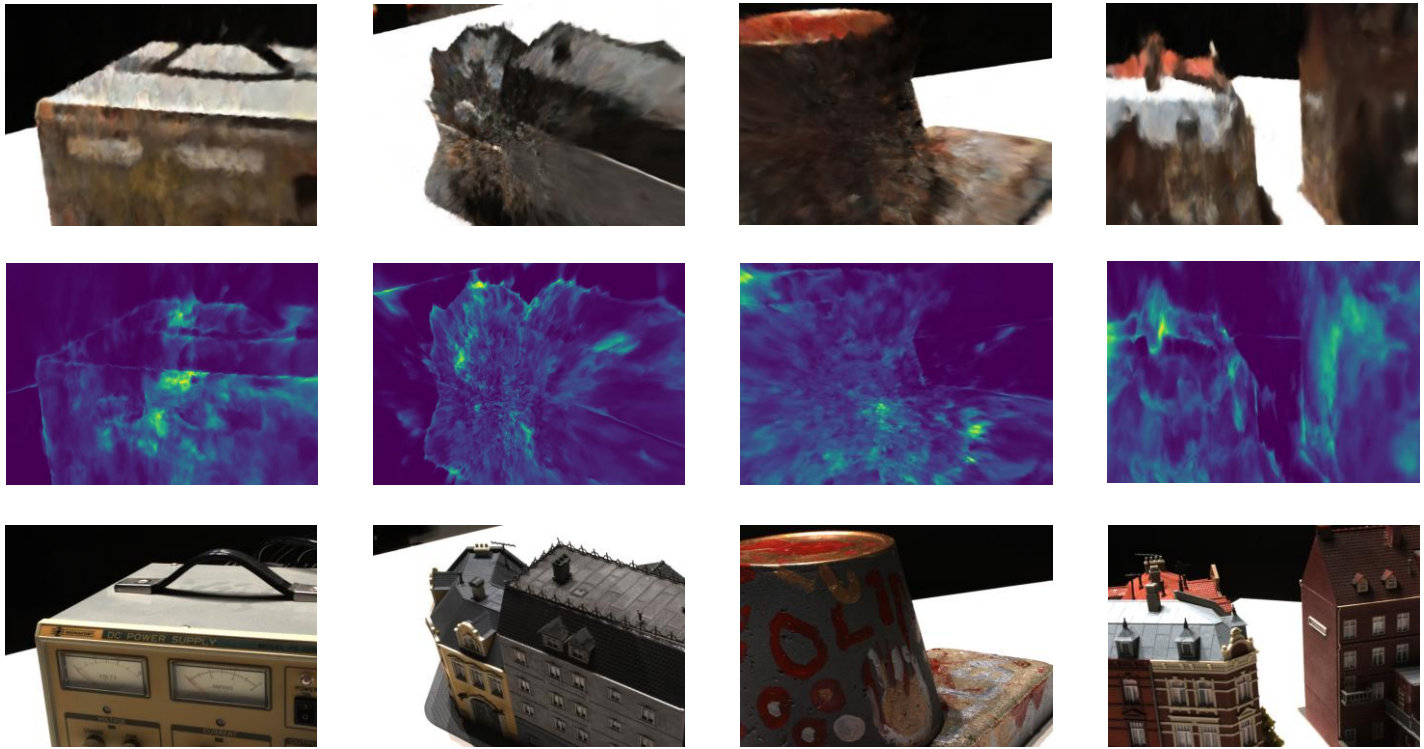


Challenges (1)
Distinguishability



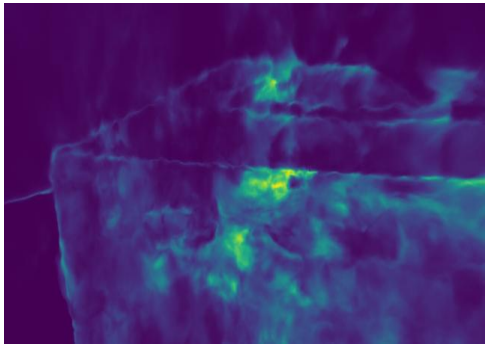
Challenges (1)

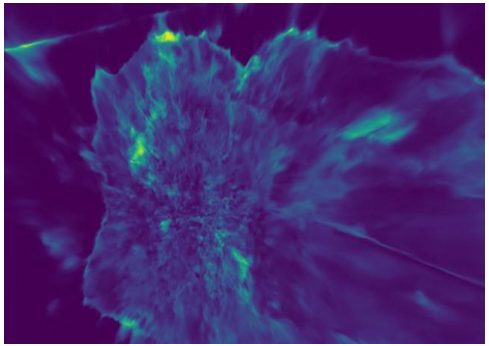
Distinguishability – Solution: One-Hot-Encoding



Challenges (2)
Overfitting

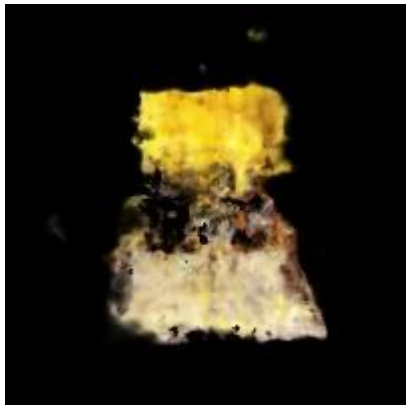
- Problem: *Images are overfitted to training views*
- The depth information is not extracted as good as in the one scene setting.
- Mitigation?
  - Smaller Networks, more regularization, dataset splitting …





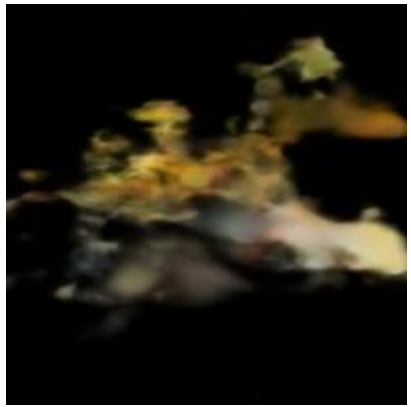
Challenges (2)
Overfitting

- Problem: *Images are overfitted to training views*



← Image close to training view

Image not so close →



Other Problems

Adaptions to get this to work

- NeRF learning becomes more complicated
  - Caveat: optimizations from the standard NeRF have to be reevaluated
- Learning is sensitive to:
  - Learning Rate
  - LR-Scheduling
  - Higher batching sizes have to be implemented
  - Dropout must be introduced
- The dataset must be split between reference views and training views

Results & Future Work

- Multi-Scene NeRFs are really resource intensive and instable to train
- A NeRF can be trained robustly on several scenes and recall them separately
- The main challenges of a generalized holistic prior are:
  - Distiniguishability
    - It is possible to distinguish scenes, but it is hard to do so without explicit data.
    - The One-Hot-Encoding did not build up implicit recognition.
  - Overfitting
    - The complicated task makes it harder to extract the right geometries.
    - Problem is in the Generative Model, a better regularization mechanism has to be implemented