Uncertainty-Guided 3D Reconstruction via Multi-Scale Gaussian Mixture Models

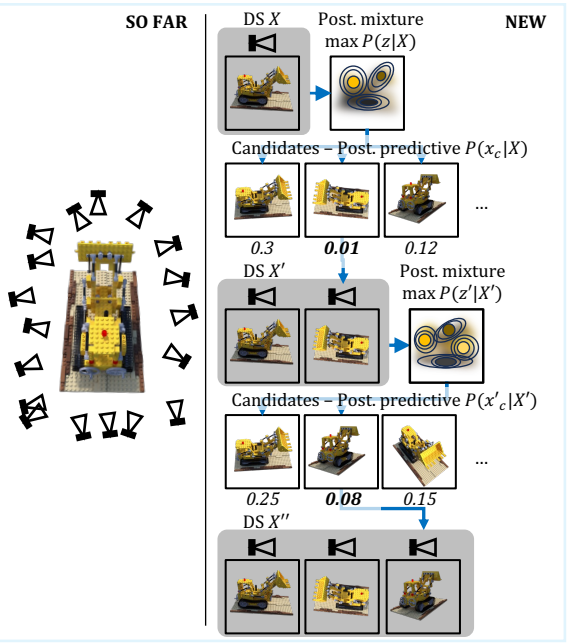
LENNART SCHULZE*

* ls3932@columbia.edu

We iteratively model the probability of new-view images belonging to old-view GMM to minimize the size of a data set explaining the maximum uncertainty in a scene. We use this set for NeRF-based 3D reconstruction.

1) Introduction

- **Motivation**: Since its introduction in 2020, NeRF[1] has become the SOTA approach to reconstruct 3D scenes from images.
- **Problem**: The required image number is an obstacle for training speed, real-world apps.
- **Solution**: Selecting only images with most explanatory value can achieve similar perf.
- **Contributions**:
  - 1) Unified model to describe multi-scale GMM for prob. modeling of images.
  - 2) Algorithm to compose dataset based on uncertainty resolved by samples.
  - 3) Experimental results from using algorithm to minimize train data for NeRF-based reconstruction.



2) Background

Neural (Radiance) Fields

- = Neural-implicit representation of a 3D scene via NN-modeled map over points: $f_\Phi: (\boldsymbol{x}, \boldsymbol{d}) \rightarrow (\sigma, \boldsymbol{c})$.
- Training via loss between reconstructed and given ground-truth 2D images.
- Training requires many (~100) object-centered images from different views, creating bottleneck for real-world apps.
- Prior approaches for speedup on data structure[2], generalization[3], not data selection.

Probabilistic modeling of images

- Images modeled probabilistically per single image, e.g. based on pixel neighborhood[4].
- Per-scene dataset modeling novel approach.

3) Method

Unified image modeling framework

x_{img} = image $\in \mathbb{R}^{w*h*3}$

$x \in$ {image, region, pixel} $= \mathbb{R}^{w_s*h_s*3}$

K = |Gaussian components| (per unit or shared)

F = |features per component| $\in$

{3 : one pixel for all pixels in pixel, region, or image;

$3 * w_s * h_s$: as many pixels as pixels in pixel, region, or image;

$(F_{before}) + 2$: model image position in Gaussian component}

Derived image modeling approaches

Method	*x*	*F*	*K*	*fit for*	*intuition*
A1	pixel	3	1	each pixel	Simple dist. per pixel
A2	pixel	3	m	each pixel	Complex dist. per pixel
A3	pixel	3	1, m	all pixels	Same dist. for all pixels
A4	pixel	3+2	npixels, m	all pixels	Each pixel finds dist.
B1	region	3	1	each region	Simple dist. per region
B2	region	3	m	each region	Complex dist. per region
B3	region	3	1, m	all regions	Same dist. for all regions
B4	region	3+2	nregions, m	all regions	Each region finds dist.
C1	img	3	1	all images	Same dist. for all imgs
C2	img	3	nimgs, m	all images	Complex dist. for all imgs

Proposed data selection algorithm

Input: $N_{train} \in \mathbf{N}$

$X = \{x_i \mid x_i \in \mathbf{R}^{3*w*h}\}_{i=0}^{N_{imgs}-1}$

Output: $X_{out} = \{x_j \mid x_j \in \mathbf{R}^{3*w*h}\}_{j=0}^{N_{train}-1}$

X_{out} = [random choice(X)];
X.pop(X_{out}[-1]);
for $(l = 0; l < N_{train}; l++)$ **do**
  $(\theta, \boldsymbol{\beta}, \mathbf{z}) = (\theta, \boldsymbol{\beta}, \mathbf{z}) \sim P(\theta, \boldsymbol{\beta}, \mathbf{z} | X_{out})$; /* Fit GMM LVs by Gibbs sampling */
  X_{out}.append(arg $\min_{x \in X} P(x|\theta, \boldsymbol{\beta})$) ; /* Select by posterior predictive */
  X.pop(X_{out}[-1]);
end
return X_{out}

Dataset

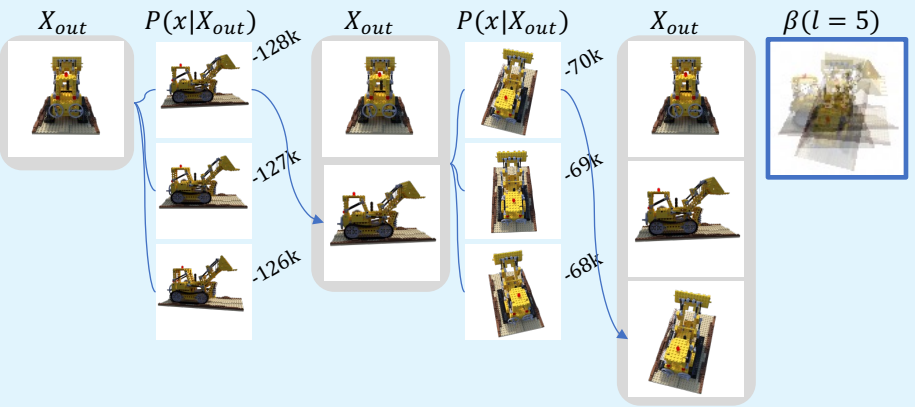
- "Blender bulldozer" dataset is 3D reconstruction benchmark.
- w=100, h=100, $x^{i,j} \in [0,1]$.
- N_imgs=100, N_val=100, N_test=25.

Implementation

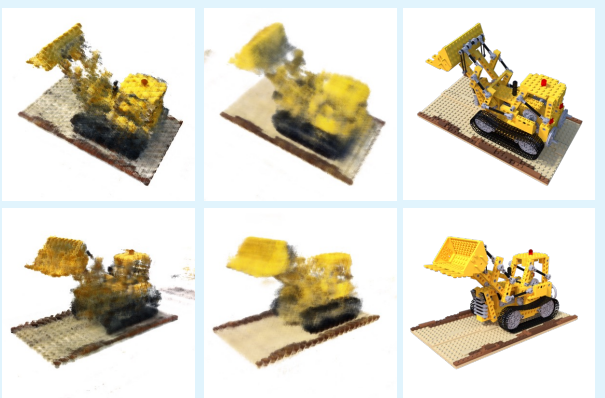
- Probabilistic image modeling implemented from scratch.
- Algorithm output fed into NeRF.
- NeRF Pytorch variant is adopted and customized[5].

4) Results

Method example: posterior predictive, beta for new images (A1)



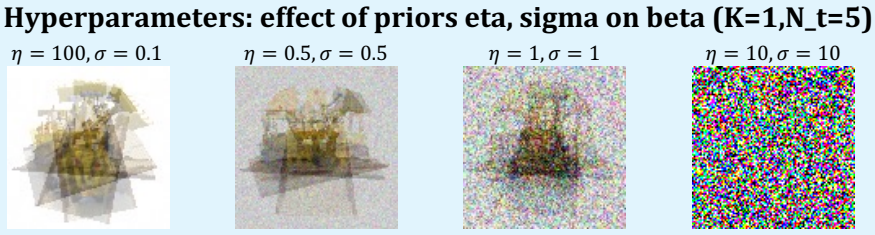
Reconstruction examples: Uncert.-guided (A1) vs Random vs GT



Reconstruction performance: test set (l2-loss ↓ / PSNR ↑)

Method	*N_train = 2*	*N_train = 5*	*N_train=10*	*N_train=20*
Uni. random	0.036 / 14.40	0.012 / 19.09	0.008 / 21.19	0.006 / 22.05
Approach A1	0.043 / 13.64	0.013 / 18.95	0.008 / 21.05	0.007 / 21.55
Approach A2	0.064 / 11.92	0.012 / 19.35	0.008 / 21.14	0.006 / 22.03

Hyperparameters: effect of priors eta, sigma on beta (K=1,N_t=5)



5) Conclusion

- **Can uncertainty-guided data selection reduce the number of training images while maintaining reconstruction quality?**
  - Per visual inspection: reconstruction quality higher than random approach, more details with same num. of samples.
  - Per loss: performance comparable; however, worst-case of random is worse than of uncertainty-guided approach.
- **Applications:** On pre-collected data sets; and knowledge can be used to guide future footage collection in real life and simulation.

References

[1] Mildenhall, B., et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." ECCV. 2020.
[2] Müller, Thomas, et al. "Instant neural graphics primitives with a multiresolution hash encoding." ACM Transactions on Graphics (ToG) 41.4 (2022): 1-15.
[3] Yu, Alex, et al. "pixelnerf: Neural radiance fields from one or few images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[4] Li, S. Z. (2009). Markov random field modeling in image analysis. Springer Science & Business.
[5] Yen-Chen, L. (2020). NeRF-pytorch. https://github.com/yenchenlin/nerf-pytorch/