

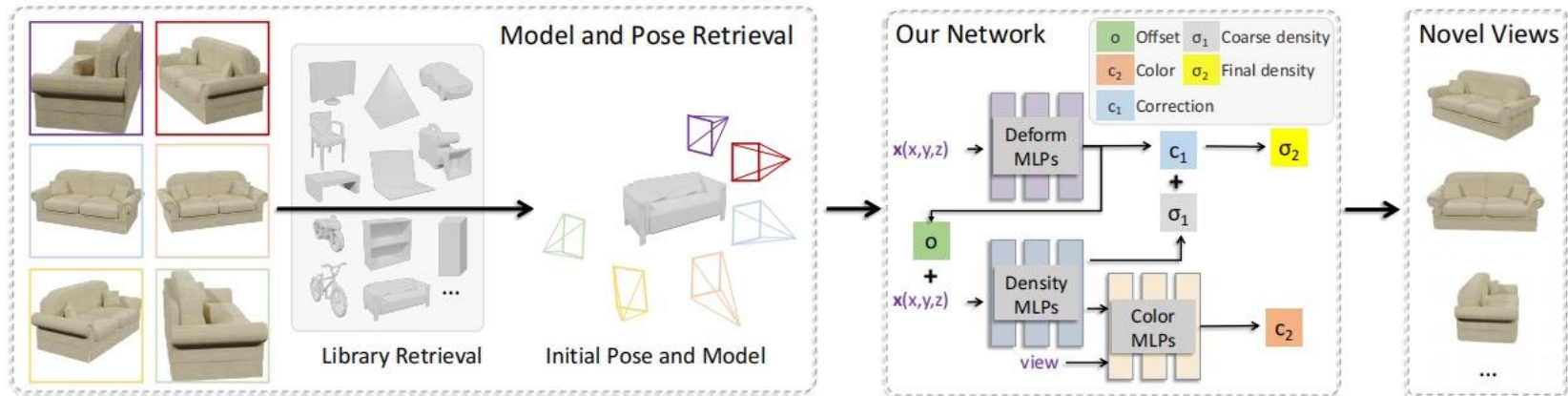
CAD-NeRF: Learning NeRFs from Uncalibrated Few-view Images by CAD Model Retrieval

**Xin WEN, Xuening ZHU, Renjiao YI ,
Zhifeng WANG, Chenyang ZHU, Kai XU**

Frontiers of Computer Science, DOI: [10.1007/s11704-024-40417-7](https://doi.org/10.1007/s11704-024-40417-7)

Problems & Ideas

- Problems of Sparse-view NeRFs approaches:
 - Most works require accurate initial camera poses of input images, which are derived from COLMAP, but COLMAP often fails in few-view and feature-sparse scenes.
 - Some works require rough camera poses from fixed views as inputs, which is limiting.
- Ideas: A multi-view pose retrieval and jointly training method to effectively reconstruct the neural radiance field from less than 10 images without any known poses.



The CAD-NeRF pipeline. Input images are used to retrieve the model of poses from the mini library. The CAD model is treated as the supervision to pre-train the initial density field (phase one). In phase two, sampled rays are sent from retrieved poses to generate 3D points along the rays, for the deformation network to predict the offset and correction of each point. Then a color network is added and three networks are trained together. Poses are optimized for a certain number of iterations to prevent overfitting.

Main Contributions

- Contributions:
 - We propose a method that reconstructs NeRFs from very few images without camera poses, with the help of prior shapes from CAD datasets. The method effectively boosts the generalization ability of NeRFs;
 - We propose a multi-view pose retrieval method, that considers the ordering of the input images, to prevent pose conflicts of input images;
 - On both synthetic and real objects, CAD-NeRF demonstrates significant improvement over the state-of-the-arts.

	PSNR \uparrow			SSIM \uparrow			LPIPS \downarrow			Average \downarrow		
	3-view	6-view	9-view	3-view	6-view	9-view	3-view	6-view	9-view	3-view	6-view	9-view
BARF	18.860	18.968	20.161	0.881	0.881	0.901	0.311	0.309	0.236	0.115	0.115	0.092
NeRS	18.687	18.561	19.070	0.885	0.884	0.895	0.210	0.193	0.181	0.104	0.102	0.095
SCNeRF	17.371	18.648	19.259	0.852	0.882	0.887	0.293	0.190	0.175	0.128	0.097	0.089
SPARF	18.459	20.781	20.759	0.880	0.933	0.933	0.220	0.094	0.092	0.104	0.059	0.059
Ours	21.137	21.669	21.931	0.918	0.929	0.930	0.105	0.082	0.081	0.067	0.058	0.056

Method	3-view	6-view	9-view
BARF	21.486	19.750	16.681
NeRS	60.534	12.091	9.327
SCNeRF	5.943	7.083	8.252
SPARF	21.470	4.876	6.086
Ours	5.273	3.960	3.914

Method	3-view	6-view	9-view
BARF	60.003	43.144	34.925
NeRS	194.467	36.103	21.564
SCNeRF	19.607	23.402	27.795
SPARF	18.500	9.920	14.660
Ours	14.960	8.140	7.900

The first row: evaluation of ShapeNet data with the initial identity poses obtained by our retrieval poses.

The second row: The comparison of rotation errors (left) and translation errors (right) of optimized poses for 3/6/9 input images, here the rotation errors are in degrees, the translation errors are multiplied by 100.