Official Reviews

Paper ID: 1399

Paper Title: NeLF-Pro: Neural Light Field Probes

Reviewer 1

Paper Summary:

The paper replaces the traditional volume-based neural representation with local light field probes. In this way, a relatively large sense can be decomposed into several regions and each region is guarded by a light field probe. Each light field probe encodes some local features for its own region and among all probes they share a common feature to ensure consistency across regions. This new representation can reduce the memory footprint for the model and also has better capability than most of the single-model representations, especially for large scenes. On the other hand, the light field representation naturally has less computation complexity so the training time and be reduced as well.

The paper compares its performance against Instant-NGP, F^2-NeRF on the mip-NeRF360 dataset and Free dataset. The NeLF-Pro implementation is 20x faster than other approaches at training time with comparable quality. It shows the advantage of large scenes to better preserve structured patterns.

Paper Strengths:

The paper introduces a new neural scene representation based on the light field. By decomposing the scene into regions, this approach can overcome an inherited limitation in light field representation, which means the light fields cannot model rays that are blocked by the scene itself. However, distributing the light fields across regions in this paper can resolve this issue. The VMM factorization technique can be useful to make the representation compact.

The paper is well-written and the video clearly demonstrates both the algorithm as well as the rendered animations.

Paper Weaknesses:

I don’t see any major weaknesses. The paper will likely have more impact if the authors release their code so it can be compared against future publications.

Please report the model size and rendering time which is also very important for the application.

Overall Recommendation: 5: accept

Justification For Recommendation And Suggestions For Rebuttal:

The paper elegantly blends the light field representation and scene decomposition and solves an inherent issue for light field representation. This design point consumes significantly less memory than volume-based methods which makes it an attractive choice for Neural Radiance Fields.

I don't have any concerns or issues that need to be addressed in the rebuttal.

Some literature reviews on recent neural light fields are missing:

1. R2L: Distilling Neural Radiance Field to Neural Light Field for Efficient Novel View Synthesis
2. NeuLF: Efficient Novel View Synthesis with Neural 4D Light Field
3. Light Field Networks: Neural Scene Representations with Single-Evaluation Rendering

It will be better to explicitly discuss how this method **can** model rays that are blocked by the scene itself and show a demo by moving the camera through objects in the scene.

Confidence Level: 5: The reviewer is absolutely certain that the evaluation is correct and very familiar with the relevant literature.

Final Rating: 5: Accept

Final Rating Justification:

The authors addressed my question on the blocked ray modeling quite clearly. I am looking forward to seeing a better discussion in the revision.

Reviewer 2

Paper Summary:

The paper proposes a radiance field representation and optimization scheme that is designed to flexibly handle both small and large-scale scenes, including scenes with arbitrary topology and no distinction between center/foreground and background, such as roads or hallways. The main idea is to combine aspects of light field representations and volume rendering, placing spherical light field probes sparsely throughout the scene and then performing volume rendering using features from nearby light field probes. These light field probes offer improvements in efficiency of representation and rendering because they are sparse in the scene and only nearby probes need to be considered when rendering each viewpoint, and including volume rendering makes the results more consistent with respect to occlusions and reduces jumping artifacts.

Paper Strengths:

The evaluation shows clear improvement over prior methods, especially for the large and irregularly structured scenes. Parts of the paper are written clearly, but others less so (see weaknesses). From my understanding, the main ideas that seem novel and interesting (and I'd like to see emphasized / explained clearly) are:

- How the ideas of light field probes and volume rendering are combined, and what technical benefits it yields
- How each light field probe is layered (like an onion?), and nearby probes are combined, to provide a semblance of mip-mapping (i.e. building on the explanation in lines 384-394)
- How the tensor factorization and weight sharing reduce memory usage while providing some tunable control over resolution and representation capacity
- How the probe locations are selected for the scene (i.e. building on lines 325-330), and how the probes that contribute to a rendered view are selected for each ray

Paper Weaknesses:

The method strikes me as convoluted, likely more than it needs to be. However, it's possible that the method is less convoluted than I think it is, and it's the description of the method that is convoluted. Details such as hyper parameter settings are provided, but more important descriptions of core aspects of the optimization and representation are not clearly described. Some specific concerns:

- It's not clear how the probes that contribute to each ray sample are selected, other than that it has to do with the camera pose.
- The description of aggregation is confusing; at the time it's presented (around Figure 2), it wasn't clear what was being aggregated.
- I'm not sure why the projection network is so named; from my understanding of its function I expected this would be called a decoder.
- The way each probe represents its portion of the scene is only partially clear. Lots of notation is used but not explained clearly. For example, I'm not sure what is added by equations 5 and 6, as 5 would appear to be superseded by 6 and 6 would appear to be superseded by 9. I suspect that these equations were presented in this order in an attempt to gradually increase complexity as the method is described, but I found it confusing to see multiple slightly different/inconsistent equations for the same quantity tau_l. I was also initially confused by the description of the core factors as global, and only on my second reading did I realize that these core factors are also localized, it's just that there are fewer of them. Likewise, I'm not sure what the term "modes" means on line 324. I would also appreciate some clearer description of how the tensor representation is interpreted as spherical (is this solely through the sinusoidal featurization?)
- I would suggest removing some of the hyper parameter settings details (to the appendix) and using the space to expand on the ablation studies and what insights can be gleaned from them.
- There are frequent typos and incomplete sentences, possibly due to copy and paste or splicing? e.g. lines 079 (extend -> extent), 119 (principle -> principled), 169 (quicking -> quick), 232 (Splitting -> Splatting), 235 (overfitting -> overfit), 242 (missing a preposition?), etc. there are too many to list them all

I also have a few concerns regarding the evaluation:

- I would be curious to see how the performance compares to Gaussian Splatting, which I expect should be capable of representing the scenes that are considered.
- The caption of Table 1 states that "the scores of baseline methods are taken from the original papers," yet many of the baseline methods were not evaluated on these datasets in the original papers (e.g. NeRF++, Plenoxels, DVGO). Perhaps the caption means to say that the original parameter settings or code distributions were used? Given the numbers, I also wonder if the original version of DVGO was used (as this is what is cited), or if the authors intended to refer to DVGOv2?
- Table 3 (large scale scenes) does not compare to BlockNeRF, which to my knowledge is a natural baseline for this setting.

Overall Recommendation: 2: weak reject

Justification For Recommendation And Suggestions For Rebuttal:

The results are impressive both quantitatively and qualitatively, on a challenging and important task of modeling large and irregularly shaped scenes. However, the method strikes me as convoluted and poorly exposited, making me worry that (1) interesting insights may be lost regarding how and why this particular set of design decisions works, and (2) it may be challenging for other researchers to understand the method, reproduce it, and build on it. The authors promise to release their code, but this does not absolve them of the necessity to make at least the core of their method understandable and reproducible from the

paper text alone.

I am not comfortable publishing the paper in its current state because of the lack of clarity of the exposition in terms of how the method works and why many of the design decisions were made. I would recommend restructuring the method description following a top-down approach, providing first a high level description and intuition of the model components and the purpose of each, followed by more details about the implementation, tensor dimensions, aggregation, etc. If the exposition is made clear, I am open to raising my score as I do believe the results are compelling and the problem setting is important.

Confidence Level: 4: The reviewer is confident but not absolutely certain that the evaluation is correct.

Final Rating: 4: Weak Accept

Final Rating Justification:

Thanks for the clarifications and the comparison to 3DGS. I am raising my score based on the rebuttal and the visual and quantitative quality of the results. I encourage the authors to make the promised clarifications in the camera-ready manuscript so that the method will be easily understood by future readers.

Reviewer 3

Paper Summary:

This paper proposed a novel neural scene representation using light field probes. The light field of a scene is presented as a set of local light field feature probes, and a point feature can be reconstructed by weighted blending of probes close to the camera position. Experiment results showed that the proposed approach achieved better reconstruction quality compared to feature-grid-based representations, while using on par or slightly larger compute cost.

Paper Strengths:

- The proposed approach achieved better reconstruction quality than state of the art feature-grid-based approaches while using on par compute, especially doing better on thin structures and fine details.
- The light field probe factorization idea for scene presentation is elegant and effective.

Paper Weaknesses:

- The paper could benefit from more comprehensive comparison with the state of the art approaches for large scale scenes, such as Block-NeRF.
- While being claimed as a main benefit compared to feature-grid-based approaches, the authors have not demonstrated the proposed approach scale better with large scenes (e.g., avoid tiling effects and artifacts).

Overall Recommendation: 4: weak accept

Justification For Recommendation And Suggestions For Rebuttal:

- The paper proposed a novel approach to neural scene representation with light field probes and the results seem to show improved reconstruction quality compared to feature-grid-based approach
- The evaluation could be improved by comparing to more state-of-the-art approach, such as Block-NeRF

Confidence Level: 3: The reviewer is fairly confident that the evaluation is correct.

Final Rating: 4: Weak Accept

Final Rating Justification:

The rebuttal has addressed my questions and concerns very well, for BlockNeRF comparison I understand it's challenging during the rebuttal period but worth adding to the revision.