IS1-026

# BlipNeRF: MLLM による不可視視点の予測を用いた少数の物体正面画像からの 3 次元再構成

松浦史明 , 中溝雄斗 , 田邊光 , 柳井啓司（電気通信大学）

## 概要

### 背景

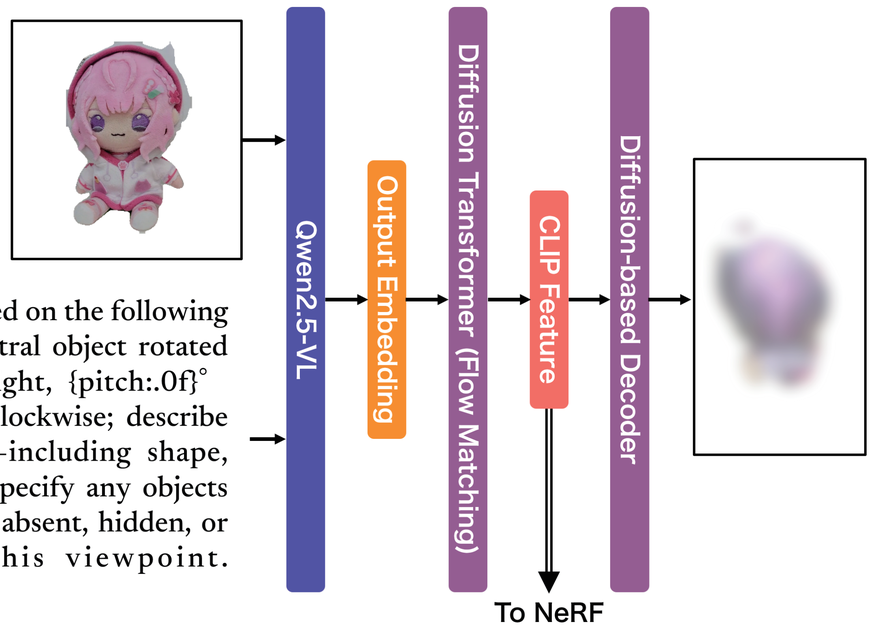
- 正面側 view, Few-Shot での NeRF 再構成
- 意味的な正則化を用いた場合、背面の出力が現在見えている範囲の特徴量に影響され、偽の目などのアーティファクトが出現する
  →ヤヌス問題（[1] に関する議論より）

### 目的

**意味的な正則化のみでの、ヤヌス問題の解決**

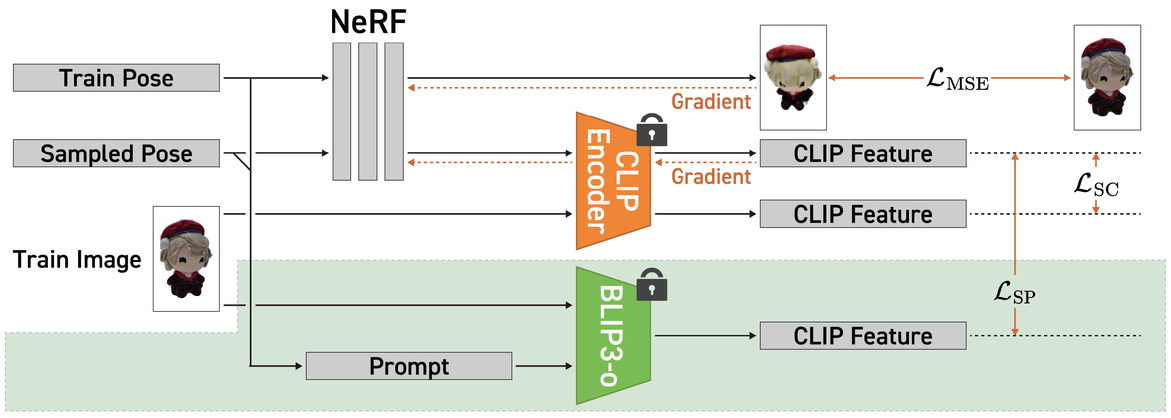
## 手法

### 視点予測




- MLLM の一種である BLIP3-o[2] を用いて、入力画像中の物体を指定角度回転させたときの見え方を推測
- 画像生成プロセス中に中間表現として生成される、CLIP 特徴 [3] を NeRF 学習へ利用

### NeRF との統合




$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{MSE}} + \mathcal{L}_{\mathrm{SC}} + \mathcal{L}_{\mathrm{SP}}$$

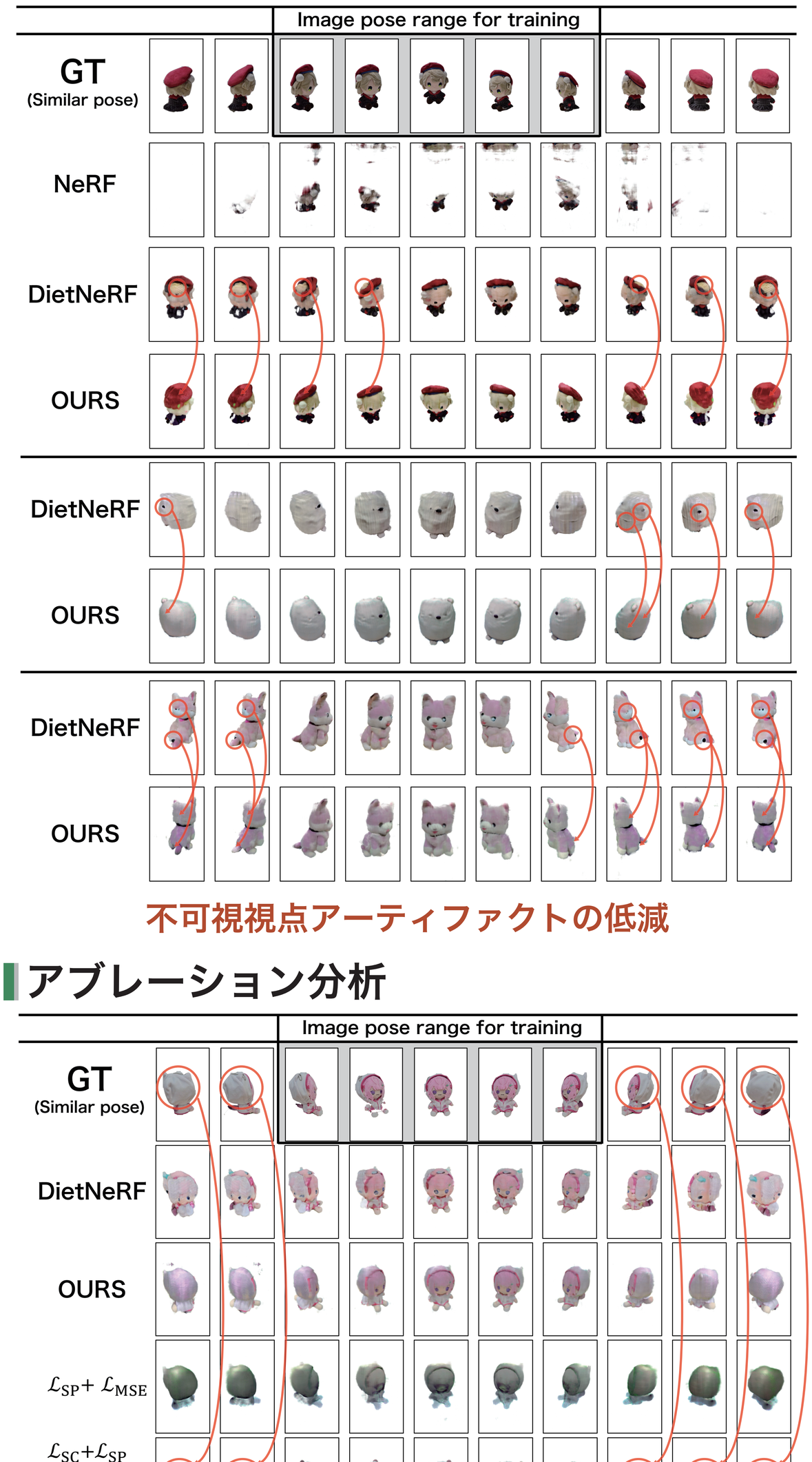
- **意味一貫性損失：** $\mathcal{L}_{\mathrm{SC},\ell_2}\left(I, \hat{I}_{\mathbf{p}_\mathrm{r}}\right) = \lambda_{\mathrm{sc}} \left|\left|\phi\left(I\right) - \phi\left(\hat{I}_{\mathbf{p}_\mathrm{r}}\right)\right|\right|_2^2$
  [4] より。レンダリング結果と学習画像間で、意味的な一貫性を持たせる（$\phi$: 画像エンコーダ）
- **意味予測損失：** $\mathcal{L}_{\mathrm{SP},\ell_2}\left(I, \hat{I}_{\mathbf{p}_\mathrm{r}}\right) = \lambda_{\mathrm{sp}} \left|\left|B(I, prompt) - \phi\left(\hat{I}_{\mathbf{p}_\mathrm{r}}\right)\right|\right|_2^2$
  レンダリング結果を、意味的に BLIP3-o の視点推測結果に近づける（B: BLIP3-o）
- その他の構造は、ベースとした DietNeRF[4] と同様

## 結果

### プロジェクトページ

- 以下記載の再構成結果を、右記リンクへ掲載 （https://mm.cs.uec.ac.jp/matsuura-f/blipnerf/）

### ベース手法との比較




**不可視視点アーティファクトの低減**

### アブレーション分析




- （５段目）プロンプトを、「パンダの画像」に
  **BLIP3-o による、不可視視点の学習誘導の実証**

[1] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. DreamFusion: Text-to-3D using 2D Diffusion. In ICLR, 2023.
[2] Jiuhai Chen, ZhiyangXu, Xichen Pan, Yushi Hu, Can Qin, Tom Goldstein, Lifu Huang, Tianyi Zhou, Saining Xie, Silvio Savarese, Le Xue, Caiming Xiong, and Ran Xu. BLIP3-o: A Family of Fully Open Unified Multimodal Models-Architecture, Training and Dataset. arXivpreprint arXiv:2505.09568, 2025.
[3] Quan Sun, Yuxin Fang, Ledell Wu, Xinlong Wang, and Yue Cao. EVA-CLIP: Improved Training Techniques for CLIP at Scale. arXivpreprint arXiv:2303.15389, 2023.
[4] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting NeRF on a Diet: Semantically Consistent Few-Shot View Synthesis. In ICCV, pages 5885–5894, 2021.

miru 2025.07