

# U-Net を用いた点光源・透視投影下における単一視点画像からの反射率・法線・深度推定

EP22119 原 圭佑

指導教授：宇佐美 裕康

## 1. はじめに

Computer Vision における三次元形状推定は重要課題であり、単一画像による復元は低コストかつ機材制約のある環境で有用である。従来の Shape from Shading [1] 法は理想化された前提（平行光線・平行投影）に依存しており、実環境への適用に限界がある。そこで本研究では、点光源照明および透視投影下での輝度変化を考慮した、反射率・法線・深度の推定を行う。これは Iwahori [2] らの研究に基づくアプローチである。

## 2. 提案手法

### 2-1. データセット構築

Mitsuba3 を用いて、Poly Haven や Scanned Objects で取得した 3D モデルから RGB 画像・反射率・法線・深度を合成した。原点にカメラと点光源を配置し、透視カメラ（50mm, 256×256）で 65 種の物体に対し回転と奥行きを変化させ、全 4680 シーンを生成した。輝度が極端な画像は除外し、学習に適したデータを構築した。また、Mitsuba3 出力の深度は不十分であるため、Lambertian 反射モデルに基づき Z 値を最適化し、高精度な深度マップを作成した。さらにヒストグラムで有効範囲を抽出し、0～1 に正規化してノイズを抑制した。

### 2-2. U-Net 構造

本研究では、RGB 画像から反射率・法線・深度を同時に推定するため、マルチ出力型の U-Net [3] を構築した（図 1）。エンコーダは ReLU と MaxPooling により特徴を抽出し、ボトルネック層を経て 3 つのデコーダに分岐する。各デコーダは対応するエンコーダとの skip connection により高解像度情報を活用し、法線には Leaky ReLU, 反射率と深度には ReLU を適用する。出力は Sigmoid または Tanh でスケールリングし、教師データと同一範囲に正規化される。

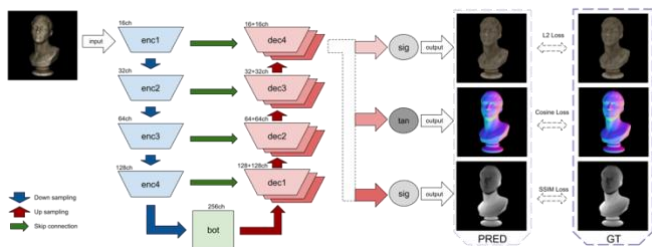


図 1 U-Net 構造

### 2-3. 損失関数

各出力（反射率・法線・深度）に対して適切な損失関数を設計し、背景を除くためにマスクを適用した。最終損失は以下の通りである：

$$\mathcal{L}_{total} = \mathcal{L}_a + \mathcal{L}_n + \mathcal{L}_d \quad (1)$$

反射率には平均二乗誤差（MSE）を、法線にはコサイン類似度に基づく損失を、深度推定では、構造的類似度指標（SSIM）により、構造的整合性を保ちながら誤差を最小化している。

## 3. 実験

実装には PyTorch を用い、GPU 環境（RTX 2070 SUPER）で学習を実施した。Adam 最適化と学習率  $1 \times 10^{-4}$ , 最大 200 エポックで訓練し、50 エポック損失が改善しない場合は Early Stopping を適用して過学習を抑制した。反射率および深度の定量的性能評価には平均絶対誤差（MAE）、法線には平均角度誤差（Angular MAE）を用いた。これらの指標に基づいて、テストデータセットに対する提案手法の性能を評価し、先行研究と比較した結果を表 1 に、また可視化結果を図 2 に示す。

表 1 比較結果

Method	MAE		Angular MAE Normal [°]
	Albedo	Depth	
Depth Anything V2	-	0.3155	-
DepthPro	-	0.4373	-
Marigold	-	0.2368	19.24
Ours	0.059	<b>0.1664</b>	<b>16.47</b>

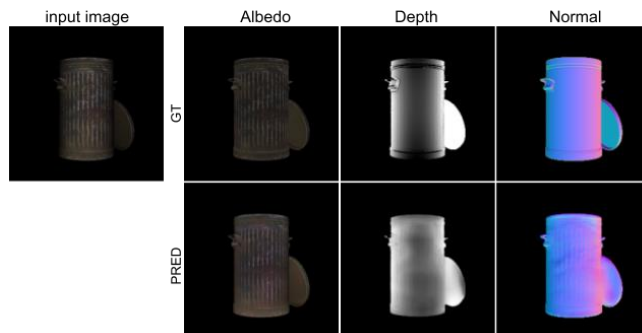


図 2 推定結果

表 1 から、提案手法は深度および法線推定において先行研究を上回る精度を達成した。特に深度の MAE は 0.1664 と最も低く、法線の平均角度誤差も 16.47°と優れている。図 2 の可視化結果からも、提案手法は構造を保った自然な再構成が可能であり、形状や材質の違いにも安定して対応できている。

## 4. むすび

本研究では、点光源照明および透視投影下での単一画像から、反射率・法線・深度を推定する手法を提案した。今後は、実写画像や非ランバート材質などの対応が課題である。

## 参考文献

- [1] B.K.P. Horn, "Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View," MIT Technical Report, 1970.
- [2] Yuji Iwahori et al., "Point Source Illumination Stereo for Objects with Uniform Reflectance," Computers & Graphics, vol. 14, no. 2, pp. 199–207, 1990.
- [3] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, pp. 234–241, 2015.