青山学院大学理工学部
情報テクノロジー学科
２０２２年度卒業研究論文

単眼RGBカメラを用いた食品の形状把握

２０２３年1月２６日提出

指導教員　鷲見和彦

１５８１９０１９　遠藤琳太

単眼 RGB カメラを用いた食品の形状把握

遠藤　琳太 (15819019)
鷲　見　研　究　室

1.　はじめに

近年技術の発達に伴い，自動あたため機能をもつ電子レンジが数多く発売されている．電子レンジの自動あたため機能では赤外線・湿度・蒸気・重量・温度などを測定するセンサのいずれか，もしくはその複数を組み合わせて自動で調理を行っている．しかし食品によっては，加熱後も食品内部が冷たい場合や，逆に外側部分が温まりすぎる場合がある．これらの問題は物体形状に適した加熱方法を行うことで，対処が可能である．これらの背景から，庫内の物体形状を正確に把握する技術が必要とされており，その方法の 1 つとして 3 次元再構成が存在する．

実際に 3 次元再構成を用いた物体形状把握をするにあたり，調理家電は消費者向け製品であるため，高度なセンサを搭載することや，複数のカメラを用いることはコストの面から考えて難しい．そのため，カメラ一つだけで物体形状把握が可能な単眼 3 次元再構成に着目した．

単眼 3 次元再構成は単一視点画像から 3 次元モデルを再構成する技術で，コンピュータビジョン分野において重要な課題となっている．単一視点画像からの 3 次元再構成は物体のある一面の情報から，見えていない部分も推定する必要があるので，コンピュータにとって非常に難易度の高い課題となっている．

2.　提案手法

本研究では，背景のある実画像から，CG データセットのみを用いて，被写体を 3 次元モデルに再構成する．再構成はまず Ronnerberger らの手法 [1] をベースに用いてフォアグラウンドセグメンテーションで被写体のマスク画像を生成する．その後 Hu らの手法 [2] をベースに用いて単眼深度推定を行い，マスク画像と合成することで，深度情報を含むマスキングされた被写体の画像を得る．その後 Wang らの手法 [3] をベースに用いて 3 次元メッシュモデルを生成する．またフォアグラウンドセグメンテーション，単眼深度推定，単眼 3 次元再構成のモデルの学習に使用するデータセットの作成にはドメイン適応 [4] を用いる．

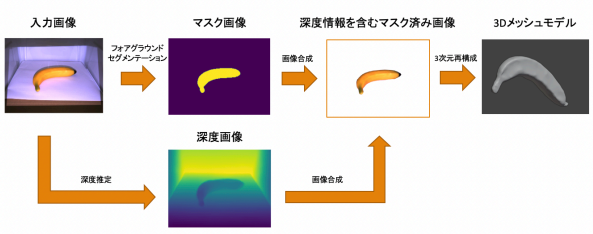

図 1　提案手法概要

本研究で用いるセグメンテーションネットワークは Ronnerberger らのネットワーク [1] を採用しており，バックボーンには ResNet34 [5] を使用した．はみ出しなどを少なく，厳密に評価したいので損失関数には Jaccard 係数を使用した．また，学習の際にデータ拡張として，ランダムにノイズ加算，色相彩度の変更を行った．

三次元再構成には Wang らの手法 [3] をベースとして用いるが，この手法では再構成に深度画像を使用のに対して，本研究では深度画像を新しいモダリティとして加えて推定を行いたいため，入力画像のチャンネル数を RGB の 3 チャンネルから RGBD の 4 チャンネルに変更した．

3.　評価実験

表 1　提案手法評価結果平均

	chamfer loss	edge loss	normal loss	laplacian loss
バナナ	0.199	0.00358	**0.0221**	0.00454
クロワッサン	0.159	0.00361	0.0320	0.00479
梨	**0.131**	**0.00269**	0.0298	**0.00389**
桃	0.333	0.00320	0.0351	0.00421

表 1 から全ての項目で高い loss や，低い loss をもつ物体はなく，物体ごとの特徴があることが分かった．normal loss を除いて，全体的な精度が最も高かったのが梨だった．バナナ・クロワッサン・桃は暖色系の色合いをもつのに対して，梨は寒色系の色合いをもつ

ため，コンピュータが混乱しづらかったのが要因だと考える．また，桃は chamfer loss が他の物体に比べて高い結果となった．

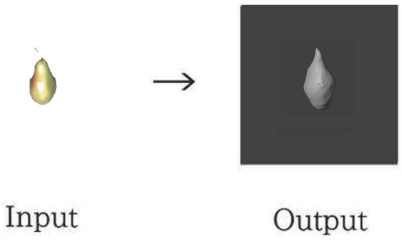

図 2　梨の 3 次元再構成結果例（左：入力画像，右：出力モデル）

表 2　深度情報の有無による結果の比較

	chamfer loss	edge loss	normal loss	laplacian loss
深度情報有り平均	0.206	**0.00327**	**0.0298**	**0.00435**
深度情報無し平均	**0.199**	0.00411	0.1047	0.00787

表 2 から，chamfer loss を除いて深度情報がある方が，深度情報が無い場合より精度が高くなった．chamfer loss の精度の差も 0.007 と非常に小さく，深度情報がある方が良い結果を得られたと考えられる．

4.　おわりに

本研究では，Blender を基軸とした 3DCG 制作ソフトウェアによって生成したデータセットのみを用いて背景あり画像の 3 次元再構成を行う方法を提案した．

今後の課題として，まずデータセットの拡充が求められる．特に 3 次元再構成には大規模なデータセットが必要だと考える．本研究のベースラインである Wang らの研究 [3] では，3 次元再構成のデータセットとして大規模な 3D モデルデータセットである ShapeNet を使用しており，13 種類の物体に対して計 50,000 個の 3D モデルからレンダリングした画像と点群情報を用いて学習している．一方この研究では 4 種類の物体に対して計 20 個の 3D モデルのみを使用しているため，データセットのバリエーションが不足している．現時点ではデータセットのバリエーションの少なさから，角度によって再構成の精度が大きく影響してしまっているため，全方位からのデータを満遍なく生成することでこのような事態を防ぐことができる．

またレンダリングを行ったレンダラの都合上，光源から直接物体表面に当たる光のみレンダリングしており，照り返しなどの複雑な照明効果が無かったため，物体が置かれている面を物体と誤認識する場合があった．このような点を考慮しつつデータセットを作成すればより良い結果が得られると考える．

また現時点では，実際に電子レンジで調理を行うようなもの，例えば，お椀に入ったスープやお皿に乗っているおかず，冷凍食品など実際に温める機会が多そうなについては形状が複雑なため再構成には至れなかったため，このような食品に対しても再構成を行う必要がある．

参考文献

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. pp. 234–241, 2015.

[2] J. Hu, M. Ozay, Y. Zhang, and T. Okatani. Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries. *IEEE Winter Conference on Applications of Computer Vision*, pp. 1043–1051, 2019.

[3] N Wang, Y Zhang, Z Li, Y Fu, W Liu, and Y Jiang. Pixel2mesh:Generating 3D Mesh Models from Single RGB Images. *Proceedings of the European Conference on Computer Vision*, 2018.

[4] 富永樹, et al. 既知の三次元環境内における物体形状把握. 青山学院大学理工学部 情報テクノロジー学科 2021 年度卒業研究論文, 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '16, pp. 770–778. IEEE, June 2016.

目 次

図 目 次

表 目 次

第1章　序論

1.1　研究背景

近年技術の発達に伴い，自動あたため機能をもつ電子レンジが数多く発売されている．電子レンジの自動あたため機能では赤外線・湿度・蒸気・重量・温度などを測定するセンサのいずれかもしくはその複数を組み合わせて自動で調理を行っている．しかし食品によっては，加熱後も食品内部が冷たい場合や，逆に外側部分が温まりすぎる場合がある．これらの問題は物体形状に適した加熱方法を行うことで，対処が可能である．したがって，庫内の物体形状を正確に把握する技術が必要とされており，その方法の 1 つとして 3 次元再構成が存在する．

実際に 3 次元再構成を用いた物体形状把握をするにあたり，調理家電は消費者向け製品であるため，高度なセンサを搭載することや，複数のカメラを用いることはコストの面から考えて難しい．そのためカメラ一つだけで可能な，単眼 3 次元再構成による物体形状把握に着目した．

単眼 3 次元再構成は単一視点画像から 3 次元モデルを再構成する技術で，物体のある一面の情報から見えていない部分も推定する必要があるので，コンピュータにとって非常に難易度が高く，重要な課題となっている．

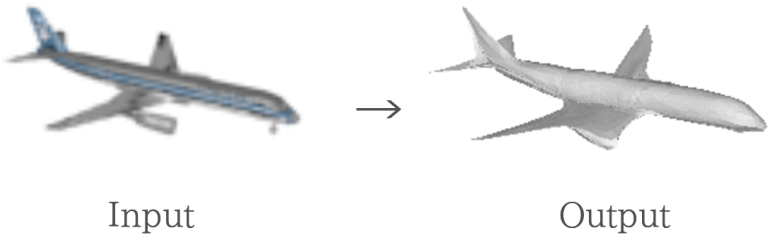

図 1-1: 単眼 3 次元再構成の例

1.2 研究目的

本研究では，CG で作成したデータセットの学習のみを用いて，背景有り画像の被写体の 3 次元モデル再構成を目的とする．被写体と背景の領域を推定するフォアグラウンドセグメンテーション，CG データセットの様々なパラメータをランダム化することでモデルの汎化性能を向上させるドメイン適応，単一視点画像から画像の深度を推定する単眼深度推定などの技術を取り入れることで，背景有り画像の単眼 3 次元再構成を可能にする．

1.3 本論文の構成

本論文では，第 2 章で画像の領域分割，単眼 3 次元再構成，単眼深度推定，ドメイン適応について紹介し，本研究のベースとなる手法について説明する．次に第 3 章で提案手法について詳細に述べ，第 4 章で提案手法の結果と定量的評価を紹介する．最後に第 5 章で本研究の結論と今後の課題について述べる．

第２章　関連研究

2.1　画像の領域分割

画像の領域分割とは，画像内で似た特徴量をもつグループごとにまとめることで，画像内の物体領域をピクセル毎に分類する技術のことである．ピクセル毎に意味をラベル付けしていくセマンティックセグメンテーションや個体ごとにラベル付けを行うインスタンスセグメンテーションなどが存在するが，本研究では図 2-1 のように前景と背景の領域を分割するフォアグラウンドセグメンテーションを使用する．

図 2-1: セマンティックセグメンテーションの例

セグメンテーションの手法の 1 つである全層畳み込みネットワーク（FCN） [1] は，クラス分類ニューラルネットワークでの全結合層を畳み込み層に置き換え，畳み込み層とプーリング層でネットワークを構成することで物体が何であるかという出力ではなく，物体がどこにあるかという出力（確率マップ）を得ている．

Mask R − CNN [2] は領域畳み込みニューラルネットワーク（R-CNN）をベースとした手法で，物体検出を応用しセグメンテーションを行っている．このネットワークでは物体が存在する領域（RoI）を抽出し，物体を囲む枠とクラスを推定する物体検出を行い，その枠内に対してピクセルごとの領域分類を行っている．

エンコーダ・デコーダ構造のネットワークは，画像から特徴を抽出するエンコーダと，抽出した特徴から確率マップを出力するデコーダ部分で構成される．エンコーダで特徴を抽出する過程で失った情報を，デコーダ部分で再構成することで精度を向上させる仕組みになっている．この構造をもった代表的なネットワークとして SegNet [3] や U-Net [4] がある．

本研究の画像の領域分割を行うベースラインである U-Net [4] では，図 2-2 のように入力された画像に対して畳み込み積分と最大値プーリングを行うことで特徴量を抽出し，抽出した特徴量を逆畳み込み積分することで入力画像と同サイズの確率マップを出力している．このネットワークの特徴としてスキップ接続があり，これはエンコーダでの特徴マップを複製してデコーダの特徴マップに接続することで，分類精度を向上させている．

また，物体検出やセグメンテーションにおいては，エンコーダに画像分類のモデルの構造をほぼそのまま使用することができ，このとき基本とするネットワーク構造のことをバックボーンと呼ぶ．この研究ではバックボーンに ResNet34 [5] を使用している．ResNet ではネットワーク内での手前の層の入力を後ろの層に足し合わせる Shortcut Connection を導入することで，深い層での学習を可能にしている．

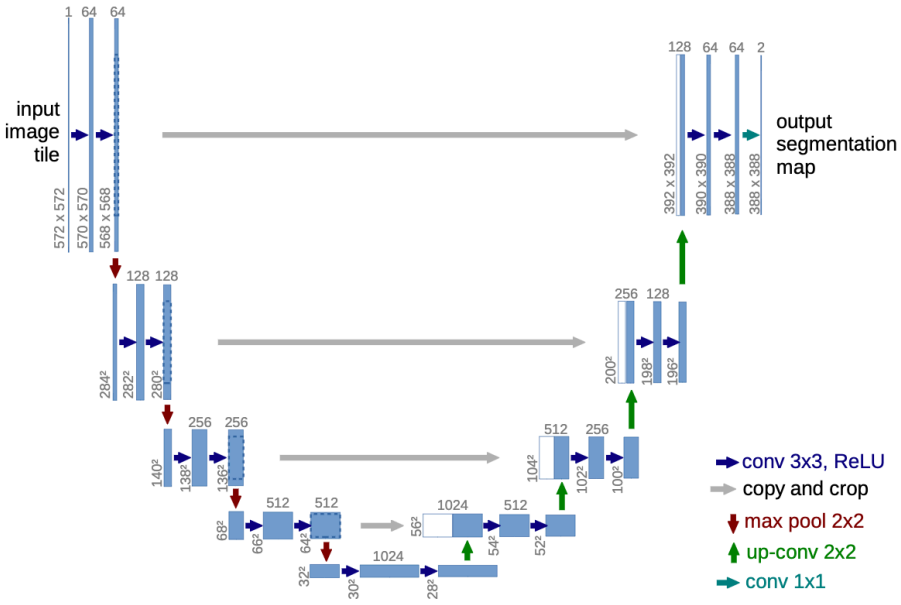

図 2-2: Ronnerberger らの提案手法概要

2.2 単眼 3 次元再構成

単眼 3 次元再構成とは，単一視点画像から 3 次元モデルを再構成する技術である．人間は単一視点を見るだけで，目に見えない部分状態を経験的に推定できる．同様に，コンピュータが単一視点から見えていない部分を推定するためには，その物体の典型的な形状を学ぶ必要がある．このような問題に対して深層学習ベースの手法などが考えられている．

単眼 3 次元再構成にはボクセルの再構成，点群の再構成，メッシュの再構成の 3 種類ある．ボクセルの再構成 [6] は GPU のメモリ効率が悪く，解像度の高い再構成は難しい．点群の再構成 [7] は物体形状を得るために点同士の接続を別途計算する必要がある．メッシュの再構成はボクセルの再構成に比べメモリ効率が良く，また点群の再構成と違い点同士の接続情報をもつため，本研究ではメッシュの再構成を目的とする．

本研究の 3 次元再構成のベースラインである Wang らの研究 [8] はメッシュテンプレートを使用してメッシュ再構成する手法で，図 2-3 のように変形元である楕円体に対して画像から抽出した特徴量に応じたメッシュ変形とグラフ逆プーリングを繰り返すことで再構成している．

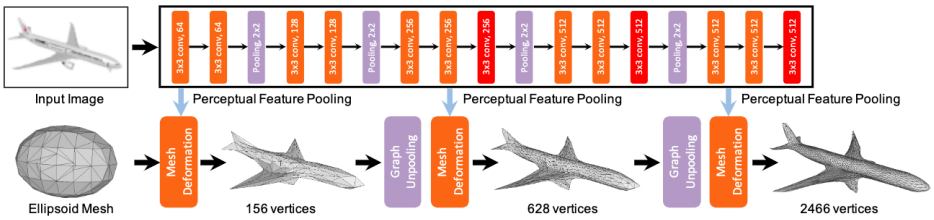

図 2-3: Wang らの提案手法概要

また，このネットワークでは，初期メッシュである楕円体を移動しながら分割することで，目的の形状を得ている．図 2-4 のように各面の中心ではなく，辺上に頂点を増やすことで均一な密度で面を増やしている．

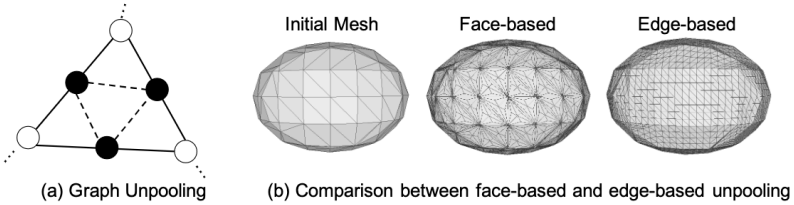

図 2-4: グラフ逆プーリング手法概要

また Kato らの研究 [9] では，三角形メッシュの頂点位置，メッシュ上のカラーテクスチャ，カメラ，ライトのパラメータを調整することで，目標とする画像とレンダリング結果のロスを最小にし，目的のメッシュを生成している．

2.3 単眼深度推定

単眼深度推定は，単一視点画像から物体までの距離を推定する技術のことである．本研究では深層学習をベースとした手法に着目しており，Eigen らの手法 [10] では global coarse-scale network で画面全体の深度特徴を抽出し，抽出した深度特徴と入力画像を連結処理する local fine-scale network を用いることで詳細な深度画像を出力している．

本研究の単眼深度推定のベースラインである Hu らの手法 [11] は Eigen らの手法を基に提案されており，図 2-5 のように入力 RGB 画像から縮尺を変えた特徴を抽出しサンプリングすることで異なる解像度の特徴マップの情報を集約している．

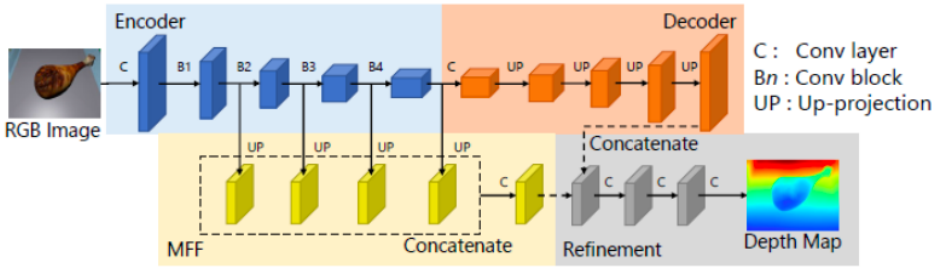

図 2-5: Hu らの提案手法概要

2.4 ドメイン適応

ドメイン適応とは，転移学習の一種で，十分な教師ラベルをもつソースドメインから得られた知識を，教師ラベルの不十分なターゲットドメインに適応することである．一般的には，Goodfellow らの提案したネットワークである Generative Adversarial Networks（GAN） [12] を学習に用いることが多い．富永らの研究 [13] では CG 画像のデータセットを生成する際に，テクスチャや光源，座標などをランダムにレンダリングすることで，実画像自体もそのバリエーションの一部として扱えるようにする手法を用いている．この手法を用いることで，実画像における光源などの条件の変化にもロバストな推定を行える．本研究では同様に，学習の段階ではなくデータセット

の作成の際に，様々な要素をランダムに変更しながらレンダリングすることで，実画像の推定を行えるようにしている．

第3章　提案手法

3.1　概要

本研究では，背景のある実画像から，CG データセットのみを用いて，被写体を 3 次元モデルに再構成する．再構成はまず Ronnerberger らの手法 [4] をベースに用いてフォアグラウンドセグメンテーションで被写体のマスク画像を生成する．その後 Hu らの手法 [11] をベースに用いて単眼深度推定を行い，マスク画像と合成することで，深度情報を含むマスキングされた被写体の画像を得る．その後 Wang らの手法 [8] をベースに用いて 3 次元メッシュモデルを生成する．またフォアグラウンドセグメンテーション，単眼深度推定，単眼 3 次元再構成のモデルの学習に使用するデータセットの作成にはドメイン適応 [13] を用いる．

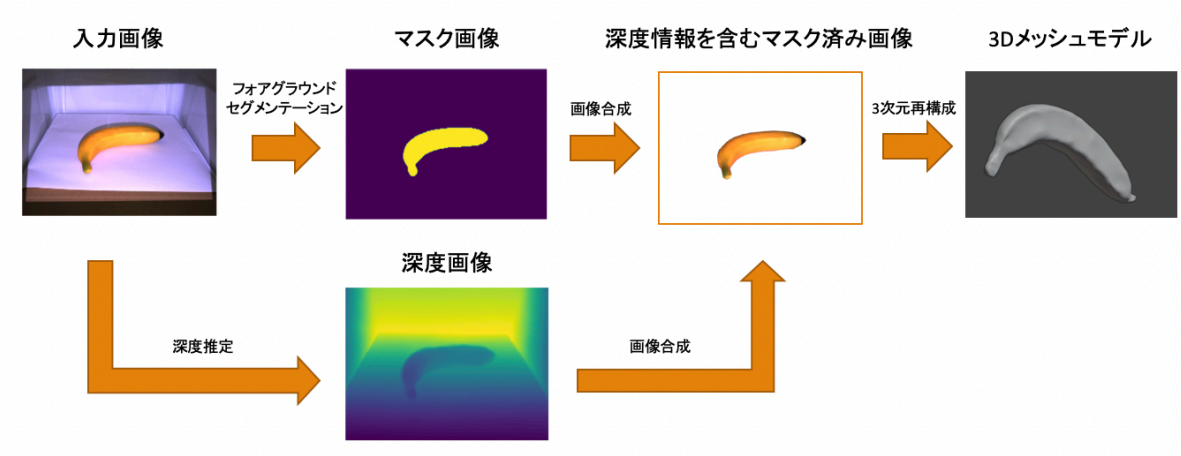

図 3-1: 提案手法概要

3.2 準備

3.2.1 実画像の撮影

3 次元再構成を行う対象である背景のある実画像の撮影を行った．対象とした食品はバナナ，クロワッサン，梨，桃の 4 種類で，食品の位置と回転をランダムに各食品 5 枚撮影した．

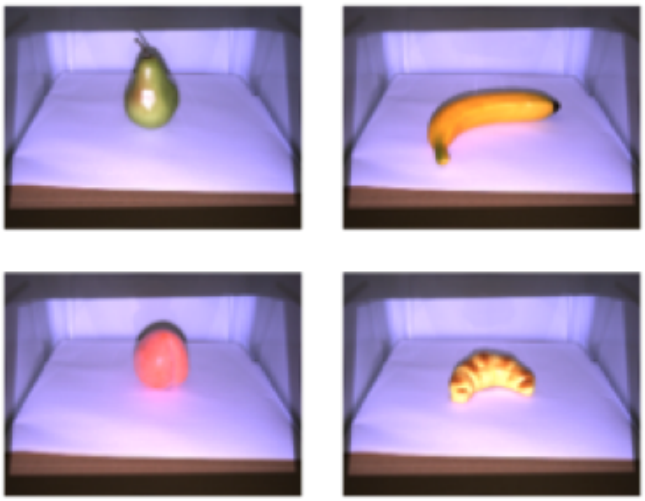

図 3-2: 撮影した食品の実画像例

3.2.2 食品サンプルの 3D スキャン

3 次元再構成を行った後に評価実験を行うため，正解 3D モデルは正確に計測できる 3D スキャンを用いて撮影した．スキャンには対象物との距離を測定する LiDAR を使用し，スキャンの際に生じた物体表面の凹凸を緩和するために表面にはスムージングをかけている．3D スキャンを行った結果が図 3-3 である．

3.3 フォアグラウンドセグメンテーションと深度推定

3.3.1 正解マスク・正解深度付きデータセット

正解マスク・正解深度付きデータセットはフォアグラウンドセグメンテーションのモデルと単眼深度推定の 2 つのモデルの訓練に使用する．

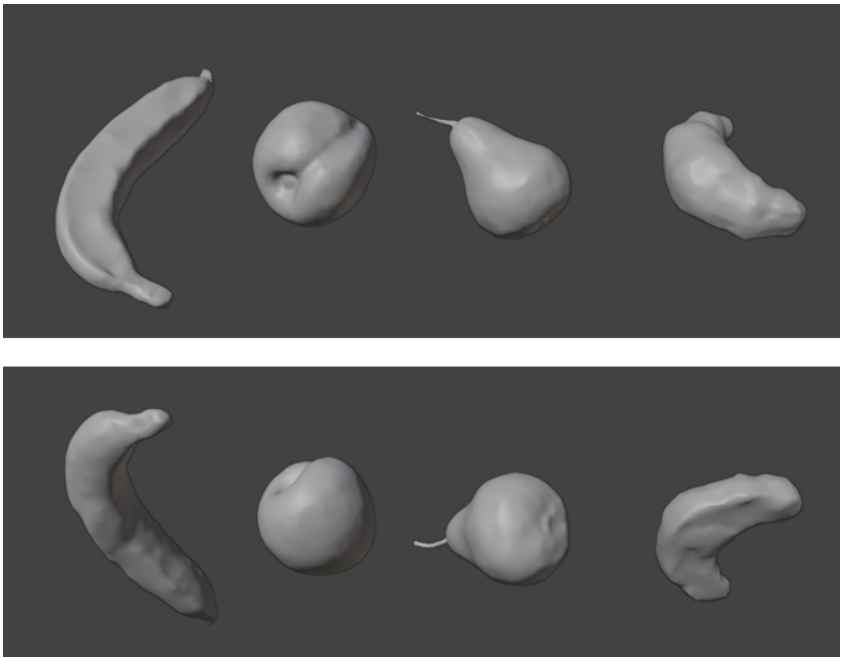

図 3-3: 3D スキャン結果画像（上：上図，下：下図）

データセットは 3DCG 制作ソフトウェアである Blender [14] とプロシージャル Blender パイプラインである BlenderProc2 [15] を使用して生成した．各ピクセルに対して RGB 情報と，写っているものが背景か否かの情報，カメラから物体までの距離情報（深度情報）を含めるように作成した．

実画像の再現にあたり，実画像に対してパース解析ツール fSpy を用いることで，図 3-4 のように撮影時のカメラ設定を再現した．食品の 3D モデルは富永の研究 [13] で用いられていたものと同様のもの（9 種類の食品）を使用し，食品の座標と回転や光源の座標と強さ，背景テクスチャなどをランダムに変更したものを各食品 100 枚生成した．生成された 900 枚のうちランダムに選択した 800 枚が学習用で残りの 100 枚はテスト用とした．

3.3.2 セグメンテーションモデル

本研究で用いるセグメンテーションネットワークは Ronnerberger らのネットワーク [4] を採用しており，バックボーンには ResNet34 [5] を使用した．はみ出しなどを少なく，厳密に評価したいので損失関数には Jaccard 係数を使用した．また，学習の際にデータ拡張として，ランダムにノイズ加算，色相彩度の変更を行った．学習結果は図 3-6 のようになった．

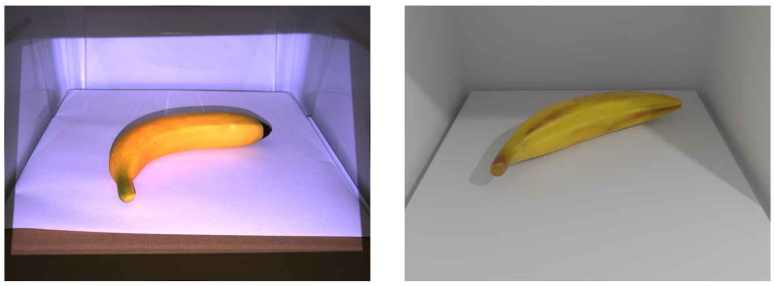

図 3-4: 画角と背景の再現結果画像（左：実画像，右：再現画像）

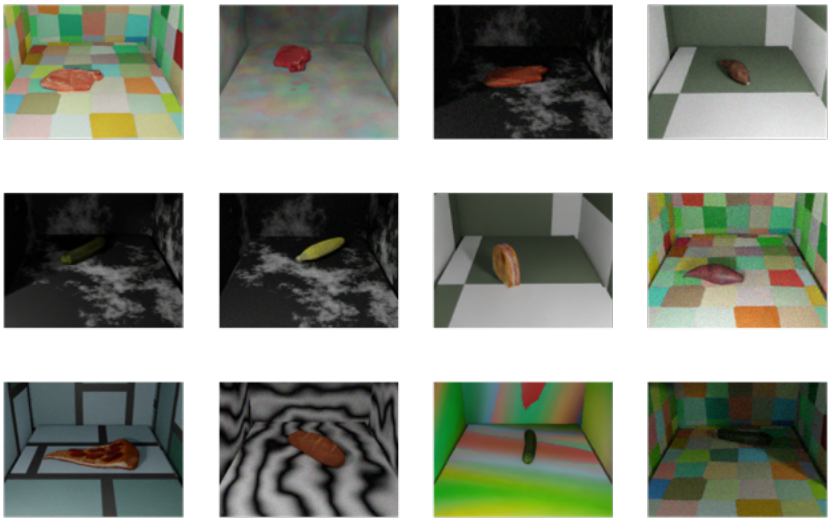

図 3-5: 正解マスク・正解深度付きデータセット例

3.3.3 深度推定モデル

本研究で用いるセグメンテーションネットワークは Hu らのネットワーク [11] をベースに改良された富永らのネットワーク [13] を使用する．またアーキテクチャには ResNet [5] を使用した．

3.4 3 次元再構成

3.4.1 正解深度・正解点群付きデータセット

正解深度・正解点群付きデータセットは単眼 3 次元再構成のモデルの訓練に使用する．

データセットは Blender [14] と trimesh ライブラリを使用して生成した．各視点からの背景透

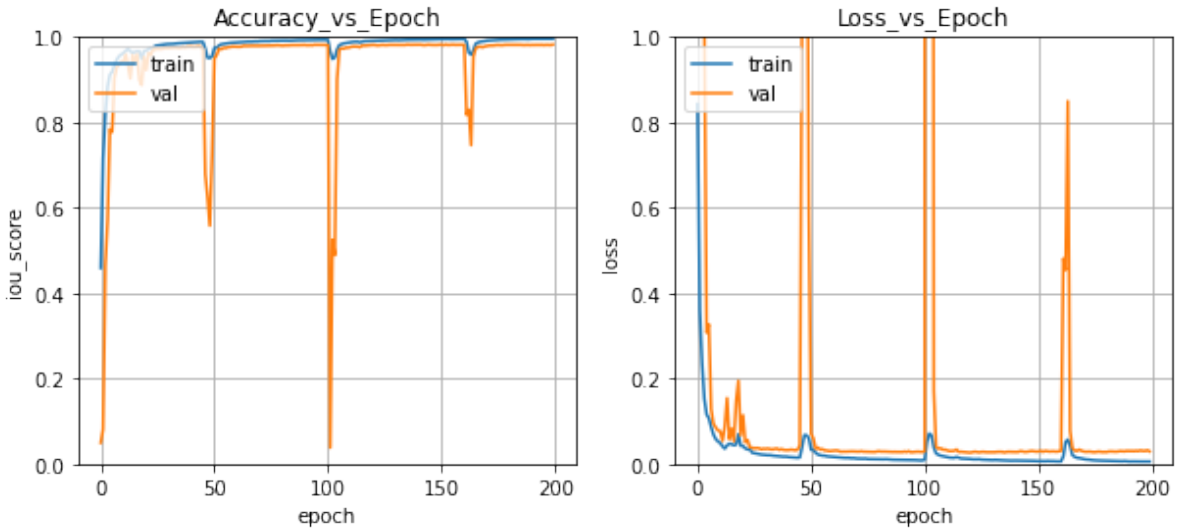

図 3-6: セグメンテーションモデルの学習過程グラフ

過 RGBA 画像と，それに対する 3 次元点群データをもつ．また Blender を用いてカメラから物体までの距離を取得し，アルファチャンネル（不透明度チャンネル）に反転して入力した．これにより図 3-7 のようにカメラに近づくほどアルファ値が高く，遠ざかるほどアルファ値が低くなるようにレンダリングを行った．

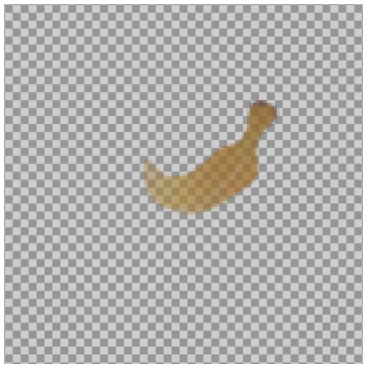

図 3-7: 深度情報をアルファチャンネルに入力した結果画像

バナナ，クロワッサン，梨，桃の 4 種類の食品に対して，回転，大きさ，色合いをランダムに変更した 5 つの状態を用意し，それぞれ方位角 5 度刻みで一周撮影したので，総枚数は 1440 枚になった．

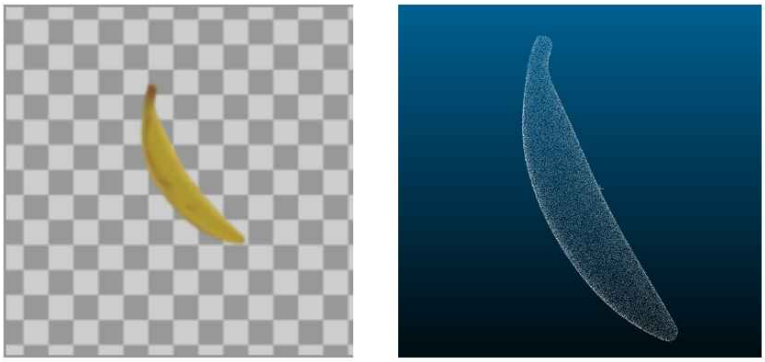

図 3-8: 正解点群付きデータセット例（左：背景透過画像，右：正解点群データ）

3.4.2 3 次元再構成モデル

ベースラインである Wang らの手法 [8] では再構成に深度画像を使用しないが，本研究では深度画像を新しいモダリティとして加えて推定を行いたいため，入力画像のチャンネル数を RGB の 3 チャンネルから RGBA（アルファチャンネルに深度情報を入力しているため実質的には RGBD）の 4 チャンネルに変更した.

第4章　評価実験

4.1　セグメンテーションモデル推定結果

学習させたセグメンテーションモデルを用いて，フォアグラウンドセグメンテーションを行った結果，図 4-2 のようになった．図を見ると，全体的に認識は出来ているが，一部床面を認識している場合や，物体であるはずの部分が物体として認識されておらず，物体の領域に穴が空いている箇所が見られた．原因としては，現実の画像では床面からの照り返しや，強い光による白飛びなどが発生しているのに対して，作成したデータセットではそのような現象は発生しておらず，そういったパターンを学習できていないためだと考えられる．

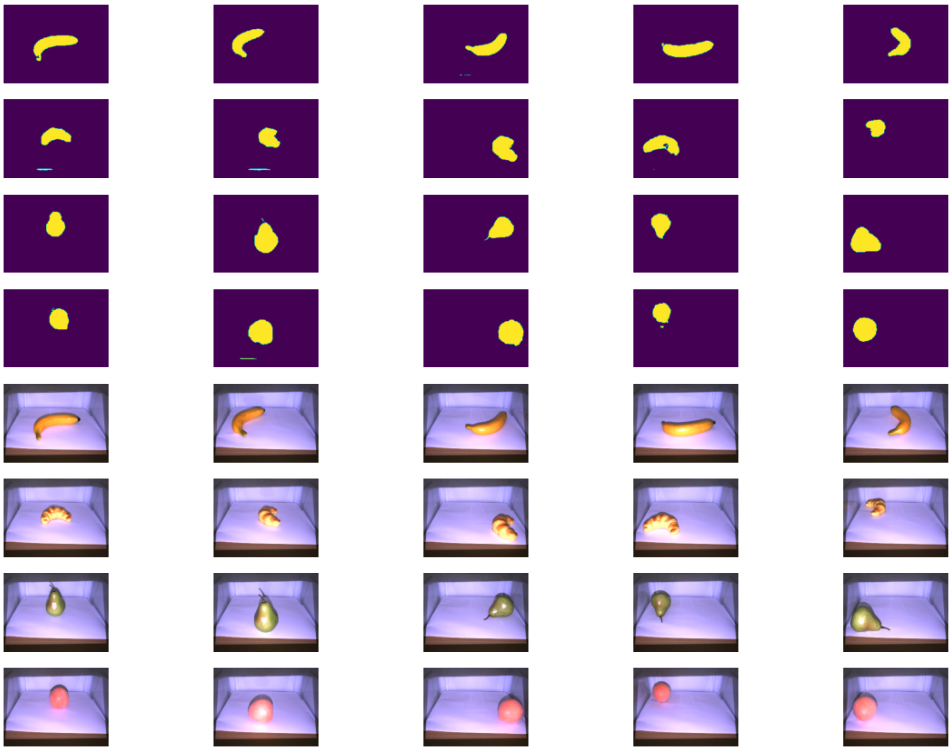

図 4-1: 実画像のフォアグラウンドセグメンテーション結果（上 4 列：セグメンテーション結果，下 4 列: 実画像）

4.2 深度推定結果

学習させた深度推定モデルを用いて，深度推定を行った結果，図 4-2 のようになった．図を見ると，背景の形状はもちろん，物体の中でも深度の深い部分と浅い部分があり，物体の深度が適切に推定できていることが分かる．図 4-2 のセグメンテーション結果とは違い，照り返しや白飛びの影響を受けていない．

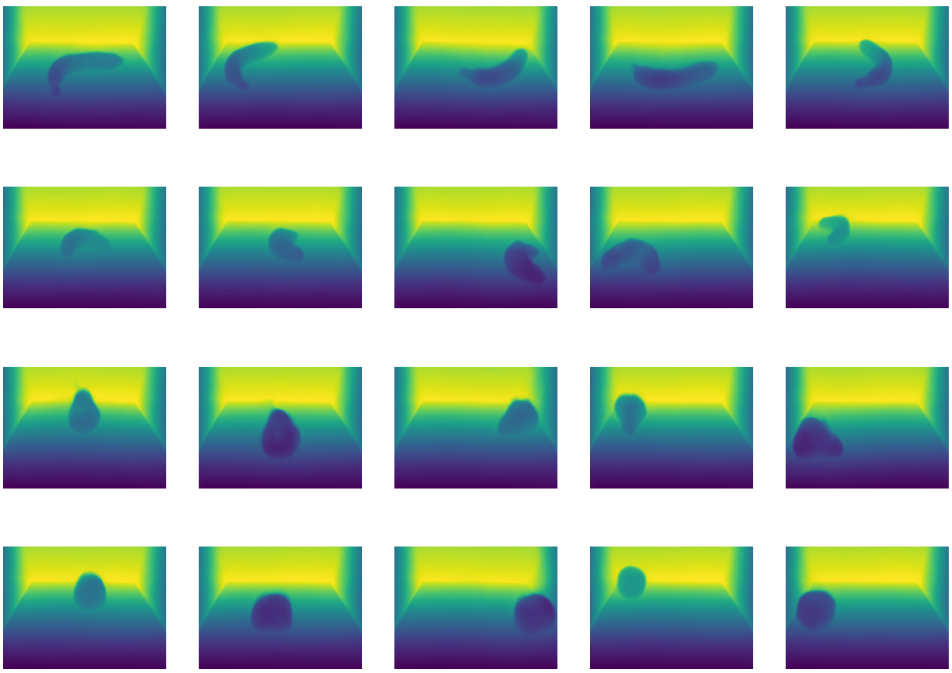

図 4-2: 実画像の深度推定結果

4.3 3 次元再構成評価

3 次元再構成の定量的評価には，3 次元モデルの比較で用いられる以下の評価指標を用いる．

- chamfer loss
- edge loss
- normal loss
- laplacian loss

表 4-1: 提案手法評価結果平均

	chamfer loss	edge loss	normal loss	laplacian loss
バナナ	0.199	0.00358	**0.0221**	0.00454
クロワッサン	0.159	0.00361	0.0320	0.00479
梨	**0.131**	**0.00269**	0.0298	**0.00389**
桃	0.333	0.00320	0.0351	0.00421

評価には PyTorch3D [16] を用いた.

表 4-1 から全ての項目で高い loss や，低い loss をもつ物体はなく，物体ごとの特徴があることが分かった．normal loss を除いて，全体的な精度が最も高かったのが梨だった．実際に図 4-3 の再構成例を見ると，細かな部分は再構成出来ていないが，大まかな形は再構成できているように感じた．精度が高かった原因としてはバナナ・クロワッサン・桃は暖色系の色合いをもつのに対して，梨は寒色系の色合いをもつため，コンピュータが混乱しづらかったのが要因だと考えられる．また，桃は chamfer loss が他の物体に比べて高い結果となった.

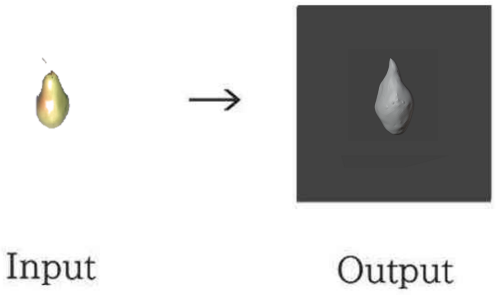

図 4-3: 梨の 3 次元再構成結果例（左：入力画像，右：出力モデル）

表 4-2: バナナの個別評価結果

	chafer loss	edge loss	normal loss	laplacian loss
バナナ 1	0.170	0.00324	0.0109	0.00392
バナナ 2	0.222	0.00357	0.0311	0.00510
バナナ 3	0.310	0.00392	0.0346	0.00488
バナナ 4	0.151	0.00370	0.0235	0.00446
バナナ 5	0.142	0.00347	0.0105	0.00432
平均	**0.199**	**0.00358**	**0.0221**	**0.00454**

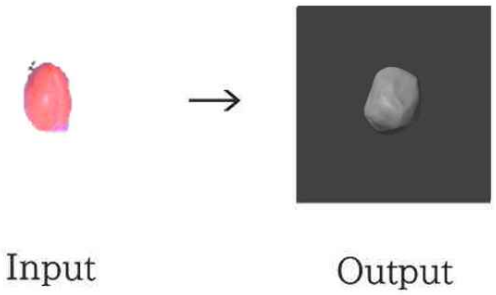

図 4-4: 桃の 3 次元再構成結果例（左：入力画像，右：出力モデル）

表 4-2 で特定の物体について着目してみると，chamfer loss と normal loss は同じ物体でも最小と最大で倍以上の精度の違いが見られた．実際に再構成された物体を見てみると，実画像での被写体の写り方によっては上手く再構成できていない場合があった．これはデータセット内に類似した角度のデータが無く，再構成に失敗したためだと考えられる．

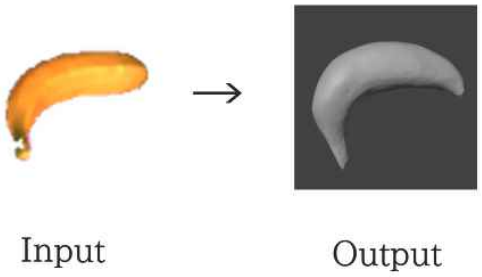

図 4-5: バナナの 3 次元再構成結果例（左：入力画像，右：出力モデル）

表 4-3 から深度情報を含めた場合と同様に，精度が高い項目が多かったのは梨だった．加えて桃の chamfer loss が他の物体と比べて高いのも同様だったので，深度情報の有無に関わらず同様の傾向があることが分かった．

表 4-4 から，chamfer loss を除いて深度情報がある方が精度が高くなった．chamfer loss の精度の差も 0.007 と非常に小さく，深度情報がある方が良い結果を得られたと考えられる．これは深度情報を与えることで，カメラから見えている物体の面に関しては形状の情報が得られるためであると考える．

表 4-3: 深度情報を含めなかった場合の評価結果平均

	chamfer loss	edge loss	normal loss	laplacian loss
バナナ	0.159	0.00405	0.0816	**0.00689**
クロワッサン	0.172	0.00376	**0.1237**	0.00718
梨	**0.113**	**0.00361**	0.1242	0.00772
桃	0.353	0.00503	0.0893	0.00971

表 4-4: 深度情報の有無による結果の比較

	chamfer loss	edge loss	normal loss	laplacian loss
深度情報有り平均	0.206	**0.00327**	**0.0298**	**0.00435**
深度情報無し平均	**0.199**	0.00411	0.1047	0.00787

第5章　結論

本研究では，Blender [14] を基軸とした 3DCG 制作ソフトウェアによって生成したデータセットのみを用いて背景あり画像の 3 次元再構成を行う方法を提案した．

今後の課題として，まずデータセットの拡充が求められる．特に 3 次元再構成には大規模なデータセットが必要だと考える．本研究のベースラインである Wang らの研究 [8] では，3 次元再構成のデータセットとして大規模な 3D モデルデータセットである ShapeNet を使用しており，13 種類の物体に対して計 50,000 個の 3D モデルからレンダリングした画像と点群情報を用いて学習している．一方この研究では 4 種類の物体に対して計 20 個の 3D モデルのみを使用しているため，データセットのバリエーションが不足している．現時点ではデータセットのバリエーションの少なさから，角度によって再構成の精度が大きく影響してしまっているため，全方位からのデータを満遍なく生成することでこのような事態を防ぐことができる．

またレンダリングを行ったレンダラの都合上，光源から直接物体表面に当たる光のみレンダリングしており，照り返しなどの複雑な照明効果が無かったため，物体が置かれている面を物体と誤認識する場合があった．このような点を考慮しつつデータセットを作成すればより良い結果が得られると考える．

また現時点では，実際に電子レンジで調理を行うようなもの，例えば，お椀に入ったスープやお皿に乗っているおかず，冷凍食品など実際に温める機会が多そうなについては形状が複雑なため再構成には至れなかったため，このような食品に対しても再構成を行う必要がある．

謝 辞

本研究を行うにあたり, 数多くの的確で有益な御助言をいただきました， 青山学院大学 理工学部 情報テクノロジー学科 鷲見 和彦 教授， 金子 直史 助教に心より感謝するとともに御礼申し上げます.

また研究の方針や発表面で様々なアドバイスをいただきました， 青山学院大学 理工学研究科 理工学専攻 知能情報コース 加藤 那由多さん，綿引 凌さん，田坂 光司さん，佐久間 絢子さん，菊池 真美さん，田中 亮さん，光林 優菜さん，牟 耕さん，リク テツインさん，Wisan Dhammatorn さんに心より感謝いたします.

そして研究生活を共にした青山学院大学 理工学部 情報テクノロジー学科 伊藤さん，大可さん，片平さん，佐野さん，玉串さん，中島さん，村田さん，森山さん，鈴木さんに感謝いたします.

青山学院大学 理工学部 情報テクノロジー学科

鷲見研究室 遠藤 琳太

令和 5 年 1 月 26 日

参考文献

[1] J. Long, E. Shelhamer and T. Darrell: “Fully convolutional networks for semantic segmentation”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015).

[2] K. He, G. Gkioxari, P. Dollar and R. Girshick: “Mask R-CNN”, IEEE international conference on computer vision, p. 2961–2969 (2017).

[3] V. Badrinarayanan, A. Kendall and R. Cipolla: “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”, IEEE transactions on pattern analysis and machine intelligence, pp. 2481–2495 (2017).

[4] O. Ronneberger, P. Fischer and T. Brox: “U-net: Convolutional networks for biomedical image segmentation”, pp. 234–241 (2015).

[5] K. He, X. Zhang, S. Ren and J. Sun: “Deep Residual Learning for Image Recognition”, Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR ’16, IEEE, pp. 770–778 (2016).

[6] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao: “3D ShapeNets: A deep representationfor volumetric shapes R-CNN”, IEEE Computer Vision and Pattern Recognition, p. 1912–1920.

[7] H. Fan, H. Su and L. J. Guibas: “A point set genera-tion network for 3d object reconstruction from a sin-gle image”, Proc. of IEEE Computer Vision and PatternRecognition, p. 605–613.

[8] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu and Y. Jiang: “Pixel2mesh:Generating 3D Mesh Models from Single RGB Images”, Proceedings of the European Conference on Computer Vision (2018).

[9] H. Kato, Y. Ushiku and T. Harada: “Neural 3d mesh renderer”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018).

[10] D. Eigen, C. Puhrsch and R. Fergus: “Depth map prediction from a single image using a multi-scale deep network”, NIPS, pp. 2366–2374 (2014).

[11] J. Hu, M. Ozay, Y. Zhang and T. Okatani: “Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries”, IEEE Winter Conference on Applications of Computer Vision, pp. 1043–1051 (2019).

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio: “Generative adversarial nets”, Advances in neural information processing systems, pp. 2672–2680 (2014).

[13] 富永, et al：“既知の三次元環境内における物体形状把握”, 青山学院大学理工学部 情報テクノロジー学科 2021 年度卒業研究論文 (2021).

[14] B. O. Community: “Blender - a 3D modelling and rendering package”, Blender Foundation, Stichting Blender Foundation, Amsterdam (2018).

[15] M. Denninger, M. Sundermeyer, D. Winkelbauer, Y. Zidan, D. Olefir, M. Elbadrawy, A. Lodhi and H. Katam: “Blenderproc”, arXiv preprint arXiv:1911.01911 (2019).

[16] N. Ravi, J. Reizenstein, D. Novotny, T. Gordon, W.-Y. Lo, J. Johnson and G. Gkioxari: “Accelerating 3d deep learning with pytorch3d”, arXiv:2007.08501 (2020).