Vol.2022-CVIM-229 No.11
2022/3/10

高周波形状復元のための微分可能レンダラーを用いたデータ拡張の最適化

時枝 康大[1,a)]　岩口 堯史[1]　川崎 洋[1]

概要：構造化光を用いたアクティブステレオ法によるワンショット 3 次元形状復元は，計測結果が疎であり，高周波な形状は復元できないという問題がある．深層学習を用いて画像と疎な形状を入力して高密度な物体表面の形状を推定する手法が提案されているが，学習のためには実際に計測した 3 次元形状データセットから高周波形状を学習に用いるか，CG によるシミュレーションにおいて周波数や振幅などのパラメータを調整して実データに近いデータを作成することが必要になる．本論文では，CG におけるパラメータ調整の問題を，最近提案された深層学習における勾配降下法を用いたデータ拡張の最適化を用いて解消する手法を提案する．その際，従来のデータ拡張手法では，アフィン変換や色変換など基本的な画像処理のみしか適用できなかったのに対して，微分可能レンダラーを用いて 3 次元形状とその配置を変化させてレンダリングすることで，より広範な学習データを作成することができる．実験では，本手法により高周波形状復元のための深層学習が効率化されたことを実データによる実験で確認した．

キーワード：微分レンダラー，3 次元復元，深層学習，データ拡張最適化

1. はじめに

構造化光を用いたアクティブステレオ法による 3 次元復元は，システム構成が簡単で，高精度な形状再構成が可能なことから，その重要性が高まってきている．時間的符号化方式として知られるグレイコードやフェーズシフトが代表的な手法であり，これまで広く利用されてきた．これらの手法は，画素の位置を特定するために複数のパターンを必要とするため，高速に移動する物体を計測することができない．この問題を解決するために，1 つのパターンで識別するためのコードを構成する疎な構造化光に基づいて，1 つまたはいくつかのパターンの投影から形状再構成を実現するワンショット計測法が注目されている．疎な構造化光の共通の問題点は，再構成の分解能が時間的符号化法よりも大幅に低く，疎なパターン領域周辺の形状しか復元できないことである．従来の手法では，移動平均や放射基底関数による補間が一般的であったが，パターン投影のない領域は，たとえ高周波数の形状を持つ表面であっても，低周波数の形状，すなわち滑らかで平坦な表面として再構成されていた．文献 [23] では，構造化光源と単一カメラに基づく RGB-D センサで撮影した，疎な深度情報と密な陰影情報を用いて，密で高周波な形状を復元している．学習データセットには，実現が困難な実環境での高精度なデータ収集ではなく，CG で描画された合成高周波形状画像を用いて深層ニューラルネットワークを学習させることで，高精度な形状復元を実現している．

合成画像を用いることで，大規模なデータセットを用いた学習が可能になる一方で，実際に推定を行うデータセットとのドメイン・特性の違いによる性能悪化が発生する．そこで文献 [23] では，実データを用いた追加学習によりドメイン適応を行い，推定誤差の低減に成功した．しかし，この 2 段階の学習により，最初の学習重みが消滅することが懸念や，実データのドメインに近い合成データを適切に生成するためのデータ拡張のハイパーパラメータを調整する手法は不明であった．

本論文では，限られた実データから効率的に学習データセットを構築することで，高周波形状推定の性能を向上させる．このような目的のために，少数の学習データに対してドメイン適用を実現するデータ拡張が広く用いられているが，ランダムなデータ拡張では，パラメータ数が増加すると学習データ総数が急激に増加するという問題がある．このため，グリッドサーチや勾配降下法によりデータ拡張のためのハイパーパラメータを最適化する方法が研究されており，画像のアフィン変換や色変換を効率よく拡張できることが示されている [19]．提案手法では，これを 3 次

[1] 九州大学
[a)] tokieda.kodai.281@s.kyushu-u.ac.jp

元形状を拡張するために微分レンダラーを用いて，復元すべき実データに対する損失に応じて，勾配降下法により形状パラメータを直接最適化する．微分可能なレンダリングを用いることで，2 次元描画画像による損失を逆伝播することができ，データ拡張パラメータを最適化することができる．

本論文の要点は以下の 2 つである：

- 微分可能なレンダラーを導入することで，形状と位置からなる 3D シーンのパラメータを勾配降下法により最適化し，学習用データセットを自動生成することで効率的にデータ拡張を行うことを可能にした．
- 実データを用いた定量的・定性的な実験を行い，本手法の優位性を示した．

2. 関連研究

2.1 Shape from shading とフォトメトリックステレオ

Shape from shading は古くから研究されている手法であり [14], [27]，また，陰影を利用して正規分布を解析的に計算するフォトメトリックステレオは Woodham らによって提案されて以来，広く研究されている [26]．近年，機械学習を用いた手法がいくつか提案されており，例えば Ikehata らは相互反射がある場合のフォトメトリックステレオのための CNN を提案している [16]．

単一光源環境で撮影された画像から形状を一意に決定できない，bas-relief ambiguity[4] と呼ばれるものがある．Barron と Malik は，物体の特性に対して強い事前分布を課すことで，1 枚の陰影画像から最も可能性の高い形状，照明，反射率を復元する手法 [2] を提案した．Henderson と Ferrari は，形状モデルを生成し，レンダリング画像の陰影を評価することによって，単一の画像から形状，姿勢，陰影を推定する手法 [12] を提案している．

2.2 距離画像の超解像

RGB-D カメラで得られる低解像度の距離画像の解像度を上げる手がかりとして，高解像度の RGB 画像の利用が研究されている．Barron らは fast bilateral filter を用いて，RGB 画像をガイドとすることで深度を伝搬させる方法を提案した [3]．Lutio らは多層パーセプトロンを用いて，ピクセル間のマッピングにより超解像を実現するガイド付き超解像を提案した [18]．これらは教師なし手法であるが，Hui らによって学習型アプローチが提案されている [15]．この手法では，マルチスケールガイドネットワークと呼ばれる CNN がエンコーダで異なるスケールの特徴を抽出し，デコーダでアップサンプリングする．

これらの手法では，自然光の下で撮影した RGB 画像をガイドとしているが，提案手法では，プロジェクター光源で撮影した陰影画像をガイドとする．

2.3 合成画像による学習データ

機械学習，特に深層学習を用いた手法では，十分な量の学習データセットを用意することが重要である．しかし，その手法に必要なデータ，例えば特定のシステムで計測されたデータやアノテーションは十分に存在しないことが多い．また，実環境での撮影が困難な場合もある．この問題を解決するために，合成画像によりデータセットを構築し，学習する手法が近年重要な研究課題となっている．

Weinmann らは，物質分類の学習に bidirectional texture 関数を用いた合成データを使用した [25]．実験では，実環境と同じ設定でレンダリングした画像で学習することで，この課題での分類が可能であることが示されている．Dosovitskiy らは，CNN によるオプティカルフローの学習のための合成画像データセットを作成した [7]．飛行中の椅子という非現実的なシーンで構成されているにもかかわらず，実環境で計測したデータセットで高い推定精度を達成している．Ikehata らは，フォトメトリックステレオ学習用の合成データを用いて，実測で大量に集めることが難しい複雑な陰影を持つ凹形状のデータセットを構築し，学習済みモデルによって実データセットが再構成できることを示した [16]．

2.4 データ拡張の最適化

深層学習におけるデータ拡張は，モデルの性能を向上させるために必須の技術だが，ハイパーパラメータの組み合わせは膨大で，最適な値を設定するためには専門的な知識が必要である．最適なデータ拡張を実現するために，検証データセットと同じ分布にデータを変形することで適切なデータ拡張を達成する GAN ベースの手法 [1], [20], [24] がある．また，AutoAugment[6] や PBA[13] などのその派生型 [10], [17] などでは，モデルの学習を繰り返し，検証データセットでの精度を評価し，性能を向上させるようにデータ拡張のパラメータを調整するため，性能は高いが計算コストが高くなる．一方で，勾配降下法ベースの手法は，データ拡張のパラメータを直接最適化する手法であるため，効率的な学習が期待されており，近年注目されている．CNN とデータ拡張を同時に最適化する手法として，データ拡張の方針の勾配をノイマン級数を用いて近似する MADAO[11] がある．また，Mounsaveng らは，勾配降下法を用いて CNN とデータ拡張の深層ネットワークを同時に最適化する手法を提案した [19]．これらの手法は，クラス分類タスクであるため，画像処理や色変換によるデータ拡張を行う．しかし，我々は陰影を用いた深度推定を目的としており，陰影は 3D シーンと密接な関係があるため，物理的に誤った画像を生成してしまうため，これら方法は適さない．そこで，従来の画像処理だけでなく，形状情報を含む 3D シーンのパラメータに基づくレンダリングにより学習データを拡張する．

図 1 実験のセットアップ.

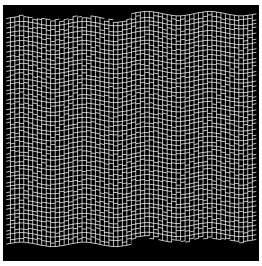

図 2 3 次元復元に用いる疎な格子状パターン.

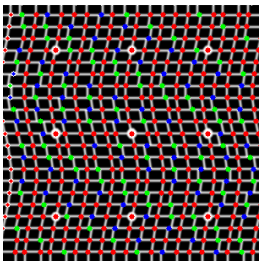

図 3 対応点を検出するためにパターン内に埋め込まれた 3 種類のコード.

3. 陰影情報を用いた深層学習による高周波形状復元

3.1 構造化光を用いたアクティブステレオ法によるワンショット 3 次元復元

ワンショット 3 次元復元手法として，パターンを投影するプロジェクターとカメラで構成されたシステムを用いたアクティブステレオ法 [9] を用いる．計測のセットアップを**図 1** に示す．

プロジェクターで投影されるパターンは，格子状のパターン（**図 2**）を用いており，このパターンは隣接する格子点間の位置関係として埋め込まれたコード（**図 3**）で構成されている．コードを用いて格子点を特定することで，カメラで撮影した画像との対応点を検出し，三角測量により距離を算出する．この計測手法では，パターンを構成する線上の形状のみを計測することができる．この疎な形状を放射基底関数で補間し，低周波の形状を得る．

本論文では，低周波形状については，この計測手法 [9] を用いるが，我々のフレームワークにおける形状計測手法は，これに限定されるものではない．

3.2 深層学習を用いた shape from shading

文献 [23] に従い，陰影情報を用いて，疎な距離画像から密な距離画像を推定する．Shape from shading による，深層学習を用いた距離画像の超解像手法の概要を**図 4** に示す．まず，固定照明下で物体の陰影画像と低周波形状の計測を行う．次に，低周波形状，パターン投影画像，陰影画像を U-net[22] 構造を持つ CNN（**図 5**）に入力して物体表面の高周波形状を推定し，測定した低周波形状に高周波成分を加えることで高精度な形状を復元する．深層学習を用いた手法であるので，学習には多くのデータが必要であるが，多様な形状の物体の測定には時間がかかり，また，真値となる高周波形状の測定も難しいという課題がある．そこで，以下の式のように，複数の正弦波を組み合わせて高周波形状を生成した．

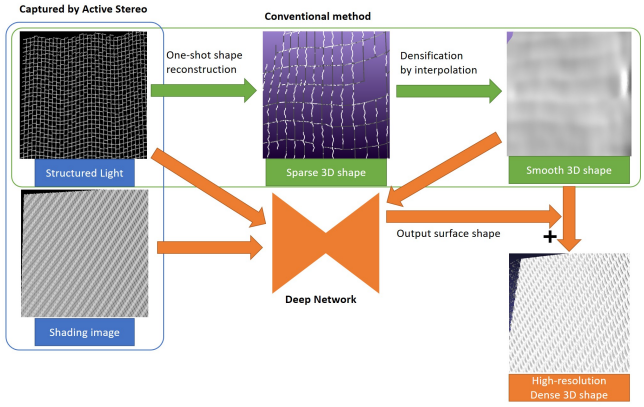

図 4 深層学習を用いた shape from shading．まず，パターン投影画像と陰影画像を撮影する．パターン投影画像から疎な形状を復元し，補間して低周波形状とする．2 枚の画像と低周波形状をネットワークに入力し，高精度な形状を得る．

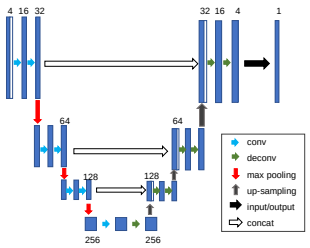

図 5 ネットワーク構造.

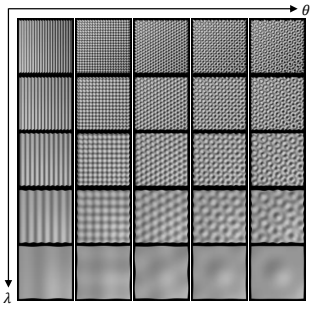

図 6 正弦波を組み合わせた高周波形状の生成.

$$I(x,y) = \sum_{i=0}^{N} \alpha_i \cos\left(2\pi x' \lambda_i + \psi_i\right), \tag{1}$$

$$x' = x\cos\theta_i + y\sin\theta_i \tag{2}$$

ここで，α は振幅，λ は波長，ψ は位相であり，ランダムに選択される．N は重ね合わせる波の数である．パラメータは，目的の実データに近い形状を得るために手動で調整する．規則的に変化する θ と λ を持つ形状生成の例を**図 6** に示す．このシーンは，実環境での計測と同じ内部パラメータを持つカメラのビューからラスタライズアルゴリズムでレンダリングされる．

合成画像は実環境の画像とは異なる特性を持つため，合成データのみで学習したモデルは実データに活用しにくいとされている．そこで，実データの入力に対応するために，

CNN の重みを fine-tuning するための追加学習も行う．この追加学習では，数枚の実データのみでネットワークを学習させ，全層の重みを更新する．

4. 提案手法

4.1 微分可能レンダラーを用いた 3 次元形状のデータ拡張

実環境におけるアクティブステレオ法による計測データを学習するために，陰影画像，パターン投影画像，距離画像を微分可能レンダラーにより生成する．陰影は光源や物体配置の影響を強く受けるため，従来の画像処理によるデータ拡張では物理的に正しくない画像が生成されてしまう．そこで，様々なシーンパラメータに基づいた 3 次元シーンをレンダリングすることで，物理的に正しいデータ拡張を行う．

高周波の形状は，波の周波数，振幅，角度などによって決定される．１つの高周波の 3 次元形状を用意し，この形状を拡大縮小や回転などで変形させることで，様々な高周波形状を作り出すことができる．具体的には，形状を X 軸と Y 軸方向に同じ倍率で拡大縮小させると周波数が，Z 軸方向に変形させると振幅が，Z 軸で回転させると角度が変化する．このように，3 次元形状を変形した後にレンダリングを行うことで，物理的に正しいデータ拡張を行うことができる．

また，勾配降下法を用いた最適化を行うために，微分可能なレンダリング処理によって画像を生成する必要がある．微分可能なレンダラーとして PyTorch3D[21] を用い，アクティブステレオ法の合成データを作成するために，PointLights を修正してプロジェクターライトを実装した．レンダリングした陰影画像と構造化光パターン（図 2）を投影した画像を**図 7** に示す．プロジェクターで光を投影しているため，物体の形状や角度によって陰影や投影されたパターンが変化していることが分かる．

図 8 は拡大縮小と回転を適用してデータ拡張した例であり，周波数や角度によって見え方が大きく変わるため，基本的な画像処理が適用できないことが分かる．この方法は，これらの変形に限らず，x 軸や y 軸の回転，光源の位置，3 次元形状などを変えることで様々なデータ拡張に対応でき，さらにレンダリング後の画像に色調整などの基本的な画像処理によるデータ拡張も適用することができる．

4.2 Bi-level optimization によるデータ拡張の最適化

クラス分類タスクにおいて，画像分類のための CNN のパラメータを θ，データ拡張する augmenter（パラメータ ω）を A とする．内部ループでは，学習データ X_T を A で変形し，損失 f を最小化するように θ を更新する．また，外部ループでは，検証データ X_V において損失 g を最小化するように ω を更新する．このような，CNN とデータ拡張の同時最適化は以下のように書ける：

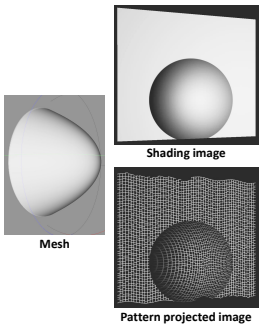

図 7 PyTorch3D を用いたレンダリング．

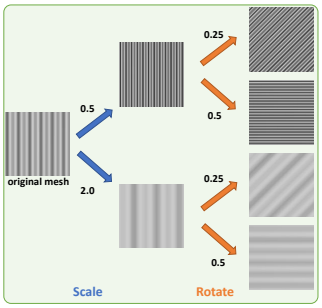

図 8 3 次元形状のデータ拡張の例．

$$\theta^* = argmin\, f\left(A_\omega\left(X_T\right), \theta\right) \tag{3}$$

$$\omega^* = argmin\, g\left(X_V, \theta^*\right) \tag{4}$$

Mounsaven らは，データ拡張において微分可能な画像処理を用い，augmenter として深層ネットワーク（augmentation network）を用い，また，勾配により式 (3) と式 (4) を近似することで CNN と augmentation network を同時に最適化する online bilevel optimization 手法を提案している [19]．

このデータ拡張の最適化手法に，微分可能な 3 次元形状変形処理と微分可能なレンダラーを用いたデータ拡張を適応することで，合成データを用いた 3 次元形状推定における深層学習のデータ拡張の最適化手法を提案する．提案手法のアルゴリズム概要を**図 9** に示す．Augmentation network はノイズベクトルからパラメータを出力し，レンダラーはそのパラメータと 3 次元形状を用いて拡張した画像を生成する．内部ループでは，augmentation network とレンダラーにより学習データを生成し，学習データで CNN を学習する．外部ループでは，検証データで損失を計算し，augmentation network の重みを更新する．学習データ X_T と式 (3) は以下のように表される．

$$X_T = D\left(A_\omega\left(\varepsilon\right), m\right) \tag{5}$$

$$\theta^* = argmin\, f\left(D\left(A_\omega\left(\varepsilon\right), m\right), \theta\right), \tag{6}$$

ここで，CNN のパラメータを θ，A（パラメータ ω）を augmentation network，D を微分可能レンダラー，m を 3 次元形状のメッシュ，ε をノイズベクトルとする．

4.3 実装

Augmentation network と bi-level optimization を用いた学習処理は，[19] で使用されているコードを参照して PyTorch で実装し，3 次元形状の変形とレンダリングは PyTorch3D で実装した．3 次元形状の変形には，5 つのパラメータ（x・y 軸についての拡大縮小，z 軸についての拡大縮小，x・y・z 軸それぞれについての回転）を用い，レンダリング後の画像には，明るさとコントラストを変える 2 つのパラメータを用い，合計 7 つのパラメータを用いた．

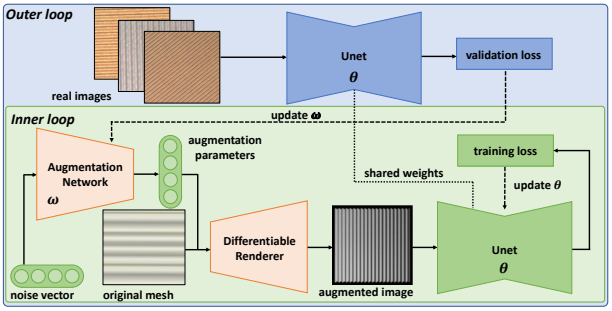

図 9 アルゴリズムの概要.
内側のループでは，augmentation network がノイズベクトルからパラメータを出力し，微分可能レンダラーがパラメータと 3 次元形状から拡張された画像を出力し，CNN を学習させる．外側のループでは，内側のループと同じ重みを持つ CNN を使用して，検証データにおける損失により augmentation network の重みを更新する．

また，レンダリング画像には，実データと合成データの見た目の違いをなくすために，パーリンノイズを適用した．

Augmentation network は多層ニューラルネットワークであり，最適化アルゴリズムは Adam[8] を用い，学習率は 0.01 に設定した．CNN の最適化アルゴリズムは SGD を用い，学習率は 0.01，momentum は 0.9 に設定した．

5. 実験

5.1 セットアップ

Unet の学習では，波の数は $N = 1$ に設定し，高周波形状の周波数，振幅，角度などの形状パラメータ，物体の X 軸や Y 軸における角度の姿勢パラメータ，レンダリング画像の明るさやコントラストのパラメータの合計 7 パラメータを変化させて拡張データを生成した．光源の位置と方向は実環境と同じ設定に固定し，1024 × 1024 の陰影画像とパターン投影画像をレンダリングし，同時に得られた深度を Ground Truth とした．画像を 128 × 128 のパッチに切り出し，背景のないパッチを学習用として使用した．

実環境で撮影する物体として，高周波形状を持つダンボールについては，周波数の異なる 3 種類（A，B，C）を用意し，より複雑な表面形状を持つ物体を 3 種類（靴，メッシュ，バスケット）用意し，合計 6 種類の物体を様々な距離や角度で計測を行った．解像度 1024 × 768 のパターンをプロジェクターで投影し，解像度 1200 × 1200 の画像をカメラで撮影した．実環境で計測したデータは，学習データセットとテストデータセットに分けられる．学習データセットには，ダンボール A が 16 枚，B が 4 枚，そして複雑な形状は，靴が 4 枚，メッシュが 2 枚，バスケットが 1 枚の合計 27 枚が含まれている．テストデータセットには，ダンボール B と C が 12 枚，靴が 4 枚，メッシュが 6 枚，カゴが 2 枚の合計 24 枚が含まれており，ダンボール A は含まれていない．学習や fine-tuning には，解像度 1200 × 1200 の画像を解像度 120 × 120 のパッチに分割し，0.5〜1.5 の倍率で画像の輝度値をランダムに拡張してネットワークに入力し，元の解像度で推論処理を行う．

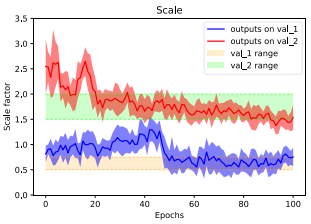

図 10 スケールパラメータに関する augmentation network の出力.

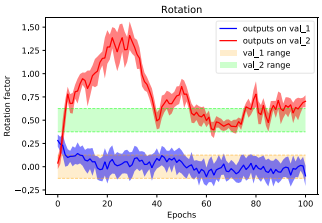

図 11 回転パラメータに関する augmentation network の出力.

定量的な評価としては，GT に対する推定値の RMSE（Root Mean Square Error）を算出する．陰影を利用して物体の形状を推定するため，凹凸のスケールや bas-relief ambiguity[4] として知られている凹凸の不確定性が問題になることがある．このスケールの不確定性に対処するため，出力パッチの平均と標準偏差を GT のパッチと一致するように調整して評価を行うことで，凹凸が推定できているか正しく評価できるようにした．この計算のためのパッチサイズは，すべてのデータで高周波形状の 1 周期が完全に含まれるように 49 × 49 にした．

5.2 データ拡張の最適化

まず，合成画像において，augmentation network が目標とするデータセットのドメインに適合するパラメータを推定できるか確認する実験を行った．目標とするデータセットとして，スケールと回転についてパラメータの範囲の異なる 2 つのデータセット（val_1 と val_2）を用意した．スケールのパラメータについては，X 軸と Y 軸において，この値を用いて同じ倍率で拡大縮小を行った．val_1 は 0.5〜0.75 の範囲のパラメータを持ち，val_2 は 1.5〜2.0 の範囲のパラメータを持つ．回転のパラメータについては，この値に π を乗算した値を用いて z 軸に対して形状を回転させた，つまり，波の角度を変化させた．val_1 は-0.125〜0.125 の範囲，val_2 は 0.375〜0.625 の範囲のパラメータを持つ．これらのデータセットを検証用データセットとして使用した場合の学習時における，augmentation network の出力の変化を可視化した．**図 10** はスケールについての結果を示し，**図 11** は回転についての結果を示す．出力値の平均と標準偏差を示し，検証用データセットにおけるパラメータの範囲を帯状に示す．Augmentation network は，これらの検証用データセットの領域に近いパラメータを出力できていることが分かる．

次に，提案する augmentation network を用いた 3 次元データ拡張の最適化が，Unet の学習において，どの程度学習を効率的にするかを確認する実験を行った．**図 12** は，val_1

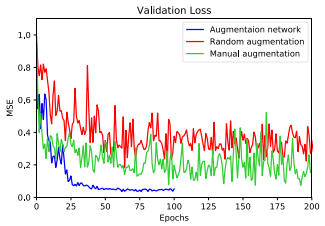

図 12 合成データにおける損失グラフの比較.

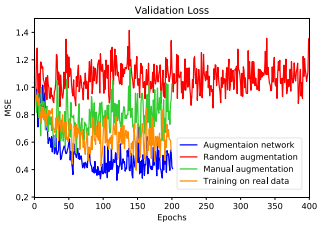

図 13 実データにおける損失グラフの比較.

を検証用データセットとして，3 つの学習過程の損失を比較したグラフである．提案手法を適用した Augmentation network は，損失が早い段階で収束しており，高い性能であることを示している．一方，Random augmentation は，ランダムにデータ拡張して学習した結果であり，200 エポックまで学習しても提案手法の性能に達していないことが分かる．Manual augmentation は，パラメータを手動で調整してデータ拡張した結果であり，提案手法に近い性能ではあるが，効率的な学習を行うためには，さらにパラメータを調整する必要があることが分かる．

また，検証用データセットとして実データを用いた実験も行った（**図 13**）．高周波形状のダンボールの実データを学習データと検証用データに分け，この検証用データを用いた 4 つの学習過程において，その結果を比較した．データ拡張では，波の周波数，振幅，角度，物体の X 軸と Y 軸の回転，明るさ，コントラストという 7 つのパラメータを用いた．Random augmentation と Manual augmentation は，データ拡張によって合成データを作成し，パーリンノイズを加えた画像を用いて学習させた結果である．Train on real data は，画像処理やノイズは加えず，実データの学習データのみで学習させた結果である．これらの結果から，Random augmentation と Manual augmentation では損失が下がらず，合成データに過学習していることが示唆される．より精度の高いパラメータ調整や過学習を抑制するための手法が必要であると考えられる．また，実データを用いた学習も性能が悪く，学習用の多様で膨大なデータセットを用意する必要があると言えるが，実環境での計測でデータセットを用意するのはコストがかかるという問題がある．一方，提案手法の性能が最も良く，これはパラメータを適切に調整できた結果であり，データ拡張における専門的な知識に基づく調整が不要になるとともに，合成データによる大規模な学習が可能であることを示している．この後の実験では，合成データの学習モデルは，この学習済みモデルを用いた．

5.3 実データにおける形状推定

本実験では，合成データで学習したモデル，実データで学習したモデル，および fine-tuning したモデルの 3 つのモデルを異なるテストデータで比較した．合成データでのモデルは，データ拡張の最適化処理で学習させ，実データでのモデルは，実データを用いて 50 エポックで学習させ，また，fine-tuning モデルは，合成データで事前学習させたモデルを実データを用いて 50 エポックで再学習させた．それぞれのモデルについて，検証損失が最も小さくなった時点でのモデルを用いて評価を行った．ダンボールのテストデータにおける RMSE を**表 1**，複雑な形状を持つ物体のテストデータにおける RMSE を**表 2** に示す．Low-res は [9] での計測結果を示す．

	Test data (RMSE[mm])		
	Cardboard-B	Cardboard-C	Avg.
Low-res	0.411	0.832	0.622
Synthetic data	0.347	0.544	0.446
Real data	0.275	0.332	0.304
Fine-tuning	**0.221**	**0.284**	**0.253**

表 1 高周波形状を持つダンボールでのテスト結果. 異なる訓練データにおける性能の比較. Low-res は [9] での計測結果を示す.

	Test data (RMSE[mm])			
	Shoes	Mesh	Basket	Avg.
Low-res	1.572	2.042	**0.617**	1.410
Synthetic data	1.686	2.811	0.784	1.760
Real data	1.352	1.928	0.775	1.352
Fine-tuning	**1.138**	**1.042**	0.649	**0.943**

表 2 複雑な形状を持つ物体でのテスト結果. 異なる訓練データにおける性能の比較. Low-res は [9] での計測結果を示す.

まず，合成データでのモデルの性能を評価する．ダンボールのテストデータにおいては，Low-res よりも RMSE は小さくなっているが，靴などのより複雑な形状では改善されていない．これは，ダンボールは合成データに近い高周波の形状をしているためであり，合成データで学習したモデルは，ドメインが近い形状を推定する能力を持つが，複雑な形状には対応できないことが分かる．

次に，合成データで学習させたモデルや実データで学習させたモデルと比較することで，fine-tuning モデルの性能を評価する．ダンボールや，より複雑な形状のどちらにおいても，合成データや実データのみで学習したモデルに比べて，性能が向上した．バスケットでは若干の改善が見られるが，靴とメッシュでは，性能が大きく改善された．バスケットはデータセットに含まれるデータ数が少なく，また他の複雑な形状に比べ凸凹の高さが小さいため，効果が低いケースであると思われる．全体として，fine-tuning を行ったモデルは，実データで学習したモデルを上回っているため，合成データでの大規模な学習が実データでの正確な推定に寄与していることが示された．

また，**図 14** はダンボールのテストデータにおける Random augmentation と Augmentation network の比較であ

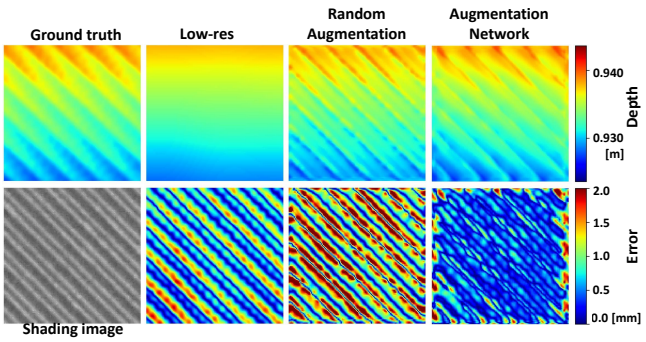

図 14 ダンボールにおける結果．Random augmentation と Augmentation network の比較．上段は深度マップ、下段は陰影画像と誤差マップを示す．提案手法の結果は GT と同様の表面形状を復元し，誤差が小さくなっている．

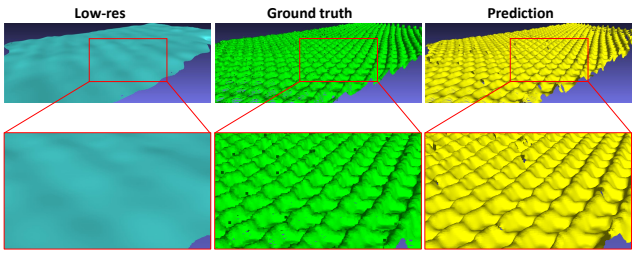

図 15 メッシュにおける 3 次元復元形状．

る．上の段は，GT，Low-res，推定結果の深度マップを示し，下の段は，陰影画像と各深度における GT との誤差マップを示す．誤差マップから，本手法が高い精度で形状を復元できていることが確認できる．さらに，メッシュデータにおける 3 次元形状（**図 15**）から，Low-res では復元できていない物体表面の形状が復元できていることが視覚的に確認できる．このように，本手法では従来の手法では復元できなかった形状を，高精度に復元できていることが分かる．

5.4 他の手法との比較

さらに，[18] や [5] といった，距離画像における最新の超解像手法との比較実験を行った．本実験では，実データから切り出した 512 × 512 の領域を使用した．[18] では，ガイド画像として陰影画像を与え，低解像度の距離画像として [9] での計測結果を与えた．[5] では，RGB 画像として陰影画像を与え，疎な深度情報として [9] での計測結果を与え，疎な計測パターンとしてパターン投影画像を与えた．その結果を**表 3** に示す．ダンボールとメッシュについて，本手法が他の手法を大きく上回り，平均でも他の手法を上回った．その理由として，他の手法では自然光でのデータを前提としているが，アクティブステレオ法による撮影データでは光源位置と形状によって陰影が変化するため，陰影と形状の関係を正しく扱えていないことが考えられる．一方で，提案手法では，物体表面の形状を陰影画像からネットワークで推定するアルゴリズムにより，陰影と

	Test data (RMSE[mm])				
	Cardboard	Shoes	Mesh	Basket	Avg.
Low-res	0.604	2.094	2.300	0.571	1.235
[18]	0.645	1.854	2.702	0.648	1.299
[5]	0.495	**1.410**	2.222	**0.556**	1.036
Ours	**0.195**	1.824	**1.109**	0.610	**0.791**

表 3 他の最新手法との比較．[18] では，陰影画像をガイド画像として使用し，[5] では，構造化光パターン画像をスパース深度パターンとして使用した．

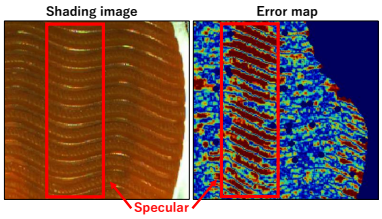

図 16 鏡面反射による推定の失敗例．

形状と関係を適切に学習することができ，また，合成データによる大規模な学習により，1 枚の陰影画像からでも十分に凹凸を推定することができた結果であると考えられる．

5.5 今後の課題

実環境での計測データには，素材によっては鏡面反射が発生するものもある．本手法は陰影から形状を推定するため，**図 16** のような強い鏡面反射があった場合，推定を失敗することがある．これは，学習時の合成データにおいて，拡散反射の素材でレンダリングを行っているためであり，鏡面反射を起こす素材を含む多様な素材で合成データを生成することで解決を図る予定である．

また，提案したデータ拡張手法では，１つの高周波形状を変形することで様々な高周波形状を生成している．より複雑な形状に対応するために，変形させる形状をより複雑な形状にすることや，形状を重ね合わせることでより複雑な形状を生成することで本手法の性能向上を図る予定である．

6. おわりに

本論文では，微分可能なレンダラーと 3 次元形状変形を用い，勾配を用いた bi-level optimization を導入することで，3 次元形状推定におけるデータ拡張の最適化手法を提案した．本手法は，ディープラーニングによる Shape from Shading 手法での学習効率を向上させ，アクティブステレオ法による計測結果の精度を向上させた．さらに，深度推定においても高い性能を達成し，他の最先端手法に対する本手法の優位性を示した．

謝辞

本研究は JSPS 科研費 JP20H00611，JP21H01457，20K19825 の助成を受けたものである。

参考文献

[1] Antoniou, A., Storkey, A. and Edwards, H.: Augmenting image classifiers using data augmentation generative adversarial networks, *International Conference on Artificial Neural Networks*, Springer, pp. 594–603 (2018).
[2] Barron, J. T. and Malik, J.: Shape, Illumination, and Reflectance from Shading, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 8, pp. 1670–1687 (2015).
[3] Barron, J. T. and Poole, B.: The Fast Bilateral Solver, *ECCV* (2016).
[4] Belhumeur, P. N., Kriegman, D. J. and Yuille, A. L.: The bas-relief ambiguity, *International journal of computer vision*, Vol. 35, No. 1, pp. 33–44 (1999).
[5] Chen, Z., Badrinarayanan, V., Drozdov, G. and Rabinovich, A.: Estimating Depth from RGB and Sparse Sensing, *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
[6] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. and Le, Q. V.: Autoaugment: Learning augmentation strategies from data, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 113–123 (2019).
[7] Dosovitskiy, A., Fischer, P., Ilg, E., Häusser, P., Hazirbas, C., Golkov, V., v. d. Smagt, P., Cremers, D. and Brox, T.: FlowNet: Learning Optical Flow with Convolutional Networks, *ICCV* (2015).
[8] Duchi, J., Hazan, E. and Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization., *Journal of machine learning research*, Vol. 12, No. 7 (2011).
[9] Furukawa, R., Morinaga, H., Sanomura, S., Tanaka, S., Yoshida, S. and Kawasaki, H.: Shape acquisition and registration for 3D endoscope based on grid pattern projection, *ECCV* (2016).
[10] Hataya, R., Zdenek, J., Yoshizoe, K. and Nakayama, H.: Faster autoaugment: Learning augmentation strategies using backpropagation, *European Conference on Computer Vision*, Springer, pp. 1–16 (2020).
[11] Hataya, R., Zdenek, J., Yoshizoe, K. and Nakayama, H.: Meta approach to data augmentation optimization, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2574–2583 (2022).
[12] Henderson, P. and Ferrari, V.: Learning single-image 3D reconstruction by generative modelling of shape, pose and shading, *International Journal of Computer Vision*, pp. 1–20 (2019).
[13] Ho, D., Liang, E., Stoica, I., Abbeel, P. and Chen, X.: Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules (2019).
[14] Horn, B. K. P.: *Obtaining Shape from Shading Information*, pp. 115–155, Winston, P. H. (Ed.) (1974).
[15] Hui, T.-W., Loy, C. C., and Tang, X.: Depth Map Super-Resolution by Deep Multi-Scale Guidance, *ECCV*, pp. 353–369 (2016).
[16] Ikehata, S.: CNN-PS: CNN-based photometric stereo for general non-convex surfaces, *ECCV*, pp. 3–18 (2018).
[17] Lim, S., Kim, I., Kim, T., Kim, C. and Kim, S.: Fast autoaugment, *Advances in Neural Information Processing Systems*, Vol. 32, pp. 6665–6675 (2019).
[18] Lutio, R., D'Aronco, S. and Wegner, J.: Guided Super-Resolution as a Learned Pixel-to-Pixel Transformation, *ICCV* (2019).
[19] Mounsaveng, S., Laradji, I., Ben Ayed, I., Vázquez, D. and Pedersoli, M.: Learning data augmentation with online bilevel optimization for image classification, *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1691–1700 (2021).
[20] Ratner, A. J., Ehrenberg, H. R., Hussain, Z., Dunnmon, J. and Ré, C.: Learning to compose domain-specific transformations for data augmentation, *Advances in neural information processing systems*, Vol. 30, p. 3239 (2017).
[21] Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W.-Y., Johnson, J. and Gkioxari, G.: Accelerating 3D Deep Learning with PyTorch3D, *arXiv:2007.08501* (2020).
[22] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *MICCAI*, pp. 234–241 (2015).
[23] Tokieda, K., Iwaguchi, T. and Kawasaki, H.: High-Frequency Shape Recovery from Shading by CNN and Domain Adaptation, *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 3672–3676 (2021).
[24] Tran, T., Pham, T., Carneiro, G., Palmer, L. and Reid, I.: A bayesian data augmentation approach for learning deep models, *arXiv preprint arXiv:1710.10564* (2017).
[25] Weinmann, M., Gall, J. and Klein, R.: Material classification based on training data synthesized using a BTF database, *ECCV* (2014).
[26] Woodham, R. J.: Photometric Method For Determining Surface Orientation From Multiple Images, *Optical Engineering*, Vol. 19, No. 1, pp. 139 – 144 (1980).
[27] Zhang, R., Tsai, P.-S., Cryer, J. E. and Shah, M.: Shape-from-shading: a survey, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 21, No. 8, pp. 690–706 (1999).