

## AI による三次元形状予測の基礎的検証

足利大学 大学院 工学研究科 情報・生産工学専攻 廣川研究室 修士1年

小島 龍輝（こじま りゅうき）

【概要】本研究では、LiDAR などの高価なデプスカメラでしか計測できない三次元形状を、AI により通常のカメラ映像から予測する手法の基礎的な検討を行った。今回の研究で AI の学習に使用した GPU 機材はコンシューマ向けのものであり、小規模な学習モデルでも特徴的な形状を視認できる精度を得ることができた。これを発展させることで、少ない学習コストで高精度な予測モデルが作成可能であると考えられる。

【栃木を元気にするには】様々な現場で三次元形状をデジタル化し、活用する機会が増えていくと予想される中で、高価なデプスカメラを安価なカメラで代用できれば、DX 化を推進することができる。本研究で開発した手法は、発展させることで小さなコストで三次元形状のデジタル化や、ドローンなどの衝突回避など様々な分野に応用が可能である。

### 1. はじめに

三次元形状のデジタル化は高価なデプスカメラが必要であり、様々な現場で DX 化を推進する上での課題となっている。また、AI により通常のカメラ画像から三次元形状を推論する手法が研究されているが、高価な GPU(画像を処理する装置)が必要となるのが現状である。

本研究では様々な現場で DX 化を推進するため、安価なカメラと GPU で三次元形状を推論する手法を提案し、基礎的な検証を行なった。

### 2. AI による三次元形状推論

デプスカメラを使用して撮影した RGB 画像と奥行き画像のペアを用意し、AI に読み込ませて学習させる。その後、AI に RGB 画像のみを与えて奥行きを推論させる(図 1)。AI には敵対的生成ネットワーク(以下、GAN)を用いた。GAN は偽物のデータを作り出す生成 AI と、入力されたデータが偽物か本物かを見分ける識別 AI を組み合わせ、互いに競わせて学習を行うことで精度の高い偽物データを作り出す手法である。偽札を精巧に作るうとする偽造団と、それを見分けようとする鑑定団のイタチごっこをさせることで、高精度な偽物を生成するようなくみである。

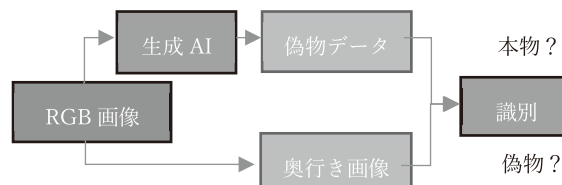


図 1 敵対的生成ネットワーク

### 3. 提案手法

本研究では GAN を対象として、AI の学習時に一度に読み込む画像データの量(以下、ミニバッチサイズ)を減らした上、画像の特定の色に着目する Self-Attention 手法(以下、SA)を取り入れることで GPU の使用メモリ量削減と推論精度を向上させる手法を提案する。

ミニバッチサイズは大きな学習データの塊を、いくつかの断片に分割し、一度の学習データ量を指定するものである。ミニバッチを小さくする程、GPU のメモリ使用量は減少するが、学習精度が悪化する傾向がある。SA は注目すべきデータの特徴を AI に発見させ、その特徴を活用することで精度を向上させる。本研究では RGB の三色のうちどの色に着目するかを AI に発見・活用させ、推論精度を向上させる。

### 4. 実験方法

デプスカメラとして Intel 社の D435、AI の学習と推論にはコンシューマ向けの安価な GPU (RTX2060

12GB)を用いた。

提案手法の有効性を検証するため、既存手法と提案手法の結果を比較する。既存手法は①ミニバッチサイズが 32、②ミニバッチサイズが 64 の 2 つの条件での結果を求めた。提案手法は③ミニバッチサイズを 16 にした。なお、②及び③は GPU のメモリが不足するため、小数点の精度を部分的に半分に低下させる混合精度を用いた。

## 5. 実験結果

図 2 の Real は RGB 画像、Depth は奥行き画像、下側は学習させた 3 つのモデルに動画を推論させ、1 フレームを抜粋した上で色情報の差分を取ったものである。Batch32 は既存手法の①バッチサイズ 32、Batch64 は②バッチサイズ 64、Batch16+SA は提案手法で③バッチサイズ 16 の結果である。白に近いものが正解画像に近いことを意味している。3 つのモデルの生成画像を比較すると、バッチサイズ 32 と 64 のものは見た目上の差異はあまりみとめられないが、SA を用いたものは変化している。これは SA によって着目する色に変化が生じたことによって生成画像に差が生じたと考えられる。

表 1 は推論結果の精度比較であり、2 つの評価指標 Mean Absolute Error (以下、MAE)と Binary Cross-Entropy (以下、BCE)を用いた。どちらの指標も画像のピクセルあたりの色に着目し、正解画像と偽画像との差を誤差として表現しており、値が小さいほど誤差が少ないことを表す。ミニバッチ 32 とくらべ、SA を組み合わせることでバッチサイズを 16 へと半減させても MAE 及び BEC が減少し、ピクセルあたりの変換精度は向上することがわかった。

## 6. 考察

検証の結果として図 2 を参照すると少ないデータセットと、軽量モデルで学習を行ったことを考慮すると輪郭などの形状を確認することができる精度が得られたことはさらなる発展性を感じるものとなった。しかし、遠方の形状については学習不足と考えられる着色のブレが多く見受けられる結果となった。低解像度

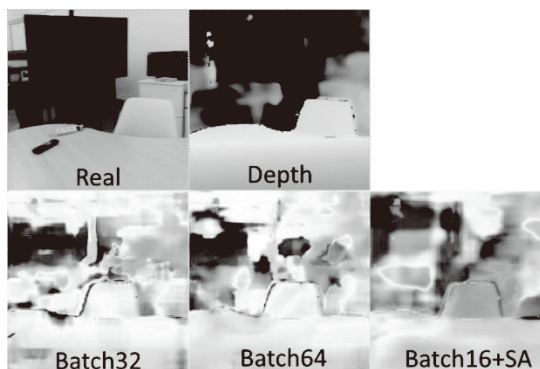


図 2 推論結果

表 1 検証データの BEC と MAE

モデル/ロス	BEC	MAE	BEC+MAE
ミニバッチサイズ32	0.6262	20.1406	20.7668
ミニバッチサイズ16+SA	0.6332	18.0735	18.7067

画像を用いたことによるものであると考えられる。

これは AI モデルを改良し、高解像度画像を使用することで改善可能であると考えられる。

## 7. まとめ

今回の検証から、コンシューマ向けの GPU でも画像の輪郭や特徴的な形状が確認可能な精度の学習が行えることがわかった。精度改良の検証としてバッチサイズの変更や SA 手法などの検証を行ったが、傾向としてバッチサイズを変更するだけで精度に差が生まれることがわかった。発展として、この結果を元にさらなるパラメータの調節を行い、データセットの拡張や、GPU を高性能なものに変更するなどの多くの改良の余地があり、精度の向上が見込める。

本検証により低コストで学習可能かつ運用可能な三次元形状推定が実現可能であることがわかった。今後、さらなる精度向上を重ねることで、カメラを手軽で高精度なセンサーとして用いることが実現できると考えられる。

## 参考文献

- 1) Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros “Image-to-Image Translation with Conditional Adversarial Network