

令和 6 年度卒業論文

大規模 3 次元モデルデータセット
ShapeNet を用いた 3 D-SRCGAN の評価

宮崎大学工学部工学科
情報通信工学プログラム
七種 唯斗

指導教員 椋木 雅之

令和 6 年 2 月 7 日

目次

1. はじめに	1
2. 3D-SRCGAN による 3 次元モデル超解像.....	3
2.1. 超解像の定義.....	3
2.2. 超解像の従来研究.....	3
2.3. 3D-SRCGAN.....	4
2.4. 3D-SRCGAN の構造.....	5
2.5. 3D-SRCGAN の学習.....	6
3. 大規模 3D モデルデータセットを用いた超解像の評価.....	7
3.1. 3D-SRCGAN の評価実験における問題点.....	7
3.2. データセット ShapeNet について.....	8
3.3. 評価の観点.....	8
4. 実験	10
4.1. 実験設定	10
4.2. 1 クラスでの学習データ数の違いによる精度の評価.....	11
4.3. 多クラスでの学習データ数の違いによる精度の評価.....	13
4.4. 多クラスを用いることの有効性の調査.....	16
4.5. クラス識別情報を用いることの有効性の調査.....	19
5. おわりに.....	22
謝辞.....	24
参考文献	25

1. はじめに

近年、3DCG 技術の発展に伴い、ゲーム、アニメーション、映画、建築、医療など多岐にわたる分野でその技術が活用されている。3次元モデルは2次元画像よりも物体の形状を把握しやすく、複雑な形状を表現することが可能である。

3次元モデルの表現方法の一つに、ボクセルと呼ばれる立方体を積み重ねて3次元形状を表現するボクセル表現がある。このボクセル表現では、ボクセルを積み重ねることで容易に3次元モデルを生成することができる。しかし、精細な3次元モデルを作成するには、多数のボクセルを積み重ねる必要があり、手間が多くかかる。粗い3次元モデルから精細な3次元モデルを生成することができればこの手間を減らすことができる。このように低解像度データから高解像度データを復元・生成する技術は超解像と呼ばれる。

3次元ボクセルモデルの超解像を実現する手法として、3D-Super Resolution Conditional Generative Adversarial Networks (3D-SRCGAN) [1]が提案されている。3D-SRCGAN は、2次元画像における超解像技術をボクセル表現の3次元モデルに拡張した手法である。この手法は、元となる 3D-Super Resolution Generative Adversarial Networks (3D-SRGAN) [2]と同様に、Generative Adversarial Network (GAN) [3]を活用し、二つのネットワークが相互に敵対的に学習することにより、性能の向上を図るものである。しかし、3D-SRGAN には問題点が存在する。例えば、椅子のボクセルモデルの超解像を行いたい場合に、データが不足しているため、他のモデル（ベッドや机など）を混ぜて学習を行うことがある。これによりモデルそれぞれの特徴が混在し、超解像精度が低下するという問題が生じる。この問題に対処するため、3D-SRCGAN では3D-SRGAN にクラス識別情報を導入した。これにより、超解像精度が向上することが示されている。

一方で、3D-SRCGAN の性能評価を行う実験に使用された3次元モデルデータセットは、総データ数が少ないものであった。また、学習に使われるデータ数が3000~4000個を超えると、データ数の増加による超解像精度の向上が止まると述べていたが、一クラス当たりのデータ数が少なかったため、評価が不十分であったと言える。3D-SRCGAN の特性を正確に評価するためには、大規模なデ

ータセットを用いることが重要であると考えられる。ShapeNet[4]は 3D-SRCGAN の論文[1]で評価用に使用していた ModelNet[5]に比べて 1 クラス当たりのデータ数が多いクラスを複数含んでおり、総データの数も多い 3 次元モデルデータセットである。

本研究では、大規模な 3 次元モデルデータセットである ShapeNet を用いて、3D-SRCGAN の評価を行う。具体的にはデータ数やクラス数が超解像精度に与える影響やクラス識別情報を用いることの有効性について調査し、その結果について考察する。

以下、2 章では本研究で評価する 3D-SRCGAN について述べる。3 章では大規模 3D モデルデータセットを用いた超解像の評価について述べる。4 章では 3D-SRCGAN の評価を行い、データ数やクラス数が超解像精度に与える影響を調査し、その結果について考察する。最後に 5 章で、結論を述べる。

2. 3D-SRCGAN による 3 次元モデル超解像

2.1. 超解像の定義

超解像とは低解像度のデータから高解像度のデータを生成・補完する技術である。解像度とは、データを構成する単位が一定の間にどれだけ存在しているかを表している。本研究では超解像の対象としてボクセル表現の 3 次元モデルを扱っている。また、超解像によって生成された高解像度データを超解像 3 次元モデルと呼ぶ。

超解像は、低解像度データから高解像度データを生成する技術であるが、この問題は不良設定問題[6]である。不良設定問題とは、解を求めるために必要な情報が一部欠けている問題のことである。低解像度データから高解像度データに変換するためには、低解像度データには存在しない部分のデータを生成する必要がある。この低解像度データに存在しない部分には無数の生成パターンがあり得る。そのため無数のパターンから適切なパターンを選択できる方法を導入することが必要である。

2.2. 超解像の従来研究

超解像の従来研究として深層学習を用いた手法が提案されている。深層学習とは、機械学習の一種で、ニューラルネットワークにおいて深い階層構造を持ったネットワークを利用する学習方法である。深層学習の 1 つに畳み込みニューラルネットワーク (Convolution Neural Network : CNN) がある。CNN は畳み込み層で複数のカーネルと呼ばれる格子状の数値データを用いて入力データに畳み込み処理を行い、特徴マップを出力する。畳み込み処理とはカーネルと、同サイズの部分画像の数値データについて、各画素値の積の和を出力とする処理である。CNN では、入力の各層でカーネルをずらしながら畳み込み処理を行っている。この CNN を使用した超解像手法である SRCNN[7]では、多数の実画像を学習に利用することで一般的な画像における高解像データの生成パターンを学習し、より自然な高解像度画像を生成する。

深層学習を活用した 2 次元超解像手法には、他にも SRGAN[8]など GAN を利用した手法が存在する。GAN を利用した手法は、SRCNN などと比較して超解像精度が高い特徴がある。

また、2 次元画像超解像手法を 3 次元データ向けに拡張することで 3 次元超解像が実現されている。3 次元超解像は、主に医療画像などの 3 次元データに適用されている。

2.3. 3D-SRCGAN

3D-SRCGAN は 3D-SRGAN を応用した手法である。3D-SRGAN での超解像結果は、本来の高解像度 3 次元モデルには存在しない細かい穴が開いていたり、外周部分が滑らかになっていなかったりなど、超解像精度に問題点があった。

例えば、椅子のボクセルモデルの超解像を行いたい場合に、データが不足しているため、他のモデル（ベッドや机など）を混ぜて学習を行うことがある。しかし、これによりモデルそれぞれの特徴が混在し、超解像精度が低下するという問題が生じる。この問題に対処するため、3D-SRGAN にクラス識別情報を導入したものが 3D-SRCGAN である。

3D-SRCGAN は、Conditional Generative Adversarial Networks(CGAN)[9]を応用して作られている。CGAN では、入力画像にクラス識別情報を付加し、生成されるデータのクラスを制御している。3D-SRCGAN では、クラス識別情報で制限をかけるのではなく、各クラスの特徴が混ざり合わないようになっている。

クラス識別情報として、3D-SRCGAN では、CGAN と同様に Generator の入力である低解像度 3 次元モデルと、Discriminator の入力である超解像 3 次元モデルと高解像度 3 次元モデルに one-hot 表現のクラス識別情報を付加する。one-hot 表現は他の手法に比べてシンプルで様々な手法に応用しやすい。クラス数分の次元を持ったベクトルの各クラスに対応する値だけを 1、それ以外を 0 とすることで、クラス識別情報を付加している。

クラス識別情報の付加は、まず各クラスのモデルにクラス番号を割り振る。次に、3 次元モデルが入力される際にクラス数分のチャンネルを持った空の 3 次元モデルを用意し、割り振られたクラス番号に相当するチャンネルのボクセルの値を

すべて 1 で埋める。そして、入力モデルと組み合わせて $1 + \text{クラス数}$ 分のチャンネルを持つ 3 次元モデルがネットワークへの入力となる。

2.4. 3D-SRCGAN の構造

3D-SRCGAN は Generator と Discriminator の 2 つのネットワークで構成(図 1)されている。入出力データとしてボクセル表現の 3 次元モデルを扱う。学習データには低解像度 3 次元モデルと高解像度 3 次元モデルのペアを使用する。Generator は低解像度 3 次元モデルから高解像度 3 次元モデルを生成し、Discriminator は入力された 3 次元モデルが学習データの高解像度 3 次元モデルなのか Generator が生成した 3 次元モデルなのかを識別する。この 2 つは敵対関係にあり、それぞれの目的関数は、Generator は Discriminator をだますように学習データと類似した 3 次元モデルを生成すること、Discriminator は学習データと生成された 3 次元モデルを見分けることである。3D-SRCGAN としては Discriminator が識別できないような 3 次元モデルを Generator が生成できることが最終目的となる。また、学習の結果得られた Generator が生成モデルとなる。

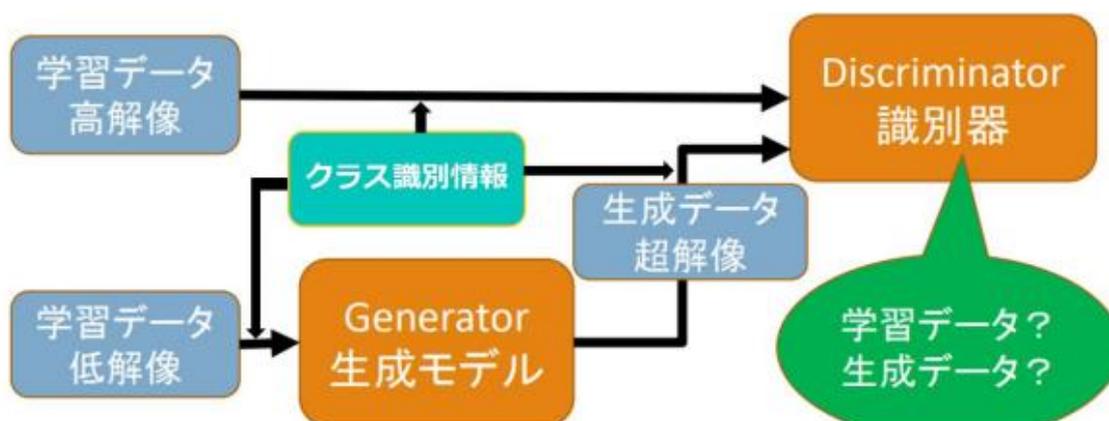


図 1 3D-SRCGAN の構造[1]

2.5. 3D-SRCGAN の学習

Generator の学習では、損失関数 l_G^{3DSR} の値が小さくなるようにネットワークのパラメータを逐次更新する収束計算を行う。Generator の損失関数 l_G^{3DSR} は式(4)で表現できる。

$$l_G^{3DSR} = \frac{1}{m} (l_{con}^{3DSR} + 10^{-3} l_{Gen}^{3DSR}) \quad (4)$$

$$l_{con}^{3DSR} = \sum_{x=1}^m \frac{1}{r^3 W H D} \sum_{x=1}^{rW} \sum_{y=1}^{rH} \sum_{z=1}^{rD} (l_{x,y,z}^{HR} - G(l^{LR})_{x,y,z})^2 \quad (5)$$

$$l_{Gen}^{3DSR} = \sum_{x=1}^m -\log D(G(l^{LR})) \quad (6)$$

式(4)の l_{con}^{3DSR} は Content Loss、 l_{Gen}^{3DSR} は Adversarial Loss、 m はバッチサイズを表している。

Content Loss は、式(5)で計算される。 W 、 H 、 D はそれぞれ 3次元空間の幅、高さ、奥行きを表している。式(5)では、学習データの高解像度 3次元モデル l^{LR} と Generator が生成した高解像度 3次元モデル $G(l^{LR})$ の平均二乗誤差を計算している。Content Loss は、元のデータの見たいをある程度維持して、別の操作を加えることができる。そのため超解像しても元の見た目から大きく変化することがない。

Adversarial Loss は、式(6)で計算される。 $G(l^{LR})$ は Discriminator で識別された結果、 $D(\cdot)$ は Discriminator の出力である。Adversarial Loss は、学習データと見分けがつかないように学習をする。Content Loss だけでは、見た目の変化の問題が解決するだけで、はっきりとした 3次元モデルが生成されず、Adversarial Loss だけでは学習データの性質を保持しているが、入力とは異なる 3次元モデルが生成される。そのため、この 2つを適切に組み合わせて超解像を行っている。

3. 大規模 3D モデルデータセットを用いた超解像の評価

3.1. 3D-SRCGAN の評価実験における問題点

3D-SRCGAN を提案した研究[1]では、クラス識別情報が付加されたことによる超解像精度の向上のほか、学習に用いる総データ数の違いやクラス数の違いによる精度の変化などについても実験がなされていた。

学習に用いる総データ数については、3000 個程度が必要で、それ以上増やしても超解像精度があまり向上しないとしていた。しかし、この実験結果は 10 クラスのデータを合わせて学習した場合のものであった。1 クラスで学習に用いるデータ数を変化させた時にも同様の傾向が見られるのか、3D-SRCGAN 手法の限界を明らかにするために、評価が必要である。また、クラス数の違いによる精度変化については、5~15 クラスについて、総データ数が同程度になるようデータを選択して学習を行い、精度を比較し、クラス数の影響は大きくないとしていた。しかし、選択するクラスがクラス数ごとにバラバラで、クラス数の影響と選択したクラス自体の超解像の難しさを切り分けられていなかった。特に、総データ数が十分な場面において、複数のクラスを学習することが超解像精度の向上に有効なのか、さらにその場面でクラス識別情報が有効に働くのか評価できていなかった。

本研究では、これらの点について、クラス数が多く、1 クラスのデータ数も多い大規模データセット ShapeNet を用いて評価する。

3.2. データセット ShapeNet について

本研究では大規模 3 次元モデルデータセット ShapeNet を用いる。データ総数は 57,448 個、クラス総数は 57 種類である。各クラスにおけるデータ数は異なり、最も多いクラスは table で 8,509 個である。本研究ではこのデータセットを基に、様々な条件で組み合わせて学習・超解像する。

PC のメモリ容量に制約があるため、データセットのすべてのモデルを使用することは困難である。このため、使用するデータは 1 クラス当たりのデータ数が多い 20 クラスからデータを選択し、学習および超解像の実験に適したサンプ

ルを選定する。選択したデータをデータ数が多い順に表 1 に示す。ShapeNet の大規模かつ多様なデータを使用することで、実験結果の一般性と精度を高めることができると考えている。

表 1 使用した ShapeNet データの内訳

クラス名	table	car	chair	airplane	rifle
データ数	8509	7496	6778	4045	2372
クラス名	lamp	vessel	bench	loud speaker	cabinet
データ数	2318	1939	1816	1618	1572
クラス名	boat	display	bus	bathtub	guitar
データ数	1137	1095	939	857	797
クラス名	faucet	clock	pot	jar	knife
データ数	744	655	602	597	423

3.3. 評価の観点

評価の目的として、まず 1 クラスでも学習データが十分あれば超解像精度は高くなるのかを確認することが挙げられる。その理由として 3D-SRGAN では例えば椅子のボクセルモデルの超解像を行いたい場合に、データが不足しているため、他のモデル（ベッドや机など）を混ぜて学習を行うことがある。これによりモデルそれぞれの特徴が混在し、超解像精度が低下するという問題が生じる。この問題に対処するため、3D-SRCGAN では 3D-SRGAN にクラス識別情報を導入した。これにより、超解像精度が向上することが示されている。これは 1 クラスで十分な学習データ数を得られなかったから考えられた手法である。そこで 1 クラスでも学習データが十分あれば超解像精度は高くなるのかを確認する。また、比較のために多クラス(10 クラス)でも学習データ数が増加すれば超解像精度は高くなるのかを確認する。

次に、十分なデータ数があれば、多クラスで学習する必要はないのかについて明らかにする。先行研究では、クラス数の違いによる超解像精度の向上はないと

結論づけていたが、評価方法に不十分な点があった。その点を改めたうえで多クラスを用いることの有効性について評価する。

次に十分な学習データがあれば、クラス識別情報は不要であるかについて確認する。十分な学習データがあれば、クラス識別情報を付加せずに複数の異なるクラスを混ぜても特徴が混ざり合うことなく、超解像を行うことができる可能性がある。十分なデータがある場合のクラス識別情報の必要性を明らかにするために実験を行う。

以上の学習データ数の違いによる精度の評価、多クラスを用いることの有効性の評価、クラス識別情報を用いることの有効性の評価の 3 点について明らかにするために実験を行う。

4. 実験

4.1. 実験設定

本研究では、学習、評価に用いる 3 次元モデルとして、データセット ShapeNet を使用した。このデータセットは table などの 57 クラスの 3 次元モデルがある。実験では、その中からデータ数の多い表 1 の 20 クラスを用いる。学習データは実験ごとにデータ数を変更して選択する。テストデータは、table クラスの中から学習に使用していない 849 個のデータをすべての実験で共通して用いる。先行研究[1]では、実験ごとに定量評価に用いるテストデータが異なっており、実験に使用するクラスが異なる場合に適切な比較が行えていなかった。本研究では、全ての実験が table クラスの学習を含み、評価にも同じテストデータを用いることで、実験間で適切な比較評価を行う。学習データ、テストデータともにボクセル数が $16 \times 16 \times 16$ の低解像度 3 次元モデルと $64 \times 64 \times 64$ の高解像度 3 次元モデルのペアに変換し、モデル内の空洞をボクセルで埋めた。学習回数は予備実験の結果や PC のメモリ容量を考慮して、バッチサイズを 8、Epoch 数はそれぞれのモデルにおいてよい精度であった Epoch 数を選定した。

超解像 3 次元モデルの評価では、以下の 4 つの項目で定量評価を行った。各評価項目は値が小さいほど良い性能である。テストデータに対するこれらの値の平均値を定量評価結果とする。

- 高解像度 3 次元モデルと超解像 3 次元モデルの連結成分数の差 (連)
- 高解像度 3 次元モデルと超解像 3 次元モデルの穴の総数の差 (穴)
- 高解像度 3 次元モデルと超解像 3 次元モデルの空洞の総数の差 (空)
- 高解像度 3 次元モデルと超解像 3 次元モデルの同じ座標間でのボクセルの保持する値 (ボクセル値) が異なるものの数 (値)

また、主観的な評価として、それぞれの超解像 3 次元モデルを同じ角度から見た結果を画像として見比べた。

4.2 節以降でこれらの評価項目にしたがって、3.3 節で述べた 3 つの評価の観点に関して 3 次元モデルの超解像結果の定量評価と画像を示して、結果について考察していく。

4.2. 1 クラスでの学習データ数の違いによる精度の評価

学習に用いるデータ数の変化による超解像精度への影響を調査するために table クラスの 1 クラスでデータ数が 800、3500、7660 で学習したモデルの超解像精度を比較する。

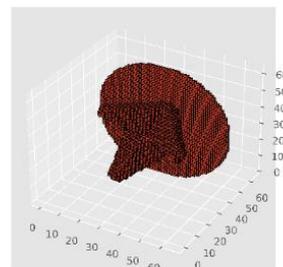
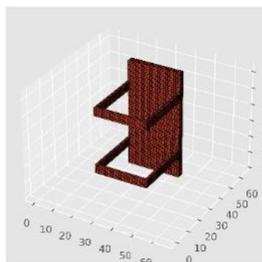
図 2 に table クラスの 2 例の 3 次元モデルの超解像結果を示す。上から 1 段目が高解像度 3 次元モデル、上から 2 段目がデータ数 800、上から 3 段目がデータ数 3500、最下段がデータ数 7660 の結果である。各列は 2 例の table クラス 3 次元モデルに対応している。左から table1、table2 とする。

図 2 を主観的に見ると、データ数が増加するにつれて、より良い 3 次元モデルを生成できている。高解像度 3 次元モデルとそれぞれを見比べてみると、大まかな形状は生成できている。データ数 800 では table1 には足にかけた部分があったり、table2 には台の部分に穴が開いている。それらがデータ数 3500 になると、きれいになっている。また、データ数 7660 とデータ数 3500 において、table2 では足の部分が、データ数 7660 の方が少し直線的になっており、高解像度 3 次元モデルに比較的近い。

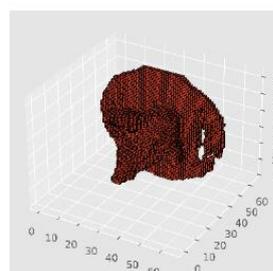
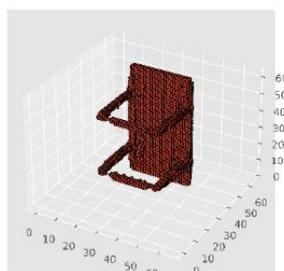
また、表 2 を見ると、連結成分に関しては、データ数 3500 が一番優れていたり、空洞に関してはデータ数 800 とデータ数 3500 に差は無いが、全体的に見て、データ数が増加するにつれて超解像精度が向上している。

以上のことにより、1 クラスのデータ数を増やすことで、超解像精度が向上することが分かった。また、先行研究の実験結果として、学習に用いる総データ数については、3000 個程度が必要で、それ以上増やしても、超解像精度はあまり向上しないと述べていた。この実験結果は 10 クラスのデータを合わせて学習した場合であったが、1 クラスで学習に用いるデータ数を変化させた場合、同様の傾向がみられるのか、評価が必要であった。先行研究の実験結果とは異なり、3000 個以上でも超解像精度が向上していた。よって、1 クラスでデータを 7660 個以上用意することが出来れば、1 クラスの学習でもさらに超解像精度が向上する可能性がある。

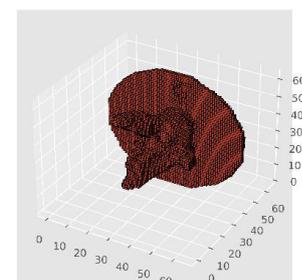
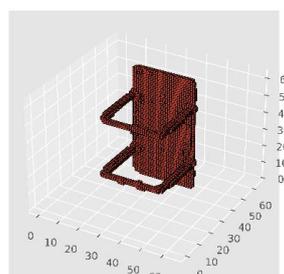
高解像度
3次元
モデル



データ数
800



データ数
3500



データ数
7660

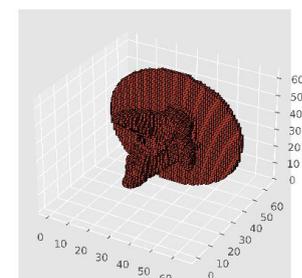
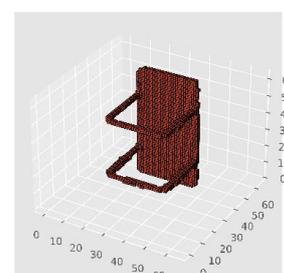


table1

table2

図2 tableクラスの超解像結果の例(1クラスの学習)

表2 1クラスでのデータ数の違いによる超解像結果の定量評価
(tableクラスのテストデータ全体に対する平均値)

学習データ数(個)	連	穴	空	値
800	4.1	7.4	2.5	22048.0
3500	2.8	6.4	2.5	21933.1
7660	3.1	5.6	2.1	21723.7

4.3. 多クラスでの学習データ数の違いによる精度の評価

4.2節では、1クラスの学習では学習データ数が多くなる程超解像精度が向上することを示した。本節では、多クラスの場合でも同様の傾向が見られるか調査する。学習には、表3に示した10クラスを用いる。学習データ数が少ない場合（各クラス766個、計7660個）と多い場合（各クラス1500個程度、計14356個）の2通りで学習を行う。

図3にtableクラスの2例の3次元モデルの超解像結果を示す。上段が高解像度3次元モデル、中段がデータ数7660、下段がデータ数14356の結果である。各列は2例のtableクラス3次元モデルに対応している。左からtable1、table2とする。

図3を主観的に見ると、データ数7660より、データ数14356の方が高解像度3次元モデルに近いものを生成できている。高解像度3次元モデルとそれぞれを見比べると、どちらも大まかな形状は生成できている。しかし、table1ではデータ数14356の方がデータ数7660に比べて、端のなめらかさが良い。また、table2に関しては、データ数7660の方は足の部分にちぎれた部分が存在している。

また、表4を見ると、連結成分、穴、空洞に関して、データ数14356の方が大幅に優れている。しかし、ボクセル値に関してはデータ数7660の方が優れていた。

以上のことにより、データ数のみの違いによる超解像精度はデータ数が多い方が優れていることが分かった。ボクセル値に関しては他の項目が優れている

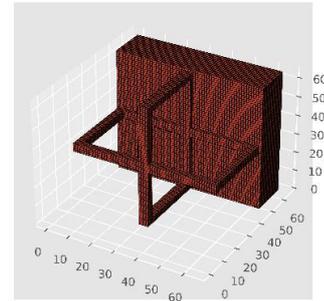
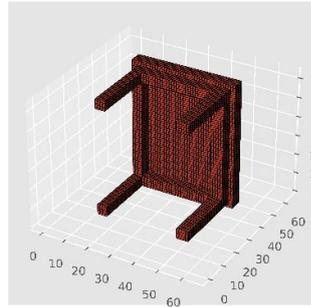
のにボクセル値だけ、精度が向上しないことが他の評価でも見られたため、理由や原因が存在しているかもしれない。それらを究明できれば、さらなる超解像精度の向上につながると考えられる。

1クラスでも学習データが増加すると、超解像精度が向上していたが、多クラスの方がより超解像精度が向上していた。

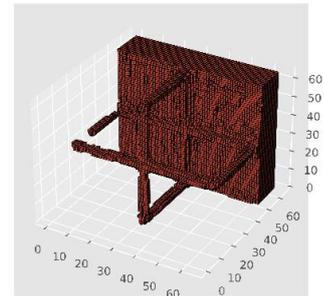
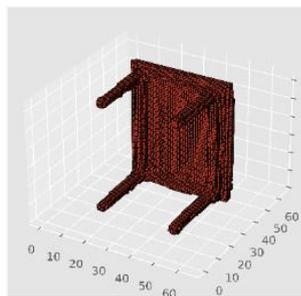
表3 実験4.2の学習に使用したクラス

使用した	airplane	bench	boat	cabinet	car
10クラス	chair	lamp	loudspeaker	rifle	table

高解像度
3次元
モデル



データ数
7660



データ数
14356

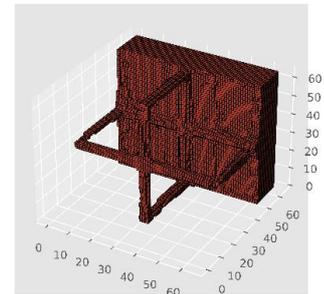
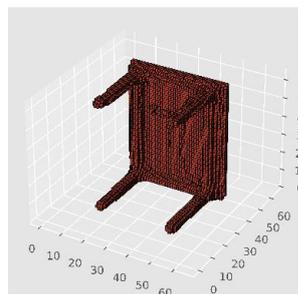


table1

table2

図3 tableクラスの超解像結果の例(10クラス学習)

表4 10クラスでのデータ数の違いによる超解像結果の定量評価
(tableクラスのテストデータ全体に対する平均値)

学習データ数(個)	連	穴	空	値
7660	2.5	6.7	3.8	21713.1
14356	1.9	3.3	0.98	23224.6

4.4. 多クラスを用いることの有効性の調査

本節では、学習データが多い状況で多クラスでの学習を行うことの有効性を調査する。学習に用いる総データ数を約 7500 個に統一した状況で、その内訳をクラス数 1、クラス数 10、クラス数 15、クラス数 20 とした 4 通りで学習を行う。使用したクラスと総データ数の詳細を表 5 に示す。クラス数 1 は、table クラスのみ (7660 個) を学習に用いる。クラス数 10,15,20 の使用クラスは各段階で 5 クラスずつ追加している。表 5 には各段階で追加したクラスのみを記載している。

図 4 に table クラスの 2 例の 3 次元モデルの超解像結果を示す。上から 1 段目が高解像度 3 次元モデル、2 段目がクラス数 1、3 段目がクラス数 10、4 段目がクラス数 15、最下段がクラス数 20 の結果である。各列は 2 例の table クラス 3 次元モデルに対応している。左から table1、table2 とする。

図 4 を主観的に見ると、いずれの超解像結果でも高解像度 3 次元モデルに近いものを生成できている。いずれのクラスでもデータ数が多く、データ数が多いと超解像精度が良いということにつながっているといえる。

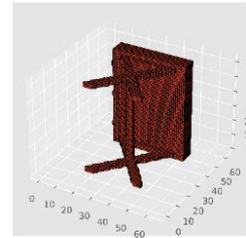
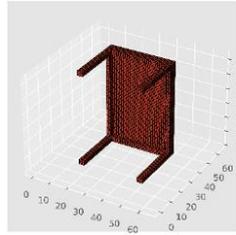
また、表 6 を見ると、穴と空洞に関しては、クラス数 1 がクラス数 10 や 15 を上回っている部分がある。ボクセル値に関しては、クラス数 10 が優れており、数値は上下して、むらがある。しかし、全体的に見て、クラス数が増加するにつれて超解像精度が向上している。

以上のことにより、クラス数を増やすことは、3 次元ボクセルモデル超解像において超解像精度向上に有効であることがわかった。

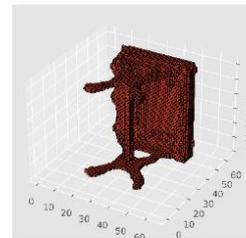
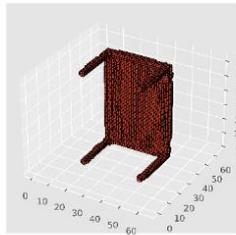
表 5 実験 4.4 の学習に使用したクラス

	使用クラス					総データ数
クラス数 10	airplane	bench	boat	cabinet	car	7660
	chair	lamp	loud speaker	rifle	table	
クラス数 15	bathtub	bus	display	guitar	vessel	7680
クラス数 20	clock	faucet	jar	knife	pot	7500

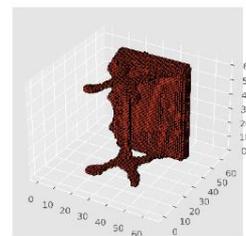
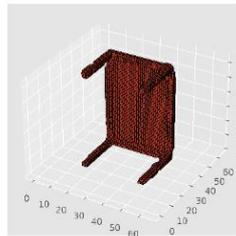
高解像度
3次元
モデル



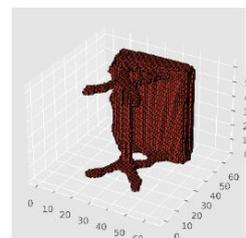
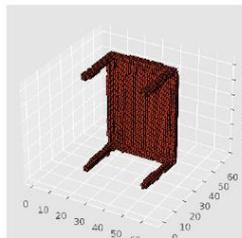
クラス数 1



クラス数 10



クラス数 15



クラス数 20

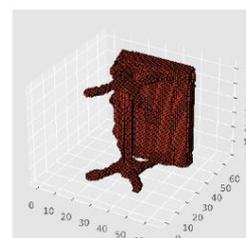
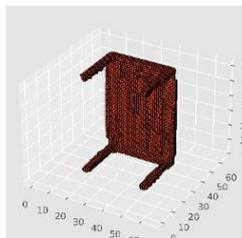


table1

table2

図4 tableクラスの超解像結果の例(クラス数の違い)

表6 クラス数の違いによる超解像結果の定量評価
(tableクラスのテストデータ全体に対する平均値)

	連	穴	空	値
クラス数 1	3.1	5.6	2.1	21723.7
クラス数 10	2.5	6.7	3.8	21713.1
クラス数 15	2.3	5.6	2.6	22197.6
クラス数 20	2.1	3.6	1.2	23439.1

4.5. クラス識別情報を用いることの有効性の調査

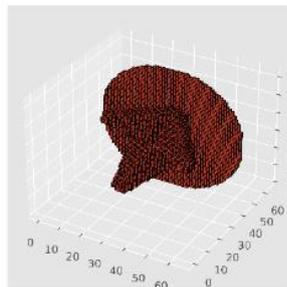
クラス識別情報の有効性を調査するために、クラス数 10、15、20 の 3 つの学習モデルについてクラス識別情報を用いた場合と用いていない場合の学習モデルで超解像精度を比較する。使用データは実験 4.4 と同様のデータである。クラス識別情報なしの学習方法として、異なるクラスを区別せず、1 クラスとして学習する。

図 5 に table クラスの 3 次元モデルの超解像結果の例を示す。上から 1 段目が高解像度 3 次元モデル、2 段目がクラス数 10、3 段目がクラス数 15、4 段目がクラス数 20 の結果である。それぞれ左がクラス識別情報あり、右がクラス識別情報なしである。

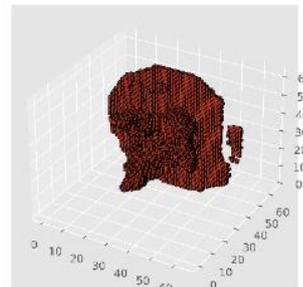
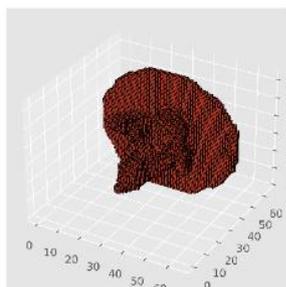
図 5 を主観的に見ると、どのクラスにおいても、クラス識別情報を用いていない場合に比べて、クラス識別情報を用いている場合の方が高解像度 3 次元モデルに近いものを生成できている。クラス識別情報を用いていない場合、穴が生じており、綺麗に生成できていない。

また、表 7 でそれぞれクラスでクラス識別情報ありとなしを比較すると、ボクセル値以外の連結成分、穴、空洞に関して、クラス識別情報ありの方が優れている。全体的に見て、クラス識別情報を用いることの超解像精度に対する有効性はあることが分かった。

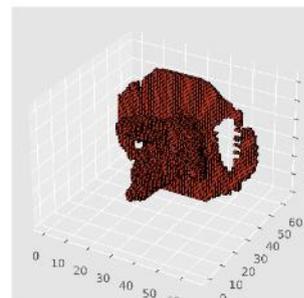
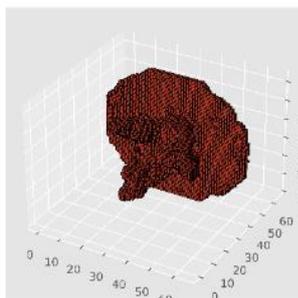
高解像度
3次元
モデル



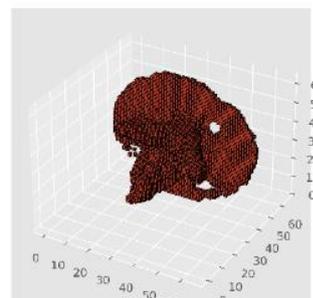
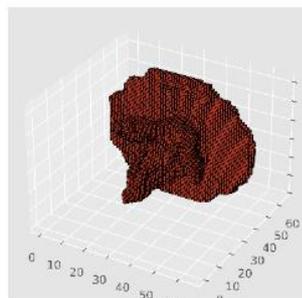
クラス数 10



クラス数 15



クラス数 20



クラス
識別情報

あり

なし

図5 table クラスの超解像結果の例(クラス識別情報の有無)

表7 クラス識別情報の有無による超解像結果の定量評価

(table クラスのテストデータ全体に対する平均値 識別あり/識別なし)

	連	穴	空	値
クラス数 10	2.5/4.7	6.7/10.8	3.8/4.6	21713.1/19528.3
クラス数 15	2.3/4.3	5.6/9.7	2.6/4.0	22197.6/19340.9
クラス数 20	2.1/5.0	3.6/11.1	1.2/4.6	23439.1/19210.4

5. おわりに

本研究では、大規模な 3 次元モデルデータセットである ShapeNet を用いて、3D-SRCGAN の評価を行った。まず、1 クラス、10 クラスでのデータ数の違いによる超解像精度の違いについて評価した。その結果、どちらにおいてもデータ数が増加するにつれて超解像精度は向上した。先行研究の実験結果として、学習に用いる総データ数については、3000 個程度が必要で、それ以上増やしても、超解像精度はあまり向上しないと述べていた。この実験結果は 10 クラスのデータを合わせて学習した場合であったが、1 クラスで学習に用いるデータ数を変化させた場合、同様の傾向がみられるのか、評価が必要であった。結果としてデータ数が 3000 個以上から増加しても超解像精度が向上していた。よって、1 クラスでデータを 7000~8000 個程度以上用意することが出来れば、1 クラスの学習でもさらに超解像精度が向上する可能性がある。

次に多クラスを用いることの有効性について評価した。その結果、学習に用いるクラス数が増加するにつれて超解像精度は向上していた。また、先行研究では、クラス数の違いによる超解像精度の変化はないと結論づけていた。これは、評価方法として学習データを構成するクラスの種類が異なっているにもかかわらず、すべてのクラスを含めたデータ全体に対する平均値で評価したためだと考えられる。本研究では、同じ一つのクラスのテストデータで評価することにより、多クラスを用いることの有効性があることを確認できた。

さらにクラス識別情報を用いることの有効性についても評価を行った。結果として、クラス識別情報を用いている場合の方が、クラス識別情報を用いていない場合よりも超解像精度が良く、高精度な超解像 3 次元モデルを生成できていた。

以上のことから、まず、1 クラスでも十分なデータがあれば、超解像精度は向上するということが分かった。次に多クラスを用いることで超解像精度がさらに向上することが分かった。また十分なデータがある場合でもクラス識別情報を用いるべきで、クラス識別情報なしだと超解像精度が低下することが確認できた。

今後の課題として、ボクセル値に関する評価の見直しが挙げられる。本研究の評価実験のなかで、評価項目として、連結成分、穴、空洞、ボクセル値の4つがあるが、ボクセル値は他の評価項目で精度が改善しているのとは反対に悪化したり、上下したりなどの不安定な傾向がみられた。この原因を究明できれば更なる超解像精度の向上につながると考えられる。また、学習データ数の違いについて評価したが、学習データ数が3500個から7660個にかけて超解像精度が向上した。このことから学習データ数による超解像精度向上の限界値は7660個以上であり、未知数である。よりデータ数が多いデータセットを用いて、限界値を明らかにすることが課題である。また、3次元モデルの複雑さによって、超解像結果に差がある。3次元モデルが複雑であっても、対応できる工夫や手法を見つけることで、さらに超解像精度が向上すると考えられる。

謝辞

本研究を進めるにあたり、多くの方々にご協力いただきました。ご指導いただいた椋木雅之教授に感謝いたします。指導教員である椋木雅之教授には研究に関しての相談や論文執筆にあたって多くのご指導をいただきました。本当にありがとうございました。

また、研究室の皆様にも研究について様々な助言をいただきました。本当にありがとうございました。

参考文献

- [1] 野村淳也,“クラス識別を導入した 3D-SRGAN によるボクセル超解像”,令和5年度 宮崎大学大学院 工学研究科 工学専攻 機械・情報系コース 情報システム工学分野 修士論文(2024)
- [2] 岡和寿,椋木雅之,“SRGAN の 3 次元ボクセルモデル超解像への適用”, 画像電子学会論文誌, vol. 48, no. 4, pp. 448-496 (2019)
- [3] I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, A Courville, Y Bengio, “Generative adversarial nets”, Advances in Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
- [4] ShapeNet, <https://shapenet.org/>
- [5] ModelNet10, ModelNet40, <http://modelnet.cs.princeton.edu/>
- [6] 池谷彰彦, 広明敏彦,“超解像ソリューション (組込みソフトウェア・ソリューション 特集)--(イメージ/音声処理コンポーネントソリューション)”,NEC 技報, vol. 60, no. 2, pp. 24-26 (2007)
- [7] C Dong, CC Loy, K He, X Tang, “Image super-resolution using deep convolutional networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 38, no. 2, pp. 295-307(2015)
- [8] C Ledig, L Theis, F Huszar, J Caballero, A Cunningham, A Acosta, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105-114 (2017)
- [9] M Mirza, S Osindero, “Conditional generative adversarial nets” arXiv:1411.1784 (2014)