

令和5年度修士論文

クラス識別を導入した  
3D-SRGANによるボクセル超解像

指導教員 棕木 雅之

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

T2203425 野村 淳也

# 概要

本論文では、3次元モデルの超解像を行う 3D-SRGAN にクラス識別情報を導入して、超解像精度を向上させた 3D-Super Resolution Conditional Generative Adversarial Networks (3D-SRCGAN) を提案する。

低解像度データから高解像度データを復元・生成する技術を超解像という。3次元ボクセルモデルの超解像を行う手法に 3D-SRGAN がある。3D-SRGAN には、1度の学習で同時に使えるデータ数が少ない、学習に使っていないクラスの超解像精度が低いといった問題点がある。これは、複数クラスのデータを混ぜて学習を行うと全く異なるモデルそれぞれの特徴が混ざり合ってしまうという 3D-SRGAN の性質が原因であった。

GAN を用いた画像生成には、各クラスのデータに個別のクラス識別情報を付加し、クラス識別を行う CGAN が存在する。3次元モデル超解像においても、クラス識別によって各クラスの特徴を制御できる可能性がある。そこで、本論文では 3D-SRGAN の学習にクラス識別情報を付加することで、超解像精度の向上を目指す。

従来の 3D-SRGAN と提案手法である 3D-SRCGAN で比較実験を行った。1度の学習で同時に使えるデータ数が少ないという問題点に対して、データ数が増えることによる超解像精度の向上を評価した。1クラスのデータで学習した 3D-SRGAN と複数クラスのデータで学習した 3D-SRCGAN で超解像結果を比較した結果、超解像精度が改善した。この結果より、クラス識別情報を導入した上で、学習データの数を増やすことは超解像精度向上に有効であることを示した。

学習に使っていないクラスの超解像精度が低いという問題点に対しては、クラス識別情報によって超解像精度を低下させずに複数クラスを同時に学習できることを評価した。複数クラスのデータで学習した 3D-SRGAN と 3D-SRCGAN で超解像結果を比較した結果、定量評価で 3D-SRCGAN の誤差が 3D-SRGAN の  $\frac{1}{4}$  以下となった。この結果より、クラス識別情報の導入により、複数クラスの特徴が混ざり合い、超解像精度が低下することを防げることを示した。

# 目次

1. はじめに .....	1
2. 従来手法 3D-SRGAN による 3次元モデル超解像 .....	3
2.1. 超解像の定義 .....	3
2.2. 2次元画像超解像の従来研究 .....	4
2.3. 3次元データ超解像の従来研究 .....	5
2.4. 3D-SRGAN .....	6
2.5. 3D-SRGAN による超解像の問題点 .....	7
2.6. CGAN .....	9
3. 提案手法 3D-SRCGAN による 3次元モデル超解像 .....	10
3.1. クラス識別情報の導入 .....	10
3.2. 3D-SRCGAN のネットワーク構造 .....	12
3.3. 3D-SRCGAN の Generator の構造 .....	13
3.4. 3D-SRCGAN の Discriminator の構造 .....	16
3.5. Generator の学習更新 .....	18
3.6. Discriminator の学習更新 .....	19
4. 実験 .....	20
4.1. 実験設定 .....	20
4.2. クラス識別情報を導入することによる精度向上の評価 .....	21
4.2.1. 実験内容 .....	21
4.2.2. 実験結果 .....	21
4.3. 複数クラスで学習することによる精度向上の評価 .....	25
4.3.1. 実験内容 .....	25
4.3.2. 実験結果 .....	25
4.4. 総データ数の違いによる超解像精度への影響調査 .....	29
4.4.1. 実験内容 .....	29
4.4.2. 実験結果 .....	29
4.5. クラス数の違いによる超解像精度への影響調査 .....	34
4.5.1. 実験内容 .....	34
4.5.2. 実験結果 .....	36

4.6. 異なるクラス識別情報が超解像精度に与える影響の調査 .....	41
4.6.1. 実験内容 .....	41
4.6.2. 実験結果 .....	41
4.7. データセットの解像度を上げた場合の超解像例 .....	50
4.7.1. 実験内容 .....	50
4.7.2. 実験結果 .....	51
5. おわりに .....	56
謝辞 .....	58
参考文献 .....	59

# 1. はじめに

近年、3次元モデルは建築や機械設計、エンターテインメント、医療等、幅広い分野で利用されている。2次元画像に比べて3次元モデルを利用すると、簡単に物体の形状を把握し、複雑な形状を設計することができる。

3次元モデルの表現方法の1つに、ボクセルと呼ばれる立方体を積み重ねて3次元モデルを表現するボクセル表現がある。このボクセル表現では、3次元空間の各ボクセルの値を2値で表し、ボクセルの有無を表現している。多数のボクセルを積み重ねることで精細な3次元モデルを作成することができるが、粗い3次元モデルと比較するとボクセルを積み重ねる手間が増える。粗い3次元モデルから精細な3次元モデルを自動で生成することができれば、この手間を軽減できる。このような粗いデータから精細なデータを生成する処理を超解像と呼ぶ。

3次元モデルの超解像の手法として、3D-SRGAN[1]がある。3D-SRGANは、深層学習を用いた生成モデルである Generative Adversarial Network(敵対的学習: GAN)[2]を用いている。GANは、生成器(Generator)と識別器(Discriminator)と呼ばれる二つのネットワークが互いに学習を行うことで性能を高めていく。これにより、粗い3次元モデルを与えることでそれに合った精細な超解像3次元モデルを生成することができる。

しかし、従来の3D-SRGANは、超解像3次元モデルの超解像精度に改善の余地がある。超解像精度が高くない原因の一つとして、学習データの数が少ないことが考えられる。深層学習において、高い精度を実現するためには大量のデータが必要となる。しかし、3D-SRGANで利用している3次元ボクセルモデルは2次元画像などと比較して大量のデータを確保することが難しい。他の原因として、学習に使うデータのクラスの問題が考えられる。過去の研究[1]や予備実験により、3D-SRGANには、学習に使っていないクラスや複数クラスを混ぜたデータセットでの超解像精度は低下する傾向があることが分かっている。1クラスの学習では、様々な3次元ボクセルモデルの超解像が行えない上、学習に使うデータ数も限られ、超解像の精度が高くない。一方、複数クラスを学習すると、異なるクラスの特徴が混ざってしまい、超解像の精度が低下する。そのため、超解像の精度を向上させることが困難であった。

本研究では、GAN の学習に条件付けを行うことで生成物を制御し、生成精度を向上させる CGAN[3]を応用し、クラス識別情報を付加することで超解像精度向上させた 3D-Super Resolution Conditional Generative Adversarial Networks (3D-SRCGAN) を提案する。CGAN は、Generator と Discriminator の入力に追加の条件情報を与えて学習させることで生成物の制御、生成精度向上を図っている。この条件情報にはクラス識別情報や画像、テキストなど様々なものがあり、用途によって適したものを使うことが重要である。本研究においては、条件情報としてクラス識別情報を付加している。これにより、超解像精度を落とすことなく複数クラスを同時に学習することができると考えた。

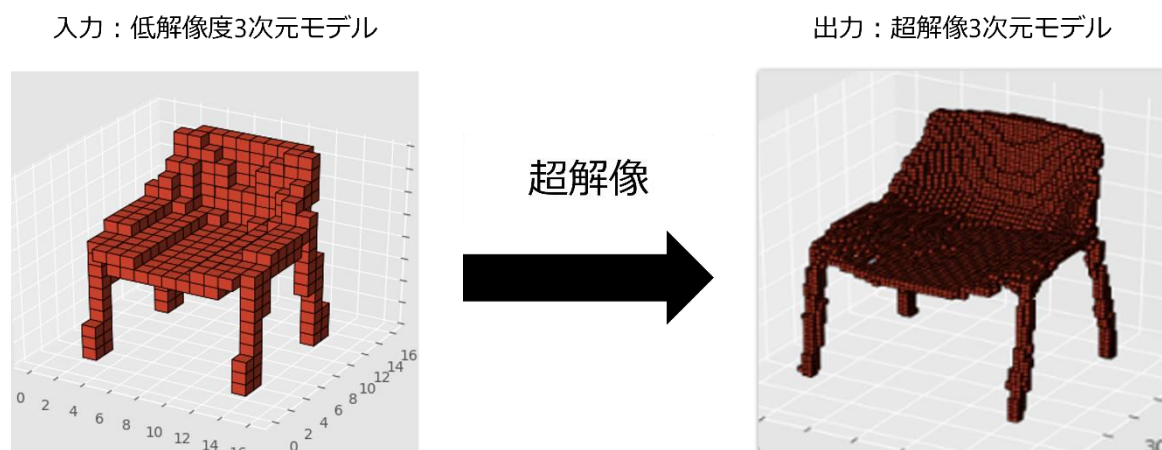
以下、2章では、本研究の基となった 3D-SRGAN について述べる。3章では、3D-SRGAN にクラス識別情報を付加した提案手法である 3D-SRCGAN について述べる。4章では 3D-SRCGAN が超解像精度の向上に有効であるか、実験により評価する。最後に 5章で、結論と今後の課題を述べる。

## 2. 従来手法 3D-SRGAN 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、画素やボクセルなどデータの構成単位が一定の区間内にどれだけ密に存在しているかを表したものである。本研究では、超解像の対象として、ボクセル表現の 3 次元モデルを扱う (図 1)。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 (ill-posed problem) [4] である。不良設定問題とは、解を求めるために必要な情報の一部が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える。さらに、多くの実データにはノイズが含まれる上、デジタルデータには量子化誤差も含まれるため、一般に低解像度データから適切な高解像度データを生成することは難しい問題である。



## 2.2. 2次元画像超解像の従来研究

2次元画像超解像には、複数の画像から行う手法と1枚の画像から行う手法がある。

複数の画像から超解像する手法では、同じ対象に対する複数の画像から得られる情報を利用して超解像を行っている。これは実画像に基づく手法なので、正しい高解像度画像が得られる。しかし、適切な設定の画像が複数枚必要である。

1枚の画像から超解像する手法には、補間の考え方に基づく信号処理的なアプローチによる手法[5][6]と、学習や事例集合に基づいて復元する手法[7]がある。前者では、元画像に高周波成分が含まれていない場合は大きな改善が期待できない。一方で後者は、学習に使った画像や事例集合内の画像の情報を利用して、1枚の画像から高周波成分を付与できる。そのため高周波成分が含まれていない画像であっても、高周波成分を付与した高解像度画像を生成することができる。しかし、低解像度画像に存在しない部分のデータは、学習や事例集合に利用した異なる画像を参考にして生成するため、必ずしも正しい高解像度画像が得られるとは限らない。

近年では深層学習を用いて超解像を行う研究がなされている。深層学習とは、機械学習の1種である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[8]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でき、より自然な高解像度画像を生成することができる。

深層学習を活用した2次元超解像手法には、他にも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。本研究では、GANを利用した超解像手法であるSRGAN[9]を応用している。



## 2.3. 3次元データ超解像の従来研究

3次元超解像は、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。

医療画像において高解像度な画像を使うことで、医師による診断や治療をより正確なものにできる。さらに、これらの画像は医者や研究者、学生にとって貴重な資料となる。これらの理由から、3次元超解像は医療画像分野において診断や治療計画において非常に重要であるだけでなく、新たな発見や治療法の開発など医療分野の発展にも重要な役割を担っている。

医療画像における3次元超解像は、主に Magnetic Resonance Imaging (MRI) や Computed Tomography (CT) で活用されている[10][11]。MRIは、強い磁石による磁場と電磁波を使って体内の水素原子を振動させ、発生した微弱な電磁波の強さを多値で表し、断層画像を得ている。一方CTは、人体にX線を照射し、透過したX線量を多値で表し、断層画像を得ている。どちらの方法もこの断層画像を重ねることで、3次元画像化している。MRIやCTにより、人体内の臓器の分布や状態を把握することができる。しかし、これらの方法で高解像度な画像を取得することは、通常、患者に大きな負担をかける。低解像度で取得した3次元画像から、超解像により高解像度の3次元画像を生成することで、患者の被曝を最小限に抑えつつ、精密な医療画像を取得することが可能となっている。

以上のように、医療画像での3次元超解像では、多値で表現した3次元画像から、物体の分布や状態を把握することが重要視されている。一方、本研究のもととなった3D-SRGANで行っている3次元モデリングでの3次元超解像では、2値で表現した3次元モデルから、物体の形状を把握し、再現することが重要である。

## 2.4. 3D-SRGAN

3D-Super Resolution Generative Adversarial Networks (3D-SRGAN)[1]は深層学習を用いた3次元モデル超解像手法であり、画像の超解像に有効な Super Resolution Generative Adversarial Networks(SRGAN)[9]の超解像の対象を、ボクセル表現の3次元モデルに拡張した手法である。SRGANはGenerative Adversarial Network(GAN)と呼ばれる生成モデルを応用している。GANはGenerator(生成器)とDiscriminator(識別器)の2つのニューラルネットワークで構成されている。Generatorはより学習データに近いデータを生成しようとし、Discriminatorは入力データが学習データかGeneratorが生成したデータかを識別しようとする。GeneratorはDiscriminatorを騙すことができるように学習し、DiscriminatorはGeneratorが生成したデータを見破ることができるように学習する。この学習は、Generatorが学習データと同じようなデータを生成できるようになることが最終目標である。この状態になると、Discriminatorは学習データと生成データの識別ができなくなり、正答率は50%になる。以上のような仕組みでGANは学習データに非常に近い新たなデータを生成することができる。3D-SRGANは、GANのこの特徴を利用して超解像を行う。

## 2.5. 3D-SRGAN による超解像の問題点

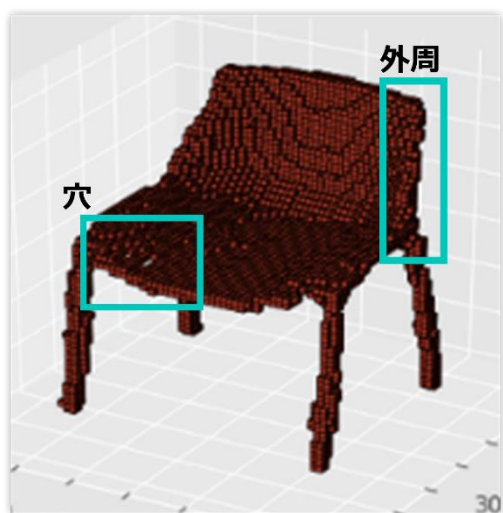
従来の 3D-SRGAN での超解像結果（図 2）は、本来の高解像度 3 次元モデルには存在しない細かい穴が開いていたり、外周部分が滑らかになっていないなど、超解像精度に改善の余地がある。

これは、学習データの数が少ないことが原因の一つだと考えられる。深層学習において、高い精度を実現するためには大量のデータが必要となる。しかし、3D-SRGAN で利用している 3 次元ボクセルモデルは、2 次元画像などと比較して大量のデータを確保することが難しい。さらに、3D-SRGAN では 1 クラス分のデータのみを使用して学習を行っている。結果として、3D-SRGAN の学習に使用できるデータ数はかなり限られる。

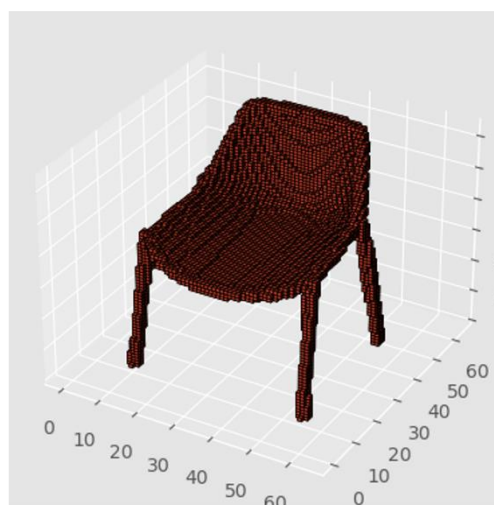
他の原因として、学習に使っていないクラスの超解像精度が低いという 3D-SRGAN の特徴が考えられる。3D-SRGAN では、1 クラスのデータのみを使用している影響で、学習に使っていない他のクラスの特徴情報を学習することができない。その結果、学習データと形状が似ていないクラスの超解像精度が低い傾向があることが先行研究[1]で示されている。

この問題に対し、複数クラスのデータを使って学習することで、データ数を増やしつつ、学習に使ったすべてのクラスの超解像精度を改善できると考えた。そこで、予備実験として複数クラスのデータを混ぜ、データ数を増やして学習した 3D-SRGAN で超解像を行った。その結果、学習に使ったクラスの超解像精度が低下してしまった（図 3）。

このように、異なるクラスのデータを合わせることでデータ数を増やすことはできる。しかし、異なるクラスのデータを混ぜて学習を行うと全く異なるモデルそれぞれの特徴が混ざり合い、結果的に超解像精度が落ちてしまう。



(a) 超解像3次元モデル



(b) 高解像3次元モデル

図2 従来の超解像結果と高解像度3次元モデルの比較

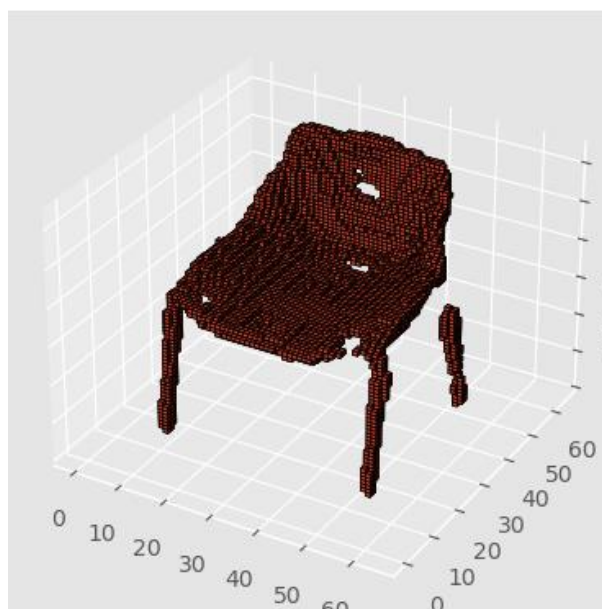


図3 複数クラス学習での超解像結果

## 2.6. CGAN

複数クラスのデータを同時に扱う関連研究として、Conditional Generative Adversarial Networks (CGAN) [3]という手法がある。

CGAN は、Generator と Discriminator の入力に追加の条件情報を与えて学習させている。条件情報には、クラス識別情報以外にも画像やテキストデータなどが使用されることもある[12][13]。通常の GAN がランダムな画像を生成するのに対し、CGAN は条件情報により生成物の制御が可能になっている。そのため、CGAN は特に条件付きのデータ生成において優れた性能を発揮する。例えば、特定の数字の手書き文字を生成する、特定のクラスの画像を生成する、条件付きの画像変換などがある。クラス識別情報の与え方には様々な手法[14][15]が存在する。Sage らの手法 [14]では、Generator の各層へ one-hot 表現のクラス識別情報を一様に与えている。これにより、出力層付近でクラス識別情報が消失しないようにしている。また、ACGAN[15]では、Discriminator にクラス識別情報を与えず、Discriminator で入力画像の真贋判定に加えてクラス識別も行っている。これにより、任意のクラスかつ高品質な画像の生成を実現している。

### 3. 提案手法 3D-SRCGAN による 3 次元モデル超解像

#### 3.1. クラス識別情報の導入

2.5 節で示したように、従来の 3D-SRGAN の超解像精度は改善の余地がある。この問題に対し、本研究では CGAN の条件情報としてクラス識別情報を与えることで、複数クラスが混在するデータセットから任意のクラスを生成できるという特徴を応用した 3D-Super Resolution Conditional Generative Adversarial Networks (3D-SRCGAN) を提案する。CGAN では、入力画像にクラス識別情報を付加して学習することによって、生成に制限をかけ、クラスを制御している。一方 3D-SRCGAN ではクラス識別情報で制限をかけるのではなく、クラス識別情報を付加することで各クラスの特徴が混ざり合わないようになっている。これにより従来の 3D-SRGAN の、異なるクラスのデータを同時に学習することで各モデルの特徴が混ざり合い、超解像精度が低下するという問題点を解消し、超解像精度が向上すると考える。さらに、複数クラスのデータを同時に学習することで、学習データ数が増え、超解像精度が向上すると考える。

条件情報の付加について、CGAN は、GAN の入力であるノイズに対して 2.6 節で示したように様々な方法で条件情報を付加している。3D-SRCGAN では、通常の CGAN[3]と同様に、3D-SRGAN の 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/高解像度 3 次元モデルそれぞれに one-hot 表現のクラス識別情報を付加する (図 4, 5)。one-hot 表現とは、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、クラスの情報を与える手法である。この手法は、他の手法に比べてシンプルであり様々な手法に応用しやすい。そのため、本研究ではこの手法を用いて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、3D 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 3D 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。例えば、10 クラスのクラス識別情報を付加する場合、10 チャンネルの空の 3D モデルを用意する。このとき、入力モデルに割り振ったクラス番号が 5 だった場合、

5に相当する6番目の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+クラス数分のチャンネルを持つ3Dモデルがネットワークへの入力となる。

ネットワーク構造については従来の3D-SRGANから変更はない。CGANでは、ネットワークが深すぎる場合、条件情報が考慮しにくくなり出力層付近で情報が失われてしまう可能性がある。実際、この問題はPoly-gan [16]で主張されており、Generatorの入力層以外にも条件を与えることで問題に対処している。本研究では、この問題に対応するために層の深さを変更して予備実験を行ったが、特に結果に変化は見られなかった。また、層を深くしすぎるとGPUの性能に依存して実験が行えない可能性がある。以上の理由により、3D-SRGANのネットワーク構造でも十分に条件情報を保持できていると考え、変更する必要はないと判断した。

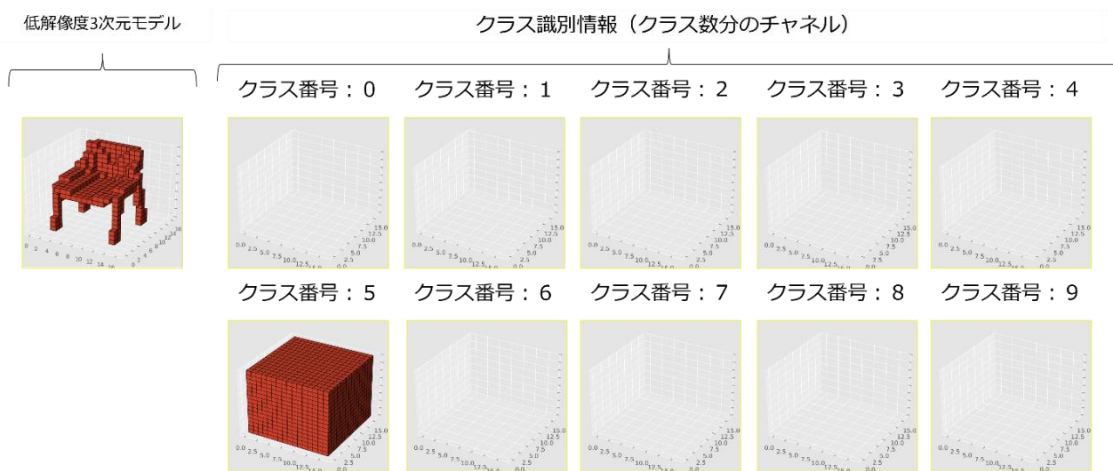


図4 Generatorの入力



図5 Discriminatorの入力

### 3.2. 3D-SRCGAN のネットワーク構造

3D-SRCGAN のネットワーク構造は、基本的に 3D-SRGAN[1]と同様である。

3D-SRCGAN は GAN と同じく Generator と Discriminator の 2つのネットワークで構成されており、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(図 6)。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つは敵対的関係にあり、それぞれの目的関数は、Generator は Discriminator を騙すように学習データと似た 3次元モデルを生成することであり、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 の最終的な目的は、Discriminator が識別できないような 3次元モデルを Generator が生成できることである。また、学習の結果得られた Generator が生成モデルとなる。

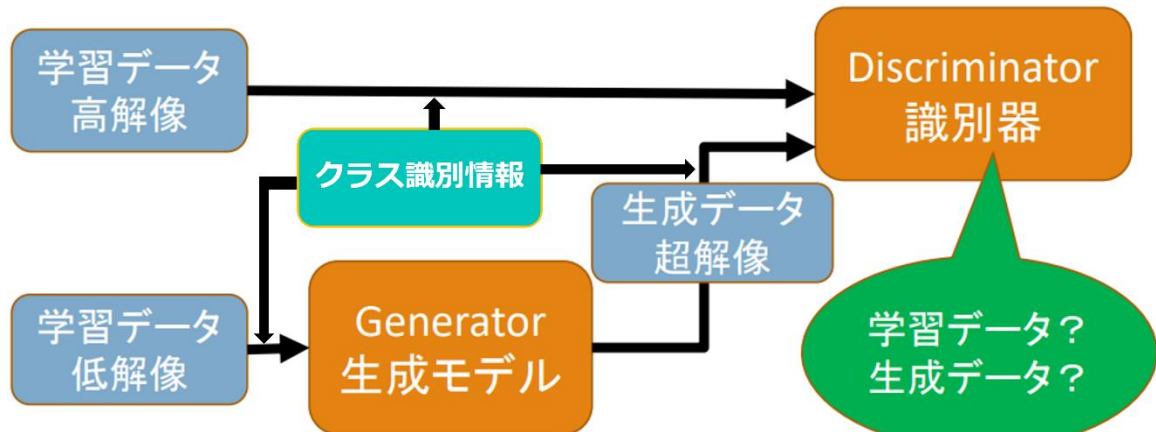


図 6 3D-SRCGAN の概要図



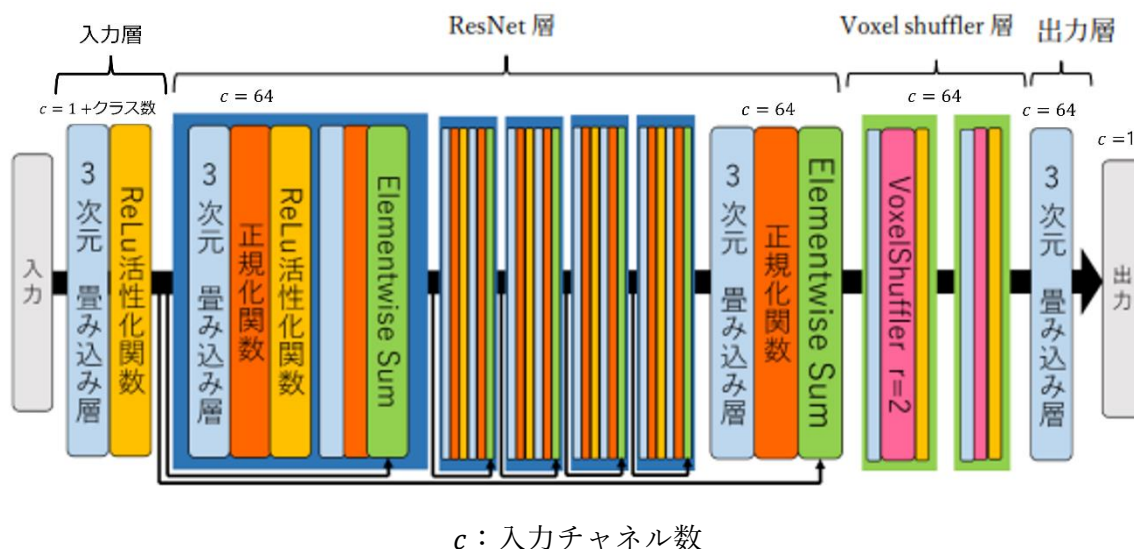


図7 Generatorの構造

### 3.3. 3D-SRCGANのGeneratorの構造

3D-SRCGANのGeneratorに関しては、入力層への入力チャンネル数以外に従来の3D-SRCGAN[1]と変化はない。

Generatorは図7のような構造になっている。以降、入力のチャンネル数を $C_{in}$ 、出力のチャンネル数を $C_{out}$ 、ストライドを $s$ 、カーネルのサイズを $k$ とする。ここでストライドとは、畳み込み処理を行うときにカーネルをずらしていく間隔のことである。Generatorの3次元畳み込み層のストライドは全て $s=1$ である。

入力には学習データの低解像度3次元モデルが与えられ、出力として高解像度3次元モデルを生成する。

Generatorの構造は大きく入力層、ResNet層、Voxel Shuffler層、出力層の4つに分けられる。入力層では、入力で与えられた3次元モデルを3次元畳み込み層によって、複数のチャンネルとしてResNet層に与えている。ResNet層では3次元モデルの特徴を抽出し、Voxel shuffler層で3次元モデルを拡大している。

最後に出力層で入力の特徴マップから3次元モデルを生成し、出力している。

3D-SRCGANのGeneratorでは、3次元畳み込み層を利用している。3次元の畳み込み層では、カーネルのサイズが $k \times k \times k$ の畳み込み処理をする。

3次元畳み込み層の後の処理に、正規化関数や活性化関数がある。正規化関数では3次元畳み込み層の出力を正規化している。この処理をすることで学習速度を速くす

ることができ、学習の安定化を図ることができる[17]。Generator では活性化関数に ReLu 活性化関数(図 8)を使用しており、式(1)で表現できる。

$$f(x) = \max(0, x) \quad (1)$$

次にそれぞれの層の処理を説明する。まず、入力層では、 $C_{in} = 1 + \text{クラス数}$ 、 $C_{out} = 64$ 、 $k = 9$  の 3 次元畳み込み層、ReLu 活性化関数を順番に処理する。

ResNet 層は 5 つの Residual Block で構成されている。Residual Block の中身は、 $C_{in} = 64$ 、 $C_{out} = 64$ 、 $k = 3$  の 3 次元畳み込み層、正規化関数、ReLu 活性化関数、 $C_{in} = 64$ 、 $C_{out} = 64$ 、 $k = 3$  の 3 次元畳み込み層、正規化関数の順番で構成されている。それぞれの Residual Block では入力を出力と加算している。即ち各 Residual Block はそれぞれの入力と出力の差を学習することになる。5 つの Residual Block の処理の後、 $C_{in} = 64$ 、 $C_{out} = 64$ 、 $k = 3$  の 3 次元畳み込み層、正規化関数で処理を行い、ResNet 層の入力を加算して出力している。

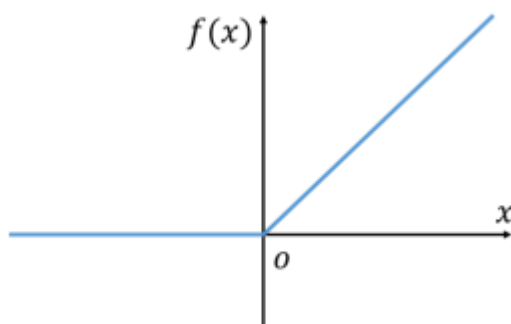


図 8 ReLu 活性化関数

Voxel shuffler 層では、実際に 3 次元モデルのサイズを拡大していく(図 9)。そのために入力の特徴マップを並び替えて 3 次元モデルを拡大することができる Voxel Shuffler を利用する。Voxel shuffler 層は 2 つのブロックで構成されており、1 つの Block で入力の特徴マップから 2 倍に拡大している。1 つのブロックの中身は  $C_{in} = 64$ 、 $C_{out} = 512$ 、 $s = 1$ 、 $k = 3$  の 3 次元畳み込み層、 $C_{in} = 512$ 、 $C_{out} = 64$  の Voxel Shuffler 層、Relu 活性化関数を順に処理している。この 2 つのブロックを通過した出力は 4 倍に拡大されている。

最後に  $C_{in} = 64$ 、 $C_{out} = 1$ 、 $s = 1$ 、 $k = 9$  の 3 次元畳み込み層で処理し、高解像度 3 次元モデルを出力している。

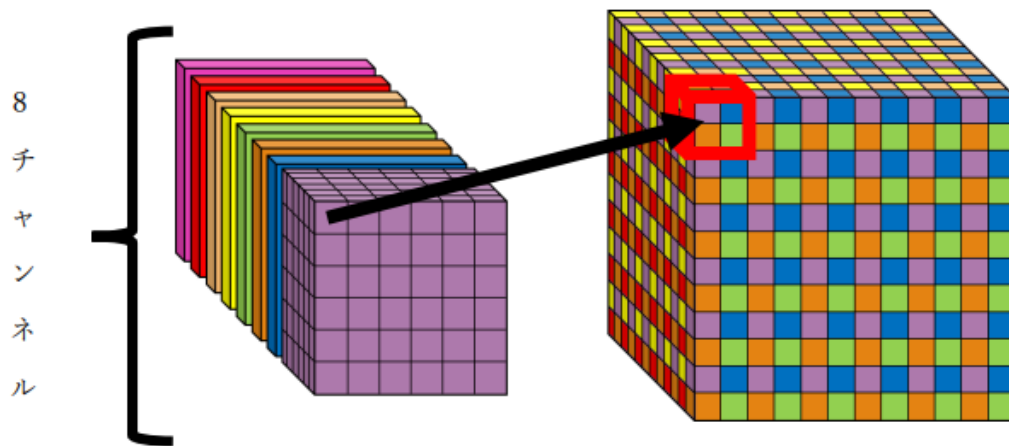


図9 Voxel Shuffler

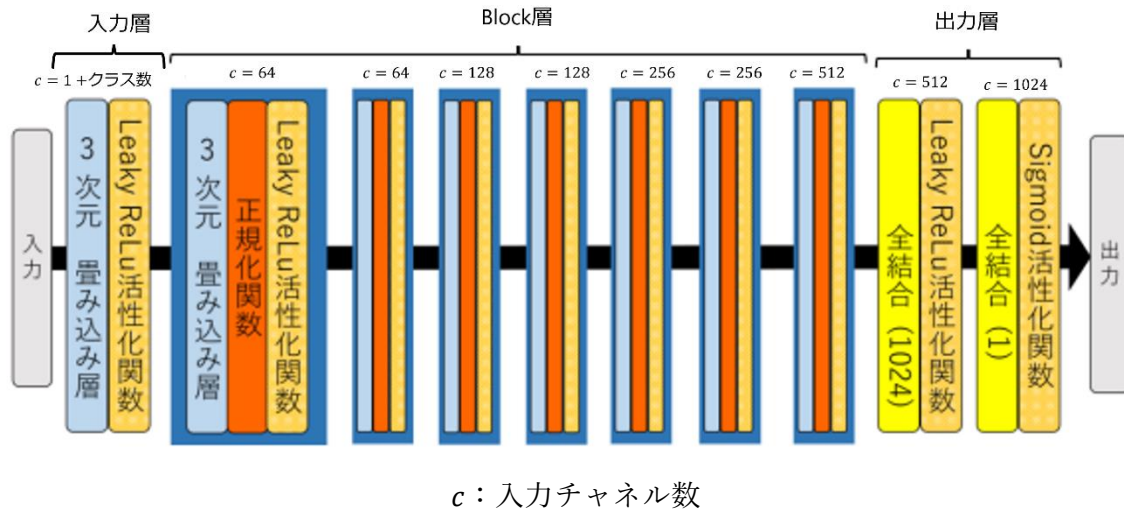


図 10 Discriminator の構造

### 3.4. 3D-SRCGAN の Discriminator の構造

3D-SRCGAN の Discriminator に関しても、入力層への入力チャネル数以外に従来の 3D-SRCGAN[1] と変化はない。

Discriminator は図 10 の構造になっている。Discriminator の構造は大きく入力層、Block 層、出力層の 3 つに分けられる。入力層では入力の 3 次元モデルから複数のチャネルを出力し、Block 層で 3 次元モデルの特徴を抽出し、出力層では入力の 3 次元モデルが学習データである確率が出力される。

3D-SRCGAN の Discriminator でも Generator と同様に 3 次元畳み込み層を利用し、3 次元モデルの畳み込み処理を行っている。

Discriminator では、活性化関数として Leaky ReLU 活性化関数(図 11)と Sigmoid 活性化関数(図 12)を使用しており、それぞれ式(2)(3)と表現できる。

$$f(x) = \max(\alpha x, x), 0 < \alpha < 1 \quad (2)$$

$$f(x) = \frac{1}{1 + \exp(-\beta x)}, 0 < \beta \quad (3)$$

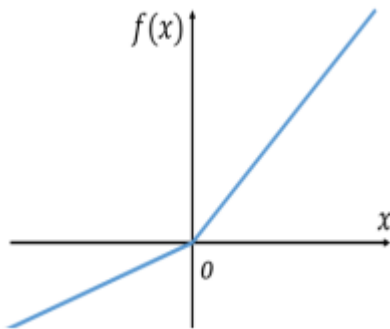


図 11 Leaky ReLu 活性化関数

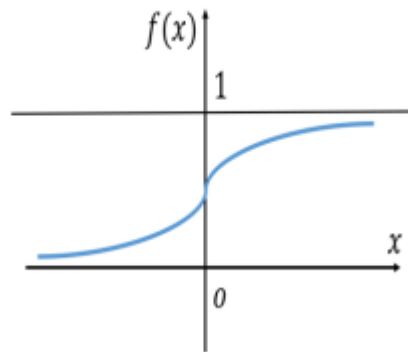


図 12 sigmoid 活性化関数

次にそれぞれの層の処理を説明する。まず、最初の層では $C_{in} = 1 + \text{クラス数}$ 、 $C_{out} = 64$ 、 $s = 1$  の 3次元畳み込み層、Leaky ReLu 活性化関数の順に処理している。次の Block 層は 7つのブロックで構成されている。1つのブロックの中身は、3次元畳み込み層、正規化関数、Leaky ReLu 活性化関数の順に処理している。また、それぞれの Block の 3次元畳み込み層のパラメータは、層毎に順に $C_{in} = 64, 64, 128, 128, 256, 256, 512$ 、 $C_{out} = 64, 128, 128, 256, 256, 512, 512$ 、 $s=2, 1, 2, 1, 2, 1, 2$ となっている。3次元畳み込み層のカーネルサイズは全て $k=3$ である。

7つのブロックの処理後、 $C_{out} = 1024$ の全結合、Leaky ReLu 活性化関数、 $C_{out} = 1$ の全結合、Sigmoid 活性化関数を順番に処理している。最終的な出力の値は $[0,1]$ の間の値となる。この値が、Discriminator の入力学習データである確率を表している。

### 3.5. Generator の学習更新

3D-SRCGAN の Generator の学習更新は 3D-SRGAN[1] と同様である。

Generator の Loss 関数  $l_G^{3DSR}$  は式(4)で表現できる。

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

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

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

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 には、元のデータの見た目がある程度そのままにして、別の操作を加えることができる特徴がある。そのため、超解像しても元のデータの見た目から大きく変化する心配がない。

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、GAN で通常使われている Loss 関数で、学習データと見分けがつかないように学習をする特徴がある。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、入力とは別の 3次元モデルが生成されてしまう。そのため、この 2つを適切に組み合わせることで超解像を行っている。

Generator の学習では、この Loss 関数の値が小さくなるよう、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。

### 3.6. Discriminator の学習更新

3D-SRCGAN の Discriminator の学習更新は 3D-SRGAN[1]と同様である

Discriminator の Loss 関数 $l_D^{3DSR}$ は式(7)で表現できる。

$$l_D^{3DSR} = \frac{1}{m} \sum^m [\log D(l^{HR}) - \log (1 - D(G(l^{LR})))] \quad (7)$$

$m$ はバッチサイズを表している。 $D(x)$ は Discriminator の出力であり、入力データ $x$ が学習データの 3次元モデルである確率を表している。そのため、Discriminator の識別がうまくいくと $D(l^{HR})$ の値が大きくなり、 $D(G(l^{LR}))$ の値は小さくなる。

Discriminator の学習も、Generator と同様の収束計算により行う。

## 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 ModelNet10[18]を使用した。このデータセットは bathtub、bed、chair、desk、dresser、monitor、night\_stand、sofa、table、toilet の 10 クラスの 3 次元モデルがあり、それぞれ学習データとテストデータが用意されている。表 1 にデータセット ModelNet10 の内訳を示す。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から、バッチサイズを 30、Epoch 数を 50 とした。

超解像 3 次元モデルの評価は、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。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。

1. 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)
2. 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)
3. 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)
4. 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)

表 1 ModelNet10 の内訳

クラス名	bathtub	bed	chair	desk	dresser
学習データ数	106	515	889	200	200
テストデータ数	50	100	100	86	86
クラス名	monitor	night_stand	sofa	table	toilet
学習データ数	465	200	680	392	344
テストデータ数	100	86	100	100	100



## 4.2. クラス識別情報を導入することによる精度向上の評価

3D-SRGAN には学習に使えるデータ数が少ない、学習に使っていないクラスの超解像精度が低いという問題がある。これらの問題は複数クラスのデータを同時に学習することで解決できると考え、予備実験を行った。しかし、複数クラスのデータを混ぜて 3D-SRCGAN を学習すると、データ数を増やすことはできるが、全く異なるモデルそれぞれの特徴が混ざり合い、超解像精度が落ちていた。クラス識別情報を付加することで各クラスの特徴が混ざり合うことなく学習ができ、超解像精度の改善が期待できる。

本実験では、複数クラスを混ぜたデータを使った超解像において、クラス識別情報が超解像精度向上に有効であるかを評価する。

### 4.2.1. 実験内容

本実験では、複数クラスを混ぜた同じデータを使用して、従来の 3D-SRGAN と 3D-SRCGAN の比較実験を行う。3D-SRCGAN にはクラス識別情報も付加することで、クラス識別情報の有無による超解像精度の差を評価する。ModelNet10 の 10 クラスのデータすべてを使用し、クラス識別情報以外のバッチサイズや学習回数といった条件は揃えている。

### 4.2.2. 実験結果

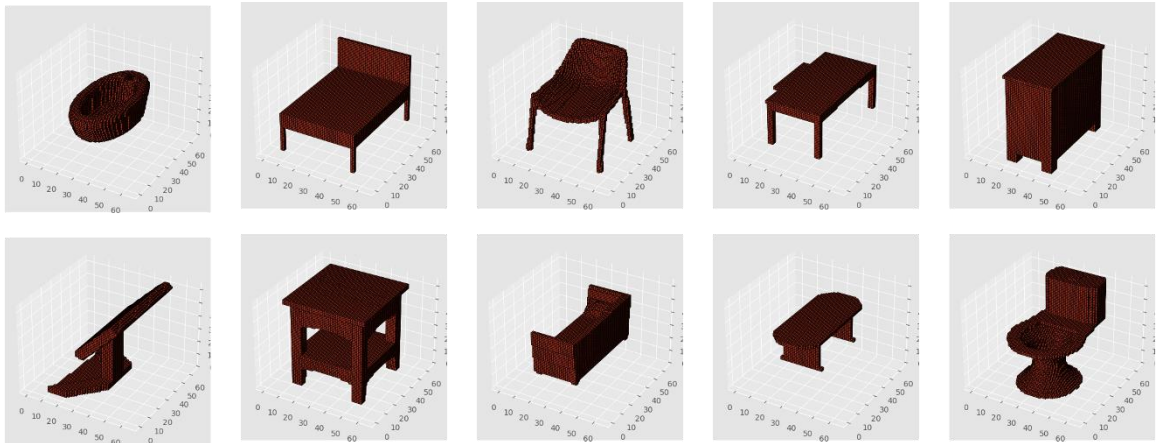
図 13 に 10 クラスのモデルに対する超解像結果の例を示す。上段 2 行が高解像度 3 次元モデル、中段 2 行が 3D-SRGAN、下段 2 行が 3D-SRCGAN の結果である。各列はクラスに対応している。1 行目は左から順に bathtub、bed、chair、desk、dresser、2 行目は左から monitor、night\_stand、sofa、table、toilet のクラスの超解像結果である。また、表 2 に図 13 に示した超解像結果に対する定量評価を、表 3 に 10 クラスのテストデータ全体の超解像結果に対する定量評価の平均を示す。表 2、3 では最も評価が高いものを太字にしている。

図 13 を主観的に見ると、3D-SRGAN に比べて 3D-SRCGAN のほうが高解像度 3 次元モデルに近いものを生成できていた。高解像度 3 次元モデルとそれぞれを見比べる

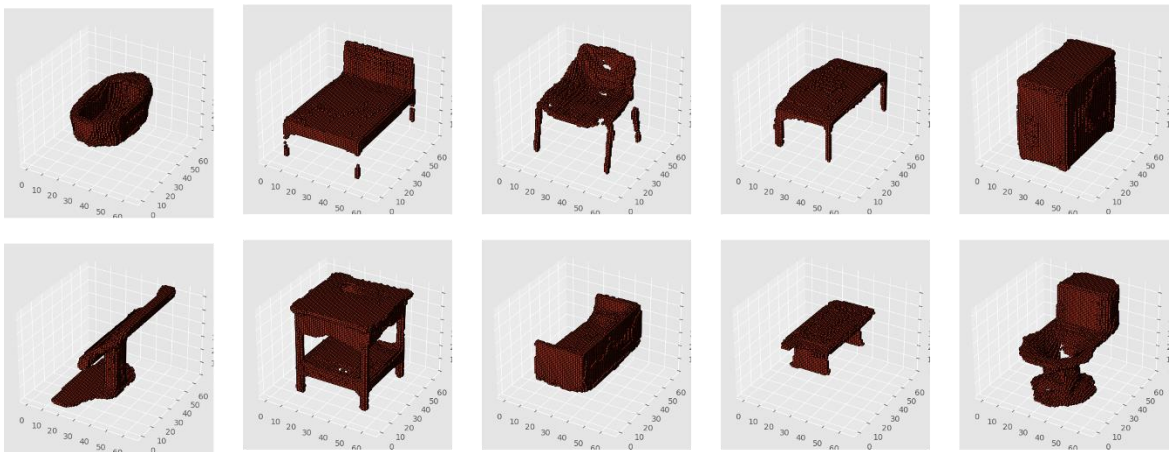
と、どちらも大まかな形状はうまく生成できている。また、直線や平面は綺麗に生成できているが、曲線や曲面、細い部分や小さい部分では、3D-SRGAN は穴が開いたり、本来繋がっている部分が途切れたりしている。一方、3D-SRCGAN では、bed と dresser の足や chair の座面、monitor の液晶部分など、3D-SRGAN では上手く生成できていない部分も綺麗に生成されており、より高解像度 3 次元モデルに近くなっている。しかし、3D-SRCGAN の超解像結果にも、chair の足や toilet の中央部分のように、穴や途切れが発生している 3 次元モデルもあった。

表 2 を見ると、連結成分、穴、空洞の誤差についてはどちらも大きな差はない。しかし、ボクセル値の誤差に関しては、すべてのクラスにおいて 3D-SRCGAN のほうが優れており、その差も大きかった。また、表 3 をみると、表 2 では大きな差がなかった連結成分、穴、空洞の誤差も、テストデータ全体で見ると 3D-SRCGAN のほうが小さい。ボクセル値の誤差に関しては 3D-SRCGAN は 3D-SRGAN の約  $\frac{1}{4}$  になっている。

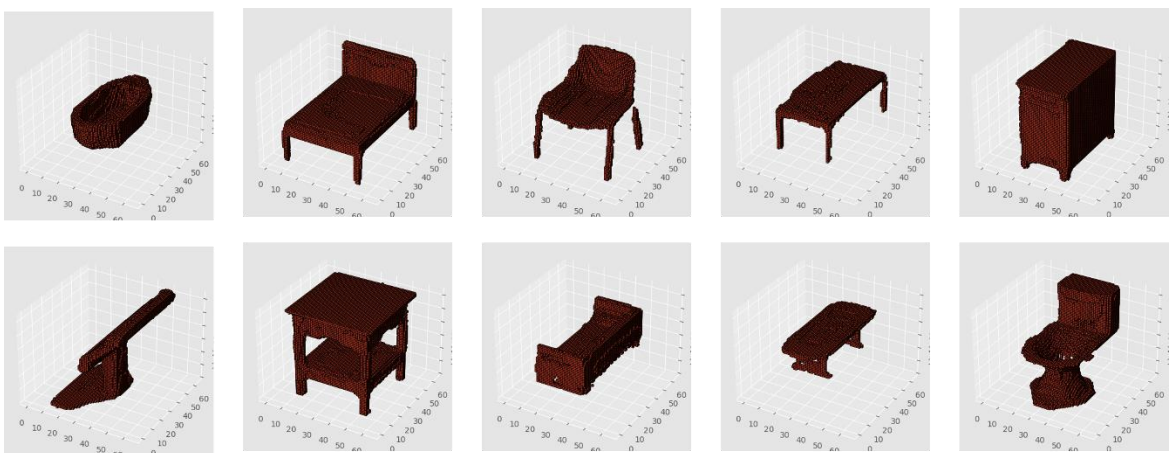
以上のことにより、複数クラスを混ぜたデータを使用した場合の 3 次元ボクセルモデル超解像におけるクラス識別情報の導入は、超解像精度向上に有効であることが分かった。



(a) 高解像度3次元モデル



(b) 3D-SRGAN



(c) 3D-SRCGAN

図 13 10 クラスの 3 次元モデルの超解像結果の例 (複数クラスでの学習結果)

表2 図13の3次元モデルの超解像結果に対する定量評価

	bathtub				bed				chair				desk			
評価項目	連	穴	空	値	連	穴	空	値	連	穴	空	値	連	穴	空	値
3D-SRGAN	0	2	0	27612	6	0	0	52545	4	21	0	11587	1	0	0	14590
3D-SRCGAN	0	19	1	<b>7022</b>	0	0	0	<b>3179</b>	3	12	0	<b>2982</b>	0	4	0	<b>4876</b>
	dresser				monitor				night_stand				sofa			
評価項目	連	穴	空	値	連	穴	空	値	連	穴	空	値	連	穴	空	値
3D-SRGAN	0	3	3	95424	0	1	1	29327	4	13	0	41798	2	8	2	33292
3D-SRCGAN	0	0	0	<b>13186</b>	0	2	2	<b>6828</b>	0	0	0	<b>9149</b>	10	27	3	<b>10657</b>
	table				toilet											
評価項目	連	穴	空	値	連	穴	空	値								
3D-SRGAN	1	6	0	12520	8	35	2	43468								
3D-SRCGAN	1	9	0	<b>2473</b>	0	47	0	<b>7832</b>								

表3 複数クラスでの学習結果の定量評価 (10クラスのテストデータ全体に対する平均値)

	連	穴	空	値
3D-SRGAN	2.7	6.9	1.8	47640.2
3D-SRCGAN	<b>2.2</b>	<b>5.6</b>	<b>1.3</b>	<b>10189.3</b>

### 4.3. 複数クラスで学習することによる精度向上の評価

4.2 節の実験から、複数クラスを混ぜたデータを使用した場合において、クラス識別情報が超解像精度向上に有効であることが分かった。しかし、複数クラスを混ぜたデータを使用した 3D-SRGAN は、従来の 1 クラスのデータで学習した場合よりも超解像精度が低いため、この実験だけでは 3D-SRCGAN が従来の 3D-SRGAN よりも優れているとは言えない。

本実験では、1 クラスで学習した 3D-SRGAN と複数クラスで学習した 3D-SRCGAN で比較を行い、クラス識別情報が超解像精度向上に有効であるかを評価する。

#### 4.3.1. 実験内容

本実験では、1 クラスで学習した 3D-SRGAN と複数クラスで学習した 3D-SRCGAN で比較実験を行う。3D-SRCGAN は実験 4.2 と同じ生成モデルを使っている。学習データとして 3D-SRGAN では ModelNet10 の chair クラスのデータのみ、3D-SRCGAN では ModelNet10 の 10 クラスすべてのデータを使用することで、複数クラスで学習することによる超解像精度の差を確認する。学習データの数は、3D-SRGAN が 889 個に対して、3D-SRCGAN は 3991 個である。学習データ、クラス識別情報の有無以外の条件は揃えている。

#### 4.3.2. 実験結果

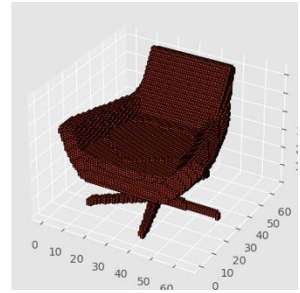
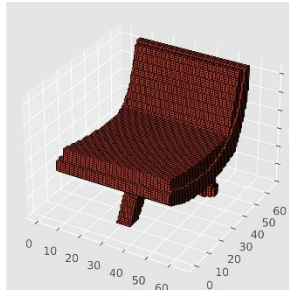
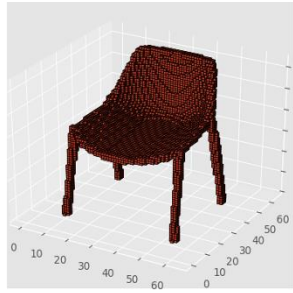
図 14 に chair クラスの 3 例の 3 次元モデルの超解像結果を示す。上段 2 行が高解像度 3 次元モデル、中段 2 行が 3D-SRGAN、下段 2 行が 3D-SRCGAN の結果である。各列は 3 例の chair クラス 3 次元モデルに対応している。左から順に chair 1、chair2、chair3 とする。また、表 4 に図 14 に示した超解像結果に対する定量評価を、表 5 に chair クラスのテストデータ全体の超解像結果に対する定量評価の平均を示す。表 4、5 では最も評価が高いものを太字にしている。

図 14 を主観的に見ると、3D-SRGAN に比べて 3D-SRCGAN のほうが高解像度 3 次元モデルに近いものを生成できていた。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うまく生成できている。しかし、3D-SRGAN では chair2 の座面や chair3 の足などが上手く生成できていなかった。一方、3D-SRCGAN では、3D-SRGAN では上手く生成できていない部分も綺麗に生成できており、より高解像度 3 次元モデルに近づいている。しかし、3D-SRCGAN の超解像結果にも、chair1 の足のように、途切れが発生している 3 次元モデルもあった。

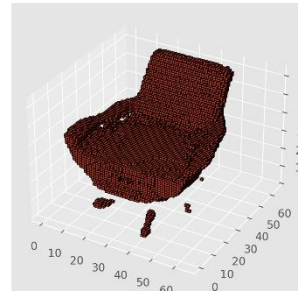
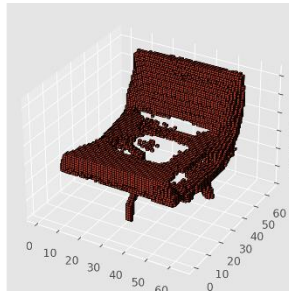
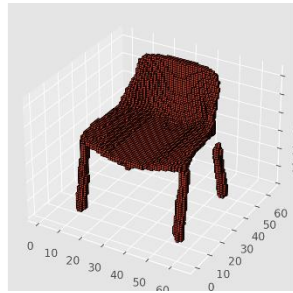
表 4 を見ると、連結成分、穴、空洞の誤差については大きな差はない。しかし、ボクセル値の誤差に関しては、3 例の 3 次元モデルすべてで 3D-SRCGAN のほうが優れていた。また、表 5 をみると、表 4 では大きな差がなかった連結成分、穴、空洞の誤差も、テストデータ全体で見ると 3D-SRCGAN の方が小さかった。ボクセル値の誤差については表 4 と同様に 3D-SRCGAN の方が小さかった。

以上のことにより、複数クラスのデータを使用し、データ数を増やすことは、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分かった。また、4.2 節の実験結果と合わせて、3D-SRCGAN は、複数クラス、1 クラスのどちらで学習した 3D-SRGAN よりも優れていることが分かった。

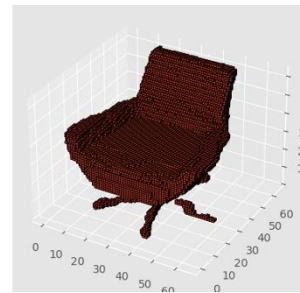
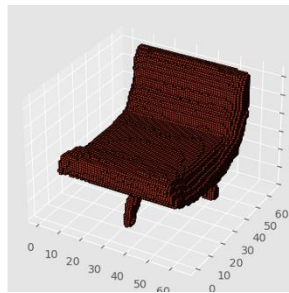
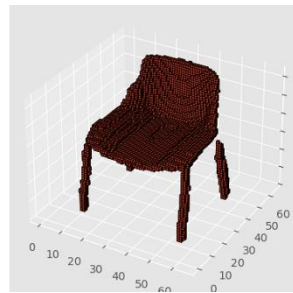
高解像度  
3次元  
モデル



3D-  
SRGAN  
(1クラス)



3D-  
SRCGAN  
(複数クラス)



chair 1

chair 2

chair 3

図 14 chair クラスの 3 例の 3 次元モデルの超解像結果 (1 クラスと複数クラスでの学習結果)

表 4 図 14 の 3 次元モデルの超解像結果に対する定量評価

評価項目	chair 1				chair 2				chair 3			
	連	穴	空	値	連	穴	空	値	連	穴	空	値
3D-SRGAN	1	0	0	4091	96	99	4	42477	51	31	7	24112
3D-SRCGAN	3	12	0	<b>2982</b>	0	2	5	<b>8876</b>	2	11	0	<b>6087</b>

表 5 1 クラスと複数クラスでの超解像結果の定量評価

(chair クラスのテストデータ全体に対する平均値)

	連	穴	空	値
3D-SRGAN (1 クラス)	5.9	8.5	0.7	7893.8
3D-SRCGAN (複数クラス)	<b>2.6</b>	<b>4.4</b>	<b>0.2</b>	<b>5968.6</b>



## 4.4. 総データ数の違いによる超解像精度への影響調査

機械学習において、高い精度を得るためには、大量のデータを学習する必要がある。しかし、実際は大量のデータを準備することが難しかったり、学習時間が長くなりすぎたり、少量のデータでも十分な精度が得られるなど、大量のデータが使われない場合もある。3D-SRCGAN が十分な超解像精度を実現するために必要なデータ数が分かれば、より効率的に学習や実験を行うことができる。

本実験では、ModelNet10 の学習データ 3991 個を、クラス数を変えることなく減らすことで 3D-SRCGAN の超解像精度にどのような影響が出るのかを調査する。

### 4.4.1. 実験内容

本実験では、ModelNet10 の学習データをすべて使った場合と、すべてのクラスのデータ数を 0.9 倍、0.7 倍、0.5 倍、0.3 倍、0.1 倍した場合の 6 パターンで比較を行い、学習データの数の違いのみによる超解像精度への影響を調査する。各パターンのデータ数は、すべて使った場合が 3991 個、0.9 倍が 3592 個、0.7 倍が 2794 個、0.5 倍が 1996 個、0.3 倍が 1197 個、0.1 倍が 399 個となっている。学習データの数以外の条件は揃えている。

### 4.4.2. 実験結果

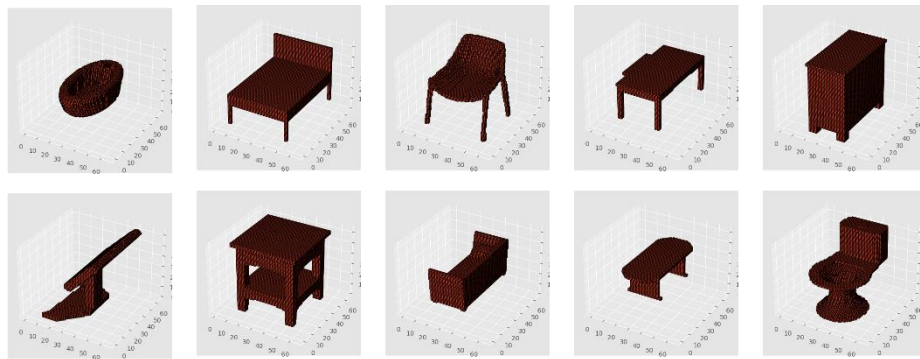
図 15 に 10 クラスの 3 次元モデルの超解像結果を示す。上から 2 行ずつ、高解像度 3 次元モデル、学習データすべて、0.9 倍、0.7 倍、0.5 倍、0.3 倍、0.1 倍の結果である。各列はクラスに対応している。1 行目は左から順に bathtub、bed、chair、desk、dresser、2 行目は左から monitor、night\_stand、sofa、table、toilet のクラスの超解像結果である。また、表 6 に図 15 に示した超解像結果に対する定量評価を、表 7 に 10 クラスのテストデータ全体の超解像結果に対する定量評価の平均を示す。表 6、7 では最も評価が高いものを太字にしている。

図 15 を主観的に見ると、大まかな形状については、学習データすべて、0.9 倍、0.7 倍、0.5 倍、0.3 倍の場合はいま生成できており、大きな差はない。一方、0.1 倍の場

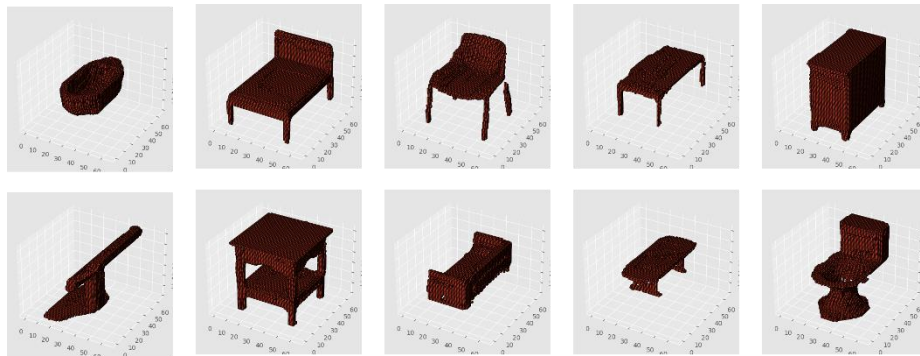
合は、大きく形状が歪んだり、細かい部分が消えてしまっているものがあった。また、直線や平面についてはどのパターンでも綺麗に生成できていたが、曲線や曲面、細い部分や小さい部分では、データ数が全体の半分以下のパターンは、bathtubの形状、bedとdresserの足、toiletの穴などが上手く生成できていなかった。一方、データ数が多いパターンでは、それらも上手く生成できていた。しかし、データ数が多いパターンでも、chairの右奥の足やsofaの下部のように、穴や途切れが発生している3次元モデルもあった。

表6を見ると、連結成分、穴、空洞の誤差には大きな差が見られなかった。ボクセル値の誤差に関しては、見た目ほどの差はなかったが、学習データをすべて使った場合と0.9倍にした場合が優れていることが多かった。表7を見ると、テストデータ全体ではすべての項目において、学習データをすべて使った場合が最も良い結果になっていた。

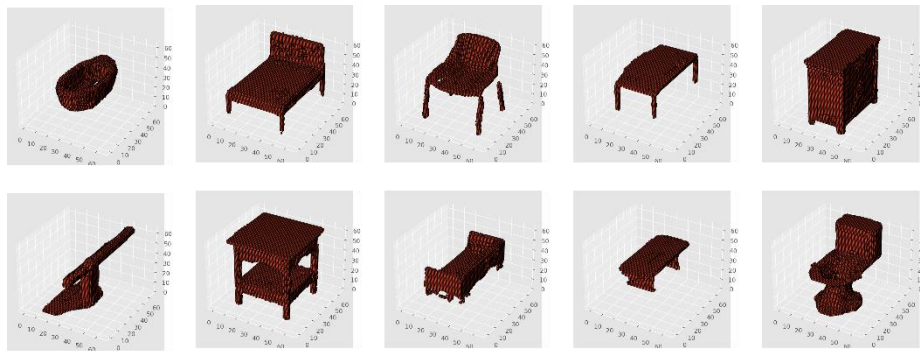
以上のことから、3D-SRCGANでは、学習データ数が多いほど超解像精度は高い。ただし、表7から、3000個ほどのデータがあれば、見た目、数値ともに実験を行うのに十分な精度が出せることが分かった。



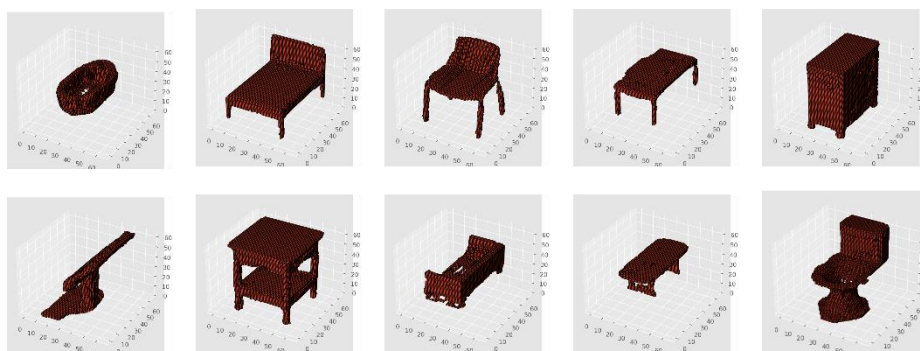
(a) 高解像度3次元モデル



(b) 1.0倍



(c) 0.9倍



(d) 0.7倍

図 15 10 クラスの 3 次元モデルの超解像結果の例 (異なるデータ数での学習結果)

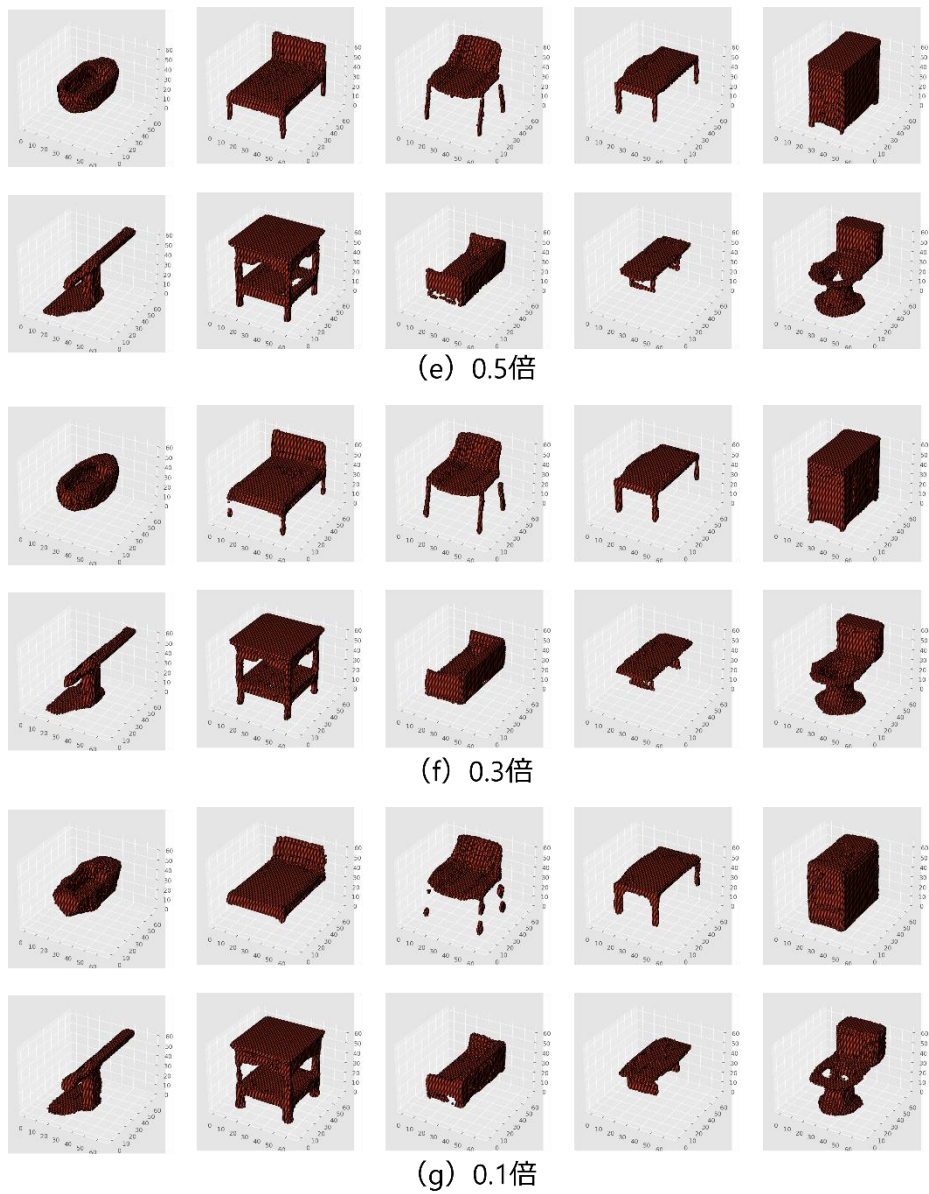


図 15 10 クラスの 3 次元モデルの超解像結果の例 (異なるデータ数での学習結果) (続き)

表 6 図 15 の 3 次元モデルの超解像結果に対する定量評価

評価項目	bathtub				bed				chair				desk			
	連	穴	空	値	連	穴	空	値	連	穴	空	値	連	穴	空	値
1.0 倍	0	19	1	7022	0	0	0	<b>3179</b>	3	12	0	<b>2982</b>	0	4	0	<b>4876</b>
0.9 倍	2	9	0	6533	1	0	0	4798	2	13	0	3354	0	0	0	5612
0.7 倍	1	13	0	6728	0	1	1	4373	2	6	0	3714	0	1	0	5200
0.5 倍	0	11	0	<b>6257</b>	0	0	0	4112	4	4	0	4089	0	0	0	5031
0.3 倍	0	<b>0</b>	0	6648	2	0	0	3409	2	<b>0</b>	0	4924	0	1	0	8172
0.1 倍	2	2	0	9557	2	20	18	7371	8	1	0	7464	0	0	0	13866
評価項目	dresser				monitor				night_stand				sofa			
	連	穴	空	値	連	穴	空	値	連	穴	空	値	連	穴	空	値
1.0 倍	0	0	0	13186	0	2	2	<b>6828</b>	0	<b>0</b>	0	9149	10	27	3	10657
0.9 倍	0	0	0	<b>12179</b>	0	0	0	7274	0	5	0	<b>8301</b>	15	22	8	9975
0.7 倍	0	0	0	12955	0	0	0	7520	0	6	0	8888	17	29	2	<b>9518</b>
0.5 倍	0	0	0	13476	0	2	2	7311	1	15	0	10876	17	19	2	11766
0.3 倍	0	0	0	12613	0	0	0	7424	9	15	0	9075	<b>5</b>	38	2	11762
0.1 倍	0	6	6	38475	3	7	5	10676	0	4	0	13774	6	<b>2</b>	1	12069
評価項目	table				toilet											
	連	穴	空	値	連	穴	空	値								
1.0 倍	1	9	0	<b>2473</b>	<b>0</b>	47	0	7832								
0.9 倍	<b>0</b>	5	0	2590	3	56	0	8029								
0.7 倍	1	8	0	2625	9	14	0	8325								
0.5 倍	2	17	0	4562	1	43	0	<b>7508</b>								
0.3 倍	7	4	0	4301	5	<b>13</b>	0	8535								
0.1 倍	1	<b>1</b>	0	4957	11	23	4	12433								

表 7 異なるデータ数での学習結果の定量評価 (10 クラスのテストデータ全体に対する平均値)

学習データ (個数)	連	穴	空	値
1.0 倍 (3991 個)	<b>2.2</b>	<b>5.6</b>	<b>1.3</b>	<b>10189.3</b>
0.9 倍 (3592 個)	2.5	5.8	1.4	10812.8
0.7 倍 (2794 個)	3.4	6.6	<b>1.3</b>	10730.3
0.5 倍 (1996 個)	3.2	6.8	1.5	10550.1
0.3 倍 (1197 個)	6.2	8.9	2.7	11295.6
0.1 倍 (399 個)	7.3	9.2	4.3	12702.1

## 4.5. クラス数の違いによる超解像精度への影響調査

4.3 節の実験から、3D-SRCGAN において複数クラスのデータを使用し、データ数を増やすことは、超解像精度の向上に有効であることが分かった。しかし、これまでの実験では、クラスごとのデータ数が少なくても学習クラス数を増やしたほうが良いのか、学習クラス数を減らして各クラスのデータ数を増やすほうが良いのかが分からない。前者は、各クラスのデータ数が少ないため細かい特徴まで学習できないが、多数のクラスの大まかな特徴を学習することができる。一方後者は、学習できる特徴の種類は少ないが、それぞれの特徴を細かく学習することができる。学習クラス数とデータ数の関係性が分かれば、同じ時間でより高精度な超解像を行うことができる。

この実験では、学習データの総数を固定し、学習クラス数を変更しながら比較を行うことで、3D-SRCGAN における学習データのクラス数の違いによる超解像精度への影響調査を行う。

### 4.5.1. 実験内容

本実験では、学習データの総数を固定し、学習クラス数が 5、6、7、8、15 の 5 パターンで比較を行い、クラス数の違いによる超解像精度の差を評価する。学習データの総数については、実験 4.4 から 3000 個ほどのデータで十分な精度が得られることが分かっているため、約 3000 個になるようにクラスを選び、実験を行う。本実験では、10 クラス以上の場合と比較するために、ModelNet10 に加えてその拡張版である ModelNet40[18]の一部のデータを使用している。表 8 に ModelNet40 内の、本実験で使ったクラスの内訳を示す。学習データ以外のバッチサイズや学習回数といった条件は揃えている。

各実験で使うデータの内訳を表 9 に示す。比較のためにすべてのパターンで chair クラスを使用している。

表 8 ModelNet40 の内訳

クラス名	airplane	bench	cone	flower_pot	guitar	lamp
学習データ数	626	173	167	149	155	124
テストデータ数	100	20	20	20	100	20
クラス名	laptop	piano	sink	tent	tv_stand	
学習データ数	149	231	128	163	267	
テストデータ数	20	100	20	20	100	

表 9 異なるクラス数での学習実験での学習データの内訳

クラス名	bed	chair	monitor	sofa	table	
5 クラス	515	889	465	680	392	
クラス名	bathhtub	bed	chair	monitor	sofa	toilet
6 クラス	106	515	889	465	680	344
クラス名	chair	desk	dresser	monitor	night_stand	sofa
7 クラス	889	200	200	465	200	680
	toilet					
	344					
クラス名	bathhtub	chair	desk	dresser	night_stand	sofa
8 クラス	106	889	200	200	200	680
	table	toilet				
	392	344				
クラス名	airplane	bench	chair	cone	desk	flower_pot
15 クラス	300	173	400	167	200	149
	guitar	lamp	laptop	monitor	night_stand	piano
	155	124	149	465	200	231
	sink	tent	tv_stand			
	128	163	267			

## 4.5.2. 実験結果

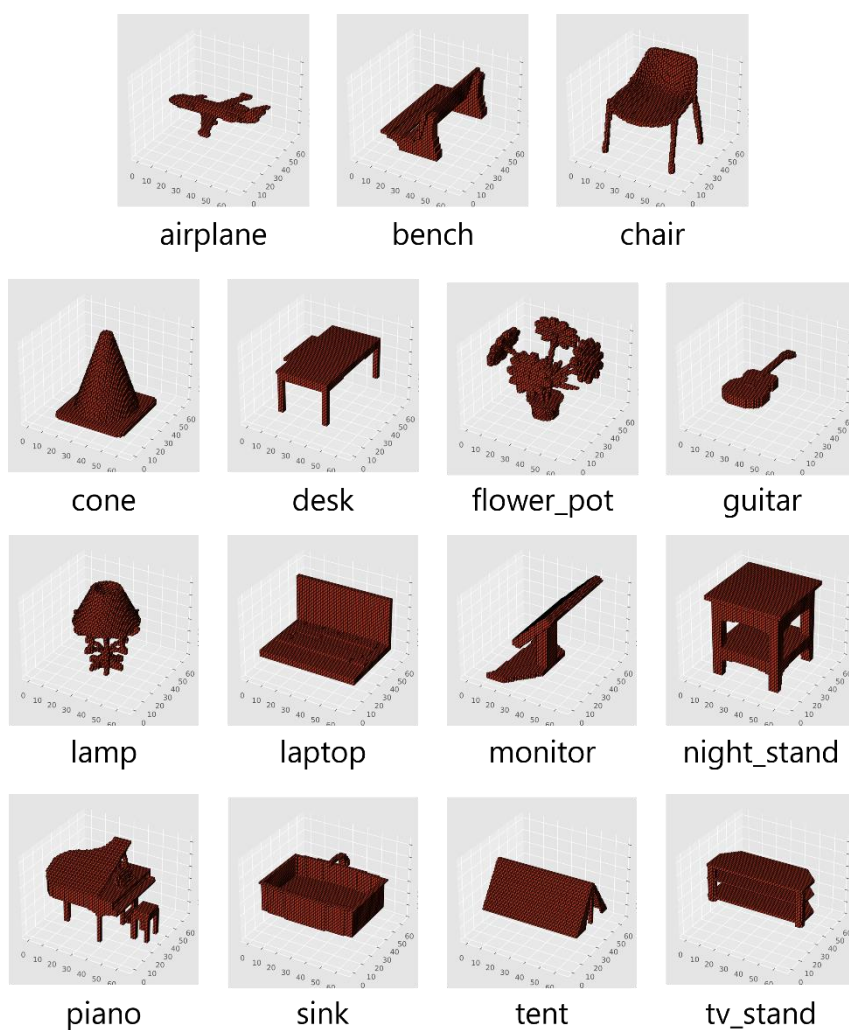
図 16 に各クラスの 3 次元モデルの超解像結果を示す。上から順に高解像度 3 次元モデル、5 クラス学習、6 クラス学習、7 クラス学習、8 クラス学習、15 クラス学習の結果である。本実験では、比較対象ごとに使用しているクラスが違うため、図 16 にクラス名も示している。また、表 10 にはテストデータ全体の超解像結果に対する定量評価の平均を示す。表 10 では最も評価が高いものを太字にしている。

図 16 を主観的に見ると、6 クラス学習と 8 クラス学習の超解像結果が高解像度 3 次元モデルに近いものを生成できていた。5 クラス学習の場合は sofa と table に穴が開いており、7 クラス学習の場合は chair、sofa、toilet に穴が開いていた。また、15 クラス学習の場合は flower\_pot、night\_stand、sink、tv\_stand で穴や途切れが発生していた。一方で、6 クラス学習と 8 クラス学習の場合ではそれらの問題があまり見られなかった。

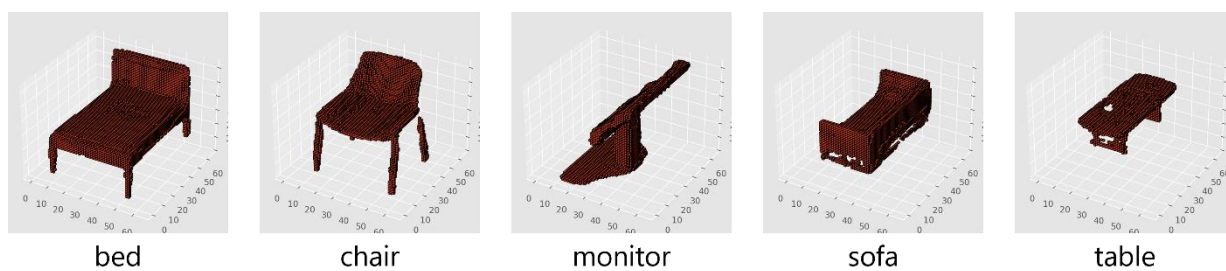
表 10 を見ると、図 16 の評価と違い、連結成分、穴、空洞の誤差は 5 クラス学習の場合が、ボクセル値の誤差は 15 クラス学習の場合が最も良い評価になっていた。連結成分、穴、空洞の誤差に関しては、8 クラス学習の場合が目立ちにくい小さな穴がいくつも存在しているのに対し、5 クラス学習の場合は大きな穴がいくつか存在していることで、見た目の評価と一致していない場合があると思われる。

以上のことより、学習に使うクラス数の違いによって超解像結果の違いはあるものの、精度については特に傾向は見られなかった。これは、総データ数を固定した影響で、クラス数を増やした際の、データ数を増やせるという利点を生かすことができなかつたためだと考える。また、クラスによっては形状に大きな差がないものも存在する。しかし、本実験では、クラスを選ぶ際に形状の多様性を評価していない。この多様性を考慮してクラスを選択することで、超解像精度の向上につながる可能性がある。



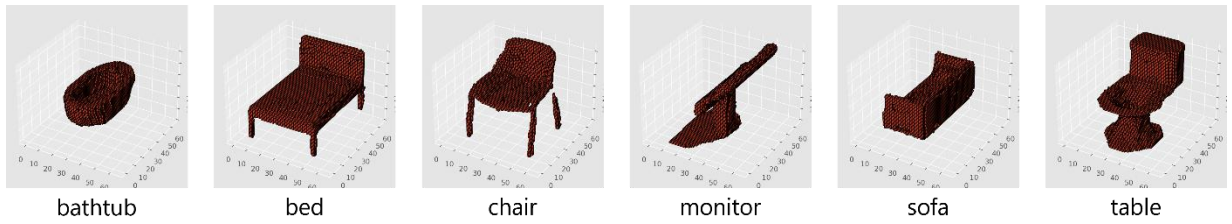


(a) 高解像度3次元モデル

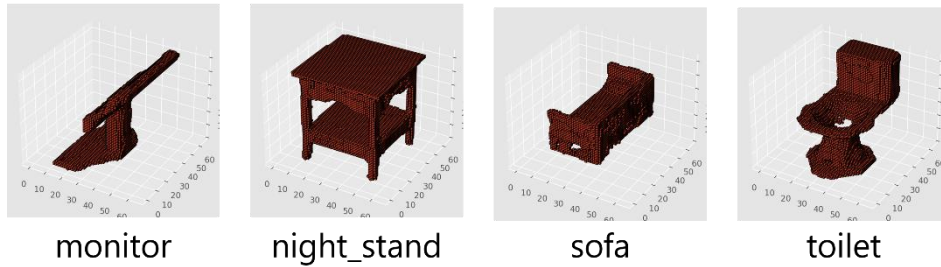
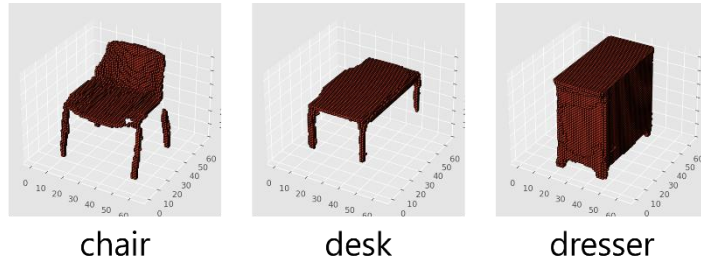


(b) 5クラス学習

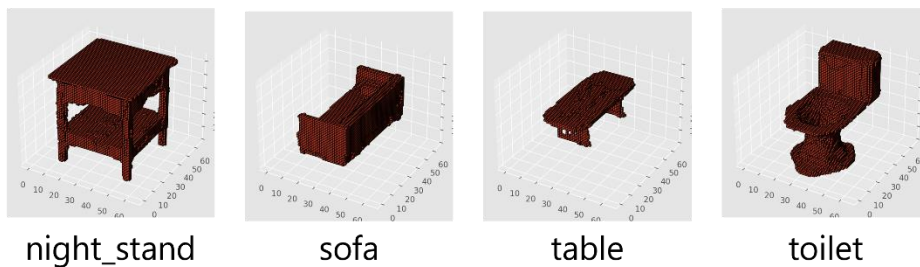
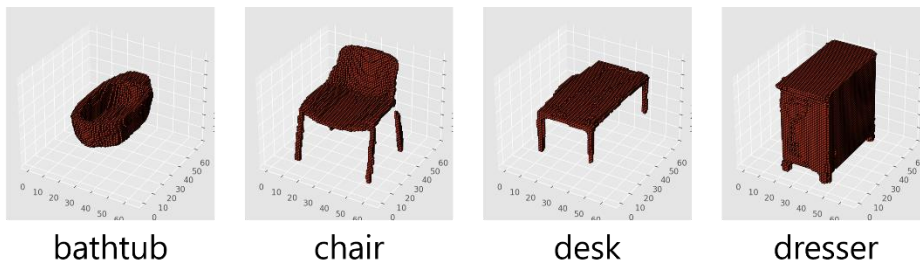
図 16 3次元モデルの超解像結果の例 (異なるクラス数での学習結果)



(c) 6クラス学習

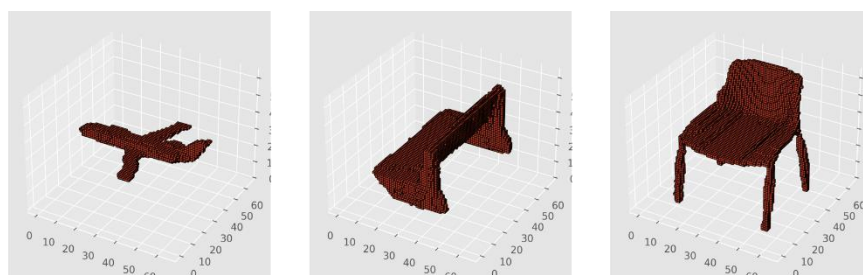


(d) 7クラス学習



(e) 8クラス学習

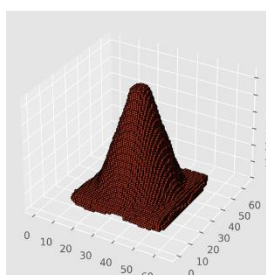
図 16 3次元モデルの超解像結果の例（異なるクラス数での学習結果）（続き）



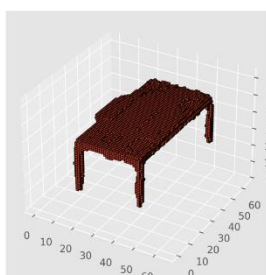
airplane

bench

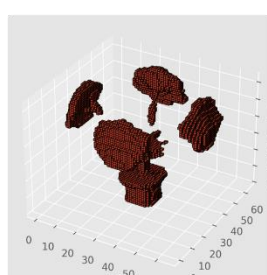
chair



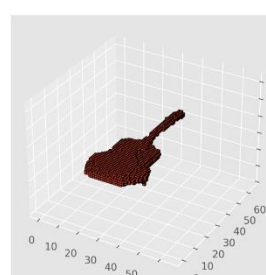
cone



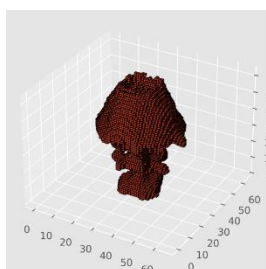
desk



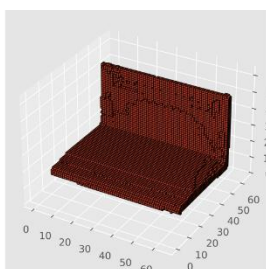
flower\_pot



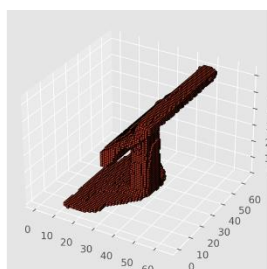
guitar



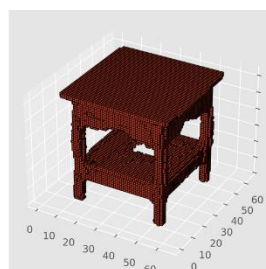
lamp



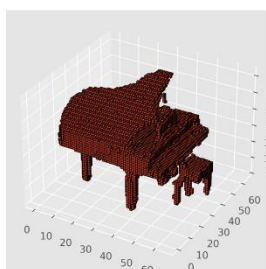
laptop



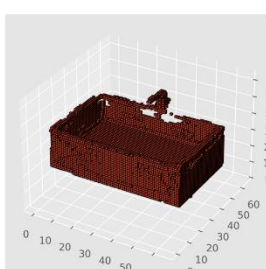
monitor



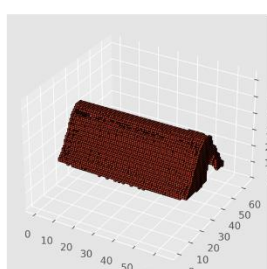
night\_stand



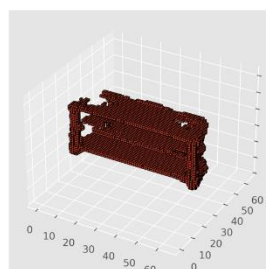
piano



sink



tent



tv\_stand

(f) 15クラス学習

図16 3次元モデルの超解像結果の例 (異なるクラス数での学習結果) (続き)

表 10 異なるクラス数での学習結果の定量評価

(各クラス数のテストデータ全体に対する平均値)

	連	穴	空	値
5 クラス学習	1.6	3.3	0.5	7263.3
6 クラス学習	1.7	4.3	0.5	7816.1
7 クラス学習	3.6	7.5	3.1	13774.8
8 クラス学習	3.1	8.1	1.9	10652.1
15 クラス学習	2.6	6.9	0.9	<b>7016.57</b>

## 4.6. 異なるクラス識別情報が超解像精度に与える影響の調査

これまでの実験で、学習時にクラス識別情報を付加することで、複数クラスの学習データを適切に学習し、超解像精度を向上できることが分かった。本実験では、超解像 3 次元モデルの生成時にクラス識別情報がどの程度超解像精度に影響を与えるかを調査する。

### 4.6.1. 実験内容

本実験では、実験 4.5 の生成モデルを使い、chair クラスの 3 次元モデルに別のクラス識別情報を与えて超解像を行い、正しいクラス識別情報を与えた場合の結果と比較する。学習データは、実験 4.5 と同様に ModelNet10 に加え、ModelNet40 の一部を使用している。

### 4.6.2. 実験結果

図 17 に chair クラスの 3 次元モデルの超解像結果を示す。上から順に高解像度 3 次元モデル、5 クラス学習、6 クラス学習、7 クラス学習、8 クラス学習、15 クラス学習の結果である。各列は与えたクラス識別情報に対応している。本実験では、比較対象ごとに与えるクラス識別情報が違うため、図 17 にクラス名も示している。また、表 11 に図 17 に示した超解像結果に対する定量評価を、表 12 に各学習クラス数について、クラス識別情報ごとの chair クラスのテストデータ全体の超解像結果に対する定量評価の平均を示す。表 11、12 では最も評価が高いものを太字にしている。

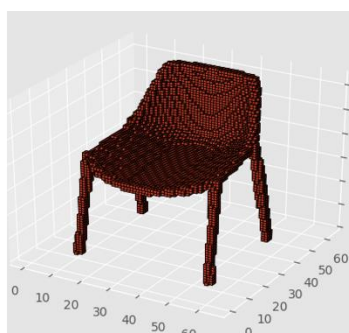
図 17 を見ると、どのクラス数での学習でも正しいクラス識別情報を与えたものが最も高解像度 3 次元モデルに近い。このことから、クラス識別情報が正しく働いていることが分かる。また、クラスに着目すると、bed や dresser などの直線や平面ばかりで構成されているクラスのクラス識別情報を与えると、曲線や曲面が多い chair の座面や背もたれを上手く生成できていなかった。また、bathtub や airplane、cone、guitar、lamp などの chair と共通点がほとんどないクラスのクラス識別情報を与えた場合でも上手く超解像で

きていなかった。一方、benchのようにchairと似ているクラスのクラス識別情報を与えた場合では、比較的高精度な超解像が行えていた。

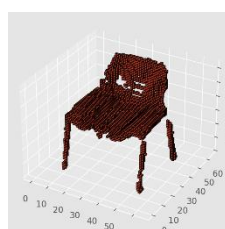
表 11 を見ると、連結成分、穴、空洞の誤差は chair クラスのクラス識別情報を与えたときに最も小さくなるが多かった。また、5 クラス学習で sofa クラスのクラス識別情報を与えた場合や、15 クラス学習で laptop クラスのクラス識別情報を与えた場合など、図 17 で上手く生成できていたものも chair クラスのクラス識別情報を与えた時と同じように良い結果が出ている。ボクセル値の誤差に関しては、学習したクラス数に関係なく、ほとんどの場合で chair クラスのクラス識別情報を与えたときの誤差が最も小さくなっている。また、表 12 から chair クラスのテストデータ全体で見ても、表 11 と同様の傾向が見られた。

このような結果から、3D-SRCGAN では、超解像対象と同じクラスのデータを使って学習することで最大限の超解像精度を実現することができる。また、同じクラスで学習していなかったとしても、ある程度似ているクラスを学習に使っていれば高精度な超解像を行うことができる。このことから、3D-SRCGAN においてクラス識別情報は、学習時だけでなく、超解像時にも有効であることが分かった。

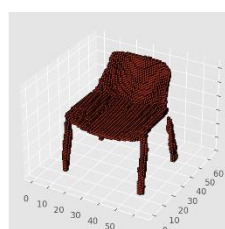




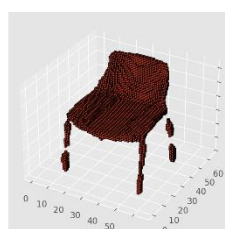
(a) 高解像度3次元モデル



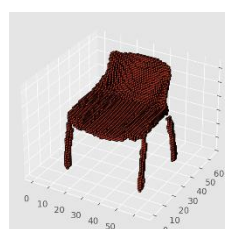
bed



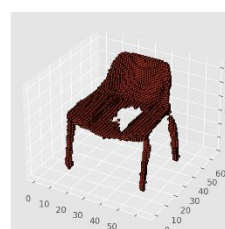
chair



monitor

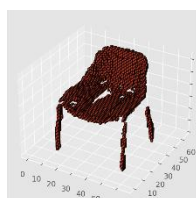


sofa

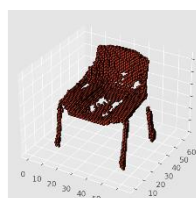


table

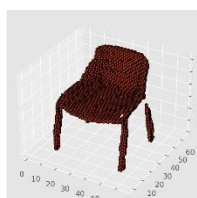
(b) 5 クラス学習



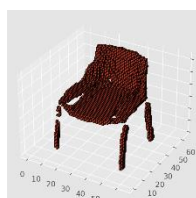
bathtub



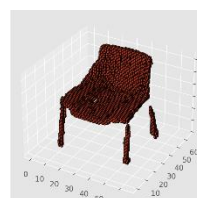
bed



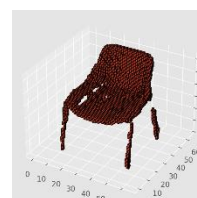
chair



monitor



sofa



table

(c) 6 クラス学習

図 17 chair クラスの 3 次元モデルの超解像結果の例 (クラス識別情報ごとの結果)

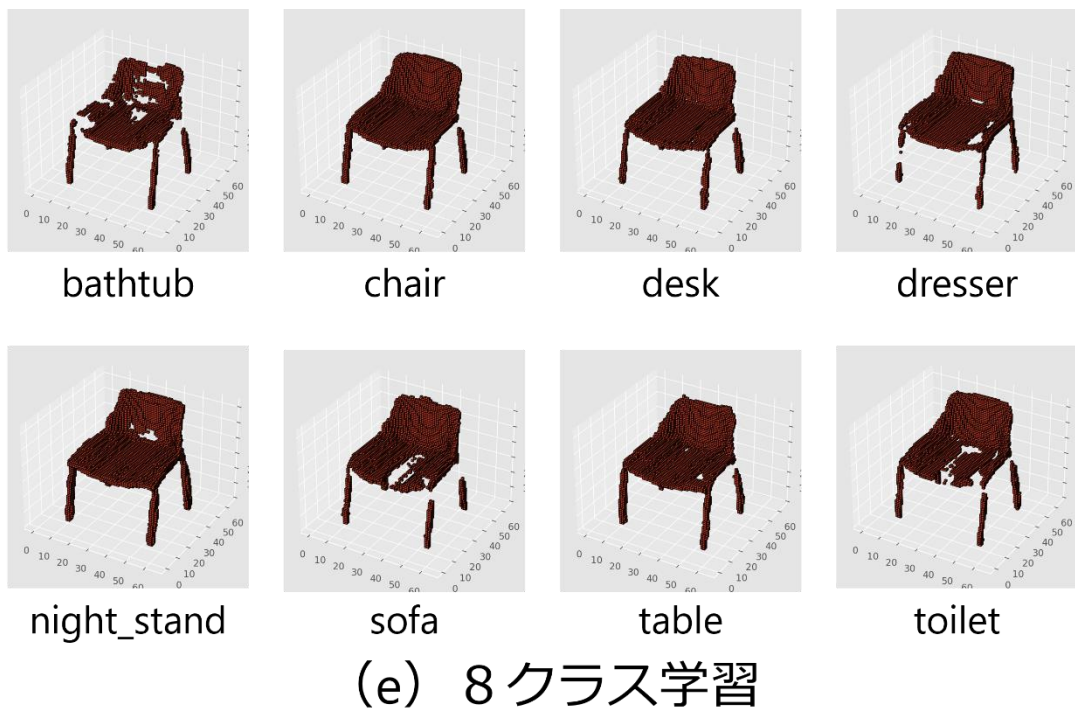
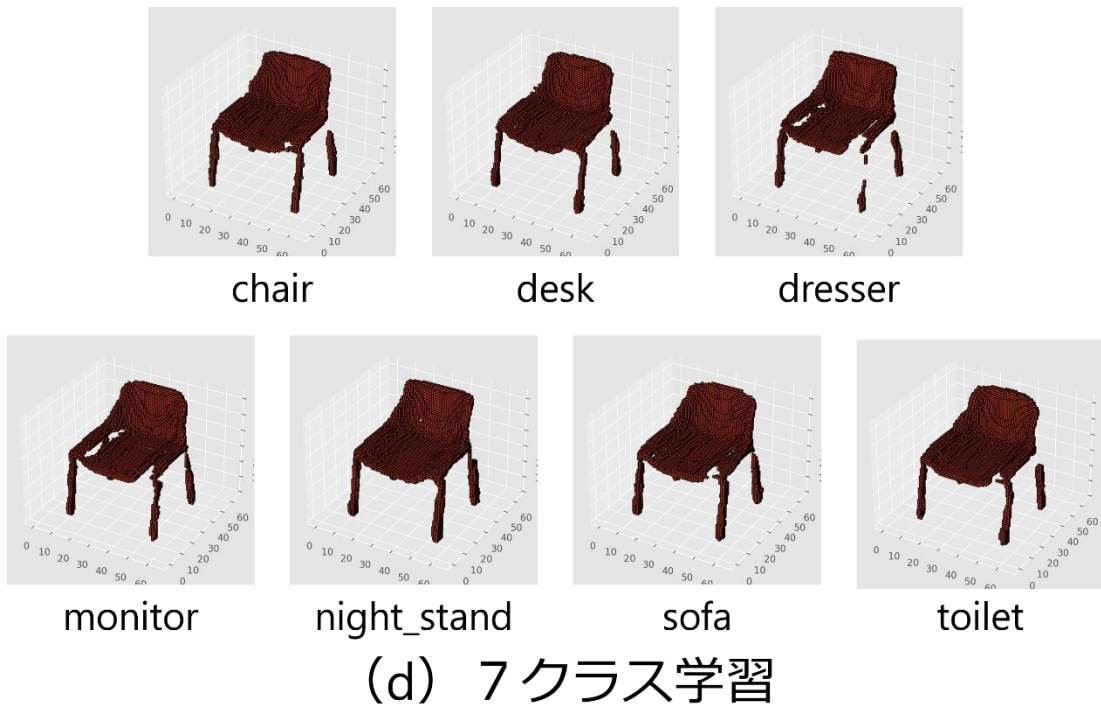
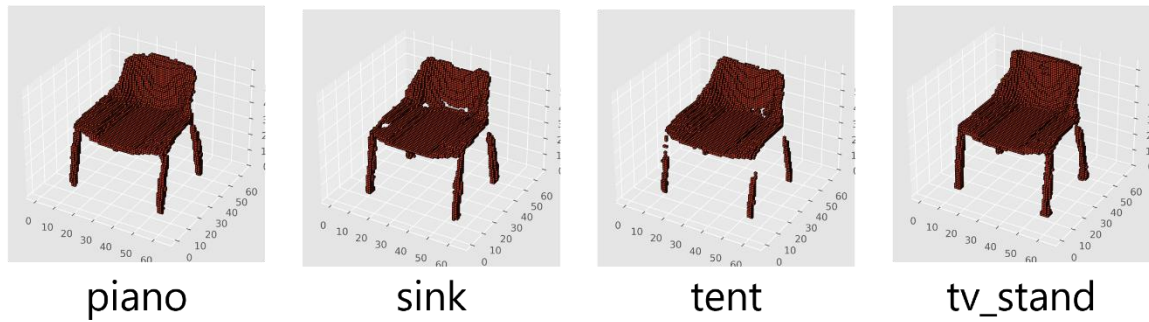
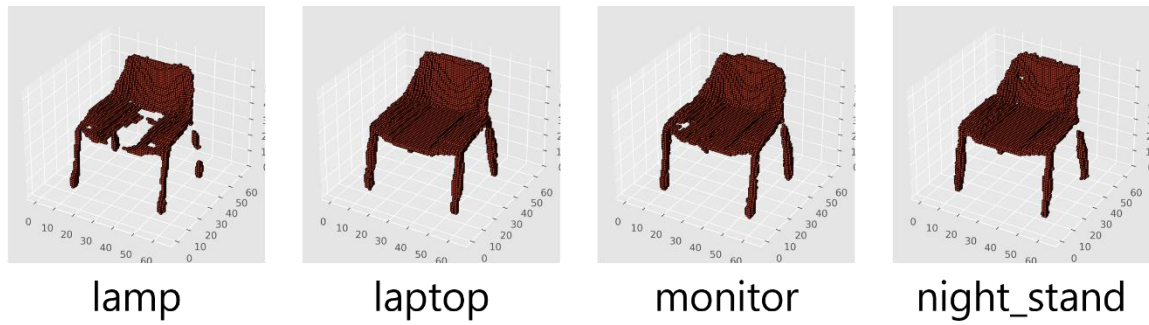
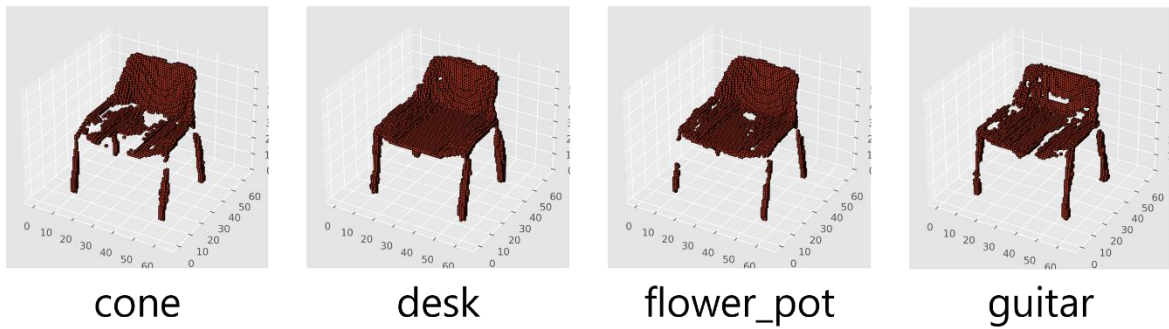
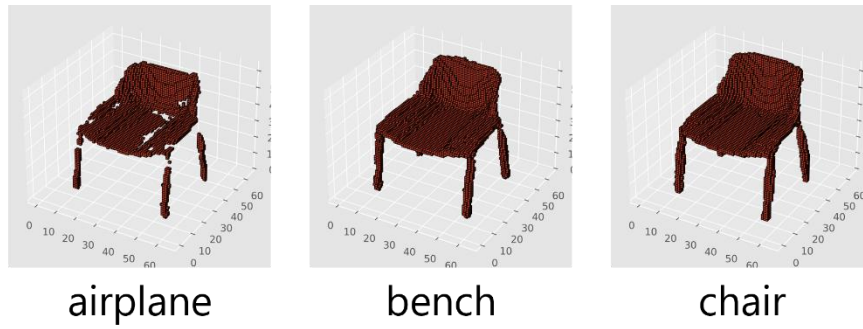


図 17 chair クラスの 3 次元モデルの超解像結果の例 (クラス識別情報ごとの結果) (続き)





(f) 15 クラス学習

図 17 chair クラスの 3 次元モデルの超解像結果の例 (クラス識別情報ごとの結果) (続き)

表 11 図 17 の 3 次元モデルの超解像結果に対する定量評価

5 クラス 学習	クラス識別情報	連	穴	空	値
	bed	11	25	0	4585
	chair	<b>2</b>	4	0	<b>3407</b>
	monitor	6	3	0	4174
	sofa	<b>2</b>	<b>2</b>	0	4450
	table	3	4	0	3799
6 クラス 学習	クラス識別情報	連	穴	空	値
	bathtub	8	38	0	4340
	bed	15	35	0	4127
	chair	<b>2</b>	12	0	<b>3123</b>
	monitor	7	<b>11</b>	0	4460
	sofa	4	26	0	4705
	toilet	3	19	0	3727
7 クラス 学習	クラス識別情報	連	穴	空	値
	chair	3	<b>7</b>	0	<b>3095</b>
	desk	4	12	0	3818
	dresser	12	10	0	3926
	monitor	<b>2</b>	9	0	4340
	night_stand	3	11	0	4012
	sofa	9	11	0	3920
	toilet	3	25	0	3193
8 クラス 学習	クラス識別情報	連	穴	空	値
	bathtub	15	23	0	4037
	chair	2	<b>3</b>	0	3462
	desk	5	16	0	3723
	dresser	6	5	0	3799
	night_stand	<b>1</b>	7	0	4161
	sofa	18	14	0	3999
	table	2	23	0	<b>3226</b>
	toilet	9	12	0	3501

表 11 図 17 の 3 次元モデルの超解像結果に対する定量評価 (続き)

15 クラス 学習	クラス識別情報	連	穴	空	値
	airplane	9	18	0	3566
	bench	2	11	0	4208
	chair	<b>1</b>	<b>0</b>	0	<b>3253</b>
	cone	14	17	0	3464
	desk	3	7	0	3619
	flower_pot	6	30	0	3496
	guitar	10	24	0	3847
	lamp	4	15	0	3526
	laptop	<b>1</b>	<b>0</b>	0	4669
	monitor	3	3	0	4154
	night_stand	2	13	0	3543
	piano	2	11	0	4452
	sink	2	24	0	3375
	tent	11	24	0	3865
tv_stand	2	2	0	3846	

表 12 クラス識別情報ごとの超解像結果の定量評価

(chair クラスのテストデータ全体に対する平均値)

5 クラス 学習	クラス識別情報	連	穴	空	値
	bed	10.5	11.5	0.2	9773.4
	chair	<b>3.1</b>	9.2	<b>0.1</b>	<b>8536.8</b>
	monitor	3.4	<b>8.9</b>	0.2	9004.8
	sofa	5.3	9.9	0.3	9026.7
	table	3.4	9.7	0.2	9701.1
6 クラス 学習	クラス識別情報	連	穴	空	値
	bathtub	10	12.8	0.4	10242.1
	bed	6.8	13.3	0.2	9744.7
	chair	4.4	10.6	<b>0.1</b>	<b>8629.2</b>
	monitor	<b>3.9</b>	10.5	0.8	9254.1
	sofa	5	10.8	0.7	9221
	toilet	5	<b>10.2</b>	0.5	9068.1
7 クラス 学習	クラス識別情報	連	穴	空	値
	chair	4.3	10	<b>0.1</b>	<b>8536.2</b>
	desk	<b>4.2</b>	12	0.7	9394.6
	dresser	6	9.9	0.3	9563.1
	monitor	5.2	10.7	0.2	9311.5
	night_stand	2.5	<b>8.4</b>	<b>0.1</b>	9088.6
	sofa	5.3	10.7	0.3	9061.2
	toilet	5.8	10	0.2	9269.9
8 クラス 学習	クラス識別情報	連	穴	空	値
	bathtub	12.1	13.1	0.3	10746.8
	chair	5	9.4	0.2	<b>8507.2</b>
	desk	4.7	11.4	0.3	9247.6
	dresser	5.3	11	0.3	9229.8
	night_stand	<b>2.9</b>	<b>9</b>	0.3	8932
	sofa	8.3	11.5	0.3	9110.5
	table	4.3	11.3	<b>0.1</b>	9565.7
	toilet	7.1	9.5	0.3	9314.2

表 12 クラス識別情報ごとの超解像結果の定量評価  
 (chair クラスのテストデータ全体に対する平均値) (続き)

15 クラス 学習	クラス識別情報	連	穴	空	値
	airplane	9	15.6	0.9	10764.7
	bench	2.6	9.9	0.3	9414.9
	chair	3.5	9.6	<b>0.1</b>	<b>8640</b>
	cone	6.5	9.8	<b>0.1</b>	9341.5
	desk	3.2	11.7	0.3	9310.5
	flower_pot	7.3	10.9	0.2	9674.4
	guitar	7.1	11.3	0.8	10595.36
	lamp	4.5	11.1	<b>0.1</b>	9186.1
	laptop	<b>2.1</b>	<b>7</b>	0.2	8992.4
	monitor	3.3	9.1	0.5	9191.9
	night_stand	2.8	10.7	0.4	9183.1
	piano	2.8	10	0.9	9073.4
	sink	5.4	10.4	0.2	9432.5
	tent	9	10.8	0.3	9839.4
tv_stand	3.7	12.8	0.7	9734.4	

## 4.7. データセットの解像度を上げた場合の超解像例

3D-SRCGAN では、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 次元モデルを使用して、超解像を行っている。しかし、 $64 \times 64 \times 64$  の 3 次元ボクセルモデルでは、完璧に超解像が行えたとしても、実物とはあまり似ていない。低解像度 / 高解像度 3 次元モデルの解像度を上げることで、特別な手法を用いることなく超解像 3 次元モデルを実物に近づけることができると考えた。

この実験では、学習データ、テストデータ両方の解像度を上げることで、超解像結果にどのような影響があるかを確認する。

### 4.7.1. 実験内容

本実験では、学習データ、テストデータ両方の解像度を、低解像度 3 次元モデルは  $32 \times 32 \times 32$  に、高解像度 3 次元モデルは  $128 \times 128 \times 128$  に変更する。データセットは変わらず ModelNet10 を使用する。学習データの解像度を上げた影響で、ModelNet10 のすべてのデータを使って学習しようとする、プログラムの実行時にメモリ不足になる。そのため、本実験では ModelNet10 の一部を使用して実験を行った。データの内訳を表 13 に示す。学習は、chair、monitor、night\_stand の 3 クラスと、desk、monitor、night\_stand、table の 4 クラスの 2 通りを行った。また、4 クラスについては比較のために、従来の解像度での学習も行った。

表 13 解像度を上げた超解像実験に用いたデータの内訳

クラス名	chair	desk	monitor	night_stand	table
学習データ数	200	200	250	200	200
テストデータ数	100	86	100	86	100

## 4.7.2. 実験結果

図 18 に 3 クラス学習に対する超解像結果を、図 19 に解像度を上げる前後の 4 クラス学習に対する超解像結果を示す。また、表 14 に図 18、19 に示した超解像結果の定量評価を、表 15 に各クラスのテストデータ全体の超解像結果に対する定量評価の平均を示す。表 14、15 において、従来の解像度である  $64 \times 64 \times 64$  のデータを使った結果は、比較のためにボクセル値の誤差のみ 8 倍にしている。

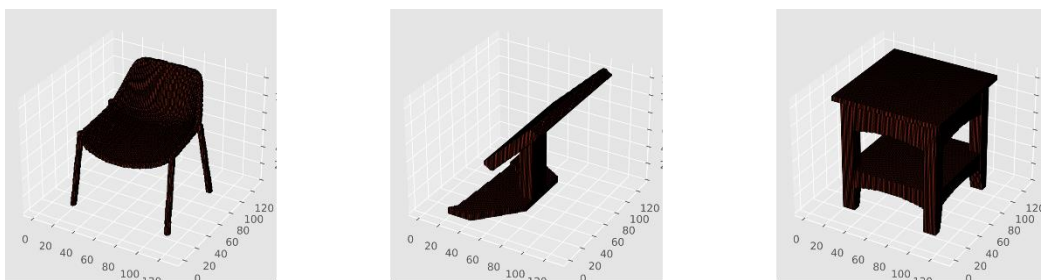
図 18 を見ると、chair 以外のモデルは、全体的に足や支えの部分が細くなっているが、高解像 3 次元モデルにかなり近く、超解像精度はかなり高い。一方、chair は背もたれに穴が開いていたり、足が途切れたりしている。これは、chair 以外のモデルがほとんど直線や平面で構成されているのに対し、chair には座面や背もたれなど曲面が多いことが原因だと考えられる。また、図 19 を見ると、 $128 \times 128 \times 128$  のモデルについては図 18 と同様の傾向が見られた。一方  $64 \times 64 \times 64$  のモデルでは、全体的に歪んでおり、desk や table の足が消えているなど、 $128 \times 128 \times 128$  と比較して、かなり超解像精度が低い。

表 14 の 3 クラスの結果と  $128 \times 128 \times 128$  のデータを使って 4 クラスで学習した場合を見ると、連結成分、穴、空洞の誤差に関しては超解像結果の品質と同じく chair だけが特に悪い。一方、ボクセル値の誤差では、比較的品質が良かった monitor と night\_stand の評価が低い。これは、3 次元モデルを構成するボクセル数によるものだと考えられる。誤差の小さかった chair は 5187 個なのに対し、monitor は 17647 個、night\_stand が 29040 個と、多くのボクセルで構成されている。そのため、3 次元モデルの位置がずれた場合の誤差が大きくなってしまう。ボクセル数と誤差の比率で考えると、chair よりも monitor と night\_stand の評価が高いため、これらを考慮することで、より正確な評価を行える可能性がある。一方、 $64 \times 64 \times 64$  の結果を見ると、全ての項目で図 19 の超解像結果の品質と異なり、desk が高評価になっていた。これは、足などの細かい部分はボクセル数が少ないため、評価に影響が出にくいことが原因と考えられる。 $128 \times 128 \times 128$  と  $64 \times 64 \times 64$  を比較すると、連結成分、穴、空洞の誤差に関しては特に差が見られなかったが、ボクセル値の誤差はすべてのモデルで  $128 \times 128 \times 128$  のほうが良い結果となっていた。また、表 15 から、テストデータ全体で見ても、表 14 と同様に解像度が高い

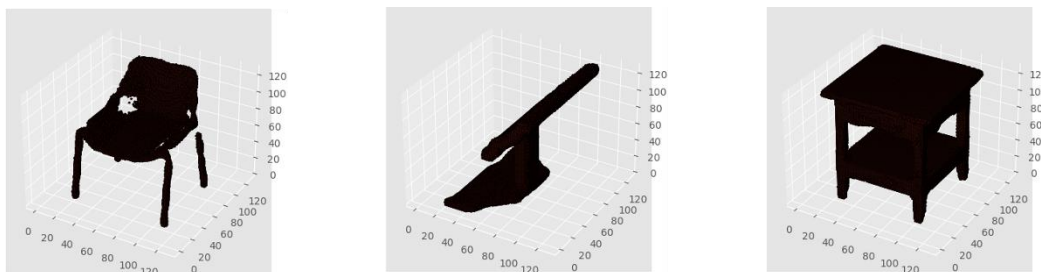
ほうが、超解像精度も高くなることが分かった。これは、同じデータ数では  $128 \times 128 \times 128$  のほうがモデル 1 つから学習できる情報が多いためだと考えられる。

以上のことから、学習データの解像度を上げることでより高品質な 3 次元モデルを高精度に生成することができると分かった。しかし、曲線や曲面が多いモデルに関しては逆に超解像精度が落ちてしまう可能性がある。また、解像度を上げると学習に必要なメモリ量も増大する。



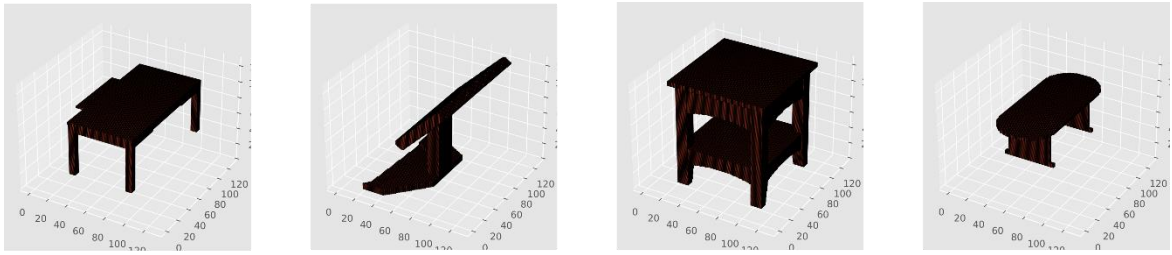


(a) 高解像度3次元モデル

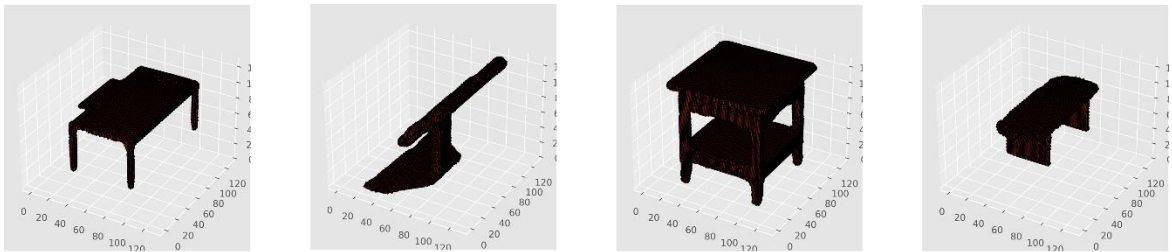


(b) 3クラス学習

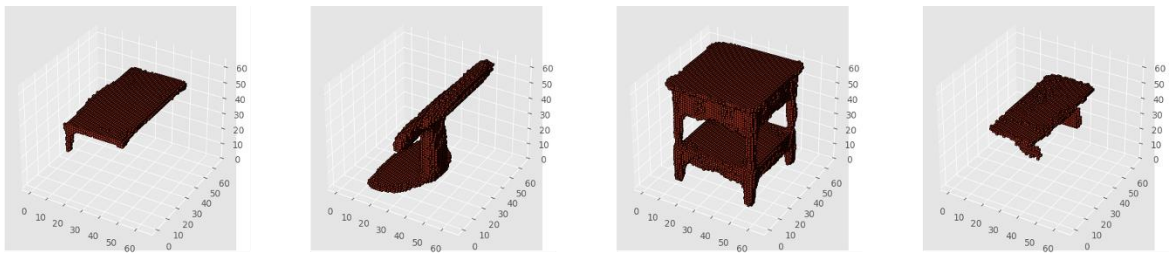
図 18 解像度を上げた3次元モデルの超解像結果の例 (3クラスでの学習結果)



(a) 高解像度3次元モデル



(b) 4クラス学習 (128×128×128)



(c) 4クラス学習 (64×64×64)

図 19 解像度を上げた 3 次元モデルの超解像結果の例 (異なる解像度での学習結果)

表 14 図 18、図 19 の 3 次元モデルの超解像結果の定量評価

3 クラス学習	連	穴	空	値
chair	110	146	0	17430
monitor	5	1	0	30659
night_stand	2	0	0	58062
4 クラス学習 (128×128×128)	連	穴	空	値
desk	10	6	0	20397
monitor	15	7	0	31977
night_stand	5	0	0	58659
table	8	9	0	17593
4 クラス学習 (64×64×64)	連	穴	空	値
desk	1	0	0	50984
monitor	17	9	1	71704
night_stand	0	8	0	76840
table	15	14	0	37520

表 15 異なるクラス数、解像度での超解像結果の定量評価

(各クラス数、解像度のテストデータ全体に対する平均値)

	連	穴	空	値
3 クラス学習	133.9	112.9	25.3	61012.6
4 クラス学習 (128×128×128)	95.4	99	21.7	55872.2
4 クラス学習 (64×64×64)	143.2	157.6	49.6	90023.2

## 5. おわりに

本研究では、3D-SRGAN による 3 次元ボクセルモデル超解像において、学習に使えるデータ数が少ない、学習に使っていないクラスの超解像精度が低いという問題に対し、クラス識別情報を導入することで超解像精度を改善した 3D-SRCGAN を提案した。まず、従来の 3D-SRGAN と、クラス識別情報を導入した 3D-SRCGAN で、複数クラスのデータを使って学習を行い、複数クラスが混ざることによる超解像精度の低下がクラス識別情報によって改善するかを評価した。その結果、クラス識別情報の導入によって超解像精度を低下させずに複数クラスを同時に学習できることが確認できた。次に、1 クラスのデータで学習を行った 3D-SRGAN と 3D-SRCGAN で比較を行い、クラス識別情報の導入によって超解像精度が向上するかを評価した。結果として超解像精度が向上し、3D-SRCGAN が複数クラス、1 クラスどちらにおいても従来の 3D-SRGAN よりも優れていることが確認できた。その後、3D-SRCGAN において、データ数の違いが超解像精度にどのような影響を与えるかを調査した。その結果、3D-SRCGAN ではデータ数が多いほうが超解像精度は高くなるが、データ数 3000 個～4000 個付近ではあまり差が見られなかった。次にデータ数を約 3000 個に固定し、クラス数の違いによる超解像精度への影響を調査した。この実験では、クラス数の違いによる超解像精度の差は見られなかった。これは、クラス数を増やす利点の 1 つである学習データ数の増加がこの実験ではできなかったことによるものと思われる。さらに、この実験の生成モデルを使って、本来のクラスとは別のクラス識別情報を与え、超解像を行った。その結果、正しいクラスでなくとも、特徴が似ているクラスであればそれほど超解像精度は落ちないことが分かった。このことから、3D-SRCGAN は未知のクラスの超解像であっても、似ているクラスを学習に使っていれば、高品質な超解像 3 次元モデルを生成できる可能性があることが分かった。最後に、超解像 3 次元モデルをより実物に近づけるために、データセットの解像度を上げて超解像を行った。その結果、従来の解像度に比べて高品質かつ高精度な超解像 3 次元モデルを生成できることが分かった。

今後の課題として、クラス識別情報の付加方法の見直しが挙げられる。本研究で用いた手法は Generator と Discriminator の入力層への入力に one-hot 表現のクラス識別情報を付加するという非常に単純なものである。この手法は出力層付近で条件情報が消失しやすいという特徴がある。3D-SRCGAN においては、条件情報が消失することはなかつ

たが、他の手法を使うことでより多くの条件情報を保持しておける可能性もある。また、one-hot 表現のクラス識別情報では、情報量に対してデータ量が非常に大きい。この条件情報をより小さなデータ量で表すことができれば、学習にかかる時間を短縮し、より多くのデータを使って学習ができると考えられる。

## 謝辞

本研究を行うにあたり、多くの方々にご協力いただきました。熱心にご指導いただいた椋木雅之教授に深く感謝いたします。指導教員である椋木雅之教授には研究や論文執筆に関し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。

山森教授、伊達准教授には、副査を務めていただきました。お忙しい中、貴重なお時間を割いてくださりましてありがとうございます。

椋木研究室の皆様には、研究を進めるにあたって、様々な助言をいただきました。皆様のおかげで充実した研究生活を過ごすことができました。大変感謝しております。

## 参考文献

- [1] 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)
- [2] 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)
- [3] M Mirza, S Osindero, “Conditional generative adversarial nets” arXiv:1411.1784 (2014)
- [4] 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)
- [5] H Chang, D-Y Yeung, Y Xiong, “Super-resolution through neighbor embedding”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 275-282 (2004)
- [6] D Khaledyan, A Amirany, K Jafari, MH Moaiyeri, AZ Khuzani, N Mashhadi, “Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution”, IEEE Global Humanitarian Technology Conference (GHTC), pp. 1-5 (2020)
- [7] WT Freeman, TR Jones, EC Pasztor, "Example-based super-resolution", IEEE Computer Graphics and Applications (CGA), vol. 22, no. 2, pp. 56-65 (2002)
- [8] 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)
- [9] 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)

- [10] J Wang, Y Chen, Y Wu, J Shi, J Gee, “Enhanced generative adversarial network for 3D brain MRI super-resolution”, IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 3627-3636 (2020)
- [11] J Hatvani, A Basarab, JY Tournier, M Gyöngy, D Kouamé, “A tensor factorization method for 3-D super resolution with application to dental CT”, IEEE Transactions on Medical Imaging (TMI), vol. 38, no. 6, pp. 1524-1531 (2018)
- [12] S Reed, Z Akata, X Yan, Logeswaran L, Schiele B, Lee H, “Generative adversarial text to image synthesis”, International Conference on Machine Learning (ICML), pp. 1060-1069 (2016)
- [13] P Isola, JY Zhu, T Zhou, AA Efros, “Image-to-image translation with conditional adversarial networks” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1125-1134 (2017)
- [14] A Sage, E Agustsson, R Timofte, LV Gool, “Logo synthesis and manipulation with clustered generative adversarial networks,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.5879-5888 (2018)
- [15] A Odena, C Olah, J Shlens, “Conditional image synthesis with auxiliary classifier GANs,” International Conference on Machine Learning (ICML), vol. 70, pp. 2642–2651 (2017)
- [16] N Pandey, A Savakis, “Poly-gan: Multi-conditioned gan for fashion synthesis,” Neurocomputing, vol. 414, no. 13, pp. 356–364 (2020)
- [17] A Radford, L Metz, S Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", International Conference on Learning Representations (ICLR), (2016)
- [18] ModelNet10, ModelNet40, <http://modelnet.cs.princeton.edu/>