

令和2年度卒業論文

3次元モデルの超解像における
敵対学習の有効性

宮崎大学工学部 情報システム工学科

金田 健太郎

指導教員 棕木 雅之

目次

1. はじめに.....	1
2. 3D-SRGAN による 3 次元モデルの超解像	2
2.1. 超解像の定義.....	2
2.2. 2 次元画像の超解像.....	2
2.3. ボクセル表現の 3 次元モデル	3
2.4. 3D-SRGAN による 3 次元モデルの超解像.....	3
2.5. 3D-SRGAN の学習	4
3. 敵対学習の有効性調査.....	6
3.1. 有効性調査の必要性.....	6
3.2. 検証の方法	7
4. 実験.....	8
4.1. SRGAN での 2 値画像超解像.....	8
4.1.1. 実験手法	8
4.1.2. 実験データ.....	8
4.1.3. 比較方法	9
4.1.4. 実験結果・考察	10
4.2. 3D-SRGAN での敵対学習の検証	14
4.2.1. 実験手法	14
4.2.2. 実験データ.....	14
4.2.3. 比較方法	15
4.2.4. 実験結果・考察	16
5. おわりに.....	20
謝辞.....	21
参考文献	22

1. はじめに

3次元モデルの表現方法に、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、多数のボクセルを積み重ねることにより精細な3次元モデルを作成することができるが、当然粗い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次元モデルを生成することができればこの手間が軽減できる。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-Super Resolution Generative Adversarial Networks (3D-SRGAN)[1]がある。3D-SRGANは2次元画像の超解像に有効なSuper Resolution Generative Adversarial Networks(SRGAN)[2]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。この手法は拡張元のSRGANと同じくGenerative Adversarial Network(GAN)[3]を用いて敵対学習することにより、優れた超解像結果を得ようとしている。しかし、ボクセル表現の3次元モデルを構成するボクセルはカラー画像を構成するピクセルよりも保持する情報が少なく組み合わせが単純なため、敵対学習を行わずとも視覚的品質の高い超解像結果が得られる可能性がある。GANの学習では、勾配消失問題等により学習が不安定になるという問題がある。3D-SRGANと同等の超解像結果が得られるのであれば、GANを用いない方がよいといえる。

そこで、本研究では3D-SRGANにおいてGANを用いた敵対学習が視覚的品質の向上において有効に働いているのか調査を行い、3次元モデルの超解像に敵対学習が必要かどうかを考察する。

2. 3D-SRGAN による 3 次元モデルの超解像

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、2次元画像とボクセル表現の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、画像は超解像画像、3次元モデルは超解像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[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手法を導入することが必要である。

2.2. 2次元画像の超解像

カラーの2次元画像は通常 $8\text{bit} \times 3$ の情報量を持っているピクセルの組み合わせで構成される。超解像を行うことによりより多くのピクセルを用いた精細な2次元画像が復元・生成される。

2次元画像の超解像手法に、有名なものとして SRGAN という手法がある。特徴として、超解像画像を生成する Generator と、入力された画像が Generator の生成した超解像画像なのか学習データの high-resolution 画像なのかを識別する

Discriminator という 2 つのニューラルネットワークによって構成されていることが挙げられる。Generator は Discriminator を騙すことができるように、Discriminator は入力データを識別することができるように敵対学習を行う。Generator のみで超解像 3 次元モデルについて学習を行っていた従来手法と比較すると、Discriminator を追加して敵対学習を行う SRGAN の方が視覚的品質が優れている。

2.3. ボクセル表現の 3 次元モデル

ボクセル表現の 3 次元モデルは、3 次元空間をさいの目状に分割したボクセルと呼ばれる立方体の集まりで表現される。ボクセルは 0 か 1 かの 1bit の情報を保持し、3 次元空間中でその位置が 3 次元物体の表面又は内部であれば 1 を、外部であれば 0 をとる。超解像を行うことにより、より多くのボクセルを用いた (3 次元空間をより細かく分割した) 精細な 3 次元モデルが生成・補完される。

2.4. 3D-SRGAN による 3 次元モデルの超解像

3D-SRGAN は深層学習を用いた 3 次元モデル超解像の手法であり、2 次元画像の超解像で視覚的品質の向上に有効性が知られている SRGAN を、3 次元モデルが扱えるように拡張したものである。

3D-SRGAN は SRGAN と同じく Generator と Discriminator の 2 つのネットワークで構成されている (図 1)。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。学習データにはペアとなる低解像度と高解像度の 3 次元モデルを使用する。Generator は低解像度 3 次元モデルから超解像 3 次元モデルを生成

し、Discriminator は入力された 3次元モデルが Generator によって生成された 3次元モデルなのか学習データの高解像度 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、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が、Discriminator は生成された 3次元モデルと学習データとを見分けることが目的となっている。3D-SRGAN 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。

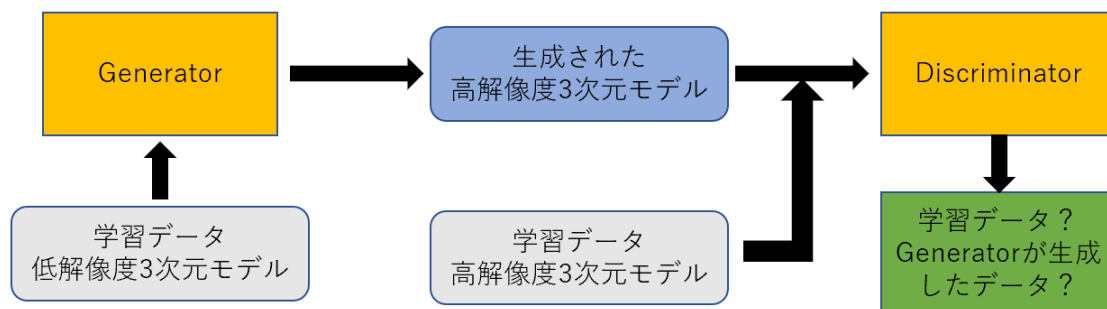


図 1 3D-SRGA の構成

2.5. 3D-SRGAN の学習

超解像生成モデルとなる Generator の学習では、Loss 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い、優れた超解像結果を目指す。Generator の Loss 関数 l_G^{3DSR} を式(1)に示す。

$$l_G^{3DSR} = l_{Con}^{3DSR} + \lambda l_{Gen}^{3DSR} \quad (1)$$

l_{Con}^{3DSR} が Content Loss、 l_{Gen}^{3DSR} が Adversarial Loss、 λ が係数を表す。Content Loss は学習データの高解像度 3次元モデルと Generator が生成した超解像 3次元モデルの平均二乗誤差を計算しており、学習において低解像度データが入力され

た時、ペアとなる高解像度データに近い結果を出力できるように学習する。Adversarial Loss は Discriminator の 識別の結果であり、Generator が Discriminator を騙せるよう、学習データと見分けがつかない出力が得られるように学習をする。この 2 つを係数 λ で調整し足し合わせて使用している。 λ の値は通常 10^{-3} が設定される。

3. 敵対学習の有効性調査

3.1. 有効性調査の必要性

3D-SRGAN では敵対学習を行うことにより、超解像結果の視覚的品質の向上を図っている。しかし、3次元モデルの超解像に敵対学習を適用することに、以下の点から疑問がある。

1つ目は、GAN を用いた敵対学習が3次元モデルの超解像結果の視覚的品質の向上に本当に有効なのかという点である。SRGAN では、生成モデル (Generator)のみで学習を行った超解像結果と敵対学習を行った超解像結果を比較して視覚的品質の向上に対する有効性が検証されている。これは超解像画像の生成・補完部分のバリエーションが多様であり、その中から視覚的品質の良いものを選ぶのに敵対学習が必要であるからだ。しかし3次元モデルを構成するボクセルは保持する情報が1bitと、カラー画像を構成するピクセルの情報8bit×3と比較すると小さく、ボクセル同士の組み合わせが単純である。そのため生成・補完部分の多様性がカラー画像と比較すると少なく、生成モデルのみでの学習でも視覚的品質の良い超解像結果が得られる可能性がある。

2つ目は、GAN の学習は安定しないということが挙げられる。GAN の学習では、勾配消失問題等が原因で収束計算が破綻することがある。GAN を用いたSRGAN、3D-SRGAN も同様に収束計算の破綻が発生することがある[5]。収束計算が破綻すると学習が進まず、良い結果が得られない。生成モデルのみならば学習を十分に進めることができるので、よい出力結果が得られる可能性がある。

以上の点から3次元モデルの超解像において、GAN を用いた敵対学習が超解

像結果の視覚的品質の向上に有効でない場合は、学習が不安定な GAN を用いるのは不適切だと考える。そのため GAN を用いた敵対学習が視覚的品質の向上に有効であるのか調査し、3次元モデルの超解像に敵対学習が必要かどうか考察する必要がある。

3.2. 検証の方法

超解像結果を生成する Generator は、式 (1) の Loss 関数 l_G^{3DSR} を小さくするように学習する。Loss 関数 l_G^{3DSR} は、Content Loss と Adversarial Loss が係数 λ で重み付けされて足し合わされたものである。Content Loss は、低解像度データに存在しない部分のデータを生成して超解像データを出力しようとする。一方、Adversarial Loss は、このデータの生成の際に、Discriminator を騙せるようなデータを生成しようとする。つまり、Content Loss のみで低解像度データから超解像データを生成することは可能で、Adversarial Loss は、その生成されるデータに影響を与えるのみといえる。Adversarial Loss が生成されるデータにどの程度影響を与えるかは、係数 λ で決まる。すなわち、この Loss 関数 l_G^{3DSR} を構成する Adversarial Loss の係数 λ を変更することで、Discriminator の結果をどのくらい反映するのかを調整することができる。Adversarial Loss の係数を $\lambda = 0$ にした場合には Generator の学習の更新には Discriminator の結果は考慮されず、GAN を用いていないことになる。よって、Adversarial Loss の係数 λ を変化させ、超解像結果を比較することにより、GAN を用いた敵対学習の視覚的品質に対する有効性を検証する。

4. 実験

4.1. SRGAN での 2 値画像超解像

この実験では、ピクセルの持つ情報をボクセルと同様に 2 値にした 2 値画像を超解像対象とした SRGAN において、GAN を用いた敵対学習の視覚的品質向上の有効性を検証し、3D-SRGAN での敵対学習の有効性の検証についての予備調査を行った。

4.1.1. 実験手法

SRGAN も 3D-SRGAN 同様に、Generator の Loss 関数は Content Loss と Adversarial Loss を係数 λ で重み付けした和となっている。よって節 3.2.で説明した方法を用いることによって敵対学習の有効性検証が行える。

Generator の Loss 関数を構成する Adversarial Loss の係数 λ を 0 、 10^{-3} 、 10^{-2} 、 10^{-1} と変更し、それぞれの係数 λ について学習データを用いて 172epoch 学習を行った。学習が終了した生成モデルを用いて評価データを超解像し、係数 $\lambda = 0$ (敵対学習なし) とその他の係数の超解像結果を比較し、視覚的品質に違いがあるのかを検証した。

4.1.2. 実験データ

学習データにはデータセット RAISE_HR、LR[6]を使用した。RAISE_HR、LR はカラー画像や濃淡画像であり、高解像度と低解像度の画像のペアが与えられている。評価データには Linnaeus 5[7]を使用した。Linnaeus 5 は 256×256 のカラー画像であり、変換を加えないものを高解像度画像として、 64×64 に変換

を行ったものを低解像度画像として扱う。今回は 2 値画像の超解像なので学習データ、評価データともにカラー画像を濃淡画像に変換したうえで、閾値 128 未満のピクセル値を 0、閾値 128 以上のピクセル値を 255 の 2 値に変換した。

4.1.3. 比較方法

比較の方法としては、視覚的品質という主観的な品質を比較するために、以下の評価項目①について定性評価を行った。定性評価では、○を良い、△を普通、×を悪いとする。

① 2 値画像のエッジの細かさ

また、PSNR と SSIM を用いた定量評価も行った。PSNR とは、同じ位置のピクセル値がどのくらい変わったかを示す指標になっており、式(2)で計算される。 MAX_I は画像が取りうる最大ピクセル値を、 MSE は Generator が生成した超解像画像と学習データの高解像度画像のピクセル値の平均二乗誤差を表している。SSIM とは、PSNR よりも人間の視覚的評価を反映できるように考えられた指標になっており、式(4)で計算できる。式(3)では x と y は超解像画像と高解像度画像の一定範囲内(ウィンドウ)の各ピクセル値を要素とするベクトルを、 μ_x 、 μ_y はピクセル値の平均、 $2\sigma_x$ 、 $2\sigma_y$ はピクセル値の標準偏差、 $2\sigma_{xy}$ はピクセル値の共分散、 c_1 、 c_2 は定数を表している。式(4)では X と Y が超解像画像と高解像度画像を、 M がウィンドウの数、 x_i と y_i は超解像画像と高解像度画像のウィンドウの各ピクセル値を要素とするベクトルを表している。PSNR と SSIM とともに値が大きい方が優れている。

$$PSNR = 10 \cdot \log_{10} \frac{MAX_I^2}{MSE} \quad (2)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (3)$$

$$MSSIM(X, Y) = \frac{1}{M} \sum_{i=1}^m SSIM(x_i, y_i) \quad (4)$$

4.1.4. 実験結果・考察

図 2 に 3 種類の 2 値画像に対する超解像結果を示す。上から順に高解像度画像、Adversarial Loss の係数 λ を 0 、 10^{-3} 、 10^{-2} 、 10^{-1} で学習を行った超解像画像である。各列は画像の種類に対応している。左の列から順に画像 1、画像 2、画像 3 と呼ぶ。また、表 1 には図 2 に示した超解像結果の定性評価結果を、表 2 には PSNR と SSIM を用いた定量評価結果を示す。

表 1 を見ると $\lambda = 0$ と $\lambda = 10^{-3}$ とでは、定性評価に違いがみられない。これは $\lambda = 10^{-3}$ が小さすぎて Generator の学習に Discriminator の結果が十分に反映されていないためだと考えられる。 $\lambda = 0$ と $\lambda = 10^{-2}$ とでは、 $\lambda = 10^{-2}$ の方がよりエッジが細かくなっており精細な印象を受け、定性評価は高い。 $\lambda = 0$ と $\lambda = 10^{-1}$ とでは、 $\lambda = 10^{-2}$ よりもよりエッジが細かくなっており定性評価はより高いものになっている。この結果より、本実験の超解像においては Adversarial Loss の係数 $\lambda = 10^{-2}$ 、 $\lambda = 10^{-1}$ が適切な値であった。Adversarial Loss を $\lambda = 10^{-2}$ 、 $\lambda = 10^{-1}$ の重みで加えた方が、敵対学習なし($\lambda = 0$)より視覚的品質の向上があったことから、2 値画像の超解像において GAN を用いた敵対学習は視覚的品質の向上に有効であると考えられる。

表 2 の PSNR を用いた定量評価では、定性評価の結果と違い、 $\lambda = 0$ (敵対学

習なし)が最良な結果として評価されている。しかし、SSIM を用いた定量評価は、定性評価と同じく敵対学習ありが最良な結果として評価されている。SRGAN では、定量評価の結果が必ずしも視覚的品質の良し悪しに対応しないことが指摘されている。よって今回用いた定量評価も必ずしも視覚的品質と一致するものではないと考える。

以上のことより SRGAN を用いた 2 値画像の超解像においては、GAN を用いた敵対学習は視覚的品質の向上に有効であることが分かった。

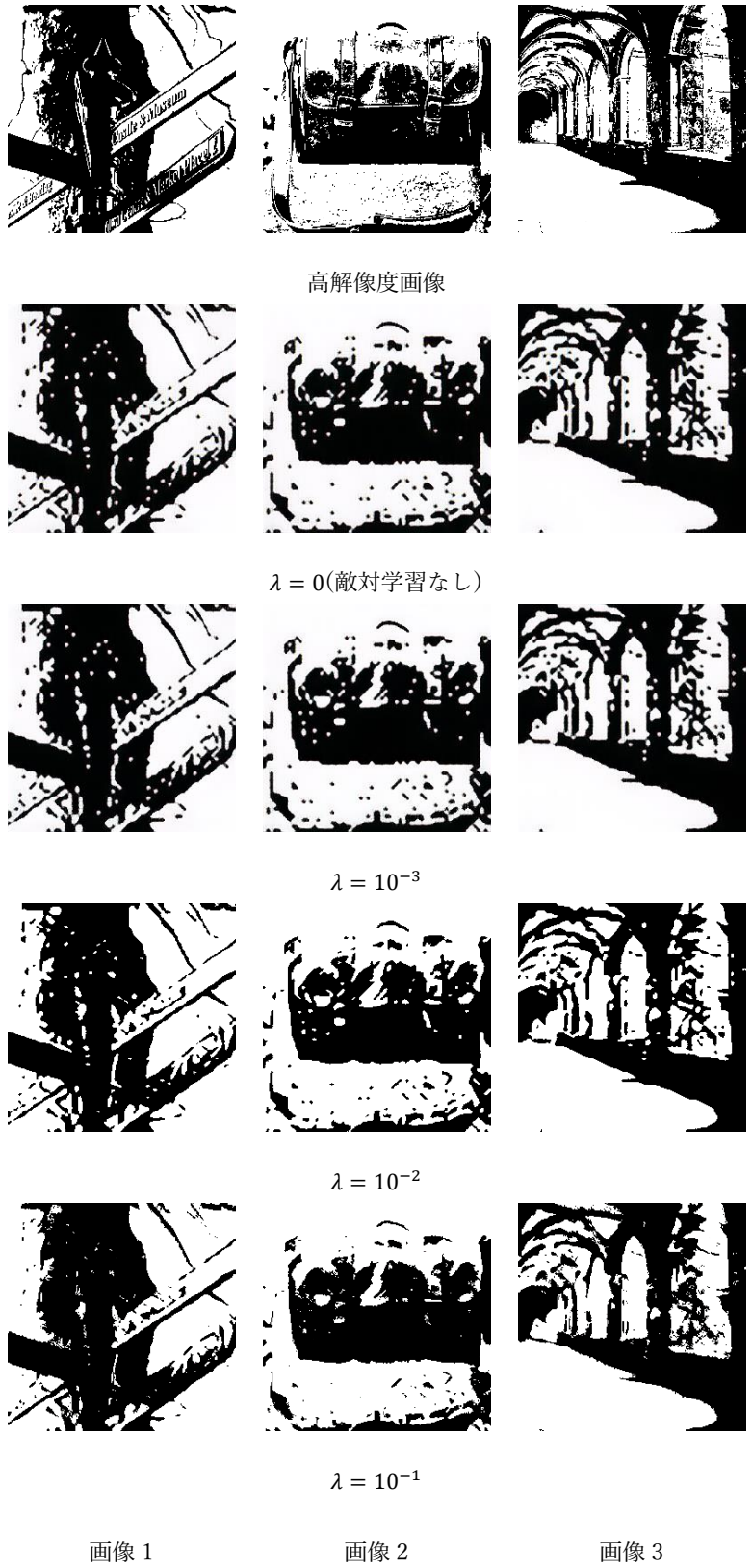


図 2 2 値画像の超解像結果

表 1 2 値画像の超解像結果に対する定性評価

	画像 1	画像 2	画像 3
$\lambda = 0$	×	×	×
$\lambda = 10^{-3}$	×	×	×
$\lambda = 10^{-2}$	△	△	△
$\lambda = 10^{-1}$	○	○	○

表 2 2 値画像の超解像結果に対する定量評価(PSNR/SSIM)

	画像 1	画像 2	画像 3
$\lambda = 0$	11.010841/0.506929	10.574608/0.470684	10.455501/0.519866
$\lambda = 10^{-3}$	10.926120/0.554123	10.496278/0.494341	10.437329/0.538595
$\lambda = 10^{-2}$	10.217995/ 0.613702	9.863261/ 0.546570	9.620683/ 0.589629
$\lambda = 10^{-1}$	10.071733/0.590231	10.021970/0.538918	9.531631/0.571144

4.2. 3D-SRGAN での敵対学習の検証

この実験では、3D-SRGAN において GAN を用いた敵対学習の視覚的品質向上の有効性を検証し、3次元モデルの超解像における GAN を用いた敵対学習の必要性について考察する。

4.2.1. 実験手法

Generator の Loss 関数を構成する Adversarial Loss の係数 λ を0、 10^{-3} 、 10^{-2} 、 10^{-1} と変更し、それぞれの係数 λ について学習データを用いて 50epoch 学習を行った。学習が終了した生成モデルを用いて評価データを超解像し、係数 $\lambda = 0$ (敵対学習なし) とその他の係数の超解像結果を比較し、視覚的品質に違いがあるのかを検証した。

4.2.2. 実験データ

学習、評価に用いる 3次元モデルとして、データセット ModelNet10[8]のイスのクラスを使用した。イスのクラスでは学習用データが 889 個、テスト(評価)用データが 100 個用意されている。学習データ、評価用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た。

4.2.3. 比較方法

比較の方法としては、視覚的品質という主観的な品質を比較するために、以下の評価項目①から④について定性評価を行った。定性評価では、○を良い、△を普通、×を悪いとする。

- ① 3次元モデルの輪郭の滑らかさ
- ② 3次元モデルの形状の滑らかさ
- ③ 3次元モデルの部分の連結性
- ④ 3次元モデルの形状の連結性

また、高解像度モデルと超解像 3次元モデルの同じ位置にあるボクセル同士で、保持する値（ボクセル値）が異なるものの数を用いた定量評価も行った。この定量評価では値が小さい方が良い性能とする。

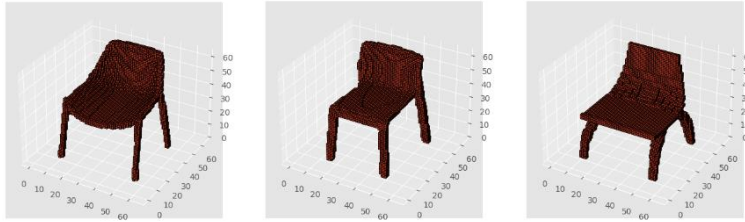
4.2.4. 実験結果・考察

図 3 に 3 種類のイスに対する超解像結果を示す。上から順に高解像度 3 次元モデル、Adversarial Loss の係数 λ を 0 、 10^{-3} 、 10^{-2} 、 10^{-1} で学習を行った超解像 3 次元モデルである。各列はイスの種類に対応している。左の列から順にモデル 1、モデル 2、モデル 3 と呼ぶ。また、表 3 には図 3 に示した超解像結果の定性評価結果を、表 4 には定量評価結果を示す。

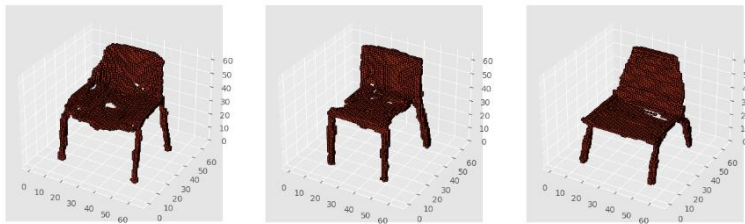
表 3 を見ると $\lambda = 0$ と $\lambda = 10^{-3}$ の超解像 3 次元モデルとでは、4 個の項目に対する定性評価ではほぼ違いがみられない。 $\lambda = 0$ と $\lambda = 10^{-2}$ とでは、 $\lambda = 10^{-2}$ の方が高解像度 3 次元モデルのように足の形がよりしっかりとした形になっており輪郭・形状の滑らかさについての定性評価は高い。加えて、 $\lambda = 0$ の超解像 3 次元モデルに発生している穴や足の欠損などが $\lambda = 10^{-2}$ の超解像 3 次元モデルでは少なくなっており形状・部分の連結性についての定性評価も高い。 $\lambda = 0$ と $\lambda = 10^{-1}$ とでは、 $\lambda = 10^{-3}$ と同様に定性評価ではほぼ違いがみられない。この結果より、本実験の超解像においては Adversarial Loss の係数 $\lambda = 10^{-2}$ が適切な値であった。Adversarial Loss を $\lambda = 10^{-2}$ の重みで加えた方が、敵対学習なし($\lambda = 0$)より視覚的品質の向上があったことから、3 次元モデルの超解像において GAN を用いた敵対学習は視覚的品質の向上に有効であると考えられる。

表 4 の定量評価でも最良の結果は、敵対学習ありのものになっている。ただし、モデル 2 に対しては定性評価と同様 $\lambda = 10^{-2}$ が最良であったが、モデル 1、3 では定性評価と異なる結果となった。拡張元である SRGAN では、定量評価の結果が必ずしも視覚的品質の良し悪しに対応しないことが指摘されている。よって今回用いた定量評価も必ずしも視覚的品質と一致するものではないと考える。

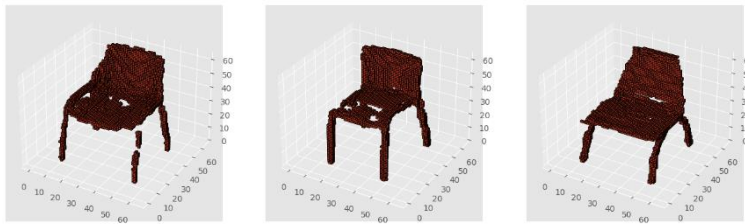
以上のことより 3D-SRGAN を用いた 3 次元モデルの超解像では GAN を用いた敵対学習は視覚的品質の向上に有効であることが分かった。そのため、3 次元モデルの超解像において視覚的品質に重点を置く場合は GAN を用いることは適切だと考える。



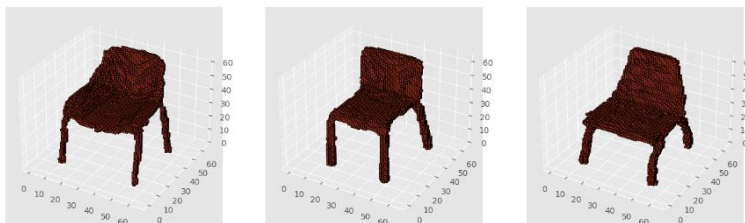
高解像度 3次元モデル



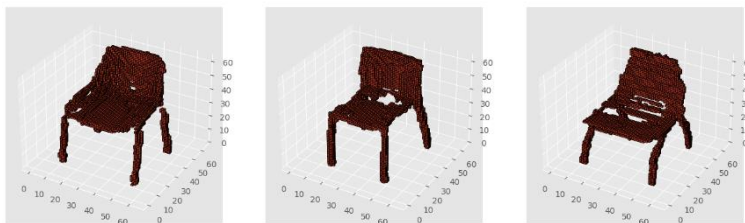
$\lambda = 0$ (敵対学習なし)



$\lambda = 10^{-3}$



$\lambda = 10^{-2}$



$\lambda = 10^{-1}$

モデル 1

モデル 2

モデル 3

図 3 3次元モデルの超解像結果

表 3 3次元モデルの超解像結果に対する定性評価

評価項目	モデル 1				モデル 2				モデル 3			
	①	②	③	④	①	②	③	④	①	②	③	④
$\lambda = 0$	△	△	○	△	△	△	○	△	△	△	○	△
$\lambda = 10^{-3}$	×	×	×	△	△	△	○	△	△	△	○	△
$\lambda = 10^{-2}$	○	○	○	○	○	○	○	○	○	○	○	○
$\lambda = 10^{-1}$	△	△	△	△	△	△	○	×	△	△	○	×

表 4 3次元モデルの超解像結果に対する定量評価

	モデル 1	モデル 2	モデル 3
$\lambda = 0$	3081	3542	6514
$\lambda = 10^{-3}$	2721	4194	5892
$\lambda = 10^{-2}$	3060	3453	5917
$\lambda = 10^{-1}$	2855	3501	6009

5. おわりに

本研究では、3D-SRGAN による 3 次元モデルの超解像において、GAN を用いた敵対学習が視覚的品質の向上に有効か調査を行った。敵対学習を行ったものを行っていないもので超解像を行い、生成された超解像 3 次元モデルに視覚的品質の差が見られるかを評価した。結果として、適切な重み付けで敵対学習を取り入れた方が視覚的品質が向上することが確認された。この結果から、GAN を用いた敵対学習は 3 次元モデルの超解像の視覚的品質向上に有効であることが分かった。

今後の課題としては、GAN を用いた敵対学習の有効性が確認されたため、より優れた超解像ができるように GAN のアーキテクチャの変更やパラメータの最適化を行うことが挙げられる。

謝辞

本研究を進めるにあたり、熱心に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ついての相談や論文についての添削等さまざまなご指導を頂きました。本当にありがとうございました。

また、研究室での活動についての相談や助言をくださった研究室内の皆様にも感謝をいたします。

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] Christian Ledig, et al., "Photo-Realistic Single Image SuperResolution Using a Generative Adversarial Network", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105-114 (2017)
- [3] I. Goodfellow, et al., “Generative adversarial nets”, Advances in Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
- [4] 池谷彰彦, 広明敏彦, “超解像ソリューション”, NEC 技報, vol. 60, no. 2, pp. 24-26 (2007)
- [5] 森芳雄, “深層学習を用いた 3 次元モデル超解像の学習安定化の検討”, 平成 31 年度 宮崎大学大学院 工学研究科修士論文(2020)
- [6] RAISE - The Raw Images Dataset, <http://loki.disi.unitn.it/RAISE/>
- [7]Chaladze, G. Kalatozishvili L, “Linnaeus 5 Dataset for Machine Learning”(2017), <http://chaladze.com/l5/>
- [8] ModelNet10, <http://modelnet.cs.princeton.edu/>