

修士論文概要書

Summary of Master's Thesis

Date of submission: (01/29/2020)

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	梅田 聖也	指 導 教 員 Advisor	渡辺 裕 印
研究指導名 Research guidance	オーディオビジュ アル情報処理研究	学籍番号 Student ID number	5118F017-4		
研究題目 Title	超解像を用いたビット深度拡張に関する研究 Research of the Bit Depth Extension by Super-Resolution				

1. まえがき

2020年現在、総務省のロードマップ[1]に従い、次世代の超高精細映像への移行が始まっている。この次世代の映像では、解像度のみならず、色の多階調表現やHDRによる広範囲の輝度の表現なども拡張される。これに伴い、映像のデータ量は増大し、従来の方式での映像を扱うための高圧縮な技術が求められている。また、従来コンテンツを次世代映像に変換する方式の需要も増加しているが、市販されている製品は解像度のみの変換にとどまっている。そこで、本研究では次世代映像における解像度とビット深度に着目し、この二つの点について超解像を用いて同時に拡張する方式を提案する。

2. 従来手法とその問題点

市販の次世代映像への変換を行う製品ではCNNを用いた超解像による解像度のみでの拡張が用いられている。解像度以外を拡張する技術としてビット深度拡張がある。この手法では、Zero-paddingなどの簡単な手法や、CNNを用いたより高い精度で拡張できる手法[2]がある。前者の手法は少ない計算量でビット深度の補間が可能だが、精度が低い。一方で後者では、高い精度での補間が可能だが計算コストが非常に高い。そのため、精度の高い次世代映像への変換を試みると、計算コストの高い、超解像とCNNによるビット深度拡張を直列的に行う必要があり、莫大な計算コストとなる問題がある。

3. 提案手法

本研究では、より計算コストを抑えるために超解像による解像度とビット深度の同時拡張を提案する。提案手法の概要図を図1に示す。この提案手法では、超解像を学習する際に、低ビット-低解像度の画像と高ビット-高解像度の画像のペアを用いることで、解像度とビット深度の同時拡張を行う。

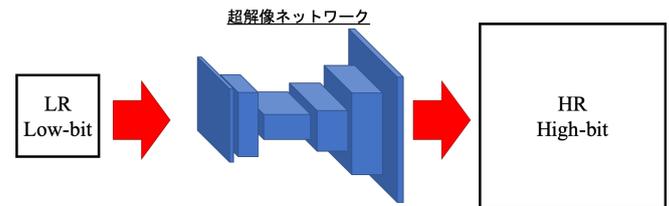


図1 提案手法

4. 評価実験

評価実験では、低解像-低ビットと高解像度-高ビットのペアからなるデータセットで学習した超解像に対して、低ビットの低解像度画像を入力し、出力された高ビットの高解像度画像から各指標を算出、既存手法との比較を行うことで提案手法の有効性を確認する。また、多くの超解像のネットワークはビット深度が8-bitを超える画像の入力に対応・最適化されていない。そのため、本実験においては、より正當に評価するために、低ビットは6-bit、高ビットは8-bitとする。また、より実際に運用される環境に近づけるため低ビット画像は6-bit画像を2-bitシフトして得られる8-bit画像とする。評価実験の概要を図2に示す。本実験では、超解像のネットワークとして、EDSRとSRGANを使用する。また、提案手法より得られる出力画像と以下の二つの画像において各指標を算出して比較する。

- ・ビット深度学習をしていない超解像を用いて得られる画像

- ・提案手法の出力を6-bitに変換し、Zero-paddingにより8-bitにした画像(以降、提案手法+Zero-padding)

学習画像及びテスト画像にはDIV2Kデータセット[3][4]を使用した。また、評価指標としてPSNR, SSIMの他に単体で画像の品質を評価することが可能なBRISQUE[5]を使用し、テスト画像100枚の平均値を算出する。この指標では、値が小さくなるほど品質が高いことを示している。EDSRでの実験結果を図3および表1に示す。また、SRGANでの結果を図4および表2に示す。

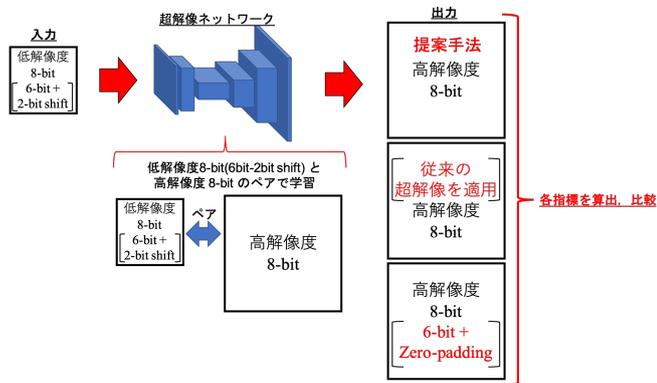


図 2 評価実験



図 3 EDSR 実験結果[3][4]

表 1 EDSR 実験結果

	PSNR dB	SSIM	BRISQUE
提案手法 +Zero-padding	29.59	0.9047	45.80
従来の超解像	29.61	0.9081	49.28
提案手法	29.76	0.9101	45.84

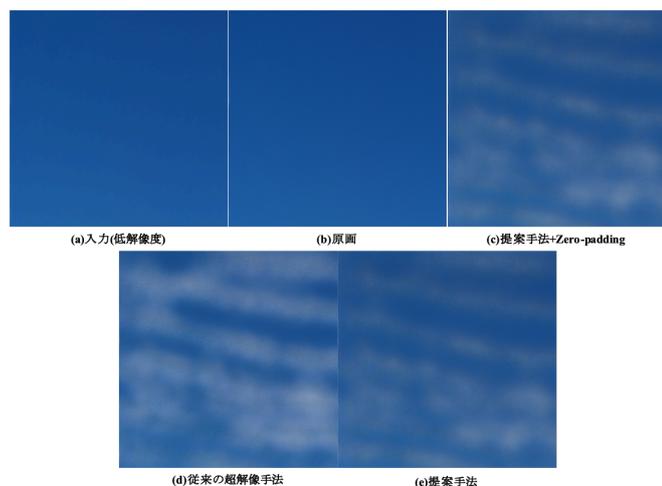


図 4 SRGAN 実験結果[3][4]

表 2 SRGAN 実験結果

	PSNR dB	SSIM	BRISQUE
提案手法 +Zero-padding	24.20	0.8335	13.51
従来の超解像	24.45	0.8480	11.39
提案手法	24.24	0.8397	12.07

図 4 および表 2 より SRGAN では画像の視覚的な比較においても指標による評価においても従来手法が優れていることが確認できる. これらのことより, 提案手法において 直接 GAN 系超解像である SRGAN を用いることは適していないと考えられる.

一方で, EDSR において各手法における各指標を比較すると, PSNR および SSIM では提案手法が, BRISQUE においては 0.04pt 差で提案手法+Zero-padding が高品質を示す値となっている. 一方で, 各出力画像を比較すると, 提案手法では, ビット深度不足による段状のノイズは完全に除去されている. また, 提案手法において BRISQUE では, 細かいテクスチャが多く含まれている画像において高い値となっている. これらのことより, 提案手法において PSNR 系超解像を用いることは適していることを実験により確認できた. また, 提案手法による解像度とビット深度の同時拡張は PSNR 系超解像用いた場合に有効であると言えると考えられる.

5. 結論

本研究では, 超解像を用いて解像度とビット深度を同時に拡張する手法を提案した. また, 評価実験により, PSNR 系の超解像である EDSR を用いた際に, 解像度の拡張およびビット深度の拡張のどちらにおいても提案手法が有効であることを確認した.

参考文献

- [1] 総務省, "4K・8K ロードマップに関するフォローアップ 会合 第二次中間報告」の公表", 平成 27 年 7 月 30 日
- [2] Jing Liu, Wanning Sun, Yuting Su, Peiguang Jing, Xiaokang Yang, "BE-CALF: Bit-Depth Enhancement by Concatenating All Level Features of DNN", IEEE Transactions on Image Processing, Volume.28, Issue.10, Oct. 2019.
- [3] Eirikur Agustsson, Radu Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July. 2017.
- [4] Eirikur Agustsson, Radu Timofte, Van Gool, Luc and Yang, Ming-Hsuan and Zhang, Lei and Lim, Bee and others, "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July. 2017.
- [5] Anish Mittal, Anush K. Moorthy and Alan C. Bovik, "Blind/Referenceless Image Spatial Quality Evaluator", 2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), April. 2012.

2019 年度

早稲田大学大学院基幹理工学研究科 情報理工・情報通信専攻

修士論文

超解像を用いたビット深度拡張に関する研究

Research of the Bit Depth Extension by Super-Resolution

梅田聖也

(5118F017-4)

提出日：2020.01.29

指導教員：渡辺裕 印

研究指導名：オーディオビジュアル情報処理研究

目次

第 1 章	序論	1
1.1	研究背景	1
1.2	本研究の目的	2
1.3	本論文の構成	2
第 2 章	関連技術	3
2.1	はじめに	3
2.2	Convolutional Neural Network (CNN)	3
2.3	超解像技術	4
2.3.1	技術概要	4
2.3.2	分類	4
2.3.2.1	超解像技術の分類	4
2.3.2.2	再構成型超解像	4
2.3.2.3	学習型超解像	4
2.3.2.3.1	学習型超解像概要	4
2.3.2.3.2	PSNR 系超解像	5
2.3.2.3.3	GAN 系超解像	6
2.3.2.3.4	問題点	7
2.4	ビット深度拡張	8
2.4.1	技術概要	8
2.4.2	分類	8
2.4.2.1	従来的手法	8
2.4.2.2	深層学習的手法	8
2.5	評価指標	9
2.5.1	画質評価指標	9
2.5.2	完全参照メトリック	9
2.5.3	非参照メトリック	10
2.6	むすび	10
第 3 章	予備実験	11
3.1	はじめに	11
3.2	予備実験 概要	11
3.3	予備実験 1	11
3.3.1	予備実験 1 概要	11
3.3.2	予備実験 1 結果	12

3.3.3	予備実験 1 考察・結論	13
3.4	予備実験 2.....	13
3.4.1	予備実験 2 概要	13
3.4.2	予備実験 2 結果	14
3.4.3	予備実験 2 考察・結論	15
3.5	予備実験 3.....	16
3.5.1	予備実験 3 概要	16
3.5.2	予備実験 3 結果	17
3.5.3	予備実験 3 考察・結論	21
3.6	むすび	21
第 4 章	提案手法	22
4.1	まえがき	22
4.2	提案手法概要	22
4.3	むすび	22
第 5 章	評価実験	23
5.1	まえがき	23
5.2	実験概要.....	23
5.3	実験 1 結果	26
5.4	実験 1 考察・結論	30
5.5	実験 2 結果	31
5.6	実験 2 考察・結論	35
5.7	むすび	36
第 6 章	結論・今後の課題	37
6.1	結論	37
6.2	今後の課題	37
	謝辞.....	38
	参考文献	39
	図一覧	42
	表一覧	43
	研究業績	44

第1章 序論

1.1 研究背景

2020年現在、総務省のロードマップ[1]に従い、2020年東京オリンピックに向けて従来の Full High Definition(フルHD: 1920x1080画素)映像から次世代の4K(3840x2160画素)や8K(7680x4320画素)などの超高精細映像への移行が始まっている。一般社団法人電子情報技術産業協会の2019年民生用電子機器国内出荷統計[2]によれば、4K対応の薄型テレビの国内出荷実績は前年比で134.5%となっており、より我々の生活に身近になってきていることがわかる。これに伴い従来コンテンツを次世代映像に変換したいという需要が高まっている。また、2018年12月よりNHKのBS4K及び民放キー局系列のBS放送各社(但し、BS日テレを除く)で4K放送が始まり、NHKではBS8Kにおいて8K放送が開始されている[3]。

これらの次世代映像は解像度以外にも、従来の映像と異なる特徴を持っている。一つ目の特徴としてあげられるのが多階調表現である。従来の映像ではRGBそれぞれに8-bitが割当てられ最大で1600万階調で映像の色を表現している。これに対し次世代映像ではRGBそれぞれに10-bitを割当てることで最大で約10億階調の表現を可能とし、より自然に近い色や明るさの変化、グラデーションの再現を実現している。二つ目の特徴として High Dynamic Range(HDR)技術を用いた輝度範囲の拡大が挙げられる。これは従来の輝度範囲と比べて広い範囲を使う技術であり、表現できる明るさの範囲が広がり、暗いところはより暗く表現されるだけでなく、暗い中でも様々な強弱をつけた表現が可能となる。この他にも従来の色域よりも広色域を用いることにより再現できる色の幅の増加や、フレームレートを従来の60fpsから120fpsまで上げることによる時間方向の解像度の拡大などの特徴も持っている。これらの特徴により次世代映像では従来の映像とは一線を画す映像を実現している。

一方でそれに伴う弊害もある。その中でも最も顕著であるのがデータ量の増加である。解像度のみで従来の情報量と比較して4倍から16倍に増加している。そのため、映像符号化方式 HEVC/H.265[4]で圧縮した場合でも80~100Mbps[5]の伝送路容量が必要となる。一方で現行の地上デジタル放送の伝送路容量は付加情報を除くと約30.9Mbps[6]しかなく、8K映像を地上波で放送するための技術開発が求められている。

近年、深層学習がこれらの問題を解決する可能性があるとして注目されている[7]。この技術では事前に大量のデータを用いてニューラルネットワークを学習しておくことで想定される入力があった際に学習データに基づいて任意の出力を得ることができる。この技術を映像の圧縮等に応用することで高い圧縮率を実現する方式も報告されている[8]。また、クライアント側で従来映像を8Kに変換するような製品も発売されている。一方でこの技術

を用いて高い精度を実現するためにはニューラルネットワークの規模を大きくする必要がある。しかしながら、実際の運用を考慮するとあまり大きなニューラルネットワーク実用には難しいという問題がある。また、市販の製品は解像度の変換のみで他の次世代映像の特徴の変換はされておらず真の意味での次世代映像への変換をできていないという問題などを抱えている。

そこで、本研究では、次世代映像の解像度とビット深度に着目し、ニューラルネットワークのサイズを抑えつつ解像度とビット深度の二つの特徴を深層学習を用いて従来映像から次世代映像に変換できる方式を提案する。また、実験により提案手法の有効性を検証する。

1.2 本研究の目的

一般的に、次世代映像は、膨大なデータ量を持っており、従来方式で直接扱うことが難しいという問題を抱えている。また、次世代映像の普及に伴い従来コンテンツを次世代映像に変換したいという需要がある。解像度を提案手法により N 倍にする方式では全体データ量は単純計算で $1/N^2$ に縮小することができる。また、ビット深度においても M ビット増加させるとすると $(\text{解像度} * M^3)$ -bit 分のデータ量を削減することができると考えられる。また、提案手法による従来映像から次世代映像までの拡張をクライアント側で実行すると想定すると、それまでの伝送経路では従来の方式をそのまま使用することができるという利点も挙げられる。

これらのことより提案手法を実現することで再生機器のみを次世代対応のものとするれば次世代映像を楽しむことができるようになり、より次世代映像の発展・普及につながると考えられる。

1.3 本論文の構成

第1章は本章であり、本研究における背景、目的について述べる。

第2章では、本研究における主要技術である CNN や超解像技術について述べる。

第3章では、超解像における問題点を検証するための実験について述べる。

第4章では、本研究で提案する手法について述べる。

第5章では、提案手法を評価する実験について述べ、提案手法の有効性について述べる。

第6章では、本研究における結論と今後の課題について述べる。

第2章 関連技術

2.1 はじめに

本章では、本研究で用いる主要技術である CNN，超解像技術，ビット深度拡張に加え，その他関連技術について述べる。

2.2 Convolutional Neural Network (CNN)

CNN とは人間の脳の神経細胞(=ニューロン)の構造を参考に構成された Neural Network に畳み込み層を追加した構造を持つコンピュータアルゴリズムである[9]。入力に対して小領域ごとに畳み込みを適用することでより正確に特徴量を抽出することが可能になり，画像認識や音声分析などのタスクにおいて従来の手法を大きく上回る精度を実現している。一般的に CNN は図 2.2.1 のような構造からなる。

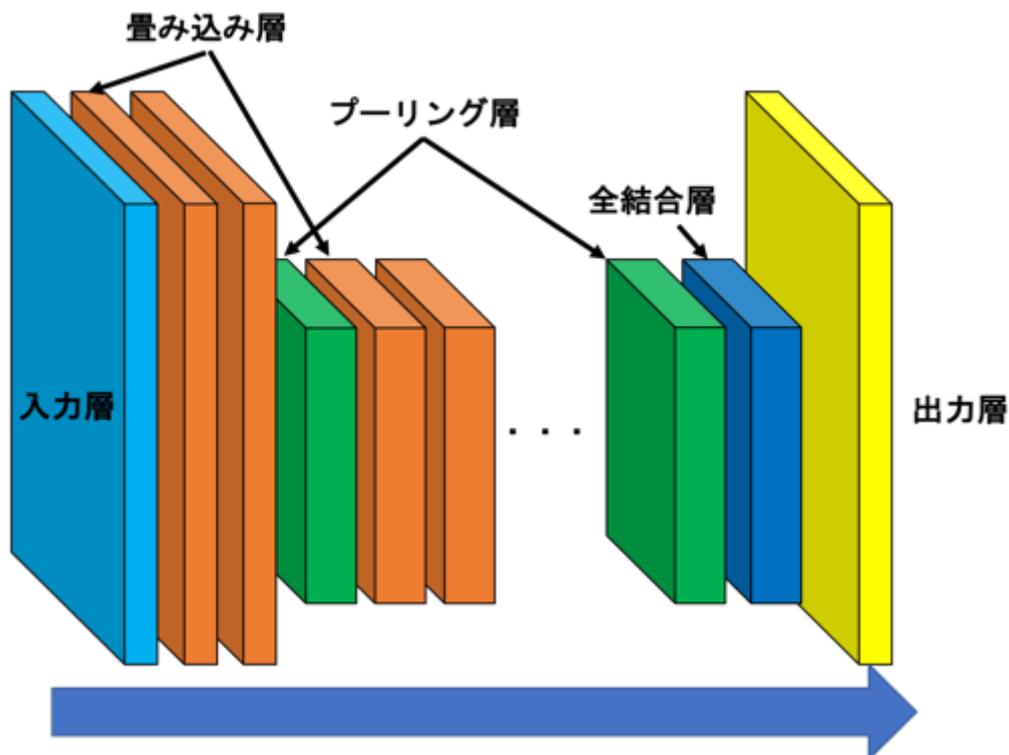


図 2.2.1 CNN の構造

CNN は入力層からの入力に対し，畳み込み層，プーリング層を適用し，これらの層から抽出した特徴に対し活性化関数をかけたのちに最後の全結合層で任意の出力を得ることができるアルゴリズムとなっている[10]。層の数，畳み込みを適用するフィルタのサイズなどを変化させることで画像処理や音声分析など様々な分野での適用可能となっている。

2.3 超解像技術

2.3.1 技術概要

超解像技術とは入力された信号に対してデジタル信号処理を施し、入力信号よりも高い解像度の出力を作り出す技術のことを指す。一般的には画像や映像などが入力され使用されることが多い。ビデオカメラや液晶テレビ、DVDの再生時に入力信号が出力画面の解像度に満たない際に、足りない画素を補う目的で超解像技術が用いられることもある。

2.3.2 分類

2.3.2.1 超解像技術の分類

超解像技術は、従来の信号処理・画像処理を使用し、足りない画素を補間する手法とディープラーニングを用いて足りない画素を推測する手法の大きく二つに分類することができる。

2.3.2.2 再構成型超解像

CNNを用いない超解像手法の一つとして再構成型超解像が挙げられる。この手法は対象の画像の他に対象に類似している画像を複数用いることで、対象画像の足りない画素を補間し、超解像を実現している。そのため、一般的な画像には適用が難しく、映像などの容易に類似画像を得ることができる場合でのみ適用することが可能となる。また、位置合わせや補間処理に多くの時間を必要とし、入力画像の劣化度合いやノイズ成分に超解像精度が大きく依存しているという問題を抱えている[11]。

2.3.2.3 学習型超解像

2.3.2.3.1 学習型超解像概要

超解像技術の中で、多階層からなるCNNに対し、事前に大量の低解像度画像と高解像度画像のペアを学習させることで未知の低解像度画像が入力された際に高解像度画像を作り出す手法を学習型超解像という。この手法では従来のような補間ではなく事前学習に基づき足りない画素を推定することで超解像を行う。そのため、低解像度画像に含まれない高周波成分の推定ができ、画像のボケやジャギーなどのノイズを削減することができる。このCNNを用いた学習型超解像は、2014年にC.Dong氏らが発表した低解像度画像の入力と高解像度画像の出力の関係をend-to-endで機械学習するSRCNN[12]がはじまりとされており、今現在もこのSRCNNのネットワークをベースに、層の数や各フィルタサイズを増減させ、ネットワーク全体をより深くすることで超解像の精度向上を図るなど、様々な超解像手法が考案されている。SRCNNのネットワークを図2.3.2.3.1.1に示す。SRCNNでは3層のネットワークで超解像を実現している。この学習型超解像は、ネットワークの学習の手法や評価関数の種類によってPSNR系超解像とGAN系超解像の二つに分けることができる。それぞれの手法で超解像した画像を図2.3.2.3.1.2に示す。

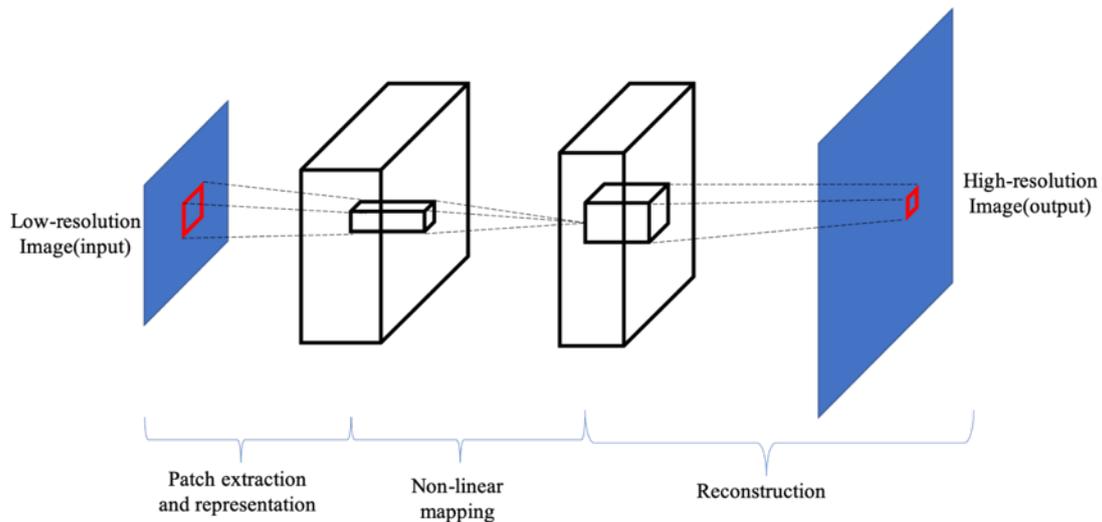


図 2.3.2.3.1.1 SRCNN のネットワーク構造



図 2.3.2.3.1.2 各超解像手法の比較

2.3.2.3.2 PSNR 系超解像

超解像手法のうちネットワークを学習する際に使用する評価の基準が原画を基にした MSE ベースのものは PSNR 系超解像に分類される。図 2.3.2.3.2.1 に PSNR 系超解像手法の一つである EDSR[13]のネットワーク構造を示す。EDSR では Residual な層を追加することで深いネットワークでも最適化を正しく行えるようにし、精度の向上を図っている。また、従来の超解像のネットワークに存在した Batch Normalization 層の削除や Pixel shuffle[14] によるアップサンプリングなどの工夫が取り入れられている。この種類の多くの超解像では、補間された画素が原画に近づくように学習される。そのため、低解像度画像を入力した際によりオリジナルに近い高解像度画像を得ることができる。一方で、内部的には原画の高解像度画像に対し Bicubic 法で任意の解像度まで縮小した画像と原画の対応関係を基に、出力画像の画面全体での PSNR が高くなるように学習されている。そのため、ネットワーク全体で平均画像を作るように動作してしまい、全体的にぼやけたような画像を出力してしまう傾向がある。実際に図 2.3.2.3.1.2-(b)をみると原画と比較すると山の雪などのテクスチャが潰れていて全体的に少しぼやけているのがわかる。このように PSNR 系

超解像では高周波成分を多く含むような画像や、細かいテクスチャなどは再現されにくいという問題がある。

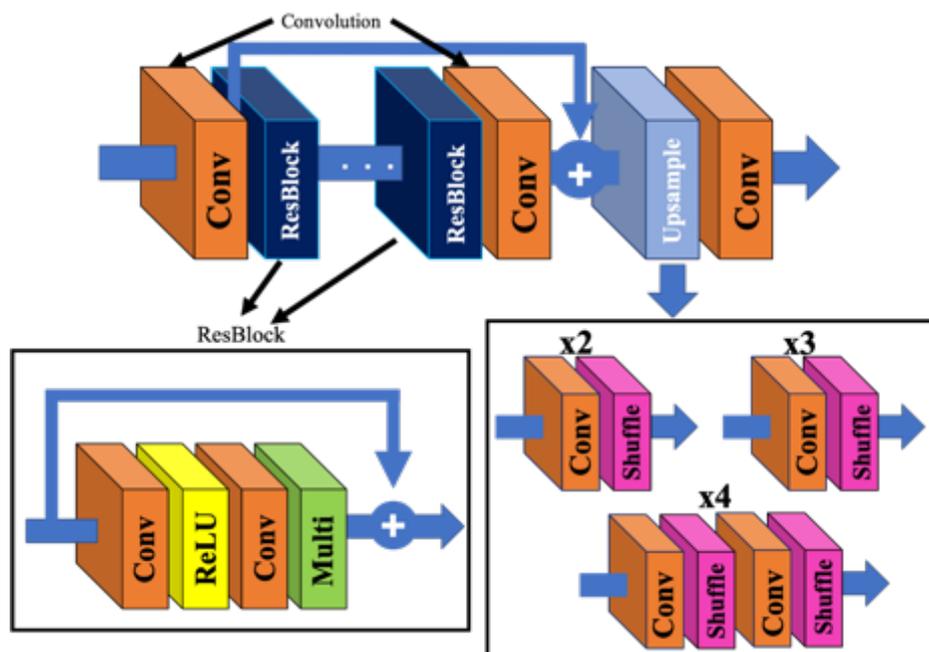


図 2.3.2.3.2.1 EDSR のネットワーク構造

2.3.2.3.3 GAN 系超解像

超解像手法のうち従来の超解像のネットワークの構成に敵対性ネットワーク(GAN)[15]を加え、評価関数に perceptual loss を使用したものは GAN 系超解像に分類される。図 2.3.2.3.3.1 に GAN 系超解像手法の一つである SRGAN[16]のネットワーク構造を示す。超解像のネットワークに perceptual loss を加えることで、原画との距離に応じた評価ではなく、人間の知覚に近い画像の品質を基準にした評価を可能にしている。また、GAN 構造をとることで補間画素を新たに作り出すことができるので、より高品質な細かいテクスチャや高周波成分を作り出すことができる可能性がある。一方で、この超解像手法で出力される高解像度画像は低解像度画像を参考に新たに作り出された高解像度画像となるので、原画には存在しないテクスチャが生成されることがある。そのため、原画等と比較する指標で評価することができず、参照画像を用いない評価指標で評価する必要がある。また、実際に図 2.3.2.3.1.2-(c)と(b)を比較すると、(c)の方が山の細かいテクスチャが再現されていることがわかる。一方で山の下の方には原画(a)には存在しないようなテクスチャが現れていることも確認できる。

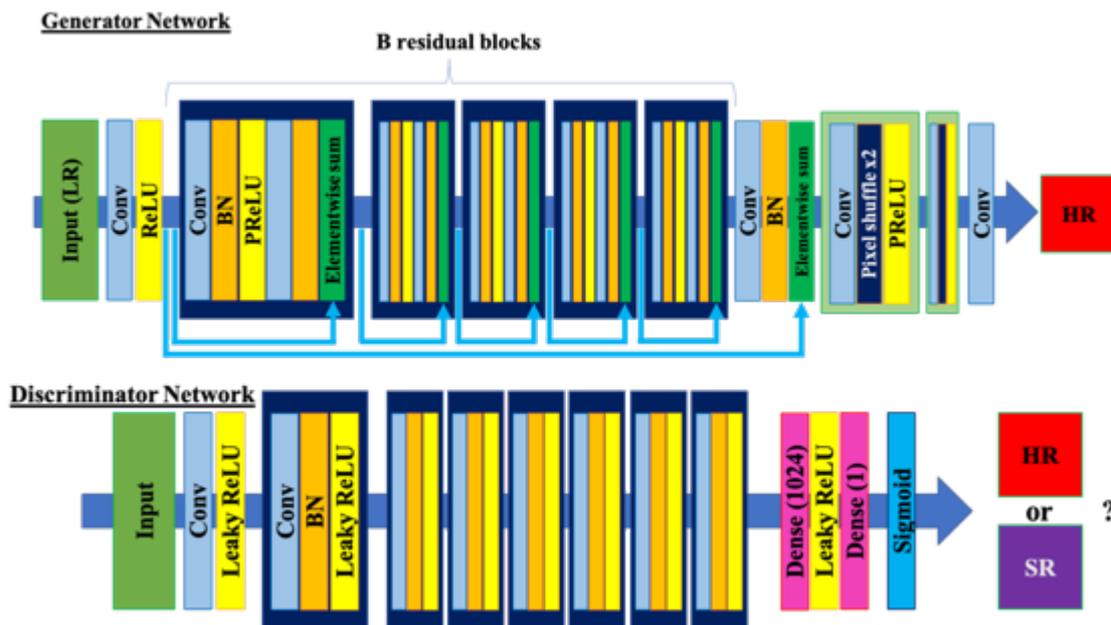


図 2.3.2.3.3.1 SRGAN のネットワーク構造

2.3.2.3.4 問題点

一般的に超解像技術は学習させる際に低解像度画像と高解像度画像のペアで学習させる。この際、高解像度画像のデータセットに対し Bicubic 法を適用し縮小することで任意の倍率の低解像度画像を得ている。そのため、超解像のネットワークは Bicubic 法による縮小の劣化過程を学習することになり、画像を入力する際も Bicubic 法で縮小した低解像度画像を入力することが前提となる。したがって、超解像には Lanczos などの異なるカーネルで縮小した低解像度画像を入力した際に高い超解像精度が得られない可能性があるという問題を抱えている。また、上述のような縮小過程を挟むことで得られる低解像度画像は一般的な低解像度画像とは異なり少なからず折り返し歪みを含んでいる。そのため、超解像はこの折り返し歪みを基に高解像度画像の高周波成分を推定しているのではないかという指摘もされている。これらの問題は超解像は縮小された画像にのみ適用することができるという問題に帰結する。これは実際に適用する場合を想定すると、超解像は 8K カメラで撮影、4K に縮小後に伝送するという方式では精度を保証できるが、4K カメラで撮影した映像を 8K にする場合では精度を保証できないという問題につながる。そのため、超解像を適用する方式ではこの問題を考慮する必要がある。

2.4 ビット深度拡張

2.4.1 技術概要

ビット深度拡張は、一般的にデータを扱う際に従来の割り当てられたビット数よりも多くのビット数を割り当てたい際に増えるビット数を補う目的で使用される。音声や画像などのデータの場合、各規格ごとにデータに使用するビット数が決定されている。そのため、規格が異なると各データに割り当てられるビット数が異なることがあり、他規格への変換の際に不足するビット数を補う目的としてもビット深度拡張が用いられる。この技術は、従来の画像処理技術を用いる手法と CNN などの深層学習を用いてビット深度拡張する手法の大きく二つに分けることができる。

2.4.2 分類

2.4.2.1 従来的手法

ビット深度拡張の最も簡単な手法の一つとして”Zero-padding”と呼ばれる手法がある。これは増えたビットに’0’を入れるという手法で、処理が非常に簡単である一方で、ビット深度拡張の精度としてはあまり高くない。この他にも Multiplication-by-ideal-gain (MIG), Bit Replication (BR) [17], Minimum Risk based Classification (MRC) [18], Contour Region Reconstruction (CRR) [19]といった手法も提案されているが、いずれも深層学習的手法を超える精度を実現していない[20]。

2.4.2.2 深層学習的手法

従来的な画像処理を用いたビット深度拡張より高い品質を実現する手法として CNN を用いた深層学習的手法がある。この手法では、ビット深度が足りない画像とオリジナル画像を大量に用意し、この対応関係を学習することで、従来手法より高い精度でのビット深度拡張を実現している。一方で、ビット深度拡張を実現するためには大量のパラメータが必要になり、ニューラルネットワークの規模も大きくなるため処理に時間がかかってしまうという問題がある。この手法の一つである Bit-depth Enhancement by Concatenating All Level Features of DNN (BE-CALF) [20]では、Zero-padding でビット深度拡張された画像とオリジナルの画像の残差を学習することでビット深度の拡張を実現している。この手法では、従来的手法と比較して PSNR で平均 3dB 以上の品質の向上を実現している。一方で、実行時間は従来的手法の 100 倍以上を要する結果となっている。

2.5 評価指標

2.5.1 画質評価指標

一般的に超解像やビット深度拡張などの性能を評価する際には得られた画像から画質評価指標を算出し比較するという手法が採られることが多い。この画質評価には完全参照メトリック (Full-Reference Metric) と呼ばれるものと非参照メトリック (No-Reference Metric) と呼ばれるものの大きく分けて 2 種類存在する。これらは、評価対象の画像がどのような特徴を持っているか、どのような過程で得られたものなのかなどによって使い分けられる。

2.5.2 完全参照メトリック

完全参照メトリックは評価したい画像と対象画像の原画(=歪みの含まない画像)を比較しその差を数値化することで画質の評価を行う指標である。そのため、対象画像は原画と比べてどの程度歪みやノイズを含んでいるかを評価する指標になるので歪みを示す指標とも呼ばれる。この指標の中で最も一般的なものとして Mean Square Error (MSE: 平均二乗誤差) が挙げられる。MSE は対象画像と参照画像の同じ位置の画素値の差分の 2 乗を全ピクセルで算出し、平均を求めることで得ることができる。解像度が $m \times n$ からなる画像 I, K において MSE は次の式(2.1)で表される。

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (2.1)$$

この MSE は 2 画像間の差分を数値化したものであるため、差が大きくなると差に鈍くなるという人の知覚特徴が加味されていない。この MSE を人の知覚特徴に合わせた画質評価指標が Peak Signal-to-Noise Ratio (PSNR: ピーク信号対雑音比) である。PSNR は次の式(2.2)で定義される。ただし、 MAX_I は画像がとりうる最大の画素値とする。

$$PSNR = 10 \log_{10} \frac{MAX_I^2}{MSE} \quad (2.2)$$

この PSNR を用いることで原画とどの程度異なるかを確実に評価することができる。一方で、対象画像の全画素が水平方向に 1 画素ずれた場合に、人の目では差を感じられないが、PSNR では著しく悪い値になってしまう問題がある。また画面全体の多くで原画と同じ画素値を持つがごく一部分局的に大きな差を含むような画像と、画面全体で少しずつ異なるような画像がある時に人の知覚では前者の方が低品質な画像と評価されるのに対し、PSNR では同じ値をとってしまい同程度の品質と評価される可能性があるという問題がある。この問題を解決するために作られた指標が Structural Similarity (SSIM: 構造的類似性指数) である。SSIM は式(2.3)で定義される。

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2.3)$$

μ_x, μ_y は平均, σ_x, σ_y は標準偏差, σ_{xy} は共分散を示す.

対象の画素のみでの評価ではなく周囲のピクセルとの相関を取り込むことで PSNR での問題の緩和を実現しつつ, 人の知覚特徴も加味された指標となっている. 一般的に, 超解像の分野のうち PSNR 系超解像ではここで述べた完全参照メトリックが用いられることが多い.

2.5.3 非参照メトリック

非参照メトリックは, 完全参照メトリックとは異なり, 統計的な特徴量や, 大量の自然画像からなるデータベースを学習したモデルを用いることで評価したい画像単体での画質の評価を実現し, より人の知覚特徴に合った評価をすることができる指標となっている. 代表的な指標として, Natural Image Quality Evaluator (NIQE)[21]や Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)[22]などがある. これらの指標の詳細な説明は本稿では省略するが NIQE および BRISQUE は基本統計量として Natural Scene Statistic(NSS)ベースの特徴量を使用している. NIQE では大量の画像からなるデータベースから得られた特徴と, 対象画像から算出した Natural Scene Statistic(NSS)ベースの特徴の距離を算出して指標としている. BRISQUE ではこの特徴量と主観的スコアを Support Vector Machine(SVM)を用いて学習, 指標の算出を行っている.

GAN 系の超解像では出力画像は Generator によって作り出されたもののため, 原画とは大きく異なるが人の目で見ると品質的に優れていることが多い. 非参照メトリックではこの差を定量的評価することが可能である. そのため, これらの指標は一般的に超解像のうち GAN 系の超解像で多く用いられる. PSNR 系の超解像の一部でもこれらの指標を用いることで正確に画像の品質評価, 超解像の性能評価を行える場合も存在する. また, この他にも Ma[23]や Perceptual score[24]と呼ばれる指標もある.

2.6 むすび

本章では, 本研究で用いる主要技術である CNN や超解像技術, ビット深度拡張, 評価指標における種類や用途, 問題点について述べた.

第3章 予備実験

3.1 はじめに

本章では、本研究の主要技術である、超解像技術の縮小過程に依存して品質が変化するという問題点について検証した予備実験について述べる。

3.2 予備実験 概要

前章でも述べたように、超解像技術は縮小過程が異なる場合、超解像精度が著しく悪くなるという問題と折り返し歪みの情報を基に高周波成分を推定している可能性があるという問題を抱えている。そこで、本研究では、この問題を予備実験により検証した。

3.3 予備実験 1

3.3.1 予備実験 1 概要

予備実験 1 では、超解像のネットワークが学習していない縮小方法で縮小された低解像度画像が入力された際に超解像の性能が低下することについて検証実験を行う。予備実験 1 の概略図を図 3.3.1.1 に示す。

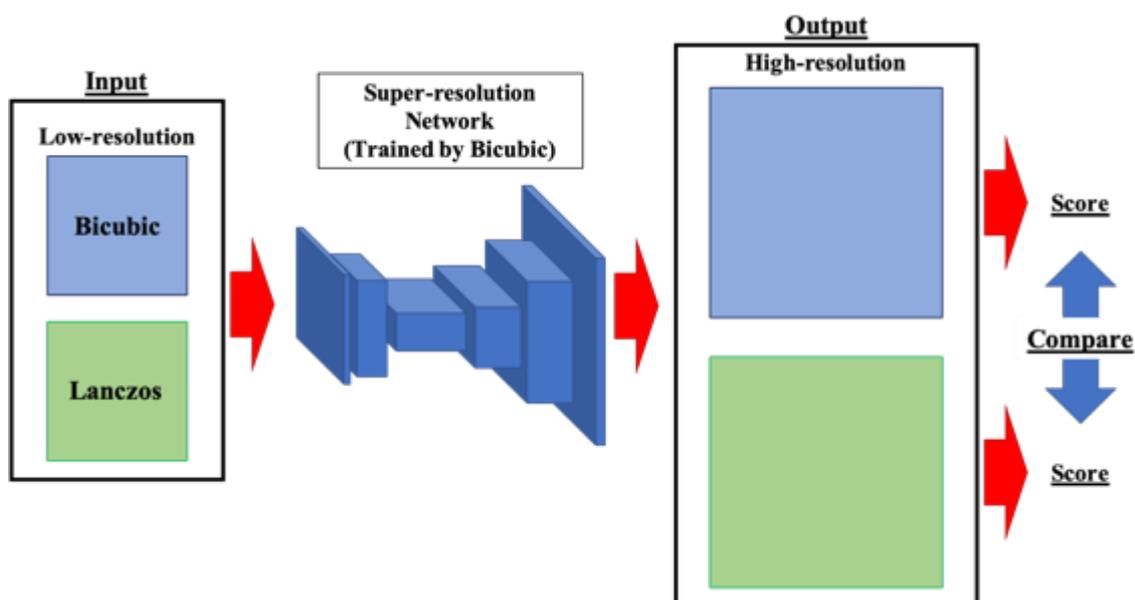


図 3.3.1.1 予備実験 1 概略図

予備実験 1 では、Bicubic 法で作成した低解像度画像と高解像度画像を用いて学習した超解像のネットワークに対し、Bicubic 法で縮小した低解像度画像と Lanczos フィルタを用いて縮小した低解像度画像をそれぞれ入力し、高解像度画像をそれぞれ作成する。ここで得られた高解像度画像に対し、PSNR および SSIM をそれぞれ算出し、これを比較することで超解像の性能の変化を検証する。今回は使用する超解像ネットワークとして EDSR を

選択し、DIV2K データセット[25][26]のうち 800 枚を学習用画像、90 枚をテスト画像とした。

3.3.2 予備実験 1 結果

予備実験 1 で得られた出力画像を図 3.3.2.1, 図 3.3.2.2 に示す。

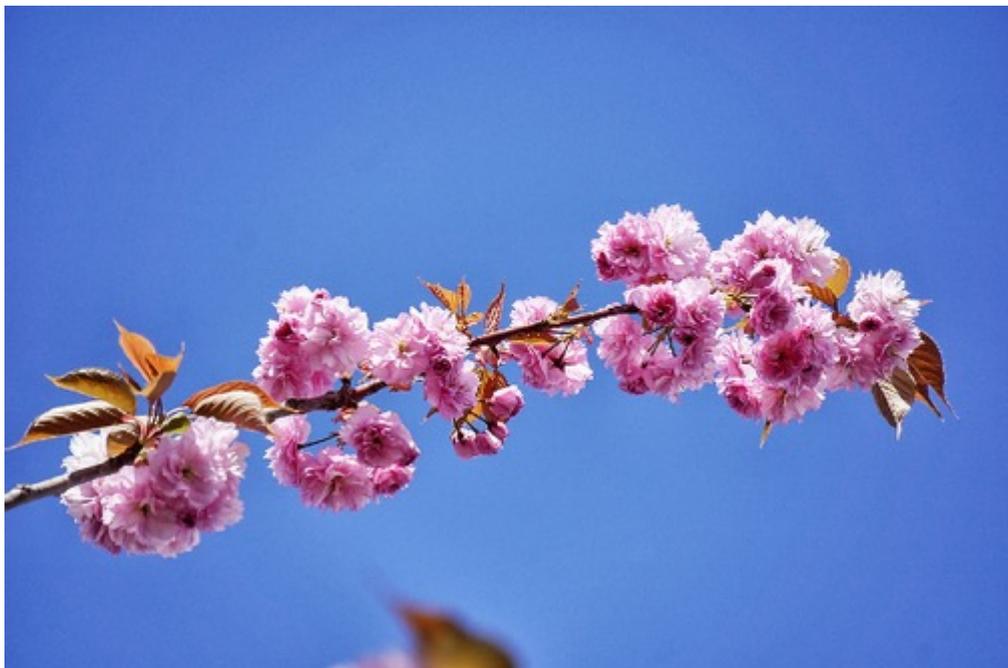


図 3.3.2.1 予備実験 1 出力画像 1(入力:Bicubic 縮小画像) (“DIV2K Dataset”[25][26])

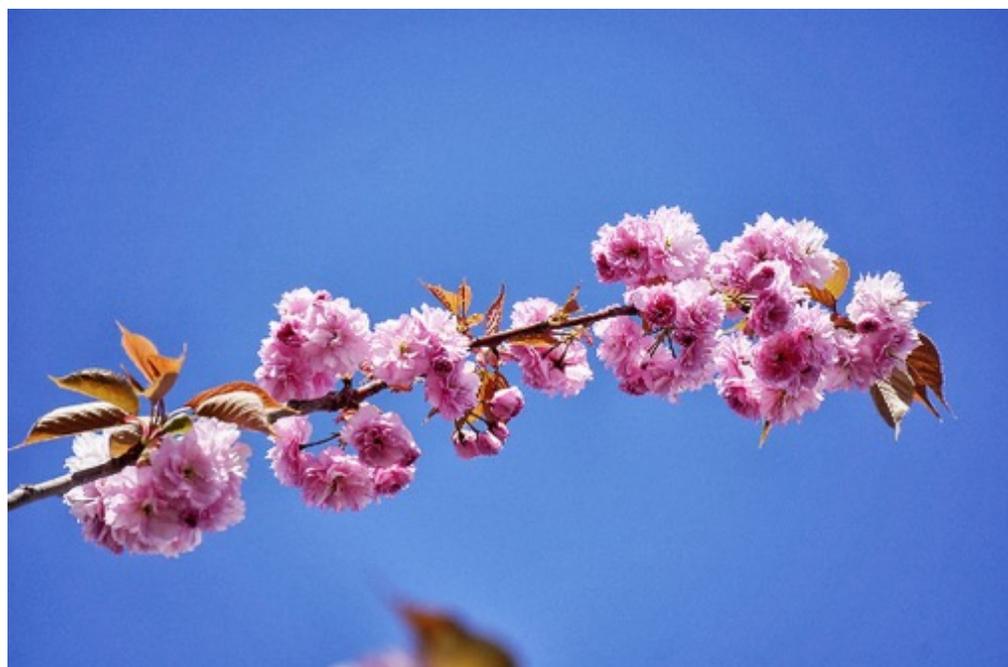


図 3.3.2.2 予備実験 1 出力画像 2(入力:Lanczos 縮小画像) (“DIV2K Dataset”[25][26])

この二つの画像を比較すると、あまり大きな差ではないが出力画像2の方がコントラストがはっきりしており、より偽輪郭が強調されていることが確認できる。次に算出した評価指標について表 3.3.2.1 に示す。ここでは入力画像 90 枚から得た出力画像 90 枚に対し PSNR と SSIM をそれぞれ算出し、それらの値の平均値を示す。

表 3.3.2.1 予備実験 1 の結果

入力画像の縮小方式	PSNR dB	SSIM
Bicubic	24.50	0.8574
Lanczos	23.97	0.8483

表 3.3.2.1 より、実験 1 では、PSNR、SSIM とともに超解像のネットワークに学習させた縮小方法である Bicubic 法で縮小した画像を入力とした方が Lanczos フィルタで縮小した画像を入力した時よりも平均で+0.53dB 品質が向上することが確認できる。また、平均ではなく各画像での比較では最大で約+1.0dB ほど高い値をとるということも確認できた。

3.3.3 予備実験 1 考察・結論

予備実験 1 で学習した縮小画像を入力した際に評価指標で高い値をとったという結果より、超解像の問題点である縮小方法に依存した超解像が行われていることが確認できた。一方で、評価指標では確かな差が生じているが視覚的には、あまり大きな差が生じていないことも確認できた。

3.4 予備実験 2

3.4.1 予備実験 2 概要

予備実験 2 では、入力画像の縮小過程を学習したネットワークを適切に選択することでより高い精度で超解像を行えることを実験により検証する。予備実験 2 の概要図を図 3.4.1.1 に示す。予備実験 1 では Bicubic 法による縮小を学習した超解像ネットワークを使用した。予備実験 2 ではこれに加えて Lanczos フィルタを用いた縮小画像で学習した超解像ネットワークを用いる。この二つの超解像ネットワークに Lanczos フィルタを用いて縮小した低解像度画像を入力し、高解像度画像を生成、これらを指標により評価する。学習に使用する画像および入力画像は予備実験 1 と同じものを使用した。

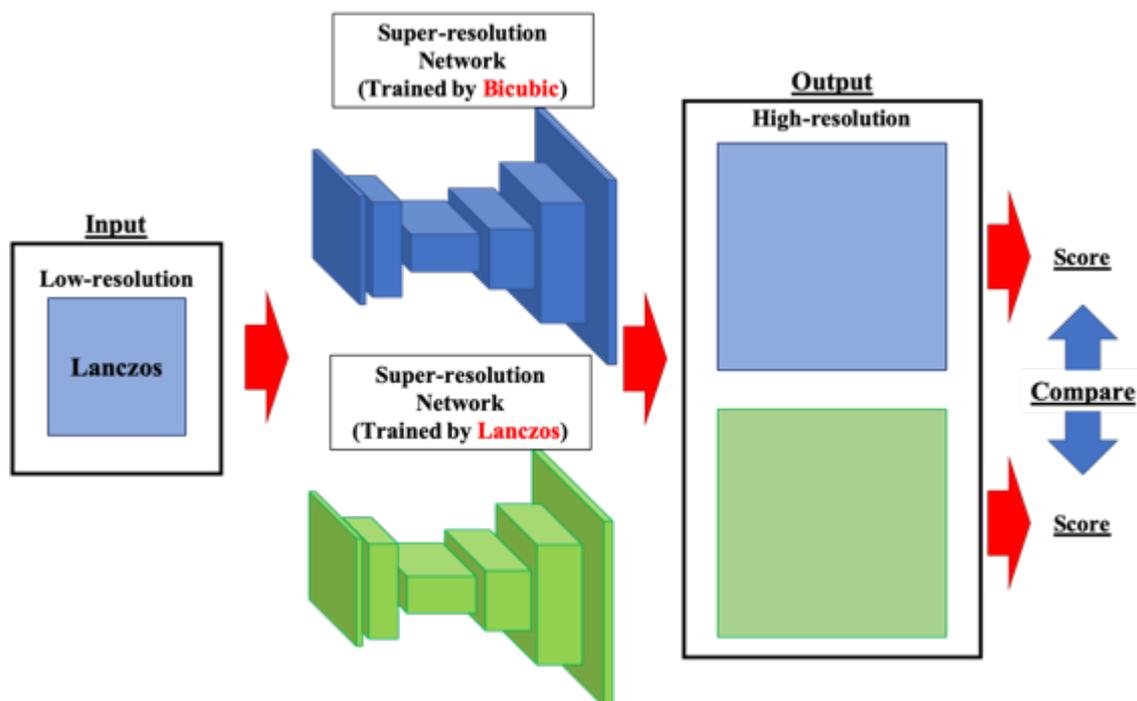


図 3.4.1.1 予備実験 2 概要図

3.4.2 予備実験 2 結果

予備実験 2 で得られた画像をそれぞれ比較したものを図 3.4.2.1 に示す。図中左側(a)は Bicubic 法による縮小画像を入力した際に得られる出力画像の一部，図中右側(b)は Lanczos フィルタによる縮小画像を入力した際に得られる出力画像の一部となっている。

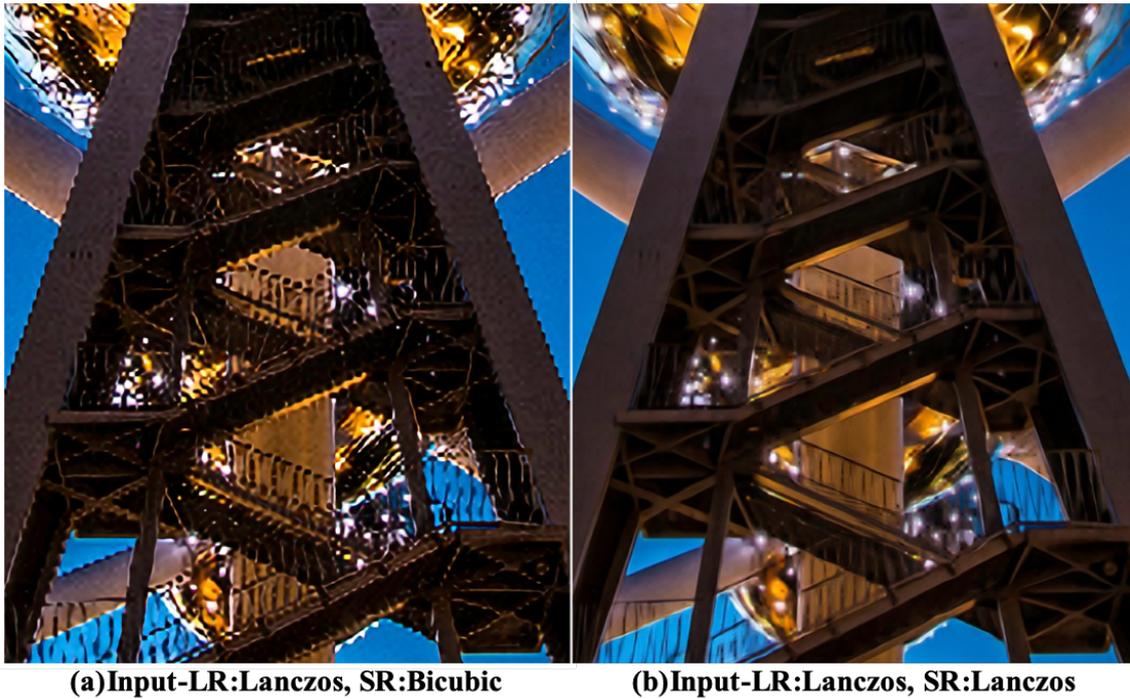


図 3.4.2.1 予備実験 2 出力画像の比較(“DIV2K Dataset”[25][26])

図中(a)と(b)を比較すると、(a)では鉄柱や階段部分などのエッジ部分に偽輪郭が生じ、本来直線の箇所にも階段上のノイズが発生してしまっていることが確認できる。一方で、(b)ではそれらのノイズが見受けられず、スムーズな直線として表現されている。次に、算出した指標を表 3.4.2.1 に示す。

表 3.4.2.1 予備実験 2 の結果

ネットワーク 入力画像	Trained by Bicubic		Trained by Lanczos	
	PSNR dB	SSIM	PSNR dB	SSIM
Lanczos	23.97	0.8483	28.45	0.9023

表 3.4.2.1 より、Lanczos フィルタにより縮小した画像に対し、Lanczos フィルタの縮小過程を学習した超解像を適用することで、Bicubic 法の縮小過程を学習した超解像を適用した時より、超解像の精度が PSNR で+4.48dB、SSIM でも 0.054pt 向上することが確認できる。

3.4.3 予備実験 2 考察・結論

予備実験 2 の結果より、入力画像の縮小過程に合わせて最適な超解像のネットワークを選択し、適用することで、超解像の精度が向上することが確認できた。また、今回使用している EDSR という超解像手法は Bicubic 法による縮小過程を学習することを前提にネットワークが最適化されている。そのため、超解像ネットワークのパラメータを Lanczos フィルタによる縮小画像に合わせて最適化をすれば超解像精度は更に向上すると考えられる。

3.5 予備実験 3

3.5.1 予備実験 3 概要

予備実験 3 では，超解像の精度が縮小画像に含まれる折り返し歪みにどの程度依存しているかの検証をする．予備実験 3 の概略図を図 3.5.1.1 に示す．

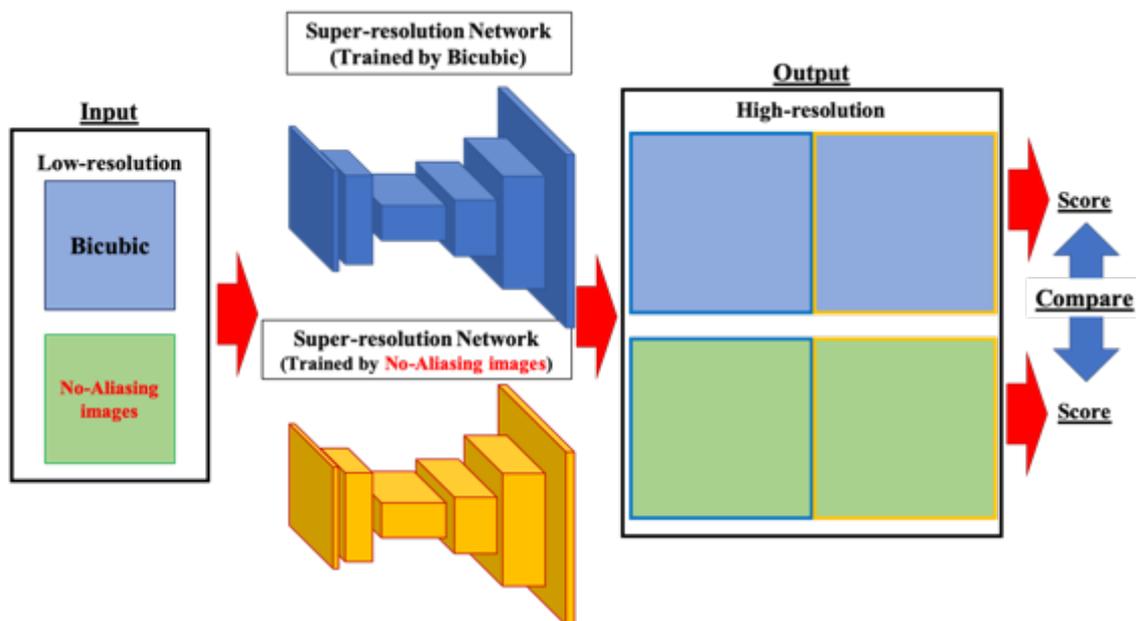


図 3.5.1.1 予備実験 3 概略図

予備実験 3 では，入力画像および超解像のネットワークの学習で使用する低解像度画像に折り返し歪みを全く含まない画像(図中 No-Aliasing images)を使用する．この画像は高解像度画像を二次元離散フーリエ変換(2D-discrete Fourier transform(DFT))を用いて空間周波数領域に変換し高周波成分を削除，これに対し，二次元逆離散フーリエ変換(2D-inverse discrete Fourier transform)を適用し，空間領域に戻したのちに，2:1 サブサンプリングすることで作成する．この No-Aliasing images の作成手順を図 3.5.1.2 に示す．データセットは予備実験 1，予備実験 2 と同じものを使用した．

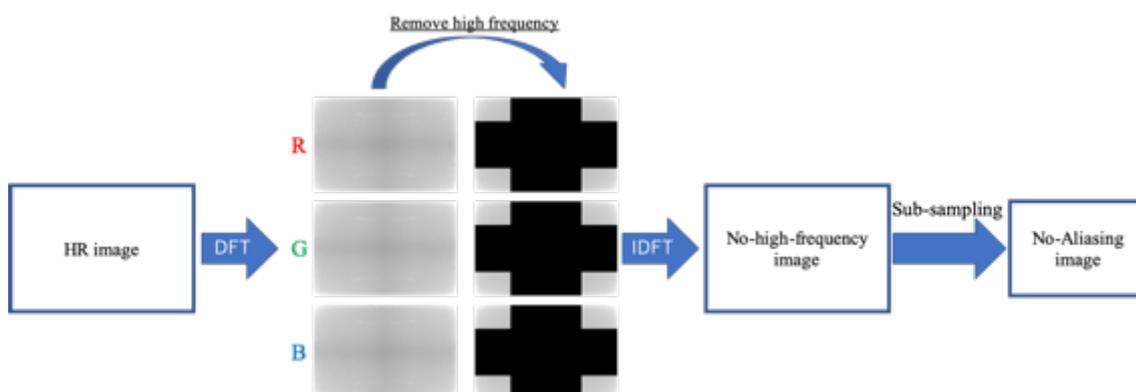


図 3.5.1.2 No-Aliasing images の作成手順

3.5.2 予備実験 3 結果

予備実験 3 で算出した各指標を表 3.5.2.1 に示す。また、得られた出力画像の比較を図 3.5.2.1 に示す。各指標は予備実験 1 と同様に出力画像 90 枚に対し評価指標をそれぞれ算出し、それらの値の平均値である。出力結果を比較するとあまり大きな差は見受けられないが、(a),(d)の出力画像が(b),(c)それぞれと比較すると若干ではあるが画像が鮮明にみえ、(b)(c)はややぼやけていることが確認できる。また、算出した指標をみると、折り返し歪みを含まないような低解像度画像を使用しても予備実験 1、予備実験 2 と同様に入力した低解像度画像と同じ縮小過程を学習した超解像のネットワークを使用した時の方が高い精度で超解像を行えており、+1.97dB の超解像精度の向上が得られることが確認できる。一方で、全体で比較すると Bicubic 法による縮小過程を学習した超解像ネットワークを使用した際の方が最も良いという結果となっている。

表 3.5.2.1 予備実験 3 結果 1

	Trained by bicubic PSNR dB	Trained by No-aliasing PSNR dB
Bicubic	31.78	30.60
No-Aliasing images	27.62	29.59



図 3.5.2.1 予備実験 3 出力画像の比較 1(“DIV2K Dataset”[25][26])

次に高解像度画像から高周波成分を除去した画像と、折り返し歪みを含まない縮小画像を学習した超解像に対応する画像を入力した際に得られる出力画像を比較する。算出した各指標を表 3.5.2.2 に、得られた出力画像の比較を図 3.5.2.2 にそれぞれ示す。

表 3.5.2.2 予備実験 3 結果 2

	No-Aliasing image SR by Trained No-aliasing images SR	No-High-frequency HR
PSNR dB	29.59	32.152
BRISQUE	32.85	33.72



(a) No-Aliasing image SR
by Trained No-aliasing images SR

(b) No-High-frequency HR

図 3.5.2.2 予備実験 3 出力画像の比較 2(“DIV2K Dataset”[25][26])

表 3.5.2.2 において PSNR を比較すると、超解像を適用した画像の方が -2.5dB ほど PSNR が低い値となっていることが確認できる。一方で図 3.5.2.2 の得られた出力画像を比較すると PSNR では優れているはずの図中(b)の方が(a)よりも全体的にぼやけたものとなっている。そこで PSNR に加えて、視覚的な差をより正しく評価できる指標である BRISQUE を算出した。この指標は画像の品質が高いほど低い値を取る。BRISQUE で比較をすると、超解像を適用した時の方が低い値を取り、画像の品質として優れているという結果となっている。次に、この 2 枚の画像をそれぞれ周波数領域に変換した際のスペクトルを図 3.5.2.3, 図 3.5.2.4 にそれぞれ示す。

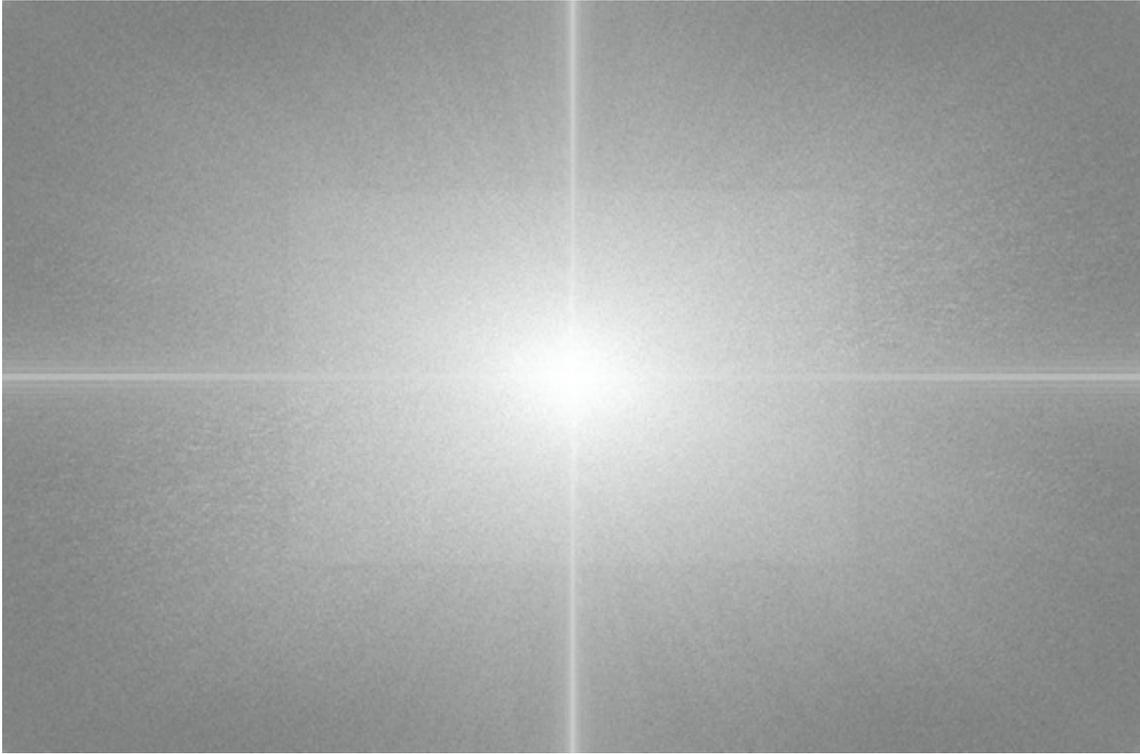


図 3.5.2.3 No-Aliasing image SR by Trained No-aliasing images SR の周波数スペクトル

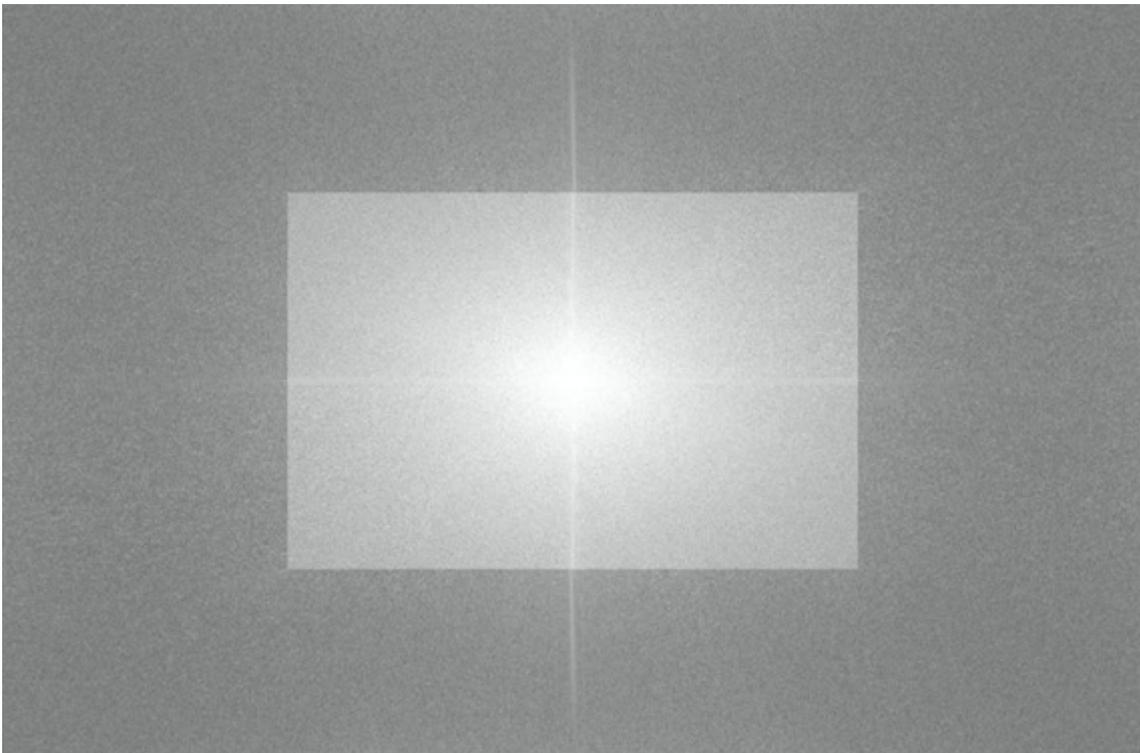


図 3.5.2.4 No-High-frequency HR の周波数スペクトル

この二つの画像を比較すると後者は高周波成分が除去されているのに対し、前者では折り返し歪みを含まない縮小画像を入力したにもかかわらず高周波成分が含まれていることが確認できる。

3.5.3 予備実験 3 考察・結論

予備実験 3 の結果より、入力画像が高周波成分の折り返し歪みを含まない場合でも、それに対応したデータセットで超解像のネットワークを学習することで精度の高い超解像を行うことができることが確認できた。学習型超解像の超解像精度が入力する縮小画像に含まれる折り返し歪みに完全に依存していると仮定した場合、学習データセットの変更のみで+1.97 dB という高い利得は得られないと考えられる。このことより学習型超解像において行なわれる超解像は折り返し歪みの情報に完全に依存しているわけではないと考えられる。

また、高解像度画像から高周波成分を除去した画像と、折り返し歪みを含まない縮小画像を対応する超解像のネットワークに入力した際に得られる出力画像を比較した際には、後者の超解像を適用した画像は視覚的にも評価指標 BRISQUE においても優れた結果となった。さらに、周波数領域での比較においても折り返し歪みを含まない縮小画像を入力したにもかかわらず出力画像は高周波成分を含んでいることが確認できた。これらのことより、今回使用した学習型超解像では、折り返し歪みを基に高周波成分を推定して高解像度画像を作成しているのではなく、事前に学習したデータセットに基づいて品質が高くなるように高周波成分の推定および高解像度画像の作成をしていると考えられる。

また、全体的な品質では Bicubic 法による縮小過程を学習した超解像ネットワークを使用した際の方が良いという結果となった要因として、EDSR という超解像手法が Bicubic 法による縮小を学習することを前提にネットワークが最適化されていることに依存していると考えられる。そのため、この問題も超解像のネットワークのパラメータを実験条件に合わせて最適化することで解決できると考えられる。したがって、今回の結果では高周波画像を削減した高解像度画像と折り返し歪みを含まない縮小画像に対応する超解像を適用した時を比較すると後者の方が PSNR が低くなっているが、ネットワークのパラメータの最適化でこれは改善できると考えられる。

3.6 むすび

本章では、本研究で用いる超解像技術における問題点について 3 種類の予備実験を行い、問題の検証をした。

第4章 提案手法

4.1 まえがき

本章では、本研究における提案手法について述べる。

4.2 提案手法概要

本研究では、次世代映像のデータ量が膨大になるという問題を解決する手法として超解像技術を用いて解像度とビット深度を同時に拡張する手法を提案する。提案手法の概要図を図 4.2.1 に示す。

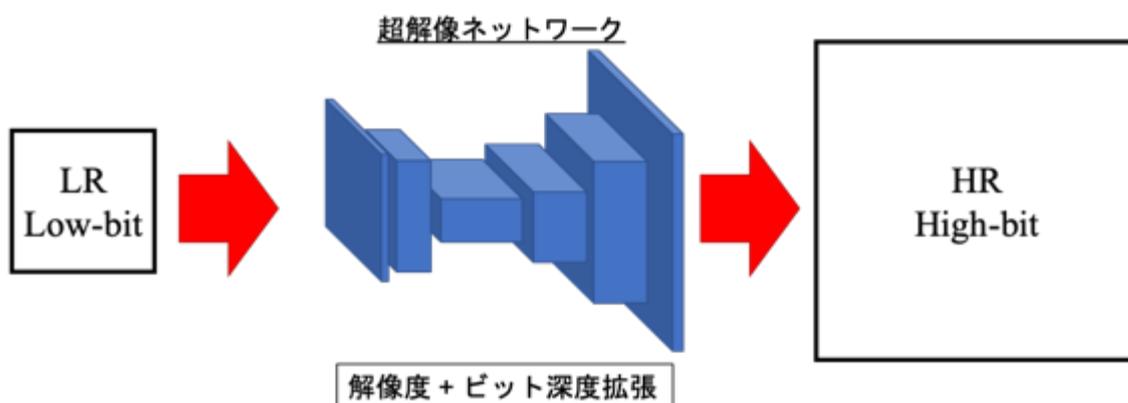


図 4.2.1 提案手法 概要図

一般的に超解像ネットワークは解像度のみを拡張する手法として利用されるが根本的には低解像度画像と高解像度画像のペアの2画像間の関係を学習することで解像度の拡張を実現している。そのため、学習するデータセットの画像ペアを低解像度-低ビットからなる画像と高解像度-高ビットからなる画像のペアに変更することで超解像のネットワークは解像度とビット深度の二つの特徴を学習でき、解像度とビット深度の同時拡張が実現できると考えられる。本研究では、この手法を用いることで超解像のネットワークを用いた解像度とビット深度の同時拡張を試みる。

4.3 むすび

本章では、本研究における提案手法について述べた。

第5章 評価実験

5.1 まえがき

本章では，提案手法の有効性を検証するための評価実験について説明する．

5.2 実験概要

評価実験では，低解像度-低ビットと高解像度-高ビットのペアからなるデータセットで学習した超解像を用いる．この超解像ネットワークに対し，低ビットの低解像度画像を入力し，出力された高ビットの高解像度画像から各指標を算出，既存手法との比較を行うことで提案手法の有効性を確認する．本来，次世代映像としてはビット深度が 10-bit または 12-bit で解像度が 4K または 8K からなる映像が想定される．しかしながら，多くの超解像のネットワークはビット深度が 8-bit を超える画像の入力に対応・最適化されておらず，この状況で 8-bit を超える画像での超解像を行えるようにネットワークを変更すると問題がより複雑になってしまい提案手法の評価の正当性が損なわれると考えられる．そこで本研究では，より正当に評価を行うために低ビットは 6-bit，高ビットは 8-bit とすることでより実験環境を単純にし，他の問題から切り離す．また実際に実験を行う際には，より実際に運用される環境に近づくため低ビット画像は 6-bit 画像を 2-bit シフトして得られる 8-bit 画像とする．評価実験の概要を図 5.2.1 に，低ビット画像の例を図 5.2.2 に，高ビット画像の例を図 5.2.3 にまたその比較を図 5.2.4 にそれぞれ示す．

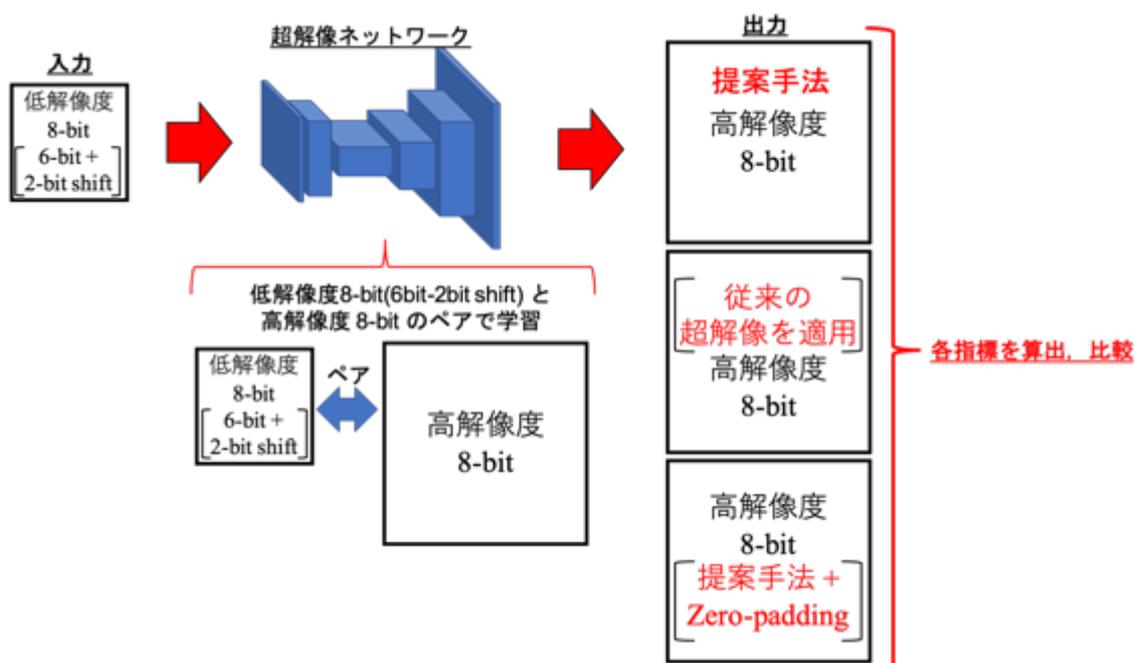


図 5.2.1 評価実験の概要図



図 5.2.2 低ビット画像の例(“DIV2K Dataset”[25][26])

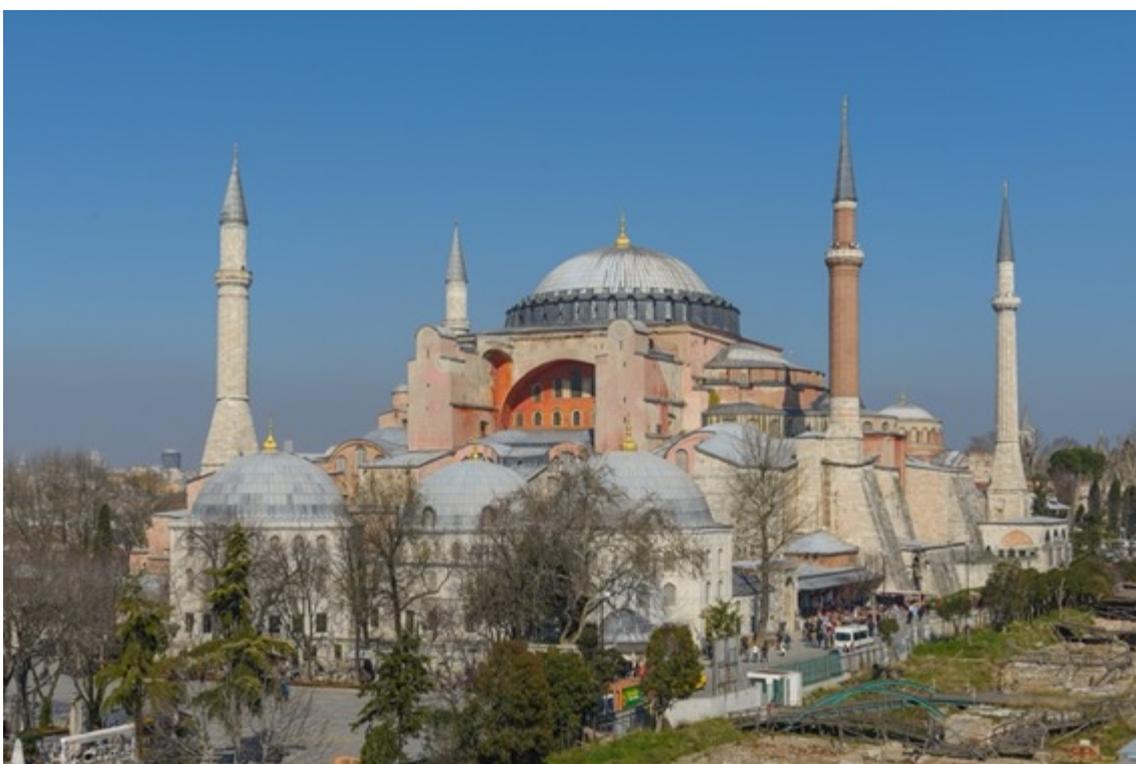


図 5.2.3 高ビット画像の例(“DIV2K Dataset”[25][26])

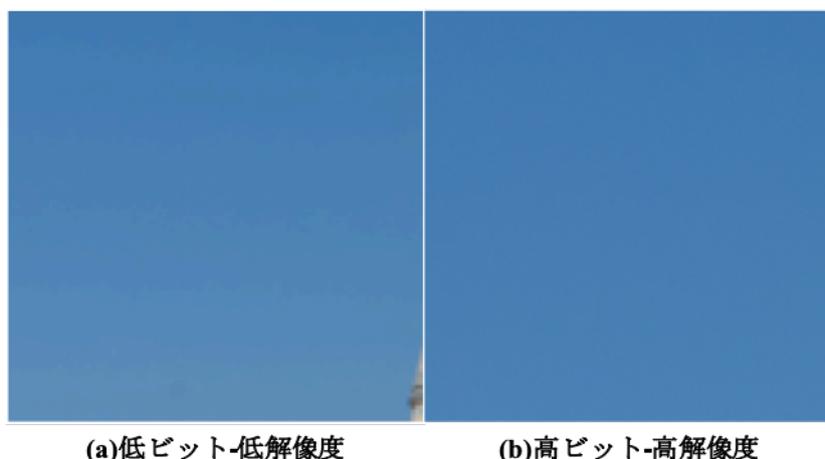


図 5.2.4 低ビット画像と高ビット画像の比較(“DIV2K Dataset”[25][26])

低ビット画像と高ビット画像を比較すると、低ビット画像は色の移り変わるような箇所で表現するビット数が足りずに段状のノイズが生じていることが確認できる。本実験では、超解像のネットワークとして、PSNR系のEDSRとGAN系のSRGANを使用する。また、提案手法より得られる出力画像と以下の二つの画像において各指標を算出して比較する。

- ・ビット深度学習をしていない超解像を用いて得られる画像
- ・提案手法の出力を 6-bit に変換し、Zero-padding により 8-bit にした画像(以降、提案手法+Zero-padding)

超解像のネットワークを学習するためのデータセットは予備実験と同様に DIV2K のうち 800 枚を使用し、テスト画像は DIV2K のうちの一部とその他の 4K データセットの一部からなる画像セットを使用する。実験で使用した PC の環境を表 5.2.1 に示す。

表 5.2.1 実験環境

OS	Ubuntu16.04
CPU	Intel(R) Core(TM) i7-8700K CPU @ 3.70GHz
GPU	NVIDIA GeForce GTX 1080 Ti x2

5.3 実験 1 結果

実験 1 では、超解像手法として SRGAN[16]を使用した。SRGAN の実験条件を表 5.3.1 に示す。

表 5.3.1 SRGAN の実験条件

使用言語		Python3.6
学習用データセット (DIV2K)	低解像度	0.5K, 6-bit-2bit shift, 800 枚
	高解像度	2K, 8-bit, 800 枚
テスト画像		0.5K, 6-bit-2bit shift, 100 枚
Epoch	Generator	200
	Discriminator	5000
拡大倍率		4 倍

次に入力したテスト画像を図 5.3.1 に、この画像の原画を図 5.3.2 にそれぞれ示す。また、提案手法+Zero-padding によって得られた画像を図 5.3.3 に、従来の超解像で得られた出力画像を図 5.3.4 に、提案手法の超解像によって得られた出力画像を図 5.3.5 にそれぞれ示す。またこれらの一部を切り取り拡大して比較したものを図 5.3.6 に示す。

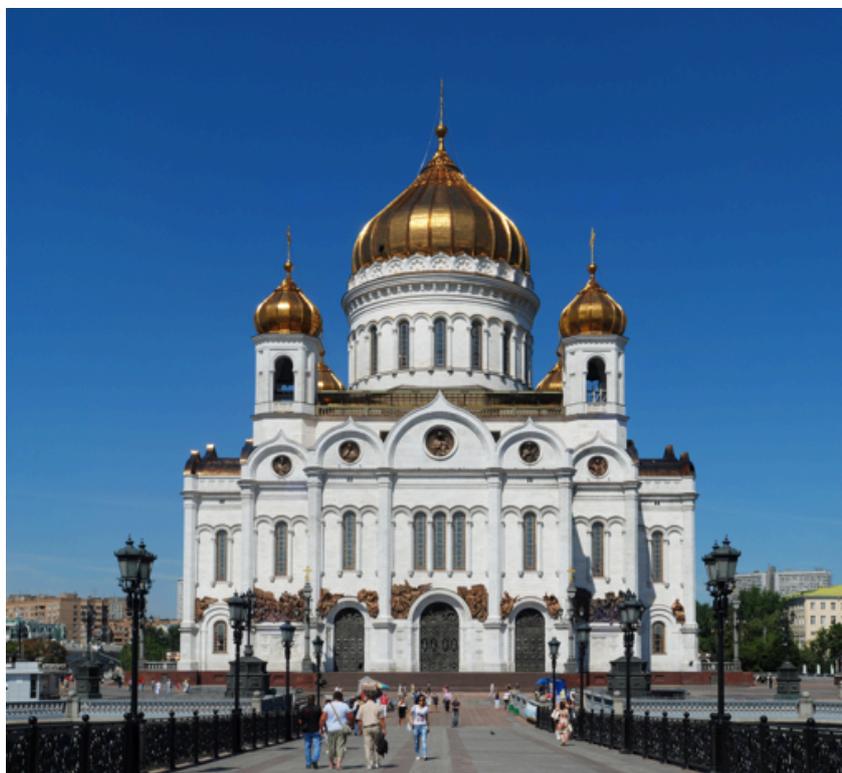


図 5.3.1 入力したテスト画像(“DIV2K Dataset”[25][26])

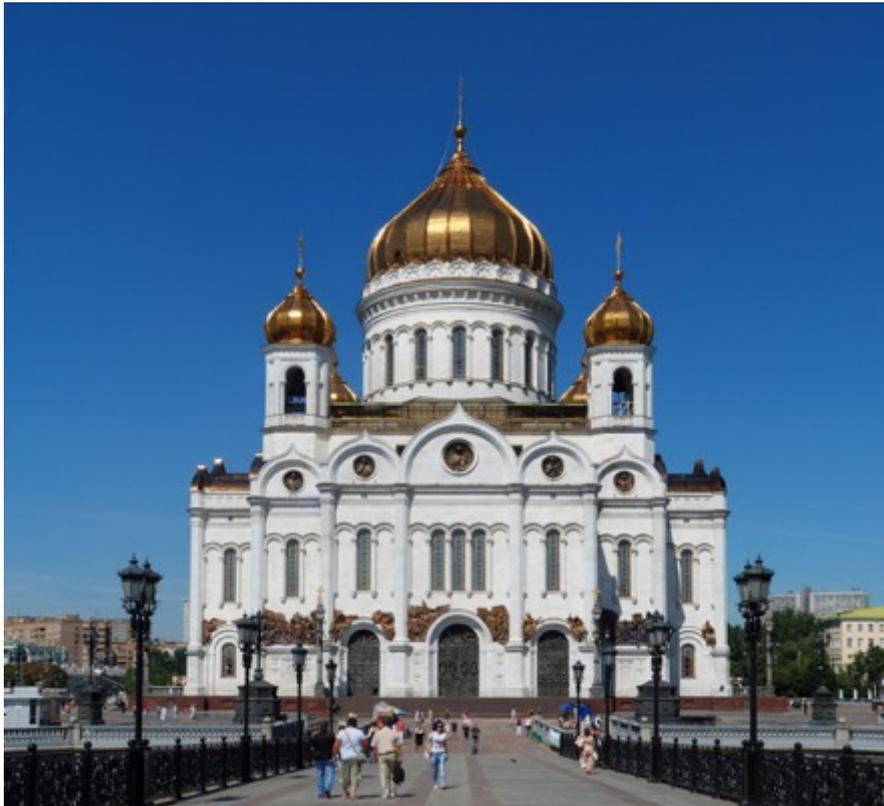


図 5.3.2 入力画像の原画(“DIV2K Dataset”[25][26])

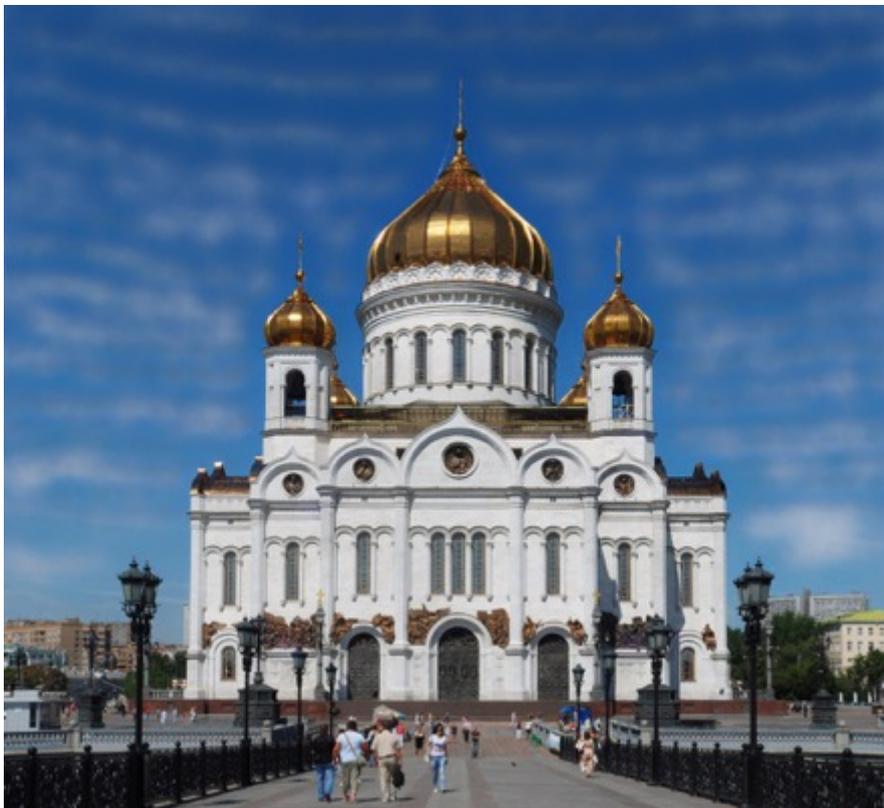


図 5.3.3 提案手法+Zero-padding によって得られた画像(“DIV2K Dataset”[25][26])

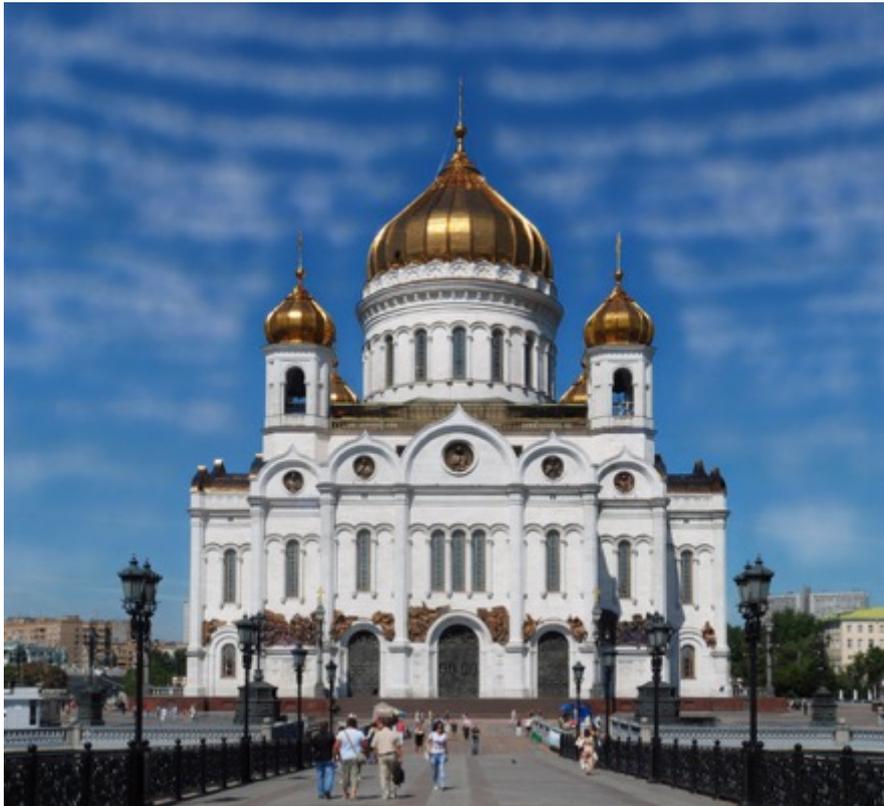


図 5.3.4 従来の超解像手法によって得られた出力画像(“DIV2K Dataset”[25][26])

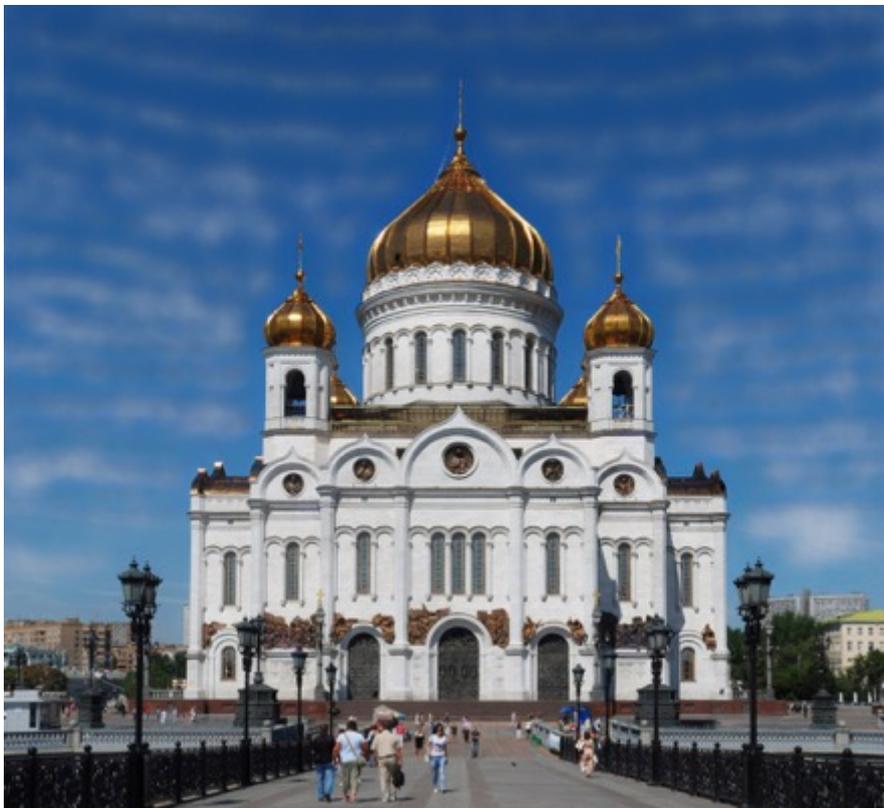


図 5.3.5 提案する超解像手法によって得られた出力画像(“DIV2K Dataset”[25][26])

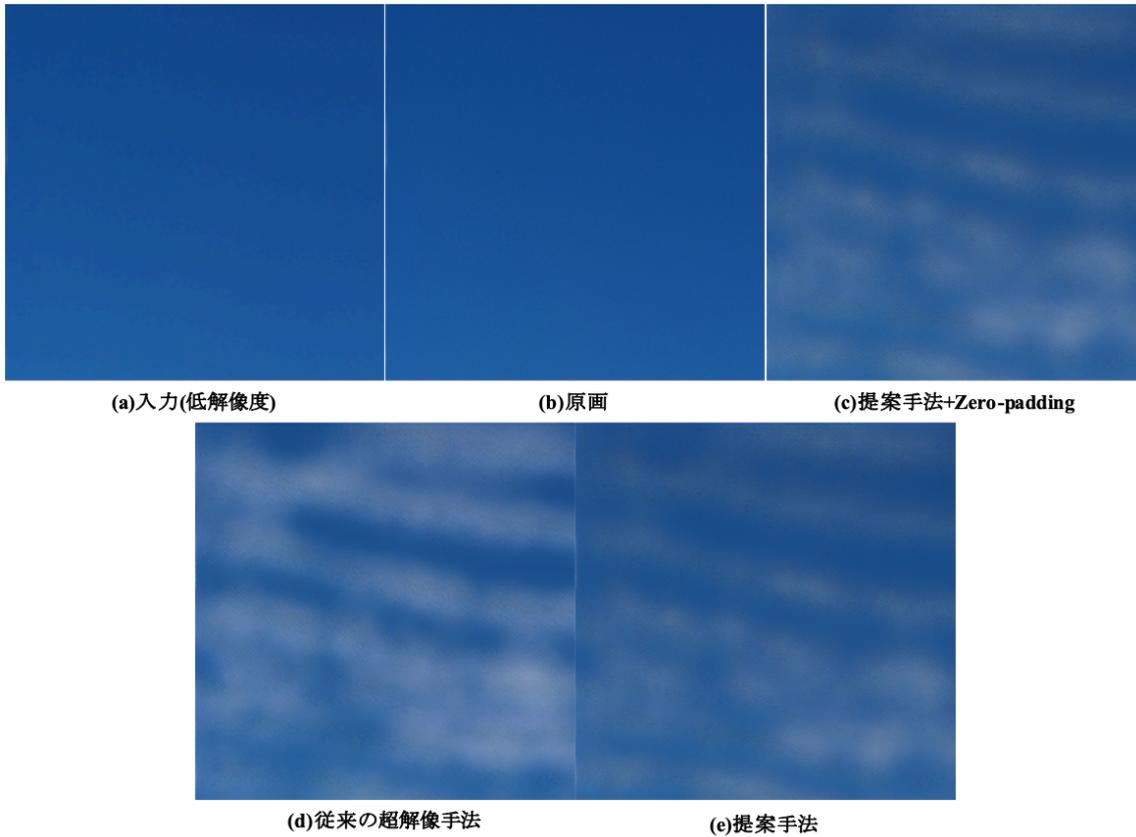


図 5.3.6 各画像の比較(“DIV2K Dataset”[25][26])

次に算出した各指標のテスト画像 100 枚での平均値を表 5.3.2 に示す.

表 5.3.2 実験 1 で算出した各指標

	PSNR dB	SSIM	BRISQUE
提案手法 + Zero-padding	24.20	0.8335	13.51
従来の超解像	24.45	0.8480	11.39
提案手法	24.24	0.8397	12.07

5.4 実験1 考察・結論

はじめに、各手法における各指標を比較すると、従来の超解像を適用して得られる出力画像が全ての指標において最も良い結果となっていることがわかる。また、出力された画像を見ると全ての超解像を適用した画像において原画には存在しないような雲のようなテクスチャが新たに生じていることが確認できる。これは、GAN系の超解像の特徴の一つで、入力された画像の中の情報と学習した情報を紐付けてテクスチャを推定し、画像を作り出しているからである。今回の場合は、入力画像のビット深度が足りないことによる段状のノイズをSRGANが空上に存在している何かしらのテクスチャであると判断し、学習データの中の空上のテクスチャ=雲と結びづけを行い、雲のテクスチャを新たに作り出していると考えられる。

提案手法+Zero-paddingの出力と提案手法の出力画像を比較すると、あまり大きな差は見受けられないが、提案手法の出力の方が若干ではあるが雲のノイズが少ないことが確認できる。また、PSNRやSSIMでもあまり差は生じていないが、BRISQUEでは、提案手法の品質が高いという値となっている。これは、前述した、雲のテクスチャが提案手法の方が薄いことに起因していると考えられる。次に、提案手法と従来手法の出力画像を比較すると、見た目では提案手法の方が雲のテクスチャが少ないことが確認できる。この原因として、提案手法ではビット深度も含めたデータセットで学習しているため、空のテクスチャと判断された入力画像のビット深度不足による段状ノイズの量が減ったからであると考えられる。しかしながら、各指標で比較すると、全ての指標において従来手法の方が高品質であると示している。これは、提案手法ではビット深度と解像度の学習を同時に行ったことにより、SRGANが推定する問題がより難化していることが原因として考えられる。これにより、学習の最適化ができず、超解像の精度が悪くなり出力画像の品質が低下していると考えられる。そのため、実験1において提案手法の精度を向上させるには、大規模なネットワークの変更や、パラメータの変更を加える必要があると推測できる。これらのことより、提案手法において直接GAN系超解像であるSRGANを用いることは適していないと考えられる。

5.5 実験 2 結果

実験 2 では，超解像手法として EDSR[13]を使用した．EDSR の実験条件を表 5.5.1 に示す．

表 5.5.1 EDSR の実験条件

使用言語		Python3.6
学習用データセット (DIV2K)	低解像度	0.5K, 6-bit-2bit shift, 800 枚
	高解像度	2K, 8-bit, 800 枚
テスト画像		0.5K, 6-bit-2bit shift, 100 枚
Epoch		300
拡大倍率		4 倍

次に入力したテスト画像を図 5.5.1 に，この画像の原画を図 5.5.2 にそれぞれ示す．また，提案手法+Zero-padding によって得られた画像を図 5.5.3 に，従来の超解像で得られた出力画像を図 5.5.4 に，提案手法の超解像によって得られた出力画像を図 5.5.5 にそれぞれ示す．またこれらの一部を切り取り拡大して比較したものを図 5.5.6 に示す．



図 5.5.1 入力したテスト画像(“DIV2K Dataset”[25][26])



図 5.5.2 入力画像の原画(“DIV2K Dataset”[25][26])



図 5.5.3 提案手法+Zero-padding によって得られた画像(“DIV2K Dataset”[25][26])



図 5.5.4 従来の超解像手法によって得られた出力画像(“DIV2K Dataset”[25][26])



図 5.5.5 提案する超解像手法によって得られた出力画像(“DIV2K Dataset”[25][26])



図 5.5.6 各画像の比較(“DIV2K Dataset”[25][26])

次に算出した各指標のテスト画像 100 枚での平均値を表 5.5.2 に示す.

表 5.5.2 実験 2 で算出した各指標

	PSNR dB	SSIM	BRISQUE
提案手法 + Zero-padding	29.59	0.9074	45.80
従来の超解像	29.61	0.9081	49.28
提案手法	29.76	0.9101	45.84

5.6 実験2 考察・結論

はじめに各手法における各指標を比較すると、PSNR および SSIM では提案手法が、BRISQUE においては 0.04pt 差で提案手法+Zero-padding が高品質を示す値となっている。次に、各出力画像を比較すると、提案手法+Zero-padding と従来手法では、入力画像に含まれるビット深度不足による段状のノイズがそのまま残っていることがわかる。一方で、提案手法では、この段状のノイズは完全に除去されており、ほぼ原画のような滑らかな色のグラデーションが表現されていることがわかる。このように見た目では、提案手法+Zero-padding より提案手法の方が優れていることが確認できるが、BRISQUE では提案手法+Zero-padding 低い値となっている。この原因として、グラデーション以外の細かいテクスチャを持つ領域で生じている超解像の補間誤差が Zero-padding において丸められ、BRISQUE が低い値になるように作用したことが考えられる。そこで、テスト画像 100 枚の平均値ではなく提案手法と提案手法+Zero-padding の各画像における BRISQUE の値を図 5.6.1 に示す。

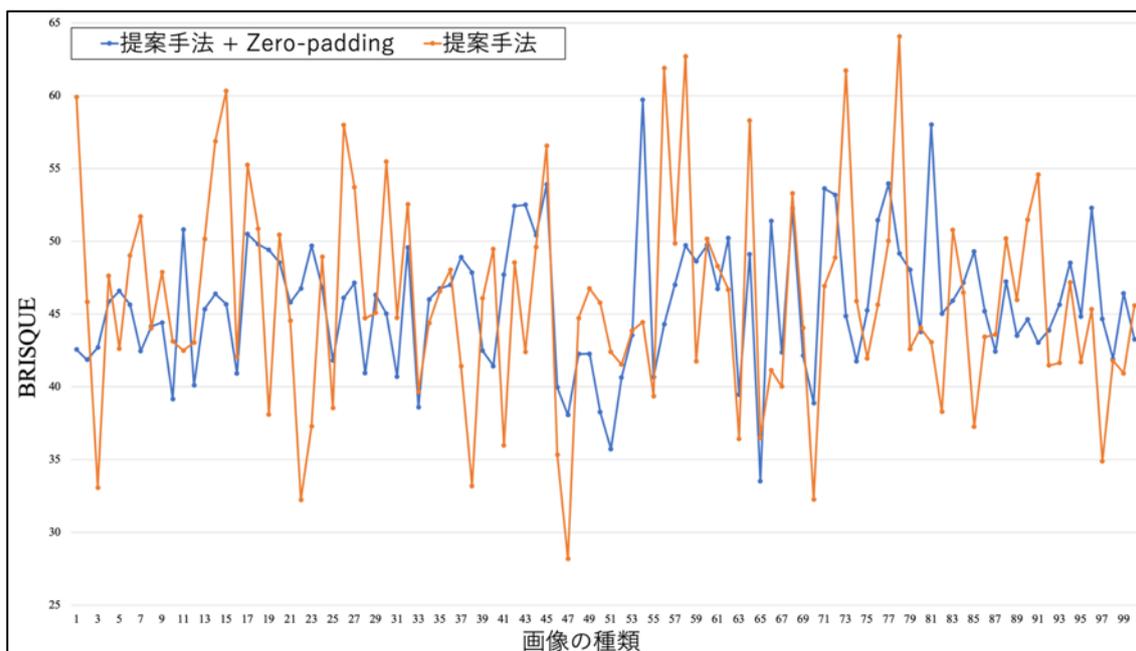


図 5.6.1 提案手法と提案手法+Zero-padding における各画像の BRISQUE

この図より、画像の種類によっては提案手法(図中オレンジ線)の方が優れた値(低い値)を取っていることがわかる。実際に、提案手法+Zero-padding が良い値となっている画像をみると、空や海などのビット深度不足の影響を受ける領域が少なく、森やビル街など細かいテクスチャを多く含む画像であることが確認できた。そのため、ビット深度をより正確に補うという点においては提案手法の方が提案手法+Zero-padding より優れていると考えられる。

これらのことより，提案手法において PSNR 系超解像を用いることは非常に適していることが実験により確認できた．また，提案手法による解像度とビット深度の同時拡張は PSNR 系超解像用いた場合に有効であると言えると考えられる．

5.7 むすび

本章では，提案手法の有効性を検討する評価実験を行い本研究における提案手法の有効性を確認した．

第6章 結論・今後の課題

6.1 結論

本研究では、解像度を拡張する技術である超解像を用いて、解像度とビット深度を同時に拡張する手法を提案した。また、提案手法は超解像手法として、GAN系の超解像であるSRGANを用いた際にはうまく適用できず悪い精度での超解像となり、PSNR系の超解像であるEDSRを用いた際には、視覚的にも指標的にも優れた出力を得ることができ、有効であることを評価実験により検証した。

6.2 今後の課題

今後の課題として、GAN系の超解像での適用を実現することと、8-bit以上の画像での適用を可能にするものの二つが挙げられる。近年、超解像の分野ではPSNR系の超解像では補間できる画素とその精度に限界があるとされ、GAN系超解像が注目を浴びている。GAN系の超解像では、PSNR系超解像とは異なり、細かいテクスチャを再現でき、視覚的に優れた画像を得ることができる。そのため、本研究においてもGAN系超解像において精度を向上させることは望ましいと考えられる。後者に関しては、次世代映像は8-bitを超える映像となっているのに対し、本研究では上限を8-bitとしている。そのため、実際の運用を考慮すると、8-bitの低解像度から10-bitあるいは12-bitの高解像度画像を生成する方式の方が望ましい。したがって、超解像手法を8-bit以上の画像に適用できるように調整したのちに、8-bitの低解像度と10-bitあるいは12-bitの高解像度画像のペアからなるデータセットを用いて学習をし、提案手法の有効性の検証を行う必要があると考えられる。

謝辞

本研究を行うにあたり，素晴らしい実験設備および研究環境を整えて下さり，研究内容に関して丁寧にご指導を頂きました渡辺裕教授に深謝いたします。

また，本研究のアイデアやアドバイス等，様々な面でご協力下さいました SHARP 株式会社の伊藤典男様，猪飼知宏様，中條健様，佐々木瑛一様に心より感謝いたします。

日頃から，研究について議論や相談，ご指摘等で協力していただいた研究室の皆様に厚く御礼を申し上げ，感謝の意を表します。

最後に，学費の工面や普段の生活でのサポートをしたくださった家族に対して感謝いたします。

参考文献

- [1] 総務省, "4K・8K ロードマップに関するフォローアップ会合 第二次中間報告」の公表, " 平成 27 年 7 月 30 日
- [2] 一般社団法人電子情報技術産業協会(JEITA), "2019 年民生用電子機器国内出荷統計, " <https://www.jeita.or.jp/japanese/stat/shipment/2019/index2.htm> (2019 年 12 月現在)
- [3] 総務省, "総務省トップ > 政策 > 情報通信(ICT 政策) > 放送政策の推進 > 4K 放送・8K 放送情報サイト, " http://www.soumu.go.jp/menu_seisaku/ictseisaku/housou_suishin/4k8k_suishin.html (2019 年 12 月現在)
- [4] MPEG The Moving Picture Experts Group, "High Efficiency Video Coding(HEVC), " <https://mpeg.chiariglione.org/standards/mpeg-h/high-efficiency-video-coding> (2019 年 12 月現在)
- [5] 齊藤知弘, "8K スーパーハイビジョンの伝送技術, " NHK 技研 R&D, No.152, p56-65, 2015.8.
- [6] 白井規之, " 地上放送高度化方式における信号帯域幅拡張に関する検討, " NHK 技研 R&D, No.172, p40-47, 2018.
- [7] 河村 圭, "次世代動画像符号化方式 VVC の標準化動向と深層学習の適用, " 情報処理学会研究報告オーディオビジュアル複合情報処理 (AVM), 2019-AVM-104, vol.5, p1-1, Feb. 2019.
- [8] Seiya Umeda, Niai Yano, Hiroshi Watanabe, Tomohiro Ikai, Takeshi Chujoh and Norio Ito, "Deep Learning Based Super-Resolution for Future Video Coding, " International Workshop on Advanced Image Technology (IWAIT2018), No.89, pp.1-4, Jan. 2018.
- [9] MathWorks, "ニューラルネットワーク, " <https://jp.mathworks.com/discovery/neural-network.html> (2019 年 12 月現在)
- [10] MathWorks, "畳み込みニューラルネットワークとは, " <https://jp.mathworks.com/discovery/convolutional-neural-network.html> (2019 年 12 月現在)
- [11] 合志清一, "再構成型超解像技術の限界, " 映像情報メディア学会誌, Vol.68, No.4, pp.J169-J173, 2014.
- [12] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang, "Learning a Deep Convolutional Network for Image Super-Resolution, " Proc. of European Conference on Computer Vision (ECCV), 2014.

- [13] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution, " *New Trends in Image Restoration and Enhancement workshop and challenge on image super-resolution 2017 (NTIRE2017 Super-Resolution Challenge)*, July.2017.
- [14] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, " *Computer Vision and Pattern Recognition 2016 (CVPR 2016)*.
- [15] Jamie Beckett, "Generative Adversarial Network とは——トップ研究者が解説, " June 2017, <https://blogs.nvidia.co.jp/2017/06/21/generative-adversarial-network/>
- [16] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network, " *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [17] R. A. Ulichney and S. Cheung, "Pixel bit-depth increase by bit replication, " in *Proc. Int. Soc. Optical Eng. (SPIE)*, 1998, pp. 232–241.
- [18] G. Mittal, V. Jakhetya, S. P. Jaiswal, O. C. Au, A. K. Tiwari, and D. Wei, "Bit-depth expansion using minimum risk based classification, " *Visual Communications and Image Processing (VCIP)*, 2012, pp. 1–5.
- [19] C.Cheng, O.C.Au, C.Liu,andK.Yip, "Bit-depth expansion by contour region reconstruction, " *The IEEE International Symposium on Circuits and Systems (ISCAS)*, 2009, pp. 944–947.
- [20] Jing Liu, Wannin Sun, Yuting Su, Peiguang Jing, Xiaokang Yang, "BE-CALF: Bit-Depth Enhancement by Concatenating All Level Features of DNN, " *IEEE Transactions on Image Processing*, Volume.28, Issue.10, Oct. 2019.
- [21] Anish Mittal, Rajiv Soundararajan, Alan C. Bovik, "Making a "Completely Blind" Image Quality Analyzer, " *IEEE Signal Processing Letters*, Volume. 20, Issue.3, March. 2013.
- [22] Anish Mittal, Anush K. Moorthy and Alan C. Bovik, "Blind/Referenceless Image Spatial Quality Evaluator, " *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, April. 2012.
- [23] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, Ming-Hsuan Yang, "Learning a No-Reference Quality Metric for Single-Image Super-Resolution, " *Computer Vision and Image Understanding*, Dec. 2016

- [24] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor, "The 2018 PIRM Challenge on Perceptual Image Super-resolution," Workshop and Challenge on Perceptual Image Restoration and Manipulation, 2018.
- [25] Eirikur Agustsson, Radu Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July. 2017.
- [26] Eirikur Agustsson, Radu Timofte, Van Gool, Luc and Yang, Ming-Hsuan and Zhang, Lei and Lim, Bee and others, "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July. 2017.

図一覧	
図 2.2.1 CNN の構造.....	3
図 2.3.2.3.1.1 SRCNN のネットワーク構造.....	5
図 2.3.2.3.1.2 各超解像手法の比較.....	5
図 2.3.2.3.2.1 EDSR のネットワーク構造.....	6
図 2.3.2.3.3.1 SRGAN のネットワーク構造.....	7
図 3.3.1.1 予備実験 1 概略図.....	11
図 3.3.2.1 予備実験 1 出力画像 1(入力:Bicubic 縮小画像).....	12
図 3.3.2.2 予備実験 1 出力画像 2(入力:Lanczos 縮小画像).....	12
図 3.4.1.1 予備実験 2 概要図.....	14
図 3.4.2.1 予備実験 2 出力画像の比較.....	15
図 3.5.1.1 予備実験 3 概略図.....	16
図 3.5.1.2 No-Aliasing images の作成手順.....	16
図 3.5.2.1 予備実験 3 出力画像の比較 1.....	18
図 3.5.2.2 予備実験 3 出力画像の比較 2.....	19
図 3.5.2.3 No-Aliasing image SR by Trained No-aliasing images SR の周波数スペクトル.....	20
図 3.5.2.4 No-High-frequency HR の周波数スペクトル.....	20
図 4.2.1 提案手法 概要図.....	22
図 5.2.1 評価実験の概要図.....	23
図 5.2.2 低ビット画像の例.....	24
図 5.2.3 高ビット画像の例.....	24
図 5.2.4 低ビット画像と高ビット画像の比較.....	25
図 5.3.1 入力したテスト画像.....	26
図 5.3.2 入力画像の原画.....	27
図 5.3.3 提案手法+Zero-padding によって得られた画像.....	27
図 5.3.4 従来の超解像手法によって得られた出力画像.....	28
図 5.3.5 提案する超解像手法によって得られた出力画像.....	28
図 5.3.6 各画像の比較.....	29
図 5.5.1 入力したテスト画像.....	31
図 5.5.2 入力画像の原画.....	32
図 5.5.3 提案手法+Zero-padding によって得られた画像.....	32
図 5.5.4 従来の超解像手法によって得られた出力画像.....	33
図 5.5.5 提案する超解像手法によって得られた出力画像.....	33
図 5.5.6 各画像の比較.....	34
図 5.6.1 提案手法と提案手法+Zero-padding における各画像の BRISQUE.....	35

表一覧

表 3.3.2.1 予備実験 1 の結果.....	13
表 3.4.2.1 予備実験 2 の結果.....	15
表 3.5.2.1 予備実験 3 結果 1.....	17
表 3.5.2.2 予備実験 3 結果 2.....	18
表 5.2.1 実験環境.....	25
表 5.3.1 SRGAN の実験条件.....	26
表 5.3.2 実験 1 で算出した各指標.....	29
表 5.5.1 EDSR の実験条件.....	31
表 5.5.2 実験 2 で算出した各指標.....	34

研究業績

国際学会

- [1] S. Umeda, H. Watanabe, T. Ikai, T. Hashimoto, T. Chujo, and N. Ito, “Joint Super-Resolution and Bit Depth Extension by DNN,” IWAIT-FMIA 2019, No.201 Jan. 2019.(査読あり)
- [2] Seiya Umeda, Niai Yano, Hiroshi Watanabe, Tomohiro Ikai, Takeshi Chujoh and Norio Ito, “Deep Learning Based Super-Resolution for Future Video Coding,” International Workshop on Advanced Image Technology (IWAIT2018), No.89, pp.1-4, Jan. 2018. (査読あり)

国内学会

- [1] 堀 隼也, 糞 子臣, 梅田聖也, 渡辺裕, 中條健, 佐々木瑛一, 猪飼 知宏, 伊藤典男, “縮小拡大と符号化によるノイズを含む画像の超解像に関する検討 (A Study on Super-resolution of Noisy Images by Scaling and Coding),” 2019 年画像符号化シンポジウム・2019 年映像メディア処理シンポジウム (PCSJ/IMPS2019), P-4-12, Nov. 2019.
- [2] 梅田, 渡辺, 猪飼, 橋本, 中條, 伊藤, “超解像によるビット深度拡張の特性 (Characteristics of Bit Depth Extension by Super Resolution),” 電子情報通信学会総合大会, D-12-7, Mar. 2019.
- [3] 梅田, 矢野, 渡辺, “符号化画像に対する超解像処理の特性改善について (On Quality Improvement of Super Resolution Processing for Coded Images),” 2018 年映像情報メディア学会冬季大会, 14D-4, Dec. 2018.
- [4] 梅田, 矢野, 渡辺, “映像符号化における超解像処理画像の画質改善 (Quality improvement of Super-Resolution Image for Video Coding),” 2018 年画像符号化シンポジウム・2018 年映像メディア処理シンポジウム (PCSJ/IMPS2018), P-1-8, Nov. 2018.
- [5] 梅田, 矢野, 渡辺, 猪飼, 中條, 伊藤, “FVC 量子化パラメータ制御による超解像の特性 (Quantize parameter dependent super resolution characteristics in Future Video Coding),” 情報処理学会全国大会, 2U-03, Mar. 2018.
- [6] 矢野, 梅田, 渡辺, 猪飼, 中條, 伊藤, “方向性を考慮した超解像処理技術による FVC 符号化画像の画質改善 (Quality Improvement of FVC Encoded Image Using Super-Resolution Considering Feature Directionality),” 情報処理学会全国大会, 2U-02, Mar. 2018.
- [7] 矢野, 梅田, 渡辺, 猪飼, 中條, 伊藤, “画像の方向性を考慮した超解像処理技術 (Super-Resolution Technology Considering Directionality of Image),” 2017 年映像情報メディア学会冬季大会, 14B-2, Dec. 2017.