

修士論文概要書

Master's Thesis Summary

Date of submission: 01/23/2023 (MM/DD/YYYY)

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	矢野 礼美菜	指導 教員 Advisor	渡辺 裕 印 Seal
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	5121F089-2		
研究題目 Title	符号化映像のブロック分割を用いた超解像ネットワークの研究 Research on Super-Resolution Network Using Partitioned Block of Coded Video				

1. まえがき

近年、新 4K8K 衛星放送が開始し、動画配信でも 4K 映像が見られるなど、高解像度な映像に対する需要が高まっている。しかし、実際に利用される通信を介した映像では伝送効率のための圧縮符号化による符号化劣化が発生する。本研究では、Versatile Video Coding (VVC) によって符号化及び縮小された映像に対して、超解像を用い、高解像度及び高品質化を目指す。

2. 関連技術

2.1. Versatile Video Coding (VVC)

VVC は ISO/IEC と ITU-T の共同チームである Joint Video Experts Team (JVET) によって、2020 年に発表された HEVC の後継の動画圧縮規格である。VVC に新たに追加された機能の一つに Reference Picture Resampling (RPR) がある。RPR とは、シーケンス内でフレーム毎に符号化解像度を変更する機能である。RPR を利用することで、異なる解像度間でのフレーム参照が可能となる。また、VVC では HEVC と同様に Coding Tree Unit (CTU) と呼ばれるブロック単位で符号化処理を行う。実際の符号化処理は、CTU を画像特性に合わせて再帰的に分割した Transform Unit (TU) 単位で行われる。

2.2. SRResNet[1]

SRResNet は SRGAN[1] の論文内で提案された、ResNet[3] を参考とした超解像ネットワークである。図 1 のように、SRResNet はいくつかの Residual Block を連ねたものである。本研究では、Residual Block を 16 個連ねている。Residual block では Skip connection を採用し、Residual (残差) 学習をすることで、勾配消失及び勾配爆発を抑制しながら安定した学習を期待できる。

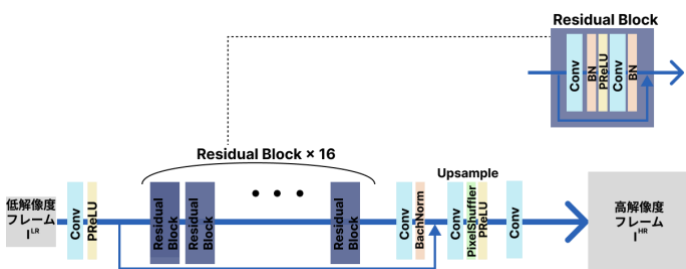


図 1 SRResNet

3. 提案手法

3.1. TU 画像

2.1 項で述べた VVC の TU ブロック分割を用いて、ブロックサイズが小さいときは大きな値を、大きいときには小さい値を与えることでグレースケール画像を作成した。このグレースケール画像を TU 画像と称し、その例を図 2 に示す。

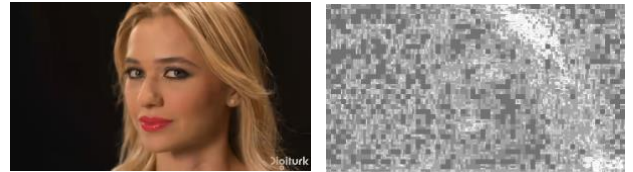


図 2 符号化フレーム (UVG dataset [2]) (左) とその TU 画像 (右) の例 (2023 年電子情報通信学会総合大会において発表予定)

3.2. 超解像ネットワーク

本論文で提案する超解像ネットワークを図 3 に示す。本ネットワークは 2.2 項で説明した SRResNet を基盤として用い、TU 画像と符号化フレームをチャンネル方向に結合している。損失関数は、SRResNet と同様に MSE Loss を用いている。

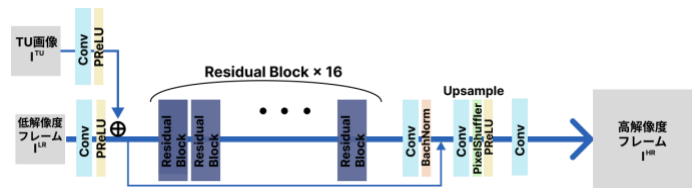


図 3 提案手法

3.3. モデル作成条件

提案手法を用いて 4 つのモデルを作成した。学習データには、VVC 符号化における I ピクチャのみのフレーム (io : intra-only) と I ピクチャ含む全てのフレーム (ra : random-access) の 2 種類を用いた。ra では、SJTU dataset [2] と UVG dataset [3] を用い、io では、SJTU dataset と UVG dataset に加え、BVI-DVC dataset [4] を用いた。4 つのモデルの作成条件を表 1 に示す。

表 1 モデル作成条件

モデル	ours1	ours2	ours3	ours4
QP	22	22	32	32
ra/io	ra	io	ra	io

3840x2160 のシーケンスである。

4. 実験

4.1. 実験内容

テストデータは、JVET Common Test Conditions (CTC) for SDR. [5]のデータ (CTC data)のうち、原画像の解像度が 3840x2160 と 832x480 のデータを用いた。QP 値を 22, 32, 42 として VVC 符号化及び縮小したテストデータとその TU 画像を低解像度の入力データとし、図 3 に示す提案手法を用いて拡大した超解像フレームを出力として得た。得られた出力画像を、原画像との PSNR を算出することで、結果を比較した。また、既存手法である Bicubic 補間と SRResNet においても同様に PSNR を算出し、その結果を比較した。

4.2. 実験結果及び考察

表 2 及び表 3 に提案手法を適用した超解像画像の PSNR 結果を示す。

表 2 から、原画像の解像度が 832x480 のテストデータにおいて、QP 値が 22 及び 32 で符号化されたデータでは、ours2 の精度が良いことがわかる。ただし、QP 値が 42 で符号化されたデータでは SRResNet の精度が良くなっている。QP 値が 22 で符号化されたデータに比べ、QP 値が 42 で符号化されたデータは TU ブロック分割が荒いため、QP 値が 42 で符号化されたテストフレームでは、その TU 画像が提案する超解像ネットワークにとって不要な情報として見られているのではないかと考えられる。

表 3 から、原画像の解像度が 3840x2160 のテストデータにおいて、ours1 の精度が良いことがわかる。学習データとテストデータの原画解像度がほぼ一致しており、テストデータも ra であることからデータ性質の近い ours2 の精度が良くなったと考えられる。

表 2 実験結果に対する PSNR[dB] (832x480)

QP	Bicubic	SRResNet	ours1	ours2	ours3	ours4
22	25.1969	25.7304	25.5481	26.0899	25.7382	25.8167
32	24.5273	24.9815	24.8825	25.1040	25.0055	24.9452
42	22.8945	23.3745	23.3548	23.0841	23.3196	23.0323

表 3 実験結果に対する PSNR[dB] (3840x2160)

QP	Bicubic	SRResNet	ours1	ours2	ours3	ours4
22	31.9636	32.9097	32.9979	32.2731	32.8385	32.2353
32	30.5548	31.8201	31.9163	31.0703	31.8422	31.0532
42	27.9683	29.3268	29.3729	28.8555	29.3619	28.9086

4.3. 実験結果画像

図 4 及び図 5 に QP 値が 22 で符号化したテストデータの結果画像の一例を示す。BasketballDrillText, Campfire はそれぞれ原画の解像度が 832x480,



図 4 BasketballDrillText (QP22) の実験結果

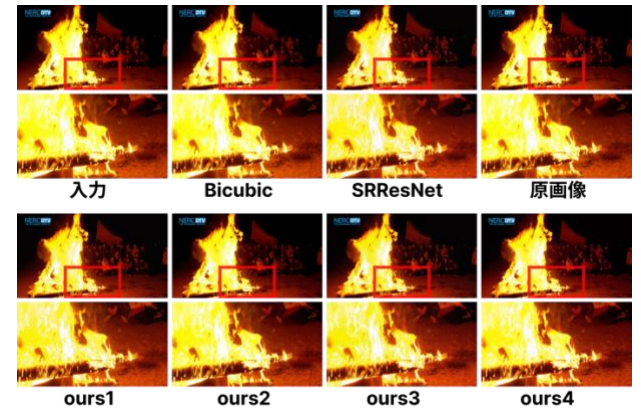


図 4 Campfire (QP22) の実験結果

5. むすび

本研究では、超解像ネットワークの一つである SRResNet を改良し、VVC 符号化映像の高解像度化及び高品質化の精度を向上させる提案を行った。

Bicubic 補間や SRResNet, その他の超解像手法においては、符号化の特性や条件を考慮していなかった。そこで、符号化に関する情報を TU 画像として表し、超解像ネットワークに組み込むことで、より高品質な高解像度映像を再構成できることを確認した。

参考文献

- [1] C. Ledig, et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network" Computer Vision and Pattern Recognition (CVPR), July 2017.
- [2] A. Mercat, et al. "UVG dataset: 50/120fps 4K sequences for video codec analysis and development" in Proc. ACM Multimedia Syst. Conf., Istanbul, Turkey, June 2020.
- [3] L. Song, et al. "The SJTU 4K Video Sequence Dataset" the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), Klagenfurt, Austria, July 3rd-5th, 2013.
- [4] D. Ma, et al. "BVI-DVC: a training database for deep video compression" IEEE Transactions on Multimedia, 2021.
- [5] F. Bossen, et al. "JVET common test conditions and software reference configurations for SDR video" Joint Video Experts Team Document, JVET-N1010, Geneva, 2019.

2022 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

符号化映像のブロック分割を用いた
超解像ネットワークの研究

Research on Super-Resolution Network
Using Partitioned Block of Coded Video

矢野 礼美菜
(5121F089-2)

提出日：2023.01.23

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	はじめに.....	1
1.2	研究目的.....	1
1.3	本論文の構成.....	2
第2章	次世代映像符号化技術.....	3
2.1	はじめに.....	3
2.2	H.264/MPEG-4 AVC.....	3
2.3	High Efficiency Video Coding (HEVC).....	3
2.4	Versatile Video Coding (VVC).....	3
2.4.1	技術概要.....	3
2.4.2	Reference Picture Resampling (RPR).....	4
2.4.3	Transform Unit (TU).....	4
2.5	むすび.....	4
第3章	超解像技術.....	5
3.1	まえがき.....	5
3.2	超解像技術の概要.....	5
3.3	SRResNet.....	5
3.3.1	Pixel Shuffler.....	6
3.3.2	損失関数.....	6
3.4	むすび.....	6
第4章	提案手法.....	7
4.1	まえがき.....	7
4.2	TU 画像.....	7
4.3	超解像ネットワーク.....	8
4.4	作成モデルおよび学習条件.....	8
4.5	むすび.....	9
第5章	実験内容及び結果考察.....	10
5.1	まえがき.....	10
5.2	実験内容.....	10

5.3	実験結果及び考察.....	10
5.2.1	実験結果画像.....	13
第6章	結論.....	17
6.1	結論.....	17
6.2	今後の展望.....	17
	謝辞.....	18
	参考文献.....	19
	図一覧.....	20
	表一覧.....	21
	研究業績.....	22

第1章 序論

1.1 はじめに

2023年現在，すでに4K(3840x2160画素)や8K(7680x4320画素)映像による新4K8K衛星放送が運用されており，これらの高解像度映像を出力するためのデバイスに対する需要が大きくなっている．出力機器の中でも特に4K対応テレビの国内販売台数は累計1,000万台を超える[1]．また，各種動画ストリーミングサービスや家庭用ゲーム機，グラフィックボードを用いるゲームなどでも，4Kや8Kといった解像度への対応が進んでおり，高解像度映像はより身近なものとなっている．

新4K8K衛星放送の運用や高解像度映像の利用における問題点の一つとして，伝送効率のための圧縮符号化による符号化劣化が挙げられる．特に，4Kや8Kといった高解像度映像では，Full High Definition(フルHD:1920x1080画素)の映像と比べてそれぞれ4倍，16倍の画素数があり，膨大なデータ量が必要となる．そのため，限られた通信速度で膨大なデータを送るために，大きなデータ圧縮が求められる．現在幅広く利用されている映像圧縮符号化に関する規格にはH.264/MPEG-4 AVCがあるが，より効率と品質の良い次世代映像符号化技術の研究開発がJoint Video Experts Team(JVET)とInternational Telecommunication Union Telecommunication Standardization Sector(ITU-T)により行われている．

一方，画像処理分野に限らず近年様々な分野において，ヒトの神経を模倣したニューロンを利用したニューラルネットワークによって，高精度な問題解決が行われている．問題解決方法の一つである超解像技術は，入力の低解像度画像を，画像の高さと幅を2倍や4倍などに拡大することで，高解像度画像を出力する技術である．空間画素数が4倍や16倍などに増えることで，入力の低解像度画像では存在しなかった画素値を補間する必要があるが，ニューラルネットワークの利用によって，従来よりも高精細な画素値補間を可能としている．

1.2 研究目的

従来の超解像技術の研究では，低解像度の画像・映像を対象にすることが多いが，符号化後の画像に対して，超解像技術によって高解像度かつ高品質の映像を得られれば，より圧縮率の高い符号化方式を利用することが可能である．本研究ではJVETが発表した次世代映像符号化方式であるVersatile Video Coding(VVC)で符号化および縮小された映像に対して，超解像技術を適用し，その品質評価を行うことで，超解像技術の有効性を示す．また，既存の超解像ネットワークに，VVCの符号化時に得られるビットストリーム内の情報を利用することで，VVC符号化の特性を捉えた超解像ネットワーク

を提案し，その有効性を示す．

1.3 本論文の構成

以下に本論文の構成を示す．

- 第1章 本研究の背景，およびその目的について述べる．まず研究背景について記述した後，本研究の目的について述べる．
- 第2章 本研究において採用している規格である VVC を含めた次世代映像符号化技術や，その基盤となる規格について述べる．
- 第3章 本研究で用いた超解像技術と実際に基盤として用いた SRResNet の概要について述べる．
- 第4章 本研究の提案手法について述べる．
- 第5章 本研究の実験内容及びその結果と考察を述べる．
- 第6章 本研究のまとめおよび今後の課題を述べる．

第 2 章 次世代映像符号化技術

2.1 はじめに

本章では、次世代映像符号化技術について述べる。まず次世代映像符号化の基盤となっている符号化方式である H.264/MPEG-4 AVC, High Efficiency Video Coding (HEVC) について述べた後、本研究で採用している VVC について述べる。

2.2 H.264/MPEG-4 AVC

H.264 と MPEG-4 Part 10 Advanced Video Coding (MPEG-4 AVG) は、それぞれ ITU-T と国際標準化機構 (ISO) / 国際電気標準会議 (IEC) によって策定された動画圧縮規格の一つであり、両者は技術的には同一であるため、H.264/MPEG-4 AVC あるいは MPEG-4 AVC/H.264 と記述されることが多い。ファイルの拡張子は「.mp4」で表される。H.264 は従来規格である MPEG-2 と比べて、2 倍以上の圧縮率となっており、現在最も多く利用されている動画圧縮規格の一つである。

2.3 High Efficiency Video Coding (HEVC)

HEVC は H.264/MPEG-4 AVC の後継の動画圧縮規格の一つであり、ISO/IEC と ITU-T が共同で開発し 2013 年に ITU-T によって承認された。HEVC は H.264/MPEG-4 AVC に比べて約 2 倍の圧縮率でありながら、同程度あるいはそれ以上の品質を保つことが特徴である。また、H.264 では 4K までしかサポートされていなかったが、HEVC では 8K までサポートしていることも一つの利点である。ファイルの拡張子は「.mov」で、正しくファイルを表示させるためには、互換性のあるハードウェアが必要である。最新のスマートフォンに搭載されているビデオレコーディング機能や、動画ストリーミングサービスなどではすでに HEVC が採用されはじめている。

2.4 Versatile Video Coding (VVC)

2.4.1 技術概要

VVC は ISO/IEC と ITU-T の共同チームである Joint Video Experts Team (JVET) によって、2020 年に発表された HEVC の後継の動画圧縮規格である。VVC を利用することで、HEVC と同品質の映像を提供するために必要なビットレートを約 40%削減可能である。また、360° 映像や VR 映像に対応する機能も追加されるなど、幅広い用途での使用が期待される。VVC で符号化された 8K 映像をリアルタイムで復号するソフトウェアも開発されており、今後主流になる可能性のある規格だと言える。

2.4.2 Reference Picture Resampling (RPR)

VVC に新たに追加された機能の一つに Reference Picture Resampling (RPR) がある。RPR とは、シーケンス内でフレーム毎に符号化解像度を変更する機能である。そのため図 2.1 のように、異なる解像度間でのフレーム参照が可能となる。また、映像の解像度をシーケンス内で適宜に変化させることができるため、ストリーミングなどの帯域変動が大きい状況でも比較的安定した画像品質を保つことが可能となる。

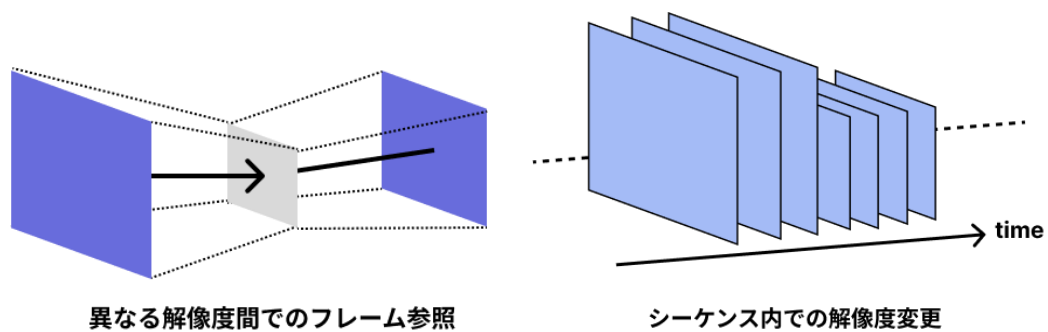


図 2.1 Reference Picture Resampling (RPR)

2.4.3 Transform Unit (TU)

VVC では HEVC と同様に Coding Tree Unit (CTU) と呼ばれるブロック単位で符号化処理を行う。CTU のブロックサイズは HEVC では最大 64×64 であったが、VVC では 128×128 となっている。実際の符号化処理は、CTU を画像特性に合わせて再帰的に分割した Transform Unit (TU) 単位で行われる。TU の最小ブロックサイズは 1×16 , 16×1 , 2×8 , 8×2 , 4×4 のいずれかであり、最大ブロックサイズは 64×64 である。ただしスキップ処理が行われる場合や、特定の Inter においては変換係数が発生せずに最大ブロックサイズが 128×128 として出力されることがある。

2.5 むすび

本章では次世代映像符号化技術の概要と、その用途に関して述べた。また、VVC の機能の一つである RPR に関して述べた。本論文では 2.4 節で述べた VVC を適用した映像に対して、超解像技術を適用する。

第3章 超解像技術

3.1 まえがき

本章では超解像技術の概要について述べた後、具体的な超解像ネットワークの構成に関して述べる。

3.2 超解像技術の概要

超解像技術とは、低解像度の画像あるいは映像を高解像度に変換する技術である。解像度を上げる技術として Bilinear 補間や Bicubic 補間があるが、こうした補間ベースの手法では、低解像度の画像に含まれる画素を利用するため劣化が起きてしまう。超解像技術では、低解像度の画像には存在しない高周波成分を推定するため、補間ベースの手法よりも品質の良い高解像度の画像の生成が期待できる。

3.3 SRResNet

SRResNet は SRGAN[2]の論文内で提案された、ResNet[3]を参考とした超解像ネットワークである。図 3.1 に SRResNet の概要を示す。図 3.1 のように、SRResNet はいくつかの Residual Block を連ねたものである。本研究では、Residual Block を 16 個連ねている。Residual block では Skip connection を採用し、Residual(残差) 学習をすることで、勾配消失及び勾配爆発を抑制しながら安定した学習を期待できる。また、Residual Block に続いて Pixel Shuffler を用いて特徴マップを空間方向に拡大している。

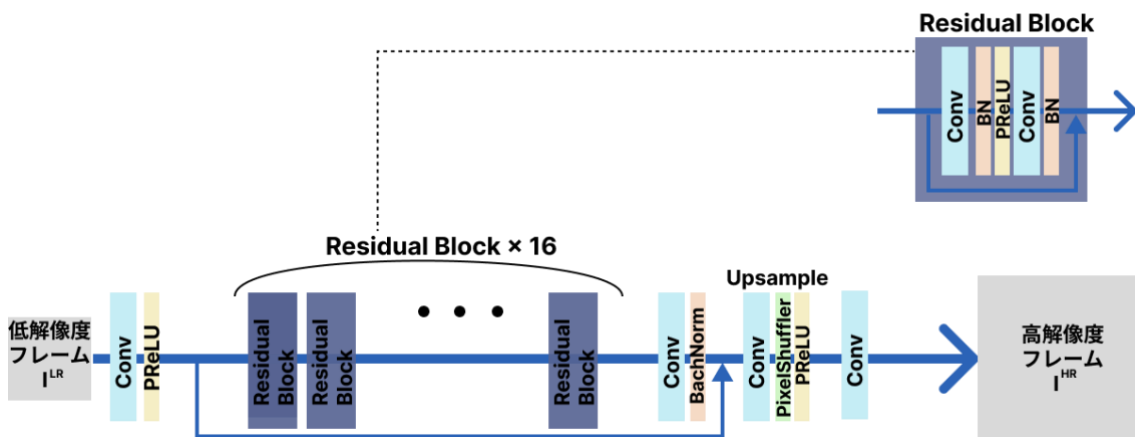


図 3.1 SRResNet

3.3.1 Pixel Shuffler

Pixel Shuffler は、チャンネル方向に連なった特徴マップの各値を、空間方向に並び替えることで、特徴マップを拡大する。図 3.2 のように 4 ($=2^2$) チャンネルの特徴マップを Pixel Shuffler に通すことで、縦横それぞれ 2 倍に拡大された特徴マップを取得することができる。

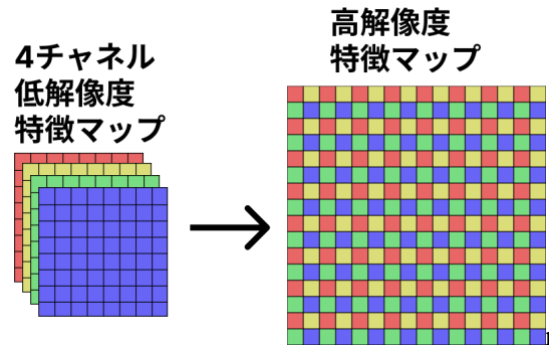


図 3.2 Pixel Shuffler

3.3.2 損失関数

SRResNet の損失関数は、下式 (3.1) に示す MSE Loss を用いている。ただし I^{LR} , I^{HR} はそれぞれ高解像度化対象の符号化フレーム、原画像の画素値であり W , H はそれぞれ高解像度画像の幅と高さである。

$$\text{MSE Loss} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H (I_{x,y}^{HR} - f_{SR}(I_{x,y}^{LR}))^2 \quad (3.1)$$

3.4 むすび

本章では、超解像技術の概要と分類および、SRResNet の概要に関して述べた。

第4章 提案手法

4.1 まえがき

本章では，VVC 符号化のビットストリームから得られる TU 画像の概要と，その TU 画像を組み込んだ超解像ネットワークについて提案する．

4.2 TU 画像

2.4.3 項で述べた TU ブロック分割を用いて，ブロックサイズが小さいときは大きな値を，大きいときには小さい値を与えることでグレースケール画像を作成した．実際に与えた値は表 3.1 のとおりである．このグレースケール画像を TU 画像と称し，その例を図 4.1 に示す．

表 4.1 TU 分割ブロックサイズと割当画素値

TU ブロックサイズ [pix ²]	画素値
16	255
32	242
48	229
64	216
96	204
128	191
192	178
256	165
384	153
512	140
768	127
1024	114
1536	102
2048	89
3072	76
4096	63
8192	51
16384	38



図 4.1 符号化フレーム (UVG dataset[5]) (左) とその TU 画像 (右) の例 (2023 年電子情報通信学会総合大会において発表予定)

4.3 超解像ネットワーク

本論文で提案する超解像ネットワークを以下の図 4.2 に示す. 図 4.2 を見るとわかるように, 本ネットワークは 3.3 項で説明した SRResNet を基盤として用い, TU 画像と符号化フレームをチャンネル方向に結合している. 損失関数は, SRResNet と同様に式 (4.1) に示す MSE Loss を用いている. ただし, I^{TU} は TU 画像の画素値, \oplus はチャンネル方向の結合を示す.

$$\text{MSE Loss} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \left(I_{x,y}^{HR} - f_{SR}(I_{x,y}^{TU} \oplus I_{x,y}^{LR}) \right)^2 \quad (4.1)$$

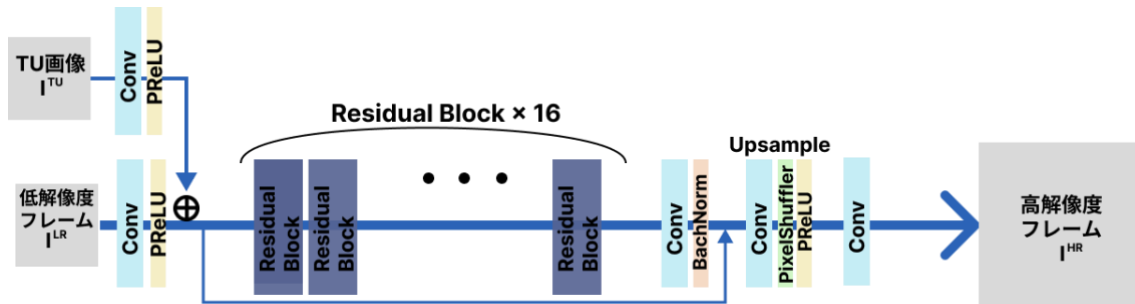


図 4.2 提案する超解像ネットワーク

4.4 作成モデルおよび学習条件

本研究では, 4.3 項で記述した超解像ネットワークを用いてモデルを作成した. 学習データには, VVC 符号化における I ピクチャのみのフレーム (io : intra-only) と I ピクチャ含む全てのフレーム (ra : random-access) の 2 種類を用いた. ra では, SJTU dataset[4] と UVG dataset[5] を用い, io では, SJTU dataset と UVG dataset に加え, BVI-DVC dataset[6] を用いた. ra と io の詳細を表 4.1 に示す. また, ra と io それぞれにおいて, QP 値の異

なる符号化データを用いて，計4つのモデルを作成した．表4.2に各モデルの作成条件を示す．また，本研究で用いるVVC符号化データは全て2.4.2項で示したRPRの機能を用いて縮小されたデータである．

表 4.2 ra と io のデータ詳細

	ra	io
データセット	UVG SJTU	UVG SJTU BVI-DVC
シーケンス数	15 (UVG) 14 (SJTU)	15 (UVG) 14 (SJTU) 200 (BVI-DVC)
フレーム数	960 (UVG) 896 (SJTU)	120 (UVG) 112 (SJTU) 1600 (BVI-DVC)

表 4.3 モデル作成時の学習条件

モデル	ours1	ours2	ours3	ours4
QP	22	22	32	32
ra/io	ra	io	ra	io

4.5 むすび

本章では，VVC符号化時のTU分割ブロックをもとに作成したTU画像の作成方法及び，提案する超解像ネットワークについて述べた．

第 5 章 実験内容及び結果考察

5.1 まえがき

本章では，作成した各モデルに対して，JVET Common Test Conditions (CTC) for SDR. [7]のデータ (CTC data) を，VVC によって符号化および縮小したデータを入力としたときの実験結果について PSNR とともに示し，その考察について記述する．

5.2 実験内容

実験において用いた学習データセットは 4.4 項で記述した通りである．テストデータセットは，CTC data のうち，原画像の解像度が 3840x2160 と 832x480 のデータを用いた．QP 値を 22, 32, 42 として VVC 符号化及び縮小したテストデータとその TU 画像を低解像度の入力データとし，図 4.2 に示す提案手法を用いて拡大した超解像フレームを出力として得た．得られた出力画像を，原画像との PSNR を算出することで，結果を比較した．また，既存手法である Bicubic 補間と SRResNet においても同様に PSNR を算出し，その結果を比較した．

5.3 実験結果及び考察

表 5.1 及び 5.2 に提案手法を適用した超解像画像の PSNR 結果を示す．

表 5.1 から，原画像の解像度が 832x480 のテストデータにおいて，QP 値が 22 及び 32 で符号化されたデータでは，ours2 の精度が良いことがわかる．ours2 と ours4 の結果を比べることで，QP 値が 32 で符号化された学習データを用いるよりも QP 値が 22 で符号化された学習データを用いる方が，優れたモデルを作成することができると言える．また，ours1 と ours2 の結果を比べることで，ra のデータを用いるよりも io のデータのみを用いて学習したモデルの性能が良いことがわかる．ただし，QP 値が 42 で符号化されたデータでは SRResNet の精度が良くなっている．図 5.1 は，BasketballDrill シーケンスを例にとり，QP 値が異なる同じフレームの TU 画像を示している．図 5.1 を見ると，QP 値が 22 で符号化されたデータに比べ，QP 値が 42 で符号化されたデータは TU ブロック分割が荒いことがわかる．そのため，QP 値が 42 で符号化されたテストフレームでは，その TU 画像が提案する超解像ネットワークにとって不要な情報として見られているのではないかと考えられる．

表 5.2 から，原画像の解像度が 3840x2160 のテストデータにおいて，ours1 の精度が比較的良いことがわかる．しかし，DaylightRoad2 や FoodMarket4 のシーケンスの結果を見ると QP 値が 32 あるいは 42 で符号化されたテストデータに対しては，ours1 の精度は低くなっているなど，データ依存も確認できる．学習データに用いた原画像の解像

度が、テストデータとほぼ一致しているため、データ性質の近いシーケンスの精度がよくなったと考えられる。

表 5.1 各モデルの実験結果に対する PSNR [dB] (832x480)

テストデータの QP 値	シーケンス	Bicubic	SRResNet	ours1	ours2	ours3	ours4
22	BasketballDrill	26.6902	27.2116	27.2354	27.2793	27.1096	26.9338
	BasketballDrillText	24.4092	24.8481	24.8498	25.2969	24.8647	25.0122
	BQMall	26.4342	27.1835	26.7167	27.7945	27.2398	27.5661
	PartyScene	22.3341	22.5293	22.1772	23.0964	22.5965	22.9252
	RaceHorsesC	26.1165	26.8797	26.7615	26.9827	26.8804	26.6459
	平均	25.1969	25.7304	25.5481	26.0899	25.7382	25.8167
32	BasketballDrill	25.7827	26.1973	26.2518	26.1100	26.1588	25.8883
	BasketballDrillText	23.8399	24.2324	24.2685	24.5015	24.2681	24.2825
	BQMall	25.8262	26.4668	26.1555	26.7148	26.5262	26.5868
	PartyScene	22.1003	22.3463	22.1073	22.6753	22.3889	22.5490
	RaceHorsesC	25.0872	25.6646	25.6295	25.5183	25.6853	25.4191
	平均	24.5273	24.9815	24.8825	25.1040	25.0055	24.9452
42	BasketballDrill	23.8313	24.2247	24.2091	23.9128	24.1392	23.8046
	BasketballDrillText	22.4256	23.1322	23.1355	22.8206	23.0389	22.6808
	BQMall	24.0827	24.6815	24.6273	24.3844	24.6233	24.3624
	PartyScene	20.9778	21.3100	21.3442	21.1281	21.3307	21.0875
	RaceHorsesC	23.1553	23.5243	23.4578	23.1744	23.4658	23.2259
	平均	22.8945	23.3745	23.3548	23.0841	23.3196	23.0323

表 5.2 各モデルの実験結果に対する PSNR [dB] (3840x2160)

テストデータの QP 値	シーケンス	Bicubic	SRResNet	ours1	ours2	ours3	ours4
22	Campfire	28.9593	30.2856	30.1821	30.2323	30.0510	30.8312
	CatRobot1	31.8828	33.0433	33.0872	32.1215	32.8275	31.8780
	DaylightRoad2	32.0819	32.9831	32.8305	32.9111	32.9319	32.8365
	FoodMarket4	37.0740	37.0943	37.3975	36.3269	37.3450	36.2389
	ParkRunning3	26.0911	28.3113	28.3860	28.2611	28.1900	27.8912
	Tango2	35.6927	35.7404	36.1042	33.7858	35.6856	33.7361
	平均	31.9636	32.9097	32.9979	32.2731	32.8385	32.2353
32	Campfire	28.0370	31.0045	31.0488	30.3413	30.8955	30.2315
	CatRobot1	30.6772	31.8454	31.8506	31.4090	31.8183	31.3589
	DaylightRoad2	31.0395	33.6526	33.7258	33.2331	33.8162	33.2519
	FoodMarket4	34.5176	31.4696	31.6400	30.2067	31.4865	30.2971
	ParkRunning3	25.1755	29.4696	29.5505	29.0918	29.4718	29.0244
	Tango2	33.8822	33.4788	33.6819	32.1398	33.5646	32.1556
	平均	30.5548	31.8201	31.9163	31.0703	31.8422	31.0532
42	Campfire	26.0132	28.9800	29.0088	28.6391	29.0039	28.6602
	CatRobot1	28.2937	29.4709	29.4866	29.2579	29.5416	29.3286
	DaylightRoad2	28.6125	29.2527	29.2921	28.9876	29.3024	29.0176
	FoodMarket4	30.7712	26.6071	26.6495	25.9576	26.6138	26.0367
	ParkRunning3	23.3161	30.0671	30.1365	29.3735	30.1136	29.5124
	Tango2	30.8034	31.5831	31.6640	30.9175	31.5961	30.8958
	平均	27.9683	29.3268	29.3729	28.8555	29.3619	28.9086



図 5.1 BasketballDrill シーケンスの 100 フレーム目の TU 画像
(左から, QP22, QP32, QP42)

5.2.1 実験結果画像

次に, 各シーケンスに対する, 入力画像と Bicubic 補間による拡大画像, 実験により得られた画像, 原画像を比較する. 図 5.2~図 5.4 に原画像の解像度が 832x480 のテスト結果を, 図 5.5~図 5.7 に原画像の解像度が 3840x2160 のテスト結果を示す.

図 5.2 と図 5.3 より, ours2 及び ours4 の出力画像が他の手法の結果よりも, 物体の輪郭部分がくっきりと見えることがわかる. ただし図 5.4 の QP 値が 42 で符号化されたデータに対しては, 拡大画像においても目で見えてわかる違いは無い.

図 5.5~図 5.7 より, 原画像の解像度が 3840x2160 のテストデータに対しては, 拡大画像を見ても, 目で見えてわかる違いは無いことがわかる.

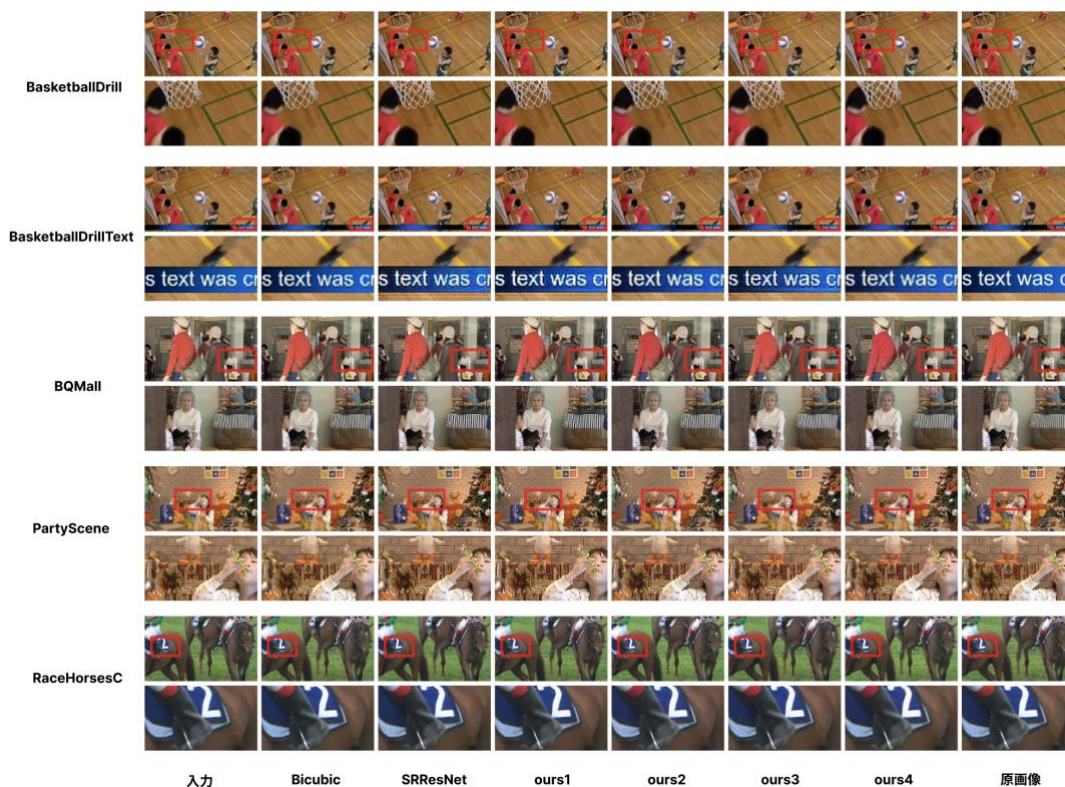


図 5.2 原画像 832x480 (QP22) テスト結果

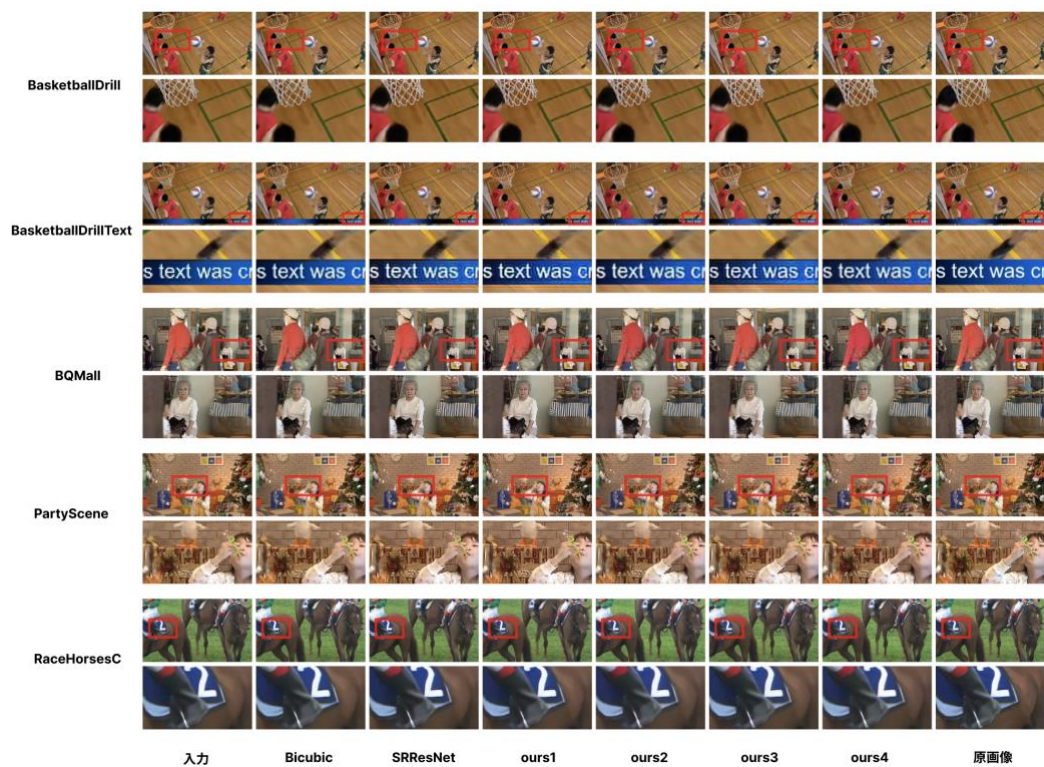


図 5.3 原画像 832x480 (QP32) テスト結果

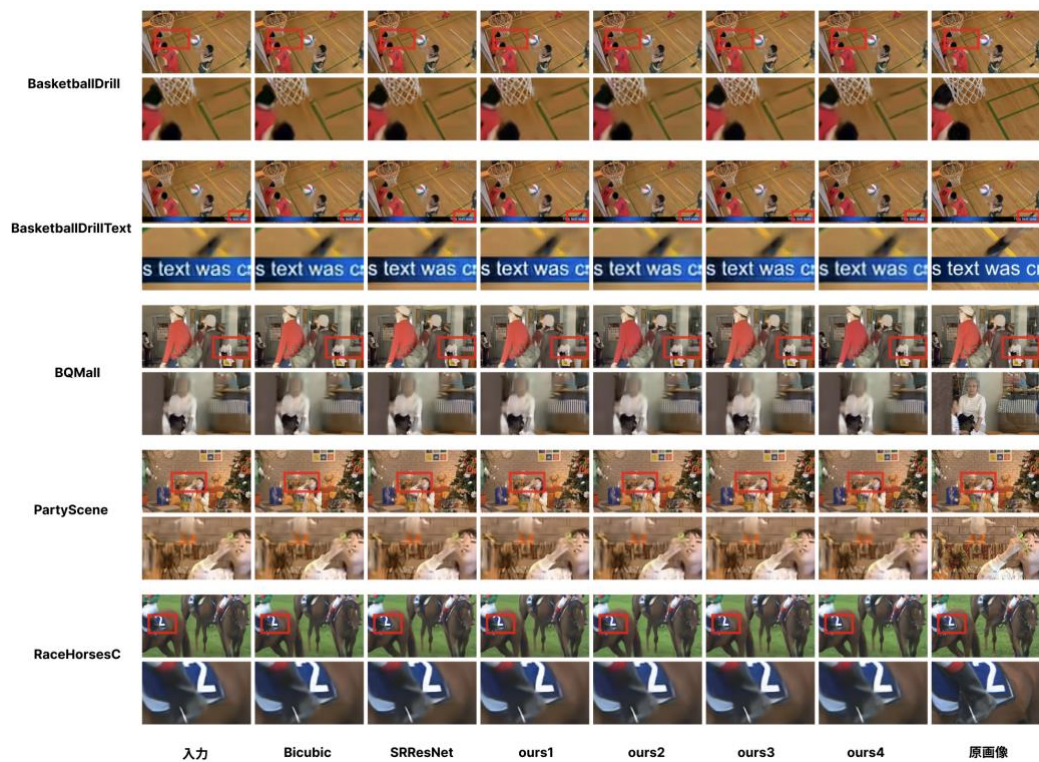


図 5.4 原画像 832x480 (QP42) テスト結果

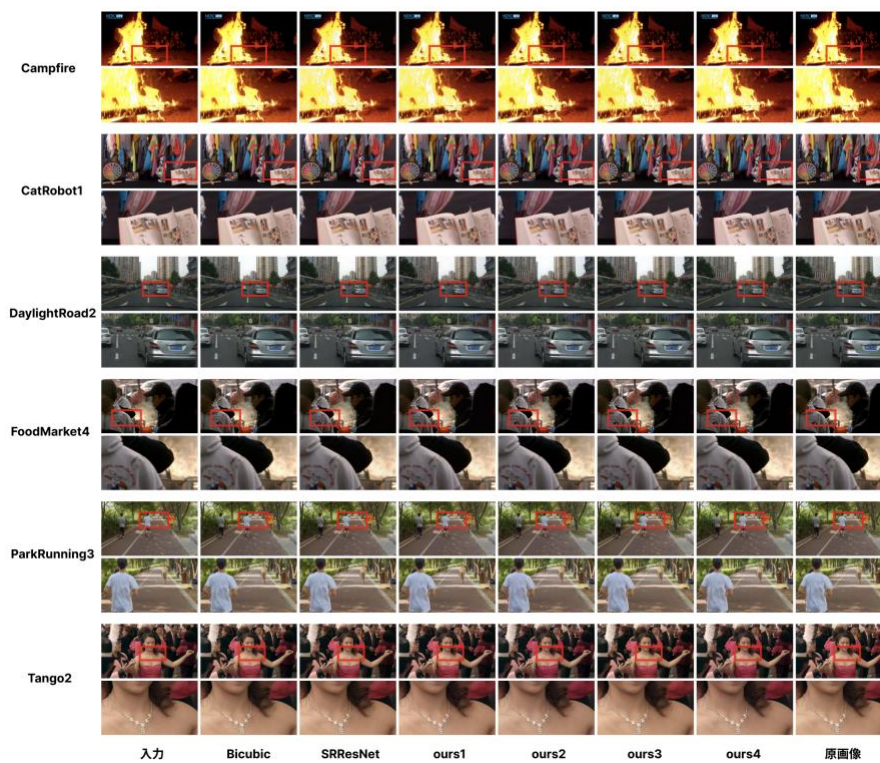


図 5.5 原画像 3840x2160 (QP22) テスト結果

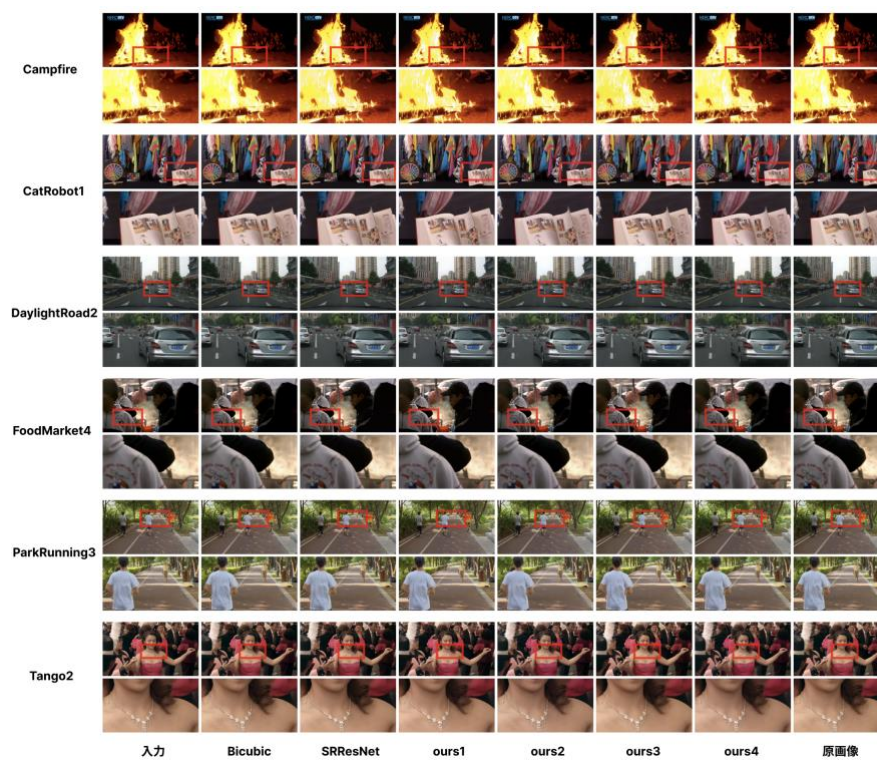


図 5.6 原画像 3840x2160 (QP32) テスト結果

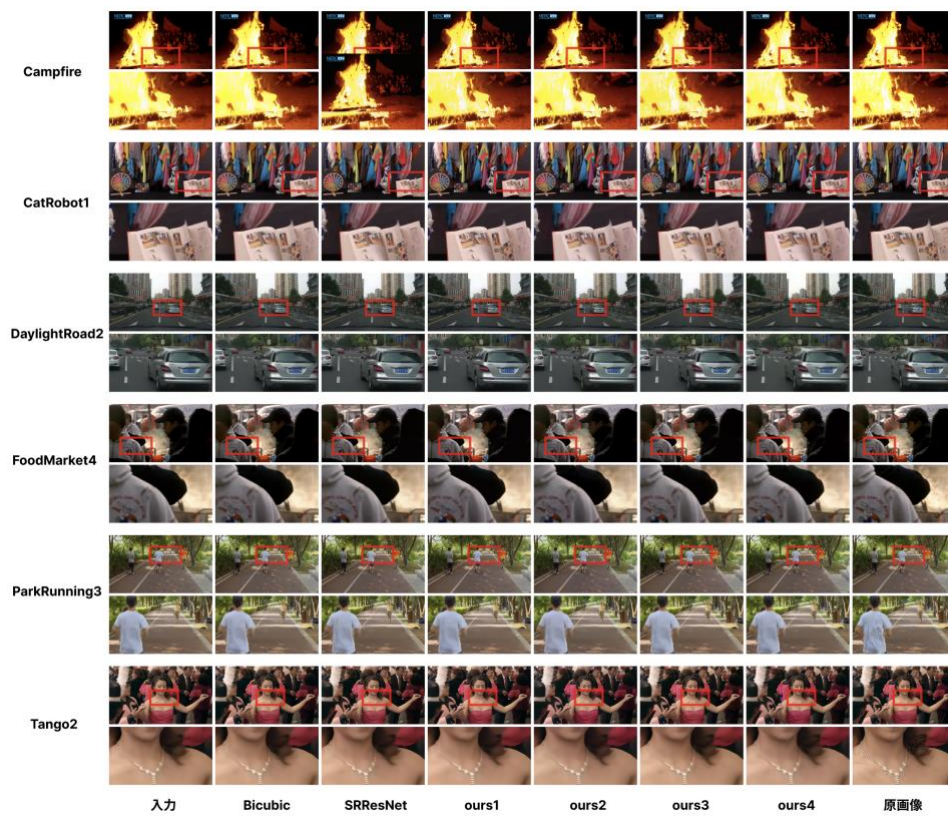


図 5.7 原画像 3840x2160 (QP42) テスト結果

第6章 結論

6.1 結論

本研究では、超解像ネットワークの一つである SRResNet を改良し、VVC 符号化映像の高解像度化及び高品質化の精度を向上させる提案を行った。

Bicubic 補間や SRResNet, その他の超解像手法においては、符号化の特性や条件を考慮していなかった。そこで、符号化に関する情報を TU 画像として表し、超解像ネットワークに組み込むことで、より高品質な高解像度映像を再構成できることを確認した。

6.2 今後の展望

従来の超解像手法では、ガウシアンフィルタなどの VVC に比べて簡素な手法を用いて縮小された画像あるいは映像に対する高解像度化及び高品質化を前提としており、我々が実際にテレビやスマートフォンで目にする通信を介した映像に対する手法は、ほとんど提案されていない。しかし、第5世代移動通信システム (5G) を利用した通信の進歩など、通信を介した映像の需要は今後も増え続けると考えられる。

本研究では、TU 画像を超解像ネットワークに組み込むことで、実用化予定の VVC で符号化及び圧縮された映像に対して超解像を行っている。実際に送受信される形の映像に対して実験を行ったことにより、提案手法は VVC 及び超解像技術の実用化に向けた技術であるといえる。さらに、TU 画像は、HEVC においても定義されるため、提案手法は VVC 以外の符号化映像に対しても有効であると考えられる。

謝辞

本研究をご指導していただいた渡辺教授に深く感謝いたします。

本研究に際して、アドバイスをしてくださったシャープ株式会社の方々に心より感謝いたします。

また、コロナ禍で充分に対面できなかつた中、オンラインでの素早い対応や実験環境を整えていただいた渡辺研究室の皆様感謝いたします。

最後に、健康面、精神面ともにサポートしていただいた家族に感謝いたします。

参考文献

- [1] 一般社団法人 電子情報技術産業協会 「民生用電子機器国内出荷統計」
<https://www.jeita.or.jp/japanese/stat/shipment/>
(2023年1月閲覧)
- [2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” *Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [4] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, “The SJTU 4K Video Sequence Dataset,” *the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013)*, Klagenfurt, Austria, July 3rd-5th, 2013.
- [5] A. Mercat, M. Viitanen, and J. Vanne, “UVG dataset: 50/120fps 4K sequences for video codec analysis and development,” in *Proc. ACM Multimedia Syst. Conf.*, Istanbul, Turkey, June 2020.
- [6] D. Ma, F. Zhang, and D. Bull “BVI-DVC: a training database for deep video compression,” *IEEE Transactions on Multimedia*, 2021.
- [7] Bossen, F., Boyce, J., Sühring, K., Li, X. and Seregin, V., “JVET common test conditions and software reference configurations for SDR video,” *Joint Video Experts Team Document, JVET-N1010*, Geneva, 2019.

図一覧

図 2.1 Reference Picture Resampling (RPR).....	4
図 3.1 SRResNet.....	5
図 3.2 Pixel Shuffler	6
図 4.1 符号化フレーム (UVG dataset[5]) (左) とその TU 画像 (右) の例 (2023 年電 子情報通信学会総合大会において発表予定)	8
図 4.2 提案する超解像ネットワーク	8
図 5.1 BasketballDrill シーケンスの 100 フレーム目の TU 画像.....	13
図 5.2 原画像 832x480 (QP22) テスト結果.....	13
図 5.3 原画像 832x480 (QP32) テスト結果.....	14
図 5.4 原画像 832x480 (QP42) テスト結果.....	14
図 5.5 原画像 3840x2160 (QP22) テスト結果.....	15
図 5.6 原画像 3840x2160 (QP32) テスト結果.....	15
図 5.7 原画像 3840x2160 (QP42) テスト結果.....	16

表一覧

表 4.1 TU 分割ブロックサイズと割当画素値	7
表 4.2 ra と io のデータ詳細	9
表 4.3 モデル作成時の学習条件.....	9
表 5.1 各モデルの実験結果に対する PSNR [dB] (832x480)	11
表 5.2 各モデルの実験結果に対する PSNR [dB] (3840x2160)	12

研究業績

- [1] R. Yano, Y. Liu, H. Watanabe, T. Suzuki, T. Chujoh, and T. Ikai “A Study of Video Super-Resolution Method Using Video Coded Data as Training Data,” 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE 2021), Oct. 2021.
- [2] T. Shindo, T. Watanabe, R. Yano, M. Arimoto, M. Takahashi, and H. Watanabe, “Super Resolution for QR code Images,” in IEEE Global Conference on Consumer Electronics (GCCE), pp. 281-284, Oct. 2022.
- [3] 矢野礼美菜, 劉韞, 渡辺裕, 鈴木拓矢, 范哲銘, 中條健, 猪飼知宏 “符号化映像のブロック分割を用いた超解像ネットワークの検討,” 電子情報通信学会, Mar. 2023. (2023年3月掲載予定)