

卒業論文概要書

Summary of Bachelor's Thesis

Date of submission: 02/02/2020 (MM/DD/YYYY)

学科名 Department	情報理工	氏名 Name	矢野礼美菜	指導 教員 Advisor	渡辺 裕 ㊞
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	1W173130-1		
研究題目 Title	符号化映像に対する超解像技術の検討 A Study on Super-Resolution Technology to Coded Video				

1. まえがき

2021年現在では、すでに新4K8K衛星放送が運用されており、それらの高解像度映像を出力するためのデバイスや、効率のよい符号化方式が模索されている。

一方、ニューラルネットワークの一つの問題解決手法として、超解像技術が画像・映像の解像度や品質の向上のための手段として広く研究されている。

2. 関連研究

2.1. Deformable Convolution

Deformable Convolution は畳み込み位置・距離(オフセット)を最適化する技術である。入力画像内の物体のスケールと形状を学習するため、図1のように、対象の畳み込み箇所によって受容野が柔軟に変化している。Deformable Convolution は映像超解像の分野においては、時間方向のアライメント、すなわち動き推定と動き補償を一度に行うことができ、オクルージョンやアーティファクトの低減に有効な手法として注目され始めている。

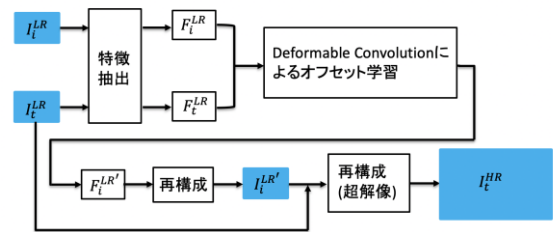


図1 Deformable Convolutionにおける畳み込み箇所(黄点)とその受容野(赤点)[1]

2.2. TDAN [2]

図2に示すTDANはDeformable Convolutionを映像超解像ネットワークに組み込んでいる。TDANでは、対象フレームの超解像を行うために、前後2フレームを利用してDeformable Convolutionによってオフセット、すなわちフレーム間の動き

を推定し、対象の高解像度フレームを取得する。Deformable Convolutionのオフセット学習と、超解像処理は別々の損失関数を用いているが、実際にはそれぞれの損失関数を加算しているものを利用しており、end-to-endで学習可能である。



※ $i = t-2, t-1, t+1, t+2$

図2 TDANのネットワーク概要

3. 提案手法

本研究ではTDANの映像超解像ネットワークに、次世代符号化方式の一つであるVVCによって符号化および縮小された映像を入力し、得られた映像の品質を画像評価指標PSNRにより検証する。

TDANで学習を行う際には、まず符号化を行っていないデータを利用してモデルAを作成し、その後、モデルAを符号化したデータによってファインチューニングを行い、モデルA-dec1とA-dec2を作成した。モデルA-dec1とA-dec2の作成にあたって使用したデータの概要を表1に示す。

表1 モデルA-dec1, A-dec2の作成に使用した符号化データ1,2の概要

フレームレート	符号化データ1	符号化データ2
30fps	3シーケンス	14シーケンス
50fps	8シーケンス	8シーケンス
120fps	7シーケンス	7シーケンス

4. 結果および考察

本研究の実験結果を表 2 に示す. 表 2 において BI とはバイキュービック補間を指しており, シーケンス名の括弧内はフレームレートを表している.

三つのシーケンスにおいては提案手法 A-dec2 が最も良い結果となったが, シーケンス BQSquare においては A-dec1 において最も良い結果となった. 4 つのシーケンスは全てオリジナルの解像度は 416×240 であり, VVC によって解像度が 208×120 に変換されたものを入力としている. また, シーケンス RaceHorsesD, BQSquare における実験結果画像を図 3 に示す.

表 2 各モデルの実験結果に対する PSNR[dB]

モデル シーケンス\	BI	A (既存手法)	A-dec1 (提案手法)	A-dec2 (提案手法)
RaceHorsesD (30fps)	23.8385	23.8955 (+0.0570)	24.0766 (+0.2381)	24.5842 (+0.7457)
BasketballPass (50fps)	25.5524	25.3686 (-0.1838)	25.6329 (+0.0805)	25.8385 (+0.2861)
BlowingBubbles (50fps)	22.7777	22.8004 (+0.0227)	23.0079 (+0.2302)	23.1049 (+0.3271)
BQSquare (60fps)	20.6018	20.6027 (+0.0009)	21.1228 (+0.5210)	20.5757 (-0.0261)

シーケンス BQSquare は, フレームレートが他のシーケンスよりも大きく, 他のシーケンスがなにか動作をしている中で, 風景画となっているため, フレーム間の動きが少ないという特徴を持っている. また表 1 より A-dec2 ではフレームレートが 30fps のシーケンスが 14 シーケンスと, 多く含まれており, A-dec2 はフレーム間で大きな動きを持つデータにより作成されていることがわかる. 従って, シーケンス BQSquare では, A-dec2 の PSNR 値が小さくなってしまったと考えられる.

5. まとめ

本研究では, 符号化映像に対して TDAN という超解像ネットワークを適用することで, 画像の解像度向上および符号化劣化の改善を行う手法を提案した. 実験結果より, 学習データと同程度の動きを持つ映像では, 高解像度かつ符号化劣化を抑えた映像を取得することが確認できた. しかし, 実験に用いたテストデータはオリジナルの解像度が 416×240 と, 4K や 8K の映像と近い条件とは言えないため, 実用化するには十分な精度を確かめられていない.

今後の課題としては, 新 4K8K 衛星放送で利用されている映像と同条件のテストデータで, 提案手法の有効性を検証することが挙げられる. また, 対象映像のフレームレートやフレーム間の動きの多さに合わせて学習したモデルを適用するなどし

て, 様々な映像に対して柔軟な超解像ネットワークを考案する必要がある.

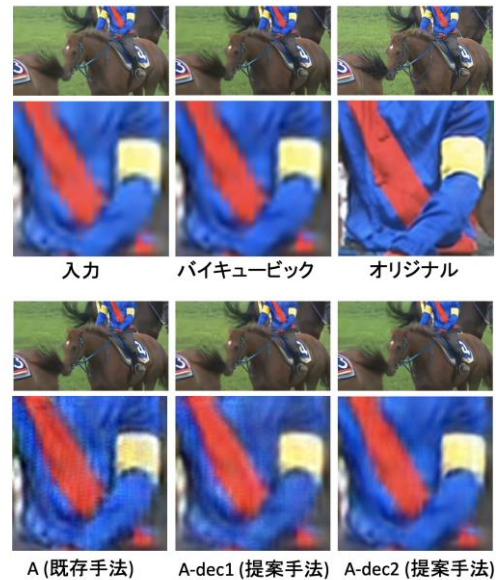


図 3 実験結果画像 (RaceHorsesD)



図 4 実験結果画像 (BQSquare)

参考文献

- [1] Jifeng Dai, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei: “Deformable Convolutional Networks” arXiv preprint arXiv:1703.06211, 2017.
- [2] Tian, Yapeng and Zhang, Yulun and Fu, Yun and Xu, Chenliang: “TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.

2020 年度 卒業論文

符号化映像に対する超解像技術の検討

A Study on Super-Resolution Technology to Coded Video

指導教員 渡辺 裕 教授

早稲田大学基幹理工学部 情報通信学科

1W173130-1

矢野 礼美菜

目次

第1章	序論.....	1
1.1	研究背景.....	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
2.5	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	超解像技術の分類.....	5
3.3.1	学習型超解像.....	5
3.3.2	再構成型超解像.....	5
3.4	映像超解像.....	6
3.4.1	オプティカルフロー.....	6
3.4.2	Deformable Convolution.....	6
3.5	むすび.....	7
第4章	提案手法.....	8
4.1	まえがき.....	8
4.2	提案手法において用いたネットワークの概要.....	8
4.3	ネットワークの詳細.....	8
4.3.1	特徴抽出.....	8
4.3.2	Deformable Convolution によるアライメント.....	9
4.3.3	再構成.....	9
4.3.4	損失関数.....	9
4.4	作成モデルおよび学習条件.....	10
4.5	むすび.....	11

第5章	実験結果・考察.....	12
5.1	まえがき.....	12
5.2	実験結果.....	12
5.2.1	実験結果画像.....	12
5.3	考察.....	14
第6章	結論.....	15
6.1	結論.....	15
6.2	今後の課題.....	15
謝辞	16
参考文献	17
図一覧	18
表一覧	19
研究業績	20

第1章 序論

1.1 研究背景

2021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新4K8K衛星放送テレビの販売台数は図1のように増加傾向にある[1].

新4K8K衛星放送の運用においての問題点の1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.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1つである超解像技術は,低解像度画像から高解像度画像へ変換する際の高周波成分の復元において,高い精度を出していることで知られている.

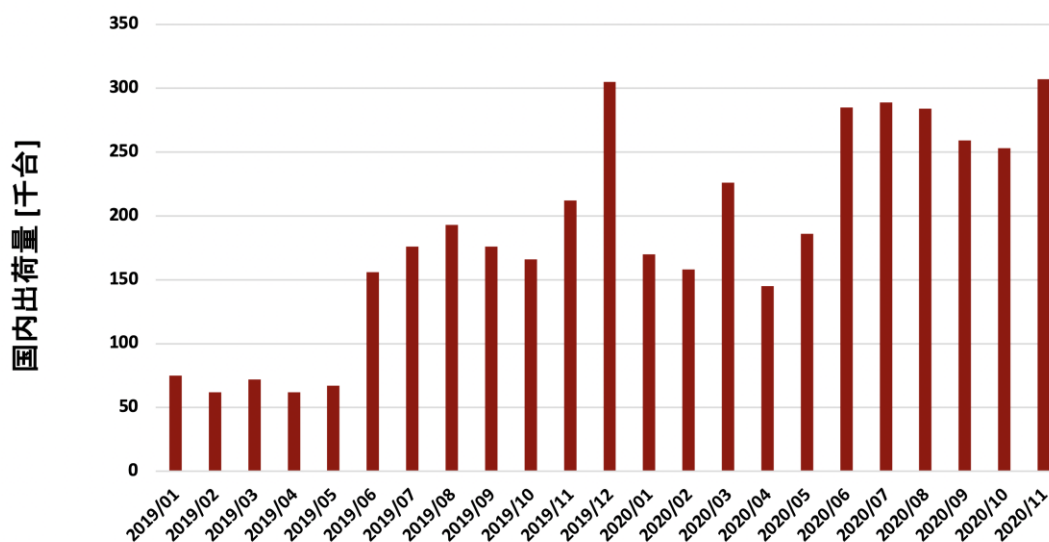


図 1.1 新4K8K衛星放送対応テレビの国内出荷量

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。

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、本研究の目的および従来研究について述べる。
- 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 本研究の主要なトピックである超解像技術について、その定義や分類に関して述べる。
- 第4章 本研究の提案手法について述べる。
- 第5章 本研究の実験結果および考察を述べる。
- 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。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)

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。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であると言える。

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．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およびその分類や定義について述べた後、映像超解像で利用されている技術について述べる。

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 超解像技術の分類

超解像技術は大きく分けて、マルチフレーム超解像と単一画像超解像の2種類がある。マルチフレーム超解像は、解像度を上げたい対象画像に対して、対象画像と似た画像を複数利用する。単一画像超解像は、あらかじめ高解像度と低解像度の画像ペアを用意して学習し、その学習結果を対象画像に適用する。マルチフレーム超解像と単一画像超解像はそれぞれ、再構成型超解像、学習型超解像と呼ばれる。

3.3.1 学習型超解像

学習型超解像では、あらかじめ高解像度画像と低解像度画像のペアを用意しておき、それぞれをパッチと呼ばれる小さなブロックに分割する。次に低解像度パッチから高解像度パッチへの非線形マッピングを学習することにより、辞書を作成する。入力した低解像度画像は、パッチごとに作成済みの辞書から対応するパッチを検索して、高解像度画像に変換される。

3.3.2 再構成型超解像

再構成型超解像では、映像内の連続フレームあるいは、対象物体が含まれる複数枚のフレームを利用して、画像間においてピクセル単位で差分を取り、他のフレームから標本点を対象フレームに増やしていくことで解像度を向上させる。フレーム間の差分を取得する方法としては、オプティカルフローやブロックマッチング法が用いられている。

3.4 映像超解像

映像超解像では、過去フレームや未来フレームを解像度向上のためのデータとして利用できるが、単一画像超解像とは異なり、時間方向の一貫性を保つ必要がある。そのため、多くの既存手法では、超解像処理を行う前に時間方向でのアライメント (位置合わせ) 処理を行っている。映像超解像における主なアライメント手法としては、オプティカルフローと Deformable Convolution が多く利用されている。

3.4.1 オプティカルフロー

オプティカルフローは、フレーム間において微小空間での輝度勾配を利用して、ピクセル単位で動きを求める手法である。オプティカルフローの概要を図 3.1 に示す。オプティカルフローを用いる場合、時間方向でのアライメントは、動き推定と動き補償の二つのモジュールに分けられる。つまりオプティカルフローによる動き推定に、アライメントの精度が依存する。

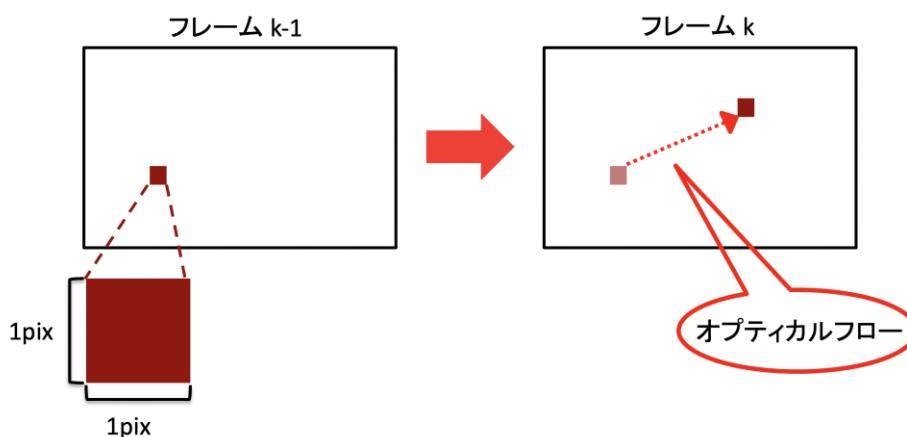


図 3.1 オプティカルフローの概要

3.4.2 Deformable Convolution

Deformable Convolution は畳み込み位置・距離 (オフセット) を最適化する技術である。入力画像内の物体のスケールと形状に合わせて畳み込みの受容野を動的に変化させることができるため、映像のような時系列データに適用することで、前後データとの差分を画像内で動的に取得することが可能である。図 3.2 に Deformable Convolution における畳み込み箇所とその受容野を表す。図 3.2 を参照すると、背景部分の畳み込みには画像全体が、オートバイや自動車といった物体部分の畳み込みには物体内の箇所が利用されていることを確認できる。また、映像超解像における時間方向のアライメントにおいては、Deformable Convolution を用いることで、動き推定と動き補償を一度に行うこと

ができ、オクルージョンやアーティファクトの低減に有効である。



図 3.2 Deformable Convolution における畳み込み箇所 (黄点) とその受容野 (赤点)
[2]

3.5 むすび

本章では、超解像技術の概要と分類および、映像超解像に利用される動き推定と動き補償の手法について説明した。

第4章 提案手法

4.1 まえがき

本章では、次世代符号化技術を利用した映像内の連続フレームに対して適用する、Deformable Convolution を含んだネットワークに関して記述する。

4.2 提案手法において用いたネットワークの概要

本研究で用いたネットワークである TDAN[3]は、次世代圧縮符号化技術によって縮小および符号化劣化を含んだ映像すなわち連続フレームに対して、拡大および品質向上を行う。そのネットワークの概略図を図4.1に示す。

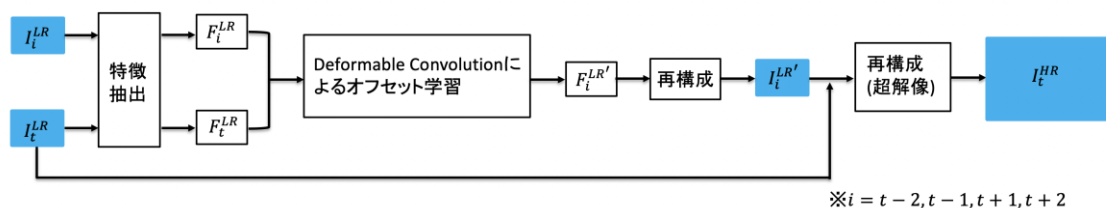


図 4.1 提案手法で用いたネットワークの概略

本研究で用いた手法は、まとめると式(4.1), (4.2)で表せるが、主に「特徴抽出」、 「Deformable Convolution によるアライメント」および「再構成」の3つのモジュールからなる。

$$I_i^{LR'} = f_{TDAN}(I_t^{LR}, I_i^{LR}) \quad (i = t-2, t-1, t, t+1, t+2) \quad (4.1)$$

$$I_i^{HR} = f_{SR}(I_{t-2}^{LR'}, I_{t-1}^{LR'}, I_t^{LR}, I_{t+1}^{LR'}, I_{t+2}^{LR'}) \quad (4.2)$$

4.3 ネットワークの詳細

4.3.1 特徴抽出

上述のネットワークではまず、5つの連続フレーム $I_i^{LR} (i = t-2, t-1, t, t+1, t+2)$ に対して、特徴抽出により特徴マップ $F_i^{LR} (i = t-2, t-1, t, t+1, t+2)$ を取得する。特徴抽出には EDSR[4]で提案されている残差ブロックを含むネットワークを利用している。

4.3.2 Deformable Convolution によるアライメント

Deformable Convolution によるアライメントでは、特徴マップ F_i^{LR} と F_t^{LR} を入力としてサンプリングパラメータ θ を推測する.

$$\theta = f_{\theta}(F_i^{LR}, F_t^{LR}) \quad (4.3)$$

ここで、 $\theta = \{\Delta p_n | n = 1, \dots, |R|\}$ であり、畳み込みにおける受容野 (オフセット) を表す. R は、 3×3 のカーネルサイズより、 $R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$ である. 推測したオフセット θ と特徴マップ F_t^{LR} から Deformable Convolution の処理を行うことでアライメントされた特徴マップ $F_t^{LR'}$ を得る.

$$F_t^{LR'} = f_{dc}(F_t^{LR}, \theta) \quad (4.4)$$

次に、アライメント済み特徴マップ $F_t^{LR'}$ からアライメント済みフレーム $I_t^{LR'}$ を得る. この処理では 3×3 の畳み込み層を利用する. 畳み込みにおいては式(4.5)のように、入力した対象フレーム I_t^{LR} とアライメント済みの前後 2 フレーム $I_i^{LR'} (i = t-2, t-1, t, t+1, t+2)$ の差分を損失関数 \mathcal{L}_{align} とする. ここで、 $N = 2$ である.

$$\mathcal{L}_{align} = \frac{1}{2N} \sum_{i=t-N, \neq t}^{t+N} \|I_i^{LR'} - I_t^{LR}\|_2^2 \quad (4.5)$$

4.3.3 再構成

最後に、対象フレーム I_t^{LR} とアライメント済みフレーム $I_i^{LR'} (i = t-2, t-1, t, t+1, t+2)$ から、超解像処理によって、目的のフレーム I_t^{HR} を得る.

$$I_t^{HR} = f_{SR}(I_{t-2}^{LR'}, I_{t-1}^{LR'}, I_t^{LR}, I_{t+1}^{LR'}, I_{t+2}^{LR'}) \quad (4.6)$$

超解像処理において、損失関数は式(4.7)である.

$$\mathcal{L}_{SR} = \|I_t^{HR'} - I_t^{HR}\|_2^2 \quad (4.7)$$

4.3.4 損失関数

4.2.2, 4.2.3 項で記述した損失関数(4.5), (4.6)は、実際の処理では式(4.8)のように、加算したものを利用しており、アライメントと超解像処理のパラメータは同時に更新されるため、提案手法で用いたネットワークは end-to-end で学習可能である.

$$\mathcal{L} = \mathcal{L}_{align} + \mathcal{L}_{SR} \quad (4.8)$$

4.4 作成モデルおよび学習条件

本研究では，4.2 項で記述した TDAN のネットワークを用いてモデル A，A-dec1，および A-dec2 の3つのモデルを作成した．表 4.1 にモデル A 作成時の学習条件を示す．ここで，Vimeo seplet dataset[5]は，1 シーケンスあたり 7 フレームを含んだ，6 万 4612 シーケンスからなるデータセットである．

表 4.1 モデル A 作成時の学習条件

モデル名	A
学習データ / 検証データ	Vimeo seplet dataset / Temple
バッチサイズ	64
学習率	1e-4
エポック数	600

モデル A-dec1，A-dec2 は，モデル A を同様のネットワークを用いてファインチューニングして得たモデルである．表 4.2 にモデル A-dec1，A-dec2 作成時の学習条件を示す．ここで，符号化データ 1 および符号化データ 2 は，SJTU[6]と UVG[7]のデータを VVC によって符号化および縮小したデータである．符号化データ 1，符号化データ 2 の内容について表 4.3 に示す．

表 4.2 モデル A-dec，A-dec2 作成時の学習条件

モデル名	A-dec1	A-dec2
学習データ / 検証データ	符号化データ 1 (学習用) / 符号化データ 1 (検証用)	符号化データ 2 (学習用) / 符号化データ 2 (検証用)
バッチサイズ	64	64
学習率	1e-4	1e-4
エポック数	1200	1200

表 4.3 符号化データの内容

フレーム レート	符号化データ 1 (18 シーケンス)		符号化データ 2 (29 シーケンス)	
	学習データ (15 シーケンス)	検証データ (3 シーケンス)	学習データ (26 シーケンス)	検証データ (3 シーケンス)
30fps	-	3 シーケンス	12 シーケンス	2 シーケンス
50fps	8 シーケンス	-	8 シーケンス	-
120fps	7 シーケンス	-	6 シーケンス	1 シーケンス

4.5 むすび

本章では，低解像度の対象フレームを，対象フレームと連続した前後2フレームからとともに超解像ネットワークに対する入力とすることで，高解像度のフレームを得る仕組みおよび，モデル作成時の学習条件について記述した．

第5章 実験結果・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. [8]のデータを，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 実験結果

実験に使用した4シーケンスに対する結果のPSNRを表5.1に示す．3つのシーケンスにおいては提案手法のA-dec2が最も良い結果となったが，シーケンスBQSquareにおいてはA-dec1において最も良い結果となった．四つのテストシーケンスは全て，オリジナルデータの解像度は 416×240 であり，VVCによって解像度は 208×120 に変換されている．

表 5.1 各モデルの実験結果に対するPSNR

シーケンス\モデル	バイキュービック	A (既存手法)	A-dec1 (提案手法)	A-dec2 (提案手法)
RaceHorsesD	23.8385	23.8955 (+0.0570)	24.0766 (+0.2381)	24.5842 (+0.7457)
BasketballPass	25.5524	25.3686 (-0.1838)	25.6329 (+0.0805)	25.8385 (+0.2861)
BlowingBubbles	22.7777	22.8004 (+0.0227)	23.0079 (+0.2302)	23.1049 (+0.3271)
BQSquare	20.6018	20.6027 (+0.0009)	21.1228 (+0.5210)	20.5757 (-0.0261)

5.2.1 実験結果画像

次に，各シーケンスに対する，入力画像とバイキュービック補間による拡大画像，実験により得られた画像，オリジナル画像を以下の図5.1から図5.4に示す．

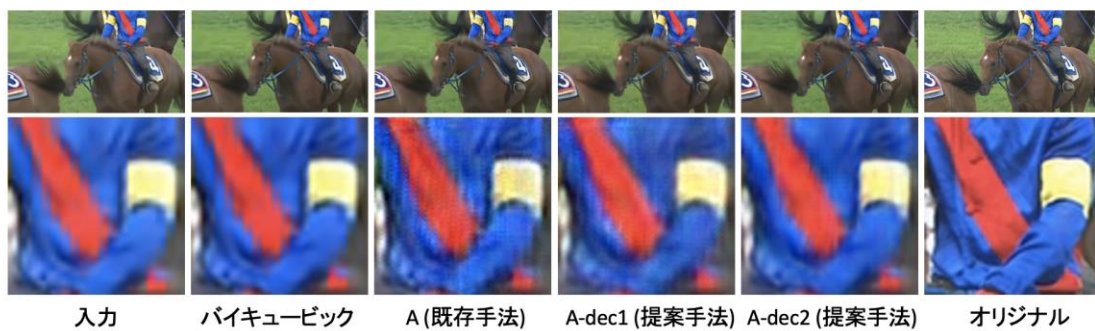


図 5.1 結果画像 (RaceHorsesD)



図 5.2 結果画像 (BasketballPass)



図 5.3 結果画像 (BlowingBubbles)

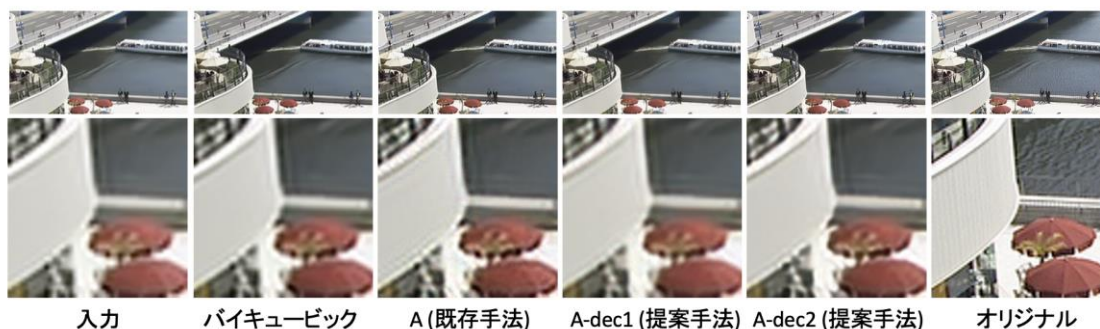


図 5.4 結果画像 (BQSquare)

5.3 考察

提案手法のモデル A-dec2 が、シーケンス BQSquare において良い PSNR 値が出なかった理由について考察する。表 5.2 は 4 つのテストシーケンスの概要を示したものである。表 5.2 を参照することにより、シーケンス BQSquare は他のシーケンスよりもフレームレートが高いことが確認できる。また、他の 3 シーケンスでは人物あるいは物体が激しい動作をしている映像であるが、シーケンス BQSquare は画角が広く、動きの少ない映像である。よってシーケンス BQSquare はフレーム間の動きが少ないという特徴を持っている。表 4.3 よりモデル A-dec2 では、26 シーケンスの学習データのうち、フレームレートが 30fps の映像が 12 シーケンス含まれているので、モデル A-dec2 は、フレーム間の動きが多いという特徴を学習したモデルであると考えられる。従って、シーケンス BQSquare は、モデル A-dec2 ではなく A-dec1 で良い PSNR 値が出たと考えられる。

表 5.2 テストシーケンスの概要

シーケンス	フレームレート [fps]	映像内容
RaceHorsesD	30	競馬
BasketballPass	50	バスケットボールの試合
BlowingBubbles	50	子供 2 人がシャボン玉をしている
BQSquare	60	川辺とテラスの風景

第6章 結論

6.1 結論

本研究では、符号化映像に対して TDAN という超解像ネットワークを適用することで、画像の解像度向上および符号化劣化の改善を行う手法を提案した。実験結果より、学習データと同程度の動きを持つ映像では、高解像度かつ符号化劣化を抑えた映像を取得することが確認できた。しかし、実験に用いたテストデータはオリジナルの解像度が 416×240 と、4K や 8K の映像と近い条件とは言えないため、実用化するには十分な精度を確かめられていない。

6.2 今後の課題

今後の課題としては、6.1 項で述べた通り、新 4K8K 衛星放送で利用されている映像と同条件のテストデータで、提案手法の有効性を確かめることが挙げられる。また、対象映像のフレームレートやフレーム間の動きの多さに合わせて学習したモデルを適用するなどして、様々な映像に対して柔軟な超解像ネットワークを考案する必要がある。

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った中、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/>
(2021年1月閲覧)
- [2] Jifeng Dai, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei: “Deformable Convolutional Networks” arXiv preprint arXiv:1703.06211, 2017.
- [3] Tian, Yapeng and Zhang, Yulun and Fu, Yun and Xu, Chenliang: “TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.
- [4] Lim, Bee and Son, Sanghyun and Kim, Heewon and Nah, Seungjun and Lee, Kyoung Mu: “Enhanced Deep Residual Networks for Single Image Super-Resolution” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July 2017.
- [5] Xue, Tianfan and Chen, Baian and Wu, Jiajun and Wei, Donglai and Freeman, William T: “Video Enhancement with Task-Oriented Flow” International Journal of Computer Vision (IJCV), Vol.127, No.8, pp.1106-1125, 2019.
- [6] L. Song, X. Tang, W. Zhang, X. Yang, P. Xia, “The SJTU 4K Video Sequence Dataset” the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), Klagenfurt, Austria, July 2013.
- [7] 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.
- [8] 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 1.1	新 4K8K 衛星放送対応テレビの国内出荷量	1
図 3.1	オプティカルフローの概要	6
図 3.2	Deformable Convolution における畳み込み箇所 (黄点) とその受容野 (赤点) [2].....	7
図 4.1	提案手法で用いたネットワークの概略.....	8
図 5.1	結果画像 (RaceHorsesD).....	13
図 5.2	結果画像 (BasketballPass).....	13
図 5.3	結果画像 (BlowingBubbles).....	13
図 5.4	結果画像 (BQSquare).....	14

表一覧

表 4.1	モデル A 作成時の学習条件.....	10
表 4.2	モデル A-dec, A-dec2 作成時の学習条件	10
表 4.3	符号化データの内容	10
表 5.1	各モデルの実験結果に対する PSNR.....	12
表 5.2	テストシーケンスの概要	14

研究業績

[1] T.Hori, R.Yano, H.Watanabe, T. Chujoh, T. Ikai, E. Sasaki, T. Suzuki and N. Ito: “Multi Frame Super-Resolution for Versatile Video Coding” Proc.International Workshop on Advanced Imaging Technology (IWAIT2021), No.45, Mar. 2021. (掲載決定)