卒業論文概要書

Summary of Bachelor's Thesis

-			Date of su	.0111351011.	01/30/2024 (IVIIVI/DD/1	111)
学科名 Department	情報通信	氏名 Name	速見 泰雅	指導	演辺 松	r:n
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	1W202285-0	教 貞 Advisor	汉 之 府	Ð
研究題目 Title	研究題目 TitleNeRF を基にしたフレーム補間手法の研究Research on NeRF-based Frame Interpolation Method					

1. まえがき

近年,インターネットやスマートフォンの普及 により動画を扱う機会が増加しており,高品質な 動画の需要が高まっている.フレーム補間は動画 のフレームレートを向上させるための技術であり, 動画の品質改善に用いられる.一方,Novel View Synthesis は複数の視点から新しい視点の画像を生 成するタスクで,Neural Radiance Fields (NeRF)[1] はその一例である.

本研究では, NeRF をフレーム補間手法として取り入れ,補間フレームの品質改善を図る.

2. 関連研究

2.1 NeRF

NeRF は静止した画像群から自由視点画像を生成する技術であり,任意の視点からの画像を得ることができる. NeRF では三次元空間位置座標x,y,z と視線方向 θ,ϕ を入力,色R,G,Bと密度 σ を出力とした多層パーセプトロンを学習する. 学習済みの多層パーセプトロンを用いてレンダリングをすることで自由視点画像が生成可能となる.

2.2 NeRFLiX

NeRFLiX[2]は NeRF のレンダリング画像の品質 を改善する技術である. NeRF は学習に使用する画 像群の視点数が少ない場合,レンダリング画像に 特有のアーティファクトが発生する.このアーテ ィファクトは一般のノイズ除去モデルの学習で用 いられるノイズとは異なるため,これらのモデル による除去は容易ではない.NeRFLiX では NeRF 特有のアーティファクトを劣化プロセスにより再 現する.劣化プロセスには Splatted Gaussian Noise (SGN)の付加,画像のピクセルに対する Repositioning (Re-Pos.), ぼかしパターンを再現する Anisotropic Blur (A-Blur)の三種類を用いる. 任意の データセットに対して劣化プロセスを介すること で疑似的な NeRF のレンダリング画像のデータセ ットを作成し, ノイズ除去モデルを学習すること でアーティファクトの除去を達成している.

1/20/2024 (AR UDD

3. 提案手法

提案するフレーム補間手法は大きく NeRF Block と Denoise Block に分かれる. NeRF Block では NeRF を用いたフレーム補間を行う. 任意の枚数の参照 フレームを NeRF の入力画像として扱い, NeRF を 学習する. 学習後,補間したい視点からの画像を レンダリングにより生成し,補間フレームとする.

Denoise Block では NeRF のレンダリング画像の 品質を改善する. NeRF は静的空間を対象としてい るため、本提案手法のように動物体を含んだフレ ームを NeRF に入力すると、動物体を中心に特有 のアーティファクトが発生する.そこで NeRFLiX の劣化プロセスに参照画像を一部重ねる Overlap と対象ピクセルの周辺範囲修正 Region Repositioning (RRe-Pos.)を追加し、ノイズ除去モデル を学習する.学習させたノイズ除去モデルを用い て NeRF のレンダリング画像の品質改善を図る.

4. 実験

Denoise Block におけるノイズ除去モデルの学習 は事前に行う.データセットには Vimeo90K[3]を 用いる.このデータセットはシーケンスごとに 7 枚の連続するフレームが提供されている.最初と 最後のフレームを参照画像とし,中間の 5 枚の画 像に対してそれぞれ劣化プロセスを施し劣化画像 を作成する.学習するノイズ除去モデルの構造は, ESRGAN[4]の生成器を参考にする.

Method NeRF Block	N. DE D1 1-	Denoise Block					DENID 1	CON ^	
	SGN	Re-Pos.	A-Blur	Overlap	RRe-Pos.	PSNR	551M	LFIP5 ↓	
Method-1	1						25.23	0.783	0.241
Method-2	1	1	1	1			25.40	0.783	0.282
Method-3	1	1	1	1	1	1	25.24	0.791	0.196

表1 各フレーム補間手法の評価比較

学習済みのノイズ除去モデルを用いて NeRF に 基づくフレーム補間を行う.データセットは Neural 3D Video Synthesis Dataset[5]を用いる.この データセットは 21 台の異なる位置のカメラで被 写体を撮影した動画で,その内 10 台の動画から特 定の時刻のフレームを取り出し,動画を作成する. 使用する 10 台のカメラは図 1 のように被写体を 囲むように配置されており,添え字の数字は使用 するフレームの時刻を示している.つまり,右側 から左側に移動しながら被写体を撮影するような 動画が得られる.これらのフレームを参照フレー ムとし,各参照フレーム間 15 枚,計 135 枚の補間 フレームを, NeRF を用いて生成する.FILM[6]に より生成したフレームを正解フレームとする.

各フレーム補間手法の評価結果の比較を表1に 示す. Method-1 では Denoise Block を使用せず, NeRF Block で生成されたフレームの品質を評価す る. Method-2, Method-3 では,その補間フレーム に対し Denoise Block でノイズ除去を行い,フレー ム品質を評価する。各ノイズ除去モデルは,表1に



図 2 補間フレームの一例, 左: GT, 中左: Method-1, 中右: Method-2, 右: Method-3

示した劣化プロセスにより学習する.また,補間 フレームの一例を図2に示す.

表1より,主観評価に近いとされている LPIPS において,提案手法が優れた結果を示すことが分 かる.また図2より, NeRF のレンダリング画像で は動物体である手を中心とした領域にアーティフ ァクトが発生しているが,提案手法ではアーティ ファクトが除去され,手のテクスチャが復元され ていることが分かる.

5. 結論

本研究では, NeRF に基づくフレーム補間手法を 提案した.また,劣化プロセスを用いたノイズ除 去モデルを使用して,本手法により生成したフレ ームの品質の改善を図った.

本手法は、現実での適用時において改善点が残 されている.実験ではカメラ視点のパラメータを 既知のものとして使用したが、実際にはカメラパ ラメータを推定する必要がある.推定したカメラ パラメータを本手法に使用した際の、補間フレー ムの品質についての精査が求められる.

参考文献

- B. Mildenhall *et al.*, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in ECCV, pp. 405-421, Nov. 2020.
- [2] K. Zhou *et al.*, "NeRFLiX: High-Quality Neural View Synthesis by Learning a Degradation-Driven Inter-viewpoint MiXer," in IEEE CVPR, pp. 12363-12374, Jun. 2023.
- [3] T. Xue *et al.*, "Video Enhancement with Task-Oriented Flow," in IJCV, pp. 1106-1125, Feb. 2019.
- [4] X. Wang *et al.*, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in ECCVW, pp. 63-79, Sep. 2018.
- [5] T. Li *et al.*, "Neural 3D Video Synthesis from Multi-view Video," in IEEE CVPR, pp. 5521-5531, Jun. 2022.
- [6] F. Reda *et al.*, "FILM: Frame Interpolation for Large Motion," in ECCV, pp. 250-266, Nov. 2022.

2023 年度 卒業論文

NeRF を基にしたフレーム補間手法の研究

Research on NeRF-based Frame Interpolation Method

指導教員 渡辺裕 教授

提出日:2024年1月30日

早稻田大学 基幹理工学部 情報通信学科

1W202285-0

速見泰雅

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	f 関連研究	3
2.1	まえがき	3
2.2	NeRF	3
2.3	NeRFLiX	4
2.4	FILM	5
2.5	むすび	5
第3章	f 提案手法	6
3.1	まえがき	6
3.2	提案手法	6
3.3	むすび	11
第4章	『 実験結果と考察	12
4.1	まえがき	12
4.2	Denoise Block の学習・結果	12
4.3	データセット	12
4.4	実験	13
4.4	実験結果	13
4.5	考察	16
4.6	むすび	16
第5章	『 結論と今後の課題	17
5.1	結論	17
5.2	課題	17
謝辞		18
参考文	〔献	19
表一覧	<u>k</u>	20
図一覧	± 	21
研究業	美積	22

第1章 序論

1.1 研究背景

近年,スマートフォンやインターネットの普及により動画を扱う機会が増えており,高 品質な動画の需要が高まっている.フレーム補間は動画のフレームレートを向上させるた めの技術であり,動画の品質改善に用いられる.一般に隣接しているフレームは似ている ことから,複数の参照フレームを用いてその間のフレームを予測することができる. Neural Network (NN)の普及によりフレーム補間を NN によって実現し,より高精度なフレームが生 成可能となっている.

一方, Novel View Synthesis は複数の視点から新しい視点の画像を生成するタスクである. NNを用いた Novel View Synthesis の例として Neural Radiance Fields (NeRF) [1]が挙げられる. NeRF や LLFF[2]といった Novel View Synthesis における技術は,既知な視点の画像から未知 な視点の画像を予測することで, VR などの自由視点映像に活用されることが期待されてい る.

1.2 研究目的

NN を用いたフレーム補間手法は参照フレームから得られる特徴量を用いて,動きベクト ル (フロー)を予測し補間フレームを生成する手法が一般である.しかし参照フレームの差 分が大きい場合,生成されるフレームの品質が大きく劣化する.これは差分が大きいこと で,参照フレーム間の同一の物体に対する対応点に誤りが生じるためである.参照フレー ム 2 枚を用いたフレーム補間では二次元の特徴量しか得られないため対応点が適切にとら れなければ生成画像に歪みが生じやすい.また既存の研究では,画像サイズが大きい場合 参照フレーム間の差分が小さくても対応する画素間の距離が大きくなり生成画像に歪みが 生じる場合がある.

そこで自由視点画像生成技術である NeRF をフレーム補間に用いることで三次元の特徴 を取り入れることができ、参照フレームの差分が大きい場合も、生成フレームの品質劣化 を抑えることができると考えられる.また、NeRF は入力画像のサイズに寄らず生成画像の 品質に影響を及ぼさない.本研究では、NeRF に基づくフレーム補間手法を提案し、生成さ れるフレームの品質改善を検討する.生成フレームの品質改善として NeRFLiX[3]に基づい た劣化プロセスを用いてノイズ除去モデルを学習し、フレームのアーティファクトを除去 する.

1.3 本論文の構成

本論文の構成を以下に示す.

第1章では本研究の研究背景,研究目的について述べる. 第2章では本研究で用いる関連研究について述べる. 第3章では本研究の提案手法について述べる. 第4章では実験の結果及び考察について述べる 第5章では本論文の結論及び今後の課題について述べる.

第2章 関連研究

2.1 まえがき

本章では、自由視点画像生成技術である NeRF, NeRF のレンダリング画像の品質改善技術である NeRFLiX, フレーム補間技術である FILM[4]について説明する.

2.2 NeRF

NeRF は多視点の画像群から自由視点の画像を生成する技術である.静止した被写体の複数視点からの画像を入力とし NN を学習する,学習した NN に指定した視点情報を入力する ことでその視点からの画像を出力として得る.

NeRF では式 2.1 のように三次元空間位置座標*X* = (*x*, *y*, *z*)と視線方向*d* = (θ , ϕ)から色 *R*, *G*, *B*と密度 σ を, NN のひとつである多層パーセプトロン (MLP: Multilayer perceptron)を 用いて学習する. 図 2.1 にあるように密度 σ は三次元空間位置座標*X*にのみより, 視線方向*d* は色*R*, *G*, *B*に使用される. これにより鏡などの同一の座標に対して異なる視線方向から見た ときに色が変化する状況に対応している. 関数 γ は Positional encoding で, 正弦波を用いて入 力の次元数を増やす. これにより画像の高周波成分が消失するのを防ぐ. Positional encoding を式(2.2)に示す. *L*はハイパーパラメータである.

$$F_{\theta}(\gamma(\boldsymbol{X},\boldsymbol{d})) \to (R,G,B,\sigma) \tag{2.1}$$

$$\gamma(p) = (\sin(2^0 \pi p), \cos(2^0 \pi p), \cdots, \sin(2^{L-1} \pi p), \cos(2^{L-1} \pi p))$$
(2.2)



図 2.1 NeRF の MLP

画像生成時は学習された MLP を用いて三次元空間の各位置座標,各方向の色と密度から画像の各ピクセルの色をレンダリングにより算出する.レンダリングでは生成画像の各ピクセルの色を式(2.3),式(2.4)のように求める.

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt$$
(2.3)

$$T(t) = exp\left(-\int_{t_n}^t \sigma(\mathbf{r}(s))ds\right)$$
(2.4)

積分の計算や計算量の面から、実際には光線r(t)上の点をサンプリングし、それらの密度 σ と 色cを遠点 t_f から近点 t_n まで足し合わせることで求める.

2.3 NeRFLiX

NeRFLiX は NeRF のレンダリング画像の品質を改善する技術である. NeRF は学習に用い る多視点画像が豊富にある場合,高品質な画像を生成できるが,少ない場合レンダリング 画像に特有のアーティファクトが生じる.また,NeRF はパラメータ数が多く学習するのに 時間がかかり,不十分な学習ではアーティファクトが生じやすい.このアーティファクト は,一般のノイズ除去モデルの学習で用いられるノイズとは異なるため,これらのモデル による除去は容易ではない.また,NeRF のレンダリング画像をデータセットとして用い, NeRF 専用のノイズ除去モデルを学習することは多くの時間と手間を費やすため困難であ る.そこで NeRFLiX では NeRF のレンダリング画像に現れるアーティファクトを劣化プロ セスにより再現し,疑似的な NeRF のレンダリング画像のデータセットを作成している.

劣化プロセスは図 2.1 に示すとおり,ガウシアンノイズを付与する Splatted Gaussian noise (SGN), ピクセルの再配置を行う Re-positioning (Re-Pos,), ぼかしを加える Anisotropic blur (A-Blur)を画像の一部分に適用する Region adaptive strategy (RA)によりそれぞれ施す.図 2.2 のように劣化プロセスを施した画像を用いてノイズ除去モデルを学習することで,NeRFのレンダリング画像のアーティファクトが除去される.ノイズ除去モデルでは劣化画像の他に,劣化プロセスを施していない参照画像 2 枚を入力として用いる.



図 2.2 劣化プロセス



図 2.3 ノイズ除去モデルの概要

2.4 FILM

FILM は参照フレームの特徴量を用いて補間フレームを予測するフレーム補間手法である.式(2.2)のように2枚の参照フレーム I_0 , I_1 から間の補間フレーム I_t をネットワークMで学習する. \hat{I}_t は予測フレームである.

$$\hat{I}_t = M(I_0, I_1)$$
(2.2)

予測フレーム*Î*_tを用いて再帰的にフレームを予測していくことで 2 枚の参照フレームから 2ⁿ – 1枚のフレームを補間できる.補足情報として深度マップやオプティカルフローを使用 せず,学習には 2 枚の参照フレームと 1 枚の正解画像のみを使用する.小さな動きと大き な動きの両方を表現するために,ネットワークMでは入力画像からピラミッド構造をしたマ ルチスケールの特徴量を抽出し,スケール間での畳み込みの重みを共有している.参照フ レームからそれぞれ特徴量を抽出し,それらを用いてフローを推定することで補間フレー ムを予測する.

2.5 むすび

本章では、本研究に関連する自由視点画像生成技術である NeRF と NeRF のレンダリング 画像のアーティファクトを除去する NeRFLiX,特徴量を用いたフレーム補間手法である FILM について説明した.

第3章 提案手法

3.1 まえがき

本章では NeRF に基づくフレーム補間手法を提案する.また,生成フレームの品質改善と して NeRFLiX に基づいた劣化プロセスを用いてノイズ除去モデルを学習し,フレームのア ーティファクトを除去する.NeRF に基づくフレーム補間を NeRF Block,生成フレームの品 質改善を Denoise Block とする.NeRF Block により補間フレームを生成し,各補間フレーム に Denoise Block を適用する.

3.2 提案手法

本研究では NeRF に基づくフレーム補間手法を提案する.提案手法全体の概要を図 3.1 に示す.図 3.1 にある通り提案手法では、NeRF を用いたフレーム補間手法の NeRF Block と、NeRF Block の生成画像に含まれるアーティファクトを除去する Denoise Block に大きく分かれる. NeRF Block の概要を図 3.2 に示す.図 3.3 に Denoise Block の概要を示す.

NeRF Block では図 3.1, 図 3.2 にある通り,任意の枚数の参照フレームを NeRF の入力画像とし,NeRF の MLP を学習する.NeRF は学習されたモデルを用いて任意の視点からの画像を生成できるため,補間したい視点の画像を学習された MLP を用いて生成する。

Denoise Block では図 3.3 にある通り,連続する参照フレーム 2 枚とその視点の間にある NeRF で生成された劣化画像 1 枚を入力として劣化画像のアーティファクト除去を行い,出 力画像を得る. Denoise Block は劣化画像それぞれに適用する. また,入力の 3 枚の画像に 対して Sobel filter を適用し,それぞれのエッジ画像を得る. これを用いて正解画像のエッジ 画像を予測する. エッジ側のネットワークの中間特徴量を画像側のネットワークと結合さ せ,正解画像を予測する. 学習するノイズ除去モデルの構造は ESRGAN [5]の Residual in Residual Dense Block (RRDB)を参考にする. RRDB の構造を図 3.4, 3.5 に示す. ノイズ除去 モデルの学習は事前に行う.

Denoise Block のノイズ除去モデルは事前に劣化プロセスを用いて学習を行う. NeRF は静 的空間を対象としているが、本研究では動的空間を対象としているため、NeRF の生成画像 にて動物体を中心にアーティファクトが発生する. そこで劣化プロセスでは NeRFLiX 既存 のプロセスに加え、図 3.6 のように、参照画像の一部を重ねる Overlap と対象ピクセルの周 辺範囲修正 Region Re-positioning (RRe-Pos.)を適用する. 本ネットワークの学習にはデータ セットとしてフレーム補間等で用いられる Vimeo90K[6]を用いる. Vimeo90K ではシーケン スごとに 7 枚の連続するフレームが提供されている. 最初と最後のフレームを参照画像と し、中間の5 枚の画像に対してそれぞれ劣化プロセスを施し劣化画像を作成する. また,このデータセットは大きく分けて以下の三つの特徴をもつシーケンスに分けられる.

I. 動物体を含まないかつ,視点が移動する

II. 動物体を含むかつ,視点は移動しない

III. 動物体を含むかつ,視点も移動する

NeRF の生成画像において動物体にアーティファクトが発生しやすいことを再現するため, 劣化プロセスの Overlap は動物体の領域にのみ適用する.参照画像 2 枚からオプティカルフ ローを RAFT[7]を用いてより求め,図 3.7 のフローチャートに従って Overlap の適用を判別 する.初めにオプティカルフローの分散がdより小さい場合,I.動物体を含まないかつ,視 点が移動する,場合としその対象画像は Overlap を適用しない. dはハイパーパラメータで d = 0.5とする.フローの分散がdより大きい場合,対象画像内に動物体が存在すると仮定す る.次に対象画像について,各座標の対応するオプティカルフローが,画像全体のオプテ ィカルフローの平均よりも大きいものに Overlap を適用する.Overlap の適用では参照画像 2 枚の同座標のピクセル値を対象画像のピクセル値と合成する.オプティカルフローの平均 より小さい座標は,Overlap を適用しない.オプティカルフローの平均と比較することで II. 動物体を含むかつ,視点は移動しない,III.動物体を含むかつ,視点も移動する,において 動物体のみを判別する.

NeRFLiX で用いられた Re-Pos.ではピクセルごとに再配置を行っていたのに対し, RRe-Pos. では対象のピクセルの縦横 3×3 の範囲に対し再配置を行う. Re-Pos.と RRe-Pos.の比較を図 3.8, 図 3.9 に示す.

損失は式(3.1),式(3.2),式(3.3)のように画像のL₁損失と VGG 損失,エッジ画像のL₁損失 の和をとる. VGG 損失は VGG-16 の事前学習済みのネットワーク 22 層目特徴量の平均二 乗誤差をとる.

$$Loss_{total} = L_{image} + L_{vgg} + L_{edge}$$
(3.1)

$$L_{image} = L1(out_{image}, gt_{image})$$
(3.2)

$$L_{vgg} = VGG(out_{image}, gt_{image})$$
(3.3)

$$L_{edge} = L1(out_{edge}, gt_{edge})$$
(3.4)



図 3.3 Denoise Block 概要図



図 3.4 RRDB の構造



図 3.5 Dense Block の構造







図 3.7 Overlap の概要図



図 3.8 Re-Pos.



図 3.9 RRe-Pos.

3.3 むすび

本章では、本研究で NeRF に基づくフレーム補間手法と参照フレームを用いた NeRF のレ ンダリング画像のアーティファクト除去手法について述べた.アーティファクト除去手法 では NeRFLiX で提案された劣化プロセスを用いてノイズ除去モデルを学習する.

第4章 実験結果と考察

4.1 まえがき

本章では、提案手法に基づく実験の概要、結果及び考察について述べる. まず、Denoise Block におけるノイズ除去モデルの学習の結果を述べる. その後、本実験で使用するデータ セット、実験方法について述べ、その実験結果を示す.

4.2 Denoise Block の学習・結果

Denoise Block のノイズ除去モデルの学習は事前に行う.ノイズ除去モデルの学習では 3 章で述べた通りデータセットとして Vimeo90K を用いる.図3.3 における畳み込み層(Conv) と RRDB のパラメータを学習により決定する.入力画像は参照画像 2 枚と,劣化プロセス を施した劣化画像 1 枚とし,劣化プロセスを施す前の画像を正解画像とする.図 4.1 に Denoise Block の学習結果例を示す. Overlap を適用した鉛筆が出力で復元できていることが 分かる.



図 4.1 Denoise Block の学習結果例 左:入力画像,中:出力画像,右:正解画像

4.3 データセット

本実験では、データセットとして Neural 3D Video Synthesis Dataset [8]を用い、cook_spinach と coffee_martini の二つのシーケンスを使用する. このデータセットは 21 台の異なる方向か ら被写体を撮影した動画データセットで、各動画の時刻は同期されている. また、21 台の カメラの配置がパラメータとして用意されている. その内の 10 台のカメラから撮影され た動画の特定の時刻のフレーム、計 10 枚の画像を用いる. 10 台のカメラの配置は図 4.2、 図 4.3 のように被写体を囲むように配置されている. 各動画から使用するフレーム画像の時 刻は図 4.2、図 4.3 の添え字にある通り、右側から左側に移るにつれ経過している. これら 10 枚のフレームをつなぎ合わせることで、右側から左側に移動しながら被写体を撮影する ような動画フレームとする. なお coffee_martini では撮影環境の不具合により図 4.3 の添え 字の 2 と 3 に間隔が空いている.



4.4 実験

入力画像として 10 枚のフレームを使用し, NeRF の MLP を学習する. 学習後, 各フレー ム間 15 枚, 計 135 枚のフレームを生成する. 生成されたフレームに対し, ノイズ除去を施 す. 評価においては, FILM により生成した画像を正解画像とし, 評価指標は PSNR, SSIM, LPIPS を使用する. PSNR は客観評価として扱われ, SSIM は PSNR に比べると主観評価に なる. LPIPS は主観評価に近い評価指標とされおり、学習済み画像分類ネットワークの VGG に画像を入力した時の中間特徴量を用いて算出する.

4.4 実験結果

各手法における評価結果を表 4.1,表 4.2 に示す. Mehod-1 では Denoise Block を使用せず, NeRF Block で生成されたフレームの品質を評価する. Method-2, Method-3 ではその補間フ レームに対し Denoise Block でノイズ除去を行い,フレーム品質を評価する。各ノイズ除去 モデルは,表 4.1,表 4.2 に示した劣化プロセスにより学習する.

また、補間フレームの一例を図 4.4、図 4.5、図 4.6、図 4.7 に示す.

Matha d	Method NeRF Block	Denoise Block				DENID 1	CCDA ↑		
Method		SGN	Re-Pos.	A-Blur	Overlap	RRe-Pos.	P2NK	551WI	LFIF5↓
Method-1	1						25.23	0.783	0.241
Method-2	1	1	1	1			25.40	0.783	0.282
Method-3	1	1	1	1	1	1	25.24	0.791	0.196

表4.1 各フレーム補間手法の評価比較(cook_spinach)

表4.2 各フレーム補間手法の評価比較(coffee_martini)

Method NeRF Block	NoDE Dio dr	Denoise Block					DENID 1	CCDA ↑	
	SGN	Re-Pos.	A-Blur	Overlap	RRe-Pos.	PSINK	331WI		
Method-1	1						25.23	0.783	0.241
Method-2	1	1	1	1			25.40	0.783	0.282
Method-3	1	1	1	1	1	1	25.24	0.791	0.196



図 4.4 補間フレームの一例 (cook_spinach) 左: GT, 中左: Method-1, 中右: Method-2, 右: Method-3



図 4.5 補間フレームの一例 (cook_spinach, その 2)



図 4.6 補間フレームの一例 (coffee_martini)



図 4.7 補間フレームの一例 (coffee_martini, その 2)

4.5 考察

PSNR, SSIM, LPIPS の評価指標による評価結果について,表4.1,表4.2より,主観評価 に近いとされている LPIPS において提案手法が優れた結果を示すことが分かる.本実験で はFILM により生成した画像を正解画像として用いているため,画像の視点,画角に対して 評価できていない. そのため客観評価に近いとされている PSNR においては3つの手法い ずれも差が生じていない.

また図 4.4 より, NeRF のレンダリング画像では動物体である手を正確に再現できていな い.しかし提案手法では手の詳細部分が復元され,手の周囲に見られたアーティファクト を除去できている.一方,図 4.5 のように NeRF のレンダリング画像が大きく劣化している 場合,提案手法を施しても手の詳細部分を復元できていない.そのため,ノイズ除去モデ ルによる処理画像の品質は NeRF のレンダリング画像の品質に依存し,品質を改善できる場 合とそうでない場合がある.図 4.6 より,NeRF のレンダリング画像では帽子のロゴが二重 になっているが,提案手法ではその重なりが解消されている.図 4.7 では提案手法により全 体的にぼけた画像が出力された.本画像は図 4.2 におけるカメラ配置 2 と 3 の間を補間した ものであり,参照フレーム間に依存して品質を改善できないケースを確認した.Denoise Block の学習時は,Vimeo90K に参照フレーム間差分が大きいものと小さいものそれぞれが 存在し,学習によりある程度の変位に対応できている.しかし,その変位量が学習時の範 囲を超える場合に,図 4.7 のような全体的にぼけた画像が得られたと考えられる.

4.6 むすび

本章では、NeRFに基づくフレーム補間手法と劣化プロセスを用いて学習したノイズ除去 モデルによる生成画像の品質を評価した.ノイズ除去モデルの学習では従来手法と提案手 法の劣化プロセスを比較した.提案した劣化プロセスにより NeRF に基づくフレーム補間手 法の生成画像のアーティファクトを除去することを確認した.

第5章 結論と今後の課題

5.1 結論

本研究では、NeRF に基づくフレーム補間手法を提案した.また,提案手法で生成された フレームのアーティファクトを除去するノイズ除去モデルを提案した. Overlap と RRe-Pos. を加えた劣化プロセスでノイズ除去モデルを学習することで,アーティファクトを除去し 詳細部分が復元されることを確認した.

5.2 課題

本研究では、現実での適用時においていくつかの課題がある. NeRF の学習では用いる画像のカメラパラメータを既知のものとして使用したが、実際にはカメラパラメータを推定する必要がある. 複数視点の画像群からカメラパラメータを推定する COLMAP[9][10]などの手法は存在するが、画像の枚数や視点関係により誤差が生じる. そのため、本研究で提案したフレーム補間手法ではカメラパラメータの推定により生成画像の品質低下が予想される. 今後、NeRF の学習と同時にカメラパラメータの推定も同時に行う研究 (NeRF--[11]など)により改善が見込まれる.

謝辞

本研究をご指導していただいた渡辺裕教授に感謝いたします.

また本研究に際して,アドバイスをしてくださった Sharp 株式会社の中條健氏,八杉将伸 氏,洪秀俊氏,范哲銘氏,猪飼知宏氏に心より感謝いたします.

また,常日頃から相談に乗ってくださり,研究に関する助言やご指摘をくださった研究 室の皆様にお礼申し上げます.

最後に、精神面、金銭面で多くの力添えをしていただいた家族に心より感謝いたします.

参考文献

- B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," in ECCV, pp. 405-421, Nov. 2020.
- [2] B. Mildenhall, P. Srinivasan, R. Ortiz-Cayon, N. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines," in ACM TOG, pp. 1-14, Jul. 2019.
- [3] K. Zhou, W. Li, Y. Wang, T. Hu, N. Jiang, X. Han, and J. Lu, "NeRFLiX: High-Quality Neural View Synthesis by Learning a Degradation-Driven Inter-viewpoint MiXer," in IEEE CVPR, pp. 12363-12374, Jun. 2023.
- [4] F. Reda, J. Kontkanen, E. Tabellion, D. Sun, C. Pantofaru, and B. Curless, "FILM: Frame Interpolation for Large Motion," in ECCV, pp. 250-266, Nov. 2022.
- [5] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in ECCVW, pp. 63-79, Sep. 2018.
- [6] T. Xue, B. Chen, J.Wu, D.Wei, and W. Freeman, "Video Enhancement with Task-Oriented Flow," in IJCV, pp. 1106-1125, Feb. 2019.
- [7] Z. Teed, J. Deng, "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow," in ECCV, pp. 402-419, Nov. 2020.
- [8] T. Li, M. Slavcheva, M. Zollhoefer, S. Green, C. Lassner, C. Kim, T. Schmidt, S. Lovegrove, M. Goesele, R. Newcombe and Z. Lv, "Neural 3D Video Synthesis from Multi-view Video," in IEEE CVPR, pp. 5521-5531, Jun. 2022.
- [9] J. Schonberger, J. Frahm, "Structure-From-Motion Revisited," in IEEE CVPR, pp. 4104-4113, Jun. 2016.
- [10] J. Schonberger, E. Zheng, J. Frahm, and M. Pollefeys, "Pixelwise View Selection for Unstructured Multi-View Stereo," in ECCV, pp. 501-518, Sep. 2016.
- [11] Z. Wang, S. Wu, W. Xie, M. Chen, and A. Prisacariu, "NeRF--: Neural Radiance Fields Without Known Camera Parameters," arXiv preprint arXiv: 2102.07064, 2021.

表一覧

表 4.1	各フレーム補間手法の評価比較	(cook_spinach)	
表 4.2	各フレーム補間手法の評価比較	(coffee_martini)	

図一覧

⊠ 2. 1 NeRF 𝒫 MLP
図 2.2 劣化プロセス4
図 2.3 ノイズ除去モデルの概要5
図 3.1 提案手法概要図7
図 3.2 NeRF Block 概要図
図 3.3 Denoise Block 概要図
図 3.4 RRDB の構造
図 3.5 Dense Block の構造
図 3.6 提案する劣化プロセス9
図 3.7 Overlap の概要図
⊠ 3. 8 Re-Pos
図 3.9 RRe-Pos10
図 4.1 Denoise Block の学習結果例11
図 4.2 カメラ配置 (cook_spinach)12
図 4.3 カメラ配置(coffee_martini)12
図 4.4 補間フレームの一例 (cook_spinach)13
図 4.5 補間フレームの一例 (cook_spinach, その 2)13
図 4.6 補間フレームの一例 (coffee_martini)14
図 4.7 補間フレームの一例 (coffee_martini, その 2)14

研究業績

- 速見泰雅,金洛旭,渡辺裕,中條健,青野友子,八杉将伸,洪秀俊,范哲銘,猪飼知宏: "NeRF および特徴マップに基づくフレーム補間手法の特性評価 (Characterization of frame interpolation methods based on NeRF and feature maps)", 2023 年画像符号化シンポジ ウム・2023 年映像メディア処理シンポジウム (PCSJ/IMPS2023), P2-05, Nov. 2023.
- [2] Luoxu Jin, Hiroshi Watanabe, Taiga Hayami, Takeshi Chujoh, Tomoko Aono, Yukinobu Yasugi, Sujun Hong, Zheming Fan, and Tomohiro Ikai: "Post-processing Based Image Coding via Stable Diffusion," 2023 Picture Coding Symposium, Image Media Symposium (PCSJ/IMPS2023), P3-08, Nov. 2023.
- [3] Hiroshi Watanabe, Luoxu Jin, Taiga Hayami, Takeshi Chujoh, Tomoko Aono, Yukinobu Yasugi, Sujun Hong, Zheming Fan, and Tomohiro Ikai: "Prompt-based Image Coding with Edge Information," 2023 Picture Coding Symposium, Image Media Symposium (PCSJ/IMPS2023), P1-12, Nov. 2023.
- [4] 速見泰雅, 金洛旭, 渡辺裕: "NeRF に基づくフレーム補間手法の品質改善 (Quality Improvement of NeRF-based Frame Interpolation Method)", 電子情報通信学会総合大会, Mar. 2024 (発表予定).
- [5] 渡辺裕,金洛旭,速見泰雅,中條健,八杉将伸,洪秀俊,范哲銘,猪飼知宏: "エッジ・色情報を反映したプロンプトベースの画像符号化 (Prompt-based Image Coding with Edge and Color Information)",電子情報通信学会総合大会, Mar. 2024 (発表予定).