次世代 AI を活用したマルチモーダル情報通信基盤

研究代表者　甲藤　二郎
（基幹理工学部　情報通信学科　教授）

１．研究課題

この 10 年間で、AI 技術の研究開発が急速に発展している。AI 技術の情報通信応用も急速に進化しており、検出、認識、管理、圧縮、予測、復元から生成に至るまで多岐に亘る。本研究課題では、AI 技術の最近の進展を踏まえつつ、情報通信技術として無線通信とマルチモーダル情報処理に重点を置いた研究開発を進める。AI 技術は、当初は CNN（Convolutional Neural Network）が基盤技術として用いられてきたが、最近は Transformer が主流になり、近年の大規模言語モデルの発展をもたらすと共に、状態空間モデル、フローモデル、拡散モデル（Diffusion Model）などが新たな学習モデルとして注目されている。また、大規模言語モデルの情報加工・編集機能は、情報生成に拡張性と柔軟性を提供すると共に、テキスト・音声・画像等の相互変換を容易にしている。また、無線通信への AI 技術の適用は、無線通信の高度化と安定化、無線通信へのセマンティクスの導入などの展開が期待されている。2025 年度の成果としては、NeRV モデルを用いた未来映像予測、ニューラル色表現を用いた 3DGS の圧縮、MogaNet を用いた未来映像予測に関する研究成果の紹介を行う。

２．主な研究成果

2.1 NeRV モデルの知識蒸留を用いた軽量化とフレーム外挿予測応用

2.1.1 はじめに

近年、深層学習を用いた映像処理技術は急速に発展している。一方で、映像データは高解像度化に伴い、取り扱う情報量が増加の一途をたどっており、それに対応するためモデルの大規模化・高計算量化が進んでいる。しかし、モデル規模や計算量の増大は、学習・推論コストや実装環境の制約を通じて、実運用上の大きな課題となっている。本稿では、主に映像圧縮モデルとして用いられる NeRV（Neural Representations for Videos）を対象に、3 種類の知識蒸留手法を適用することで、性能を維持しつつパラメータ数の削減を行う。さらに、NeRV の学習ネットワーク構造をフレーム外挿が可能となるよう拡張し、将来フレームを生成できる映像予測モデルとしても活用可能な手法を提案する。加えて、同一データセット上で既存のオプティカルフローベースの映像予測手法である DMVFN（Dynamic Multi-scale Voxel Flow Network）と性能を比較し、提案手法の有効性を検証した。

2.1.2 関連研究

2.1.2.1 NeRV

NeRV は映像をニューラルネットワークの中に符号化する新しい映像のニューラル表現手法である。フレーム列として映像を扱う従来の表現とは異なり、フレーム番号を入力として受け取り RGB 画像を出力する関数として映像を表現する。また NeRV を拡張した D-NeRV を本稿で

は主に用いる。D-NeRV は、より広範な映像の圧縮に対応したモデルである。各映像クリップのキーフレームに条件付けすることで、異なる映像を単一モデル内で表現できる。また、クリップ固有の視覚内容を動き情報から分離し、ニューラル表現に時間的推論を導入し、空間冗長性を減らすために中間出力としてタスク指向フローを用いる。

2.1.2.2 知識蒸留

知識蒸留とは、大きい教師モデルの出力や中間表現を知識として、小さい生徒モデルを学習させることで生徒が教師を模倣して同等に近い性能を目指す技術である。教師から知識を生徒に移すことで小型ネットワークに圧縮でき、推論時の計算コストやメモリ使用量を減らしつつ、精度低下を抑えた性能を狙うことができ、深層学習におけるモデル圧縮や高速化の代表的手法として用いられている。本稿では出力ベース、特徴ベース、関係ベースの 3 種類の知識蒸留に分けて考える。

出力ベースの蒸留は、教師モデルの最終出力を画素ごとに損失を取る手法である。特徴ベースの蒸留は、教師の中間表現を生徒に移すことによって損失を取る手法である。特に本稿では、AT（Attention Transfer）という手法を参考にする。中間表現の attention map の教師と生徒を合わせるように学習する手法である。最後に関係ベースの蒸留について説明する。本手法は、特徴間・層間・サンプル間の関係を知識として転移する。特に本稿では RKD (Relational Knowledge Distillation) を参考にする。従来の出力ベースの蒸留損失と異なり、距離、角度ベースの蒸留損失を取る手法である。

2.1.3 提案手法と評価

2.1.3.1 NeRV の知識蒸留を用いたモデルの軽量化

第一の実験では、既存の NeRV モデル（D-NeRV）に対して知識蒸留を用いたネットワークモデルの軽量化を行い、蒸留損失を取らない手法と比較することで、映像精度の改善を示す。データセットは UVG を用いる。結果は表 1 の通りである。表 1 の α、 β、 γ は 3 種類の蒸留損失を用いた場合を表し、損失関数は以下の(1)式のように表せる。

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \alpha\,\mathcal{L}_{resp} + \beta\mathcal{L}_{AT} + \gamma\,\mathcal{L}_{RKD} \tag{1}$$

表 1: 3 種類の蒸留損失を加えた時の客観評価

	蒸留なし	α	β	γ
PSNR[dB]	32.79	**33.16**	32.85	**33.00**
MS-SSIM	0.9383	**0.9389**	0.9380	**0.9386**
	$\alpha+\beta$	$\alpha+\gamma$	$\beta+\gamma$	$\alpha+\beta+\gamma$
PSNR[dB]	32.80	**33.18**	32.81	32.88
MS-SSIM	0.9375	**0.9383**	0.9380	0.9384

蒸留なしの場合と、各蒸留損失を単独で導入した場合を比較すると、出力ベース蒸留および関係ベース蒸留の導入によって画質が顕著に向上した。さらに、出力ベースと関係ベースを併用した場合も同様に改善が確認された。一方で、特徴ベース蒸留を単独で適用した場合

や、他の蒸留損失と組み合わせた場合には、画質の明確な向上は観察されなかった。これは、本タスクにおいて中間特徴の整合よりも、最終出力の再現性やサンプル間関係の保持を直接促す制約の方が有効であること、また特徴ベース蒸留は教師と生徒間の表現空間の影響を受けやすく、最適化が十分に進みにくいことが要因として考えられる。以上より、本研究の設定では、特徴ベース蒸留の画質改善への寄与は限定的であり、出力ベース蒸留および関係ベース蒸留がより有効な軽量化手法であると考えられる。

2.1.3.2 NeRV のフレーム外挿応用

第二の実験では、既存の NeRV 系モデル（D-NeRV）におけるフレーム内挿部のアーキテクチャを改変しフレーム外挿予測が可能なモデルを構築したうえで映像予測精度を評価する。比較対象として、近年提案された近隣フレームのオプティカルフローを活用する映像予測手法である DMVFN を用い、同一条件下で精度比較を行う。具体的には、UVG データセットの各映像(全 600 フレーム)に対し、先頭 300 フレームを学習に用い、学習後に 301～304 フレームを予測生成させる。

次に、D-NeRV をフレーム外挿予測に対応させるための実装上の変更点を述べる。まず、元の D-NeRV におけるフレーム内挿の仕組みを整理する。生成したい時刻(フレーム番号)をt、基準となる 2 枚のキーフレームに対応する時刻(フレーム番号)をt_0, t_1とすると、正規化時刻 $r(t;t_0,t_1)$ は(2)式のように表せる。ここで、内挿時は $r(t;t_0,t_1)$ を $[0,1]$ としていたが、$r(t;t_0,t_1)$ が 1 を超える値も取り得るように拡張した。

$$r(t;t_0,t_1)=\frac{t-t_0}{t_1-t_0} \tag{2}$$

提案モデルの結果を表 2、DMVFN の結果を表 3 に示す。フレーム間の動きが小さい HoneyBee では、DMVFN と提案モデルの PSNR 差が相対的に大きい結果となった。一方で、Beauty、ReadySteadyGo、Jockey の 3 映像については、DMVFN には及ばないものの、NeRV 系の圧縮モデルを基盤とした外挿として一定の予測性能を確認した。

表 2: 提案モデルのフレーム外挿結果

Frame Number	301	302	303	304
Honey Bee	35.11	34.45	34.38	34.04
Beauty	30.19	28.09	27.24	26.58
Ready SteadyGo	20.89	19.67	17.60	15.99
Jockey	20.34	17.46	15.92	14.80

表 3: DMVFN を用いた時のフレーム外挿結果

Frame Number	301	302	303	304
Honey Bee	50.05	46.34	43.70	42.06
Beauty	38.23	33.07	30.47	29.13
Ready SteadyGo	29.79	25.24	22.45	20.44
Jockey	24.62	20.04	17.91	16.63

2.1.4 まとめ

本稿では、NeRV(D-NeRV)を対象として、(1) 知識蒸留によるモデル軽量化、(2) フレーム外挿予測への応用、の 2 点を提案した。前半の軽量化では、再構成損失のみで学習していた既存モデルに対し、蒸留損失を導入することで画質指標の値の改善を確認し、性能を維持しながらモデルの効率化が可能であることを示した。後半の外挿予測では、D-NeRV の内挿機構を外挿に拡張することで、将来フレームを生成可能な予測モデルを構築し、既存の映像予測手法である DMVFN と比較しても大きく劣らない予測精度を達成できた。以上より、NeRV を圧縮のみならず映像予測へと応用できることを示し、適用範囲を拡張した。今後の課題として、より多様な知識蒸留手法の導入を検討し、さらなる性能向上を目指す。また、フレーム外挿予測については、ワーピングや時間方向のモデル化を改善し、長期予測における画質劣化の抑制と予測精度の向上に取り組む。

2.2 ニューラル色表現を用いた 3D Gaussian Splatting モデルの圧縮

2.2.1 はじめに

3D Gaussian Splatting（3DGS）は、多視点画像から任意視点の画像を生成する任意視点合成（Novel View Synthesis）手法として高い性能を示している。3DGS は、多数のガウスを用いてシーンを表現し、高速かつ高品質なレンダリングが可能である。一方で、各ガウスが多数のパラメータを持つため、学習済みモデルのサイズは数百 MB～1GB と大きく、実用上の制約となっている。特に色を表現する球面調和パラメータは大きなサイズを占めている。そこで本研究は、3DGS のレンダリング品質を保ちながら、学習済みモデルのメモリ使用量を大幅に削減することを目的とする。特にボトルネックとなっている方向依存の色について効率的な表現方法を追求する。具体的には、従来の球面調和関数を用いた色表現を、ニューラルネットワークを使った色表現に置き換える手法を提案する。さらに、このニューラル色表現を導入した 3DGS に対して、Compressed 3D gaussian Splatting（Compressed3DGS）を基にした、ガウスのプルーニング・パラメータの量子化・エントロピー符号化などを行い、実験によってその性能を評価する。

2.2.2 関連研究

2.2.2.1 3D Gaussian Splatting (3DGS)

3DGS は、多視点画像から新規視点画像を高速かつ高品質に生成する手法であり、シーンを数十万～数百万の 3D ガウスで表現する。各ガウスは、3 次元の中心座標、3 次元のスケール、4 次元の回転、1 次元の不透明度、48 次元の球面調和係数（SH 係数）を持つ。そのため、色表現に用いられるパラメータがモデルサイズの大部分を占めている。

2.2.2.2 ニューラル色表現

ニューラル色表現とは、任意視点合成においてニューラルネットワークにより色を推定する手法である。近年では、Point-NeRF や DVGO に代表されるように、点やボクセルなどの幾何プリミティブに学習可能な特徴ベクトル $\boldsymbol{f}$ を持たせ、(3)式のように、特徴ベクトル $\boldsymbol{f}$ と視点方向ベクトル $\boldsymbol{d}$ を入力とし、MLP によって、サンプリング点の色 $\boldsymbol{c}$ を推定する構成が一般的である。

$$\boldsymbol{c} = MLP(\boldsymbol{f}, \boldsymbol{d}) \tag{3}$$

特徴ベクトル $\boldsymbol{f}$ はレンダリング画像と正解画像との差に基づく損失関数を通じて MLP と同時に最適化される。このようなニューラル色表現は、3DGS のように各プリミティブに高次元の色係数を直接保持する手法と比較して、モデルサイズを抑えられる傾向がある。

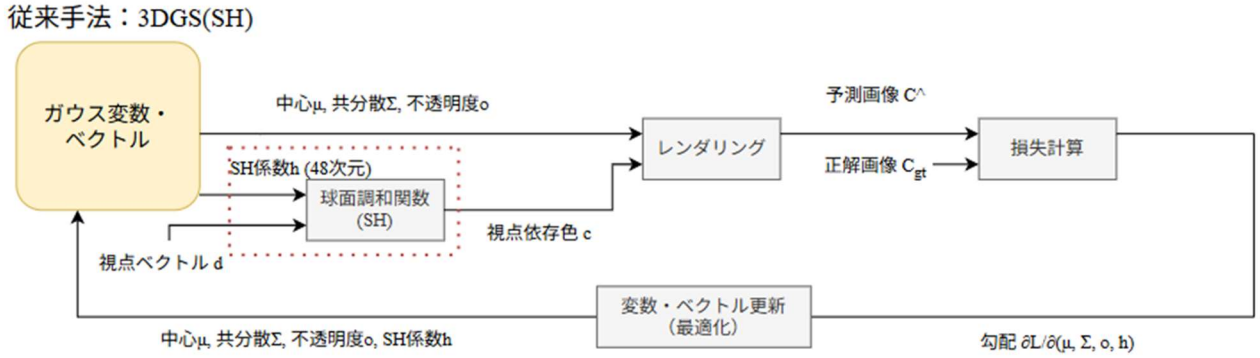
2.2.2.3 Compressed 3D Gaussian Splatting（Compressed3DGS）

Compressed3DGS は、3DGS のモデルサイズ削減を目的とした圧縮手法である。以下の 4 つのステップで圧縮を行う。

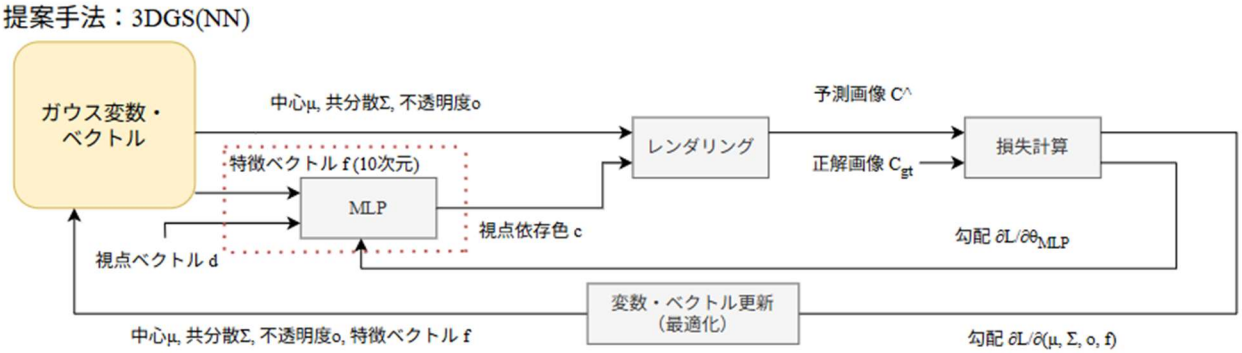
1. ガウスが持つ SH 係数と形状パラメータ（スケール＋回転）の各成分の出力画像に対する重要度を計算する。各ベクトルの重要度は、ベクトルの成分の最大値とする。SH 係数の重要度が 0 であるガウスをプルーニングする。
2. SH 係数と形状パラメータに対して、重要度で重み付けした k-means を使ったベクトル量子化（VQ）を行う。ただし、重要度が閾値以上のパラメータは VQ 対象外とする。
3. ガウスのパラメータを Quantization aware Training（QAT）を用いてファインチューニングし、低ビット化して保存する。
4. ガウスをモートン順（Z-order）に従って並べ替えるモートンソートを行い、空間的に近接したガウスが連続するように配置する。これによりデータの連続性を高め、LZ77 による圧縮効率を向上させる。その後、LZ77 とハフマン符号化を組み合わせた DEFLATE 圧縮アルゴリズムを用いて、モデル全体を圧縮する。

2.2.3 提案手法

オリジナルの 3DGS と提案手法の処理の流れを図 1 に示す。本手法では、3DGS のモデルサイズ削減を目的として、ニューラルネットワークを用いた色表現を導入した 3DGS（以下、3DGS（NN））を提案する。従来の 3DGS では、各ガウスに 48 次元の SH 係数を持たせ、視点方向に応じた色を SH 関数によって計算していた。これに対し提案手法では、各ガウスに 10 次元の特徴ベクトルを割り当て、これと視点方向ベクトルを MLP に入力することで、視点依存の色を推定する。この特徴ベクトル fは、レンダリングによって得られる予測画像と正解画像との差を最小化する損失関数に基づき、他のガウスのパラメータや色推定を行う MLP の重みと同時に最適化される。さらに、学習済みの 3DGS（NN）モデルに対して、Compressed3DGS を用いた圧縮を適用したモデル（以下、Compressed3DGS(NN)）を提案する。圧縮の流れは前節と同様の 4 ステップを行い、従来の SH 係数を特徴ベクトルに置き換えた形で各処理を行う。各属性に対する最終的な保存形式を表 1 に示す。



（a） 3DGS



（b） 提案手法

図 1: 3DGS と提案手法の処理の流れ

表 4: Compressed3DGS（NN）の各属性の保存形式

各ガウスが持つパラメータ	保存形式
中心座標	FP16 で保存
形状（スケール＋回転）	VQ して代表値は INT8 で保存
特徴ベクトル	VQ して代表値は FP32 で保存
不透明度	INT8 で保存

2.2.4 実験

従来の SH 関数を使った 3DGS（SH）と、提案手法である 3DGS（NN）、Compressed3DGS（NN）の画像品質、モデルサイズを比較する。データセットは、実写の多視点画像である Mip-NeRF 360 を用いる。9 つのシーンがあり、屋内 4 シーン、屋外 5 シーンで構成される。画像の解像度は慣例通り、横幅 1600px に制限している。3DGS の学習と Compressed3DGS のファインチューニングのイテレーションは、ともに 30,000 である。色表現に用いるニューラルネットワークは、隠れ層 1 層で、64 ユニットである。Compressed3DGS 圧縮手法の VQ で用いる、特徴ベクトルとガウス形状のコードブックサイズは、ともに 4096 である。VQ 対

象外のベクトルを決定するための重要度の閾値は、特徴ベクトルが$\beta_c = 3.0 \times 10^6$。ガウスの形状が、$\beta_g = 3.0 \times 10^6$である。結果は表 5 に示す。この結果から、3DGS（SH）と比較して、3DGS（NN）と Compressed3DGS（NN）は、画像品質指標は僅かに低下するが、モデルサイズをそれぞれ 2.37×、22.7×に圧縮することができた。Compressed3DGS（NN）におけるガウスのプルーニング率は 8.22%、特徴ベクトルの VQ 対象外は 1.35%、ガウスの形状の VQ 対象外は、9.22%であった。レンダリング画像を比較すると、3DGS（NN）と Compressed3DGS（NN）は 3DGS（SH）と比較して目立った画質劣化は見られなかったが、屋外シーンの詳細部分にぼやけなどの多少の劣化が見られた。

表 5: 3DGS の圧縮結果

	PSNR↑	SSIM↑	LPIPS↓	Size(MB)↓	圧縮率
3DGS（SH）	**27.55**	**0.813**	**0.221**	624.2	1.0x（基準）
3DGS（NN）	26.92	0.785	0.247	224.8 (NN:11.3KB)	2.37x
Compressed3DGS（NN）	26.38	0.782	0.263	**27.5** (NN:11.3KB)	22.7x

2.2.5 まとめ

本研究では、3DGS の圧縮を目的として、SH 関数を用いた色表現からニューラルネットワークを用いた色表現へ置き換える手法を提案した。さらに、このモデルに対して、Compressed3DGS を基にしたガウスのプルーニング・各パラメータの量子化・エントロピー符号化などを行うことで、画像品質をほぼ維持したまま約 22.7×のモデル圧縮を達成した。一方で従来の 3DGS と比較すると、屋外シーンでの細部表現に課題が見られた。今後は、特徴ベクトルの量子化手法や、色推定に用いるニューラルネットワークの構成を見直すことで、さらなる画質向上が期待される。

2.3 MogaNet を用いた SimVP のモデル改良

2.3.1 はじめに

近年、未来フレームを出力する映像予測モデルが数多く提案されている。深層学習を用いた映像予測モデルは、過去のフレームのシーケンスを入力とし未来のフレームを予測する。このタスクは、自動車の自動運転や低遅延映像伝送システム、映像伝送のパケットロス対策などの多くの応用先が考えられ、注目を集めている。一方で、モデルの複雑化、計算コストの増加、汎用性の低さなどの課題が生じている。また、直感とは異なり、予測するフレーム数が増えるにつれて予測の精度が低下する場合がある。そこで本研究では、CNN で構成されたシンプルかつ長期でも精度の高い映像予測モデルの実現を目指す。具体的には、CNN のみで構成された SimVP の時間方向の特徴量抽出コンポーネントを MogaNet というネットワークに変更し、損失関数の改善、skip connection の追加、オプティカルフローの追加によって長期でも予測精度の良いモデルを提案する。検証方法としては、KTH、KITTI といった代表的な映像予測データセットを使用し、PSNR・SSIM・LPIPS などの客観評価に加え、予測フレームの可視化による主観評価を行う。

2.3.2 関連研究

2.3.2.1 SimVP

図 2 に SimVP の構成を示す。SimVP は CNN のみで構成された映像予測モデルである。SimVP はエンコーダ、トランスレータ、デコーダの 3 つのモジュールで構成され、入力されたフレームの特徴量変化を捉え、任意のフレーム数を一度に出力する。エンコーダは、各入力フレームに対してチャネル方向への畳み込み・正規化・活性化関数という ConvSC ブロックを複数回繰り返すことで、空間的特徴量を抽出する。トランスレータは、複数サイズのカーネルでの畳み込みを並列で行う Inception ブロックを階層的に重ね、時間方向に畳み込みを行い、各フレーム間での時間的特徴量変化を抽出する。デコーダでは、転置畳み込み・正規化・活性化関数のブロックを複数回繰り返し、抽出した特徴量から予測フレームを生成している。

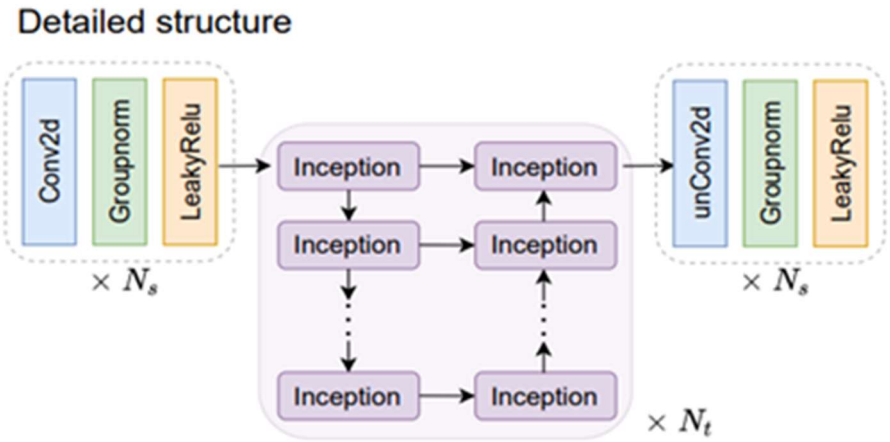


図 2: SimVP の構成

2.3.2.2 MogaNet

MogaNet は、低次元および高次元の特徴量間の相互作用を効率的に捉える CNN アーキテクチャである。このモデルは、空間集約ブロックとチャネル集約ブロックから構成され、局所的特徴と広域的特徴を統合する。空間集約ブロックは、FD（feature decomposition）モジュールと Moga（multi-order gated aggregation）モジュールから構成される。FD モジュールでは、1×1 畳み込みにより局所的特徴を抽出し、GAP（global average pooling）によりグローバルな特徴を取得する。入力特徴量を X、出力 Y とすると FD の処理は

$$Y = \mathrm{Conv}_{1\times1}(X), Z = \mathrm{GELU}\left(Y + \gamma_s \odot \left(Y - \mathrm{GAP}(Y)\right)\right) \tag{4}$$

で表される。ここで、γ_sは学習可能なスケーリング係数である。Moga モジュールでは、複数の深さ方向畳み込みを用いて異なる受容野の特徴を統合し、ゲーティング機構により空間的特徴を融合する。

チャネル集約ブロックでは、1×1 畳み込みおよび深さ方向畳み込みを用いてチャネル情報を再構成する。入力特徴量 X に対し、

$$Y = \mathrm{GELU}\left(\mathrm{DW}_{3\times3}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Norm}(X)\right)\right)\right) \tag{5}$$

$$Z = \mathrm{Conv}_{1\times1}\left(\mathrm{CA}(Y)\right) + X \tag{6}$$

と定義される。ここで、チャネル集約関数 CA は

$$\mathrm{CA}(X) = X + \gamma_C \odot \left(X - \mathrm{GELU}(XW_r)\right) \tag{7}$$

で表され、γ_Cはチャネルごとのスケーリング係数を表す。MogaNet では、これら二つのブロックを繰り返し適用することで、高次の特徴表現を獲得する。

2.3.3 提案手法

本研究では、SimVP の改良モデルを提案する。提案手法のアーキテクチャを図 3 に示す。提案手法では時間的特徴量変化を抽出するトランスレータを Inception から Moga Block に変更する。この変更により、動画内の動的物体部を強調して特徴量抽出が行われる。 また、長期予測でも物体のエッジや形状を保持するために損失関数に VGG の中間出力の差分の項目を加える。加えて、エンコーダからデコーダへの skip connection を挿入する。この際、物体の動きを捉えるために入力された最終 2 フレームから RAFT を用いてオプティカルフローを算出し、ワーピングを行う。

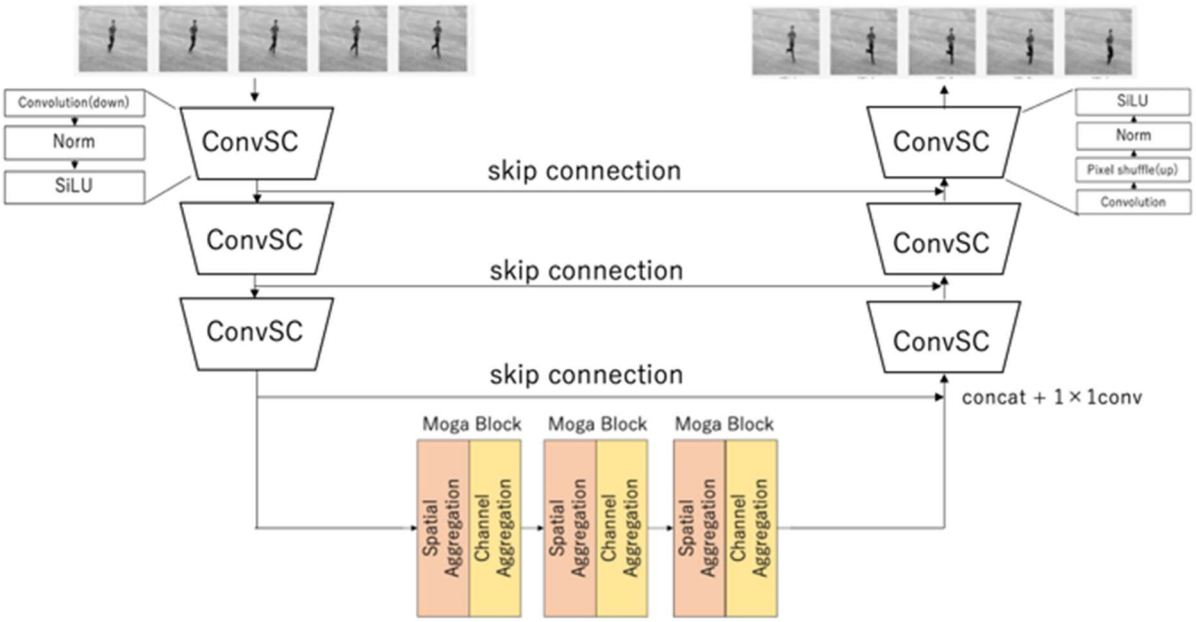


図 3: 提案手法のアーキテクチャ

2.3.4 実験

前節に示した提案手法を KTH、KITTI データセットを用いて学習評価する。比較として、従来の SimVP、 ConvLSTM、PredRNN も同様に学習を行う。それぞれ 10 フレームを入力し、20 フレームを出力させている。KTH データセットに対する客観評価結果を表 6 に、KITTI

データセットに対する客観評価結果を表 7 に示す。また、KTH データセットに対する予測画像例を図 4 に、KITTI データセットに対する予測画像例を図 5 に示す。

表 6: KTH データセットに対する評価結果

モデル名	PSNR(↑)	SSIM(↑)	LPIPS(↓)	GFLOPs
ConvLSTM	23.24	0.7167	0.4567	120.3
PredRNN	25.14	0.7387	0.4021	350.5
SimVP(IncepU)	27.66	0.7764	0.3112	35.45
SimVP(MogaNet) +VGG loss+skip+flow	**30.04**	**0.8425**	**0.2432**	45.31

表 7: KITTI データセットに対する評価結果

モデル名	PSNR(↑)	SSIM(↑)	LPIPS(↓)	GFLOPs
ConvLSTM	15.98	0.4368	0.6789	180.5
PredRNN	16.40	0.4531	0.6564	430.2
SimVP(IncepU)	20.34	0.5064	0.5628	43.64
SimVP(MogaNet) +VGG loss+skip+flow	**22.95**	**0.6353**	**0.4472**	64.37

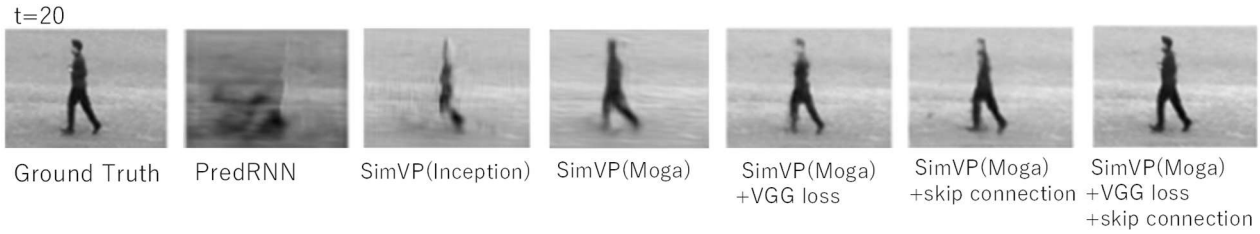


図 4: KTH データセットに対する予測フレーム例

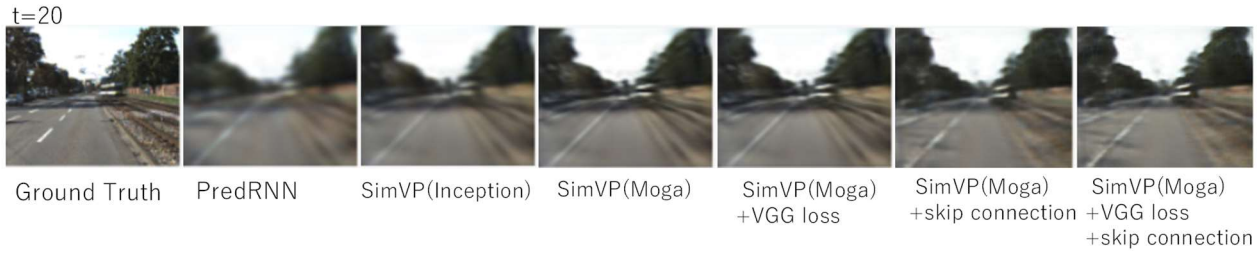


図 5: KITTI データセットに対する予測フレーム例

上記の結果から、KTH データセット、KITTI データセットどちらに対しても提案手法が PSNR、SSIM、LPIPS 全ての評価指標で最高数値となっている。計算量に関して、従来のモデルからは増加となっているものの、ConvLSTM、PredRNN と比較すると半分以下に抑えることができている。図 4 では、従来の SimVP ではフレーム数が増えることで人の輪郭が曖昧になり、最終的には形が確認できなくなってしまっているのに対して、提案手法では形状

が維持された出力が確認される。同様に、図 5 では、提案モデルでは、フレーム数が進んでも道路の輪郭、中心の白線の切れ目、線路など、物体のエッジや形が保持されて出力できている。

2.3.5 おわりに

本研究では、CNN ベースの映像予測モデルである SimVP の改良案を提案し、実験で精度が向上することを示した。今後は予測精度を維持しながらモデルの軽量化を目指すなど、映像予測の実用化に向けた改良案を模索していくことが重要であると考える。

３． 共同研究者

亀山　渉（早稲田大学・教授）
渡辺　裕（早稲田大学・教授）
文　鄭（早稲田大学・准教授）
佐藤　俊雄（早稲田大学・理工学術院総合研究所・客員上級研究員）
為末　和彦（早稲田大学・理工学術院総合研究所・客員上級研究員）
山崎　恭（北九州市立大学・准教授）
孫　鶴鳴（東京科学大学・准教授）

４． 研究業績

4.1 学術論文

【査読付き論文誌】

[1] Ran Wang, Yongqiang Wang, Heming Sun, and Jiro Katto：“Variable rate compression with Uniform Spatial-Frequency Residual Bottleneck Adapter for learned image compression”, EURASIP Journal on Advances in Signal Processing, Dec.2025, DOI: 10.1186/s13634-025-01268-x.

[2] Ao Luo, Diego Fujii, Keisuke Nonaka, Heming Sun, Jiro Katto: “Storage-and-Memory-Efficient Learned Image Compression with Quality-aware Hyperprior Pruning,” IET Image Processing, Oct.2025, DOI: 10.1049/ipr2.70231.

[3] Shuaibu Yau, Suphakit Awiphan, Jakramate Bootkrajang, Jiro Katto: “A Robust Throughput Estimation in Edge-Assisted Adaptive Bitrate Streaming Networks,” IEEE Access, Vol.13, Aug.2025, DOI: 10.1109/ACCESS.2025.3602651.

[4] Ran Wang, Wen Jiang, Heming Sun, Jiro Katto: “Single model learned image compression utilizing multiple scaling factors,“ Journal of Visual Communication and Image Representation, July 2025, DOI: 10.1016/j.jvcir.2025.104541.

[5] Yongqiang Wang, Feng Liang, Hang Chen, Haisheng Fu, Jiro Katto : “Towards Multi-Task Perception for Remote Sensing Imagery via Compression and Prompt Tuning,” IEEE Geoscience and Remote Sensing Letters, July 2025, DOI: 10.1109/LGRS.2025.3589030.

[6] Ao Luo, Linxin Song, Keisuke Nonaka, Jinming Liu, Kyohei Unno, Kohei Matsuzaki, Heming Sun, Jiro Katto : “MDLPCC: Misalignment-aware Dynamic LiDAR Point Cloud Compression,” Journal of Visual Communication and Image Representation, July 2025,

DOI: 10.1016/j.jvcir.2025.104481.
[7] Sanxin Jiang, Jiro Katto, Heming Sun: “RDDM: A Rate-Distortion Guided Diffusion Model for Leaned Image Compression Enhancement,” IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Early Access, April 2025, DOI: 10.1109/JETCAS.2025.3563228.
[8] Heming Sun, Lu Yu, and Jiro Katto: “Q-LIC: Quantizing Learned Image Compression with Channel Splitting,” IEEE Transactions on Circuits and Systems for Video Technology, Vol.35, No.4, pp.3798-3811, April 2025, DOI: 10.1109/TCSVT.2022.3231789.

【査読付き国際学会】
[1] Kasidis Arunruangsirilert, Jiro Katto: “Evaluation of NVENC Split-Frame Encoding (SFE) for UHD Video Transcoding,” PCS 2025, Dec.2025.
[2] Qingyue Ling, Zhengxue Cheng, Donghui Feng, Shen Wang, Chen Zhu, Guo Lu, Heming Sun, Jiro Katto, Li Song: “A Multi-Grid Implicit Neural Representation for Multi-View Videos,” PCS 2025, Dec.2025.
[3] Shimon Murai, Fangzheng Lin, Jiro Katto: “FlashGMM: Fast Gaussian Mixture Entropy Model for Learned Image Compression,” IEEE VCIP 2025, Dec.2025.
[4] Kasidis Arunruangsirilert, Jiro Katto: “Evaluation of GPU Video Encoder for Low-Latency Real-Time 4K UHD Encoding,” IEEE VCIP 2025, Dec.2025.
[5] Eiko Nakajima, Fangzheng Lin, Kasidis Arunruangsirilert, Jiro Katto: “Evaluation of 2D Video Interpolation and Extrapolation Methods for Real-Time V-PCC Error Concealment,” IEEE VCIP 2025, Dec.2025.
[6] Yongqiang Wang, Feng Liang, Heming Sun, Jiro Katto: “CGICM: CLIP-Guided Semantic Frequency Adaptation in Image Compression for Machines,” IEEE VCIP 2025, Dec.2025.
[7] Kasidis Arunruangsirilert, Pasapong Wongprasert, Jiro Katto: “Evaluations of High Power User Equipment (HPUE) in Urban Environment,” ICCCN 2025, Aug.2025.
[8] Zhang Chun, Heming Sun, Jiro Katto: “FLAVC: Learned Video Compression with Feature Level Attention,” IEEE CVPR 2025, June 2025.

【国内学会・研究会】
[1] 清水颯人・甲藤二郎: “リアルタイム点群圧縮における空間コンテキスト強化の検討および伝送・可視化実験による性能評価,” 電子情報通信学会 IE 研究会, Mar.2026.
[2] 田中晃誠・甲藤二郎: “3D Gaussian Splatting 圧縮技術 HAC++の性能改善,” 電子情報通信学会 IE 研究会, Mar.2026.
[3] 村井史門・甲藤二郎: “ウェーブフロント並列化を用いた学習型画像圧縮,” 電子情報通信学会 IE 研究会, Mar.2026.
[4] 内芝謙允・村井史門・甲藤二郎: “FlashGMM による LALIC の圧縮性能向上,” 電子情報通信学会 IE 研究会, Mar.2026.
[5] 橋本夕輝・甲藤二郎: “2 段階学習によるパラメータ効率的な参照画像セグメンテーションの提案,” 電子情報通信学会 IE 研究会, Mar.2026.

[6] 柳川健太・村井史門・甲藤二郎: “線形アテンションを用いた意味的学習型画像圧縮,” 電子情報通信学会 IE 研究会, Mar.2026.
[7] 佐藤俊雄, 甲藤二郎: “予測モデルによるビデオストリーミングの複数フレームの修復,” 電子情報通信学会総合大会, Mar.2026.
[8] Ao Luo, Keisuke Nonaka, Jiro Katto: “Structure-from-Motion Method Comparisons for 3DGS,” 電子情報通信学会総合大会, Mar.2026.
[9] Kasidis Arunruangsirilert, Jiro Katto: “Feasibility Study of Server-Side AI Rate Control for 5G MBS using Probabilistic Channel Modeling, “ 電子情報通信学会総合大会, Mar.2026.
[10] 村井史門・甲藤二郎: “混合ガウスモデルと平均シフト量子化を用いた学習型画像圧縮,” 電子情報通信学会総合大会, Mar.2026.
[11] 辻井若葉・甲藤二郎: “ボリュメトリックビデオストリーミングにおける適応型 ABR アルゴリズムの提案と評価,” 電子情報通信学会 IN 研究会, Mar.2026.
[12] 杉本遼太・金井謙治・甲藤二郎: “３D オブジェクト形状を考慮した SHAP に基づく適応型点群ダウンサンプリング手法,” 電子情報通信学会 IE 研究会, Feb.2026.
[13] 池邉諭次郎・甲藤二郎: “NeRV モデルの知識蒸留を用いた軽量化とフレーム外挿予測応用,” 電子情報通信学会 IE 研究会, Feb.2026.
[14] 中原将希・甲藤二郎: “高精度な CNN ベースの長期映像予測モデルの検討”, 電子情報通信学会 IE 研究会, Dec.2025.
[15] 野口陽生・甲藤二郎: “ニューラル色表現を用いた 3D Gaussian Splatting モデルの圧縮,” 電子情報通信学会 IE 研究会, Feb.2026.
[16] 清水颯人・甲藤二郎: “リアルタイム 3 次元 LiDAR 点群圧縮のための階層的アニメーション伝送手法,” PCSJ/IMPS 2025, Nov.2025.
[17] 杉本遼太・金井謙治・甲藤二郎: “3 次元物体分類のためのグローバルとローカルな領域を考慮した SHAP に基づくダウンサンプリング手法,” PCSJ/IMPS 2025, Nov.2025.
[18] 村井史門・林方正・甲藤二郎: “学習型画像圧縮のための高速な混合ガウスエントロピーモデル,” PCSJ/IMPS 2025, Nov.2025.
[19] 中島瑛子・甲藤二郎: “V-PCC のパッチ連続性を考慮した三次元点群符号化制御,” PCSJ/IMPS 2025, Nov.2025.
[20] Zhi-Han Xue, Jiro Katto: “Scene Hand-Drawn Sketch to Image Generation with Adaptive Multi-Conditional Diffusion Model,” 映像情報メディア学会年次大会, Aug.2025.
[21] 杉本遼太・金井謙治・甲藤二郎: “3 次元物体分類のための SHAP を用いたダウンサンプリング手法,” 電子情報通信学会 IE 研究会, June 2025.
[22] 中原将希・甲藤二郎: “軽量かつ高精度な CNN ベースの映像予測モデルの検討,” 電子情報通信学会 IE 研究会, June 2025.

4.2 総説・著書

なし

4.3 招待講演

[1] 甲藤二郎: “IP マルチキャスト放送の無線伝送に向けた技術開発,” AXIES 高品質・セキュ

リティ ICT 部会会合, Oct.2025,

4.4 受賞・表彰

[1] 杉本遼太, 電子情報通信学会・画像工学研究会, IE 賞

[2] 村井史門, 画像符号化シンポジウム 2025, ベストポスター賞

4.5 学会および社会的活動

[1] 甲藤二郎： 総務省 放送システム委員会 主査代理

５． 研究活動の課題と展望

2025 年度は、上記に紹介した映像予測と 3DGS 圧縮に加え、可変レート符号化、高速演算、三次元点群符号化、コンピュータビジョン統合等、深層学習を用いた画像符号化（学習型画像符号化）に関する各種の成果発表を行った。また、映像と三次元点群の伝送に関する基盤技術としての成果発表も進めている。現在、マルチモーダル大規模言語モデルに代表されるように、深層学習の意味解析や情報生成に関する技術開発の進展が著しく、学習型符号化においても、信号レベルの符号化に加え、意味レベルの符号化の開発を加速する。並行して、セマンティックな情報伝送を支える通信技術の開発を進めて行く。