

公益財団法人矢崎科学技術振興記念財団
奨励研究助成 成果報告書

公益財団法人矢崎科学技術振興記念財団
理事長 殿

研究助成期間終了にあたり、下記の通り成果を報告します。

2025 年 6 月 6 日

氏名 孫 鶴鳴

所属 横浜国立大学 大学院工学研究院

職位 准教授

1. 申請研究の題目

アルゴリズムとアーキテクチャの協調最適化による学習型画像圧縮システム

2. 研究の目的

画像圧縮は、画像の保存・再利用および伝送にかかる負担を軽減するために不可欠な技術である。従来より、画像圧縮規格は 30 年以上にわたって開発が進められてきたが、近年 7 年間ではニューラルネットワークを用いた学習型画像圧縮が急速に進展し、PSNR や MS-SSIM など多くの指標において従来の国際標準規格を上回る性能を示している。

一般に、ニューラルネットワークを活用することで圧縮率の向上が可能となる一方で、演算量が非常に大きくなるという課題がある。こうした処理をエネルギー効率よく高速化するには、ハードウェアアクセラレーションが不可欠である。さまざまなアクセラレーターの中でも、以下の理由から FPGA による実装が最も有望である。汎用アクセラレーターである GPU と比べて、FPGA はハードウェア資源の利用効率が高く、専用アクセラレーターである ASIC と比べても再構成可能で柔軟性が高く、新たな手法にも適用しやすい。

FPGA は理論上、再構成可能性によってあらゆるハードウェアアルゴリズムに対応可能であるが、実際にはアルゴリズムによってはレイアウト(配置配線)処理が完了しないことも少なくない。FPGA チップには LUT や BRAM、DSP といったリソースが固定配置されており、データフローがこれらの配置と適合しない場合、局所的な配線混雑や高ファンアウトといった問題が発生する。このような問題を回避するためには、FPGA に適したアルゴリズム設計が求められるが、同時にアルゴリズム本来の性能を損なわないことも重要な課題である。

本研究では、アルゴリズムと FPGA アーキテクチャを個別に設計してきた従来研究とは異なり、両者を連携させた協調的な最適化によって、最適な FPGA 実装型学習型画像圧縮システムの実現を目指す。

3. 研究の内容

本研究の内容は以下の二つである。

1. ハードウェアのチャンネルの並列性に制約を加えることで、DSP の利用率を向上させる。配置配線フェーズでボトルネックの一つである活性化バッファの LUT を削減することで、配線エラーを発生させることなく、より多くの DSP を使用することができる。
2. ニューラルネットワーク探索により DSP 効率を向上させる。各畳み込みニューラルネットワークレイヤーのチャンネル数を調整することで、割り当てられた DSP はより高い効率(すなわちクロックサイクルのバブルが少ない)で動作することができる。

4. 研究の成果と結論、今後の課題

ニューラルネットワーク探索の結果、分析変換および合成変換におけるチャンネル数は表 I に示されている。分析変換と合成変換にはそれぞれ 5 つの層があり、最初の 4 層のチャンネル数は、分析変換と合成変換でそれぞれ{128, 128, 128, 192}および{192, 128, 128, 128}である。ZC706 における最適化後のチャンネル数は、分析変換が {128, 120, 128, 190}、合成変換が {198, 124, 120, 124} である。KU115 の場合は {128, 136, 128, 184} および {192, 112, 144, 128} である。VCU118 の場合、結果は {120, 120, 144, 192} および {192, 128, 128, 128} となっている。なお、VCU118 の合成変換に関しては、探索されたネットワークの FPS が元のネットワークと同一であるため、ネットワークの調整は行っていない。

表 I: ニューラルネットワーク探索の結果

方法	分析変換のチャンネル数	合成変換のチャンネル数
参考文献[1]	128, 128, 128, 192	192, 128, 128, 128
ZC706 向け提案	128, 120, 128, 190	198, 124, 120, 124
KU115 向け提案	128, 136, 128, 184	192, 112, 144, 128
VCU118 向け提案	120, 120, 144, 192	192, 128, 128, 128

続いて、探索されたネットワークの圧縮性能を表 II にて評価する。Kodak データセットを用いた評価において、ZC706 における提案ネットワークと [1] を比較すると、ビットレート(bpp)が低いほど PSNR が悪化している。一方、KU115 と [1] を比較すると、PSNR の向上および bpp の低減が確認できる。VCU118 の傾向は ZC706 と同様である。さらに、より明確な比較のために R-D コストのメトリックも提示する。ZC706 および KU115 における探索ネットワークは、R-D コストの点で [1] よりも優れていることがわかる。VCU118 の探索ネットワークは、R-D コストにおいて [1] よりやや劣っている。

表 II: 圧縮性能の比較

方法	PSNR (dB) ↑	bpp ↓	R-D cost ↓
参考文献[1]	29.94	0.2542	0.6373
ZC706 向け提案	29.86	0.2434	0.6344
KU115 向け提案	30.06	0.2531	0.6227
VCU118 向け提案	29.73	0.2367	0.6424

また、[1] とのハードウェア速度性能の比較を表 III にて評価する。提案手法は、きめ細かなパイプライン・アーキテクチャを採用した従来研究 [1] と比較して、最大で 1.5 倍高速なスループットを達成できることを示している。

表 III: ハードウェア速度性能の比較

Board	Enc/dec	参考文献[1]	提案
ZC706	Enc	360P@25.68FPS	360P@33.91FPS
	Dec	360P@28.55FPS	360P@35.00FPS
KU115	Enc	720P@25.04FPS	720P@37.74FPS
	Dec	720P@24.66FPS	720P@30.14FPS
VCU118	Enc	1080P@19.83FPS	1080P@25.53FPS
	Dec	1080P@20.42FPS	1080P@25.53FPS

また、Xilinx DPU により高速化された従来研究 [2] と比較して、我々の手法は、より優れた圧縮性能とより高速なスループットを両立できる。[2] との比較結果を表 IV に示す。エンコードにおい

ては 1.98 倍、デコードにおいては 8.51 倍のスループットを実現している。圧縮性能に関しても、2 種類の圧縮品質(ラムダ)において [2] を上回る結果が得られている。特に、ラムダが 0.01 の場合には、より高い PSNR とより低い bpp を達成している。

表 IV: [2]との圧縮性能およびハードウェア性能比較

Method	Proposal	参考文献[2]
Platform	VCU118	ZCU104
Technology	16nm	16nm
Encoding FPS $\uparrow$	25.53	12.87
Decoding FPS $\uparrow$	25.53	3.00
PSNR ($\lambda=0.005$) $\uparrow$	29.73	29.89
bpp ($\lambda=0.005$) $\downarrow$	0.2367	0.2565
PSNR ($\lambda=0.01$) $\uparrow$	31.75	31.50
bpp ($\lambda=0.01$) $\downarrow$	0.4055	0.4165

今後の課題としては、探索空間を拡大し、異なる層とチャンネルを持つより多くのニューラルネットワークを探索する計画である。

参考文献

- [1] H. Sun, Q. Yi, F. Lin, L. Yu, J. Katto, and M. Fujita, “F-lic: Fpga-based learned image compression with a fine-grained pipeline”, IEEE Asian Solid-State Circuits Conference (A-SSCC), pp. 1–3, Nov. 2022.
- [2] C. Jia, X. Hang, S. Wang, Y. Wu, S. Ma, and W. Gao, “Fpx-nic: An fpga accelerated 4k ultra-high-definition neural video coding system”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, pp. 6385–6399, Sept. 2022.

5. 成果の価値

5-1. 学術的価値

実際、本手法であるアルゴリズムとアーキテクチャの協調最適化の方法論は、本研究で扱った学習型画像圧縮に限らず、他のアプリケーションにも適用可能である。

さらに、本研究の発展は、JPEG-AI や MPEG-VCM の標準化にも貢献し得る。

5-2. 社会的価値

次世代のマルチモーダル AI 社会においては、動画の通信量が爆発的に増加すると予測されており、その課題解決に向けて、高効率な圧縮技術の導入により通信量および消費電力を大幅に削減することが期待されている。このように、本研究は、カーボンニュートラル社会の実現にも大きく貢献し得る重要な分野を切り拓いている。

6. 研究成果

・「研究論文(原著)」

- [1] **Heming Sun**, Qingyang Yi, Masahiro Fujita, “FPGA Codec System of Learned Image Compression with Algorithm-Architecture Co-Optimization”, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol. 14, No. 2, pp. 334–347, June 2024.
- [2] Zhibo Chen, **Heming Sun**, Li Zhang, Fan Zhang, “Survey on Visual Signal Coding and Processing with Generative Models: Technologies, Standards and Optimization”, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol. 14, No. 2, pp. 149–171, June 2024.

・「国際会議発表」

[3] Heming Sun, Jing Wang, Silu Liu, Shinji Kimura, Masahiro Fujita, “Learned Image Codec on FPGA: Algorithm, Architecture and System Design”, Asia and South Pacific Design Automation Conference (ASPDAC), pp. 328–331, Jan. 2025.

・「受賞」

[4] 孫鶴鳴、若手科学者賞、文部科学省、2024 年度

[5] 孫鶴鳴、2023 年度最優秀発表賞、情報処理学会システムと LSI の設計技術研究会、2024 年度

以上