

C-006

FPGA 実装に向けた超解像 CNN のパラメータ数削減 Super-Resolution CNN with Fewer Parameters for FPGA Implementation

森 一紀[†] 黒木 修隆[†] 廣瀬 哲也^{††} 沼 昌宏[†]
Kazuki Mori Nobutaka Kuroki Tetsuya Hirose Masahiro Numa

1. はじめに

近年、物体認識や画像・映像の高画質化を畳み込みニューラルネットワークにより実現する技術が注目を集めている。膨大な計算量を必要とする畳み込みニューラルネットワーク (CNN: Convolution Neural Network) [1] の処理を高速に実行するためのアクセラレータとして、従来は GPGPU (General Purpose Graphic Processing Unit) が一般的に利用されていたが、汎用プロセッサのため無駄な処理が多く、消費電力が大きいという問題があった。そこで、GPGPU の約 10 分の 1 程度の消費電力で動作可能な FPGA 上にニューラルネットワーク専用の回路を実装することで、低消費電力化を実現する研究が注目されている。FPGA 上の限られたリソースで実装可能な超解像 CNN として、Y. Kim らが提案した HF-SRN (Hardware-Friendly Super-Resolution Network) [2] が、少ないパラメータ数で超解像を実現している。しかし、他のネットワーク構造と比べて層が浅く受容野が狭いため、画質が劣化する問題がある。

そこで本稿では、FPGA 上のリソースで実装可能な規模で高い PSNR が得られる超解像ネットワークを提案する。まず、HF-SRN の 1 層目の畳み込みのフィルタサイズを 3×3 から 5×5 にすることで、提案するネットワークの受容野 (receptive field) の拡大を行い、PSNR の向上を図る。次に、 5×5 のフィルタを用いたことによるパラメータ数の増大を抑制するため、 2×2 のフィルタを用いた多層畳み込みによって、FPGA で実装可能な大きさのネットワークに規模を縮小する。

2. 多層畳み込みに基づくパラメータ数削減

2.1 $r \times r$ の受容野をもつ畳み込み層

図 1 (a) に 1 層の畳み込みの概略を示す。1 層の畳み込みで、 $r \times r$ サイズの受容野を得るには、 $r \times r$ のフィルタを用いる必要がある。 $r \times r$ のユニットを 1 つのユニットに結合するため、各結合の重みとして $r \times r$ 個のパラメータを必要とする。パラメータはフィルタサイズ分存在するため、パラメータ数は r^2 である。

図 1 (b) に 2×2 のフィルタを用いた多層畳み込みの概略を示す。 $r \times r$ のユニットが 2×2 の畳み込みを行うことで、 $(r-1) \times (r-1)$ のユニットに結合される。これを $r-1$ 層繰り返す

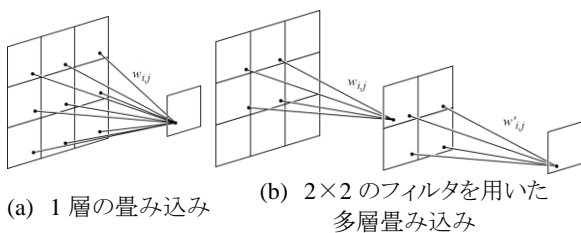


図 1 畳み込みのユニットの結合

[†] 神戸大学, Kobe University

^{††} 大阪大学, Osaka University

すことで、 $r \times r$ の受容野を得ることが可能である。 2×2 のユニットが 1 つのユニットに結合されており、この 4 つの結合それぞれに重みが存在する。この結合が層数分存在するため、パラメータ数は $2^2 \times (r-1)$ である。

2×2 のフィルタを用いた多層畳み込みは、1 層の畳み込みと同じ受容野をもちながら少ないパラメータ数で構成可能であり、層の深いネットワークを構築することができる。

2.2 提案ネットワーク

図 2 に提案するネットワーク構造を示す。1 層目の畳み込みのフィルタサイズに関して、HF-SRN では 3×3 としているが、画質の向上を目的とした受容野拡大のため、提案ネットワークでは 5×5 のフィルタを適用する。フィルタのサイズを縦横 2 ずつ拡大したことにより、受容野を 7×15 から 9×17 まで拡大することが可能となる。また、 5×5 と 3×3 の畳み込み層に対して、2.1 節で述べた 2×2 のフィルタを用いた多層畳み込みを用いる。これにより、HF-SRN と比べて広い受容野をもちながら、 5×5 のフィルタを用いた場合と比べてパラメータ数の少ないネットワークを実現する。また、 2×2 の多層畳み込みの導入による畳み込み層の増加に伴い、HF-SRN と比べて深いネットワークを構成する。

2.3 フィルタサイズの異なる畳み込み回路

本節では、 5×5 の受容野をもつ畳み込み回路の、異なるサイズのフィルタによる構成方法について述べる。本回路で想定する入力には $1,920 \times 1,080$ pixel, 60 fps の映像の Y 成分 (16 bit) とする。そのため、148.5 MHz で動作可能な回路を設計する。また、従来手法で不足していた DSP のリソースは乗算部分に対してのみ用いることで、レイテンシが小さい回路を設計した。

シリアル・パラレル変換部では、シリアル入力をフィルタサイズに応じたパラレル入力へ変換する。図 3 に積和演算部の

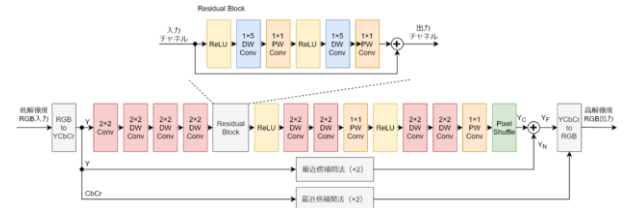


図 2 提案ネットワーク

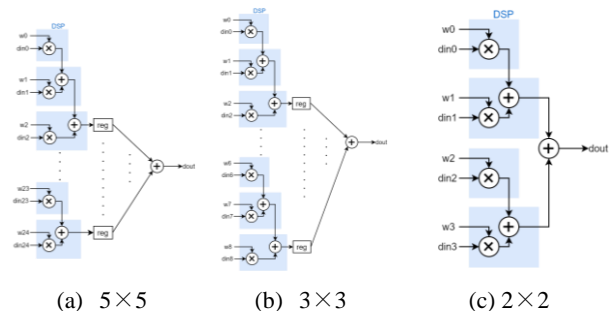


図 3 畳み込みユニットの結合

表 1 ネットワーク構造に対する PSNR 評価結果

評価対象手法	パラメータ数	各評価用データセットに対する PSNR [dB]			
		Set5	Set14	General100	Urban100
Bicubic 法	—	33.67	30.03	33.17	26.86
従来手法	2,560	36.57	32.11	36.05	29.31
従来手法 + 2×2 の畳み込み	2,480	36.57	32.14	36.07	29.28
従来手法 + 受容野の拡大	3,098	36.59	32.14	36.09	29.37
提案手法	2,736	36.67	32.18	36.19	29.39

回路を示す。積和演算部では、入力データと重みとの乗算を行い、それらの結果を足し合わせる。DSP は積和演算が可能なリソースであり、LUT を消費せずに乗算と加算が可能である。そこで、乗算結果の加算を DSP 内部で行うことにより、加算に必要な LUT を削減している。また、従来手法は DSP のリソース利用率が高いという問題があるため、DSP の外で行う加算は LUT を用いて行う。

5×5 の受容野を得るため、3×3 のフィルタは 2 層、2×2 は 4 層接続されている。

3. シミュレーション評価と考察

3.1 実験内容

本節では、HF-SRN をもとに 2×2 のフィルタを用いた多層畳み込みと受容野の拡大を行った提案手法に関して、HF-SRN を従来手法としてシミュレーションによる比較評価を行う。また、HF-SRN の 3×3 のフィルタをもつ畳み込み層すべてに 2×2 のフィルタを用いた多層畳み込みを適用したネットワーク構造と、1 層目の畳み込み層のフィルタを 3×3 から 5×5 にすることにより受容野を拡大したネットワーク、Bicubic 法についても評価を行った。なお、超解像は画像の Y 成分(輝度成分)にのみ適用し、評価も Y 成分のみについて行う。元画像を Bicubic 法により 1/2 にダウンサンプリングした画像を入力し、各ネットワークにより 2 倍拡大した画像を出力する。評価項目は、主観的評価とパラメータ数、PSNR とする。

さらに、2×2 のフィルタを用いた多層畳み込みに基づく畳み込み回路に関して、FPGA に実装した際のリソース数を評価する。

3.2 結果と考察

表 1 に PSNR とパラメータ数に関する比較結果を示す。提案手法では、2×2 のフィルタを用いた多層畳み込みを導入したことにより、従来手法と同等の PSNR を保ちつつ、パラメータ数を削減可能であることを確認した。

提案手法は、受容野の拡大のみを導入した手法と同じ受容野のサイズでありながら、すべてのデータセットにおいて PSNR が 0.02 dB から 0.1 dB 向上した。提案手法は、受容野の拡大のみを導入した手法と比べて層の深いネットワーク構造であることが、PSNR の向上に貢献したと考えられる。

受容野の拡大のみを行った手法では、パラメータ数が従来手法と比べ 21% 増加した。提案手法は、受容野の拡大を導入した手法と同じ受容野のサイズを確保しつつ、パラメータ数の増加率は従来手法と比べて 6.9% に留まった。この結果から、2×2 の多層畳み込みを用いることで、受容野の拡

表 2 畳み込み回路のマッピング結果

畳み込み回路の フィルタサイズ	LUT	レジスタ	BRAM	DSP
2×2	64	240	4	16
3×3	32	272	4	18
5×5	88	340	4	25

大に伴うパラメータ数の増加を抑制できることを確認した。

表 2 に、フィルタサイズの異なる各畳み込み回路に対するマッピング結果を示す。2×2 のフィルタを用いた多層畳み込みを用いることで、5×5 のフィルタを用いた場合と比較して LUT が約 27%、レジスタが約 29%、DSP が約 28% 削減可能であることを確認した。また、3×3 のフィルタと比べてレジスタが約 12%、DSP が約 11% 削減可能であることを確認した。その一方で、LUT については 100% の増加が確認された。しかし、HF-SRN の FPGA 実装時の LUT 利用数は 110,000 であり、実装を想定している FPGA の LUT 搭載数は 607,200 であることから、LUT の利用率は高くなるが FPGA 実装は可能であると考えられる。

4. まとめ

本稿では、CNN を用いたリアルタイム超解像処理の FPGA 上での実現を目的として、FPGA の限られたリソースで実装可能な、パラメータ数の少ない超解像 CNN のネットワーク構造を提案した。

提案したネットワークに対して実験評価を行った結果、従来手法に対して高い PSNR が得られる効果を確認した。また、2×2 のフィルタを用いた多層畳み込みを導入することで、受容野拡大にともなうパラメータ数の増加率が抑えられていることを確認した。さらに、設計した畳み込み回路に対する評価の結果、2×2 のフィルタを用いた多層畳み込み導入によるリソース削減効果を確認した。

今後の課題として、より品質の高いネットワークの構築や、提案ネットワーク全体の FPGA 実装が挙げられる。

参考文献

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
- [2] Y. Kim, J.-S. Choi, and M. Kim, "A real-time convolutional neural network for super-resolution on FPGA with application to 4K UHD 60fps video services," IEEE Transactions on Circuits and Systems for Video Technology, 2018.