C-002

# Separable 畳み込みを用いた単眼深度推定ネットワークの軽量化
# Lightweight Monocular Depth Estimation Network Using Separable Convolution

沼田 和樹† 黒木 修隆† 沼 昌宏†
Kazuki Numata Nobutaka Kuroki Masahiro Numa

## 1. はじめに

近年，2 次元の画像や映像から 3 次元位置を推定する技術について，娯楽など一部の分野・用途で実用化が進み，ビジネスや医療，自動運転など様々な分野での実用化に向けて研究が進んでいる。画像から 3 次元位置を推定するためには，既存の画像情報に加えて奥行きを表す深度情報が必要となる。深度情報の推定技術としては，1 つのカメラにより撮影された 1 枚の画像のみから深度情報を推定する単眼深度推定技術が注目されている。なかでも，Laina らが提案した畳み込みニューラルネットワーク（CNN：Convolutional Neural Network）を用いた Fully Convolutional Residual Network（FCRN）[1] という手法が，高い単眼深度推定精度を実現している。FCRN の処理実行には膨大な演算量が必要であり，その処理を高速化するアクセラレータとして GPU（Graphic Processing Unit）が用いているが，消費電力が大きいという問題がある。そこで低消費電力化のために，FPGA 上のハードウェアとしての実装が考えられる。FPGA は GPU より消費電力が小さく，書き換えによって環境に応じた深度推定の実現が可能であると見込まれる。その一方で，FPGA は実装可能な回路規模に制限があるため，実装対象とする FCRN の軽量化が必要である。

そこで本稿では，空間方向とチャネル方向とで別個に畳み込みを行うことで，一般的な畳み込みよりも少ない演算量で特徴の抽出が可能である Separable 畳み込みを用いて，演算量を低減した軽量な単眼深度推定ネットワークを提案する。

## 2. 提案手法

### 2.1 Separable 畳み込み

図 1 に Separable 畳み込みの概要を示す。通常の畳み込みが，入力特徴マップの空間方向とチャネル方向に同時に畳み込みを行うのに対して，Separable 畳み込みは空間方向とチャネル方向とで別個に畳み込みを行うことで，計算量を減らすことが可能となる。空間方向の畳み込みは Depth-wise 畳み込みと呼ばれ，単一の入力特徴マップから出力特徴マップを計算する。チャネル方向の畳み込みは Pointwise 畳み込みと呼ばれ，通常の 1×1 の畳み込みと同様の畳み込み演算を行う。

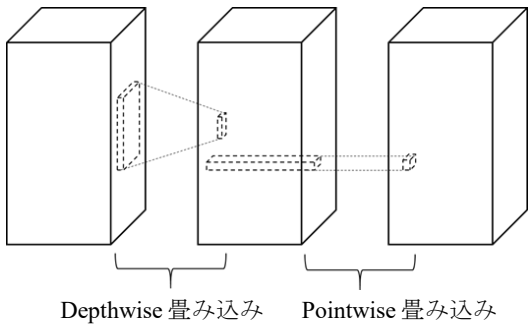



図 1　Separable 畳み込みの概要

Separable 畳み込みの演算量削減効果について説明する。$K$ はカーネルサイズ，$H$，$W$ は特徴マップの高さと幅，$C_{\text{in}}$，$C_{\text{out}}$ はそれぞれ入力特徴マップ数と出力特徴マップ数とする。一般的な畳み込み処理における積和演算数 $N_{\text{conv}}$ は，

$$N_{\text{conv}} = K^2 HWC_{\text{in}}C_{\text{out}} \quad (1)$$

で表される。一方，Separable 畳み込みによる積和演算数は，

$$N_{\text{conv}} = K^2 HWC_{\text{in}} + HWC_{\text{in}}C_{\text{out}} \quad (2)$$

で表される。これらより，$K$×$K$ の Separable 畳み込みは従来の $K$×$K$ の畳み込みと比較して，積和演算数とパラメータ数が削減される。例えば，$K$ = 5，$C_{\text{out}}$ = 512 の畳み込みに Separable 畳み込みを適用すれば，パラメータ数を約 1/25 に削減可能である。

### 2.2 提案するネットワーク構造

従来手法と提案手法のネットワーク構造をそれぞれ図 2，図 3 に示す。提案手法は従来手法のネットワーク構造をもとに，ネットワークの前半部分を ResNet-50 [2] から Separable 畳み込みを用いた軽量な CNN である MobileNet [3] に置き換え，また，Up-Projection 層の畳み込み演算部分を Separable 畳み込み（提案 Up-Projection）とした構造である。さらに，Up-Projection 後の 3 層の通常の畳み込み層についても，演算量削減のため Separable 畳み込み層に置き換えている。

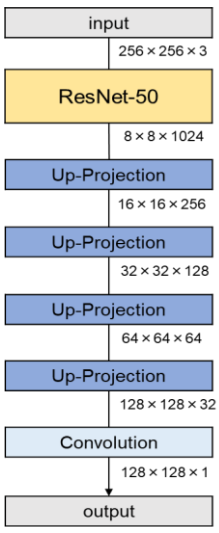



図 2　従来手法の概要

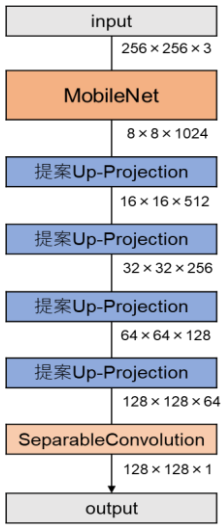



図 3　提案手法の概要

† 神戸大学，Kobe University

表 1　評価関数の結果とパラメータ数

| 評価手法 | RMS | Thresholded accuracy | | | パラメータ数 |
|---|---|---|---|---|---|
| | | $\delta$ <1.25 | $\delta$ <1.25$^2$ | $\delta$ <1.25$^3$ | |
| 従来手法 | 0.780 | 0.733 | 0.935 | 0.982 | $5.27\times10^7$ |
| 提案手法 | 0.788 | 0.732 | 0.945 | 0.987 | $5.10\times10^6$ |

## 3. 評価実験と考察

### 3.1　実験内容

従来手法と提案手法に対する評価実験を行った。評価項目は 2 種類の評価関数（RMS，Thresholded accuracy）と主観的評価である。RMS は数値が低い程，全体的に誤差が少ないことを示す。Thresholded accuracy は数値が高い程，誤差の大きい画素の割合が少ないことを示す。

### 3.2　実験結果と考察

表 1 に，各手法による単眼深度推定の評価関数の結果と，モデルのパラメータ数を示す。精度について，提案手法と従来手法とで 2 種類の評価関数いずれも 0.1 pt 以内の差でほぼ変わらない結果となった。

次にパラメータ数について，提案手法は従来手法と比較して $5.27\times10^7$ から $0.51\times10^7$ へと約 1/10 に削減し，かつ同程度の精度を実現できていることを確認した。

図 4 に深度推定結果の例を示す。提案手法の結果は，従来手法と比較して，全体的な深度推定に関しては大きな誤差がないということを確認した。これにより，提案手法は従来手法に対して大きな精度低下は無く，深度推定が可能であることを確認した。

### 3.3　FPGA 実装に向けた課題に関する考察

本稿では単眼深度推定の軽量化ネットワークの提案を行ったが，実際の FPGA 実装を見据えた回路設計や利用リソース数の算出には至っていない。そこで本節では，佐田ら [4] による単眼深度推定を FPGA に実装した結果と比較して，FPGA 実装に向けた今後の検討項目を述べる。

本稿の提案モデルの積和演算数は $2.28\times10^9$ である。一方，佐田らの提案モデルの積和演算数は圧縮前で $5.67\times10^8$ であり，量子化等により圧縮を行うことで FPGA 実装を達成している。このことより，本稿の提案モデルは量子化等により圧縮を図ったとしても，実装できる FPGA の利用リソース数に収まる可能性が低い。そこで，さらなる軽量化が必要と考えられる。

検討項目としては，ネットワーク後半部分のさらなる軽量化が挙げられる。本稿を通して Up-Projection への Separable 畳み込み適用が，精度を維持しつつパラメータ数を削減することに大きく貢献していることが明らかとなった。しかし，Up-Projection の積和演算数は依然として大きい。このことから，ネットワーク後半のさらなる軽量化によって，精度低下を抑えつつパラメータ数を削減可能であると考えられるため，今後の課題とする。

## 4. まとめ

本稿では，FPGA 実装に向けた単眼深度推定ネットワークの軽量化を目的として，精度低下を抑えつつ演算量を低減できる Separable 畳み込みを用いて，軽量な単眼深度推定ネットワークを提案した。評価実験を行った結果，提案手法は従来手法と比較して，RMS，Thresholded accuracy の 2 種類の評価関数のいずれについても同等な精度を保ちつつ，パラメータ数を約 1/10 に削減することができた。また，主観的評価においても提案手法の精度維持を確認した。

提案手法のネットワークでは，まだ演算量が多くネットワーク全体を FPGA 実装することは難しいと考え，FPGA 実装に向けて考察を行った。今後は，その考察をもとに，ネットワーク後半のさらなる軽量化手法について検討することを課題とする。

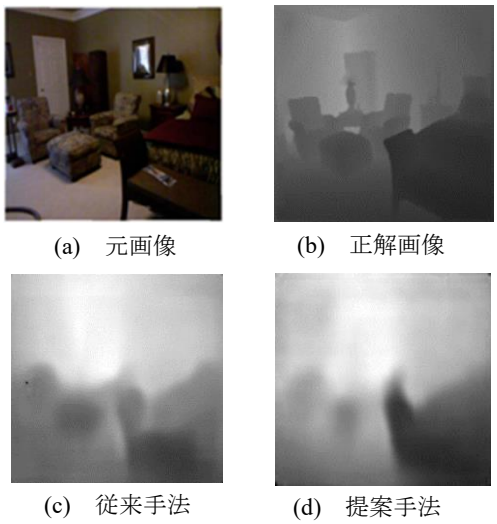
### 参考文献

[1] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, N. Navab, “Deeper depth prediction with fully convolutional residual networks,” 2016 Fourth International Conf. on 3D Vision, pp. 239-248, 2016.

[2] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition,” arXiv preprint arXiv:1512.03385, 2015.

[3] A. G. Howard, M. Zhu, B. Che, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “MobileNets: efficient convolutional neural networks for mobile vision application,” arXiv:1704.0486, 2017.

[4] 佐田悠生，下田将之，佐藤真平，中原啓貴，“畳み込みニューラルネットワークを用いた単眼深度推定の FPGA 実装,” 電子情報通信学会技術研究報告, vol. 119, no. 371, pp. 73-78, 2020 年 1 月.



(a)　元画像　(b)　正解画像

(c)　従来手法　(d)　提案手法

図 4　深度推定結果