

Skip-Connection と Separable 畳み込みを用いた CNN による画像の単眼深度推定 Monocular Depth Estimation by CNN using Skip-Connection and Separable Convolution

吉田 昇馬[†] 黒木 修隆[†] 廣瀬 哲也[‡] 沼 昌宏[†]
Shoma Yoshida Nobutaka Kuroki Tetsuya Hirose Masahiro Numa

1. はじめに

近年、被写体の 3 次元情報を復元するために、画像の深度 (奥行き) 情報を推定する手法が研究されている。なかでも、深度情報を撮影済みの 1 枚の画像のみから推定する、単眼深度推定という技術に需要が見込まれている。従来の単眼深度推定手法として Fully Convolutional Residual Networks (FCRN) を用いた手法が挙げられる。しかし、従来手法では推定結果である Depth 画像の輪郭部分が曖昧であるという問題がある。そこで本稿では、より精度の高い単眼深度推定手法の実現を目的に、Skip-Connection と Separable 畳み込みを導入したニューラルネットワーク (CNN) による単眼深度推定手法を提案する。

2. CNN による単眼深度推定

まず、従来手法で用いられている FCRN [1] の構成を図 1 に示す。FCRN はネットワークの前半部分で入力画像の特徴を抽出し、後半部分で抽出した特徴から深度情報を推定し、Depth 画像として出力する。ネットワークの前半部分は、2015 年の画像認識の精度を競う ILSVRS [2] で非常に優秀な成績を残した、ResNet-50 [3] と呼ばれるネットワークを用い、後半部分は、Up-Projection [1] と呼ばれる、特徴マップを二種類の方法で拡大し、足し合わせる手法を用いる。

2.1 問題点

従来手法では、中間層の解像度が低いため、物体の輪郭部分の深度情報が曖昧となることや、細い物体の深度情報が推定できないという問題点がある。

3. 提案手法

より精度の高い単眼深度推定手法の実現を目的に、Skip-Connection と Separable 畳み込みを導入した CNN による単眼深度推定手法を提案する。

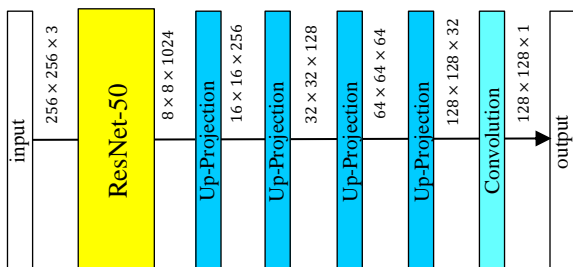


図 1 FCRN の構造

[†] 神戸大学大学院工学研究科,
Graduate School of Engineering, Kobe University
[‡] 大阪大学大学院工学研究科,
Graduate School of Engineering, Osaka University

3.1 Skip-Connection

Skip-Connection とは、ネットワークの前半部分の特徴マップを後半部分のサイズの等しい特徴マップに連結する手法である。CNN の特徴マップは CNN の層が浅いほど画像の局所的特徴を多く含み、層が深くなるにつれ画像の全体的特徴を多く含む。Skip-Connection により、特徴マップの拡大時に全体的画像の特徴に併せて画像の局所的特徴を考慮した学習が可能となり、従来手法における物体の境界部分の曖昧さを低減する効果が期待できる。

3.2 Separable 畳み込み

図 2 に、Separable 畳み込みの概要図を示す。Separable 畳み込みとは、空間方向とチャンネル方向に独立に畳み込みを行う手法である。通常の畳み込みが、入力特徴マップの空間方向とチャンネル方向に対し同時に畳み込みを行うのに対し、空間方向とチャンネル方向に独立に畳み込みを行うことで通常の畳み込みよりも計算量を減らすことが可能となり、その余力でフィルタの数と大きさを増やすことが可能となる。

3.3 提案手法のネットワーク構造

提案手法のネットワーク構造を図 3 に示す。提案手法は従来手法のネットワーク構造を基に、ネットワークの前半部分を ResNet-50 から Separable 畳み込みを用いたネットワークである Xception [4] に置き換え、また、Xception の中間層から対応する Up-Projection 層に向けて Skip-Connection を追加した構造である。

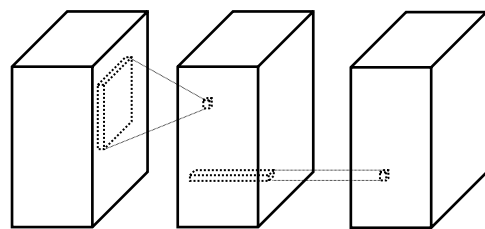


図 2 Separable 畳み込みの概要図

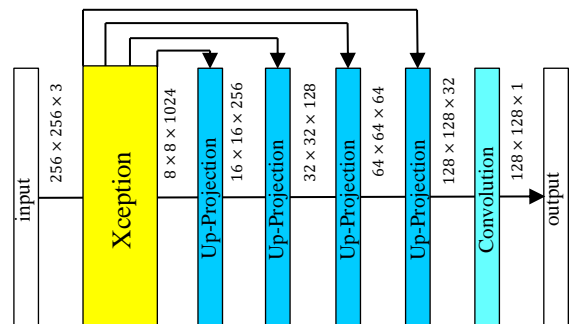


図 3 提案手法の構造

表 1 深度の推定結果に対する評価関数の平均値

評価手法	rel	rms	Thresholded accuracy		
			$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
従来手法	0.172	0.778	0.719	0.934	0.983
提案手法	0.153	0.707	0.783	0.958	0.990

4. 評価実験

4.1 実験内容

FCRN を従来手法として提案手法との比較を行った。学習には、単眼深度推定用画像データセットである NYU Depth Dataset v2 [5] を用いた。NYU Depth Dataset v2 は Microsoft Kinect で撮影された屋内での写真と Depth 画像のペアであり、学習用画像は 1160 枚、評価用画像は 289 枚とし、学習回数は 10 万回とした。評価項目は主観的評価と後述する評価関数 3 種である。

4.2 評価関数

評価に用いた 3 種の評価関数について示す。ここで、 y 、 y' はそれぞれ正解 Depth 画像、出力 Depth 画像とする。相対誤差、二乗平均平方根は数値が低いほど精度が高く、Thresholded accuracy [1] は数値が高いほど精度が高い。

4.2.1 rel (相対誤差)

$$\text{rel} = \frac{1}{n} \sum \frac{|y - y'|}{y} \quad (1)$$

4.2.2 rms (二乗平均平方根)

$$\text{rms} = \sqrt{\frac{1}{n} \sum (y - y')^2} \quad (2)$$

4.2.3 Thresholded accuracy

Thresholded accuracy は以下の式を満たす画素の割合で表される。

$$\max\left(\frac{y}{y'}, \frac{y'}{y}\right) = \delta < t \quad t \in [1.25, 1.25^2, 1.25^3] \quad (3)$$

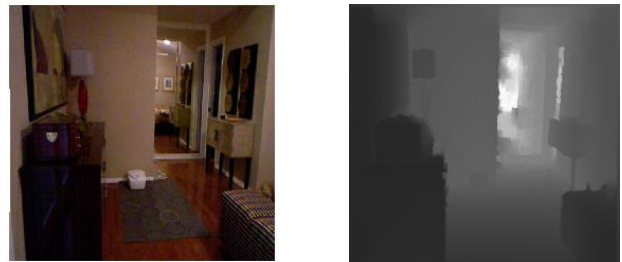
4.3 実験結果

表 1 に各手法の評価結果を示す。3 種の評価関数全てにおいて提案手法が最も良い結果となった。

図 4, 5 に深度の推定結果を示す。提案手法において Depth 画像における物体の輪郭が明瞭となったことを確認した。これにより、提案手法がより高精度な深度推定ができていないことを確認した。

5. まとめ

本稿では、従来手法において発生していた物体輪郭部分に見られる曖昧さを改善し、より高精度な単眼深度推定を行うことを目的に、Skip-Connection と Separable 畳み込みを導入した CNN を提案した。その結果、評価値の向上が確認できた。また、主観的評価においても従来手法よりも提案手法がより高精度な深度推定ができていないことを確認した。今後は、屋外のシーンへの対応を検討する予定である。



原画像

正解画像



従来手法



提案手法

図 4 深度の推定結果 (i)



原画像

正解画像



従来手法



提案手法

図 5 深度の推定結果 (ii)

参考文献

- [1] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," *The Fourth International Conference on 3D Vision (3DV)*, 2016, pp.1119-1127, 2016.
- [2] Large Scale Visual Recognition Challenge (LSVRC), <http://www.image-net.org/challenges/LSVRC/>.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, 2016.
- [4] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1251-1258, 2017.
- [5] "NYU Depth Dataset V2", https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html