4K-01

FPGAによるFast and Accurate Networkを用いた深度推定処理の性能評価

田　寿藝　　大川　猛

東海大学情報通信学部組込みソフトウェア工学科

１　はじめに

深度推定は自動車やロボットの深度センサを用いた 3D イメージングや環境マッピングにおいて高い注目を集めている[1]。それらの技術は、物体や環境の高精度な 3D モデルを作成することで、ロボットや自動車のナビゲーションや地図作成などに活用される。深度推定には、レーザーやカメラを用いたものがあり、ニューラルネットワークを用いる手法も研究されている。そのため、高速で低消費電力な深度推定処理が求められ、FPGA を用いたハードウェアアクセラレーション技術が注目されているが、FPGA によるニューラルネットワークを用いた深度推定処理の性能および精度の報告に乏しい[2][3]。

そこで、本研究ではニューラルネットワックを用いて深度推定モデルを生成し、FPGA にデプロイして性能を評価し、更に二つの深度推定法の精度を評価する。

２　FADNet深度推定ニューラルネットワーク

FPGA による深度推定ニューラルネットワークはいくつか提案されているが、FADNet (Fast and Accurate Network) [3]は、効率と精度のバランスが取れた視差推定モデルであり、メモリ使用量が少ない。FADNet は以下の 3 つの手順で FPGA による深度推定処理を行う.

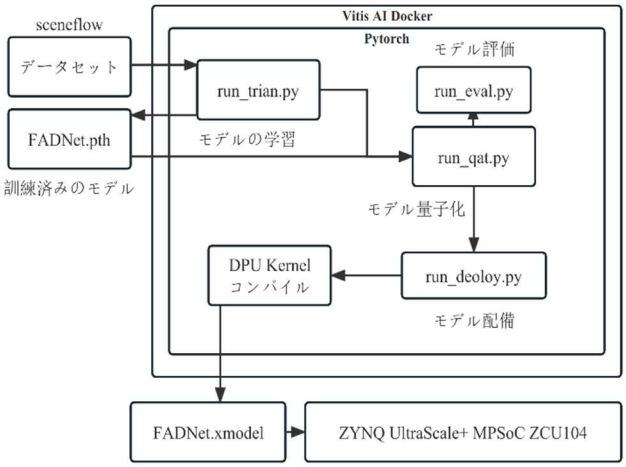
- 手順１．深度推定モデルの学習（PC 環境）
- 手順２．量子化・コンパイル（PC 環境）
- 手順３．深度推定処理（FPGA）

手順１では、SceneFlow[4]のステレオ対応および光流アルゴリズムのトレーニングのデータセットを準備し、Xilinx 社 Vitis AI[5]の pytorch 環境で FADNet 深度推定処理のモデル学習を行う。

手順２では Vitis AI で、訓練済みのモデルを量子化して、FPGA 向けにコンパイルを行う。

手順３では作成したモデルを FPGA に転送して、深度推定を行う。Vitis AI 開発ツールには、ニューラルネットワークモデルの量子化ツールと FPGA で運用・実行できるコンパイルツールが含まれる。

図１に Vitis AI による FADNet 開発環境の構成を示す。GPU を利用可能な docker を作り 、Pytorch 環境を構築する。そして、FADNet のニューラルネットワークを使って、データを訓練する。最後、量子化、コンパイルをする。



【図１】Vitis AI による FADNet 開発環境の構成

３　評価

(1) FADNet モデル処理性能評価

前節の手順で生成した FADNet モデルの処理性能評価を行った。まず、作成したモデルを Zynq UltraScale+ MPSoC ZCU104 評価キットにデプロイし、左右の視差画像を入力して、深度推定の処理時間を測定した。評価に用いたＰＣ環境を表１に、モデルの性能評価結果を表２に示す。

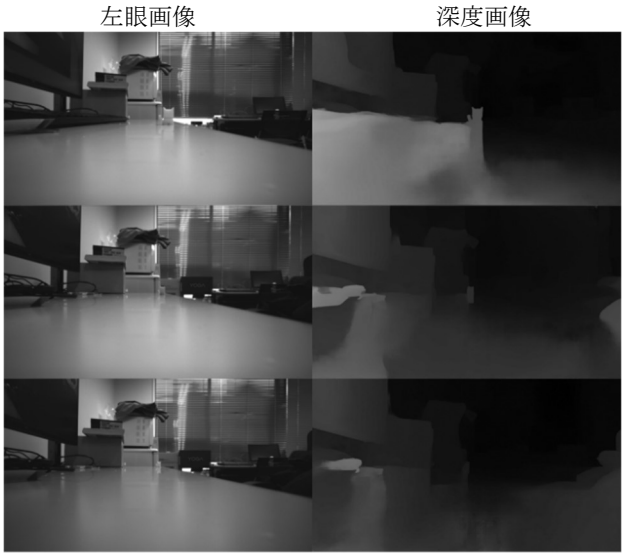
【表１】PC 環境

CPU	Intel i9-10900 (10 コア)
CPU クロック	5.20GHz
メモリ	DDR4 3200 (128GB)
GPU	RTX2080 SUPER / RTX 3080
GPU メモリ	8GB / 11GB

【表２】FADNet モデルの性能評価結果

画像サイズ	576×960
フロート精度	0.901 EPE
量子化精度	1.164 EPE
訓練 GPU メモリ使用量	6 GB
量子化 GPU メモリ使用量	9.5 GB

Performance evaluation of depth estimation using Fast and Accurate Network on FPGA
Shouyi Tian, Takeshi Ohkawa
Tokai University



【図 2】FPGA 処理 FADNet モデルによる深度推定画像

図２に、FADNet モデルによる深度推定結果画像を示す。上から下に 30cm,60cm,90cm、左の画像がに左眼画像であり、右が深度画像である。いずれも適切に深度推定が出来ている。

FPGA による深度画像生成時間は、一枚あたり平均 6.33 秒（経過時間）であった。time コマンドでの測定でユーザ時間が 3.01 秒、システム時間は 2.86 秒であった。すなわち CPU 時間が 5.87 秒で、残りの 0.46 秒が FPGA 処理時間や他プロセスの待ち時間であると考えられる。

(2) 精度比較評価

FPGA で FADNet モデルによる深度推定の精度を評価するために、Intel 社の Realsence D435 を使用して、結果比較テストを実施した。

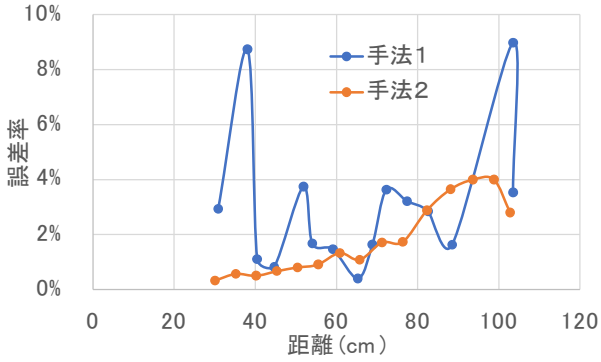
・手法１:FPGA で FADNet モデルによる深度推定
・手法２:Intel Realsence D435

異なる２つの手法の精度の比較を行うために、実際測量値 cm を単位として比較評価をした。手法１では、depth 画像が出力されるのみで定量的な比較が行えない。そこで、depth 画像の明度値から距離に変換する式（１）を作成した。

式（１）は次のように求めた．まず，距離が異なる（30cm～100cm、5cm 間隔）15 セットの画像を撮影し，その画像を手法 1 で推定する．手法 1 により、推定結果として出力された depth 画像から明度値を取得する．取得した明度値の平均を取り、近似を取ると 別々の距離での明度の近似式となる．この近似式を距離について求めるように変形することで式（１）を求めた。

式（１）深度画像の明度 x とする、距離 y とする。

$$y = 1874.8x^{-0.937} \quad (1)$$



【図 3】２つの手法(手法１：FPGA による FADNet モデル、手法２：Realsence D435)の誤差率

精度を比較するために誤差率を求めた。ここでは誤差率を推定値から正解値を引いた値を正解値で割った値とした。２つの手法の誤差率を図３に示す。図３より、手法１は誤差率の平均は 3.09%、誤差率の最大値は 8.98%、誤差率の最低値は 0.40%となった。手法２は誤差率の平均値は 1.80%、最大値は 4.00%、最低値は 0.33%となった。手法２は距離が遠くなるにつれて誤差率が大きくなる。精度は手法２の方が優れていることがわかった。

5 おわりに

FADNet モデルは、Vitis AI を使って作成し、大量の計算リソースが必要であるが、組込みデバイス FPGA にデプロイできる。深度画像生成には約 6.33 秒がかかる。精度比較の結果、Intel RealSense D435 の測量精度の誤差率は 1.80%であり、FADNet モデルの処理結果は 3.09%であった。しかし、FADNet モデルは FPGA でデプロイでき、データセットが増加し訓練時間が長くなればより精度が高く幅広い場面での汎用性が期待できる。なお、ステレオカメラを用いた FPGA 処理による FADNet モデルの、自動運転ロボットへの搭載は今後の課題である。

参考文献

[1] Hu, Gibson, et al. "A robust rgb-d slam algorithm." *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012.

[2] Perez-Vicente, Alejandro, et al. "Point Cloud Generation with Stereo and Monocular Depth Estimation on FPGA." *2022 IEEE International Conference on Electro Information Technology (eIT)*. IEEE, 2022.

[3] Wang, Qiang, et al. "Fadnet: A fast and accurate network for disparity estimation." *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020.

[4] Mayer, Nikolaus, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[5] Xilinx, “Vitis AI ユーザー ガイド (UG1414) - 2.5”Vitis AI 概要