

ConvDeconvNet の効率的モバイル実装による 画像変換・物体検出・領域分割リアルタイムiOS アプリ群

丹野 良介^{1,a)} 泉 裕貴^{2,b)} 柳井 啓司^{1,c)}

1. はじめに

Deep Learning を用いてサーバー側で学習後、エッジ側で推論を行うといった形で深層学習の Edge Computing 上での利用が今後加速していくと予想される。しかし、多くの畳込み演算を内部で行う必要があるなど、膨大な計算量をどう処理するかが問題となる。そうした中で、我々は高速化の工夫及び省メモリ化により Deep Learning を用いたリアルタイム画像認識システムがモバイル上でも実現できる段階まで来た。

そこで、本デモでは、畳込みニューラルネットワークを用いたリアルタイム物体検出の他、図 1 のような Convolution Deconvolution Neural Network(ConvDeconvNet) の構造をもつ領域分割、画像生成のアルゴリズムを iOS 上に実装し、リアルタイムに動作可能であることを実演を通して実証する。

2. モバイル実装ワークフロー

CNN をモバイルに実装する上で高速化は必須である。我々は今まで効率的な CNN のモバイル実装 [9] や深層学習フレームワーク用コンバータ [7] の研究成果がある。端的にまとめると、

- DNN を直接 C 言語コードに変換し、コンパイラ的に実行 (Caffe2C/Chainer2C)。
- Multithread 化による NEON/BLAS の効率的な利用。
- CNN に掛かる演算の可能な限りの事前計算の実行。であるが、詳細については、[7,9] を参照願いたい。

3. 物体検出

Fast RCNN や Faster RCNN の派生系である YOLO [2], SSD [6] を用いて、図 2 のような結果が得られるデモを行う。

YOLO [2]

画像をグリッド分けして同時にクラス分類, Bounding Box, 信頼度を予測することで高速に物体検出をすることが可能

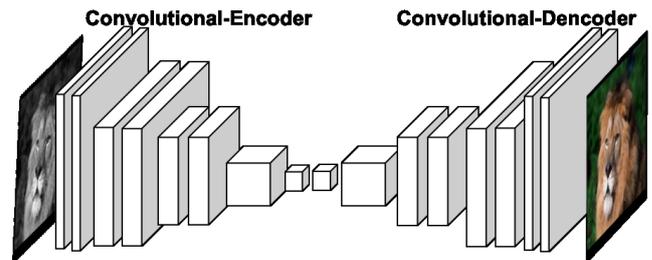


図 1 ConvDeconvNet の構造

である。Fast RCNN より精度は劣るが、100 倍高速なネットワーク。

SSD [6]

Faster RCNN の 2 段階処理 (RPN, 識別) を同時に行うことで高速な物体検出を可能にしたネットワーク。また、深さの異なる複数の feature map に適用することで、YOLO よりも高速高精度な物体検出を可能にした。



図 2 SSD による検出例 ([6] より引用)

4. 領域分割

CNN を用いた領域分割アルゴリズムは数多く考案されているが、どれも計算量がボトルネックとなってしまう、実際にアプリケーションに応用するのは困難である。そこで、本デモでは、軽量かつ高精度に領域分割可能な RCC-Net [1] のネットワークを用いて、図 6 のような結果が得られるデモを行う。

RCC-Net [1]

Inception と ResNet の仕組みを利用した軽量かつ高精度で路上の領域分割をすることが可能なネットワーク (図 3)。

5. 画像生成

近年, Variational AutoEncoder(VAE) や Generative Ad-

¹ 電気通信大学 大学院情報理工学研究科 情報学専攻

² 電気通信大学 情報理工学部 総合情報学科

a) tanno-r@mm.inf.uec.ac.jp

b) izumi-y@mm.inf.uec.ac.jp

c) yanai@mm.inf.uec.ac.jp

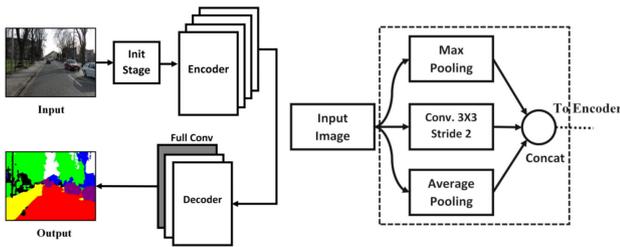


図 3 RCC-Net の概要図 ([1] より引用)

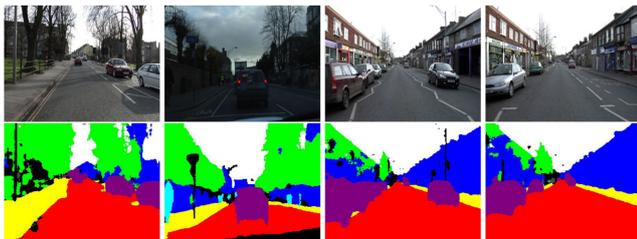


図 4 RCC-Net による検出例 ([1] より引用)

versarial Network(GAN)をはじめとする生成モデルが注目を集めている。生成モデル系では、訓練データと生成データの分布が一致するように学習することで、新しいデータを生成することを可能とするようなモデルを指す。生成系に関するネットワークは多数存在するが、その中から本デモでは、pix2pix [4], CycleGAN [10] のデモを行う。また、スタイル変換の拡張である MultiStyleTransfer [8] のデモも行う。

pix2pix [4]

GAN で提案された Adversarial Loss と ConvDeconvNet を組み合わせることで、画像のペア集合間の変換方法を学習することが可能となり、線画 ↔ カラー画像, 航空写真 ↔ 地図写真などの変換をすることができる。

CycleGAN [10]

pix2pix では入出力に画像ペアが必要だったところを、ペアが無くとも上手く画像生成できるように cycle consistency loss を導入して拡張したネットワークである。

MultiStyleTransfer [8]

事前に特定のスタイルの変換を feed-forward で行う CNN を学習しておくことで、ほぼリアルタイムで画像を変換可能にする手法が提案された [5]。この手法では、事前に学習した単一のスタイルのみを転送することしかできなかった。そこで、本研究では、[5] のネットワークを複数種類のスタイルを任意の重みで合成して、リアルタイムで合成したスタイルを転送する手法に拡張した。

参考文献

[1] Ardiyanto, I. and Adji, T. B.: Deep residual coalesced convolutional network for efficient semantic road segmentation, *IPSN Transactions on Computer Vision and Applications* (2017).
 [2] Divvala, J. A., Girshick, S. A., Farhadi, R. A. and , A.: You Only Look Once: Unified, Real-Time Object Detection, *Proc. of IEEE Computer Vision and Pattern Recognition* (2016).

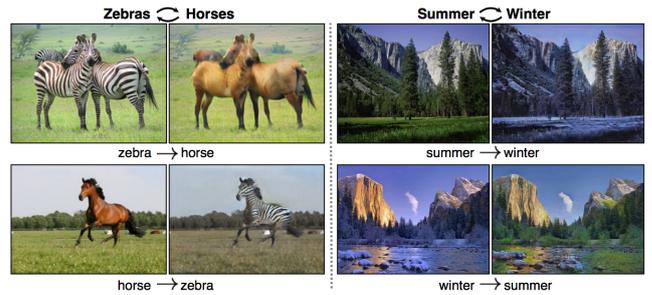


図 5 CycleGAN による生成例 ([10] より引用)

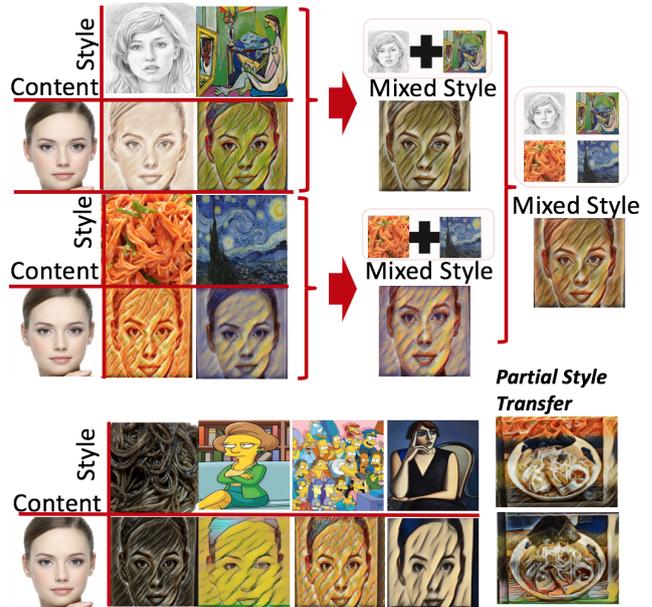


図 6 MultiStyleTransfer [8] による変換例

[3] Gatys, L., Ecker, A. and Bethge, M.: Image Style Transfer Using Convolutional Neural Networks, *Proc. of IEEE Computer Vision and Pattern Recognition* (2016).
 [4] Isola, P., Zhu, J., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, *Proc. of IEEE Computer Vision and Pattern Recognition* (2017).
 [5] Johnson, J., Alahi, A. and Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution, *Proc. of European Conference on Computer Vision* (2016).
 [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector, *Proc. of European Conference on Computer Vision* (2016).
 [7] Tanno, R. and Yanai, K.: Caffe2C: A Framework for Easy Implementation of CNN-based Mobile Applications, *Proc. of International Workshop On Mobile Ubiquitous Systems, Infra-structures, Communications, And Applications (MUSICAL 2016)* (2016).
 [8] Tanno, R. and Yanai, K.: DeepStyleCam: A Real-time Style Transfer App on iOS, *Proc. of International MultiMedia Modeling Conference (MMM)* (2017).
 [9] Yanai, K., Tanno, R. and Okamoto, K.: Efficient Mobile Implementation of A CNN-based Object Recognition System, *Proc. of ACM Multimedia* (2016).
 [10] Zhu, J. Y., Park, T., Isola, P. and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, *Proc. of arXiv:1703.10593* (2017).