7D-03

効率的ニューラルアーキテクチャ自動探索のセマンティックセグメンテーションへの適用

伊藤 多一† 魏 崇哲† 村里 圭祐† 齋藤 彰儀† 太田 満久† 若槻 祐貴‡

株式会社ブレインパッド† 高知工科大学大学院‡

あらまし 強化学習により訓練された再帰的ニューラルネットを用いて、画像識別や文書生成などのタスクに特化した深層学習ネットワークのアーキテクチャを探索する試みが近年盛んに研究されている。特に、少ないGPUリソースで効率的に探索できる手法がGoogle Brainにより提唱されている。本発表では、上述のアーキテクチャ探索手法をセマンティックセグメンテーションに適用した結果について報告する。

1 はじめに

近年、深層学習と強化学習を利用した最適化問題へのアプローチが注目を集めている。例えば、画像識別に特化した最適なネットワーク構造を、深層強化学習モデルを使って探索する試みがある[1]。一方、画像異常検知では、深層学習ネットワークによるセマンティックセグメンテーションが高精度を実現しているが、ネットワーク構造を人手で最適化するのに多大な手間と時間を要する。こうした背景の下、ネットワーク重み共有により、従来のネットワーク構造探索を飛躍的に効率化する手法としてENASが提唱された[2]。本研究では、ENASをセマンティックセグメンテーションに適用することを試みる。

2 ENASについて

強化学習では、エージェント（agent）が自ら選択した行動を通して環境（environment）に働きかけ、環境から報酬信号を受け取って自身を改善しながら最適な行動指針（方策, policy）を学習する。強化学習をネットワーク構造探索(NAS, Neural Architecture Search)[1]に適用する場合、再帰型ニューラルネットによるエージェントが、環境モデル（例えばCNN, Convolutional Neural Networkによる画像識別モデル）のアーキテクチャを生成し、環境モデルはそのアーキテクチャの精度評価を報酬信号としてエージェントに返すことで最適方策の学習が進む。その際、生成したアーキテクチャ全てを最初から学習して精度評価を行うので、探索が収束するまでに450個のGPUを使って3-4日の計算が必要であった。そこで、計算コストを削減する工夫としてENAS (Efficient Neural Architecture Search)[2]が提案された。

ENASでは、有向非巡回グラフ(DAG, Directed Acyclic Graph）による探索空間表現と環境モデルのネットワーク重み共有により、計算リソースの節約を図る。実際、NASを適用したのと同じ環境モデルにENASを適用した場合、GPU 1個による半日の計算でアーキテクチャ探索が収束し、GPU時間で1,000倍の効率化が実現した。

3 ENASのセマンティックセグメンテーションへの適用

セマンティックセグメンテーションとは、画像領域の物体クラスをピクセル単位で予測するタスクである。画像中の物体領域をクラスごとに異なる色で塗り分けたアノテーション画像を用いた教師あり学習を行い、モデルパラメータを推定する。入力画像をencoder networkで圧縮して画像特徴量を抽出し、それをdecoder networkに入力してアノテーション画像を再現するように学習を行う。ENASをセマンティックセグメンテーションに適用する場合、環境モデルはセマンティックセグメンテーションを行う深層ネットワークであり、エージェントは環境モデルのネットワーク構造を記号列として生成・探索する再帰型ネット―ワークとして定義される。

ネットワークの基本設計：

エージェントのネットワークには、言語モデルで定評のある長期・短期記憶セルを用いた再帰型ネットワーク（LSTM）を採用した。環境モデルのネットワークは、convolutionセルを積み上げたencoderと、その鏡像対称の構造を持つdecoderを連結した構造として設計される。encoderは、決まった数の層数（今回は12層）からなり、各層は6種類のconvolutionセルの

一つとして決まる。さらに encoder における特徴マップの縮小（stride 2 の pooling）と decoder における特徴マップの拡大（stride 2 の upsampling）が必要であるが、その際、encoder 側の特徴マップの情報を対応する decoder 側の特徴マップに long skip 接続 を介して伝達する。エージェントは、encoder の最適なアーキテクチャ（convolution セルの最適な組合せ）を強化学習により探索する。アーキテクチャ探索では、6 種類の convolution セルから一つずつサンプリングして選択しながら 12 層のセル系列を生成する。6 種類の convolution セルは、共通構造である conv 1x1 + BN + ReLU の上にそれぞれ以下の層を積んだものとして定義される[2]。

0：conv 3x3 + BN + ReLU
1：separable conv 3x3 + BN + ReLU
2：conv 5x5 + BN + ReLU
3：separable conv 5x5 + BN + ReLU
4：average pooling 3x3
5：max pooling 3x3

報酬関数の設計：
強化学習においては報酬関数の定義が重要である。報酬関数として accuracy を考えた場合、多くの画像で背景クラスの寄与が大きく、背景クラスを予測しやすいと懸念される。そこで、物体クラスごとの IOU を平均した mean-IOU (mIOU)、背景クラスと物体クラス全体の 2 クラスにまとめて IOU を算出して平均した改良版 mean-IOU（mIOU+）を報酬関数の候補に追加した。

４　実験

実験には、VOC2012[3]を使用した。これは、22 クラス（背景と境界線を含む）2,913 枚からなる画像データで、元画像とアノテーション画像がセットになっている。これを 7：2：1 の比率で訓練用、検証用、テスト用に分割して使用した。また、ENAS による探索で得られた環境モデルの評価指標として、accuracy の他、テスト画像全体のピクセル群で計算した mean-IOU を採用した。参考までにクラスごとの precision, recall の平均値も算出した。

検証結果：
報酬関数を変えて ENAS によるアーキテクチャ探索を行い、その結果得られた最適アーキテクチャの予測精度で報酬関数を比較した（表１）。報酬関数が accuracy の場合が、最も精度が高く次いで mIOU+, mIOU の順となった。報酬関数が accuracy の場合、環境モデルの最適アーキテクチャは、0,5,5,2,2,2,4,5,3,5,3,4 であり、0～3 の convolution セルのうちカーネルサイズが 5x5 のセルがほとんどを占めている。VOC2012 の画像のように、物体クラス領域が背景と同程度に広い場合、より広い受容野の方がクラス領域全体を捉えるのに適していることを示唆している。

表 1：精度比較

Metric\Model	reward: accuracy	reward: mIOU	reward: mIOU+
accuracy	74.42%	73.19%	71.83%
mean-precision	40.02%	38.20%	38.39%
mean-recall	34.04%	24.42%	28.25%
mean-IOU	23.70%	18.26%	19.02%
Params (million)	0.27	0.18	0.31

５　まとめ

ENAS をセマンティックセグメンテーションに適用した。現時点の SOTA である DeepLab v3+[4] のパラメータ数が 4,100 万以上であるのに対し、ENAS で探索したアーキテクチャは、その 1/100 未満と圧倒的に少ない。ENAS により、少ないパラメータ数でも実用的な精度が得られる可能性を示唆している。今後の課題としては、通常の convolution セルを dilated convolution を含んだ並列構造[4]をもつセルに拡張して、そのセル構造を探索するといった改良が挙げられる。

参考文献

[1] B. Zoph and Q.V. Le, Neural architecture search with reinforcement learning, In ICLR2017, arxiv: 1611.01578
[2] H. Pham, M. Guan, B. Zoph, Q. Le and J. Dean, Efficient Neural Architecture Search via Parameters Sharing, In ICML 2018, arxiv:1802.03268.
[3] Visual Object Classes Challenge 2012 (VOC2012), http://host.robots.ox.ac.uk/pascal/VOC/voc2012/
[4] L-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, In ECCV_2018, arxiv:1802.02611

Efficient neural architecture search for semantic segmentation
† BrainPad Inc.
‡ Kochi-University of Technology