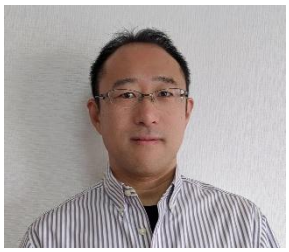
自己蒸留を用いたセグメンテーションの高精度化



堀田　一弘（Kazuhiro HOTTA, Ph. D.）

名城大学　理工学部　電気電子工学科　教授

(Professor at Meijo University, Faculty of Science and Technology, Department of Electrical and Electronic Engineering)

IEEE, 電子情報通信学会、情報処理学会、人工知能学会

受賞：

・坂井泰吾, 光岡日菜子, 堀田一弘, 外観検査アルゴリズムコンテスト 最優秀賞（2025 年）

・加藤聡太, 堀田一弘, 画像の認識・理解シンポジウム(MIRU2022)インタラクティブ発表賞（2022 年）

・本多慶伍, 堀田一弘, 精密工学会外観検査アルゴリズムコンテスト　優秀賞, レゾナンスバイオ賞（2018 年）

・堀田一弘, 映像情報メディア学会誌 ベストオーサー賞（2010 年）他

著書：

・高橋治久、堀田一弘、学習理論, コロナ社 (2009 年)

・金森 由博、井尻 敬、堀田 一弘、五十嵐 悠紀、徳吉 雄介、安田 廉、山本 醍田、向井 智彦、梅谷 信行、Computer Graphics Gems JP 2013/2014 コンピュータグラフィックス技術の最前線、ボーンデジタル（2013 年）

研究専門分野：画像認識、深層学習、人工知能、機械学習

あらまし

近年提案されているセマンティックセグメンテーション法は、新たな計算の追加などにより精度の向上を図っているため、計算コストが増大している。そこで本研究では、セグメンテーションモデルの出力から Axial-Attention を用いることにより大域的な情報を用いて精度を向上させた教師出力と、工夫を加えてない生徒出力を生成し、２つの間で知識蒸留を行う手法を提案する。これにより計算量の増加を抑制しつつ精度を向上させる。複数の学習モデルに提案手法を用いた結果、多くのモデルにおいて精度の向上を確認した。

１．はじめに

画像認識分野では Convolutional Neural Network (CNN)の有効性が広く知られている。画像認識の１つにセマンティックセグメンテーションと呼ばれる問題があり、入力画像の全ての画素にラベルを付与する。この技術は細胞画像や医用画像[1,2]などに広く応用されている。セマンティックセグメンテーションは入力画像中の全て画素を識別する必要があるため、画素同士の関係性が重要となる。そのため、最近の研究ではネットワークに新たな計算機構を追加して特徴を得るものが多いが、計算量が増大してしまうという問題が発生する。

計算量の問題を解決する手法はいくつか提案されており、その一つである知識蒸留は計算量の少ない生徒モデルの出力を計算量の多い教師モデルの出力に近づけるように学習することにより、少ない計算量で教師モデルに近い精度を目指す手法である。本研究では、計算量をあまり増やすことなく精度を向上させるために、画像の空間的情報を獲得することができる手法である Attention 機構[3]とこの知識蒸留を組み合わせた手法を提案する。具体的には、学習モデルの出力に対し Axial-Attention[4]を用いることにより大域的な情報を付与して精度を向上させた教師モデルを、元の計算量が増加していない生徒出力の学習に利用することにより、計算コストを増やすことなく精度の向上を達成する。

評価実験では、ショウジョウバエの細胞画像データセットと Covid-19 の肺炎画像データセットを用いて行う。複数の学習モデルに提案手法を用いた結果、mIoU が生徒出力では 2.5%、教師出力が 2.6%改善するなど､多くのモデルにおいて精度の向上を確認した。

２．提案手法

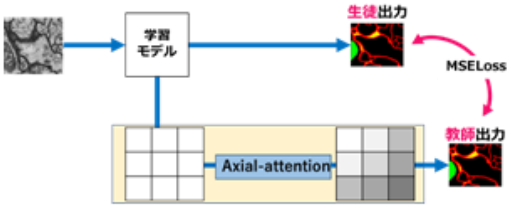


図１ 提案手法の概要

図１に提案手法の概要を示す。セグメンテーション

モデルの出力から 2 つの特徴マップを作成する。1 つ目の出力はセグメンテーションモデルの出力そのままの「生徒出力」であり、2 つ目の出力は注意機構により空間的情報が強化された「教師出力」である。 提案手法では、これらの 2 つの出力の間で自己蒸留を行う。計算量が多く、高い精度が期待できる教師出力を生徒出力が真似ることにより、従来のセグメンテーションモデルと同じ計算コストである生徒出力の精度を向上させることができる。

また、提案手法では教師出力を強化するために注意機構を利用している。各画素同士の内積計算から空間的関係性を取得する Self-Attention[3]を利用しても精度の向上は期待できるが、画像サイズが大きい場合では多大な計算コストと大量のメモリが必要となってしまい、現実的とは言えない。

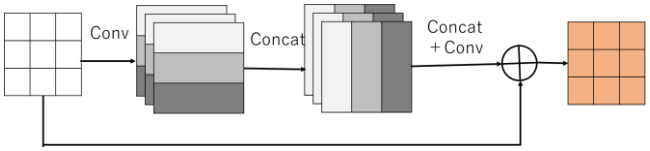


図 2　Axial-Attention

そこで、図 2 のように画像を縦横の短冊状に分割し、画像の縦軸及び横軸同士の内積計算から空間的関係性を取得することにより、計算量とメモリ使用量を低減させた Axial-Attention を利用する。これにより、計算コストをあまり増やすことなく高い精度の教師出力を得ることができ、これを真似る生徒出力の精度も向上させることができる。また、提案手法はセグメンテーションモデルにより得られる出力をそのまま利用するため、多くのモデルに適用することができる。

提案手法の最終的な損失関数は以下のようになる。

Loss = StudentLoss + TeacherLoss + MSELoss

ここで、StudentLoss は生徒出力から算出する Softmax Cross Entropy loss、TeacherLoss は教師出力から算出する Softmax Cross Entropy loss、MSELoss は生徒出力と教師出力間で自己蒸留するための二乗誤差損失である。

３．評価実験

実験では、5 クラスのショウジョウバエの細胞画像データセットと 4 クラスの Covid-19 の肺炎画像のデータセットを用いた。また、評価指標には mean Intersection over Union (mIoU)を採用した。

学習モデルには Unet [2]の他に、Unet++ [5]、Multi-scale Attention Network（Manet）[6]，Linknet [7]，Feature Pyramid Networks（FPN）[8]，Pyramid Scene Paring Network（PSPnet）[9]，Pyramid AttentionNetwork（PAN）[10]，DeepLabV3 [11]，DeepLabV3+[12]を採用し、これらに提案手法を適用する。

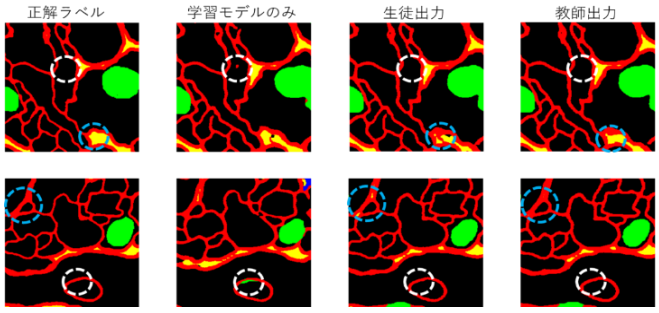
9 種類のセグメンテーションモデルに提案手法を適用した場合としなかった場合におけるショウジョウバエの細胞画像の精度を示す。表 1 より Unet では提案手法を適用した場合、生徒出力の mIoU は 73.07%、教師出力は 73.18%となり、提案手法を適用しなかった場合と比較してそれぞれ 0.71%、0.78%精度が向上した。また Linknet では提案手法を適用した場合、生徒出力、教師出力の mIoU はそれぞれを 73.61%、73.71%となり、提案手法を適用しなかった場合と比較して 2.52%、2.62%精度が向上した。しかし、PSPnet では、提案手法を適用した場合、生徒出力が 68.72%、教師出力が 68.77%となってしまい、いずれも提案手法を適用しなかった場合よりも精度が低下している。

FPN と PAN では教師出力の精度が向上したものの、生徒出力では精度が低下してしまった。しかし、Unet++、Manet、DeepLabV3、DeepLabV3+でも生徒出力、教師出力ともに精度が向上しており、提案手法の有効性を示すことができた。獲得できる情報や精度は学習モデルごとに異なっているため、 Axial-Attention により獲得される大域的情報が有益に働くモデルもあれば、十分に作用しないモデルもある。特に、元から大域的な情報を抽出しているネットワークでは、Axial-Attention による大域的な情報が加わることにより局所的な情報が重視されなくなる。したがって、これを知識蒸留している生徒モデルにも精度の差が生まれるのではないかと考えられる。

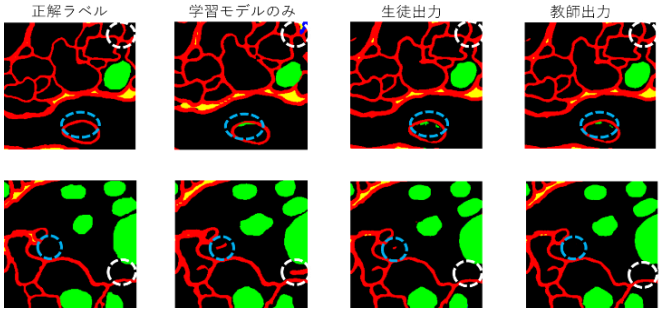
次に、Covid-19 データセットに対する精度を表 2 に示す。Unet では提案手法を適用した場合、生徒出力では 0.58%、教師出力では 1.95%精度が向上した。 ま

た、Linknet では提案手法を適用した場合、生徒出力では 0.27%、教師出力では 1.17%精度が向上した。しかし、PSPnet では提案手法を適用した場合、生徒出力では 0.59%、教師出力では 1.02%精度が低下している。これらはショウジョウバエの細胞画像における精度結果と同様の傾向であり、他の学習モデルにおいてもショウジョウバエの細胞画像データセットと同様の傾向が確認できた。

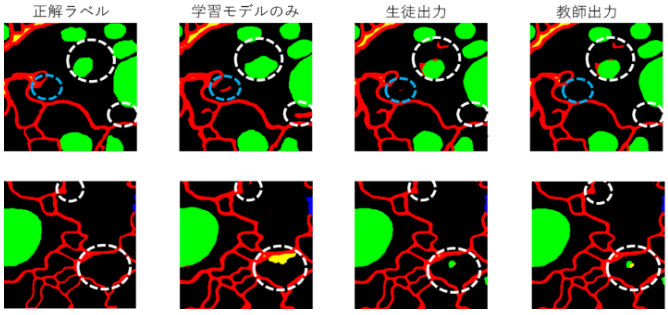
次に、セグメンテーション結果の可視化を行う。こ



(a) Unet



(b) Linknet



(c) Manet

図 3 各学習モデルにおける出力結果

こでは、標準的なモデルである Unet、最も精度が向上した Linknet、精度の向上が小さかった Manet のショウジョウバエの細胞画像におけるセグメンテーション結果を図 3 に示す。図 3(a)の白丸で囲った部分を見ると、元のモデルでは細胞膜を誤って検出しているが、提案手法により改善していることが分かる。また、青丸で囲った部分を見ると、生徒出力で誤っている細胞膜の形状が教師画像では改善されている。このため、実行環境や用途に応じて計算量が少ない生徒出力を利用するか、精度の高いが計算量も高くなる教師出力を選んで利用することが有効である。

図 3(b)の Linknet でも提案手法を用いることにより精度が改善されていることが分かる。図 3(c)より、Manet では提案手法により改善しきれなかった箇所もあるが、青丸で囲った箇所のように提案手法により誤識別が少なくなることが確認できた。

表 1　ショウジョウバエの細胞画像における精度

赤字は最高精度，青字は 2 番目に高い精度を表す。

model	提案手法 有/無		IoU(%)					
			class0	class1	class2	class3	class4	mean
Unet	無		72.27	82.11	48.54	67.11	91.76	72.36
	有	生徒	73.21	83.67	49.21	66.23	92.94	73.07
		教師	73.40	83.90	49.38	66.36	92.88	73.18
Unet++	無		73.14	82.39	50.22	70.13	92.31	73.64
	有	生徒	73.73	83.54	50.87	68.43	92.80	73.87
		教師	73.77	83.61	50.93	68.49	92.82	73.92
Manet	無		72.12	81.26	47.49	68.47	92.02	72.27
	有	生徒	72.97	82.73	49.68	67.19	91.88	73.61
		教師	73.09	83.52	50.83	68.37	92.77	73.72
Linknet	無		71.51	80.76	47.65	68.95	91.56	71.09
	有	生徒	73.56	81.67	51.14	68.68	93.01	73.61
		教師	73.65	81.78	51.25	68.79	93.09	73.71
FPN	無		69.69	85.17	51.33	66.67	90.88	72.75
	有	生徒	71.94	80.76	50.01	68.21	92.02	72.61
		教師	72.23	81.27	50.47	68.58	92.24	73.03
PSPnet	無		65.93	80.82	48.38	65.70	89.37	70.04
	有	生徒	62.71	80.98	45.44	65.71	88.74	68.72
		教師	62.75	81.04	45.49	65.77	88.78	68.77
PAN	無		66.66	79.55	46.93	65.96	90.08	69.84
	有	生徒	65.70	82.06	45.98	65.62	89.06	69.69
		教師	66.55	82.93	46.96	66.67	89.82	70.60
Deep LabV3	無		67.10	80.66	46.65	67.72	90.12	70.45
	有	生徒	63.42	84.85	49.68	69.20	88.88	71.21
		教師	62.57	83.63	48.57	68.15	88.10	70.53
Deep LabV3+	無		70.22	77.57	43.51	66.10	91.52	69.96
	有	生徒	69.78	83.40	53.07	69.33	92.20	73.56
		教師	69.15	82.64	52.14	68.47	91.42	72.79

自己蒸留を用いたセグメンテーションの高精度化

表 2　Covid-19 の肺炎画像における精度結果
赤字は最高精度，青字は 2 番目に高い精度を表す。

model	提案手法 有/無		IoU(%)				
			class0	class1	class2	class3	mean
Unet	無		94.03	33.76	42.22	1.29	42.83
	有	生徒	95.45	34.78	42.15	1.26	43.41
		教師	96.48	35.47	45.21	1.95	44.78
Unet++	無		92.11	26.76	44.99	3.69	41.89
	有	生徒	94.14	26.13	45.07	5.37	42.68
		教師	94.65	26.54	45.32	5.24	42.94
Manet	無		93.33	33.58	43.82	3.85	43.65
	有	生徒	94.01	34.19	43.54	3.05	43.70
		教師	95.62	35.28	43.68	4.06	44.66
Linknet	無		94.19	39.56	45.83	2.93	45.66
	有	生徒	95.90	37.50	47.55	2.76	45.93
		教師	96.97	38.12	48.55	3.67	46.83
FPN	無		92.46	31.41	36.78	3.67	41.08
	有	生徒	92.91	30.21	35.88	5.10	41.03
		教師	93.78	30.56	35.96	5.24	41.39
PSPnet	無		93.88	8.22	27.59	0.86	32.64
	有	生徒	94.12	7.59	26.50	0.00	32.05
		教師	94.60	6.79	25.10	0.00	31.62
PAN	無		92.53	30.57	35.86	4.26	40.81
	有	生徒	93.14	30.14	35.46	3.22	40.59
		教師	93.79	31.68	35.79	4.24	41.38
Deep LabV3	無		90.93	26.90	31.41	5.86	38.78
	有	生徒	91.32	27.96	31.88	6.69	39.71
		教師	91.78	27.21	30.84	5.24	38.99
Deep LabV3+	無		91.54	30.16	34.28	3.73	39.93
	有	生徒	92.39	31.56	36.32	5.01	41.63
		教師	93.63	30.67	35.77	4.41	40.81

4．おわりに

本稿では Axial-Attention を教師として用いた自己蒸留法を提案した。元のモデルのセグメンテーション出力に Axial-Attention を適用し、大域的な情報を強化した教師出力を作成し、生徒出力に知識蒸留を行うことにより、計算量のあまり増やすことなく精度の向上を行った。実験では多くのモデルで計算量を増やすことなく精度を向上させることができ、提案手法の有効性を示すことができた。

用語解説

*1 Axial Attention : 画像を縦横の各軸で短冊状に区切って Attention を行うことにより計算コストを削減した方法

参考文献

[1] M.Havaei, A.Davy, D.W.Farley, An.Biard, A.Courville, Y.Bengio, C.Pal, P.M.Jodoin, and H.Larochelle. Brain Tumor Segmentation with Deep Neural Networks. Medical Image Analysis, Vol. 35, pp. 18–31, 2017.

[2] O.Ronneberger, P.Fischer, and T.Brox. U-net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of Medical Image Computing and Computer-Assisted Intervention, pp. 234–241., 2015.

[3] A.Vaswani, N.Shazeer, N.Parmar, J.Uszkoreit, L.Jones, A.N.Gomez, L.Kaiser, and I.Polosukhin. Attention is All You Need. Advances in Neural Information Processing Systems, Vol. 30, 2017.

[4] H.Wang, Y.Zhu, B.Green, H.Adam, A.Yuille, L.C.Chen. Axial DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation. In Proceedings of European Conference on Computer Vision, pp.108-126, 2020.

[5] Z.Zhou, M.R.Siddiquee, N.Tajbakhsh, and J.Liang. Unet++: A nested U-net Architecture for Medical Image Segmentation. In Proceedings of Workshop on Deep Learning in Medical Image Analysis, pp. 3–11, 2018.

[6] T.Fan, G.Wang, Y.Li, and H.Wang. Ma-net: A Multi-Scale Attention Network for Liver and Tumor Segmentation. IEEE Access, Vol. 8, pp. 179656–179665, 2020.

[7] A.Chaurasia and E.Culurciello. Linknet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In Proceedings of IEEE Visual Communications and Image, pp. 1–4, 2017.

[8] T.Y.Lin, P.Dollar, R.Girshick, K.He, B.Hariharan, S.Belongie. Feature Pyramid Network for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117-2125,2017.

[9] H.Zhao, J.Shi, X.Qi, X.Wang, and J.Jia. Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890, 2017.

[10] H.Li, P.Xiong, J.An, and L.Wang. Pyramid Attention Network for Semantic Segmentation. In Proceedings of British Machie Vision Conference, 2018.

[11] L.C.Chen, G.Papandreou, F.Schroff, and H.Adam. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint arXiv:1706.05587, 2017.

[12] L.C.Chen, Y.Zhu, G.Papandreou, F.Schroff, and H.Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision, pp. 801–818, 2018.

この研究は、令和３年度ＳＣＡＴ研究助成の対象として採用され、令和４～６年度に実施されたものです。