Vol.2019-MUS-122 No.6
Vol.2019-EC-51 No.6
2019/2/22

深層学習による物体検出を利用した簡易な手書き譜面演奏装置

石曽根 奏子[1]　馬場 哲晃[1,a)]

概要：著者らはこれまで五線と符頭により構成される簡易な手書き譜面を演奏インタフェースに利用したシステム Gocen に関する開発を行ってきた．当時は技術面から即時的な演奏対話設計を実現するために，簡易な譜面表記としていたが，深層学習を利用することで，印刷譜面に対してもある程度の実時間実行が可能であることと考えた．本稿では既存の Gocen における演奏方法をそのままに，認識部分を深層学習による物体検出を利用したものとし，実際の楽譜をスキャンしながら演奏が可能であるかを評価するためのデータ・セットおよびプロトタイプ作成を報告する．

キーワード：深層学習，物体検出，譜面，演奏インタフェース

Kanoko Ishisone[1]　Tetsuaki Baba[1,a)]

1. 背景

近年の深層学習による発展を振り返ると，2012 年に Deep Learning による画像識別手法が他の機械学習手法よりも高スコアを獲得したことで [1]，とりわけ物体検出についてこれまでの手続き型処理と比較し，データセットさえ揃えば比較的容易に物体検出器を作成できる環境が整ってきた．

深層学習においてデータセットの作成が重要であることはよく知られているが，すでに ImageNet[*1]や COCO[2] 等に代表されるデータセットを学習させることで，汎用的な物体検出器開発は比較的容易になった．本稿では一般的な印刷譜面を対象に，従来の Gocen デバイス（図 1 参照）を用いてスキャン動作をさせることでリアルタイムに譜面の演奏を，デバイス操作を活用することでスラーやビブラートなどの一部の表情付けを可能にすることを目的とする．対話設計についてはすでに文献 [3] にある通り実装済みであり，今回は認識部分に関する改良を主な目的としている．

図 1　Gocen デバイス外観

2. 関連研究

2.1 物体検出手法

物体検出手法においては，様々な種類があるなか，著者らの研究グループはこれまで視覚障害者支援のプロジェクトにて，高速な自作データセット作成とリアルタイム物体検出処理に関してすでにいくらかの知見を得ている [4][5][6]．リアルタイム性を重視した学習ネットワークとして，SSD[7] や YOLO[8] が挙げられる．さらにそれを高速化させた MobileNet-SSD[9] や Yolo-tiny を本研究で利用する学習ネットワークとする．これらネットワークの実行速度に

[1] 首都大学東京
Tokyo Metropolitan University, Asahigaoka, Hino, Tokyo 191–0065, Japan
[a)] baba@tmu.ac.jp
[*1] http://imagenet.stanford.edu

関する考察は Huang ら [10] による COCO データセットをベースにした報告を参照されたい．なお，これらネットワークは OpenCV Version3.3 以降，dnn モジュールとして標準で OpenCV 側から利用することができる．

2.2 データセット

楽譜のデータ・セットとして，Musescore[*2]による musicXML ファイルがよく知られている．musicXML を工学楽譜認識（Optical Music Regognition）のデータセットとして利用し，楽譜認識を行った事例が報告されている [11]．一方で musicXML 単体では物体検出に必要なバウンディングボックス情報が提供されていない．そこで本研究ではまずデータセットの作成から始める必要があった．

3. プロトタイピング

プロトタイピングの詳細な過程は文献 [12] にて示している．まず Gocen システムに必要なラベリングの検討から始め，最初は 6 クラス（符頭，ナチュラル，フラット，シャープ，1,2 分音符，五線）の認識ネットワークを試作した．一見して同じような印刷楽譜であっても，学習済みデータによっては精確に認識結果が出ない．そこで，本プロトタイプではピアノ教本としてよく知られるブルクミュラー及びソナチネから学習データセットを作成した．ブルクミュラーからは 8 曲分，ソナチネからは 4 曲分を，既存 Gocen デバイスを利用して手作業でアノテーション作業を行った．学習データセットに利用した譜面は全音楽譜出版社による動作を確認した後，ラベル数を 15 に増やし，データセットを拡充させ再度ネットワークを学習させた．その後，更に登録数を増やした結果を表 1 に示す．これまでの報告同様に，依然としてダブルフラットやダブルシャープの登録数が少ないが，初学者を対象とした譜面においてはほとんど出現することはない．作成したデータ・セットを Yolo.v2-tiny 及び SSD-MobileNet ネットワークにて学習させ，実際に認識性能を確認した．その内容を次節で述べる．

4. 認識結果

データ学習登録していないソナチネアルバムから，無作為に gocen デバイスで演奏箇所を 15 箇所撮影した．その 15 枚の画像において，それぞれ SSD-MobileNet 及び Yolo.v2 にて認識処理を行った結果を図 2 と図 3 にそれぞれまとめる．検出対象とした信頼度閾値は 0.4 としている．物体検出アプリケーションは Openframeworks 及び OpenCV を利用して実装し，いずれも実行速度は 40fps 程度であり（Macbook Pro 15 インチ 2016 モデル），Gocen デバイスの実行基準 FPS である 30FPS を上回る．グラフから，SSD-MobileNet より Yolo.v2-tiny が物体検出率のみに関しては高い結果となっている．SSD-MobileNet では登録数が 1 万を超えている符頭でさえ，認識できていない箇所が多く見られた．物体検出の精度計測に関しては一般的に mAP（meam Average Pricision）を用いて，検出されたバウンディングボックス位置の精確性で検証する必要があるため，これらについては今後の議論としたい．実験観察中においても，Yolo.v2-tiny のほうが本来認識するべきの範囲から多少ズレがあり，検出率が少ない中でも SSD-MobileNet のほうが正確なバウンディングボックスの位置を示している．また画像から見切れている対象物に対して，SSD-MobileNet では，比較的多く認識していたが，今回の検出率には見切れている対象物は対象外としているため，SSD-MobileNet の検出率が下がっている一因であると考えられる．参考までに同一画像での MobileNet-SSD 及び Yolo.v2-tiny の認識結果をそれぞれ図 4，5 に示す．図 4 では和音箇所の符頭位置が未検出であるが，精確に記号位置を検出しているのに対し，Yolo.v2-tiny では Gocen 演奏に必要なすべての記号を検出している一方，和音部分の符頭検出位置に大きなズレがあるのが見て取れる．

一見すると Yolo.v2-tiny による物体検出が妥当と考えられるが，今回実装対象としている Gocen システムでは符頭位置と五線位置の精確性が重要となるため，物体検出位置精度（mAP）が低ければ音高位置推定の誤検出となるため，この点については引き続き SSD および YOLO，その他の手法について継続的に検討していく必要がある．

表 1 ラベル一覧と登録したバウンディングボックス数

番号	クラス名	概要	登録数
0	note head	符頭	11285
1	natural	ナチュラル記号	339
2	flat	フラット記号	451
3	sharp	シャープ記号	696
4	white head	1,2 分音符	634
5	staff	5 線	8647
6	p	ピアノ	252
7	m	メゾ	24
8	f	フォルテ	211
9	clef g	ト音記号	232
10	clef c	ハ音記号	0
11	clef f	ヘ音記号	174
12	bar	小節線	1841
13	d sharp	ダブルシャープ	5
14	d flat	ダブルフラット	0

謝辞 本研究は JSPS 科研費 JP18H03486 の助成を受けたものです。

参考文献

[1] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *Proceedings of the 25th International Conference on Neural Information Processing*

[*2] https://musescore.org/

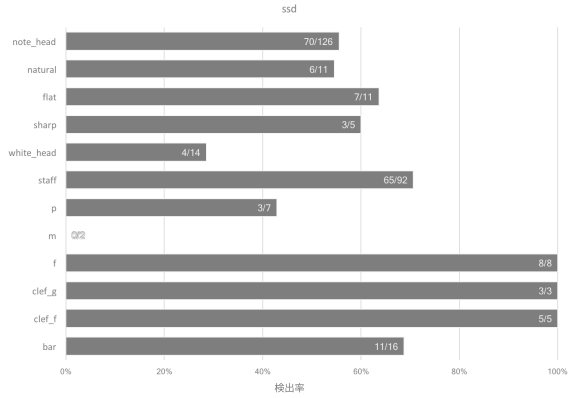

図 2 SSD-MobileNet による，ソナチネアルバムの未登録箇所の検出結果．棒グラフ上の数値は正答数を示す．

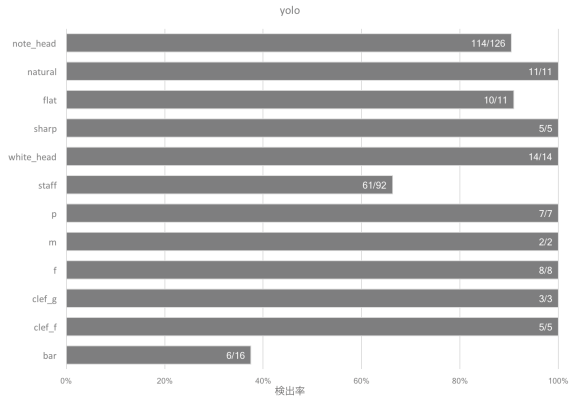

図 3 Yolo.v2-tiny による，ソナチネアルバムの未登録箇所の検出結果．棒グラフ上の数値は正答数を示す．

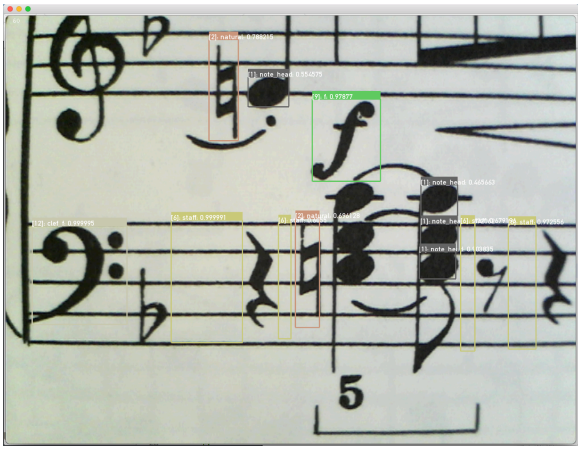

図 4 SSD-MobileNet による認識時の様子．フラット記号や和音符頭箇所が検出できていない．

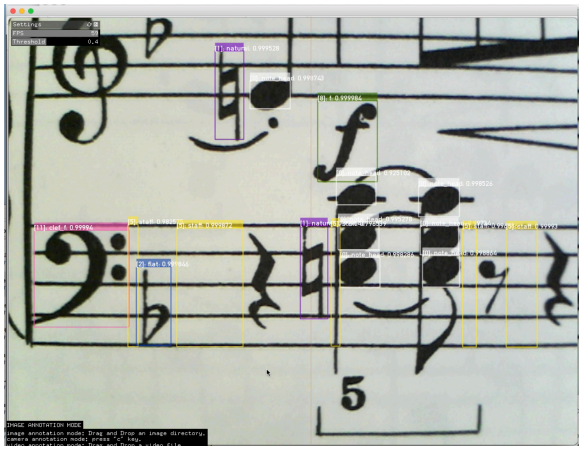

図 5 Yolo.v2-tiny による認識時の様子．ある程度検出はできているが，フォルテシモ直下のドの符頭位置が大きくずれている

Systems - Volume 1, NIPS'12, USA, Curran Associates Inc., pp. 1097–1105 (online), available from ⟨http://dl.acm.org/citation.cfm?id=2999134.2999257⟩ (2012).

[2] Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: Microsoft COCO: Common Objects in Context, *CoRR*, Vol. abs/1405.0312 (online), available from ⟨http://arxiv.org/abs/1405.0312⟩ (2014).

[3] 馬場哲晃，菊川裕也，串山久美子，青木　允：簡易な手書き譜面を利用した演奏システム Gocen の設計，情報処理学会論文誌，Vol. 54, No. 4, pp. 1327–1337（オンライン），入手先 ⟨https://ci.nii.ac.jp/naid/110009579543/⟩ (2013).

[4] 石曽根奏子，馬場哲晃，渡邉英徳，釜江常好：視覚障害者の屋外移動支援に向けた物体検出データセットの基礎検討とプロトタイピング，技術報告 9，首都大学東京, 首都大学東京, 東京大学, 東京大学／スタンフォード大学 (2018).

[5] 馬場哲晃，渡邉英徳，釜江常好：深層学習による物体検出を用いた視覚障害者の屋外活動支援システムにおけるデザイン指針の検討とプロトタイピング，技術報告 8，首都大学東京, 東京大学, 東京大学／スタンフォード大学 (2018).

[6] 石曽根奏子，馬場哲晃，渡邉英徳，釜江常好：ユーザ参加型アノテーションにおける UI 及びデータオーグメンテーションのデザイン，技術報告 1，首都大学東京, 首都大学東京, 東京大学, 東京大学／スタンフォード大学 (2018).

[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector, *ArXiv e-prints* (2015).

[8] Redmon, J. and Farhadi, A.: YOLOv3: An Incremental Improvement, *arXiv* (2018).

[9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, *CoRR*, Vol. abs/1704.04861 (online), available from ⟨http://arxiv.org/abs/1704.04861⟩ (2017).

[10] Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S. and Murphy, K.: Speed/accuracy trade-offs for modern convolutional object detectors, *CoRR*, Vol. abs/1611.10012 (online), available from ⟨http://arxiv.org/abs/1611.10012⟩ (2016).

[11] van der Wel, E. and Ullrich, K.: Optical Music Recognition with Convolutional Sequence-to-Sequence Models, *CoRR*, Vol. abs/1707.04877 (online), available from ⟨http://arxiv.org/abs/1707.04877⟩ (2017).

[12] 石曽根奏子，馬場哲晃：深層学習の画像識別と識別位置検出を用いた Gocen の譜面認識システムの再設計とプロトタイプ，ADADA 5th, 第 5 回アジアデジタルアンドアートデザイン国内大会，Asia Digital Art and Design Association (2018).