

深層学習モデルによる古琴譜の識別の試み To Recognize to Guqin Music by Deep Learning Models

呉本 堯[†] 藤野航馬[†] 高橋浩雄[†] 呉本 舜^{†, **, †††}

小柴満美子^{**} 稗田浩雄^{†††} 間普真吾^{**}

Takashi Kuremoto Kazuma Fujino Hirokazu Takahashi Shun Kuremoto
Mamiko Koshiba Hiroo Hieda Shingo Mabu

1. はじめに

3000 年以上の歴史を持つ古琴（七絃琴）の楽譜は後唐（奈良）時代に文字譜から「減字譜」へ変貌し、弾く弦の番号と指の動き方を記録するものである[1][2]。「減字譜」を解読し、琴曲の演奏の仕方を定めることは「打譜」と称され、現代において、打譜に詳しい専門家は数少ない[3]。これまで、中国の研究者らはコンピューターによる減字譜（Jian-Zu Pu）の識別を行ってきた[4]-[6]が、自動打譜の実現には至っていない。古琴曲の数は、これまで 600 曲があると言われ、実際減字譜曲は 100 曲過ぎず、さらに良く演奏される古琴曲はより少ない。従って、自動打譜システムの開発より、多く眠っている古代の減字譜曲を再現することにつながる。

これまで、著者らは深層学習モデルである VGG16, VGG16+SVM を用いて、最も簡易な古琴曲「仙翁操」の最初の 2 行の減字譜の識別を行った[7][8]が、単字認識に留まる。本研究では、YOLOv5 などの深層学習モデルを用いて、減字譜の曲譜画像より、自動的に単字を抽出・識別する上、単字映像・音声データを用いて、AI による自動打譜システムの構築を試みる。簡易な琴曲である「仙翁操」の減字譜を用いたデータセット作成、深層学習モデルの識別性能比較、映像・音声出力などの実験結果より、AI 打譜の可能性及び問題点を明らかにした。

2. 減字譜と自動打譜

古琴（こきん、クーチン、Guqin）は、七絃琴（しちげんきん）とも呼ばれる撥弦楽器である（図 1 と図 2）。箏などの構造と異なるところは、7 本の弦を支える琴柱（ことじ）はなく、徽（き）と呼ばれるマークが 13 個表面にある。琴弦の順番は、演奏者から遠くから第一弦、第二弦…、琴徽の順番は、右から左へ第一徽、第二徽…と呼ばれる。両手のそれぞれの指の位置と振る舞いを記載し、琴の音を表す曲譜である減字譜は、中国の後唐時代（奈良時代）に作られた古琴の奏法譜である（図 3 と図 4）。図 3 では、最も平易な琴曲である「仙翁操」の減字譜を示す。一般的には、減字譜には漢字で表す歌詞があり、その歌詞の意味を理解する上で、緩急のリズムや音の強弱を吟味して古琴曲を演奏するものである。

減字譜を読み、五線譜などの楽譜で表現し、古琴曲を演奏できる解読過程を「打譜」（だふ、Da-Pu）といい、専

門家でないと一般の古琴愛好者にとっても極めて困難である[3]。

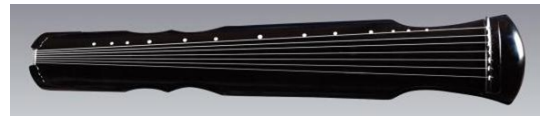


図 1 古琴
Fig. 1 Guqin



図 2 古琴の演奏
Fig. 2 Playing Guqin

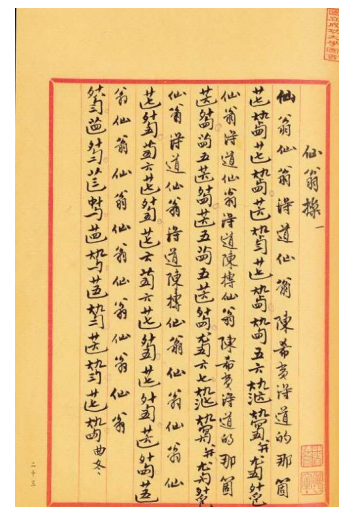


図 3 減字譜「仙翁操」
Fig. 3 Jian-Zi-Pu (Guqin music): Sen-O-So

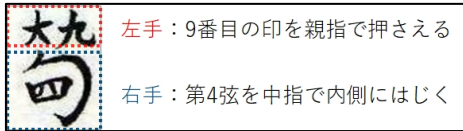


図 4 減字譜単字の例
Fig. 4 A sample of Jian-Zi-Pu

3. 提案システム

本研究では、減字譜の原画像から自動的に古琴曲の映像・音声を出力するシステムを提案する(図4)。具体的には、①減字譜の単字識別器である深層学習モデルYOLOv5を構築する；②学習済みのYOLOv5を用いて原画像から単字を推定・抽出し、位置確定・動画(映像と音声)結合を行い、原曲の動画を出力する。

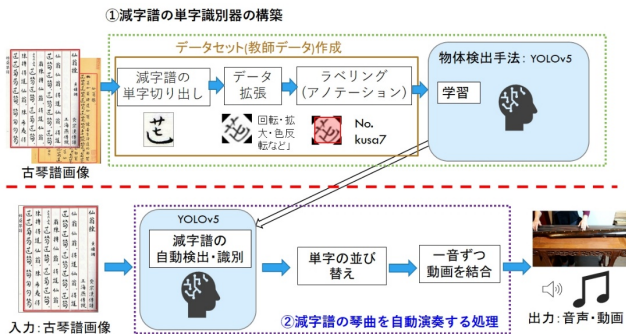


図 4 提案システム

Fig. 4 An AI system for Guqin music representation

3.1 データベースの作成

単字認識するためのデータベースの作成は先行研究[7]の手順によって行う。即ち、まずは複数の同一琴曲の減字譜画像を収集し、それぞれに対し、単字の切り出し及びラベリング(アノテーション)を行い、回転、反転、平行移動、拡大・縮小(ズームイン/アウト)、変形、フィルタなどの処理画像によって単字データセットを作成する(図5)。実験においては、インターネットより琴曲「仙翁操」の23種類減字譜(図3)を収集し、55種類の単字(図6)計343個を切り出し、データ拡張より4,951枚単字画像セットを作成した。また、それぞれの単字画像に対応した動画(映像・音声)データを作成した(図7)。

3.2 深層学習モデルYOLOv5

YOLO(You Only Look Once)はDarknet50をベースとした物体検出のための深層学習モデルで、動画やオンライン処理に適するため、本研究に導入する[9]。これまでの深層学習モデルによる減字譜認識の研究においては、スタンフォード大学のChen Shiの「Guqin Notation and Music Style Recognition」[2]の論文があり、減字譜の弾く弦を示す漢数字部分に着目して分類する畳み込みニューラルネットワ

ーク(CNN)を開発し、10クラスの識別率は89.3%であった。また、著者らの先行研究[7][8]において、VGG16及びVGG16とSVMを併用したハイブリッドモデルを用いて、「仙翁操」の最初二行15クラスの単字認識の識別率(val_acc)はそれぞれ85.09%と88.33%であった。1枚の減字譜画像をそのまま単字を自動的に抽出・認識するため、本研究では、物体検出に設計されたYOLOv5sモデル[9]を導入する。

ファイル名	画像	画像処理の詳細
7(14).jpg		元の画像
7(14).jpg_30.jpg		高さ&幅を30pxに変更
7(14).jpg_100.jpg		高さ&幅を100pxに変更
7(14).jpg_200.jpg		高さ&幅を200pxに変更
7(14).jpg_blur.jpg		カーネルサイズ(7, 7) 標準偏差1.5で平滑化
7(14).jpg_brighter.jpg		色の値を1.5倍
7(14).jpg_darker.jpg		色の値を0.9倍
7(14).jpg_noise.jpg		ランダムな場所を対象にRGBランダムに選んでその値をランダムな少数分とする
7(14).jpg_r45.jpg		反時計回りに45度回転
7(14).jpg_r90.jpg		反時計回りに90度回転
7(14).jpg_r135.jpg		反時計回りに135度回転
7(14).jpg_r180.jpg		反時計回りに180度回転
7(14).jpg_r225.jpg		反時計回りに225度回転
7(14).jpg_r270.jpg		反時計回りに270度回転
7(14).jpg_r315.jpg		反時計回りに315度回転

図 5 データ拡張の例

Fig. 5 A sample of data augmentation

芭	筇	筇	筇	筇	也
筇	筇	筇	筇	筇	清
苜	旬	筇	筇	筇	鼎
芭	旬	筇	筇	筇	鼎
芭	筇	筇	筇	筇	筇
芭	筇	筇	筇	筇	
筇	勻	筇	筇	筇	
筇	勻	筇	筇	筇	
筇	旬	筇	筇	筇	
筇	筇	筇	筇	筇	

図 6 仙翁操の55種単字画像

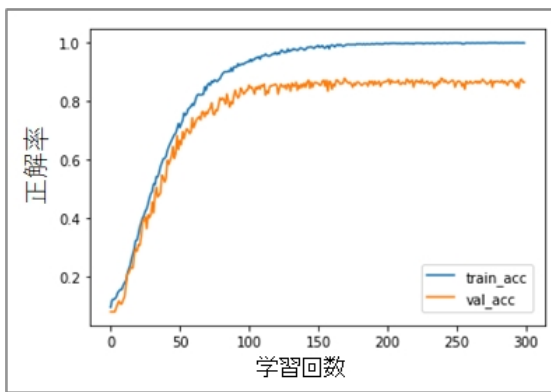
Fig. 6 A sample of 55 single characters of Sen-O-So



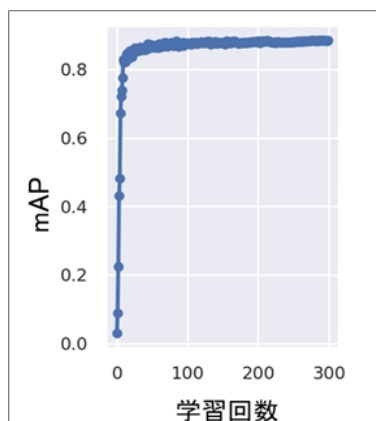
図 7 単字動画データの例
Fig. 7 A sample of video data

4. 実験結果

3.2 節で述べた「仙翁操」の 55 種類の単字データセットを用いた識別実験において、VGG16 と YOLOv5 の学習後の平均識別率(mAP)はそれぞれ 87.50%と 88.47%であった。また、それぞれの学習性能を図 8 に示す。図 8 より、VGG16 の識別率の収束は学習回数 200 要することに対し、YOLOv5 の場合は 100 回の学習で高い識別率に達したことがわかる。



(a) VGG16



(b) YOLOv5

図 8 VGG16 と YOLOv5 の学習性能の比較
Fig. 8 Learning curves of VGG16 and YOLOv5



図 9 YOLOv5 の出力
Fig. 9 An output of YOLOv5

図 9 において、YOLOv5 の 1 枚の減字譜画像に対する各単字の識別結果を示す。各単字の位置も推定できるため、その順番を動画データと結びつことにより、「仙翁操」の動画（映像と音声）を出力することができた（図 10）。

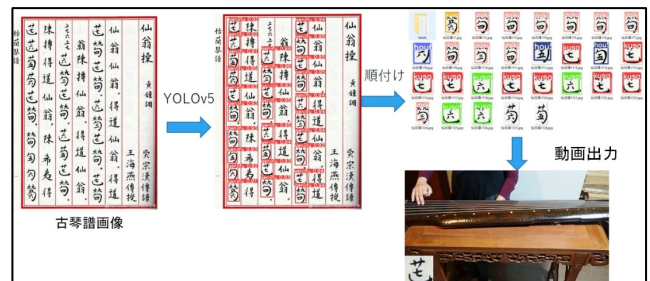


図 10 減字譜から動画（映像と音声）への自動変換
Fig. 10 Transform Jian-Zi-Pu to Video

5. まとめ

古琴の楽譜である減字譜の自動打譜に向けて、本研究では、深層学習モデル YOLOv5 を用いて打譜システムの構築を試みた。平易な琴曲である「仙翁操」の減字譜画像データセットを作成し、減字譜から動画（映像・音声）への自動変換システムが開発できた。一方、減字譜には漢字による歌詞があり、その歌詞の意味を理解し、出力する音声に強弱や緩急などの音楽要素をつけることは琴曲の復元につながるが、今後の課題に残る。

謝 辞

本研究は科研費 (No. 22H03709, No.22K12152, No.20K20500)の助成を受けたものである。

参考文献

- [1] 中国スクリプト, 「琴と箏の歴史と構造の違い」 : <http://chugokugo-script.net/chugoku-bunka/kin.html> (2016)
- [2] 日本古琴振興会, 「古琴の紹介」 : <https://www.guqin.jp/about>
- [3] 梅尾亮子 : 「打譜に関する調査報告」, お茶の水音楽論集, 第 10 号 (2008)
- [4] Zhi-xiao Pan, Chang-le Zhou : “Text Segmentation and Extraction from Images of Guqin Jianzi Pu”, 心智与計算, Vol.1, No.2, pp. 286-295 (2007) (in Chinese)
- [5] Chen Shi: “Guqin Notation and Music Style Recognition”, Computer Science (2016)
- [6] 王 利, 孫 洋, 罗 兆麟, 張輝 : 「AI 自动翻译“减字谱”——以《流水》和《不染》为例」, 芸術教育, Vol.3 (2019)
- [7] 楊 博文, 佐藤真琴, 吳本舜, 小柴満美子, 間普真吾, 稗田浩雄, 吳本 堯 : 「深層学習を用いた古琴譜の識別」, 2022 年電気学会電子・情報・システム部門大会論文集, pp.512-515 (2022)
- [8] Bowen Yang, Shun Kuremoto, Mamiko Koshiba, Shingo Mabu, Hiroo Hieda, Takashi Kuremoto, “Recognition of Guqin Notation using Deep Learning”, in Proceedings of Innovative Application Research and Education (ICIARE2022), pp.61-64 (2022)
- [9] YOLOv5: <https://github.com/ultralytics/yolov5>