NEURAL BAND-TO-PIANO SCORE ARRANGEMENT WITH STEPLESS DIFFICULTY CONTROL

Moyu Terao[1] *Eita Nakamura*[1,2] *Kazuyoshi Yoshii*[1,3]

[1]Graduate School of Informatics, Kyoto University, Japan
[2]The Hakubi Center for Advanced Research, Kyoto University, Japan
[3]PRESTO, Japan Science and Technology Agency (JST), Japan

ABSTRACT

This paper describes a music arrangement method of popular music that can convert a band score into a piano score with a steplessly-specified level of performance difficulty. The basic strategy of band-to-piano score arrangement is to select notes from an augmented band score obtained by up- and down-shifting the notes of an original band score by one octave. Given band scores and the corresponding piano scores with elementary- and advanced-levels, one can train a deep neural network (DNN) that estimates note masks conditioned by the difficulty levels. Conditioned by an intermediate level at run-time, however, the DNN tends to generate an advanced-level score. To solve this problem, assuming that an easier piano score is a subset of harder one, we estimate the basic importance of each note with a difficulty-agnostic DNN and then warp it with a power function depending on a specified difficulty level. To achieve the fine controllability of the difficulty level, we propose a training method that subjects the DNN to generating piano scores with various intermediate levels, where the note-level loss for those scores is evaluated using only the ground-truth elementary- and advanced-level scores. Considering the non-uniqueness of piano arrangement, the statistic-level loss with respect to the note density and polyphony level is also computed according to the given levels. The experimental results showed that the proposed method attained both the performance gain and the stepless difficulty control.

Index Terms— Automatic piano arrangement, score reduction, symbolic music processing, deep learning

1. INTRODUCTION

One typical way of music arrangement is to change the instrumentation of a musical piece while preserving its essential musical content (rhythm, melody, and harmony). Much effort has been devoted to automatic piano arrangement [1–5], guitar arrangement [6–8], and orchestration [9–11]. In this paper we tackle piano arrangement of band music with stepless difficulty control.

One of the most important requirements in band-to-piano score arrangement is to allow a user to steplessly control the difficulty level of the generated piano score according to his or her skill and preference (Fig. 1). If a user gradually moves the slider of the difficulty level from the elementary level to the advanced level, it would be better in terms of usability that the generated piano score also changes gradually, *i.e.*, notes are added gradually. The underlying assumption is that an easier-level piano score can be obtained as a subset of a harder-level score if both scores originate from the same band score. Since the music arrangement is a challenging one-to-many mapping task, this assumption can effectively reduce the search space.

This work is supported in part by JST PRESTO No. JPMJPR20CB and JSPS KAKENHI Nos. 19H04137, 20K21813, and 22H03661.

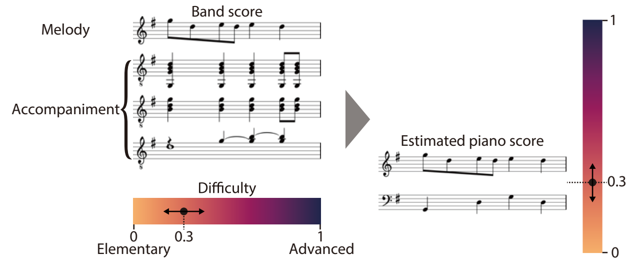

Fig. 1. The proposed band-to-piano score arrangement system with stepless difficulty control.

The score reduction approach has typically been taken for band-to-piano score arrangement. One can obtain a piano score by selecting notes from an augmented band score obtained by up- and down-shifting the notes of an original band score by one octave. Given paired data of band scores and the corresponding piano arrangements, which are typically categorized into only elementary and advanced levels, one can train a DNN that estimates masks used for note selection, conditioned by the difficulty levels [12]. Conditioned by an unseen intermediate level at run-time, however, the DNN trained with binary conditions tends to generate an advanced-level piano score.

Under these circumstances, we propose a difficulty-controllable band-to-piano score arrangement method based on the score reduction approach (Fig 2). This method estimates the *basic* importance ϕ of each note in an augmented band score with a *difficulty-agnostic* DNN and then warps it with a *difficulty-dependent* power function ϕ^α, where a harder level takes a smaller α. If the resulting importance ϕ^α exceeds a threshold, the notes is selected to be included in the piano arrangement. The estimated importance ϕ^α is thus guaranteed to monotonically increase according to the difficulty level. Thanks to the differentiable warping function, the DNN and the values of α for the elementary and advanced levels can be optimized jointly using paired data consisting of only these two levels. The value of α between these two values can then be used for an intermediate level.

To improve the controllability of the difficulty level at run-time, *i.e.*, to let the generated piano score steadily change from the elementary level to the advanced level, we subject the DNN to generating piano scores with various intermediate levels not limited to the elementary and advanced levels in the training phase. The essential problem is that such generated piano scores cannot be evaluated because only the elementary and advanced piano scores are given as the ground-truth data.

Considering the non-uniqueness of band-to-piano score arrangement as suggested in [12], we propose a regularized training method

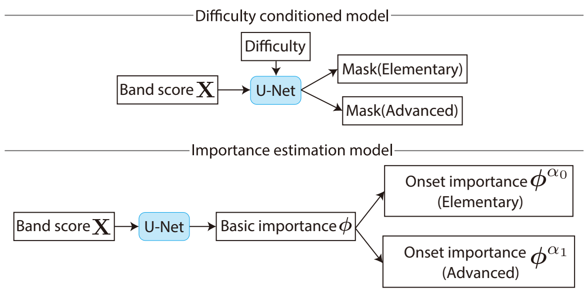

Fig. 2. Comparison of the conventional and proposed methods. The conventional method directly estimates the selection probability for each note of a band score with a DNN conditioned by a difficulty level. The proposed method first estimates the *basic* selection probability for each note with a difficulty-agnostic DNN and then warps it according to the difficulty level.

based on note- and statistic-level (instance- and distribution-level) criteria that can evaluate piano scores with arbitrary difficulty levels. When a piano score with a randomly-sampled level is generated, its note-level loss is a weighted sum of the loss using *available* ground-truth scores. To regularize the training, the statistic-level loss of the generated score is also introduced such that the distributions of the note densities and the polyphony levels are made close to those corresponding to the randomly-sampled level.

The main contribution of this study is to propose neural piano arrangement based on note importance estimation for run-time stepless difficulty control. We experimentally investigate the potential of note importance estimation in difficulty control and the effectiveness of considering the both the note- and static-level losses.

2. RELATED WORK

A standard approach to piano arrangement of multi-instrument music (*e.g.*, orchestra and popular music) is score reduction based on note selection. Huang *et al.* [11] proposed an orchestration method that generates a score of a specified musical instrument from a score of ensemble music by segmenting the instrument tracks in terms of musical content, determining their musical roles (*e.g.*, lead and pad), and selecting phrases. Takamori *et al.* [3] proposed a band-to-piano arrangement method that focuses on musical content such as melodies, chords, rhythms, and the number of notes extracted from an original score. This method selects accompaniment patterns from a database of accompaniment parts collected from piano scores of popular music. They further attempted piano arrangement from audio data [13]. Wang *et al.* [14] pointed out that the use of predefined accompaniment parts leads to a limited variety of arrangements and thus used a DNN for flexible piano arrangement. These arrangement methods focus on the constraints of piano scores in terms of the note density and the number of simultaneous notes.

The player's performance skill has recently been considered as an important factor for music arrangement. Nakamura *et al.* [4] pointed out that the weights of the constraints on piano score should be varied according to the player's skill and that more strict constraints are needed for elementary-level players. They proposed a method that represents the performance difficulty as a continuous variable and also compared the difficulty levels in terms of the frequency of performance errors. Our previous band-to-piano arrangement method can generate only elementary- and advanced-level piano scores, where the note statistics are considered to deal with the non-uniqueness of piano arrangement [12].

3. PROPOSED METHOD

This section describes the proposed method that converts a band score into a piano score with an arbitrary difficulty level.

3.1. Problem Specification

Our goal is to convert a band score $\mathbf{X} \triangleq \{\mathbf{X}_{\mathrm{A}}, \mathbf{X}_{\mathrm{M}}\}$ into a piano score $\mathbf{Y} \triangleq \{\mathbf{Y}_{\mathrm{L}}, \mathbf{Y}_{\mathrm{R}}\}$ with an arbitrary difficulty level $\gamma \in [0,1]$, where $\mathbf{X}_{\mathrm{A}} \triangleq \{\mathbf{O}_{\mathrm{A}}, \mathbf{P}_{\mathrm{A}}\}$ and $\mathbf{X}_{\mathrm{M}} \triangleq \{\mathbf{O}_{\mathrm{M}}, \mathbf{P}_{\mathrm{M}}\}$ represents the condensed accompaniment part and the melody (vocal) parts, respectively, and $\mathbf{Y}_{\mathrm{L}} \triangleq \{\mathbf{O}_{\mathrm{L}}, \mathbf{P}_{\mathrm{L}}\}$ and $\mathbf{Y}_{\mathrm{R}} \triangleq \{\mathbf{O}_{\mathrm{R}}, \mathbf{P}_{\mathrm{R}}\}$ represent the left- and right-hand parts, respectively. Each part is represented as an onset matrix $\mathbf{O}_* \in \{0,1\}^{P\times N}$ and a pitch matrix $\mathbf{P}_* \in \{0,1\}^{P\times N}$ $(* \in \{\mathrm{A},\mathrm{M},\mathrm{L},\mathrm{R}\})$, where P represents the number of pitches ($P = 128$) and N represents the number of tatums (sixteenth notes) of the piece. In the melody part of the band score, for example, $\mathbf{O}_{\mathrm{M}}(p,n) = 1$ represents the presence of an onset at pitch p and tatum n and $\mathbf{P}_{\mathrm{M}}(p,n) = 1$ represents the presence of pitch p at tatum n. Let $\mathrm{h} \in \{\mathrm{L},\mathrm{R}\}$ denote the left or right hand part.

To train a DNN-based conversion model, we use pairs of band and piano scores as training data. Note that each $\mathbf{X}$ in the training data is associated with a piano score $\mathbf{Y}$ with either of the elementary and advanced levels, *i.e.*, intermediate-level piano scores are not included in the training data.

3.2. Score Reduction Approach

We take the score reduction approach to band-to-piano arrangement. The statistical analysis of existing piano arrangements showed that on average, 92% of the notes on the right-hand part and 81% of the left-hand part are derived from the notes from an *augmented* band score obtained by up- and down-shifting the notes of an original band score by one octave [12]. This fact justifies the assumption underlying our score reduction approach that a reasonable piano score can be obtained by selecting important notes, with octave shifts if necessary, from an augmented band score, *i.e.*, a piano score can be obtained as a subset of an augmented band score. Let $\mathbf{O}_{\mathrm{B}} \in \{0,1\}^{P\times N}$ be an *augmented* onset matrix given by

$$\mathbf{O}_{\mathrm{B}}(p,n) = \max_{j\in\{-12,0,12\}} \left(\mathbf{O}_{\mathrm{A}}(p+j,n), \mathbf{O}_{\mathrm{M}}(p+j,n)\right), \quad (1)$$

where $j \in \{-12, 0, 12\}$ represents the octave shift.

Given a difficulty level γ, we estimate selection probability (note importance) matrices $\boldsymbol{\pi} \triangleq \{\boldsymbol{\pi}_{\mathrm{L}}, \boldsymbol{\pi}_{\mathrm{R}}\}$ for the notes of an augmented band score, where $\boldsymbol{\pi}_{\mathrm{L}}, \boldsymbol{\pi}_{\mathrm{R}} \in [0,1]^{P\times N}$ represent the importance matrices used for estimating the left- and right-hand parts of the piano score. The note importance is defined at only the note onsets of the augmented band score, *i.e.*, if $\mathbf{O}_{\mathrm{B}}(p,n) = 0$, $\boldsymbol{\pi}_{\mathrm{L}}(p,n) = \boldsymbol{\pi}_{\mathrm{R}}(p,n) = 0$. The piano onset matrices $\mathbf{O}_{\mathrm{L}}$ and $\mathbf{O}_{\mathrm{R}}$ are obtained by binarizing $\boldsymbol{\pi}_{\mathrm{L}}$ and $\boldsymbol{\pi}_{\mathrm{R}}$ with a threshold, respectively. The piano pitch matrices $\mathbf{P}_{\mathrm{L}}$ and $\mathbf{P}_{\mathrm{R}}$ are determined by referring to the onsets and durations of the selected notes of the augmented band score.

3.3. Supervised Training

In the training phase, a pair of a band score $\mathbf{X}$ and the corresponding elementary- *or* advanced-level piano score $\mathbf{Y}$ is given. A lot of such pairs are given as training data in practice, but no intermediate-level piano scores are available, *i.e.*, only elementary- and advanced-level piano scores can be used as target data. Using such training data, we jointly optimize the U-Net parameters and the power exponents α_0 and α_1 corresponding to the elementary and advanced levels. A key

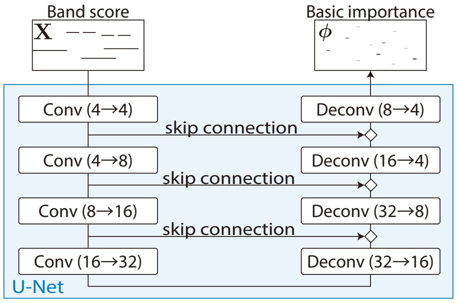

Fig. 3. The detailed architecture of the U-Net. The numbers in the parentheses represent the change in the number of channels. Skip connection denotes concatenation of channels.

feature of the proposed method is that piano scores with randomly-sampled difficulty levels not limited to the elementary and advanced levels are generated in the training phase.

We first estimate *difficulty-independent* importance matrices $\boldsymbol{\phi} \triangleq \{\boldsymbol{\phi}_{\mathrm{L}}, \boldsymbol{\phi}_{\mathrm{R}}\}$ with a convolutional neural network (CNN) called the U-Net [15] that takes the band score $\mathbf{X}$ as input (Fig. 3), where $\boldsymbol{\phi}_{\mathrm{L}}, \boldsymbol{\phi}_{\mathrm{R}} \in [0,1]^{P\times N}$ represent the basic importance values over the pitch-tatum space used for estimating the left- and right-hand parts, respectively. To estimate the *difficulty-dependent* matrices $\boldsymbol{\pi}$, we then warp $\boldsymbol{\phi}$ with an element-wise power function as follows:

$$\boldsymbol{\pi}_{\mathrm{L}} = \boldsymbol{\phi}_{\mathrm{L}}^{\alpha_\gamma} \odot \mathbf{O}_{\mathrm{B}}, \quad \boldsymbol{\pi}_{\mathrm{R}} = \boldsymbol{\phi}_{\mathrm{R}}^{\alpha_\gamma} \odot \mathbf{O}_{\mathrm{B}}, \tag{2}$$

where $\odot$ denotes the element-wise product and α_γ is a power exponent depending on the difficulty level $\gamma \in [0,1]$. Note that a lower α_γ is used for a higher level γ, *i.e.*, $\alpha_0 \geq \alpha_\gamma \geq \alpha_1$ and $\boldsymbol{\phi}_{\mathrm{h}}^{\alpha_0} \leq \boldsymbol{\phi}_{\mathrm{h}}^{\alpha_\gamma} \leq \boldsymbol{\phi}_{\mathrm{h}}^{\alpha_1}$ (Fig. 4). The power exponent α_γ corresponding to an intermediate level γ is computed with linear interpolation between α_0 and α_1 corresponding to the elementary and advanced levels (optimized as explained later) as follows:

$$\alpha_\gamma = (1-\gamma)\alpha_0 + \gamma\alpha_1. \tag{3}$$

Using a difficulty-independent threshold for $\boldsymbol{\pi}_{\mathrm{L}}$ and $\boldsymbol{\pi}_{\mathrm{R}}$, a larger number of notes are selected for a higher-level piano score.

Considering the non-uniqueness of piano arrangement, the importance matrices $\boldsymbol{\pi}$ given by Eq. (2) are evaluated at the note and statistic levels (instance and distribution-level matching). When a target level γ is equal to the difficulty level (0 or 1) of the ground-truth piano score, the note- and statistic-level losses can be computed in the same way as [12]. We here propose a novel regularized training method that minimizes the sum of the note- and statistic-level losses computable for an arbitrary level γ.

3.3.1. *Note-Level Loss*

We define the note-level loss for evaluating the importance matrices $\boldsymbol{\pi}$ estimated for a difficulty level $\gamma \in [0,1]$. If the ground-truth onset matrices $\mathbf{O} \triangleq \{\mathbf{O}_{\mathrm{L}}, \mathbf{O}_{\mathrm{R}}\}$ of the left- and right-hand parts with the same difficulty level γ are available, one can use the modified binary cross entropy defined as follows:

$$\mathrm{BCE}(\boldsymbol{\pi}, \mathbf{O}) = -\sum_{h\in\{\mathrm{L,R}\}} \sum_{p=1}^{P} \sum_{n=1}^{N} \big(w \cdot \mathbf{O}_{\mathrm{h}}(p,n) \log \boldsymbol{\pi}_{\mathrm{h}}(p,n) + \big(1-\mathbf{O}_{\mathrm{h}}(p,n)\big) \log\big(1-\boldsymbol{\pi}_{\mathrm{h}}(p,n)\big)\big), \tag{4}$$

where $w \geq 0$ is a weighting factor used for compensating for the imbalance of the numbers of onset and non-onset frames ($w = 4$ in this paper). This strategy, however, cannot be applied to the case of $\gamma \notin \{0,1\}$. We thus propose to use the following loss:

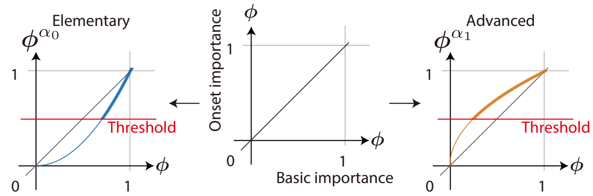

Fig. 4. Note importance warping with power functions.

$$\mathcal{L}^{\mathrm{nt}} = \begin{cases} (1-\gamma)\mathrm{BCE}(\boldsymbol{\pi}, \mathbf{O}) \text{ if level}(\mathbf{O}) = 0, \\ \gamma\mathrm{BCE}(\boldsymbol{\pi}, \mathbf{O}) \text{ if level}(\mathbf{O}) = 1, \end{cases} \tag{5}$$

where $\mathrm{level}(\mathbf{O})$ represents the difficulty level of $\mathbf{O}$. For example, when $\gamma = 0.3$ (closer to the elementary level), the binary cross entropy of $\boldsymbol{\pi}$ for the elementary- or advanced-level piano score is considered with a wight of 0.7 or 0.3, respectively.

3.3.2. *Statistic-Level Losses*

We define the statistic-level losses used for regularizing the supervised training of the U-Net with the note-level loss. Let $\mathbf{C}_{\mathrm{h}}^{\mathrm{lv}}(n)$ denote the polyphony level (the number of concurrent pitches) at tatum n, and let $\mathbf{C}_{\mathrm{h}}^{\mathrm{ds}}(m)$ denote the note density (the number of onsets in a measure) at measure m, in the estimated piano onset matrices $\mathbf{O}$ ($\mathrm{h} \in \{\mathrm{L,R}\}$), which are given by

$$\mathbf{C}_{\mathrm{h}}^{\mathrm{lv}}(n) = \sum_{p=1}^{P} \mathbf{O}_{\mathrm{h}}(p,n), \ \mathbf{C}_{\mathrm{h}}^{\mathrm{ds}}(m) = \sum_{n\in G(m)} \sum_{p=1}^{P} \mathbf{O}_{\mathrm{h}}(p,n). \tag{6}$$

where $G(m)$ represents the set of tatum indices in measure m. Note that the piano onset matrices $\mathbf{O}_{\mathrm{L}}$ and $\mathbf{O}_{\mathrm{R}}$ are stochastically determined according to $\boldsymbol{\pi}_{\mathrm{L}}$ and $\boldsymbol{\pi}_{\mathrm{R}}$, respectively, in a differentiable manner with the gumbel sigmoid function [16], instead of performing simple thresholding used at run-time.

We then compute the empirical distribution (histogram) $\mathbf{Q}_{\mathrm{h}}^{\mathrm{d}} \in [0,1]^{I^{\mathrm{d}}+1}$ by normalizing the tatum- or measure-level frequencies $\mathbf{C}_{\mathrm{h}}^{\mathrm{d}}$ over all possible values, where I^{d} represents the maximum polyphony level or the maximum note density and $\mathrm{d} \in \{\mathrm{lv}, \mathrm{ds}\}$ represents the polyphony level or the note density.

We aim to make the histogram $\mathbf{Q}_{\mathrm{h}}^{\mathrm{d}}$ close to the ground-truth histogram $\bar{\mathbf{Q}}_{\mathrm{h},\gamma}^{\mathrm{d}}$ corresponding to the difficulty level γ in a distributional sense. What we can do with the training data, however, is to estimate the ground-truth histograms $\bar{\mathbf{Q}}_{\mathrm{h},0}^{\mathrm{d}}$ and $\bar{\mathbf{Q}}_{\mathrm{h},1}^{\mathrm{d}}$ from the existing elementary- and advanced-level piano scores, respectively. Another key feature of the proposed method is to estimate the intermediate histogram $\bar{\mathbf{Q}}_{\mathrm{h},\gamma}^{\mathrm{d}}$ with Wasserstein interpolation between $\bar{\mathbf{Q}}_{\mathrm{h},0}^{\mathrm{d}}$ and $\bar{\mathbf{Q}}_{\mathrm{h},1}^{\mathrm{d}}$ [17]. The static loss with respect to $\mathrm{d} \in \{\mathrm{lv}, \mathrm{ds}\}$ is given by the Jensen-Shannon (JS) divergence as follows:

$$\mathcal{L}^{\mathrm{d}} = \sum_{\mathrm{h}\in\{\mathrm{L,R}\}} \mathcal{D}_{\mathrm{JS}}\left(\bar{\mathbf{Q}}_{\mathrm{h},\gamma}^{\mathrm{d}} \parallel \mathbf{Q}_{\mathrm{h}}^{\mathrm{d}}\right). \tag{7}$$

4. EVALUATION

This section reports an experiment conducted for evaluating the effectiveness of the proposed method.

4.1. Experimental Conditions

We collected 184 pairs of band and piano scores (85 elementary-level arrangements and 99 advanced-level arrangements). We used randomly-chosen 138 pairs as a training set, and the remaining 46 pairs as a test set. We set the analysis window size to 12 measures

Table 1. The F-measures were evaluated for the ground-truth scores (left side of slash) and for their modified version obtained by removing notes not included in the augmented scores (right side of slash). The statistic-level losses were measured for three difficulty levels $\gamma = 0, 0.5, 1$

	Losses						$\mathcal{F}$ [%]		$\mathcal{L}^{\mathrm{lv}}$ $(\times 10^2)$			$\mathcal{L}^{\mathrm{ds}}$ $(\times 10^2)$		
	$\mathcal{L}^{\mathrm{nt}}$	$\mathcal{L}^{\mathrm{lv}}$	$\mathcal{L}^{\mathrm{ds}}$	$\mathcal{L}^{\mathrm{nt}}_{\mathrm{med}}$	$\mathcal{L}^{\mathrm{lv}}_{\mathrm{med}}$	$\mathcal{L}^{\mathrm{ds}}_{\mathrm{med}}$	Left	Right	Elm	Med	Adv	Elm	Med	Adv
Conventional	✓						20.9 / 22.8	55.3 / 58.0	15.3	12.6	11.4	5.54	17.2	10.1
Proposed	✓						**25.3 / 28.0**	56.2 / 58.6	15.9	8.82	**8.43**	5.43	**3.55**	**5.35**
	✓	✓	✓				25.2 / 27.5	**57.9 / 60.7**	**11.0**	**7.41**	9.66	**4.81**	4.11	9.70
				✓			24.8 / 26.7	57.8 / 60.6	11.4	9.29	13.4	9.65	7.63	13.9
				✓	✓	✓	24.8 / 26.8	55.9 / 58.4	11.4	7.96	13.7	6.87	5.43	14.6

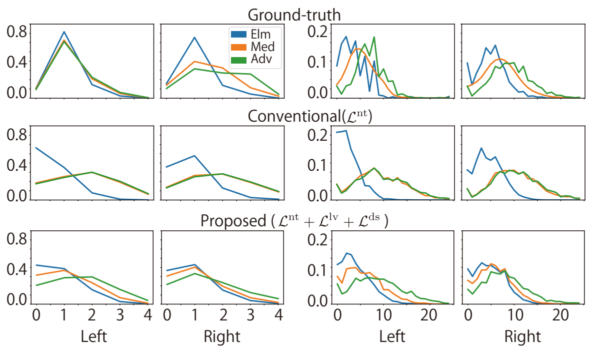

Fig. 5. The distributions of polyphony levels for the left- and right-hand parts with difficulty levels $\gamma = 0, 0.5, 1$. The top row shows the ground-truth distributions for $\gamma = 0, 1$ and their barycenter for $\gamma = 0.5$. The middle and bottom rows show the distributions computed from piano scores estimated by the conventional and proposed methods.

and assumed the time signature to be 4/4. When the length of a measure was shorter than 16 tatums, it was fill with silent. Conversely, when a measure was longer than 16 tatums, the first 16 tatums were preserved. Data augmentation was performed on the training set by transposing the keys by from 1 to 11 semitones.

For the U-Net, we set the kernel size to 4, stride to 2, and padding to 1. Dropout ($p = 0.5$) was applied to all the deconvolutional layers. The Adam optimizer with a learning rate of 10^{-4} was used for training [18]. We set the threshold to 0.5. The note-level and statistic-level losses were treated equally.

4.2. Experimental Results

We examined ground-truth piano scores and investigated how much notes of the elementary-level scores were included in the advanced-level scores. In 55 pairs of elementary- and advanced-level piano scores, 75% of the right-hand notes and 47% of the left-hand notes of the elementary-level scores were included in the advanced-level scores. This result shows that for the right-hand part the assumption that an elementary-level score is a subset of an advanced-level score is appropriate, and for the left-hand part the assumption might not hold. Considering the non-uniqueness of piano arrangement, this assumption still seemed reasonable for our task to effectively reduce the solution space.

Table 1 shows the evaluation results. The proposed method outperformed the conventional method in terms of not only $\mathcal{F}$ but also $\mathcal{L}^{\mathrm{lv}}$ and $\mathcal{L}^{\mathrm{ds}}$. We found that $\mathcal{L}^{\mathrm{lv}}$ and $\mathcal{L}^{\mathrm{ds}}$ were reduced for intermediate levels. Using intermediate scores ($\mathcal{L}_{\mathrm{med}}$) with random difficulty levels was not found to be effective. Fig. 5 shows that the proposed method is capable of fine control at intermediate levels.

Fig. 6 shows examples of piano arrangement estimated with the proposed model trained with the statistic-level losses without the intermediate losses, where γ represents the difficulty level. It was observed that the more difficult the piano score, the greater the number of notes (see other examples on our webpage[1]).

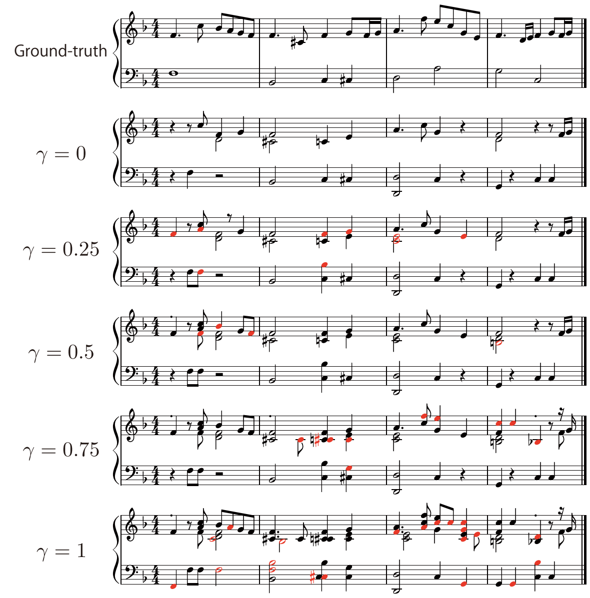

Fig. 6. The output piano score $\mathbf{Y}$ estimated by the model optimized with $\mathcal{L}^{\mathrm{nt}} + \mathcal{L}^{\mathrm{lv}} + \mathcal{L}^{\mathrm{ds}}$ for a difficulty level γ. The larger γ, the more difficult the piano score. The red notes represent the notes added to the one-step easier score.

5. CONCLUSION

We proposed a neural piano arrangement method that can control the difficulty levels of output piano scores in a stepless manner. This method uses a difficulty-agnostic U-Net for estimating the basic importance of each note in an augmented band score and then warps it with a power function depending on a specified difficulty level. Given a band score, an easier-level piano score can thus be obtained as a subset of harder-level one.

Experimental results showed that the regularized training with the note- and statistic-level losses improved the onset match rate and statistical characteristics. We quantitatively confirmed that the proposed method can control the difficulty level steplessly. On the other hand, we have observed that the intermediate loss does not effectively improve the onset match rate and statistic distance. We plan to conduct a subjective evaluation experiment to examine the validity of the output piano scores for human performance.

[1] https://teraomoyu.github.io/difficulty-controllable-piano-arrangement.github.io/

6. REFERENCES

[1] S. Chiu, M. Shan, and J. Huang, "Automatic system for the arrangement of piano reduction," in *Proc. International Symposium on Multimedia*, 2009, pp. 459–464.

[2] S. Onuma and M. Hamanaka, "Piano arrangement system based on composers' arrangement processes," in *Proc. International Computer Music Conference*, 2010, pp. 191–194.

[3] H. Takamori, H. Sato, T. Nakatsuka, and S. Morishima, "Automatic arranging musical score for piano using important musical elements," in *Proc. Sound and Music Computing Conference*, 2017, pp. 35–41.

[4] E. Nakamura and K. Yoshii, "Statistical piano reduction controlling performance difficulty," *APSIPA Transactions on Signal and Information Processing*, vol. 7, no. e13, pp. 1–12, 2018.

[5] Z. Wang, K. Chen, J. Jiang, Y. Zhang, M. Xu, S. Dai, X. Gu, and G. Xia, "Pop909: A pop-song dataset for music arrangement generation," in *Proc. International Society for Music Information Retrieval*, 2020, pp. 38–45.

[6] D. Tuohy and WD. Potter, "A genetic algorithm for the automatic generation of playable guitar tablature," in *Proc. International Computer Music Conference*, 2005, pp. 499–502.

[7] Y. Yoshinaga, S. Fukayama, H. Kameoka, and S. Sagayama, "Automatic arrangement for guitars using hidden Markov model," in *Proc. Sound and Music Computing Conference*, 2012, pp. 450–456.

[8] G. Hori, H. Kameoka, and S. Sagayama, "Input-output HMM applied to automatic arrangement for guitars," *Journal of Information Processing*, vol. 21, no. 2, pp. 264–271, 2013.

[9] H. Maekawa, N. Emura, M. Miura, and M. Yanagida, "On machine arrangement for smaller wind-orchestras based on scores for standard wind-orchestras," in *Proc. International Conference on Music Perception and Cognition*, 2006, pp. 268–273.

[10] L. Crestel and P. Esling, "Live orchestral piano, a system for real-time orchestral music generation," in *Proc. Sound and Music Computing Conference*, 2017, pp. 434–442.

[11] J. Huang, S. Chiu, and M. Shan, "Towards an automatic music arrangement framework using score reduction," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 8, no. 1, pp. 8:1–8:23, 2012.

[12] M. Terao, Y. Hiramatsu, R. Ishizuka, Y. Wu, and K. Yoshii, "Difficulty-aware neural band-to-piano score arrangement based on note- and statistic-level criteria," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 196–200.

[13] H. Takamori, T. Nakatsuka, S. Fukayama, M. Goto, and S. Morishima, "Audio-based automatic generation of a piano reduction score by considering the musical structure," in *Proc. International Conference on Multimedia Modeling*, 2019, pp. 169–181.

[14] Z. Wang, D. Xu, G. Xia, and Y. Shan, "Audio-to-symbolic arrangement via cross-modal music representation learning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 181–185.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[16] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with Gumbel-Softmax," in *Proc. International Conference on Learning Representations*, 2017.

[17] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," in *SIAM Journal on Mathematical Analysis*, 2011, pp. 904–924.

[18] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *The International Conference on Learning Representations (ICLR)*, 2015.