Music generation using Deep Learning

Gábor Lant
Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics
Budapest, PA 1111
lant.gabor98@gmail.com

Dániel M. Szalai
Department of Networked Systems and Services
Budapest University of Technology and Economics
Budapest, PA 1111
szalaimd@gmail.com

Attila Kádár
Department of Measurement and Information Systems
Budapest University of Technology and Economics
Budapest, PA 1111
attilaka98@gmail.com

Abstract

In this paper we try to demonstrate how we can generate music using neural networks from sound files converted to images. The goal is to compare the performance of different types of neural networks and also different types of music to image conversions and how they affect the generated output.

1 Problem summary

Our idea was to convert our music data into pictures and then use some kind of neural network to produce similar pictures. Then we would convert our generated images back to music format and rate the quality. Originally we hoped we could use different styles of music to use as an input for our model and then it would generate a similar music but in a different style, but later we found this to be much more difficult. We will discuss this later in section **??**.

2 Data acquisition, description

For our data-set we had two criteria in mind: we needed .midi files and we wanted to use tracks similar to each other in style. The .midi files allow us to work with large data-sets (25- 100GB as uncompressed .wav files) without actually having to store them as .wavs thus taking up only minimal storage capacity (not counting the preprocessed files). Them being similar in style is not necessary, however we thought that it would be best for our first approach to only consider tracks from the same style. Luckily we found a well-made data-set called Magenta: Maestro (1), that satisfies all our needs. Maestro is a dataset composed of over 200 hours of virtuosic piano performances.

3 Data exploration

Our data has the following main features:

- Time: Absolute time, in terms of MIDI clocks, at which this event occurs. Meta-events for which time is not meaningful (for example, song title, copyright information, etc.) have an absolute time of 0.

- Type/Event: Name identifying the type of the record. Record types are text consisting of upper and lower case letters and the underscore ("_"), contain no embedded spaces, and are not enclosed in quotes.

- Channel: The channel identifier.

- Note: The currently played note (integer 21-109, a piano has 88 keys)

- Velocity: The matching note's volume at the moment.

Further details at midicsv's website [1]

4 Data preprocessing

Here we only explain the main steps and ideas of our preprocessing, because we include a well-commented source code, that produces our actual training (70%), validation (20%) and test (10%) data. First we read the .midi files and transform them to only contain the relevant informations we need (notes length, instrument type, tempo, volume), then we transform them into 10 sec long parts (if the music is not a multiple of 10 sec long, then we "add silence" to the end). The fixed time long parts can then be transformed into pictures, where the RGBA channels are the notes length, the instrument type, the tempo and the volume respectively. The pixels height indicates the note itself. In
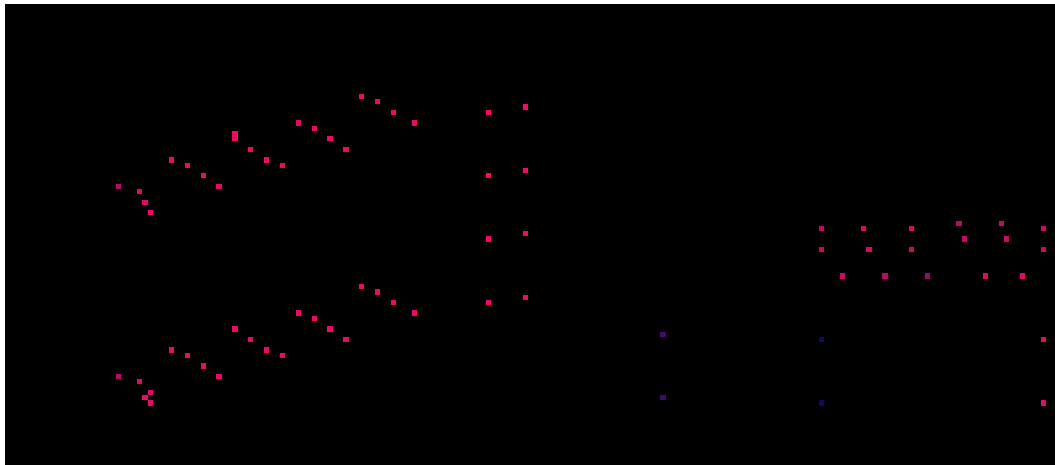

Figure 1: Sample figure caption.

our first approach this will be the way we process the data, but we would like to note here, that we are thinking about different approaches as well. Another promising way can be the usage of Hilbert's curves. In a nuttshell: Hilbert's curves (2) is a way to map the 1D-like spectrogram to a 2D picture, in such way, that the notes close to each other on the spectrogram will also be close on the 2D picture. This approach might work better with convolutional networks since the small matrices, parsing the picture can actually extract features that correlates more.

---

[1] MIDICSV: Convert MIDI File to and from CSV - Fourmilab. `https://www.fourmilab.ch/webtools/midicsv/`

5   folyt

6   Submission of papers to NIPS 2016

There is a new style file for papers submitted in 2016!

NIPS requires electronic submissions. The electronic submission site is

https://cmt.research.microsoft.com/NIPS2016/

Please read carefully the instructions below and follow them faithfully.

6.1   Style

Papers to be submitted to NIPS 2016 must be prepared according to the instructions presented here. Papers may only be up to eight pages long, including figures. Since 2009 an additional ninth page *containing only acknowledgments and/or cited references* is allowed. Papers that exceed nine pages will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2016 are the same as since 2007, which allow for $\sim 15\%$ more words in the paper compared to earlier years.

Authors are required to use the NIPS LaTeX style files obtainable at the NIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

6.2   Retrieval of style files

The style files for NIPS and other conference information are available on the World Wide Web at

http://www.nips.cc/

The file `nips_2016.pdf` contains these instructions and illustrates the various formatting requirements your NIPS paper must satisfy.

The only supported style file for NIPS 2016 is `nips_2016.sty`, rewritten for LaTeX 2_ε. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

The new LaTeX style file contains two optional arguments: `final`, which creates a camera-ready copy, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

At submission time, please omit the `final` option. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `nips_2016.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in Sections 7, 8, and 9 below.

7   General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section 9 regarding figures, tables, acknowledgments, and references.

8 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

8.1 Headings: second level

Second-level headings should be in 10-point type.

8.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a \paragraph command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

9 Citations, figures, tables, references

These instructions apply to everyone.

9.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

> http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf

Of note is the command \citet, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `nips_2016` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{nips_2016}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous."

Table 1: Sample table title

	Part		Size (μm)
Name	Description		
Dendrite	Input terminal		$\sim$100
Axon	Output terminal		$\sim$10
Soma	Cell body		up to 10^6

9.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[2] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[3]

9.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.
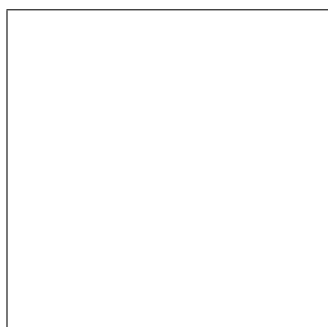

Figure 2: Sample figure caption.

9.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

```
https://www.ctan.org/pkg/booktabs
```

This package was used to typeset Table 1.

---

[2] Sample of the first footnote.

[3] As in this example.

10 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

11 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

    ```
    \usepackage{amsfonts}
    ```

    followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

    ```
    \newcommand{\RR}{I\!\!R} %real numbers
    \newcommand{\Nat}{I\!\!N} %natural numbers
    \newcommand{\CC}{I\!\!\!\!C} %complex numbers
    ```

    Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

11.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font

size to `small` (9 point) when listing the references. **Remember that you can use a ninth page as long as it contains *only* cited references.**

References

[1] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. *Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset.*, `https://arxiv.org/abs/1810.12247` In International Conference on Learning Representations, 2019.

[2] Grant Sanderson (3Blue1Brown), *Hilbert's Curve: Is infinite math useful?*, `https://www.youtube.com/watch?v=3s7h2MHQtxc`