Explicit Pitch and Rhythm Factorization for Robust Music Generation Under Incomplete Context

Assignee Research

June 12, 2026

Abstract

Drawing an analogy with automatic image completion systems, we propose Music SketchNet, a neural network framework that allows users to specify partial musical ideas guiding automatic music generation. We focus on generating the missing measures in incomplete monophonic musical pieces, conditioned on surrounding context, and optionally guided by user-specified pitch and rhythm snippets. First, we introduce SketchVAE, a novel variational autoencoder that explicitly factorizes rhythm and pitch contour to form the basis of our proposed model. Then we introduce two discriminative architectures, Sk

1 Introduction

This paper examines: Music SketchNet: Controllable Music Generation via Factorized Representations of Pitch and Rhythm. Research question: Does explicit pitch and rhythm factorization improve the robustness of music generation models against incomplete context compared to unconstrained latent diffusion approaches?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.3/10.

3 Results

13 papers retrieved. 14 claims extracted; 10 independently verified. Quality review score: 7.3/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
MuseGan allows users to condition generated results on full-length multi-track music.	✓	0.21
DeepBach provides a constraint mechanism that allows users to limit the generated results to match composer styles.	✓	0.22
Music Transformer supports an accompaniment arrangement from an existing melody track in classical music.	✓	0.17
Prior approaches including MuseGan, DeepBach, and Music Transformer require user preference to be defined in terms of co	×	0.13
Music SketchNet allows users to specify partial musical ideas in terms of incomplete and distinct pitch and rhythm repre	✓	0.24
Music InpaintNet completes musical pieces by predicting vector representations for missing measures which are then decod	✓	0.22
SketchVAE is a variational autoencoder that explicitly factorizes rhythm and pitch contour.	✓	0.20
SketchInpainter and SketchConnector are discriminative architectures used in conjunction to perform guided music complet	✓	0.26
The model was evaluated on a standard dataset of Irish folk music.	✓	0.16
In terms of objective metrics on the Irish and Scottish monophonic music dataset, SketchNet achieved a loss of 0.516, pA	×	0.15
In terms of objective metrics on the Irish and Scottish monophonic music dataset, Music InpaintNet achieved a loss of 0.	✓	0.15
SketchNet outperforms Music InpaintNet in pitch accuracy (pAcc) on the evaluated dataset (0.651 vs 0.511).	×	0.10
SketchNet outperforms Music InpaintNet in loss reduction on the evaluated dataset (0.516 vs 0.662).	×	0.07
The proposed approach outperforms state-of-the-art models in subjective listening tests.	✓	0.19

References

- http://arxiv.org/abs/2008.01291v1
- http://arxiv.org/abs/2507.20128v2
- http://arxiv.org/abs/2010.08091v1