

第1章 生成器の構成ブロック

1.1 はじめに

この章では、生成器を構成する各種ブロックの詳細を説明する．まず有音程楽器と打楽器に共通の Base Block を説明する．そして、有音程楽器ピアノロールを生成する Tonal Block, Piano Block, Guitar Block, Bass Block, Strings Block を説明する．さらに、打楽器ピアノロールを生成する Percussion Block, Drums Block, Other Percussion Block を説明する．

1.2 Base Block

Base Block では楽曲全体に共通する特徴量を生成する．このブロックは、入力されるガウス乱数ベクトル $\boldsymbol{\rho} \in \mathbb{R}^{128}$ から、

$$\mathbf{x}_{BB} = \text{LearkyReLU}(W_{BB}\boldsymbol{\rho} + \mathbf{b}_{BB}) \in \mathbb{R}^{128} \quad (1.1)$$

を生成する．ここで、 $W_{BB} \in \mathbb{R}^{128 \times 128}$ であり、 $\mathbf{b}_{BB} \in \mathbb{R}^{128}$ である．式 (1.1) の \mathbf{x}_{BB} が全トラック共通の特徴量であり、これをもとに Tonal Block と Percussion Block がそれぞれ独立に有音程楽器と打楽器の特徴量を生成する．

1.3 Tonal Block

Tonal Block は、全トラック共通特徴量 $\mathbf{x}_{BB} \in \mathbb{R}^{128}$ から有音程楽器共通の特徴量を以下のように生成する．まず、Base Block から入力された特徴ベクトル \mathbf{x}_{BB} から、

$$\mathbf{X}_{TB}^{(0)} = \text{Reshape}_{TB}(W_{TB}^{(0)}\mathbf{x}_{BB} + \mathbf{b}_{TB}) \in \mathbb{R}^{256 \times 3 \times 1 \times 1} \quad (1.2)$$

を生成する．ここで、 $W_{TB}^{(0)} \in \mathbb{R}^{768 \times 128}$ は矩形行列であり、 $\mathbf{b}_{TB} \in \mathbb{R}^{768}$ はバイアスベクトルである．また、 Reshape_{TB} は 768 次元ベクトルを $256 \times 3 \times 1 \times 1$ の 4 階テンソルに並び替える関数である．

続いて、6 段の転置畳み込み層によって処理を継続する。まず、第 1 転置畳み込み層が $\mathbf{X}_{TB}^{(0)} \in \mathbb{R}^{256 \times 3 \times 1 \times 1}$ から

$$\mathbf{X}_{TB}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}^{(0)}, \mathcal{K}_{TB}^{(1)}, \mathbf{s}_{TB}^{(1)})) \in \mathbb{R}^{128 \times 4 \times 1 \times 1} \quad (1.3)$$

を生成する。ここで、 $\mathcal{K}_{TB}^{(1)} \in \mathbb{R}^{128 \times 256 \times 2 \times 1 \times 1}$ であり、 $\mathbf{s}_{TB}^{(1)} = (1, 1, 1)$ である。この段階で第 2 モードのサイズが 4 になり、4 小節分の特徴生成を開始する。

第 2 転置畳み込み層は、 $\mathbf{X}_{TB}^{(1)} \in \mathbb{R}^{128 \times 4 \times 1 \times 1}$ から

$$\mathbf{X}_{TB}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}^{(1)}, \mathcal{K}_{TB}^{(2)}, \mathbf{s}_{TB}^{(2)})) \in \mathbb{R}^{128 \times 4 \times 1 \times 3} \quad (1.4)$$

を生成する。ここで、 $\mathcal{K}_{TB}^{(2)} \in \mathbb{R}^{128 \times 128 \times 1 \times 1 \times 3}$ であり、 $\mathbf{s}_{TB}^{(2)} = (1, 1, 1)$ である。この段階で第 4 モードのサイズが 3 になり、音高方向におよそ低域、中域、高域に対する特徴生成を開始する。

第 3 転置畳み込み層が $\mathbf{X}_{TB}^{(2)} \in \mathbb{R}^{128 \times 4 \times 1 \times 3}$ から

$$\mathbf{X}_{TB}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}^{(2)}, \mathcal{K}_{TB}^{(3)}, \mathbf{s}_{TB}^{(3)})) \in \mathbb{R}^{128 \times 4 \times 2 \times 3} \quad (1.5)$$

を生成する。ここで、 $\mathcal{K}_{TB}^{(3)} \in \mathbb{R}^{128 \times 128 \times 1 \times 2 \times 1}$ であり、 $\mathbf{s}_{TB}^{(3)} = (1, 1, 1)$ である。この段階で第 3 モードのサイズが 2 になり、二分音符単位の特徴生成を開始する。

第 4 転置畳み込み層が $\mathbf{X}_{TB}^{(3)} \in \mathbb{R}^{128 \times 4 \times 2 \times 3}$ から

$$\mathbf{X}_{TB}^{(4)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}^{(3)}, \mathcal{K}_{TB}^{(4)}, \mathbf{s}_{TB}^{(4)})) \in \mathbb{R}^{128 \times 4 \times 4 \times 3} \quad (1.6)$$

を生成する。ここで、 $\mathcal{K}_{TB}^{(4)} \in \mathbb{R}^{128 \times 128 \times 1 \times 2 \times 1}$ であり、 $\mathbf{s}_{TB}^{(4)} = (1, 2, 1)$ である。この段階で第 3 モードのサイズが 4 になり、四分音符単位の特徴生成を開始する。

第 5 転置畳み込み層が $\mathbf{X}_{TB}^{(4)} \in \mathbb{R}^{128 \times 4 \times 4 \times 3}$ から

$$\mathbf{X}_{TB}^{(5)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}^{(4)}, \mathcal{K}_{TB}^{(5)}, \mathbf{s}_{TB}^{(5)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 6} \quad (1.7)$$

を生成する。ここで、 $\mathcal{K}_{TB}^{(5)} \in \mathbb{R}^{64 \times 128 \times 1 \times 1 \times 2}$ であり、 $\mathbf{s}_{TB}^{(5)} = (1, 1, 2)$ である。この段階で第 4 モードのサイズが 6 になり、音高方向に更に 2 倍の帯域分割での特徴生成を開始する。

第 6 転置畳み込み層が $\mathbf{X}_{TB}^{(5)} \in \mathbb{R}^{64 \times 4 \times 4 \times 6}$ から

$$\mathbf{X}_{TB}^{(6)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}^{(5)}, \mathcal{K}_{TB}^{(6)}, \mathbf{s}_{TB}^{(6)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 12} \quad (1.8)$$

を生成する。ここで、 $\mathcal{K}_{TB}^{(6)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 2}$ であり、 $\mathbf{s}_{TB}^{(6)} = (1, 1, 2)$ である。この段階で第 4 モードのサイズが 12 になり、1 オクターブ 12 半音に対する特徴生成を開始する。Tonal Block は処理をこれで終了し、 $\mathbf{X}_{TB}^{(6)}$ を Tonal Block の出力 \mathbf{X}_{TB} として出力

する： $\mathbf{X}_{TB} = \mathbf{X}_{TB}^{(6)}$. これが，有音程楽器共通の特徴量であり，打楽器以外の $M - 1$ 個の Individual Block に入力される．Tonal Block のパラメータをまとめて表 1.1 に示す．

表 1.1: Tonal Block の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	スライドベクトル
第 1 層	$\mathcal{K}_{TB}^{(1)} \in \mathbb{R}^{128 \times 256 \times 2 \times 1 \times 1}$	$\mathbf{s}_{TB}^{(1)} = (1, 1, 1)$
第 2 層	$\mathcal{K}_{TB}^{(2)} \in \mathbb{R}^{128 \times 128 \times 1 \times 1 \times 3}$	$\mathbf{s}_{TB}^{(2)} = (1, 1, 1)$
第 3 層	$\mathcal{K}_{TB}^{(3)} \in \mathbb{R}^{128 \times 128 \times 1 \times 2 \times 1}$	$\mathbf{s}_{TB}^{(3)} = (1, 1, 1)$
第 4 層	$\mathcal{K}_{TB}^{(4)} \in \mathbb{R}^{128 \times 128 \times 1 \times 2 \times 1}$	$\mathbf{s}_{TB}^{(4)} = (1, 2, 1)$
第 5 層	$\mathcal{K}_{TB}^{(5)} \in \mathbb{R}^{64 \times 128 \times 1 \times 1 \times 2}$	$\mathbf{s}_{TB}^{(5)} = (1, 1, 2)$
第 6 層	$\mathcal{K}_{TB}^{(6)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 2}$	$\mathbf{s}_{TB}^{(6)} = (1, 1, 2)$

1.4 Individual Blocks

Individual Blocks はトラック別の処理を行うブロックであり，Piano Block，Guitar Block，Bass Block，Strings Block の 4 ブロックからなる．これらの 4 ブロックは，有音程楽器共通特徴量 \mathbf{X}_{TB} を入力として，各トラックのピアノロールを生成する．

1.4.1 Piano Block

従来手法である BinaryMuseGAN における Piano Block は，ピアノの音域である C1 から B7 までの 7 オクターブ全体を一括して生成する 3 個のサブブロックで構成されていた．本手法では低域 3 オクターブと高域 4 オクターブを別々のブロックで生成する．それぞれ，Piano Low Block と Piano High Block と呼ぶことにする．順に説明する．

まず Piano Low Block を，BinaryMuseGAN の Piano Block と同様に 3 個のサブブロックで構成する．それらを Piano Low Block 1，Piano Low Block 2，および Piano Low Block 3 と呼ぶことにする．Piano Low Block 1 は，表 1.2 に示す 3 層の転置畳み込み層で構成される．第 1 転置畳み込み層は $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{PiLB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{PiLB1}^{(1)}, \mathbf{s}_{PiLB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 36} \quad (1.9)$$

を生成する．ここで， $\mathcal{K}_{PiLB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$ であり， $\mathbf{s}_{PiLB1}^{(1)} = (1, 1, 3)$ である．この段階で第 4 モードのサイズが 36 になり，低域 3 オクターブ分の特徴生成を開始する．

次に、第2転置畳み込み層が $\mathbf{X}_{PiLB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 36}$ から

$$\mathbf{X}_{PiLB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiLB1}^{(1)}, \mathcal{K}_{PiLB1}^{(2)}, \mathbf{s}_{PiLB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 36} \quad (1.10)$$

を生成する．ここで、 $\mathcal{K}_{PiLB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{PiLB1}^{(2)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．最後に、第3転置畳み込み層が $\mathbf{X}_{PiLB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 36}$ から

$$\mathbf{X}_{PiLB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiLB1}^{(2)}, \mathcal{K}_{PiLB1}^{(3)}, \mathbf{s}_{PiLB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.11)$$

を生成する．ここで、 $\mathcal{K}_{PiLB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{PiLB1}^{(3)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．以上のように、Piano Low Block 1では、第4モードの特徴生成を行ってから第3モードの特徴生成を行う．すなわち、音高方向の特徴生成の後に時間方向の特徴生成を生成する．

表 1.2: Piano Low Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{PiLB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$	$\mathbf{s}_{PiLB1}^{(1)} = (1, 1, 3)$
第2層	$\mathcal{K}_{PiLB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{PiLB1}^{(2)} = (1, 4, 1)$
第3層	$\mathcal{K}_{PiLB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{PiLB1}^{(3)} = (1, 6, 1)$

一方、Piano Low Block 2 は同じサイズの特徴量テンソルを逆の手順で生成する．すなわち、第3モードの時間方向特徴量を生成してから第4モードである音高方向の特徴生成を行う．まず、第1転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{PiLB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{PiLB2}^{(1)}, \mathbf{s}_{PiLB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 12} \quad (1.12)$$

を生成する．ここで、 $\mathcal{K}_{PiLB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{PiLB2}^{(1)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．第2転置畳み込み層が $\mathbf{X}_{PiLB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 12}$ から

$$\mathbf{X}_{PiLB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiLB2}^{(1)}, \mathcal{K}_{PiLB2}^{(2)}, \mathbf{s}_{PiLB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 12} \quad (1.13)$$

を生成する．ここで、 $\mathcal{K}_{PiLB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{PiLB2}^{(2)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．第3転置畳み込み層が $\mathbf{X}_{PiLB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 12}$ から

$$\mathbf{X}_{PiLB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiLB2}^{(2)}, \mathcal{K}_{PiLB2}^{(3)}, \mathbf{s}_{PiLB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.14)$$

を生成する．ここで， $\mathcal{K}_{PiLB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$ であり， $s_{PiLB2}^{(3)} = (1, 1, 3)$ である．この段階で第4モードのサイズが36になり，低域3オクターブ分の特徴生成を開始する．Piano Low Block 2 の転置畳み込み層における各種パラメータを表 1.3 にまとめて示す．

表 1.3: Piano Low Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{PiLB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$	$s_{PiLB2}^{(1)} = (1, 4, 1)$
第2層	$\mathcal{K}_{PiLB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$s_{PiLB2}^{(2)} = (1, 6, 1)$
第3層	$\mathcal{K}_{PiLB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$	$s_{PiLB2}^{(3)} = (1, 1, 3)$

Piano Low Block 3 では，ピアノトラックの低域3オクターブ分のピアノロールを以下のように生成する．まず，入力された特徴 $\mathbf{X}_{PiLB1}^{(3)}$ ， $\mathbf{X}_{PiLB2}^{(3)}$ を

$$\mathbf{X}_{PiLB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{PiLB1}^{(3)}, \mathbf{X}_{PiLB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 36} \quad (1.15)$$

によって結合する．ここで，Concat は第1引数が示すモードに沿って第2引数以降の入力テンソルを結合する操作である． $\mathbf{X}_{PiLB1}^{(3)}$ と $\mathbf{X}_{PiLB2}^{(3)}$ の2テンソルを結合することにより，特徴量の生成順序による影響を抑制している．続いて，転置畳み込み層が $\mathbf{X}_{PiLB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 36}$ から

$$\mathbf{X}_{PiLB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{PiLB3}^{(0)}, \mathcal{K}_{PiLB3}^{(1)}, s_{PiLB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 36} \quad (1.16)$$

を生成する．ここで， $\mathcal{K}_{PiLB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり， $s_{PiLB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が1になるため階テンソルと等価になり，ピアノトラックの低域のピアノロール生成を完了する．Piano Low Block 3 の転置畳み込み層における各種パラメータを表 1.4 にまとめて示す．

表 1.4: Piano Low Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{PiLB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1 \times 1}$	$s_{PiLB3}^{(1)} = (1, 1, 1)$

ピアノトラックの高域トラックも，上記の低域と同様に生成する．まず，Piano High Block 1 の第1転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{PiHB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{PiHB1}^{(1)}, s_{PiHB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 48} \quad (1.17)$$

を生成する．ここで、 $\mathcal{K}_{PiHB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 4}$ であり、 $s_{PiHB1}^{(1)} = (1, 1, 4)$ である．この段階で第4モードのサイズが48になり、高域4オクターブ分の特徴生成を開始する．第2転置畳み込み層が $\mathbf{X}_{PiHB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 48}$ から

$$\mathbf{X}_{PiHB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiHB1}^{(1)}, \mathcal{K}_{PiHB1}^{(2)}, s_{PiHB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 48} \quad (1.18)$$

を生成する．ここで、 $\mathcal{K}_{PiHB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり、 $s_{PiHB1}^{(2)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．第3転置畳み込み層が $\mathbf{X}_{PiHB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 48}$ から

$$\mathbf{X}_{PiHB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiHB1}^{(2)}, \mathcal{K}_{PiHB1}^{(3)}, s_{PiHB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 48} \quad (1.19)$$

を生成する．ここで、 $\mathcal{K}_{PiHB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり、 $s_{PiHB1}^{(3)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．以上のように、Piano High Block 1では第4モードの特徴生成を行ってから第3モードの特徴生成を行う．Piano High Block 1の転置畳み込み層における各種パラメータを表 1.5 にまとめて示す．

表 1.5: Piano High Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	スライドベクトル
第1層	$\mathcal{K}_{PiHB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 4}$	$s_{PiHB1}^{(1)} = (1, 1, 4)$
第2層	$\mathcal{K}_{PiHB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$s_{PiHB1}^{(2)} = (1, 4, 1)$
第3層	$\mathcal{K}_{PiHB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$s_{PiHB1}^{(3)} = (1, 6, 1)$

一方、Piano High Block 2では逆の手順で特徴生成を行う．すなわち、第3モードの特徴生成を行ってから、第4モードの特徴生成を行う．Piano High Block 2の第1転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{PiHB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{PiHB2}^{(1)}, s_{PiHB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 12} \quad (1.20)$$

を生成する．ここで、 $\mathcal{K}_{PiHB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$ であり、 $s_{PiHB2}^{(1)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．第2転置畳み込み層が $\mathbf{X}_{PiHB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 12}$ から

$$\mathbf{X}_{PiHB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiHB2}^{(1)}, \mathcal{K}_{PiHB2}^{(2)}, s_{PiHB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 12} \quad (1.21)$$

を生成する．ここで、 $\mathcal{K}_{PiHB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $s_{PiHB2}^{(2)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始す

る．第 3 転置畳み込み層が $\mathbf{X}_{PiHB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 12}$ から

$$\mathbf{X}_{PiHB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PiHB2}^{(2)}, \mathcal{K}_{PiHB2}^{(3)}, \mathbf{s}_{PiHB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 48} \quad (1.22)$$

を生成する．ここで， $\mathcal{K}_{PiHB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 4}$ であり， $\mathbf{s}_{PiHB2}^{(3)} = (1, 1, 4)$ である．この段階で第 4 モードのサイズが 48 になり，高域 4 オクターブ分の特徴生成を開始する．Piano High Block 2 の転置畳み込み層における各種パラメータを表 1.6 にまとめて示す．

表 1.6: Piano High Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{PiHB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{PiHB2}^{(1)} = (1, 4, 1)$
第 2 層	$\mathcal{K}_{PiHB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{PiHB2}^{(2)} = (1, 6, 1)$
第 3 層	$\mathcal{K}_{PiHB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 4}$	$\mathbf{s}_{PiHB2}^{(3)} = (1, 1, 4)$

Piano High Block 3 では，ピアノトラックの高域 4 オクターブ分のピアノロールを以下のように生成する．まず，入力された特徴 $\mathbf{X}_{PiHB1}^{(3)}$ ， $\mathbf{X}_{PiHB2}^{(3)}$ を

$$\mathbf{X}_{PiHB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{PiHB1}^{(3)}, \mathbf{X}_{PiHB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 48} \quad (1.23)$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}_{PiHB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 48}$ から

$$\mathbf{X}_{PiHB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{PiHB3}^{(0)}, \mathcal{K}_{PiHB3}^{(1)}, \mathbf{s}_{PiHB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 48} \quad (1.24)$$

を生成する．ここで， $\mathcal{K}_{PiHB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり， $\mathbf{s}_{PiHB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が 1 になり，ピアノトラックの高域のピアノロール生成を完了する．Piano High Block 3 の転置畳み込み層における各種パラメータを表 1.7 にまとめて示す．

表 1.7: Piano High Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{PiHB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1 \times 1}$	$\mathbf{s}_{PiHB3}^{(1)} = (1, 1, 1)$

最後に低域 $\mathbf{X}_{PiLB3}^{(1)}$ と高域 $\mathbf{X}_{PiHB3}^{(1)}$ を

$$\hat{\mathbf{X}}_{Pi} = \text{Concat}(3, \mathbf{X}_{PiLB3}^{(1)}, \mathbf{X}_{PiHB3}^{(1)}) \in \mathbb{R}^{4 \times 96 \times 84} \quad (1.25)$$

によって結合し，ピアノトラック全体のピアノロール生成を完了する．

1.4.2 Guitar Blocks

従来手法である BinaryMuseGAN における Guitar Block は、C1 から B7 までの 7 オクターブの音域を一括して生成する 3 個のサブブロックで構成されていた。しかし、7 オクターブは実際のギターの演奏範囲を大きく逸脱している。本手法では、ギターの演奏範囲である E2 から D \sharp 6 までの 4 オクターブに限定して生成を行う。Piano High Block と同様に、Guitar Block を 3 個のサブブロックで構成する。それらを Guitar Block 1, Guitar Block 2, および Guitar Block 3 と呼ぶことにする。まず、Guitar Block 1 の第 1 転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{GB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{GB1}^{(1)}, \mathbf{s}_{GB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 48} \quad (1.26)$$

を生成する。ここで、 $\mathcal{K}_{GB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 4}$ であり、 $\mathbf{s}_{GB1}^{(1)} = (1, 1, 4)$ である。この段階で第 4 モードのサイズが 48 になり、4 オクターブ分の特徴生成を開始する。第 2 転置畳み込み層が $\mathbf{X}_{GB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 48}$ から

$$\mathbf{X}_{GB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{GB1}^{(1)}, \mathcal{K}_{GB1}^{(2)}, \mathbf{s}_{GB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 48} \quad (1.27)$$

を生成する。ここで、 $\mathcal{K}_{GB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{GB1}^{(2)} = (1, 4, 1)$ である。この段階で第 3 モードのサイズが 16 になり、16 分音符単位の特徴生成を開始する。第 3 転置畳み込み層が $\mathbf{X}_{GB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 48}$ から

$$\mathbf{X}_{GB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{GB1}^{(2)}, \mathcal{K}_{GB1}^{(3)}, \mathbf{s}_{GB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 48} \quad (1.28)$$

を生成する。ここで、 $\mathcal{K}_{GB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{GB1}^{(3)} = (1, 6, 1)$ である。この段階で第 3 モードのサイズが 96 になり、32 分音符の 3 連符単位の特徴生成を開始する。以上のように、Guitar Block 1 では第 4 モードの特徴生成を行ってから第 3 モードの特徴生成を行う。Guitar Block 1 の転置畳み込み層における各種パラメータを表 1.8 にまとめて示す。

表 1.8: Guitar Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{GB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 4}$	$\mathbf{s}_{GB1}^{(1)} = (1, 1, 4)$
第 2 層	$\mathcal{K}_{GB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{GB1}^{(2)} = (1, 4, 1)$
第 3 層	$\mathcal{K}_{GB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{GB1}^{(3)} = (1, 6, 1)$

一方、Guitar Block 2 では逆の手順で特徴生成を行う。すなわち、第 3 モードの特徴生成を行ってから、第 4 モードの特徴生成を行う。Guitar Block 2 の第 1 転置畳み

込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{GB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{GB2}^{(1)}, \mathbf{s}_{GB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 12} \quad (1.29)$$

を生成する．ここで、 $\mathcal{K}_{GB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{GB2}^{(1)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．第2転置畳み込み層が $\mathbf{X}_{GB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 12}$ から

$$\mathbf{X}_{GB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{GB2}^{(1)}, \mathcal{K}_{GB2}^{(2)}, \mathbf{s}_{GB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 12} \quad (1.30)$$

を生成する．ここで、 $\mathcal{K}_{GB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{GB2}^{(2)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．第3転置畳み込み層が $\mathbf{X}_{GB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 12}$ から

$$\mathbf{X}_{GB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{GB2}^{(2)}, \mathcal{K}_{GB2}^{(3)}, \mathbf{s}_{GB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 48} \quad (1.31)$$

を生成する．ここで、 $\mathcal{K}_{GB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 4}$ であり、 $\mathbf{s}_{GB2}^{(3)} = (1, 1, 4)$ である．この段階で第4モードのサイズが48になり、4オクターブ分の特徴生成を開始する．Guitar Block 2の転置畳み込み層における各種パラメータを表 1.9 にまとめて示す．

表 1.9: Guitar Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{GB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{GB2}^{(1)} = (1, 4, 1)$
第2層	$\mathcal{K}_{GB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{GB2}^{(2)} = (1, 6, 1)$
第3層	$\mathcal{K}_{GB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 4}$	$\mathbf{s}_{GB2}^{(3)} = (1, 1, 4)$

Guitar Block 3では、ギタートラックの4オクターブ分のピアノロールを以下のよう生成する．まず、入力された特徴 $\mathbf{X}_{GB1}^{(3)}$, $\mathbf{X}_{GB2}^{(3)}$ を

$$\mathbf{X}_{GB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{GB1}^{(3)}, \mathbf{X}_{GB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 48} \quad (1.32)$$

によって結合する．続いて、転置畳み込み層が $\mathbf{X}_{GB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 48}$ から

$$\mathbf{X}_{GB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{GB3}^{(0)}, \mathcal{K}_{GB3}^{(1)}, \mathbf{s}_{GB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 48} \quad (1.33)$$

を生成する．ここで、 $\mathcal{K}_{GB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり、 $\mathbf{s}_{GB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が1になり、ギタートラックのピアノロール生成を完了する．Guitar Block 3の転置畳み込み層における各種パラメータを表 1.10 にまとめて示す．

表 1.10: Guitar Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{GB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1 \times 1}$	$\mathbf{s}_{GB3}^{(1)} = (1, 1, 1)$

最後にピアノロールのサイズをピアノのものとそろえるために、 $\mathbf{X}_{GB3}^{(1)} \in \mathbb{R}^{4 \times 96 \times 48}$ の第 3 モードの低域側の端に 16 個，高域側の端に 20 個の 0 を挿入する．この変換を関数 $Pad_G(\mathbf{X}_{GB3}^{(1)})$ で表し，出力される 3 階テンソルを $\hat{\mathbf{X}}_{Gu}$ で表す：

$$\hat{\mathbf{X}}_{Gu} = Pad_G(\mathbf{X}_{GB3}^{(1)}) \in \mathbb{R}^{4 \times 96 \times 84} \quad (1.34)$$

Guitar Block は処理をこれで終了し， $\hat{\mathbf{X}}_{Gu}$ を Guitar Block の出力として出力する．

1.4.3 Base Blocks

ベースにおいても，従来手法である BinaryMuseGAN は C1 から B7 までの 7 オクターブの音域を生成していた．しかし，これは実際のベースの演奏範囲を大きく逸脱している．本研究では，C1 から B3 までの 3 オクターブに限定して生成を行う．Piano Low Block と同様に，Bass Block を 3 個のサブブロック Bass Block 1, Bass Block 2, および Bass Block 3 で構成する．まず，Bass Block 1 の第 1 転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{BaB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{BaB1}^{(1)}, \mathbf{s}_{BaB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 36} \quad (1.35)$$

を生成する．ここで， $\mathcal{K}_{BaB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$ であり， $\mathbf{s}_{BaB1}^{(1)} = (1, 1, 3)$ である．この段階で第 4 モードのサイズが 36 になり，3 オクターブ分の特徴生成を開始する．第 2 転置畳み込み層が $\mathbf{X}_{BaB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 36}$ から

$$\mathbf{X}_{BaB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{BaB1}^{(1)}, \mathcal{K}_{BaB1}^{(2)}, \mathbf{s}_{BaB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 36} \quad (1.36)$$

を生成する．ここで， $\mathcal{K}_{BaB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり， $\mathbf{s}_{BaB1}^{(2)} = (1, 4, 1)$ である．この段階で第 3 モードのサイズが 16 になり，16 分音符単位の特徴生成を開始する．第 3 転置畳み込み層が $\mathbf{X}_{BaB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 36}$ から

$$\mathbf{X}_{BaB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{BaB1}^{(2)}, \mathcal{K}_{BaB1}^{(3)}, \mathbf{s}_{BaB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.37)$$

を生成する．ここで， $\mathcal{K}_{BaB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり， $\mathbf{s}_{BaB1}^{(3)} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 96 になり，32 分音符の 3 連符単位の特徴生成を開始する．

以上のように、Bass Block 1 では第 4 モードの特徴生成を行ってから第 3 モードの特徴生成を行う。Bass Block 1 の転置畳み込み層における各種パラメータを表 1.11 にまとめて示す。

表 1.11: Bass Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{BaB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$	$\mathbf{s}_{BaB1}^{(1)} = (1, 1, 3)$
第 2 層	$\mathcal{K}_{BaB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{BaB1}^{(2)} = (1, 4, 1)$
第 3 層	$\mathcal{K}_{BaB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{BaB1}^{(3)} = (1, 6, 1)$

一方、Bass Block 2 では逆の手順で特徴生成を行う。すなわち、第 3 モードの特徴生成を行ってから、第 4 モードの特徴生成を行う。Bass Block 2 の第 1 転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{BaB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{BaB2}^{(1)}, \mathbf{s}_{BaB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 12} \quad (1.38)$$

を生成する。ここで、 $\mathcal{K}_{BaB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{BaB2}^{(1)} = (1, 4, 1)$ である。この段階で第 3 モードのサイズが 16 になり、16 分音符単位の特徴生成を開始する。第 2 転置畳み込み層が $\mathbf{X}_{BaB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 12}$ から

$$\mathbf{X}_{BaB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{BaB2}^{(1)}, \mathcal{K}_{BaB2}^{(2)}, \mathbf{s}_{BaB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 12} \quad (1.39)$$

を生成する。ここで、 $\mathcal{K}_{BaB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{BaB2}^{(2)} = (1, 6, 1)$ である。この段階で第 3 モードのサイズが 96 になり、32 分音符の 3 連符単位の特徴生成を開始する。第 3 転置畳み込み層が $\mathbf{X}_{BaB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 12}$ から

$$\mathbf{X}_{BaB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{BaB2}^{(2)}, \mathcal{K}_{BaB2}^{(3)}, \mathbf{s}_{BaB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.40)$$

を生成する。ここで、 $\mathcal{K}_{BaB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$ であり、 $\mathbf{s}_{BaB2}^{(3)} = (1, 1, 3)$ である。この段階で第 4 モードのサイズが 36 になり、3 オクターブ分の特徴生成を開始する。Bass Block 2 の転置畳み込み層における各種パラメータを表 1.12 にまとめて示す。

表 1.12: Bass Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{BaB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{BaB2}^{(1)} = (1, 4, 1)$
第 2 層	$\mathcal{K}_{BaB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{BaB2}^{(2)} = (1, 6, 1)$
第 3 層	$\mathcal{K}_{BaB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$	$\mathbf{s}_{BaB2}^{(3)} = (1, 1, 3)$

Bass Block 3 では、ベーストラックの 4 オクターブ分のピアノロールを以下のよう
に生成する．まず，入力された特徴 $\mathbf{X}_{BaB1}^{(3)}$, $\mathbf{X}_{BaB2}^{(3)}$ を

$$\mathbf{X}_{BaB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{BaB1}^{(3)}, \mathbf{X}_{BaB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 36} \quad (1.41)$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}_{BaB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 36}$ から

$$\mathbf{X}_{BaB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{BaB3}^{(0)}, \mathcal{K}_{BaB3}^{(1)}, \mathbf{s}_{BaB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 36} \quad (1.42)$$

を生成する．ここで， $\mathcal{K}_{BaB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり， $\mathbf{s}_{BaB3}^{(1)} = (1, 1, 1)$ である．この段階
でチャンネル数が 1 になり，ベーストラックのピアノロール生成を完了する．Bass
Block 3 の転置畳み込み層における各種パラメータを表 1.13 にまとめて示す．

表 1.13: Bass Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{BaB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1 \times 1}$	$\mathbf{s}_{BaB3}^{(1)} = (1, 1, 1)$

最後にピアノロールのサイズをピアノのものとそろえるために， $\mathbf{X}_{BaB3}^{(1)} \in \mathbb{R}^{4 \times 96 \times 36}$
の第 3 モードの高域側の端に 48 個の 0 を挿入する．この変換を関数 $\text{Pad}_{Ba}(\mathbf{X}_{BaB3}^{(1)})$
で表し，出力される 3 階テンソルを $\hat{\mathbf{X}}_{Ba}$ で表す：

$$\hat{\mathbf{X}}_{Ba} = \text{Pad}_{Ba}(\mathbf{X}_{BaB3}^{(1)}) \in \mathbb{R}^{4 \times 96 \times 84} \quad (1.43)$$

Bass Block は処理をこれで終了し， $\hat{\mathbf{X}}_{Ba}$ を Bass Block の出力として出力する．

1.4.4 String Blocks

ストリングスにおいても，従来手法である BinaryMuseGAN は C1 から B7 までの
7 オクターブの音域を生成していた．しかし，これも実際のストリングスの音域を少
し逸脱しているので，本研究では C#1 から C7 までの 6 オクターブに限定する．さら
に，これを低域 3 オクターブと高域 3 オクターブに分割し，それぞれを Strings Low
Block と Strings High Block で生成する．これらのブロックは Piano Low Block と同
様に 3 個のサブブロックで構成する．順に説明する．

Strings Low Block のサブブロックをそれぞれ Strings Low Block 1, Strings Low
Block 2, および Strings Low Block 3 と呼ぶことにする．Strings Low Block 1 は，
表 1.14 に示す 3 層の転置畳み込み層で構成される．第 1 転置畳み込み層は $\mathbf{X}_{TB} \in$
 $\mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{SLB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{SLB1}^{(1)}, \mathbf{s}_{SLB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 36} \quad (1.44)$$

を生成する．ここで、 $\mathcal{K}_{SLB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$ であり、 $\mathbf{s}_{SLB1}^{(1)} = (1, 1, 3)$ である．この段階で第4モードのサイズが36になり、低域3オクターブ分の特徴生成を開始する．次に、第2転置畳み込み層が $\mathbf{X}_{SLB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 36}$ から

$$\mathbf{X}_{SLB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SLB1}^{(1)}, \mathcal{K}_{SLB1}^{(2)}, \mathbf{s}_{SLB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 36} \quad (1.45)$$

を生成する．ここで、 $\mathcal{K}_{SLB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{SLB1}^{(2)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．最後に、第3転置畳み込み層が $\mathbf{X}_{SLB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 36}$ から

$$\mathbf{X}_{SLB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SLB1}^{(2)}, \mathcal{K}_{SLB1}^{(3)}, \mathbf{s}_{SLB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.46)$$

を生成する．ここで、 $\mathcal{K}_{SLB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{SLB1}^{(3)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．以上のように、Strings Low Block 1 では、第4モードの特徴生成を行ってから第3モードの特徴生成を行う．

表 1.14: Strings Low Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{SLB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$	$\mathbf{s}_{SLB1}^{(1)} = (1, 1, 3)$
第2層	$\mathcal{K}_{SLB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{SLB1}^{(2)} = (1, 4, 1)$
第3層	$\mathcal{K}_{SLB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{SLB1}^{(3)} = (1, 6, 1)$

一方、Strings Low Block 2 は同じサイズの特徴量テンソルを逆の手順で生成する．すなわち、第3モードの時間方向特徴量を生成してから第4モードである音高方向の特徴生成を行う．まず、第1転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{SLB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{SLB2}^{(1)}, \mathbf{s}_{SLB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 12} \quad (1.47)$$

を生成する．ここで、 $\mathcal{K}_{SLB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{SLB2}^{(1)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．第2転置畳み込み層が $\mathbf{X}_{SLB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 12}$ から

$$\mathbf{X}_{SLB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SLB2}^{(1)}, \mathcal{K}_{SLB2}^{(2)}, \mathbf{s}_{SLB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 12} \quad (1.48)$$

を生成する．ここで、 $\mathcal{K}_{SLB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{SLB2}^{(2)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．第3転置畳み込み層が $\mathbf{X}_{SLB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 12}$ から

$$\mathbf{X}_{SLB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SLB2}^{(2)}, \mathcal{K}_{SLB2}^{(3)}, \mathbf{s}_{SLB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.49)$$

を生成する．ここで， $\mathcal{K}_{SLB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$ であり， $\mathbf{s}_{SLB2}^{(3)} = (1, 1, 3)$ である．この段階で第4モードのサイズが36になり，低域3オクターブ分の特徴生成を開始する．Strings Low Block 2の転置畳み込み層における各種パラメータを表 1.15 にまとめて示す．

表 1.15: Strings Low Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{SLB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{SLB2}^{(1)} = (1, 4, 1)$
第2層	$\mathcal{K}_{SLB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{SLB2}^{(2)} = (1, 6, 1)$
第3層	$\mathcal{K}_{SLB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$	$\mathbf{s}_{SLB2}^{(3)} = (1, 1, 3)$

Strings Low Block 3では，ストリングストラックの低域3オクターブ分のピアノロールを以下のように生成する．まず，入力された特徴 $\mathbf{X}_{SLB1}^{(3)}$ ， $\mathbf{X}_{SLB2}^{(3)}$ を

$$\mathbf{X}_{SLB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{SLB1}^{(3)}, \mathbf{X}_{SLB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 36} \quad (1.50)$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}_{SLB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 36}$ から

$$\mathbf{X}_{SLB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{SLB3}^{(0)}, \mathcal{K}_{SLB3}^{(1)}, \mathbf{s}_{SLB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 36} \quad (1.51)$$

を生成する．ここで， $\mathcal{K}_{SLB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり， $\mathbf{s}_{SLB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が1になり，ストリングストラックの低域のピアノロール生成を完了する．Strings Low Block 3の転置畳み込み層における各種パラメータを表 1.16 にまとめて示す．

表 1.16: Strings Low Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{SLB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1 \times 1}$	$\mathbf{s}_{SLB3}^{(1)} = (1, 1, 1)$

ストリングストラックの高域トラックも，上記の低域と同様に生成する．まず，Strings High Block 1の第1転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{SHB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{SHB1}^{(1)}, \mathbf{s}_{SHB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 36} \quad (1.52)$$

を生成する．ここで， $\mathcal{K}_{SHB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 3}$ であり， $\mathbf{s}_{SHB1}^{(1)} = (1, 1, 3)$ である．この段階で第4モードのサイズが36になり，高域3オクターブ分の特徴生成を開始する．

第2 転置畳み込み層が $\mathbf{X}_{SHB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 36}$ から

$$\mathbf{X}_{SHB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SHB1}^{(1)}, \mathcal{K}_{SHB1}^{(2)}, \mathbf{s}_{SHB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 36} \quad (1.53)$$

を生成する．ここで， $\mathcal{K}_{SHB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり， $\mathbf{s}_{SHB1}^{(2)} = (1, 4, 1)$ である．この段階で第3 モードのサイズが16 になり，16 分音符単位の特徴生成を開始する．第3 転置畳み込み層が $\mathbf{X}_{SHB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 36}$ から

$$\mathbf{X}_{SHB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SHB1}^{(2)}, \mathcal{K}_{SHB1}^{(3)}, \mathbf{s}_{SHB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.54)$$

を生成する．ここで， $\mathcal{K}_{SHB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり， $\mathbf{s}_{SHB1}^{(3)} = (1, 6, 1)$ である．この段階で第3 モードのサイズが96 になり，32 分音符の3 連符単位の特徴生成を開始する．以上のように，Strings High Block 1 では第4 モードの特徴生成を行ってから第3 モードの特徴生成を行う．Strings High Block 1 の転置畳み込み層における各種パラメータを表 1.17 にまとめて示す．

表 1.17: Strings High Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1 層	$\mathcal{K}_{SHB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 4}$	$\mathbf{s}_{SHB1}^{(1)} = (1, 1, 4)$
第2 層	$\mathcal{K}_{SHB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{SHB1}^{(2)} = (1, 4, 1)$
第3 層	$\mathcal{K}_{SHB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{SHB1}^{(3)} = (1, 6, 1)$

一方，Strings High Block 2 では逆の手順で特徴生成を行う．すなわち，第3 モードの特徴生成を行ってから，第4 モードの特徴生成を行う．Strings High Block 2 の第1 転置畳み込み層が $\mathbf{X}_{TB} \in \mathbb{R}^{64 \times 4 \times 4 \times 12}$ から

$$\mathbf{X}_{SHB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{SHB2}^{(1)}, \mathbf{s}_{SHB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 12} \quad (1.55)$$

を生成する．ここで， $\mathcal{K}_{SHB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$ であり， $\mathbf{s}_{SHB2}^{(1)} = (1, 4, 1)$ である．この段階で第3 モードのサイズが16 になり，16 分音符単位の特徴生成を開始する．第2 転置畳み込み層が $\mathbf{X}_{SHB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 12}$ から

$$\mathbf{X}_{SHB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SHB2}^{(1)}, \mathcal{K}_{SHB2}^{(2)}, \mathbf{s}_{SHB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 12} \quad (1.56)$$

を生成する．ここで， $\mathcal{K}_{SHB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり， $\mathbf{s}_{SHB2}^{(2)} = (1, 6, 1)$ である．この段階で第3 モードのサイズが96 になり，32 分音符の3 連符単位の特徴生成を開始する．第3 転置畳み込み層が $\mathbf{X}_{SHB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 12}$ から

$$\mathbf{X}_{SHB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{SHB2}^{(2)}, \mathcal{K}_{SHB2}^{(3)}, \mathbf{s}_{SHB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 36} \quad (1.57)$$

を生成する．ここで， $\mathcal{K}_{SHB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$ であり， $\mathbf{s}_{SHB2}^{(3)} = (1, 1, 3)$ である．この段階で第4モードのサイズが36になり，高域3オクターブ分の特徴生成を開始する．Strings High Block 2の転置畳み込み層における各種パラメータを表 1.18 にまとめて示す．

表 1.18: Strings High Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{SHB2}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{SHB2}^{(1)} = (1, 4, 1)$
第2層	$\mathcal{K}_{SHB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{SHB2}^{(2)} = (1, 6, 1)$
第3層	$\mathcal{K}_{SHB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 3}$	$\mathbf{s}_{SHB2}^{(3)} = (1, 1, 3)$

Strings High Block 3では，ストリングストラックの高域3オクターブ分のピアノロールを以下のように生成する．まず，入力された特徴 $\mathbf{X}_{SHB1}^{(3)}$ ， $\mathbf{X}_{SHB2}^{(3)}$ を

$$\mathbf{X}_{SHB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{SHB1}^{(3)}, \mathbf{X}_{SHB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 36} \quad (1.58)$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}_{SHB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 36}$ から

$$\mathbf{X}_{SHB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{SHB3}^{(0)}, \mathcal{K}_{SHB3}^{(1)}, \mathbf{s}_{SHB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 36} \quad (1.59)$$

を生成する．ここで， $\mathcal{K}_{SHB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり， $\mathbf{s}_{SHB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が1になり，ストリングストラックの高域のピアノロール生成を完了する．Strings High Block 3の転置畳み込み層における各種パラメータを表 1.19 にまとめて示す．

表 1.19: Strings High Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{SHB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1 \times 1}$	$\mathbf{s}_{SHB3}^{(1)} = (1, 1, 1)$

以上の操作で得られた低域 $\mathbf{X}_{SLB3}^{(1)} \in \mathbb{R}^{4 \times 96 \times 36}$ と高域 $\mathbf{X}_{SHB3}^{(1)} \in \mathbb{R}^{4 \times 96 \times 36}$ を

$$\mathbf{X}_{SB} = \text{Concat}(3, \mathbf{X}_{SLB3}^{(1)}, \mathbf{X}_{SHB3}^{(1)}) \in \mathbb{R}^{4 \times 96 \times 72} \quad (1.60)$$

によって結合し，ストリングストラックのピアノロール生成を完了する．最後にピアノロールのサイズをピアノのものとそろえるために， $\mathbf{X}_{SB3}^{(1)} \in \mathbb{R}^{4 \times 96 \times 72}$ の第3モードの

低域側の端に 1 個，高域側の端に 11 個の 0 を挿入する．この変換を関数 $Pad_S(\mathbf{X}_{SB3}^{(1)})$ で表し，出力される 3 階テンソルを $\hat{\mathbf{X}}_{S_t}$ で表す：

$$\hat{\mathbf{X}}_{S_t} = Pad_S(\mathbf{X}_{SB3}^{(1)}) \in \mathbb{R}^{4 \times 96 \times 84} \quad (1.61)$$

Strings Block は処理をこれで終了し， $\hat{\mathbf{X}}_{S_t}$ を Strings Block の出力として出力する．

1.5 Percussion Block

Percussion Block は，全トラック共通特徴量 $\mathbf{x}_{BB} \in \mathbb{R}^{128}$ から打楽器共通の特徴量を以下のように生成する．まず，Base Block から入力された特徴ベクトル \mathbf{x}_{BB} から，

$$\mathbf{X}_{PB}^{(0)} = \text{Reshape}_{PB}(W_{PB}^{(0)}\mathbf{x}_{BB} + \mathbf{b}_{PB}) \in \mathbb{R}^{128 \times 3 \times 1 \times 1} \quad (1.62)$$

を生成する．ここで， $W_{PB}^{(0)} \in \mathbb{R}^{384 \times 128}$ は矩形行列であり， $\mathbf{b}_{PB} \in \mathbb{R}^{384}$ はバイアスペクトルである．また， Reshape_{PB} は 384 次元ベクトルを $128 \times 3 \times 1 \times 1$ の 4 階テンソルに並び替える関数である．

続いて，3 段の転置畳み込み層によって処理を継続する．まず，第 1 転置畳み込み層が $\mathbf{X}_{PB}^{(0)} \in \mathbb{R}^{128 \times 3 \times 1 \times 1}$ から

$$\mathbf{X}_{PB}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PB}^{(0)}, \mathcal{K}_{PB}^{(1)}, \mathbf{s}_{PB}^{(1)})) \in \mathbb{R}^{128 \times 4 \times 1 \times 1} \quad (1.63)$$

を生成する．ここで， $\mathcal{K}_{PB}^{(1)} \in \mathbb{R}^{128 \times 128 \times 2 \times 1 \times 1}$ であり， $\mathbf{s}_{PB}^{(1)} = (1, 1, 1)$ である．この段階で第 2 モードのサイズが 4 になり，4 小節分の特徴生成を開始する．

第 2 転置畳み込み層は， $\mathbf{X}_{PB}^{(1)} \in \mathbb{R}^{128 \times 4 \times 1 \times 1}$ から

$$\mathbf{X}_{PB}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PB}^{(1)}, \mathcal{K}_{PB}^{(2)}, \mathbf{s}_{PB}^{(2)})) \in \mathbb{R}^{128 \times 4 \times 4 \times 1} \quad (1.64)$$

を生成する．ここで， $\mathcal{K}_{PB}^{(2)} \in \mathbb{R}^{128 \times 128 \times 1 \times 4 \times 1}$ であり， $\mathbf{s}_{PB}^{(2)} = (1, 1, 1)$ である．この段階で第 3 モードのサイズが 4 になり，四分音符単位の特徴生成を開始する．

第 3 転置畳み込み層が $\mathbf{X}_{PB}^{(2)} \in \mathbb{R}^{128 \times 4 \times 4 \times 1}$ から

$$\mathbf{X}_{PB}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{PB}^{(2)}, \mathcal{K}_{PB}^{(3)}, \mathbf{s}_{PB}^{(3)})) \in \mathbb{R}^{128 \times 4 \times 4 \times 2} \quad (1.65)$$

を生成する．ここで， $\mathcal{K}_{PB}^{(3)} \in \mathbb{R}^{128 \times 128 \times 1 \times 1 \times 2}$ であり， $\mathbf{s}_{PB}^{(3)} = (1, 1, 1)$ である．この段階で第 4 モードのサイズが 2 になり，ドラム打楽器とその他の打楽器の元となる特徴生成を開始する．Percussion Block は処理をこれで終了し， $\mathbf{X}_{PB}^{(3)}$ を Percussion Block の出力 \mathbf{X}_{PB} として出力する： $\mathbf{X}_{PB} = \mathbf{X}_{PB}^{(3)}$ ．これが，打楽器共通の特徴量であり，Drum

Block と Other Percussion Block に入力される．Percussion Block の転置畳み込み層における各種パラメータをまとめて表 1.20 に示す．

表 1.20: Percussion Block の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{PB}^{(1)} \in \mathbb{R}^{128 \times 128 \times 2 \times 1 \times 1}$	$\mathbf{s}_{PB}^{(1)} = (1, 1, 1)$
第 2 層	$\mathcal{K}_{PB}^{(2)} \in \mathbb{R}^{128 \times 128 \times 1 \times 4 \times 1}$	$\mathbf{s}_{PB}^{(2)} = (1, 1, 1)$
第 3 層	$\mathcal{K}_{PB}^{(3)} \in \mathbb{R}^{128 \times 128 \times 1 \times 1 \times 2}$	$\mathbf{s}_{PB}^{(3)} = (1, 1, 1)$

1.6 Drum Blocks

Drums Block は，打楽器の中でも一般的なドラムセットに含まれる打楽器を生成する．この打楽器群をドラム打楽器と呼ぶことにする．本研究では，図 1.21 のように 15 の音色をドラム打楽器として定める．

表 1.21: ピアノロールを生成する打楽器

	Drum Block	Other Percussion Block
1	バスドラム	ハンドクラップ
2	サイドスティック	タンバリン
2	スネアドラム	ビブラスラップ
4	ロータム	ボンゴ
5	ミドルタム	コンガ
6	ハイトム	ティンバール
7	ハイハットペダル	アゴゴ
8	ハイハットクローズ	カバサ
9	ハイハットオープン	マラカス
10	クラッシュシンバル	ホイッスル
11	ライドシンバル	ギロ
12	チャイニーズシンバル	クラベス
13	ライドベル	ウッドブロック
14	スプラッシュシンバル	クイカ
15	カウベル	トライアングル

Drums Block では、まず、Percussion Block から入力された特徴 \mathbf{X}_{PB} から、

$$\mathbf{X}_{DB}^{(0)} = \text{Slice}(\mathbf{X}_{PB}, 4, 1, 1) \in \mathbb{R}^{128 \times 4 \times 4 \times 1} \quad (1.66)$$

を抽出する。ここで、 $\text{Slice}(\mathbf{X}, m, n, l)$ はテンソル \mathbf{X} の第 m モードの第 n 成分から第 l 成分までで構成されるテンソルを抽出する関数である。続いて、有音程楽器と同様に 3 個のサブブロック Drums Block 1, Drums Block 2, および Drums Block 3 を用いて処理を継続する。まず、Drums Block 1 の第 1 転置畳み込み層が $\mathbf{X}_{DB}^{(0)} \in \mathbb{R}^{128 \times 4 \times 4 \times 1}$ から

$$\mathbf{X}_{DB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{DB}^{(0)}, \mathcal{K}_{DB1}^{(1)}, \mathbf{s}_{DB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 15} \quad (1.67)$$

を生成する。ここで、 $\mathcal{K}_{DB1}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 1 \times 15}$ であり、 $\mathbf{s}_{DB1}^{(1)} = (1, 1, 1)$ である。この段階で第 4 モードのサイズが 15 になり、各音色の特徴生成を開始する。第 2 転置畳み込み層が $\mathbf{X}_{DB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 15}$ から

$$\mathbf{X}_{DB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{DB1}^{(1)}, \mathcal{K}_{DB1}^{(2)}, \mathbf{s}_{DB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 15} \quad (1.68)$$

を生成する。ここで、 $\mathcal{K}_{DB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{DB1}^{(2)} = (1, 4, 1)$ である。この段階で第 3 モードのサイズが 16 になり、16 分音符単位の特徴生成を開始する。第 3 転置畳み込み層が $\mathbf{X}_{DB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 15}$ から

$$\mathbf{X}_{DB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{DB1}^{(2)}, \mathcal{K}_{DB1}^{(3)}, \mathbf{s}_{DB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 15} \quad (1.69)$$

を生成する。ここで、 $\mathcal{K}_{DB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{DB1}^{(3)} = (1, 6, 1)$ である。この段階で第 3 モードのサイズが 96 になり、32 分音符の 3 連符単位の特徴生成を開始する。以上のように、Drums Block 1 では第 4 モードの特徴生成を行ってから第 3 モードの特徴生成を行う。Drums Block 1 の転置畳み込み層における各種パラメータをまとめて表 1.22 に示す。

表 1.22: Drums Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{DB1}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 1 \times 15}$	$\mathbf{s}_{DB1}^{(1)} = (1, 1, 1)$
第 2 層	$\mathcal{K}_{DB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{DB1}^{(2)} = (1, 4, 1)$
第 3 層	$\mathcal{K}_{DB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{DB1}^{(3)} = (1, 6, 1)$

一方、Drums Block 2 では逆の手順で特徴生成を行う。すなわち、第 3 モードの特徴生成を行ってから、第 4 モードの特徴生成を行う。Drums Block 2 の第 1 転置畳み込み層が $\mathbf{X}_{DB}^{(0)} \in \mathbb{R}^{128 \times 4 \times 4 \times 1}$ から

$$\mathbf{X}_{DB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{DB}^{(0)}, \mathcal{K}_{DB2}^{(1)}, \mathbf{s}_{DB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 1} \quad (1.70)$$

を生成する．ここで、 $\mathcal{K}_{DB2}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{DB2}^{(1)} = (1, 4, 1)$ である．この段階で第3モードのサイズが16になり、16分音符単位の特徴生成を開始する．第2転置畳み込み層が $\mathbf{X}_{DB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 1}$ から

$$\mathbf{X}_{DB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{DB2}^{(1)}, \mathcal{K}_{DB2}^{(2)}, \mathbf{s}_{DB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 1} \quad (1.71)$$

を生成する．ここで、 $\mathcal{K}_{DB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{DB2}^{(2)} = (1, 6, 1)$ である．この段階で第3モードのサイズが96になり、32分音符の3連符単位の特徴生成を開始する．第3転置畳み込み層が $\mathbf{X}_{DB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 1}$ から

$$\mathbf{X}_{DB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{DB2}^{(2)}, \mathcal{K}_{DB2}^{(3)}, \mathbf{s}_{DB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 15} \quad (1.72)$$

を生成する．ここで、 $\mathcal{K}_{DB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 15}$ であり、 $\mathbf{s}_{DB2}^{(3)} = (1, 1, 1)$ である．この段階で第4モードのサイズが15になり、各音色の特徴生成を開始する．Drums Block 2の転置畳み込み層における各種パラメータをまとめて表 1.23 に示す．

表 1.23: Drums Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{DB2}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 4 \times 1}$	$\mathbf{s}_{DB2}^{(1)} = (1, 4, 1)$
第2層	$\mathcal{K}_{DB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{DB2}^{(2)} = (1, 6, 1)$
第3層	$\mathcal{K}_{DB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 15}$	$\mathbf{s}_{DB2}^{(3)} = (1, 1, 1)$

Drums Block 3 では、打楽器トラックに含まれるドラム打楽器のピアノロールを以下のように生成する．まず、入力された特徴 $\mathbf{X}_{DB1}^{(3)}$ 、 $\mathbf{X}_{DB2}^{(3)}$ を

$$\mathbf{X}_{DB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{DB1}^{(3)}, \mathbf{X}_{DB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 15} \quad (1.73)$$

によって結合する．続いて、転置畳み込み層が $\mathbf{X}_{DB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 15}$ から

$$\mathbf{X}_{DB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{DB3}^{(0)}, \mathcal{K}_{DB3}^{(1)}, \mathbf{s}_{DB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 15} \quad (1.74)$$

を生成する．ここで、 $\mathcal{K}_{DB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり、 $\mathbf{s}_{DB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が1になり、ドラム打楽器のピアノロールの生成を完了する．Drums Block 3の転置畳み込み層における各種パラメータをまとめて表 1.24 に示す．

表 1.24: Drums Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第1層	$\mathcal{K}_{DB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$	$\mathbf{s}_{DB3}^{(1)} = (1, 1, 1)$

1.7 Other Percussion Blocks

Other Percussion Block は、表 1.21 に示したドラム打楽器以外の 15 の打楽器を生成する。まず、Percussion Block から入力された特徴 \mathbf{X}_{PB} から、

$$\mathbf{X}_{OPB}^{(0)} = \text{Slice}(\mathbf{X}_{PB}, 4, 2, 2) \in \mathbb{R}^{128 \times 4 \times 4 \times 1} \quad (1.75)$$

を抽出する。続いて、Drums Block と同様に、3 個のサブブロック Other Percussion Block 1, Other Percussion Block 2, および Other Percussion Block 3 を用いて処理を継続する。まず、Other Percussion Block 1 の第 1 転置畳み込み層が $\mathbf{X}_{OPB1}^{(0)} \in \mathbb{R}^{128 \times 4 \times 4 \times 1}$ から

$$\mathbf{X}_{OPB1}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{OPB}^{(0)}, \mathcal{K}_{OPB1}^{(1)}, \mathbf{s}_{OPB1}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 4 \times 15} \quad (1.76)$$

を生成する。ここで、 $\mathcal{K}_{OPB1}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 1 \times 15}$ であり、 $\mathbf{s}_{OPB1}^{(1)} = (1, 1, 1)$ である。この段階で第 4 モードのサイズが 15 になり、各音色の特徴生成を開始する。第 2 転置畳み込み層が $\mathbf{X}_{OPB1}^{(1)} \in \mathbb{R}^{64 \times 4 \times 4 \times 15}$ から

$$\mathbf{X}_{OPB1}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{OPB1}^{(1)}, \mathcal{K}_{OPB1}^{(2)}, \mathbf{s}_{OPB1}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 16 \times 15} \quad (1.77)$$

を生成する。ここで、 $\mathcal{K}_{OPB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{OPB1}^{(2)} = (1, 4, 1)$ である。この段階で第 3 モードのサイズが 16 になり、16 分音符単位の特徴生成を開始する。第 3 転置畳み込み層が $\mathbf{X}_{OPB1}^{(2)} \in \mathbb{R}^{32 \times 4 \times 16 \times 15}$ から

$$\mathbf{X}_{OPB1}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{OPB1}^{(2)}, \mathcal{K}_{OPB1}^{(3)}, \mathbf{s}_{OPB1}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 15} \quad (1.78)$$

を生成する。ここで、 $\mathcal{K}_{OPB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{OPB1}^{(3)} = (1, 6, 1)$ である。この段階で第 3 モードのサイズが 96 になり、32 分音符の 3 連符単位の特徴生成を開始する。以上のように、Other Percussion Block 1 では第 4 モードの特徴生成を行ってから第 3 モードの特徴生成を行う。Other Percussion 1 の転置畳み込み層における各種パラメータをまとめて表 1.25 に示す。

表 1.25: Other Percussion Block 1 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{OPB1}^{(1)} \in \mathbb{R}^{64 \times 64 \times 1 \times 1 \times 15}$	$\mathbf{s}_{OPB1}^{(1)} = (1, 1, 15)$
第 2 層	$\mathcal{K}_{OPB1}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 4 \times 1}$	$\mathbf{s}_{OPB1}^{(2)} = (1, 4, 1)$
第 3 層	$\mathcal{K}_{OPB1}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 6 \times 1}$	$\mathbf{s}_{OPB1}^{(3)} = (1, 6, 1)$

一方、Other Percussion Block 2 では逆の手順で特徴生成を行う。すなわち、第 3 モードの特徴生成を行ってから、第 4 モードの特徴生成を行う。Other Percussion

Block 2 の第 1 転置畳み込み層が $\mathbf{X}_{OPB}^{(0)} \in \mathbb{R}^{128 \times 4 \times 4 \times 1}$ から

$$\mathbf{X}_{OPB2}^{(1)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{OPB}^{(0)}, \mathcal{K}_{OPB2}^{(1)}, \mathbf{s}_{OPB2}^{(1)})) \in \mathbb{R}^{64 \times 4 \times 16 \times 1} \quad (1.79)$$

を生成する．ここで、 $\mathcal{K}_{OPB2}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 4 \times 1}$ であり、 $\mathbf{s}_{OPB2}^{(1)} = (1, 4, 1)$ である．この段階で第 3 モードのサイズが 16 になり、16 分音符単位の特徴生成を開始する．第 2 転置畳み込み層が $\mathbf{X}_{OPB2}^{(1)} \in \mathbb{R}^{64 \times 4 \times 16 \times 1}$ から

$$\mathbf{X}_{OPB2}^{(2)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{OPB2}^{(1)}, \mathcal{K}_{OPB2}^{(2)}, \mathbf{s}_{OPB2}^{(2)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 1} \quad (1.80)$$

を生成する．ここで、 $\mathcal{K}_{OPB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$ であり、 $\mathbf{s}_{OPB2}^{(2)} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 96 になり、32 分音符の 3 連符単位の特徴生成を開始する．第 3 転置畳み込み層が $\mathbf{X}_{OPB2}^{(2)} \in \mathbb{R}^{32 \times 4 \times 96 \times 1}$ から

$$\mathbf{X}_{OPB2}^{(3)} = \text{LeakyReLU}(\text{TrConv}(\mathbf{X}_{OPB2}^{(2)}, \mathcal{K}_{OPB2}^{(3)}, \mathbf{s}_{OPB2}^{(3)})) \in \mathbb{R}^{32 \times 4 \times 96 \times 15} \quad (1.81)$$

を生成する．ここで、 $\mathcal{K}_{OPB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 15}$ であり、 $\mathbf{s}_{OPB2}^{(3)} = (1, 1, 1)$ である．この段階で第 4 モードのサイズが 15 になり、各音色の特徴生成を開始する．Other Percussion 2 の転置畳み込み層における各種パラメータをまとめて表 1.26 に示す．

表 1.26: Other Percussion Block 2 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{OPB2}^{(1)} \in \mathbb{R}^{64 \times 128 \times 1 \times 4 \times 1}$	$\mathbf{s}_{OPB2}^{(1)} = (1, 4, 1)$
第 2 層	$\mathcal{K}_{OPB2}^{(2)} \in \mathbb{R}^{32 \times 64 \times 1 \times 6 \times 1}$	$\mathbf{s}_{OPB2}^{(2)} = (1, 6, 1)$
第 3 層	$\mathcal{K}_{OPB2}^{(3)} \in \mathbb{R}^{32 \times 32 \times 1 \times 1 \times 15}$	$\mathbf{s}_{OPB2}^{(3)} = (1, 1, 1)$

Other Percussion Block 3 では、打楽器トラックに含まれるその他の打楽器のピアノロールを以下のように生成する．まず、入力された特徴 $\mathbf{X}_{OPB1}^{(3)}$ 、 $\mathbf{X}_{OPB2}^{(3)}$ を

$$\mathbf{X}_{OPB3}^{(0)} = \text{Concat}(1, \mathbf{X}_{OPB1}^{(3)}, \mathbf{X}_{OPB2}^{(3)}) \in \mathbb{R}^{64 \times 4 \times 96 \times 15} \quad (1.82)$$

によって結合する．続いて、転置畳み込み層が $\mathbf{X}_{OPB3}^{(0)} \in \mathbb{R}^{64 \times 4 \times 96 \times 15}$ から

$$\mathbf{X}_{OPB3}^{(1)} = \tanh(\text{TrConv}(\mathbf{X}_{OPB3}^{(0)}, \mathcal{K}_{OPB3}^{(1)}, \mathbf{s}_{OPB3}^{(1)})) \in \mathbb{R}^{4 \times 96 \times 15} \quad (1.83)$$

を生成する．ここで、 $\mathcal{K}_{OPB3}^{(1)} \in \mathbb{R}^{1 \times 64 \times 1 \times 1 \times 1}$ であり、 $\mathbf{s}_{OPB3}^{(1)} = (1, 1, 1)$ である．この段階でチャンネル数が 1 になり、ピアノロール生成を完了する．Other Percussion 3 の転置畳み込み層における各種パラメータをまとめて表 1.27 に示す．

表 1.27: Other Percussion Block 3 の転置畳み込み層における各種パラメータ

転置畳み込み層番号	フィルタカーネル	ストライドベクトル
第 1 層	$\mathcal{K}_{OPB3}^{(1)} \in \mathbb{R}^{1 \times 32 \times 1 \times 1}$	$\mathbf{s}_{OPB3}^{(1)} = (1, 1, 1)$

最後に、ドラム打楽器 $\mathbf{X}_{DB3}^{(1)}$ とその他の打楽器 $\mathbf{X}_{OPB3}^{(1)}$ から、

$$\hat{\mathbf{X}}_{Pe} = \text{Trans}(\mathbf{X}_{DB3}^{(1)}, \mathbf{X}_{OPB3}^{(1)}) \in \mathbb{R}^{4 \times 96 \times 84} \quad (1.84)$$

を生成する．ここで、Trans は、 $\mathbf{X}_{DB3}^{(1)} = (X_{DB3}^1, X_{DB3}^2, \dots, X_{DB3}^{15})$ と $\mathbf{X}_{OPB3}^{(1)} = (X_{OPB3}^1, X_{OPB3}^2, \dots, X_{OPB3}^{15})$ から打楽器トラック $\hat{\mathbf{X}}_{Pe} = (\hat{X}_{Percussion}^1, \hat{X}_{Percussion}^2, \dots, \hat{X}_{Percussion}^{84})$ を生成する関数である． $X_{DB3}, X_{OPB3}, \hat{X}_{Percussion}$ の上付きの添え字は音高軸におけるインデックスを示す．Trans は、表 1.28 が示す通りに $\mathbf{X}_{DB3}^{(1)}$ と $\mathbf{X}_{OPB3}^{(1)}$ を並び替える．なお、表 1.28 に示されていない $\hat{\mathbf{X}}_{Pe}$ の音高の部分には全て 0 を埋め込む．以上により、打楽器トラックのピアノロール生成を完了し、 $\hat{\mathbf{X}}_{Pe}$ を出力として出力する．

表 1.28: 関数 $\text{Trans}(\mathbf{X}_{DB3}^{(1)}, \mathbf{X}_{OPB3}^{(1)})$ による打楽器変換

変換前	変換後	変換前	変換後
X_{DB3}^1	$X_{Percussion}^{36}$	X_{OPB3}^1	$X_{Percussion}^{39}$
X_{DB3}^2	$X_{Percussion}^{37}$	X_{OPB3}^2	$X_{Percussion}^{54}$
X_{DB3}^3	$X_{Percussion}^{38}$	X_{OPB3}^3	$X_{Percussion}^{58}$
X_{DB3}^4	$X_{Percussion}^{41}$	X_{OPB3}^4	$X_{Percussion}^{60}$
X_{DB3}^5	$X_{Percussion}^{47}$	X_{OPB3}^5	$X_{Percussion}^{63}$
X_{DB3}^6	$X_{Percussion}^{50}$	X_{OPB3}^6	$X_{Percussion}^{65}$
X_{DB3}^7	$X_{Percussion}^{44}$	X_{OPB3}^7	$X_{Percussion}^{67}$
X_{DB3}^8	$X_{Percussion}^{42}$	X_{OPB3}^8	$X_{Percussion}^{69}$
X_{DB3}^9	$X_{Percussion}^{46}$	X_{OPB3}^9	$X_{Percussion}^{70}$
X_{DB3}^{10}	$X_{Percussion}^{49}$	X_{OPB3}^{10}	$X_{Percussion}^{71}$
X_{DB3}^{11}	$X_{Percussion}^{51}$	X_{OPB3}^{11}	$X_{Percussion}^{73}$
X_{DB3}^{12}	$X_{Percussion}^{52}$	X_{OPB3}^{12}	$X_{Percussion}^{75}$
X_{DB3}^{13}	$X_{Percussion}^{53}$	X_{OPB3}^{13}	$X_{Percussion}^{76}$
X_{DB3}^{14}	$X_{Percussion}^{55}$	X_{OPB3}^{14}	$X_{Percussion}^{79}$
X_{DB3}^{15}	$X_{Percussion}^{56}$	X_{OPB3}^{15}	$X_{Percussion}^{81}$

また，各トラックの生成ピアノロール， $\hat{\mathbf{X}}_{Pi} \in \mathbb{R}^{4 \times 96 \times 30}$ ， $\hat{\mathbf{X}}_{Gu} \in \mathbb{R}^{4 \times 96 \times 84}$ ， $\hat{\mathbf{X}}_{Ba} \in \mathbb{R}^{4 \times 96 \times 84}$ ， $\hat{\mathbf{X}}_{St} \in \mathbb{R}^{4 \times 96 \times 84}$ ， $\hat{\mathbf{X}}_{Pe} \in \mathbb{R}^{4 \times 96 \times 84}$ を

$$\hat{\mathbf{X}}_{TP} = (\hat{\mathbf{X}}_{Pi}, \hat{\mathbf{X}}_{Gu}, \hat{\mathbf{X}}_{Ba}, \hat{\mathbf{X}}_{St}, \hat{\mathbf{X}}_{Pe}) \quad (1.85)$$

によって結合する．これが生成されたマルチトラックピアノロールであり，生成器の最終的な出力 $\mathcal{G}_{TP}(\boldsymbol{\rho})$ となる．

1.8 結び

本章は，生成器を構成する各種ブロックの詳細を述べた．まず有音程楽器と打楽器に共通の Base Block を説明した後，有音程楽器ピアノロールを生成するブロックおよび打楽器ピアノロールを生成するブロックを説明した．