第1章　マルチトラック楽曲の自動生成

1.1　はじめに

本参考資料では，論文

> 有音程楽器と打楽器を分離したネットワークによるマルチトラック楽曲の自動生成

において提案した生成器および判別器の詳細を説明する．以下では，論文本体と同様に，ベクトルをアルファベット小文字の太文字で $\boldsymbol{x}$ のように表記し，行列をアルファベット大文字の細文字斜体 X で，また，3 階テンソルをアルファベット大文字の太文字斜体で $\boldsymbol{X}$ と表記する．更に，4 階テンソルをアルファベット大文字の太文字立体 $\mathbf{X}$ で表記し，5 階テンソルをアルファベット大文字のカリグラフィ体で $\mathcal{X}$ のように表記する．最後に，6 階テンソルをアルファベット大文字のスクリプトフォント体で $\mathscr{X}$ のように表記する

1.2　ネットワークの構成要素

提案法は，従来法と同様に，マルチトラック楽曲生成機構を多層のフィードフォワードネットワークで構成する．これらの層の入出力関係で重要な畳み込み層と転置畳み込み層，および活性化関数を説明する．

1.2.1　畳み込み層

まず，畳み込み層について説明する．この層では，入力データは 4 階テンソル $\mathbf{X} \in \mathbb{R}^{C_i \times B \times R \times P}$ である．ここで，C_i は入力データのチャンネル数，B は小節（Bar）数，R は 1 小節当たりの時間分解能（Resolution），P は音高（Pitch）総数である．また，畳み込みに用いるカーネルを 5 階テンソル $\mathcal{K} \in \mathbb{R}^{C_k \times C_i \times W \times D \times H}$ で表す．ここで，C_k はカーネルのチャンネル数であり，W, D, H はそれぞれ，$\mathcal{K}$ の B, R, P 方向のカーネ

ルサイズを表す．そして，$\mathcal{K}$ の $\mathbf{X}$ への畳み込みを成分ごとに

$$\begin{aligned}&\mathrm{Conv}(\mathbf{X}, \mathcal{K}, \boldsymbol{s})_{c,i,j,k}\\&= \sum_{d=1}^{C_i}\sum_{l=1}^{W}\sum_{m=1}^{D}\sum_{n=1}^{H} \mathcal{K}[c,d,l,m,n]\mathbf{X}[\mathrm{ind}(d,l,m,n)] + b_c\end{aligned} \tag{1.1}$$

と定義する．ここで，$\boldsymbol{s} = (s_w, s_d, s_h)$ はストライドと呼ばれ，出力次元を低減するためのパラメータである．また，

$$\mathrm{ind}(d,l,m,n) = (d, s_w(i-1)+l, s_d(j-1)+m, s_h(k-1)+n)$$

であり，b_c はカーネルの各チャンネルに加算されるバイアスである．

信号処理分野では式（1.1）のように，左辺における加算変数 l, m, n がそれぞれ2箇所に現れ，それが同符号の場合は相互相関と呼ばれるが，機械学習分野の慣例に従って畳み込みと呼ぶことにする．また，畳み込みは第2モード，第3モード，第4モードに対してのみ行っている．第1モードに対しては，入力チャンネルに関する加重和と捉えることができる．

畳み込みの目的は特徴抽出である．このため，畳み込みの出力を特徴マップと呼ぶことがある．式（1.1）より，出力される特徴マップの第2モード，第3モード，第4モードの次元 f_w, f_d, f_h はそれぞれ，

$$f_w = \left\lceil \frac{B-W+1}{s_w} \right\rceil, f_d = \left\lceil \frac{R-D+1}{s_d} \right\rceil, f_h = \left\lceil \frac{P-H+1}{s_h} \right\rceil \tag{1.2}$$

となる．ここで，$\lceil x \rceil$ は x 以上の最小整数である．これらの値は，たとえ $s_w = s_d = s_h = 1$ であっても，それぞれ B, R, P よりも小さい．

特徴マップの第2モードから第4モードまでのサイズを大きくするために，入力データの第2モードから第4モードまでの両端に0を追加することがある．とりわけ，ストライドが $(1,1,1)$，それぞれのモードに対する0の追加総数 p_w, p_d, p_h が，

$$p_w = W-1, p_d = D-1, p_h = H-1 \tag{1.3}$$

である場合がよく用いられる．この p_w, p_d, p_h から，各モードの各端に対する0の追加数 $p_{w1}, p_{w2}, p_{d1}, p_{d2}, p_{h1}, p_{h2}$ を

$$p_{w1} = \left\lfloor \frac{p_w}{2} \right\rfloor, p_{w2} = p_w - p_{w1} \tag{1.4}$$

$$p_{d1} = \left\lfloor \frac{p_d}{2} \right\rfloor, p_{d2} = p_d - p_{d1} \tag{1.5}$$

$$p_{h1} = \left\lfloor \frac{p_h}{2} \right\rfloor, p_{h2} = p_h - p_{h1} \tag{1.6}$$

のように決定する．ここで，$\lfloor x \rfloor$ は x 以下の最大整数である．この場合に特徴マップの第2モードから第4モードまでのサイズが入力データのサイズに一致する．よって，この畳み込みはSame畳み込みと呼ばれる．

1.2.2 転置畳み込み層

次に，転置畳み込み層について説明する．この層は，畳み込み層とは逆に，特徴量データのサイズを拡大していき，小節，時間ステップ，音高からなる音楽データへと展開していくための層である．転置畳み込み層では，入力データ $\mathbf{X}$ に特殊なパディングを行ってから畳み込み操作を行う．パディングは二段階で行われる．まず，ストライド $\boldsymbol{s} = (s_w, s_d, s_h)$ を事前に指定しておく．そして，$\mathbf{X}$ の第 2 モードから第 4 モードまでの要素数が 2 以上であれば，各要素間に $s_w - 1, s_d - 1, s_h - 1$ 個の 0 を挿入する．要素数が 1 であれば挿入は行わない．次に，カーネル $\mathcal{K}$ のサイズ W, D, H を指定しておき，第 2 モードから第 4 モードまでの両端に $W - 1, D - 1, H - 1$ 個の 0 を挿入する．この変換を関数 $\mathrm{Pad}(\mathbf{X}, \mathcal{K}, \boldsymbol{s})$ で表し，出力される 4 階テンソルを $\mathbf{J}$ で表す：

$$\mathbf{J} = \mathrm{Pad}(\mathbf{X}, \mathcal{K}, \boldsymbol{s}) \tag{1.7}$$

ここでのストライドは，パディングを調節するハイパーパラメータである．

転置畳み込み層は，この $\mathbf{J}$ とカーネル $\mathcal{K}$ をストライド $\boldsymbol{s} = \mathbf{1} := (1, 1, 1)$ で以下のように畳み込む：

$$\mathrm{TrConv}(\mathbf{X}, \mathcal{K}, \boldsymbol{s}) = \mathrm{Conv}(\mathbf{J}, \mathcal{K}, \mathbf{1}) \tag{1.8}$$

1.2.3 活性化関数

最後に活性化関数について説明する．ネットワークの各層の出力は活性化関数に入力される．活性化関数には非線形関数が用いられ，アフィン変換である畳み込み層や転置畳み込み層と組み合わせることで，ネットワークの表現力を大幅に向上させる．Binary-MuseGAN と提案法で用いられる sigmoid 関数，tanh 関数，ReLU 関数，LeakyReLU 関数を説明する．

まず，sigmoid 関数は

$$\mathrm{sigmoid}(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{1.9}$$

である．この関数は全ての実数 x において微分可能である．

次に，tanh 関数は

$$\tanh(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}} \tag{1.10}$$

である．この関数も全ての実数 x において微分可能である．

また，ReLU 関数は

$$\mathrm{ReLU}(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{1.11}$$

である．この関数は，$x = 0$ でのみ微分できないので，微分係数が必要な場合は便宜的に 1 を用いる．

最後に，Leaky ReLU 関数は，

$$\mathrm{LeakyReLU}(x) = \begin{cases} \gamma x & x < 0 \\ x & x \geq 0 \end{cases} \tag{1.12}$$

を出力する．ここで，γ は $0 < \gamma < 1$ を満たすハイパーパラメータである．論文本体の計算機シミュレーションでは $\gamma = 0.2$ に設定してある．LeakyReLU 関数も $x = 0$ でのみ微分係数は定義できないが，この値が必要な場合は便宜的に 1 とおく．

第2章　生成器の構成ブロック

2.1　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．

2.2　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，

$$\boldsymbol{x}_{BB} = \mathrm{LearkyReLU}(W_{BB}\boldsymbol{\rho} + \boldsymbol{b}_{BB}) \in \mathbb{R}^{128} \tag{2.1}$$

を生成する．ここで，$W_{BB} \in \mathbb{R}^{128\times128}$ であり，$\boldsymbol{b}_{BB} \in \mathbb{R}^{128}$ である．式（2.1）の $\boldsymbol{x}_{BB}$ 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．

2.3　Tonal Block

Tonal Block は，全トラック共通特徴量 $\boldsymbol{x}_{BB} \in \mathbb{R}^{128}$ 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 $\boldsymbol{x}_{BB}$ から，

$$\mathbf{X}_{TB}^{(0)} = \mathrm{Reshape}_{TB}(W_{TB}^{(0)}\boldsymbol{x}_{BB} + \boldsymbol{b}_{TB}) \in \mathbb{R}^{256\times3\times1\times1} \tag{2.2}$$

を生成する．ここで，$W_{TB}^{(0)} \in \mathbb{R}^{768\times128}$ は矩形行列であり，$\boldsymbol{b}_{TB} \in \mathbb{R}^{768}$ はバイアスベクトルである．また，$\mathrm{Reshape}_{TB}$ は 768 次元ベクトルを $256\times3\times1\times1$ 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 $\mathbf{X}_{TB}^{(0)} \in \mathbb{R}^{256\times3\times1\times1}$ から

$$\mathbf{X}_{TB}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}^{(0)}, \mathcal{K}_{TB}^{(1)}, \boldsymbol{s}_{TB}^{(1)})) \in \mathbb{R}^{128\times4\times1\times1} \tag{2.3}$$

を生成する．ここで，$\mathcal{K}_{TB}^{(1)} \in \mathbb{R}^{128\times256\times2\times1\times1}$ であり，$\boldsymbol{s}_{TB}^{(1)} = (1,1,1)$ 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，$\mathbf{X}_{TB}^{(1)} \in \mathbb{R}^{128\times4\times1\times1}$ から

$$\mathbf{X}_{TB}^{(2)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}^{(1)}, \mathcal{K}_{TB}^{(2)}, \boldsymbol{s}_{TB}^{(2)})) \in \mathbb{R}^{128\times4\times1\times3} \tag{2.4}$$

を生成する．ここで，$\mathcal{K}_{TB}^{(2)} \in \mathbb{R}^{128\times128\times1\times1\times3}$ であり，$\boldsymbol{s}_{TB}^{(2)} = (1,1,1)$ 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 $\mathbf{X}_{TB}^{(2)} \in \mathbb{R}^{128\times4\times1\times3}$ から

$$\mathbf{X}_{TB}^{(3)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}^{(2)}, \mathcal{K}_{TB}^{(3)}, \boldsymbol{s}_{TB}^{(3)})) \in \mathbb{R}^{128\times4\times2\times3} \tag{2.5}$$

を生成する．ここで，$\mathcal{K}_{TB}^{(3)} \in \mathbb{R}^{128\times128\times1\times2\times1}$ であり，$\boldsymbol{s}_{TB}^{(3)} = (1,1,1)$ 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 $\mathbf{X}_{TB}^{(3)} \in \mathbb{R}^{128\times4\times2\times3}$ から

$$\mathbf{X}_{TB}^{(4)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}^{(3)}, \mathcal{K}_{TB}^{(4)}, \boldsymbol{s}_{TB}^{(4)})) \in \mathbb{R}^{128\times4\times4\times3} \tag{2.6}$$

を生成する．ここで，$\mathcal{K}_{TB}^{(4)} \in \mathbb{R}^{128\times128\times1\times2\times1}$ であり，$\boldsymbol{s}_{TB}^{(4)} = (1,2,1)$ 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 $\mathbf{X}_{TB}^{(4)} \in \mathbb{R}^{128\times4\times4\times3}$ から

$$\mathbf{X}_{TB}^{(5)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}^{(4)}, \mathcal{K}_{TB}^{(5)}, \boldsymbol{s}_{TB}^{(5)})) \in \mathbb{R}^{64\times4\times4\times6} \tag{2.7}$$

を生成する．ここで，$\mathcal{K}_{TB}^{(5)} \in \mathbb{R}^{64\times128\times1\times1\times2}$ であり，$\boldsymbol{s}_{TB}^{(5)} = (1,1,2)$ 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 $\mathbf{X}_{TB}^{(5)} \in \mathbb{R}^{64\times4\times4\times6}$ から

$$\mathbf{X}_{TB}^{(6)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}^{(5)}, \mathcal{K}_{TB}^{(6)}, \boldsymbol{s}_{TB}^{(6)})) \in \mathbb{R}^{64\times4\times4\times12} \tag{2.8}$$

を生成する．ここで，$\mathcal{K}_{TB}^{(6)} \in \mathbb{R}^{64\times64\times1\times1\times2}$ であり，$\boldsymbol{s}_{TB}^{(6)} = (1,1,2)$ 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 2.1 に示す．

表 2.1: 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	ストライドベクトル
第 1 層	$\mathcal{K}_{TB}^{(1)} \in \mathbb{R}^{128\times256\times2\times1\times1}$	$\boldsymbol{s}_{TB}^{(1)} = (1,1,1)$
第 2 層	$\mathcal{K}_{TB}^{(2)} \in \mathbb{R}^{128\times128\times1\times1\times3}$	$\boldsymbol{s}_{TB}^{(2)} = (1,1,1)$
第 3 層	$\mathcal{K}_{TB}^{(3)} \in \mathbb{R}^{128\times128\times1\times2\times1}$	$\boldsymbol{s}_{TB}^{(3)} = (1,1,1)$
第 4 層	$\mathcal{K}_{TB}^{(4)} \in \mathbb{R}^{128\times128\times1\times2\times1}$	$\boldsymbol{s}_{TB}^{(4)} = (1,2,1)$
第 5 層	$\mathcal{K}_{TB}^{(5)} \in \mathbb{R}^{64\times128\times1\times1\times2}$	$\boldsymbol{s}_{TB}^{(5)} = (1,1,2)$
第 6 層	$\mathcal{K}_{TB}^{(6)} \in \mathbb{R}^{64\times64\times1\times1\times2}$	$\boldsymbol{s}_{TB}^{(6)} = (1,1,2)$

2.4 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．

2.4.1 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 2.2 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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}_{PiLB1}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{PiLB1}^{(1)}, \boldsymbol{s}_{PiLB1}^{(1)})) \in \mathbb{R}^{64\times4\times4\times36} \tag{2.9}$$

を生成する．ここで，$\mathcal{K}_{PiLB1}^{(1)} \in \mathbb{R}^{64\times64\times1\times1\times3}$ であり，$\boldsymbol{s}_{PiLB1}^{(1)} = (1,1,3)$ 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 $\mathbf{X}^{(1)}_{PiLB1} \in \mathbb{R}^{64\times4\times4\times36}$ から

$$\mathbf{X}^{(2)}_{PiLB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{PiLB1}, \mathcal{K}^{(2)}_{PiLB1}, \boldsymbol{s}^{(2)}_{PiLB1})) \in \mathbb{R}^{32\times4\times16\times36} \tag{2.10}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{PiLB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{PiLB1} = (1, 4, 1)$ 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 $\mathbf{X}^{(2)}_{PiLB1} \in \mathbb{R}^{32\times4\times16\times36}$ から

$$\mathbf{X}^{(3)}_{PiLB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{PiLB1}, \mathcal{K}^{(3)}_{PiLB1}, \boldsymbol{s}^{(3)}_{PiLB1})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.11}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{PiLB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}^{(3)}_{PiLB1} = (1, 6, 1)$ 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を生成する．

表 2.2: 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	$\mathcal{K}^{(1)}_{PiLB1} \in \mathbb{R}^{64\times64\times1\times1\times3}$	$\boldsymbol{s}^{(1)}_{PiLB1} = (1, 1, 3)$
第 2 層	$\mathcal{K}^{(2)}_{PiLB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}^{(2)}_{PiLB1} = (1, 4, 1)$
第 3 層	$\mathcal{K}^{(3)}_{PiLB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}^{(3)}_{PiLB1} = (1, 6, 1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{PiLB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{PiLB2}, \boldsymbol{s}^{(1)}_{PiLB2})) \in \mathbb{R}^{64\times4\times16\times12} \tag{2.12}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{PiLB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(1)}_{PiLB2} = (1, 4, 1)$ 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 $\mathbf{X}^{(1)}_{PiLB2} \in \mathbb{R}^{64\times4\times16\times12}$ から

$$\mathbf{X}^{(2)}_{PiLB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{PiLB2}, \mathcal{K}^{(2)}_{PiLB2}, \boldsymbol{s}^{(2)}_{PiLB2})) \in \mathbb{R}^{32\times4\times96\times12} \tag{2.13}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{PiLB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}^{(2)}_{PiLB2} = (1, 6, 1)$ 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 $\mathbf{X}^{(2)}_{PiLB2} \in \mathbb{R}^{32\times4\times96\times12}$ から

$$\mathbf{X}^{(3)}_{PiLB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{PiLB2}, \mathcal{K}^{(3)}_{PiLB2}, \boldsymbol{s}^{(3)}_{PiLB2})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.14}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{PiLB2} \in \mathbb{R}^{32\times32\times1\times1\times3}$ であり，$\boldsymbol{s}^{(3)}_{PiLB2} = (1,1,3)$ 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 2.3 にまとめて示す．

表 2.3: 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	$\mathcal{K}^{(1)}_{PiLB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$	$\boldsymbol{s}^{(1)}_{PiLB2} = (1,4,1)$
第 2 層	$\mathcal{K}^{(2)}_{PiLB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}^{(2)}_{PiLB2} = (1,6,1)$
第 3 層	$\mathcal{K}^{(3)}_{PiLB2} \in \mathbb{R}^{32\times32\times1\times1\times3}$	$\boldsymbol{s}^{(3)}_{PiLB2} = (1,1,3)$

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 $\mathbf{X}^{(3)}_{PiLB1}$，$\mathbf{X}^{(3)}_{PiLB2}$ を

$$\mathbf{X}^{(0)}_{PiLB3} = \mathrm{Concat}(1, \mathbf{X}^{(3)}_{PiLB1}, \mathbf{X}^{(3)}_{PiLB2}) \in \mathbb{R}^{64\times4\times96\times36} \tag{2.15}$$

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．$\mathbf{X}^{(3)}_{PiLB1}$ と $\mathbf{X}^{(3)}_{PiLB2}$ 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 $\mathbf{X}^{(0)}_{PiLB3} \in \mathbb{R}^{64\times4\times96\times36}$ から

$$\boldsymbol{X}^{(1)}_{PiLB3} = \tanh(\mathrm{TrConv}(\mathbf{X}^{(0)}_{PiLB3}, \mathcal{K}^{(1)}_{PiLB3}, \boldsymbol{s}^{(1)}_{PiLB3})) \in \mathbb{R}^{4\times96\times36} \tag{2.16}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{PiLB3} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{PiLB3} = (1,1,1)$ 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 2.4 にまとめて示す．

表 2.4: 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	$\mathcal{K}^{(1)}_{PiLB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{PiLB3} = (1,1,1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{PiHB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{PiHB1}, \boldsymbol{s}^{(1)}_{PiHB1})) \in \mathbb{R}^{64\times4\times4\times48} \tag{2.17}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{PiHB1} \in \mathbb{R}^{64\times64\times1\times1\times4}$ であり，$\boldsymbol{s}^{(1)}_{PiHB1} = (1,1,4)$ 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 $\mathbf{X}^{(1)}_{PiHB1} \in \mathbb{R}^{64\times4\times4\times48}$ から

$$\mathbf{X}^{(2)}_{PiHB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{PiHB1}, \mathcal{K}^{(2)}_{PiHB1}, \boldsymbol{s}^{(2)}_{PiHB1})) \in \mathbb{R}^{32\times4\times16\times48} \tag{2.18}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{PiHB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{PiHB1} = (1,4,1)$ 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 $\mathbf{X}^{(2)}_{PiHB1} \in \mathbb{R}^{32\times4\times16\times48}$ から

$$\mathbf{X}^{(3)}_{PiHB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{PiHB1}, \mathcal{K}^{(3)}_{PiHB1}, \boldsymbol{s}^{(3)}_{PiHB1})) \in \mathbb{R}^{32\times4\times96\times48} \tag{2.19}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{PiHB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}^{(3)}_{PiHB1} = (1,6,1)$ 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 2.5 にまとめて示す．

表 2.5: 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	ストライドベクトル
第 1 層	$\mathcal{K}^{(1)}_{PiHB1} \in \mathbb{R}^{64\times64\times1\times1\times4}$	$\boldsymbol{s}^{(1)}_{PiHB1} = (1,1,4)$
第 2 層	$\mathcal{K}^{(2)}_{PiHB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}^{(2)}_{PiHB1} = (1,4,1)$
第 3 層	$\mathcal{K}^{(3)}_{PiHB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}^{(3)}_{PiHB1} = (1,6,1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{PiHB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{PiHB2}, \boldsymbol{s}^{(1)}_{PiHB2})) \in \mathbb{R}^{64\times4\times16\times12} \tag{2.20}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{PiHB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(1)}_{PiHB2} = (1,4,1)$ 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 $\mathbf{X}^{(1)}_{PiHB2} \in \mathbb{R}^{64\times4\times16\times12}$ から

$$\mathbf{X}^{(2)}_{PiHB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{PiHB2}, \mathcal{K}^{(2)}_{PiHB2}, \boldsymbol{s}^{(2)}_{PiHB2})) \in \mathbb{R}^{32\times4\times96\times12} \tag{2.21}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{PiHB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}^{(2)}_{PiHB2} = (1,6,1)$ 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 $\mathbf{X}^{(2)}_{PiHB2} \in \mathbb{R}^{32\times4\times96\times12}$ から

$$\mathbf{X}^{(3)}_{PiHB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{PiHB2}, \mathcal{K}^{(3)}_{PiHB2}, \boldsymbol{s}^{(3)}_{PiHB2})) \in \mathbb{R}^{32\times4\times96\times48} \tag{2.22}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{PiHB2} \in \mathbb{R}^{32\times32\times1\times1\times4}$ であり，$\boldsymbol{s}^{(3)}_{PiHB2} = (1,1,4)$ 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 2.6 にまとめて示す．

表 2.6: 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	$\mathcal{K}^{(1)}_{PiHB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$	$\boldsymbol{s}^{(1)}_{PiHB2} = (1,4,1)$
第 2 層	$\mathcal{K}^{(2)}_{PiHB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}^{(2)}_{PiHB2} = (1,6,1)$
第 3 層	$\mathcal{K}^{(3)}_{PiHB2} \in \mathbb{R}^{32\times32\times1\times1\times4}$	$\boldsymbol{s}^{(3)}_{PiHB2} = (1,1,4)$

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 $\mathbf{X}^{(3)}_{PiHB1}$，$\mathbf{X}^{(3)}_{PiHB2}$ を

$$\mathbf{X}^{(0)}_{PiHB3} = \mathrm{Concat}(1, \mathbf{X}^{(3)}_{PiHB1}, \mathbf{X}^{(3)}_{PiHB2}) \in \mathbb{R}^{64\times4\times96\times48} \tag{2.23}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}^{(0)}_{PiHB3} \in \mathbb{R}^{64\times4\times96\times48}$ から

$$\boldsymbol{X}^{(1)}_{PiHB3} = \tanh(\mathrm{TrConv}(\mathbf{X}^{(0)}_{PiHB3}, \mathcal{K}^{(1)}_{PiHB3}, \boldsymbol{s}^{(1)}_{PiHB3})) \in \mathbb{R}^{4\times96\times48} \tag{2.24}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{PiHB3} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{PiHB3} = (1,1,1)$ 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 2.7 にまとめて示す．

表 2.7: 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	$\mathcal{K}^{(1)}_{PiHB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{PiHB3} = (1,1,1)$

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

$$\hat{\boldsymbol{X}}_{Pi} = \mathrm{Concat}(3, \boldsymbol{X}^{(1)}_{PiLB3}, \boldsymbol{X}^{(1)}_{PiHB3}) \in \mathbb{R}^{4\times96\times84} \tag{2.25}$$

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．

2.4.2 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♯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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{GB1}, \boldsymbol{s}^{(1)}_{GB1})) \in \mathbb{R}^{64\times4\times4\times48} \tag{2.26}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{GB1} \in \mathbb{R}^{64\times64\times1\times1\times4}$ であり，$\boldsymbol{s}^{(1)}_{GB1} = (1, 1, 4)$ 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 $\mathbf{X}^{(1)}_{GB1} \in \mathbb{R}^{64\times4\times4\times48}$ から

$$\mathbf{X}^{(2)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{GB1}, \mathcal{K}^{(2)}_{GB1}, \boldsymbol{s}^{(2)}_{GB1})) \in \mathbb{R}^{32\times4\times16\times48} \tag{2.27}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{GB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{GB1} = (1, 4, 1)$ 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 $\mathbf{X}^{(2)}_{GB1} \in \mathbb{R}^{32\times4\times16\times48}$ から

$$\mathbf{X}^{(3)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{GB1}, \mathcal{K}^{(3)}_{GB1}, \boldsymbol{s}^{(3)}_{GB1})) \in \mathbb{R}^{32\times4\times96\times48} \tag{2.28}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{GB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}^{(3)}_{GB1} = (1, 6, 1)$ 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 2.8 にまとめて示す．

表 2.8: 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	$\mathcal{K}^{(1)}_{GB1} \in \mathbb{R}^{64\times64\times1\times1\times4}$	$\boldsymbol{s}^{(1)}_{GB1} = (1, 1, 4)$
第 2 層	$\mathcal{K}^{(2)}_{GB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}^{(2)}_{GB1} = (1, 4, 1)$
第 3 層	$\mathcal{K}^{(3)}_{GB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}^{(3)}_{GB1} = (1, 6, 1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{GB2}, \boldsymbol{s}^{(1)}_{GB2})) \in \mathbb{R}^{64\times4\times16\times12} \tag{2.29}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{GB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(1)}_{GB2} = (1, 4, 1)$ 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 $\mathbf{X}^{(1)}_{GB2} \in \mathbb{R}^{64\times4\times16\times12}$ から

$$\mathbf{X}^{(2)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{GB2}, \mathcal{K}^{(2)}_{GB2}, \boldsymbol{s}^{(2)}_{GB2})) \in \mathbb{R}^{32\times4\times96\times12} \tag{2.30}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{GB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}^{(2)}_{GB2} = (1, 6, 1)$ 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 $\mathbf{X}^{(2)}_{GB2} \in \mathbb{R}^{32\times4\times96\times12}$ から

$$\mathbf{X}^{(3)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{GB2}, \mathcal{K}^{(3)}_{GB2}, \boldsymbol{s}^{(3)}_{GB2})) \in \mathbb{R}^{32\times4\times96\times48} \tag{2.31}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{GB2} \in \mathbb{R}^{32\times32\times1\times1\times4}$ であり，$\boldsymbol{s}^{(3)}_{GB2} = (1, 1, 4)$ 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 2.9 にまとめて示す．

表 2.9: 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	$\mathcal{K}^{(1)}_{GB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$	$\boldsymbol{s}^{(1)}_{GB2} = (1, 4, 1)$
第 2 層	$\mathcal{K}^{(2)}_{GB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}^{(2)}_{GB2} = (1, 6, 1)$
第 3 層	$\mathcal{K}^{(3)}_{GB2} \in \mathbb{R}^{32\times32\times1\times1\times4}$	$\boldsymbol{s}^{(3)}_{GB2} = (1, 1, 4)$

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に生成する．まず，入力された特徴 $\mathbf{X}^{(3)}_{GB1}$，$\mathbf{X}^{(3)}_{GB2}$ を

$$\mathbf{X}^{(0)}_{GB3} = \mathrm{Concat}(1, \mathbf{X}^{(3)}_{GB1}, \mathbf{X}^{(3)}_{GB2}) \in \mathbb{R}^{64\times4\times96\times48} \tag{2.32}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}^{(0)}_{GB3} \in \mathbb{R}^{64\times4\times96\times48}$ から

$$\boldsymbol{X}^{(1)}_{GB3} = \tanh(\mathrm{TrConv}(\mathbf{X}^{(0)}_{GB3}, \mathcal{K}^{(1)}_{GB3}, \boldsymbol{s}^{(1)}_{GB3})) \in \mathbb{R}^{4\times96\times48} \tag{2.33}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{GB3} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{GB3} = (1, 1, 1)$ 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 2.10 にまとめて示す．

表 2.10: 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	$\mathcal{K}^{(1)}_{GB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{GB3} = (1,1,1)$

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，$\boldsymbol{X}^{(1)}_{GB3} \in \mathbb{R}^{4\times96\times48}$ 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 $Pad_G(\boldsymbol{X}^{(1)}_{GB3})$ で表し，出力される 3 階テンソルを $\hat{\boldsymbol{X}}_{Gu}$ で表す：

$$\hat{\boldsymbol{X}}_{Gu} = Pad_G(\boldsymbol{X}^{(1)}_{GB3}) \in \mathbb{R}^{4\times96\times84} \tag{2.34}$$

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

2.4.3 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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{BaB1}, \boldsymbol{s}^{(1)}_{BaB1})) \in \mathbb{R}^{64\times4\times4\times36} \tag{2.35}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{BaB1} \in \mathbb{R}^{64\times64\times1\times1\times3}$ であり，$\boldsymbol{s}^{(1)}_{BaB1} = (1,1,3)$ 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 $\mathbf{X}^{(1)}_{BaB1} \in \mathbb{R}^{64\times4\times4\times36}$ から

$$\mathbf{X}^{(2)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{BaB1}, \mathcal{K}^{(2)}_{BaB1}, \boldsymbol{s}^{(2)}_{BaB1})) \in \mathbb{R}^{32\times4\times16\times36} \tag{2.36}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{BaB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{BaB1} = (1,4,1)$ 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 $\mathbf{X}^{(2)}_{BaB1} \in \mathbb{R}^{32\times4\times16\times36}$ から

$$\mathbf{X}^{(3)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{BaB1}, \mathcal{K}^{(3)}_{BaB1}, \boldsymbol{s}^{(3)}_{BaB1})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.37}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{BaB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}^{(3)}_{BaB1} = (1,6,1)$ 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 2.11 にまとめて示す．

表 2.11: 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	$\mathcal{K}_{BaB1}^{(1)} \in \mathbb{R}^{64\times64\times1\times1\times3}$	$\boldsymbol{s}_{BaB1}^{(1)} = (1,1,3)$
第 2 層	$\mathcal{K}_{BaB1}^{(2)} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}_{BaB1}^{(2)} = (1,4,1)$
第 3 層	$\mathcal{K}_{BaB1}^{(3)} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}_{BaB1}^{(3)} = (1,6,1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}_{BaB2}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}_{BaB2}^{(1)}, \boldsymbol{s}_{BaB2}^{(1)})) \in \mathbb{R}^{64\times4\times16\times12} \tag{2.38}$$

を生成する．ここで，$\mathcal{K}_{BaB2}^{(1)} \in \mathbb{R}^{64\times64\times1\times4\times1}$ であり，$\boldsymbol{s}_{BaB2}^{(1)} = (1,4,1)$ 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 $\mathbf{X}_{BaB2}^{(1)} \in \mathbb{R}^{64\times4\times16\times12}$ から

$$\mathbf{X}_{BaB2}^{(2)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{BaB2}^{(1)}, \mathcal{K}_{BaB2}^{(2)}, \boldsymbol{s}_{BaB2}^{(2)})) \in \mathbb{R}^{32\times4\times96\times12} \tag{2.39}$$

を生成する．ここで，$\mathcal{K}_{BaB2}^{(2)} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}_{BaB2}^{(2)} = (1,6,1)$ 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 $\mathbf{X}_{BaB2}^{(2)} \in \mathbb{R}^{32\times4\times96\times12}$ から

$$\mathbf{X}_{BaB2}^{(3)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{BaB2}^{(2)}, \mathcal{K}_{BaB2}^{(3)}, \boldsymbol{s}_{BaB2}^{(3)})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.40}$$

を生成する．ここで，$\mathcal{K}_{BaB2}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times3}$ であり，$\boldsymbol{s}_{BaB2}^{(3)} = (1,1,3)$ 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 2.12 にまとめて示す．

表 2.12: 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	$\mathcal{K}_{BaB2}^{(1)} \in \mathbb{R}^{64\times64\times1\times4\times1}$	$\boldsymbol{s}_{BaB2}^{(1)} = (1,4,1)$
第 2 層	$\mathcal{K}_{BaB2}^{(2)} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}_{BaB2}^{(2)} = (1,6,1)$
第 3 層	$\mathcal{K}_{BaB2}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times3}$	$\boldsymbol{s}_{BaB2}^{(3)} = (1,1,3)$

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 $\mathbf{X}^{(3)}_{BaB1}$，$\mathbf{X}^{(3)}_{BaB2}$ を

$$\mathbf{X}^{(0)}_{BaB3} = \mathrm{Concat}(1, \mathbf{X}^{(3)}_{BaB1}, \mathbf{X}^{(3)}_{BaB2}) \in \mathbb{R}^{64\times4\times96\times36} \tag{2.41}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}^{(0)}_{BaB3} \in \mathbb{R}^{64\times4\times96\times36}$ から

$$\boldsymbol{X}^{(1)}_{BaB3} = \tanh(\mathrm{TrConv}(\mathbf{X}^{(0)}_{BaB3}, \mathcal{K}^{(1)}_{BaB3}, \boldsymbol{s}^{(1)}_{BaB3})) \in \mathbb{R}^{4\times96\times36} \tag{2.42}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{BaB3} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{BaB3} = (1, 1, 1)$ 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 2.13 にまとめて示す．

表 2.13: 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	$\mathcal{K}^{(1)}_{BaB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{BaB3} = (1, 1, 1)$

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，$\boldsymbol{X}^{(1)}_{BaB3} \in \mathbb{R}^{4\times96\times36}$ 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 $Pad_Ba(\boldsymbol{X}^{(1)}_{BaB3})$ で表し，出力される 3 階テンソルを $\hat{\boldsymbol{X}}_{Ba}$ で表す：

$$\hat{\boldsymbol{X}}_{Ba} = Pad_{Ba}(\boldsymbol{X}^{(1)}_{BaB3}) \in \mathbb{R}^{4\times96\times84} \tag{2.43}$$

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

2.4.4 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 2.14 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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{SLB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{SLB1}, \boldsymbol{s}^{(1)}_{SLB1})) \in \mathbb{R}^{64\times4\times4\times36} \tag{2.44}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{SLB1} \in \mathbb{R}^{64\times64\times1\times1\times3}$ であり，$\boldsymbol{s}^{(1)}_{SLB1} = (1,1,3)$ 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 $\mathbf{X}^{(1)}_{SLB1} \in \mathbb{R}^{64\times4\times4\times36}$ から

$$\mathbf{X}^{(2)}_{SLB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{SLB1}, \mathcal{K}^{(2)}_{SLB1}, \boldsymbol{s}^{(2)}_{SLB1})) \in \mathbb{R}^{32\times4\times16\times36} \tag{2.45}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{SLB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{SLB1} = (1,4,1)$ 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 $\mathbf{X}^{(2)}_{SLB1} \in \mathbb{R}^{32\times4\times16\times36}$ から

$$\mathbf{X}^{(3)}_{SLB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{SLB1}, \mathcal{K}^{(3)}_{SLB1}, \boldsymbol{s}^{(3)}_{SLB1})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.46}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{SLB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}^{(3)}_{SLB1} = (1,6,1)$ 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 2.14: 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	$\mathcal{K}^{(1)}_{SLB1} \in \mathbb{R}^{64\times64\times1\times1\times3}$	$\boldsymbol{s}^{(1)}_{SLB1} = (1,1,3)$
第 2 層	$\mathcal{K}^{(2)}_{SLB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}^{(2)}_{SLB1} = (1,4,1)$
第 3 層	$\mathcal{K}^{(3)}_{SLB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}^{(3)}_{SLB1} = (1,6,1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{SLB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{SLB2}, \boldsymbol{s}^{(1)}_{SLB2})) \in \mathbb{R}^{64\times4\times16\times12} \tag{2.47}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{SLB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(1)}_{SLB2} = (1,4,1)$ 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 $\mathbf{X}^{(1)}_{SLB2} \in \mathbb{R}^{64\times4\times16\times12}$ から

$$\mathbf{X}^{(2)}_{SLB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{SLB1}, \mathcal{K}^{(2)}_{SLB2}, \boldsymbol{s}^{(2)}_{SLB2})) \in \mathbb{R}^{32\times4\times96\times12} \tag{2.48}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{SLB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}^{(2)}_{SLB2} = (1,6,1)$ 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 $\mathbf{X}^{(2)}_{SLB2} \in \mathbb{R}^{32\times4\times96\times12}$ から

$$\mathbf{X}^{(3)}_{SLB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{SLB2}, \mathcal{K}^{(3)}_{SLB2}, \boldsymbol{s}^{(3)}_{SLB2})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.49}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{SLB2} \in \mathbb{R}^{32\times32\times1\times1\times3}$ であり，$\boldsymbol{s}^{(3)}_{SLB2} = (1,1,3)$ 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 2.15 にまとめて示す．

表 2.15: 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	$\mathcal{K}^{(1)}_{SLB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$	$\boldsymbol{s}^{(1)}_{SLB2} = (1,4,1)$
第 2 層	$\mathcal{K}^{(2)}_{SLB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}^{(2)}_{SLB2} = (1,6,1)$
第 3 層	$\mathcal{K}^{(3)}_{SLB2} \in \mathbb{R}^{32\times32\times1\times1\times3}$	$\boldsymbol{s}^{(3)}_{SLB2} = (1,1,3)$

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 $\mathbf{X}^{(3)}_{SLB1}$，$\mathbf{X}^{(3)}_{SLB2}$ を

$$\mathbf{X}^{(0)}_{SLB3} = \mathrm{Concat}(1, \mathbf{X}^{(3)}_{SLB1}, \mathbf{X}^{(3)}_{SLB2}) \in \mathbb{R}^{64\times4\times96\times36} \tag{2.50}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}^{(0)}_{SLB3} \in \mathbb{R}^{64\times4\times96\times36}$ から

$$\boldsymbol{X}^{(1)}_{SLB3} = \tanh(\mathrm{TrConv}(\mathbf{X}^{(0)}_{SLB3}, \mathcal{K}^{(1)}_{SLB3}, \boldsymbol{s}^{(1)}_{SLB3})) \in \mathbb{R}^{4\times96\times36} \tag{2.51}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{SLB3} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{SLB3} = (1,1,1)$ 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 2.16 にまとめて示す．

表 2.16: 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	$\mathcal{K}^{(1)}_{SLB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{SLB3} = (1,1,1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{SHB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{SHB1}, \boldsymbol{s}^{(1)}_{SHB1})) \in \mathbb{R}^{64\times4\times4\times36} \tag{2.52}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{SHB1} \in \mathbb{R}^{64\times64\times1\times1\times3}$ であり，$\boldsymbol{s}^{(1)}_{SHB1} = (1,1,3)$ 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 $\mathbf{X}^{(1)}_{SHB1} \in \mathbb{R}^{64\times4\times4\times36}$ から

$$\mathbf{X}^{(2)}_{SHB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{SHB1}, \mathcal{K}^{(2)}_{SHB1}, \boldsymbol{s}^{(2)}_{SHB1})) \in \mathbb{R}^{32\times4\times16\times36} \tag{2.53}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{SHB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{SHB1} = (1, 4, 1)$ 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 $\mathbf{X}^{(2)}_{SHB1} \in \mathbb{R}^{32\times4\times16\times36}$ から

$$\mathbf{X}^{(3)}_{SHB1} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{SHB1}, \mathcal{K}^{(3)}_{SHB1}, \boldsymbol{s}^{(3)}_{SHB1})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.54}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{SHB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}^{(3)}_{SHB1} = (1, 6, 1)$ 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 2.17 にまとめて示す.

表 2.17: 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	$\mathcal{K}^{(1)}_{SHB1} \in \mathbb{R}^{64\times64\times1\times1\times4}$	$\boldsymbol{s}^{(1)}_{SHB1} = (1, 1, 4)$
第 2 層	$\mathcal{K}^{(2)}_{SHB1} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}^{(2)}_{SHB1} = (1, 4, 1)$
第 3 層	$\mathcal{K}^{(3)}_{SHB1} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}^{(3)}_{SHB1} = (1, 6, 1)$

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 $\mathbf{X}_{TB} \in \mathbb{R}^{64\times4\times4\times12}$ から

$$\mathbf{X}^{(1)}_{SHB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{TB}, \mathcal{K}^{(1)}_{SHB2}, \boldsymbol{s}^{(1)}_{SHB2})) \in \mathbb{R}^{64\times4\times16\times12} \tag{2.55}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{SHB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$ であり，$\boldsymbol{s}^{(1)}_{SHB2} = (1, 4, 1)$ 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 $\mathbf{X}^{(1)}_{SHB2} \in \mathbb{R}^{64\times4\times16\times12}$ から

$$\mathbf{X}^{(2)}_{SHB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{SHB2}, \mathcal{K}^{(2)}_{SHB2}, \boldsymbol{s}^{(2)}_{SHB2})) \in \mathbb{R}^{32\times4\times96\times12} \tag{2.56}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{SHB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}^{(2)}_{SHB2} = (1, 6, 1)$ 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 $\mathbf{X}^{(2)}_{SHB2} \in \mathbb{R}^{32\times4\times96\times12}$ から

$$\mathbf{X}^{(3)}_{SHB2} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{SHB2}, \mathcal{K}^{(3)}_{SHB2}, \boldsymbol{s}^{(3)}_{SHB2})) \in \mathbb{R}^{32\times4\times96\times36} \tag{2.57}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{SHB2} \in \mathbb{R}^{32\times32\times1\times1\times3}$ であり，$\boldsymbol{s}^{(3)}_{SHB2} = (1,1,3)$ 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 2.18 にまとめて示す．

表 2.18: 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	$\mathcal{K}^{(1)}_{SHB2} \in \mathbb{R}^{64\times64\times1\times4\times1}$	$\boldsymbol{s}^{(1)}_{SHB2} = (1,4,1)$
第 2 層	$\mathcal{K}^{(2)}_{SHB2} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}^{(2)}_{SHB2} = (1,6,1)$
第 3 層	$\mathcal{K}^{(3)}_{SHB2} \in \mathbb{R}^{32\times32\times1\times1\times3}$	$\boldsymbol{s}^{(3)}_{SHB2} = (1,1,3)$

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 $\mathbf{X}^{(3)}_{SHB1}$，$\mathbf{X}^{(3)}_{SHB2}$ を

$$\mathbf{X}^{(0)}_{SHB3} = \mathrm{Concat}(1, \mathbf{X}^{(3)}_{SHB1}, \mathbf{X}^{(3)}_{SHB2}) \in \mathbb{R}^{64\times4\times96\times36} \tag{2.58}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}^{(0)}_{SHB3} \in \mathbb{R}^{64\times4\times96\times36}$ から

$$\boldsymbol{X}^{(1)}_{SHB3} = \tanh(\mathrm{TrConv}(\mathbf{X}^{(0)}_{SHB3}, \mathcal{K}^{(1)}_{SHB3}, \boldsymbol{s}^{(1)}_{SHB3})) \in \mathbb{R}^{4\times96\times36} \tag{2.59}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{SHB3} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{SHB3} = (1,1,1)$ 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 2.19 にまとめて示す．

表 2.19: 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	$\mathcal{K}^{(1)}_{SHB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{SHB3} = (1,1,1)$

以上の操作で得られた低域 $\boldsymbol{X}^{(1)}_{SLB3} \in \mathbb{R}^{4\times96\times36}$ と高域 $\boldsymbol{X}^{(1)}_{SHB3} \in \mathbb{R}^{4\times96\times36}$ を

$$\boldsymbol{X}_{SB} = \mathrm{Concat}(3, \boldsymbol{X}^{(1)}_{SLB3}, \boldsymbol{X}^{(1)}_{SHB3}) \in \mathbb{R}^{4\times96\times72} \tag{2.60}$$

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，$\boldsymbol{X}^{(1)}_{SB3} \in \mathbb{R}^{4\times96\times72}$ 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 $Pad_S(\boldsymbol{X}^{(1)}_{SB3})$ で表し，出力される 3 階テンソルを $\hat{\boldsymbol{X}}_{St}$ で表す：

$$\hat{\boldsymbol{X}}_{St} = Pad_S(\boldsymbol{X}^{(1)}_{SB3}) \in \mathbb{R}^{4\times96\times84} \tag{2.61}$$

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

2.5 Percussion Block

Percussion Block は，全トラック共通特徴量 $\boldsymbol{x}_{BB} \in \mathbb{R}^{128}$ 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 $\boldsymbol{x}_{BB}$ から，

$$\mathbf{X}^{(0)}_{PB} = \mathrm{Reshape}_{PB}(W^{(0)}_{PB}\boldsymbol{x}_{BB} + \boldsymbol{b}_{PB}) \in \mathbb{R}^{128\times3\times1\times1} \tag{2.62}$$

を生成する．ここで，$W^{(0)}_{PB} \in \mathbb{R}^{384\times128}$ は矩形行列であり，$\boldsymbol{b}_{PB} \in \mathbb{R}^{384}$ はバイアスベクトルである．また，$\mathrm{Reshape}_{PB}$ は 384 次元ベクトルを $128\times3\times1\times1$ 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 $\mathbf{X}^{(0)}_{PB} \in \mathbb{R}^{128\times3\times1\times1}$ から

$$\mathbf{X}^{(1)}_{PB} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(0)}_{PB}, \mathcal{K}^{(1)}_{PB}, \boldsymbol{s}^{(1)}_{PB})) \in \mathbb{R}^{128\times4\times1\times1} \tag{2.63}$$

を生成する．ここで，$\mathcal{K}^{(1)}_{PB} \in \mathbb{R}^{128\times128\times2\times1\times1}$ であり，$\boldsymbol{s}^{(1)}_{PB} = (1,1,1)$ 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，$\mathbf{X}^{(1)}_{PB} \in \mathbb{R}^{128\times4\times1\times1}$ から

$$\mathbf{X}^{(2)}_{PB} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(1)}_{PB}, \mathcal{K}^{(2)}_{PB}, \boldsymbol{s}^{(2)}_{PB})) \in \mathbb{R}^{128\times4\times4\times1} \tag{2.64}$$

を生成する．ここで，$\mathcal{K}^{(2)}_{PB} \in \mathbb{R}^{128\times128\times1\times4\times1}$ であり，$\boldsymbol{s}^{(2)}_{PB} = (1,1,1)$ 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 $\mathbf{X}^{(2)}_{PB} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{X}^{(3)}_{PB} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}^{(2)}_{PB}, \mathcal{K}^{(3)}_{PB}, \boldsymbol{s}^{(3)}_{PB})) \in \mathbb{R}^{128\times4\times4\times2} \tag{2.65}$$

を生成する．ここで，$\mathcal{K}^{(3)}_{PB} \in \mathbb{R}^{128\times128\times1\times1\times2}$ であり，$\boldsymbol{s}^{(3)}_{PB} = (1,1,1)$ 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，$\mathbf{X}^{(3)}_{PB}$ を Percussion Block の出力 $\mathbf{X}_{PB}$ として出力する：$\mathbf{X}_{PB} = \mathbf{X}^{(3)}_{PB}$．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 2.20 に示す．

表 2.20: 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	$\mathcal{K}_{PB}^{(1)} \in \mathbb{R}^{128\times128\times2\times1\times1}$	$\boldsymbol{s}_{PB}^{(1)} = (1,1,1)$
第 2 層	$\mathcal{K}_{PB}^{(2)} \in \mathbb{R}^{128\times128\times1\times4\times1}$	$\boldsymbol{s}_{PB}^{(2)} = (1,1,1)$
第 3 層	$\mathcal{K}_{PB}^{(3)} \in \mathbb{R}^{128\times128\times1\times1\times2}$	$\boldsymbol{s}_{PB}^{(3)} = (1,1,1)$

2.6 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 2.21 のように 15 の音色をドラム打楽器として定める．

表 2.21: 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	ハイタム	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，

$$\mathbf{X}_{DB}^{(0)} = \mathrm{Slice}(\mathbf{X}_{PB}, 4, 1, 1) \in \mathbb{R}^{128\times4\times4\times1} \tag{2.66}$$

を抽出する．ここで，$\mathrm{Slice}(\mathbf{X}, m, n, l)$ 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 $\mathbf{X}_{DB}^{(0)} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{X}_{DB1}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{DB}^{(0)}, \mathcal{K}_{DB1}^{(1)}, \boldsymbol{s}_{DB1}^{(1)})) \in \mathbb{R}^{64\times4\times4\times15} \tag{2.67}$$

を生成する．ここで，$\mathcal{K}_{DB1}^{(1)} \in \mathbb{R}^{64\times128\times1\times1\times15}$ であり，$\boldsymbol{s}_{DB1}^{(1)} = (1, 1, 1)$ 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 $\mathbf{X}_{DB1}^{(1)} \in \mathbb{R}^{64\times4\times4\times15}$ から

$$\mathbf{X}_{DB1}^{(2)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{DB1}^{(1)}, \mathcal{K}_{DB1}^{(2)}, \boldsymbol{s}_{DB1}^{(2)})) \in \mathbb{R}^{32\times4\times16\times15} \tag{2.68}$$

を生成する．ここで，$\mathcal{K}_{DB1}^{(2)} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}_{DB1}^{(2)} = (1, 4, 1)$ 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 $\mathbf{X}_{DB1}^{(2)} \in \mathbb{R}^{32\times4\times16\times15}$ から

$$\mathbf{X}_{DB1}^{(3)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{DB1}^{(2)}, \mathcal{K}_{DB1}^{(3)}, \boldsymbol{s}_{DB1}^{(3)})) \in \mathbb{R}^{32\times4\times96\times15} \tag{2.69}$$

を生成する．ここで，$\mathcal{K}_{DB1}^{(3)} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}_{DB1}^{(3)} = (1, 6, 1)$ 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 2.22 に示す．

表 2.22: 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	$\mathcal{K}_{DB1}^{(1)} \in \mathbb{R}^{64\times128\times1\times1\times15}$	$\boldsymbol{s}_{DB1}^{(1)} = (1, 1, 1)$
第 2 層	$\mathcal{K}_{DB1}^{(2)} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}_{DB1}^{(2)} = (1, 4, 1)$
第 3 層	$\mathcal{K}_{DB1}^{(3)} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}_{DB1}^{(3)} = (1, 6, 1)$

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 $\mathbf{X}_{DB}^{(0)} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{X}_{DB2}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{DB}^{(0)}, \mathcal{K}_{DB2}^{(1)}, \boldsymbol{s}_{DB2}^{(1)})) \in \mathbb{R}^{64\times4\times16\times1} \tag{2.70}$$

を生成する．ここで，$\mathcal{K}_{DB2}^{(1)} \in \mathbb{R}^{64\times128\times1\times4\times1}$ であり，$\boldsymbol{s}_{DB2}^{(1)} = (1,4,1)$ 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 $\mathbf{X}_{DB2}^{(1)} \in \mathbb{R}^{64\times4\times16\times1}$ から

$$\mathbf{X}_{DB2}^{(2)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{DB2}^{(1)}, \mathcal{K}_{DB2}^{(2)}, \boldsymbol{s}_{DB2}^{(2)})) \in \mathbb{R}^{32\times4\times96\times1} \tag{2.71}$$

を生成する．ここで，$\mathcal{K}_{DB2}^{(2)} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}_{DB2}^{(2)} = (1,6,1)$ 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 $\mathbf{X}_{DB2}^{(2)} \in \mathbb{R}^{32\times4\times96\times1}$ から

$$\mathbf{X}_{DB2}^{(3)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{DB2}^{(2)}, \mathcal{K}_{DB2}^{(3)}, \boldsymbol{s}_{DB2}^{(3)})) \in \mathbb{R}^{32\times4\times96\times15} \tag{2.72}$$

を生成する．ここで，$\mathcal{K}_{DB2}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times15}$ であり，$\boldsymbol{s}_{DB2}^{(3)} = (1,1,1)$ 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 2.23 に示す．

表 2.23: 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	$\mathcal{K}_{DB2}^{(1)} \in \mathbb{R}^{64\times128\times1\times4\times1}$	$\boldsymbol{s}_{DB2}^{(1)} = (1,4,1)$
第 2 層	$\mathcal{K}_{DB2}^{(2)} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}_{DB2}^{(2)} = (1,6,1)$
第 3 層	$\mathcal{K}_{DB2}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times15}$	$\boldsymbol{s}_{DB2}^{(3)} = (1,1,1)$

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

$$\mathbf{X}_{DB3}^{(0)} = \mathrm{Concat}(1, \mathbf{X}_{DB1}^{(3)}, \mathbf{X}_{DB2}^{(3)}) \in \mathbb{R}^{64\times4\times96\times15} \tag{2.73}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}_{DB3}^{(0)} \in \mathbb{R}^{64\times4\times96\times15}$ から

$$\boldsymbol{X}_{DB3}^{(1)} = \tanh(\mathrm{TrConv}(\mathbf{X}_{DB3}^{(0)}, \mathcal{K}_{DB3}^{(1)}, \boldsymbol{s}_{DB3}^{(1)})) \in \mathbb{R}^{4\times96\times15} \tag{2.74}$$

を生成する．ここで，$\mathcal{K}_{DB3}^{(1)} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}_{DB3}^{(1)} = (1,1,1)$ 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 2.24 に示す．

表 2.24: 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	$\mathcal{K}_{DB3}^{(1)} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}_{DB3}^{(1)} = (1,1,1)$

2.7 Other Percussion Blocks

Other Percussion Block は，表 2.21 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，

$$\mathbf{X}_{OPB}^{(0)} = \mathrm{Slice}(\mathbf{X}_{PB}, 4, 2, 2) \in \mathbb{R}^{128\times4\times4\times1} \tag{2.75}$$

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 $\mathbf{X}_{OPB}^{(0)} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{X}_{OPB1}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{OPB}^{(0)}, \mathcal{K}_{OPB1}^{(1)}, \boldsymbol{s}_{OPB1}^{(1)})) \in \mathbb{R}^{64\times4\times4\times15} \tag{2.76}$$

を生成する．ここで，$\mathcal{K}_{OPB1}^{(1)} \in \mathbb{R}^{64\times128\times1\times1\times15}$ であり，$\boldsymbol{s}_{OPB1}^{(1)} = (1, 1, 1)$ 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 $\mathbf{X}_{OPB1}^{(1)} \in \mathbb{R}^{64\times4\times4\times15}$ から

$$\mathbf{X}_{OPB1}^{(2)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{OPB1}^{(1)}, \mathcal{K}_{OPB1}^{(2)}, \boldsymbol{s}_{OPB1}^{(2)})) \in \mathbb{R}^{32\times4\times16\times15} \tag{2.77}$$

を生成する．ここで，$\mathcal{K}_{OPB1}^{(2)} \in \mathbb{R}^{32\times64\times1\times4\times1}$ であり，$\boldsymbol{s}_{OPB1}^{(2)} = (1, 4, 1)$ 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 $\mathbf{X}_{OPB1}^{(2)} \in \mathbb{R}^{32\times4\times16\times15}$ から

$$\mathbf{X}_{OPB1}^{(3)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{OPB1}^{(2)}, \mathcal{K}_{OPB1}^{(3)}, \boldsymbol{s}_{OPB1}^{(3)})) \in \mathbb{R}^{32\times4\times96\times15} \tag{2.78}$$

を生成する．ここで，$\mathcal{K}_{OPB1}^{(3)} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{s}_{OPB1}^{(3)} = (1, 6, 1)$ 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 2.25 に示す．

表 2.25: 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	$\mathcal{K}_{OPB1}^{(1)} \in \mathbb{R}^{64\times64\times1\times1\times15}$	$\boldsymbol{s}_{OPB1}^{(1)} = (1, 1, 15)$
第 2 層	$\mathcal{K}_{OPB1}^{(2)} \in \mathbb{R}^{32\times64\times1\times4\times1}$	$\boldsymbol{s}_{OPB1}^{(2)} = (1, 4, 1)$
第 3 層	$\mathcal{K}_{OPB1}^{(3)} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{s}_{OPB1}^{(3)} = (1, 6, 1)$

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 $\mathbf{X}_{OPB}^{(0)} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{X}_{OPB2}^{(1)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{OPB}^{(0)}, \mathcal{K}_{OPB2}^{(1)}, \boldsymbol{s}_{OPB2}^{(1)})) \in \mathbb{R}^{64\times4\times16\times1} \tag{2.79}$$

を生成する．ここで，$\mathcal{K}_{OPB2}^{(1)} \in \mathbb{R}^{64\times128\times1\times4\times1}$ であり，$\boldsymbol{s}_{OPB2}^{(1)} = (1, 4, 1)$ 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 $\mathbf{X}_{OPB2}^{(1)} \in \mathbb{R}^{64\times4\times16\times1}$ から

$$\mathbf{X}_{OPB2}^{(2)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{OPB2}^{(1)}, \mathcal{K}_{OPB2}^{(2)}, \boldsymbol{s}_{OPB2}^{(2)})) \in \mathbb{R}^{32\times4\times96\times1} \tag{2.80}$$

を生成する．ここで，$\mathcal{K}_{OPB2}^{(2)} \in \mathbb{R}^{32\times64\times1\times6\times1}$ であり，$\boldsymbol{s}_{OPB2}^{(2)} = (1, 6, 1)$ 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 $\mathbf{X}_{OPB2}^{(2)} \in \mathbb{R}^{32\times4\times96\times1}$ から

$$\mathbf{X}_{OPB2}^{(3)} = \mathrm{LeakyReLU}(\mathrm{TrConv}(\mathbf{X}_{OPB2}^{(2)}, \mathcal{K}_{OPB2}^{(3)}, \boldsymbol{s}_{OPB2}^{(3)})) \in \mathbb{R}^{32\times4\times96\times15} \tag{2.81}$$

を生成する．ここで，$\mathcal{K}_{OPB2}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times15}$ であり，$\boldsymbol{s}_{OPB2}^{(3)} = (1, 1, 1)$ 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 2.26 に示す．

表 2.26: 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	$\mathcal{K}_{OPB2}^{(1)} \in \mathbb{R}^{64\times128\times1\times4\times1}$	$\boldsymbol{s}_{OPB2}^{(1)} = (1, 4, 1)$
第 2 層	$\mathcal{K}_{OPB2}^{(2)} \in \mathbb{R}^{32\times64\times1\times6\times1}$	$\boldsymbol{s}_{OPB2}^{(2)} = (1, 6, 1)$
第 3 層	$\mathcal{K}_{OPB2}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times15}$	$\boldsymbol{s}_{OPB2}^{(3)} = (1, 1, 1)$

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

$$\mathbf{X}_{OPB3}^{(0)} = \mathrm{Concat}(1, \mathbf{X}_{OPB1}^{(3)}, \mathbf{X}_{OPB2}^{(3)}) \in \mathbb{R}^{64\times4\times96\times15} \tag{2.82}$$

によって結合する．続いて，転置畳み込み層が $\mathbf{X}_{OPB3}^{(0)} \in \mathbb{R}^{64\times4\times96\times15}$ から

$$\boldsymbol{X}_{OPB3}^{(1)} = \tanh(\mathrm{TrConv}(\mathbf{X}_{OPB3}^{(0)}, \mathcal{K}_{OPB3}^{(1)}, \boldsymbol{s}_{OPB3}^{(1)})) \in \mathbb{R}^{4\times96\times15} \tag{2.83}$$

を生成する．ここで，$\mathcal{K}_{OPB3}^{(1)} \in \mathbb{R}^{1\times64\times1\times1\times1}$ であり，$\boldsymbol{s}_{OPB3}^{(1)} = (1, 1, 1)$ 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 2.27 に示す．

表 2.27: 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	$\mathcal{K}^{(1)}_{OPB3} \in \mathbb{R}^{1\times32\times1\times1\times1}$	$\boldsymbol{s}^{(1)}_{OPB3} = (1,1,1)$

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

$$\hat{\boldsymbol{X}}_{Pe} = \mathrm{Trans}(\boldsymbol{X}^{(1)}_{DB3}, \boldsymbol{X}^{(1)}_{OB3}) \in \mathbb{R}^{4\times96\times84} \tag{2.84}$$

を生成する．ここで，Trans は，$\boldsymbol{X}^{(1)}_{DB3} = (X^1_{DB3}, X^2_{DB3}, \ldots, X^{15}_{DB3})$ と $\boldsymbol{X}^{(1)}_{OPB3} = (X^1_{OPB3}, X^2_{OPB3}, \ldots, X^{15}_{OPB3})$ から打楽器トラック $\hat{\boldsymbol{X}}_{Pe} = (\hat{X}^1_{Percussion}, \hat{X}^2_{Percussion}, \ldots, \hat{X}^{84}_{Percussion})$ 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 2.28 が示す通りに $\boldsymbol{X}^{(1)}_{DB3}$ と $\boldsymbol{X}^{(1)}_{OPB3}$ を並び替える．なお，表 2.28 に示されていない $\hat{\boldsymbol{X}}_{Pe}$ 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，$\hat{\boldsymbol{X}}_{Pe}$ を出力として出力する．

表 2.28: 関数 $\mathrm{Trans}(\boldsymbol{X}^{(1)}_{DB3}, \boldsymbol{X}^{(1)}_{OB3})$ による打楽器変換

変換前	変換後
X^1_{DB3}	$X^{36}_{Percussion}$
X^2_{DB3}	$X^{37}_{Percussion}$
X^3_{DB3}	$X^{38}_{Percussion}$
X^4_{DB3}	$X^{41}_{Percussion}$
X^5_{DB3}	$X^{47}_{Percussion}$
X^6_{DB3}	$X^{50}_{Percussion}$
X^7_{DB3}	$X^{44}_{Percussion}$
X^8_{DB3}	$X^{42}_{Percussion}$
X^9_{DB3}	$X^{46}_{Percussion}$
X^{10}_{DB3}	$X^{49}_{Percussion}$
X^{11}_{DB3}	$X^{51}_{Percussion}$
X^{12}_{DB3}	$X^{52}_{Percussion}$
X^{13}_{DB3}	$X^{53}_{Percussion}$
X^{14}_{DB3}	$X^{55}_{Percussion}$
X^{15}_{DB3}	$X^{56}_{Percussion}$

変換前	変換後
X^1_{OPB3}	$X^{39}_{Percussion}$
X^2_{OPB3}	$X^{54}_{Percussion}$
X^3_{OPB3}	$X^{58}_{Percussion}$
X^4_{OPB3}	$X^{60}_{Percussion}$
X^5_{OPB3}	$X^{63}_{Percussion}$
X^6_{OPB3}	$X^{65}_{Percussion}$
X^7_{OPB3}	$X^{67}_{Percussion}$
X^8_{OPB3}	$X^{69}_{Percussion}$
X^9_{OPB3}	$X^{70}_{Percussion}$
X^{10}_{OPB3}	$X^{71}_{Percussion}$
X^{11}_{OPB3}	$X^{73}_{Percussion}$
X^{12}_{OPB3}	$X^{75}_{Percussion}$
X^{13}_{OPB3}	$X^{76}_{Percussion}$
X^{14}_{OPB3}	$X^{79}_{Percussion}$
X^{15}_{OPB3}	$X^{81}_{Percussion}$

また，各トラックの生成ピアノロール，$\hat{\boldsymbol{X}}_{Pi} \in \mathbb{R}^{4\times96\times30}$ ，$\hat{\boldsymbol{X}}_{Gu} \in \mathbb{R}^{4\times96\times84}$ ，$\hat{\boldsymbol{X}}_{Ba} \in \mathbb{R}^{4\times96\times84}$ ，$\hat{\boldsymbol{X}}_{St} \in \mathbb{R}^{4\times96\times84}$ ，$\hat{\boldsymbol{X}}_{Pe} \in \mathbb{R}^{4\times96\times84}$ を

$$\hat{\boldsymbol{X}}_{TP} = (\hat{\boldsymbol{X}}_{Pi}, \hat{\boldsymbol{X}}_{Gu}, \hat{\boldsymbol{X}}_{Ba}, \hat{\boldsymbol{X}}_{St}, \hat{\boldsymbol{X}}_{Pe}) \tag{2.85}$$

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．

2.8 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．

第3章　判別器の構成ブロック

3.1　はじめに

この章では，判別器を構成する各種ブロックの詳細を説明する．各有音程楽器ピアノロールの特徴抽出を行う Piano Block, Guitar Block, Bass Block, Strings Block を説明する．続いて，有音程楽器間に共通する特徴を抽出する Low-band Block，Mid-band Block，High-band Block，Tonal Block を説明する．さらに，打楽器ピアノロールの特徴抽出を行う Percussion Block, Drums Block, Other Percussion Block を説明する．また，有音程楽器の補助的な特徴抽出を行う Chroma Block，Piano Polyphonicity Block, Guitar Polyphonicity Block, Bass Polyphonicity Block, Strings Polyphonicity Block と各 Block の出力を統合し特徴抽出を行う Merged Block 1，Merged Block 2，Merged Block 3 について説明する．

判別器への入力はマルチトラックピアノロール $\boldsymbol{X} = (\boldsymbol{X}_{Pi}, \boldsymbol{X}_{Gu}, \boldsymbol{X}_{Base}, \boldsymbol{X}_{St}, \boldsymbol{X}_{Pe})$ である．Conv 関数での計算のために各ピアノロールを 4 階テンソルへの変換を行う．まず，ピアノトラック $\boldsymbol{X}_{Pi} \in \mathbb{R}^{4\times96\times84}$ を

$$\mathbf{X}_{Piano} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Pi}) \in \mathbb{R}^{1\times4\times96\times84} \tag{3.1}$$

へと変換する．ここで，$\mathrm{Reshape}_{expand}$ は 入力の 4×96×84 の三階テンソルを 1×4×96×84 の 4 階テンソルへと拡張する関数である．同様に，ギタートラック $\boldsymbol{X}_{Gu} \in \mathbb{R}^{4\times96\times84}$ を

$$\mathbf{X}_{Guitar} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Gu}) \in \mathbb{R}^{1\times4\times96\times84} \tag{3.2}$$

へと変換する．ベーストラック $\boldsymbol{X}_{Ba} \in \mathbb{R}^{4\times96\times84}$ を

$$\mathbf{X}_{Bass} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Ba}) \in \mathbb{R}^{1\times4\times96\times84} \tag{3.3}$$

へと変換する．ストリングストラック $\boldsymbol{X}_{St} \in \mathbb{R}^{4\times96\times84}$ を

$$\mathbf{X}_{Strings} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{St}) \in \mathbb{R}^{1\times4\times96\times84} \tag{3.4}$$

へと変換する．打楽器トラック $\boldsymbol{X}_{Pe} \in \mathbb{R}^{4\times96\times84}$ を

$$\mathbf{X}_{Percussion} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Pe}) \in \mathbb{R}^{1\times4\times96\times84} \tag{3.5}$$

へと変換する．

3.2 Individual Blocks

3.2.1 Piano Block

従来手法である BinaryMuseGAN における判別器の Piano Block は，3 個のサブブロックで構成されていた．本研究では Piano Block を，6 個のサブブロックで構成する．それらを Piano Block 1，Piano Block 2，Piano Block 3，Piano Block 4，Piano Block 5，および Piano Block 6 と呼ぶことにする．順に説明する．

Piano Block 1 は，表 3.1 に示す 2 層の畳み込み層で構成される．第 1 畳み込み層は $\mathbf{X}_{Piano} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{PiB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB1}, \boldsymbol{t}^{(1)}_{PiB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{3.6}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{PiB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$ であり，$\boldsymbol{t}^{(1)}_{PiB1} = (1, 1, 12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{PiB1} \in \mathbb{R}^{32\times4\times96\times7}$ から

$$\mathbf{Y}^{(2)}_{PiB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB1}, \mathcal{L}^{(2)}_{PiB1}, \boldsymbol{t}^{(2)}_{PiB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.7}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{PiB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{PiB1} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．以上のように，Piano Block 1 では，第 4 モードの特徴抽出を行ってから第 3 モードの特徴抽出を行う．すなわち，音高方向の特徴抽出の後に時間方向の特徴抽出を行う．

表 3.1: Piano Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{PiB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$	$\boldsymbol{t}^{(1)}_{PiB1} = (1, 1, 12)$
第 2 層	$\mathcal{L}^{(2)}_{PiB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{PiB1} = (1, 6, 1)$

一方，Piano Block 2 は同じサイズの特徴量テンソルを逆の手順で抽出する．すなわち，第 3 モードの時間方向特徴量を抽出してから第 4 モードである音高方向の特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Piano} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{PiB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB2}, \boldsymbol{t}^{(1)}_{PiB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.8}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{PiB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$ であり，$\boldsymbol{t}^{(1)}_{PiB2} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{PiB2} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(2)}_{PiB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB2}, \mathcal{L}^{(2)}_{PiB2}, \boldsymbol{t}^{(2)}_{PiB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.9}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{PiB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(2)}_{PiB2} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Piano Block 2 の畳み込み層における各種パラメータを表 3.2 にまとめて示す．

表 3.2: Piano Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{PiB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$	$\boldsymbol{t}^{(1)}_{PiB2} = (1,6,1)$
第 2 層	$\mathcal{L}^{(2)}_{PiB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$	$\boldsymbol{t}^{(2)}_{PiB2} = (1,1,12)$

続いて，Piano Block 3 は表 3.3 に示す 3 層の畳み込み層で構成される．まず，第 1 畳み込み層が $\mathbf{X}_{Piano} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{PiB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB3}, \boldsymbol{t}^{(1)}_{PiB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.10}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{PiB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$，$\boldsymbol{t}^{(1)}_{PiB3} = (1,1,1)$ である．この段階で第 4 モードのサイズが 3 のカーネルを用いることにより，隣接 3 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 3 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{PiB3} \in \mathbb{R}^{\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{PiB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB3}, \mathcal{L}^{(2)}_{PiB3}, \boldsymbol{t}^{(2)}_{PiB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.11}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{PiB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{PiB3} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{PiB3} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{PiB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiB3}, \mathcal{L}^{(3)}_{PiB3}, \boldsymbol{t}^{(3)}_{PiB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.12}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{PiB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{PiB3} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Piano Block 3 の畳み込み層における各種パラメータを表 3.3 にまとめて示す．

表 3.3: Piano Block 3 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}^{(1)}_{PiB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$	$\boldsymbol{t}^{(1)}_{PiB3} = (1,1,1)$
第 2 層	$\mathcal{L}^{(2)}_{PiB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{PiB3} = (1,6,1)$
第 3 層	$\mathcal{L}^{(3)}_{PiB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}^{(3)}_{PiB3} = (1,1,12)$

続いて，Piano Block 4 は表 3.3 に示す 3 層の畳み込み層で構成される．Piano Block 3 では第一層で隣接 3 半音から特徴を抽出したが，Piano Block 3 では隣接 2 半音から特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Piano} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{PiB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB4}, \boldsymbol{t}^{(1)}_{PiB4})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.13}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{PiB4} \in \mathbb{R}^{32\times1\times1\times1\times2}$，$\boldsymbol{t}^{(1)}_{PiB4} = (1,1,1)$ である．この段階で第 4 モードのサイズが 2 のカーネルを用いることにより，隣接 2 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 2 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{PiB4} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{PiB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB4}, \mathcal{L}^{(2)}_{PiB4}, \boldsymbol{t}^{(2)}_{PiB4})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.14}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{PiB4} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{PiB4} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{PiB4} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{PiB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiB4}, \mathcal{L}^{(3)}_{PiB4}, \boldsymbol{t}^{(3)}_{PiB4})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.15}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{PiB4} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{PiB4} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Piano Block 4 の畳み込み層における各種パラメータを表 3.4 にまとめて示す．

表 3.4: Piano Block 4 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}^{(1)}_{PiB4} \in \mathbb{R}^{32\times1\times1\times1\times2}$	$\boldsymbol{t}^{(1)}_{PiB4} = (1,1,1)$
第 2 層	$\mathcal{L}^{(2)}_{PiB4} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{PiB4} = (1,6,1)$
第 3 層	$\mathcal{L}^{(3)}_{PiB4} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}^{(3)}_{PiB4} = (1,1,12)$

Piano Block 5 では，Piano Block 3 と Piano Block 4 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(3)}_{PiB3}$，$\mathbf{Y}^{(3)}_{PiB4}$ を

$$\mathbf{Y}^{(0)}_{PiB5} = \mathrm{Concat}(1, \mathbf{Y}^{(3)}_{PiB3}, \mathbf{Y}^{(3)}_{PiB4}) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.16}$$

によって結合する．$\mathbf{Y}^{(3)}_{PiB3}$ と $\mathbf{Y}^{(3)}_{PiB4}$ の 2 テンソルを結合することにより，二種類の不協和音検出の結果を統合している．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{PiB5} \in \mathbb{R}^{64\times4\times16\times7}$ から

$$\mathbf{Y}^{(1)}_{PiB5} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{PiB5}, \mathcal{L}^{(1)}_{PiB5}, \boldsymbol{t}^{(1)}_{PiB5})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.17}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{PiB5} \in \mathbb{R}^{32\times64\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{PiB5} = (1,1,1)$ である．Piano Block 5 の畳み込み層における各種パラメータを表 3.5 にまとめて示す．

表 3.5: Piano Block 5 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{PiB5} \in \mathbb{R}^{32\times64\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{PiB5} = (1,1,1)$

Piano Block 6 では，Piano Block 1，Piano Block 2，Piano Block 5 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(2)}_{PiB1}$，$\mathbf{Y}^{(2)}_{PiB2}$，$\mathbf{Y}^{(1)}_{PiB5}$ を

$$\mathbf{Y}^{(0)}_{PiB6} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{PiB1}, \mathbf{Y}^{(2)}_{PiB2}, \mathbf{Y}^{(1)}_{PiB5}) \in \mathbb{R}^{160\times4\times16\times7} \tag{3.18}$$

によって結合する．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{PiB6} \in \mathbb{R}^{160\times4\times16\times7}$ から

$$\mathbf{Y}^{(1)}_{PiB6} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{PiB6}, \mathcal{L}^{(1)}_{PiB6}, \boldsymbol{t}^{(1)}_{PiB6})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.19}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{PiB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{PiB6} = (1,1,1)$ である．Piano Block 6 の畳み込み層における各種パラメータを表 3.6 にまとめて示す．Piano Block は処理をこれで終了し，$\mathbf{Y}^{(1)}_{PiB6}$ を Piano Block の出力 $\mathbf{Y}_{PiB}$ として出力する：$\mathbf{Y}_{PiB} = \mathbf{Y}^{(1)}_{PiB6}$．これが，ピアノトラックの特徴量であり，Band Blocks に入力される．

表 3.6: Piano Block 6 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{PiB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{PiB6} = (1,1,1)$

3.2.2 Guitar Block

従来手法である BinaryMuseGAN における判別器の Guitar Block は，3 個のサブブロックで構成されていた．本研究では Guitar Block を Piano Block と同様に 6 個のサブブロックで構成する．それらを Guitar Block 1，Guitar Block 2，Guitar Block 3，Guitar Block 4，Guitar Block 5，および Guitar Block 6 と呼ぶことにする．順に説明する．

Guitar Block 1 は，表 3.7 に示す 2 層の畳み込み層で構成される．第 1 畳み込み層は $\mathbf{X}_{Guitar} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB1}, \boldsymbol{t}^{(1)}_{GB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{3.20}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{GB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$ であり，$\boldsymbol{t}^{(1)}_{GB1} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{GB1} \in \mathbb{R}^{32\times4\times96\times7}$ から

$$\mathbf{Y}^{(2)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB1}, \mathcal{L}^{(2)}_{GB1}, \boldsymbol{t}^{(2)}_{GB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.21}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{GB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{GB1} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．以上のように，Guitar Block 1 では，第 4 モードの特徴抽出を行ってから第 3 モードの特徴抽出を行う．すなわち，音高方向の特徴抽出の後に時間方向の特徴抽出を行う．

表 3.7: Guitar Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{GB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$	$\boldsymbol{t}^{(1)}_{GB1} = (1,1,12)$
第 2 層	$\mathcal{L}^{(2)}_{GB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{GB1} = (1,6,1)$

一方，Guitar Block 2 は同じサイズの特徴量テンソルを逆の手順で抽出する．すなわち，第 3 モードの時間方向特徴量を抽出してから第 4 モードである音高方向の特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Guitar} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB2}, \boldsymbol{t}^{(1)}_{GB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.22}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{GB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$ であり，$\boldsymbol{t}^{(1)}_{GB2} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{GB2} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(2)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB2}, \mathcal{L}^{(2)}_{GB2}, \boldsymbol{t}^{(2)}_{GB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.23}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{GB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(2)}_{GB2} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Guitar Block 2 の畳み込み層における各種パラメータを表 3.8 にまとめて示す．

表 3.8: Guitar Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{GB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$	$\boldsymbol{t}^{(1)}_{GB2} = (1,6,1)$
第 2 層	$\mathcal{L}^{(2)}_{GB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$	$\boldsymbol{t}^{(2)}_{GB2} = (1,1,12)$

続いて，Guitar Block 3 は表 3.9 に示す 3 層の畳み込み層で構成される．まず，第 1 畳み込み層が $\mathbf{X}_{Guitar} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{GB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB3}, \boldsymbol{t}^{(1)}_{GB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.24}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{GB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$，$\boldsymbol{t}^{(1)}_{GB3} = (1,1,1)$ である．この段階で第 4 モードのサイズが 3 のカーネルを用いることにより，隣接 3 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 3 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{GB3} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{GB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB3}, \mathcal{L}^{(2)}_{GB3}, \boldsymbol{t}^{(2)}_{GB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.25}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{GB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{GB3} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{GB3} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{GB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{GB3}, \mathcal{L}^{(3)}_{GB3}, \boldsymbol{t}^{(3)}_{GB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.26}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{GB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{GB3} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Guitar Block 3 の畳み込み層における各種パラメータを表 3.9 にまとめて示す.

表 3.9: Guitar Block 3 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}^{(1)}_{GB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$	$\boldsymbol{t}^{(1)}_{GB3} = (1,1,1)$
第 2 層	$\mathcal{L}^{(2)}_{GB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{GB3} = (1,6,1)$
第 3 層	$\mathcal{L}^{(3)}_{GB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}^{(3)}_{GB3} = (1,1,12)$

続いて，Guitar Block 4 は表 3.9 に示す 3 層の畳み込み層で構成される．Guitar Block 3 では第一層で隣接 3 半音から特徴を抽出したが，Guitar Block 3 では隣接 2 半音から特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Guitar} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{GB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB4}, \boldsymbol{t}^{(1)}_{GB4})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.27}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{GB4} \in \mathbb{R}^{32\times1\times1\times1\times2}$，$\boldsymbol{t}^{(1)}_{GB4} = (1,1,1)$ である．この段階で第 4 モードのサイズが 2 のカーネルを用いることにより，隣接 2 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 2 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{GB4} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{GB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB4}, \mathcal{L}^{(2)}_{GB4}, \boldsymbol{t}^{(2)}_{GB4})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.28}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{GB4} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{GB4} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{GB4} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{GB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{GB4}, \mathcal{L}^{(3)}_{GB4}, \boldsymbol{t}^{(3)}_{GB4})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.29}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{GB4} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{GB4} = (1, 1, 12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Guitar Block 4 の畳み込み層における各種パラメータを表 3.10 にまとめて示す．

表 3.10: Guitar Block 4 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}^{(1)}_{GB4} \in \mathbb{R}^{32\times1\times1\times1\times2}$	$\boldsymbol{t}^{(1)}_{GB4} = (1, 1, 1)$
第 2 層	$\mathcal{L}^{(2)}_{GB4} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{GB4} = (1, 6, 1)$
第 3 層	$\mathcal{L}^{(3)}_{GB4} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}^{(3)}_{GB4} = (1, 1, 12)$

Guitar Block 5 では，Guitar Block 3 と Guitar Block 4 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(3)}_{GB3}$，$\mathbf{Y}^{(3)}_{GB4}$ を

$$\mathbf{Y}^{(0)}_{GB5} = \mathrm{Concat}(1, \mathbf{Y}^{(3)}_{GB3}, \mathbf{Y}^{(3)}_{GB4}) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.30}$$

によって結合する．$\mathbf{Y}^{(3)}_{GB3}$ と $\mathbf{Y}^{(3)}_{GB4}$ の 2 テンソルを結合することにより，二種類の不協和音検出の結果を統合している．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{GB5} \in \mathbb{R}^{64\times4\times16\times7}$ から

$$\mathbf{Y}^{(1)}_{GB5} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{GB5}, \mathcal{L}^{(1)}_{GB5}, \boldsymbol{t}^{(1)}_{GB5})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.31}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{GB5} \in \mathbb{R}^{32\times64\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{GB5} = (1, 1, 1)$ である．Guitar Block 5 の畳み込み層における各種パラメータを表 3.11 にまとめて示す．

表 3.11: Guitar Block 5 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{GB5} \in \mathbb{R}^{32\times64\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{GB5} = (1, 1, 1)$

Guitar Block 6 では，Guitar Block 1，Guitar Block 2，Guitar Block 5 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(2)}_{GB1}$，$\mathbf{Y}^{(2)}_{GB2}$，$\mathbf{Y}^{(1)}_{GB5}$ を

$$\mathbf{Y}^{(0)}_{GB6} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{GB1}, \mathbf{Y}^{(2)}_{GB2}, \mathbf{Y}^{(1)}_{GB5}) \in \mathbb{R}^{160\times4\times16\times7} \tag{3.32}$$

によって結合する．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{GB6} \in \mathbb{R}^{160\times4\times16\times7}$ から

$$\mathbf{Y}^{(1)}_{GB6} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{GB6}, \mathcal{L}^{(1)}_{GB6}, \boldsymbol{t}^{(1)}_{GB6})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.33}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{GB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{GB6} = (1,1,1)$ である．Guitar Block 6 の畳み込み層における各種パラメータを表 3.12 にまとめて示す．Guitar Block は処理をこれで終了し，$\mathbf{Y}^{(1)}_{GB6}$ を Guitar Block の出力 $\mathbf{Y}_{GB}$ として出力する：$\mathbf{Y}_{GB} = \mathbf{Y}^{(1)}_{GB6}$. これが，ギタートラックの特徴量であり，Band Blocks に入力される．

表 3.12: Guitar Block 6 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{GB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{GB6} = (1,1,1)$

3.2.3 Bass Block

従来手法である BinaryMuseGAN における判別器の Bass Block は，3 個のサブブロックで構成されていた．本研究では Bass Block を Piano Block と同様に 6 個のサブブロックで構成する．それらを Bass Block 1，Bass Block 2，Bass Block 3，Bass Block 4，Bass Block 5，および Bass Block 6 と呼ぶことにする．順に説明する．

Bass Block 1 は，表 3.13 に示す 2 層の畳み込み層で構成される．第 1 畳み込み層は $\mathbf{X}_{Bass} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB1}, \boldsymbol{t}^{(1)}_{BaB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{3.34}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{BaB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$ であり，$\boldsymbol{t}^{(1)}_{BaB1} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{BaB1} \in \mathbb{R}^{32\times4\times96\times7}$ から

$$\mathbf{Y}^{(2)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB1}, \mathcal{L}^{(2)}_{BaB1}, \boldsymbol{t}^{(2)}_{BaB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.35}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{BaB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{BaB1} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．以上のように，Bass Block 1 では，第 4 モードの特徴抽出を行ってから第 3 モードの特徴抽出を行う．すなわち，音高方向の特徴抽出の後に時間方向の特徴抽出を行う．

表 3.13: Bass Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{BaB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$	$\boldsymbol{t}^{(1)}_{BaB1} = (1, 1, 12)$
第 2 層	$\mathcal{L}^{(2)}_{BaB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{BaB1} = (1, 6, 1)$

一方，Bass Block 2 は同じサイズの特徴量テンソルを逆の手順で抽出する．すなわち，第 3 モードの時間方向特徴量を抽出してから第 4 モードである音高方向の特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Bass} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{BaB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB2}, \boldsymbol{t}^{(1)}_{BaB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.36}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{BaB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$ であり，$\boldsymbol{t}^{(1)}_{BaB2} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{BaB2} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(2)}_{BaB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB2}, \mathcal{L}^{(2)}_{BaB2}, \boldsymbol{t}^{(2)}_{BaB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.37}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{BaB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(2)}_{BaB2} = (1, 1, 12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Bass Block 2 の畳み込み層における各種パラメータを表 3.14 にまとめて示す．

表 3.14: Bass Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{BaB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$	$\boldsymbol{t}^{(1)}_{BaB2} = (1, 6, 1)$
第 2 層	$\mathcal{L}^{(2)}_{BaB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$	$\boldsymbol{t}^{(2)}_{BaB2} = (1, 1, 12)$

続いて，Bass Block 3 は表 3.15 に示す 3 層の畳み込み層で構成される．まず，第 1 畳み込み層が $\mathbf{X}_{Bass} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{BaB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB3}, \boldsymbol{t}^{(1)}_{BaB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.38}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{BaB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$，$\boldsymbol{t}^{(1)}_{BaB3} = (1, 1, 1)$ である．この段階で第 4 モードのサイズが 3 のカーネルを用いることにより，隣接 3 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 3 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{BaB3} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{BaB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB3}, \mathcal{L}^{(2)}_{BaB3}, \boldsymbol{t}^{(2)}_{BaB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.39}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{BaB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{BaB3} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{BaB3} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{BaB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{BaB3}, \mathcal{L}^{(3)}_{BaB3}, \boldsymbol{t}^{(3)}_{BaB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.40}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{BaB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{BaB3} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Bass Block 3 の畳み込み層における各種パラメータを表 3.15 にまとめて示す．

表 3.15: Bass Block 3 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}^{(1)}_{BaB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$	$\boldsymbol{t}^{(1)}_{BaB3} = (1,1,1)$
第 2 層	$\mathcal{L}^{(2)}_{BaB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{BaB3} = (1,6,1)$
第 3 層	$\mathcal{L}^{(3)}_{BaB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}^{(3)}_{BaB3} = (1,1,12)$

続いて，Bass Block 4 は表 3.15 に示す 3 層の畳み込み層で構成される．Bass Block 3 では第一層で隣接 3 半音から特徴を抽出したが，Bass Block 3 では隣接 2 半音から特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Bass} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{BaB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB4}, \boldsymbol{t}^{(1)}_{BaB4})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.41}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{BaB4} \in \mathbb{R}^{32\times1\times1\times1\times2}$，$\boldsymbol{t}^{(1)}_{BaB4} = (1,1,1)$ である．この段階で第 4 モードのサイズが 2 のカーネルを用いることにより，隣接 2 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 2 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{BaB4} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{BaB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB4}, \mathcal{L}^{(2)}_{BaB4}, \boldsymbol{t}^{(2)}_{BaB4})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.42}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{BaB4} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{BaB4} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{BaB4} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{BaB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{BaB4}, \mathcal{L}^{(3)}_{BaB4}, \boldsymbol{t}^{(3)}_{BaB4})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.43}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{BaB4} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{BaB4} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Bass Block 4 の畳み込み層における各種パラメータを表 3.16 にまとめて示す．

表 3.16: Bass Block 4 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}^{(1)}_{BaB4} \in \mathbb{R}^{32\times1\times1\times1\times2}$	$\boldsymbol{t}^{(1)}_{BaB4} = (1,1,1)$
第 2 層	$\mathcal{L}^{(2)}_{BaB4} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{BaB4} = (1,6,1)$
第 3 層	$\mathcal{L}^{(3)}_{BaB4} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}^{(3)}_{BaB4} = (1,1,12)$

Bass Block 5 では，Bass Block 3 と Bass Block 4 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(3)}_{BaB3}$，$\mathbf{Y}^{(3)}_{BaB4}$ を

$$\mathbf{Y}^{(0)}_{BaB5} = \mathrm{Concat}(1, \mathbf{Y}^{(3)}_{BaB3}, \mathbf{Y}^{(3)}_{BaB4}) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.44}$$

によって結合する．$\mathbf{Y}^{(3)}_{BaB3}$ と $\mathbf{Y}^{(3)}_{BaB4}$ の 2 テンソルを結合することにより，二種類の不協和音検出の結果を統合している．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{BaB5} \in \mathbb{R}^{64\times4\times16\times7}$ から

$$\mathbf{Y}^{(1)}_{BaB5} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{BaB5}, \mathcal{L}^{(1)}_{BaB5}, \boldsymbol{t}^{(1)}_{BaB5})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.45}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{BaB5} \in \mathbb{R}^{32\times64\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{BaB5} = (1,1,1)$ である．Bass Block 5 の畳み込み層における各種パラメータを表 3.17 にまとめて示す.

表 3.17: Bass Block 5 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{BaB5} \in \mathbb{R}^{32\times64\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{BaB5} = (1,1,1)$

Bass Block 6 では，Bass Block 1，Bass Block 2，Bass Block 5 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(2)}_{BaB1}$，$\mathbf{Y}^{(2)}_{BaB2}$，$\mathbf{Y}^{(1)}_{BaB5}$ を

$$\mathbf{Y}^{(0)}_{BaB6} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{BaB1}, \mathbf{Y}^{(2)}_{BaB2}, \mathbf{Y}^{(1)}_{BaB5}) \in \mathbb{R}^{160\times4\times16\times7} \tag{3.46}$$

によって結合する．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{BaB6} \in \mathbb{R}^{160\times4\times16\times7}$ から

$$\mathbf{Y}^{(1)}_{BaB6} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{BaB6}, \mathcal{L}^{(1)}_{BaB6}, \boldsymbol{t}^{(1)}_{BaB6})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.47}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{BaB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{BaB6} = (1,1,1)$ である．Bass Block 6 の畳み込み層における各種パラメータを表 3.18 にまとめて示す．Bass Block は処理をこれで終了し，$\mathbf{Y}^{(1)}_{BaB6}$ を Bass Block の出力 $\mathbf{Y}_{BaB}$ として出力する：$\mathbf{Y}_{BaB} = \mathbf{Y}^{(1)}_{BaB6}$. これが，ベーストラックの特徴量であり，Band Blocks に入力される．

表 3.18: Bass Block 6 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{BaB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{BaB6} = (1,1,1)$

3.2.4 Strings Block

従来手法である BinaryMuseGAN における判別器の Strings Block は，3 個のサブブロックで構成されていた．本研究では Strings Block を Piano Block と同様に 6 個のサブブロックで構成する．それらを Strings Block 1，Strings Block 2，Strings Block 3，Strings Block 4，Strings Block 5，および Strings Block 6 と呼ぶことにする．順に説明する．

Strings Block 1 は，表 3.19 に示す 2 層の畳み込み層で構成される．第 1 畳み込み層は $\mathbf{X}_{Strings} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{SB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}^{(1)}_{SB1}, \boldsymbol{t}^{(1)}_{SB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{3.48}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{SB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$ であり，$\boldsymbol{t}^{(1)}_{SB1} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{SB1} \in \mathbb{R}^{32\times4\times96\times7}$ から

$$\mathbf{Y}^{(2)}_{SB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SB1}, \mathcal{L}^{(2)}_{SB1}, \boldsymbol{t}^{(2)}_{SB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.49}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{SB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{SB1} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．以上のように，Strings Block 1 では，第 4 モードの特徴抽出を行ってから第 3 モードの特徴抽出を行う．すなわち，音高方向の特徴抽出の後に時間方向の特徴抽出を行う．

表 3.19: Strings Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{SB1} \in \mathbb{R}^{32\times1\times1\times1\times12}$	$\boldsymbol{t}^{(1)}_{SB1} = (1,1,12)$
第 2 層	$\mathcal{L}^{(2)}_{SB1} \in \mathbb{R}^{64\times32\times1\times6\times1}$	$\boldsymbol{t}^{(2)}_{SB1} = (1,6,1)$

一方，Strings Block 2 は同じサイズの特徴量テンソルを逆の手順で抽出する．すなわち，第 3 モードの時間方向特徴量を抽出してから第 4 モードである音高方向の

特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Strings} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{SB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}^{(1)}_{SB2}, \boldsymbol{t}^{(1)}_{SB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.50}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{SB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$ であり，$\boldsymbol{t}^{(1)}_{SB2} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{SB2} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(2)}_{SB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SB2}, \mathcal{L}^{(2)}_{SB2}, \boldsymbol{t}^{(2)}_{SB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.51}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{SB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(2)}_{SB2} = (1, 1, 12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Strings Block 2 の畳み込み層における各種パラメータを表 3.20 にまとめて示す.

表 3.20: Strings Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{SB2} \in \mathbb{R}^{32\times1\times1\times6\times1}$	$\boldsymbol{t}^{(1)}_{SB2} = (1, 6, 1)$
第 2 層	$\mathcal{L}^{(2)}_{SB2} \in \mathbb{R}^{64\times32\times1\times1\times12}$	$\boldsymbol{t}^{(2)}_{SB2} = (1, 1, 12)$

続いて，Strings Block 3 は表 3.21 に示す 3 層の畳み込み層で構成される．まず，第 1 畳み込み層が $\mathbf{X}_{Strings} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}^{(1)}_{SB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}^{(1)}_{SB3}, \boldsymbol{t}^{(1)}_{SB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.52}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}^{(1)}_{SB3} \in \mathbb{R}^{32\times1\times1\times1\times3}$，$\boldsymbol{t}^{(1)}_{SB3} = (1, 1, 1)$ である．この段階で第 4 モードのサイズが 3 のカーネルを用いることにより，隣接 3 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 3 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{SB3} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}^{(2)}_{SB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SB3}, \mathcal{L}^{(2)}_{SB3}, \boldsymbol{t}^{(2)}_{SB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.53}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{SB3} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}^{(2)}_{SB3} = (1, 6, 1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{SB3} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}^{(3)}_{SB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{SB3}, \mathcal{L}^{(3)}_{SB3}, \boldsymbol{t}^{(3)}_{SB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.54}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{SB3} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}^{(3)}_{SB3} = (1, 1, 12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Strings Block 3 の畳み込み層における各種パラメータを表 3.21 にまとめて示す.

表 3.21: Strings Block 3 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}_{SB3}^{(1)} \in \mathbb{R}^{32\times1\times1\times1\times3}$	$\boldsymbol{t}_{SB3}^{(1)} = (1,1,1)$
第 2 層	$\mathcal{L}_{SB3}^{(2)} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}_{SB3}^{(2)} = (1,6,1)$
第 3 層	$\mathcal{L}_{SB3}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}_{SB3}^{(3)} = (1,1,12)$

続いて，Strings Block 4 は表 3.21 に示す 3 層の畳み込み層で構成される．Strings Block 3 では第一層で隣接 3 半音から特徴を抽出したが，Strings Block 3 では隣接 2 半音から特徴抽出を行う．まず，第 1 畳み込み層が $\mathbf{X}_{Strings} \in \mathbb{R}^{1\times4\times96\times84}$ から

$$\mathbf{Y}_{SB4}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}_{SB4}^{(1)}, \boldsymbol{t}_{SB4}^{(1)})) \in \mathbb{R}^{32\times4\times96\times84} \tag{3.55}$$

を抽出する．ここで，畳み込みは Same 畳み込みであり，$\mathcal{L}_{SB4}^{(1)} \in \mathbb{R}^{32\times1\times1\times1\times2}$，$\boldsymbol{t}_{SB4}^{(1)} = (1,1,1)$ である．この段階で第 4 モードのサイズが 2 のカーネルを用いることにより，隣接 2 半音の特徴抽出を開始する．この第 1 畳み込み層は，隣接 2 半音からの不協和音検出を目的としている．次に，第 2 畳み込み層は $\mathbf{Y}_{SB4}^{(1)} \in \mathbb{R}^{32\times4\times96\times84}$ から

$$\mathbf{Y}_{SB4}^{(2)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB4}^{(1)}, \mathcal{L}_{SB4}^{(2)}, \boldsymbol{t}_{SB4}^{(2)})) \in \mathbb{R}^{32\times4\times16\times84} \tag{3.56}$$

を抽出する．ここで，$\mathcal{L}_{SB4}^{(2)} \in \mathbb{R}^{32\times32\times1\times6\times1}$ であり，$\boldsymbol{t}_{SB4}^{(2)} = (1,6,1)$ である．この段階で第 3 モードのサイズが 16 になり，微細なリズム構造の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}_{SB4}^{(2)} \in \mathbb{R}^{32\times4\times16\times84}$ から

$$\mathbf{Y}_{SB4}^{(3)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB4}^{(2)}, \mathcal{L}_{SB4}^{(3)}, \boldsymbol{t}_{SB4}^{(3)})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.57}$$

を抽出する．ここで，$\mathcal{L}_{SB4}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times12}$ であり，$\boldsymbol{t}_{SB4}^{(3)} = (1,1,12)$ である．この段階で第 4 モードのサイズが 7 になり，オクターブごとの特徴抽出を開始する．Strings Block 4 の畳み込み層における各種パラメータを表 3.22 にまとめて示す.

表 3.22: Strings Block 4 の畳み込み層における各種パラメータ

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 (Same)	$\mathcal{L}_{SB4}^{(1)} \in \mathbb{R}^{32\times1\times1\times1\times2}$	$\boldsymbol{t}_{SB4}^{(1)} = (1,1,1)$
第 2 層	$\mathcal{L}_{SB4}^{(2)} \in \mathbb{R}^{32\times32\times1\times6\times1}$	$\boldsymbol{t}_{SB4}^{(2)} = (1,6,1)$
第 3 層	$\mathcal{L}_{SB4}^{(3)} \in \mathbb{R}^{32\times32\times1\times1\times12}$	$\boldsymbol{t}_{SB4}^{(3)} = (1,1,12)$

Strings Block 5 では，Strings Block 3 と Strings Block 4 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}_{SB3}^{(3)}$，$\mathbf{Y}_{SB4}^{(3)}$ を

$$\mathbf{Y}_{SB5}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{SB3}^{(3)}, \mathbf{Y}_{SB4}^{(3)}) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.58}$$

によって結合する．$\mathbf{Y}_{SB3}^{(3)}$ と $\mathbf{Y}_{SB4}^{(3)}$ の 2 テンソルを結合することにより，二種類の不協和音検出の結果を統合している．続いて，畳み込み層が $\mathbf{Y}_{SB5}^{(0)} \in \mathbb{R}^{64\times4\times16\times7}$ から

$$\mathbf{Y}_{SB5}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB5}^{(0)}, \mathcal{L}_{SB5}^{(1)}, \boldsymbol{t}_{SB5}^{(1)})) \in \mathbb{R}^{32\times4\times16\times7} \tag{3.59}$$

を生成する．ここで，$\mathcal{L}_{SB5}^{(1)} \in \mathbb{R}^{32\times64\times1\times1\times1}$ であり，$\boldsymbol{t}_{SB5}^{(1)} = (1, 1, 1)$ である．Strings Block 5 の畳み込み層における各種パラメータを表 3.23 にまとめて示す．

表 3.23: Strings Block 5 の畳み込み層における各種パラメータ

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	$\mathcal{L}_{SB5}^{(1)} \in \mathbb{R}^{32\times64\times1\times1\times1}$	$\boldsymbol{t}_{SB5}^{(1)} = (1, 1, 1)$

Strings Block 6 では，Strings Block 1，Strings Block 2，Strings Block 5 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}_{SB1}^{(2)}$，$\mathbf{Y}_{SB2}^{(2)}$，$\mathbf{Y}_{SB5}^{(1)}$ を

$$\mathbf{Y}_{SB6}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{SB1}^{(2)}, \mathbf{Y}_{SB2}^{(2)}, \mathbf{Y}_{SB5}^{(1)}) \in \mathbb{R}^{160\times4\times16\times7} \tag{3.60}$$

によって結合する．続いて，畳み込み層が $\mathbf{Y}_{SB6}^{(0)} \in \mathbb{R}^{160\times4\times16\times7}$ から

$$\mathbf{Y}_{SB6}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB6}^{(0)}, \mathcal{L}_{SB6}^{(1)}, \boldsymbol{t}_{SB6}^{(1)})) \in \mathbb{R}^{64\times4\times16\times7} \tag{3.61}$$

を生成する．ここで，$\mathcal{L}_{SB6}^{(1)} \in \mathbb{R}^{64\times160\times1\times1\times1}$ であり，$\boldsymbol{t}_{SB6}^{(1)} = (1, 1, 1)$ である．Strings Block 6 の畳み込み層における各種パラメータを表 3.24 にまとめて示す．Strings Block は処理をこれで終了し，$\mathbf{Y}_{SB6}^{(1)}$ を Strings Block の出力 $\mathbf{Y}_{SB}$ として出力する：$\mathbf{Y}_{SB} = \mathbf{Y}_{SB6}^{(1)}$．これが，ストリングストラックの特徴量であり，Band Blocks に入力される．

表 3.24: Strings Block 6 の畳み込み層における各種パラメータ

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	$\mathcal{L}_{SB6}^{(1)} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}_{SB6}^{(1)} = (1, 1, 1)$

3.3 Band Blocks

Band Blocks は Individual Blocks の出力を一つにまとめ，帯域ごとに特徴量を抽出する．帯域は低域 2 オクターブ，中域 2 オクターブ，高域オクターブの三種類である．まず，入力された各有音程楽器特徴量 $\mathbf{Y}_{PiB} \in \mathbb{R}^{64\times4\times16\times7}$，$\mathbf{Y}_{GB} \in \mathbb{R}^{64\times4\times16\times7}$，$\mathbf{Y}_{BaB} \in \mathbb{R}^{64\times4\times16\times7}$，$\mathbf{Y}_{SB} \in \mathbb{R}^{64\times4\times16\times7}$ を

$$\mathbf{Y}_{BB} = \mathrm{Concat}(1, \mathbf{Y}_{PiB}, \mathbf{Y}_{GB}, \mathbf{Y}_{BaB}, \mathbf{Y}_{SB}) \in \mathbb{R}^{256\times4\times16\times7} \tag{3.62}$$

によって結合する．$\mathbf{Y}_{BB}$ は，Low-band Block，Mid-band Block，High-band Block に入力され，帯域ごとの特徴量抽出が行われる．

3.3.1 Low-band Block

Low-band Block では，有音程楽器特徴量 $\mathbf{Y}_{BB} \in \mathbb{R}^{256\times4\times16\times7}$ から低域 2 オクターブ分の特徴量を抽出する．まず，$\mathbf{Y}_{BB}$ から

$$\mathbf{Y}_{LB}^{(0)} = \mathrm{Slice}(\mathbf{Y}_{BB}, 4, 1, 2) \in \mathbb{R}^{256\times4\times16\times2} \tag{3.63}$$

を抽出する．続いて，畳み込み層が $\mathbf{Y}_{LB}^{(0)} \in \mathbb{R}^{256\times4\times16\times2}$ から

$$\mathbf{Y}_{LB}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{LB}^{(0)}, \mathcal{L}_{LB}^{(1)}, \boldsymbol{t}_{LB}^{(1)})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.64}$$

を生成する．ここで，$\mathcal{L}_{LB}^{(1)} \in \mathbb{R}^{64\times256\times1\times4\times2}$ であり，$\boldsymbol{t}_{LB}^{(1)} = (1, 4, 2)$ である．この段階で第 3 モードが 4，第 4 モードが 1 となり，四分音符単位の低域特徴抽出を開始する．Low-band Block は処理をこれで終了し，$\mathbf{Y}_{LB}^{(1)}$ を Low-band Block の出力 $\mathbf{Y}_{LB}$ として出力する：$\mathbf{Y}_{LB} = \mathbf{Y}_{LB}^{(1)}$．これが，低域の特徴量であり，Tonal Blocks に入力される．Low-band Block のパラメータをまとめて表 3.25 に示す．

表 3.25: Low-band Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}_{LB}^{(1)} \in \mathbb{R}^{64\times256\times1\times4\times2}$	$\boldsymbol{t}_{LB}^{(1)} = (1, 4, 2)$

3.3.2 Mid-band Block

Mid-band Block では，有音程楽器特徴量 $\mathbf{Y}_{BB} \in \mathbb{R}^{256\times4\times16\times7}$ から中域 2 オクターブ分の特徴量を抽出する．まず，$\mathbf{Y}_{BB}$ から

$$\mathbf{Y}_{MiB}^{(0)} = \mathrm{Slice}(\mathbf{Y}_{BB}, 4, 3, 4) \in \mathbb{R}^{256\times4\times16\times2} \tag{3.65}$$

を抽出する．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{MiB} \in \mathbb{R}^{256\times4\times16\times2}$ から

$$\mathbf{Y}^{(1)}_{MiB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{MiB}, \mathcal{L}^{(1)}_{MiB}, \boldsymbol{t}^{(1)}_{MiB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.66}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{MiB} \in \mathbb{R}^{64\times256\times1\times4\times2}$ であり，$\boldsymbol{t}^{(1)}_{MiB} = (1,4,2)$ である．この段階で第 3 モードのサイズが 4，第 4 モードのサイズが 1 となり，四分音符単位の中域特徴抽出を開始する．Mid-band Block は処理をこれで終了し，$\mathbf{Y}^{(1)}_{MiB}$ を Mid-band Block の出力 $\mathbf{Y}_{MiB}$ として出力する：$\mathbf{Y}_{MiB} = \mathbf{Y}^{(1)}_{MiB}$．これが，中域の特徴量であり，Tonal Blocks に入力される．Mid-band Block のパラメータをまとめて表 3.26 に示す．

表 3.26: Mid-band Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{MiB} \in \mathbb{R}^{64\times256\times1\times4\times2}$	$\boldsymbol{t}^{(1)}_{MiB} = (1,4,2)$

3.3.3 High-band Block

High-band Block では，有音程楽器特徴量 $\mathbf{Y}_{BB} \in \mathbb{R}^{256\times4\times16\times7}$ から高域 3 オクターブ分の特徴量を抽出する．まず，$\mathbf{Y}_{BB}$ から

$$\mathbf{Y}^{(0)}_{HB} = \mathrm{Slice}(\mathbf{Y}_{BB}, 4, 5, 7) \in \mathbb{R}^{256\times4\times16\times3} \tag{3.67}$$

を抽出する．続いて，畳み込み層が $\mathbf{Y}^{(0)}_{HB} \in \mathbb{R}^{256\times4\times16\times3}$ から

$$\mathbf{Y}^{(1)}_{HB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{HB}, \mathcal{L}^{(1)}_{HB}, \boldsymbol{t}^{(1)}_{HB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.68}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{HB} \in \mathbb{R}^{64\times256\times1\times4\times3}$ であり，$\boldsymbol{t}^{(1)}_{HB} = (1,4,3)$ である．この段階で第 3 モードのサイズが 4，第 4 モードのサイズが 1 となり，四分音符単位の高域特徴抽出を開始する．High-band Block は処理をこれで終了し，$\mathbf{Y}^{(1)}_{HB}$ を High-band Block の出力 $\mathbf{Y}_{HB}$ として出力する：$\mathbf{Y}_{HB} = \mathbf{Y}^{(1)}_{HB}$．これが，高域の特徴量であり，Tonal Blocks に入力される．High-band Block のパラメータをまとめて表 3.27 に示す．

表 3.27: High-band Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{HB} \in \mathbb{R}^{64\times256\times1\times4\times3}$	$\boldsymbol{t}^{(1)}_{HB} = (1,4,3)$

3.4 Tonal Block

Tonal Block では，帯域別特徴量 $\mathbf{Y}_{LB} \in \mathbb{R}^{64\times4\times4\times1}$，$\mathbf{Y}_{MiB} \in \mathbb{R}^{64\times4\times4\times1}$，$\mathbf{Y}_{HB} \in \mathbb{R}^{64\times4\times4\times1}$ 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，入力された特徴 $\mathbf{Y}_{LB}$，$\mathbf{Y}_{MiB}$，$\mathbf{Y}_{HB}$ を

$$\mathbf{Y}_{TB}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{LB}, \mathbf{Y}_{MiB}, \mathbf{Y}_{HB}) \in \mathbb{R}^{192\times4\times4\times1} \tag{3.69}$$

によって結合する．続いて，2 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{TB}^{(0)} \in \mathbb{R}^{192\times4\times4\times1}$ から

$$\mathbf{Y}_{TB}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{TB}^{(0)}, \mathcal{L}_{TB}^{(1)}, \boldsymbol{t}_{TB}^{(1)})) \in \mathbb{R}^{128\times4\times4\times1} \tag{3.70}$$

を生成する．ここで，$\mathcal{L}_{TB}^{(1)} \in \mathbb{R}^{128\times192\times1\times1\times1}$ であり，$\boldsymbol{t}_{TB}^{(1)} = (1,1,1)$ である．次に，第 2 畳み込み層が $\mathbf{Y}_{TB}^{(1)} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{Y}_{TB}^{(2)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{TB}^{(1)}, \mathcal{L}_{TB}^{(2)}, \boldsymbol{t}_{TB}^{(2)})) \in \mathbb{R}^{128\times4\times1\times1} \tag{3.71}$$

を生成する．ここで，$\mathcal{L}_{TB}^{(2)} \in \mathbb{R}^{128\times128\times1\times4\times1}$ であり，$\boldsymbol{t}_{TB}^{(2)} = (1,4,1)$ である．この段階で第 3 モードのサイズが 1 になり，全音符単位の有音程楽器共通特徴抽出を開始する．Tonal Block は処理をこれで終了し，$\mathbf{Y}_{TB}^{(2)}$ を Tonal Block の出力 $\mathbf{Y}_{TB}$ として出力する：$\mathbf{Y}_{TB} = \mathbf{Y}_{TB}^{(2)}$．これが，有音程楽器共通特徴量であり，Merged Block 1 に入力される．Tonal Block のパラメータをまとめて表 3.28 に示す．

表 3.28: Tonal Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}_{TB}^{(1)} \in \mathbb{R}^{128\times192\times1\times1\times1}$	$\boldsymbol{t}_{TB}^{(1)} = (1,1,1)$
第 2 層	$\mathcal{L}_{TB}^{(1)} \in \mathbb{R}^{128\times128\times1\times4\times1}$	$\boldsymbol{t}_{TB}^{(1)} = (1,4,1)$

3.5 Chroma Block

Chroma Block は，有音程楽器全体のクローマからハーモニー特徴量を抽出する．クローマは以下のように求める．まず，入力された特徴 $\mathbf{X}_{Piano} \in \mathbb{R}^{1\times4\times96\times84}$，$\mathbf{X}_{Guitar} \in \mathbb{R}^{1\times4\times96\times84}$，$\mathbf{X}_{Bass} \in \mathbb{R}^{1\times4\times96\times84}$，$\mathbf{X}_{Strings} \in \mathbb{R}^{1\times4\times96\times84}$ を

$$\mathbf{Y}_{Chroma}^{(0)} = \mathrm{Concat}(1, \mathbf{X}_{Piano}, \mathbf{X}_{Guitar}, \mathbf{X}_{Bass}, \mathbf{X}_{Strings}) \in \mathbb{R}^{4\times4\times96\times84} \tag{3.72}$$

によって結合する．続いて，

$$\mathscr{Y}_{Chroma}^{(1)} = \mathrm{Reshape}_{Chroma}(\mathbf{Y}_{Chroma}^{(0)}) \in \mathbb{R}^{4\times4\times8\times12\times12\times7} \tag{3.73}$$

を生成する．ここで，$\mathrm{Reshape}_{Chroma}$ は 入力テンソルを $4\times4\times8\times12\times12\times7$ の 6 階テンソルに並び替える関数である．これは，入力テンソルの第 3 モードを 8 分音符単位で 12 分割し，第 4 モードをオクターブ単位で 7 分割する操作である．続いて，$\mathscr{Y}^{(1)}_{Chroma}$ の第 1 モードのインデックスを i，第 4 モードのインデックスを j，第 6 モードのインデックスを k とすると，以下のように

$$\boldsymbol{Y}_{Chroma} = \frac{1}{4\cdot12\cdot7}\sum_{i=1}^{4}\sum_{j=1}^{12}\sum_{k=1}^{7}\mathscr{Y}^{(1)}_{Chroma}[i,:,:,j,:,k] \in \mathbb{R}^{4\times8\times12} \tag{3.74}$$

を得る．これが有音程楽器全体のクローマである．Chroma Block は，入力のクローマ $\mathbf{Y}_{Chroma} \in \mathbb{R}^{4\times8\times12}$ を

$$\mathbf{Y}^{(0)}_{CB} = \mathrm{Reshape}_{CB}(\boldsymbol{Y}_{Chroma}) \in \mathbb{R}^{1\times4\times8\times12} \tag{3.75}$$

によって変形する．ここで $\mathrm{Reshape}_{CB}$ は 入力テンソルを $1\times4\times8\times12$ の 4 階テンソルに並び替える関数である．続いて，4 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}^{(0)}_{CB} \in \mathbb{R}^{1\times4\times8\times12}$ から

$$\mathbf{Y}^{(1)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{CB}, \mathcal{L}^{(1)}_{CB}, \boldsymbol{t}^{(1)}_{CB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.76}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{CB} \in \mathbb{R}^{64\times1\times1\times1\times12}$ であり，$\boldsymbol{t}^{(1)}_{CB} = (1,1,12)$ である．この段階で第 4 モードのサイズが 1 になり，和音構造の特徴抽出を開始する．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{CB} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}^{(2)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{CB}, \mathcal{L}^{(2)}_{CB}, \boldsymbol{t}^{(2)}_{CB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.77}$$

を生成する．ここで，$\mathcal{L}^{(2)}_{CB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(2)}_{CB} = (1,2,1)$ である．この段階で第 3 モードのサイズが 4 になり，4 分音符単位の和音進行の特徴抽出を開始する．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{CB} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(3)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{CB}, \mathcal{L}^{(3)}_{CB}, \boldsymbol{t}^{(3)}_{CB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.78}$$

を生成する．ここで，$\mathcal{L}^{(3)}_{CB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(3)}_{CB} = (1,2,1)$ である．この段階で第 3 モードのサイズが 2 になり，2 分音符単位の和音進行の特徴抽出を開始する．次に，第 4 畳み込み層が $\mathbf{Y}^{(3)}_{CB} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}^{(4)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{CB}, \mathcal{L}^{(4)}_{CB}, \boldsymbol{t}^{(4)}_{CB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.79}$$

を生成する．ここで，$\mathcal{L}^{(4)}_{CB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(4)}_{CB} = (1,2,1)$ である．この段階で第 3 モードのサイズが 1 になり，全音符単位の和音進行の特徴抽出を開始する．

Chroma Block は処理をこれで終了し，$\mathbf{Y}_{CB}^{(4)}$ を Chroma Block の出力 $\mathbf{Y}_{CB}$ として出力する：$\mathbf{Y}_{CB} = \mathbf{Y}_{CB}^{(4)}$．これが，有音程楽器全体のクローマから抽出したハーモニー特徴量であり，Merged Block 1 に入力される．Chroma Block のパラメータをまとめて表 3.29 に示す．

表 3.29: Chroma Block の畳み込み層における各種パラメータ

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	$\mathcal{L}_{CB}^{(1)} \in \mathbb{R}^{64\times1\times1\times1\times12}$	$\boldsymbol{t}_{CB}^{(1)} = (1, 1, 12)$
第 2 層	$\mathcal{L}_{CB}^{(2)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{CB}^{(2)} = (1, 2, 1)$
第 3 層	$\mathcal{L}_{CB}^{(3)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{CB}^{(3)} = (1, 2, 1)$
第 4 層	$\mathcal{L}_{CB}^{(4)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{CB}^{(4)} = (1, 2, 1)$

3.6 Merged Block 1

Merged Block 1 は，ピッチ情報を持つ複数の特徴量を一つの特徴量に変換する．まず，入力された特徴 $\mathbf{Y}_{TB} \in \mathbb{R}^{128\times4\times1\times1}$，$\mathbf{Y}_{CB} \in \mathbb{R}^{64\times4\times1\times1}$ を

$$\mathbf{Y}_{M1}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{TB}, \mathbf{Y}_{CB}) \in \mathbb{R}^{192\times4\times1\times1} \tag{3.80}$$

によって結合する．続いて，畳み込み層が $\mathbf{Y}_{MB1}^{(0)} \in \mathbb{R}^{192\times4\times1\times1}$ から

$$\mathbf{Y}_{MB1}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{MB1}^{(0)}, \mathcal{L}_{MB1}^{(1)}, \boldsymbol{t}_{MB1}^{(1)})) \in \mathbb{R}^{128\times1\times1\times1} \tag{3.81}$$

を生成する．ここで，$\mathcal{L}_{MB1}^{(1)} \in \mathbb{R}^{128\times192\times4\times1\times1}$ であり，$\boldsymbol{t}_{MB1}^{(1)} = (4, 1, 1)$ である．この段階で第 2 モードのサイズが 1 になり，4 小節全体の特徴抽出を開始する．Merged Block 1 Block は処理をこれで終了し，$\mathbf{Y}_{MB1}^{(1)}$ を Merged Block 1 の出力 $\mathbf{Y}_{MB1}$ として出力する：$\mathbf{Y}_{MB1} = \mathbf{Y}_{MB1}^{(1)}$．これが，ピッチ情報を持つ特徴量であり，Merged Block 3 に入力される．Merged Block 1 のパラメータをまとめて表 3.30 に示す．

表 3.30: Merged Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}_{MB1}^{(1)} \in \mathbb{R}^{128\times192\times4\times1\times1}$	$\boldsymbol{t}_{MB1}^{(1)} = (4, 1, 1)$

3.7 Polyphonicity Blocks

Polyphonicity Blocks は Individual Blocks と同様にトラック別の処理を行うブロックであり，Piano Polyphonicity Block，Guitar Polyphonicity Block，Bass Polyphonicity Block，Strings Polyphonicity Block の 4 ブロックからなる．これらの 4 ブロックは，入力ピアノロールから同時発音数テンソルを作成し，特徴抽出を行う．

3.7.1 Piano Polyphonicity Block

Piano Polyphonicity Block は，ピアノトラックの同時発音数テンソルから特徴抽出を行う．生成器同様，低域 3 オクターブを扱う Piano Low Polyphonicity Block と高域 4 オクターブを扱う Piano High Polyphonicity Block を用いて処理を行う．順に説明する．

Piano Low Polyphonicity Block は，ピアノ低域部の同時発音数テンソル $\mathbf{Y}_{PianoLPoly}$ から特徴抽出を行う．$\mathbf{Y}_{PianoLPoly}$ をは $\mathbf{X}_{Piano}$ の第 4 モードのインデックスを i とし，以下のように求める．

$$\boldsymbol{Y}_{PianoLPoly} = \frac{1}{36}\sum_{i=1}^{36}\mathbf{X}_{Piano}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{3.82}$$

$$\mathbf{Y}_{PianoLPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{PianoLPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{3.83}$$

ここで $\mathrm{Reshape}_{Poly}$ は入力テンソルを $1\times4\times96\times1$ の 4 階テンソルに並び替える関数である．続いて，6 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{PianoLPoly} \in \mathbb{R}^{1\times4\times96\times1}$ から

$$\mathbf{Y}^{(1)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{PianoLPoly}, \mathcal{L}^{(1)}_{PiLPB}, \boldsymbol{t}^{(1)}_{PiLPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{3.84}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{PiLPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$ であり，$\boldsymbol{t}^{(1)}_{PiLPB} = (1,3,1)$ である．この段階で 32 分音符の 3 連符のリズムの特徴抽出を開始し，第 3 モードのサイズが 32 になる．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{PiLPB} \in \mathbb{R}^{32\times4\times32\times1}$ から

$$\mathbf{Y}^{(2)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiLPB}, \mathcal{L}^{(2)}_{PiLPB}, \boldsymbol{t}^{(2)}_{PiLPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{3.85}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{PiLPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(2)}_{PiLPB} = (1,2,1)$ である．この段階で 32 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 16 になる．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{PiLPB} \in \mathbb{R}^{32\times4\times16\times1}$ から

$$\mathbf{Y}^{(3)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiLPB}, \mathcal{L}^{(3)}_{PiLPB}, \boldsymbol{t}^{(3)}_{PiLPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.86}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{PiLPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(3)}_{PiLPB} = (1,2,1)$ である．この段階で 16 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 8 になる．次に，第 4 畳み込み層が $\mathbf{Y}^{(3)}_{PiLPB} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}^{(4)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{PiLPB}, \mathcal{L}^{(4)}_{PiLPB}, \boldsymbol{t}^{(4)}_{PiLPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.87}$$

を抽出する．ここで，$\mathcal{L}^{(4)}_{PiLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(4)}_{PiLPB} = (1,2,1)$ である．この段階で 8 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 4 になる．次に，第 5 畳み込み層が $\mathbf{Y}^{(4)}_{PiLPB} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(5)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{PiLPB}, \mathcal{L}^{(5)}_{PiLPB}, \boldsymbol{t}^{(5)}_{PiLPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.88}$$

を抽出する．ここで，$\mathcal{L}^{(5)}_{PiLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(5)}_{PiLPB} = (1,2,1)$ である．この段階で 4 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 2 になる．次に，第 6 畳み込み層が $\mathbf{Y}^{(5)}_{PiLPB} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}^{(6)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{PiLPB}, \mathcal{L}^{(6)}_{PiLPB}, \boldsymbol{t}^{(6)}_{PiLPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.89}$$

を抽出する．ここで，$\mathcal{L}^{(6)}_{PiLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(6)}_{PiLPB} = (1,2,1)$ である．この段階で 2 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 1 になる．Piano Low Polyphonicity Block は処理をこれで終了する．Piano Low Polyphonicity Block のパラメータをまとめて表 3.31 に示す．

表 3.31: Piano Low Polyphonicity Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{PiLPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$	$\boldsymbol{t}^{(1)}_{PiLPB} = (1,3,1)$
第 2 層	$\mathcal{L}^{(2)}_{PiLPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$	$\boldsymbol{t}^{(2)}_{PiLPB} = (1,2,1)$
第 3 層	$\mathcal{L}^{(3)}_{PiLPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$	$\boldsymbol{t}^{(3)}_{PiLPB} = (1,2,1)$
第 4 層	$\mathcal{L}^{(4)}_{PiLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(4)}_{PiLPB} = (1,2,1)$
第 5 層	$\mathcal{L}^{(5)}_{PiLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(5)}_{PiLPB} = (1,2,1)$
第 6 層	$\mathcal{L}^{(6)}_{PiLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(6)}_{PiLPB} = (1,2,1)$

Piano High Polyphonicity Block は，ピアノ高域部の同時発音数テンソル $\mathbf{Y}_{PianoHPoly}$ から特徴抽出を行う．$\mathbf{Y}_{PianoHPoly}$ は $\mathbf{X}_{Piano}$ の第 4 モードのインデックスを i とし，以下のように求める．

$$\boldsymbol{Y}_{PianoHPoly} = \frac{1}{48}\sum_{i=37}^{84} \mathbf{X}_{Piano}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{3.90}$$

$$\mathbf{Y}_{PianoHPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{PianoHPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{3.91}$$

ここで $\mathrm{Reshape}_{Poly}$ は入力テンソルを $1\times4\times96\times1$ の 4 階テンソルに並び替える関数である．続いて，6 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{PianoHPoly} \in \mathbb{R}^{1\times4\times96\times1}$ から

$$\mathbf{Y}^{(1)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{PianoHPoly}, \mathcal{L}^{(1)}_{PiHPB}, \boldsymbol{t}^{(1)}_{PiHPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{3.92}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{PiHPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$ であり，$\boldsymbol{t}^{(1)}_{PiHPB} = (1,3,1)$ である．この段階で 32 分音符の 3 連符のリズムの特徴抽出を開始し，第 3 モードのサイズが 32 になる．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{PiHPB} \in \mathbb{R}^{32\times4\times32\times1}$ から

$$\mathbf{Y}^{(2)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiHPB}, \mathcal{L}^{(2)}_{PiHPB}, \boldsymbol{t}^{(2)}_{PiHPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{3.93}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{PiHPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(2)}_{PiHPB} = (1,2,1)$ である．この段階で 32 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 16 になる．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{PiHPB} \in \mathbb{R}^{32\times4\times16\times1}$ から

$$\mathbf{Y}^{(3)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiHPB}, \mathcal{L}^{(3)}_{PiHPB}, \boldsymbol{t}^{(3)}_{PiHPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.94}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{PiHPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(3)}_{PiHPB} = (1,2,1)$ である．この段階で 16 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 8 になる．次に，第 4 畳み込み層が $\mathbf{Y}^{(3)}_{PiHPB} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}^{(4)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{PiHPB}, \mathcal{L}^{(4)}_{PiHPB}, \boldsymbol{t}^{(4)}_{PiHPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.95}$$

を抽出する．ここで，$\mathcal{L}^{(4)}_{PiHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(4)}_{PiHPB} = (1,2,1)$ である．この段階で 8 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 4 になる．次に，第 5 畳み込み層が $\mathbf{Y}^{(4)}_{PiHPB} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(5)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{PiHPB}, \mathcal{L}^{(5)}_{PiHPB}, \boldsymbol{t}^{(5)}_{PiHPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.96}$$

を抽出する．ここで，$\mathcal{L}^{(5)}_{PiHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(5)}_{PiHPB} = (1,2,1)$ である．この段階で 4 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 2 になる．次に，第 6 畳み込み層が $\mathbf{Y}^{(5)}_{PiHPB} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}^{(6)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{PiHPB}, \mathcal{L}^{(6)}_{PiHPB}, \boldsymbol{t}^{(6)}_{PiHPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.97}$$

を抽出する．ここで，$\mathcal{L}^{(6)}_{PiHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(6)}_{PiHPB} = (1,2,1)$ である．この段階で 2 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 1 になる．

Piano High Polyphonicity Block は処理をこれで終了する．Piano High Polyphonicity Block のパラメータをまとめて表 3.32 に示す.

表 3.32: Piano High Polyphonicity Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{PiHPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$	$\boldsymbol{t}^{(1)}_{PiHPB} = (1,3,1)$
第 2 層	$\mathcal{L}^{(2)}_{PiHPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$	$\boldsymbol{t}^{(2)}_{PiHPB} = (1,2,1)$
第 3 層	$\mathcal{L}^{(3)}_{PiHPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$	$\boldsymbol{t}^{(3)}_{PiHPB} = (1,2,1)$
第 4 層	$\mathcal{L}^{(4)}_{PiHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(4)}_{PiHPB} = (1,2,1)$
第 5 層	$\mathcal{L}^{(5)}_{PiHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(5)}_{PiHPB} = (1,2,1)$
第 6 層	$\mathcal{L}^{(6)}_{PiHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(6)}_{PiHPB} = (1,2,1)$

最後に低域 $\mathbf{Y}^{(6)}_{PiLPB}$ と高域 $\mathbf{Y}^{(6)}_{PiHPB}$ を

$$\mathbf{Y}_{PiPB} = \mathrm{Concat}(4, \mathbf{Y}^{(6)}_{PiLPB}, \mathbf{Y}^{(6)}_{PiHPB}) \in \mathbb{R}^{128\times4\times1\times1} \tag{3.98}$$

によって結合する．これが，ピアノトラックの同時発音数特徴量であり，Merged Block 2 に入力される．

3.7.2 Guitar Polyphonicity Block

Guitar Polyphonicity Block は，ギタートラックの同時発音数テンソル $\mathbf{Y}_{GuitarPoly}$ から特徴抽出を行う．$\mathbf{Y}_{GuitarPoly}$ は $\mathbf{X}_{Guitar}$ の第 4 モードのインデックスを i とし，以下のように求める．

$$\boldsymbol{Y}_{GuitarPoly} = \frac{1}{84}\sum_{i=1}^{84} \mathbf{X}_{Guitar}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{3.99}$$

$$\mathbf{Y}_{GuitarPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{GuitarPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{3.100}$$

ここで $\mathrm{Reshape}_{Poly}$ は入力テンソルを $1\times4\times96\times1$ の 4 階テンソルに並び替える関数である．続いて，6 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{GuitarPoly} \in \mathbb{R}^{1\times4\times96\times1}$ から

$$\mathbf{Y}^{(1)}_{GPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GuitarPoly}, \mathcal{L}^{(1)}_{GPB}, \boldsymbol{t}^{(1)}_{GPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{3.101}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{GPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$ であり，$\boldsymbol{t}^{(1)}_{GPB} = (1,3,1)$ である．この段階で 32 分音符の 3 連符のリズムの特徴抽出を開始し，第 3 モードのサイズが 32 になる．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{GPB} \in \mathbb{R}^{32\times4\times32\times1}$ から

$$\mathbf{Y}^{(2)}_{GPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GPB}, \mathcal{L}^{(2)}_{GPB}, \boldsymbol{t}^{(2)}_{GPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{3.102}$$

を抽出する．ここで，$\mathcal{L}_{GPB}^{(2)} \in \mathbb{R}^{32\times32\times1\times2\times1}$ であり，$\boldsymbol{t}_{GPB}^{(2)} = (1,2,1)$ である．この段階で 32 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 16 になる．次に，第 3 畳み込み層が $\mathbf{Y}_{GPB}^{(2)} \in \mathbb{R}^{32\times4\times16\times1}$ から

$$\mathbf{Y}_{GPB}^{(3)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(2)}, \mathcal{L}_{GPB}^{(3)}, \boldsymbol{t}_{GPB}^{(3)})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.103}$$

を抽出する．ここで，$\mathcal{L}_{GPB}^{(3)} \in \mathbb{R}^{64\times32\times1\times2\times1}$ であり，$\boldsymbol{t}_{GPB}^{(3)} = (1,2,1)$ である．この段階で 16 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 8 になる．次に，第 4 畳み込み層が $\mathbf{Y}_{GPB}^{(3)} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}_{GPB}^{(4)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(3)}, \mathcal{L}_{GPB}^{(4)}, \boldsymbol{t}_{GPB}^{(4)})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.104}$$

を抽出する．ここで，$\mathcal{L}_{GPB}^{(4)} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}_{GPB}^{(4)} = (1,2,1)$ である．この段階で 8 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 4 になる．次に，第 5 畳み込み層が $\mathbf{Y}_{GPB}^{(4)} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}_{GPB}^{(5)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(4)}, \mathcal{L}_{GPB}^{(5)}, \boldsymbol{t}_{GPB}^{(5)})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.105}$$

を抽出する．ここで，$\mathcal{L}_{GPB}^{(5)} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}_{GPB}^{(5)} = (1,2,1)$ である．この段階で 4 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 2 になる．次に，第 6 畳み込み層が $\mathbf{Y}_{GPB}^{(5)} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}_{GPB}^{(6)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(5)}, \mathcal{L}_{GPB}^{(6)}, \boldsymbol{t}_{GPB}^{(6)})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.106}$$

を抽出する．ここで，$\mathcal{L}_{GPB}^{(6)} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}_{GPB}^{(6)} = (1,2,1)$ である．この段階で 2 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 1 になる．Guitar Polyphonicity Block は処理をこれで終了し，$\mathbf{Y}_{GPB}^{(6)}$ を Guitar Polyphonicity Block の出力 $\mathbf{Y}_{GPB}$ として出力する：$\mathbf{Y}_{GPB} = \mathbf{Y}_{GPB}^{(6)}$．これが，ギタートラックの同時発音数特徴量であり，Merged Block 2 に入力される．Guitar Polyphonicity Block のパラメータをまとめて表 3.33 に示す.

表 3.33: Guitar Polyphonicity Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}_{GPB}^{(1)} \in \mathbb{R}^{32\times1\times1\times3\times1}$	$\boldsymbol{t}_{GPB}^{(1)} = (1,3,1)$
第 2 層	$\mathcal{L}_{GPB}^{(2)} \in \mathbb{R}^{32\times32\times1\times2\times1}$	$\boldsymbol{t}_{GPB}^{(2)} = (1,2,1)$
第 3 層	$\mathcal{L}_{GPB}^{(3)} \in \mathbb{R}^{64\times32\times1\times2\times1}$	$\boldsymbol{t}_{GPB}^{(3)} = (1,2,1)$
第 4 層	$\mathcal{L}_{GPB}^{(4)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{GPB}^{(4)} = (1,2,1)$
第 5 層	$\mathcal{L}_{GPB}^{(5)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{GPB}^{(5)} = (1,2,1)$
第 6 層	$\mathcal{L}_{GPB}^{(6)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{GPB}^{(6)} = (1,2,1)$

3.7.3 Bass Polyphonicity Block

Bass Polyphonicity Block は，ベーストラックの同時発音数テンソル $\mathbf{Y}_{BassPoly}$ から特徴抽出を行う．$\mathbf{Y}_{BassPoly}$ は $\mathbf{X}_{Bass}$ の第 4 モードのインデックスを i とし，以下のように求める．

$$\boldsymbol{Y}_{BassPoly} = \frac{1}{84}\sum_{i=1}^{84}\mathbf{X}_{Bass}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{3.107}$$

$$\mathbf{Y}_{BassPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{BassPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{3.108}$$

ここで $\mathrm{Reshape}_{Poly}$ は入力テンソルを $1\times4\times96\times1$ の 4 階テンソルに並び替える関数である．続いて，6 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{BassPoly} \in \mathbb{R}^{1\times4\times96\times1}$ から

$$\mathbf{Y}^{(1)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{BassPoly}, \mathcal{L}^{(1)}_{BaPB}, \boldsymbol{t}^{(1)}_{BaPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{3.109}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{BaPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$ であり，$\boldsymbol{t}^{(1)}_{BaPB} = (1,3,1)$ である．この段階で 32 分音符の 3 連符のリズムの特徴抽出を開始し，第 3 モードのサイズが 32 になる．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{BaPB} \in \mathbb{R}^{32\times4\times32\times1}$ から

$$\mathbf{Y}^{(2)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaPB}, \mathcal{L}^{(2)}_{BaPB}, \boldsymbol{t}^{(2)}_{BaPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{3.110}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{BaPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(2)}_{BaPB} = (1,2,1)$ である．この段階で 32 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 16 になる．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{BaPB} \in \mathbb{R}^{32\times4\times16\times1}$ から

$$\mathbf{Y}^{(3)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{BaPB}, \mathcal{L}^{(3)}_{BaPB}, \boldsymbol{t}^{(3)}_{BaPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.111}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{BaPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(3)}_{BaPB} = (1,2,1)$ である．この段階で 16 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 8 になる．次に，第 4 畳み込み層が $\mathbf{Y}^{(3)}_{BaPB} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}^{(4)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{BaPB}, \mathcal{L}^{(4)}_{BaPB}, \boldsymbol{t}^{(4)}_{BaPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.112}$$

を抽出する．ここで，$\mathcal{L}^{(4)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(4)}_{BaPB} = (1,2,1)$ である．この段階で 8 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 4 になる．次に，第 5 畳み込み層が $\mathbf{Y}^{(4)}_{BaPB} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(5)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{BaPB}, \mathcal{L}^{(5)}_{BaPB}, \boldsymbol{t}^{(5)}_{BaPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.113}$$

を抽出する．ここで，$\mathcal{L}^{(5)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(5)}_{BaPB} = (1,2,1)$ である．この段階で 4 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 2 になる．次に，第 6 畳み込み層が $\mathbf{Y}^{(5)}_{BaPB} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}^{(6)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{BaPB}, \mathcal{L}^{(6)}_{BaPB}, \boldsymbol{t}^{(6)}_{BaPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.114}$$

を抽出する．ここで，$\mathcal{L}^{(6)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(6)}_{BaPB} = (1,2,1)$ である．この段階で 2 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 1 になる．Bass Polyphonicity Block は処理をこれで終了し，$\mathbf{Y}^{(6)}_{BaPB}$ を Bass Polyphonicity Block の出力 $\mathbf{Y}_{BaPB}$ として出力する：$\mathbf{Y}_{BaPB} = \mathbf{Y}^{(6)}_{BaPB}$．これが，ベーストラックの同時発音数特徴量であり，Merged Block 2 に入力される．Bass Polyphonicity Block のパラメータをまとめて表 3.34 に示す．

表 3.34: Bass Polyphonicity Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{BaPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$	$\boldsymbol{t}^{(1)}_{BaPB} = (1,3,1)$
第 2 層	$\mathcal{L}^{(2)}_{BaPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$	$\boldsymbol{t}^{(2)}_{BaPB} = (1,2,1)$
第 3 層	$\mathcal{L}^{(3)}_{BaPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$	$\boldsymbol{t}^{(3)}_{BaPB} = (1,2,1)$
第 4 層	$\mathcal{L}^{(4)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(4)}_{BaPB} = (1,2,1)$
第 5 層	$\mathcal{L}^{(5)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(5)}_{BaPB} = (1,2,1)$
第 6 層	$\mathcal{L}^{(6)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(6)}_{BaPB} = (1,2,1)$

3.7.4 Strings Polyphonicity Block

Strings Polyphonicity Block は，ストリングストラックの同時発音数テンソルから特徴抽出を行う．生成器同様，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，Strings Low Block と Strings High Block と呼ぶことにする．順に説明する．

Strings Low Block は，ストリングス低域部の同時発音数テンソル $\mathbf{Y}_{StringsLPoly}$ から特徴抽出を行う．$\mathbf{Y}_{StringsLPoly}$ は $\mathbf{X}_{Strings}$ の第 4 モードのインデックスを i とし，以下のように求める．

$$\boldsymbol{Y}_{StringsLPoly} = \frac{1}{36}\sum_{i=1}^{36} \mathbf{X}_{Strings}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{3.115}$$

$$\mathbf{Y}_{StringsLPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{StringsLPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{3.116}$$

ここで $\mathrm{Reshape}_{Poly}$ は入力テンソルを $1\times4\times96\times1$ の 4 階テンソルに並び替える関数である．続いて，6 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{StringsLPoly} \in \mathbb{R}^{1\times4\times96\times1}$ から

$$\mathbf{Y}^{(1)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{StringsLPoly}, \mathcal{L}^{(1)}_{SLPB}, \boldsymbol{t}^{(1)}_{SLPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{3.117}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{SLPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$ であり，$\boldsymbol{t}^{(1)}_{SLPB} = (1,3,1)$ である．この段階で 32 分音符の 3 連符のリズムの特徴抽出を開始し，第 3 モードのサイズが 32 になる．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{SLPB} \in \mathbb{R}^{32\times4\times32\times1}$ から

$$\mathbf{Y}^{(2)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SLPB}, \mathcal{L}^{(2)}_{SLPB}, \boldsymbol{t}^{(2)}_{SLPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{3.118}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{SLPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(2)}_{SLPB} = (1,2,1)$ である．この段階で 32 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 16 になる．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{SLPB} \in \mathbb{R}^{32\times4\times16\times1}$ から

$$\mathbf{Y}^{(3)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{SLPB}, \mathcal{L}^{(3)}_{SLPB}, \boldsymbol{t}^{(3)}_{SLPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.119}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{SLPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(3)}_{SLPB} = (1,2,1)$ である．この段階で 16 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 8 になる．次に，第 4 畳み込み層が $\mathbf{Y}^{(3)}_{SLPB} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}^{(4)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{SLPB}, \mathcal{L}^{(4)}_{SLPB}, \boldsymbol{t}^{(4)}_{SLPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.120}$$

を抽出する．ここで，$\mathcal{L}^{(4)}_{SLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(4)}_{SLPB} = (1,2,1)$ である．この段階で 8 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 4 になる．次に，第 5 畳み込み層が $\mathbf{Y}^{(4)}_{SLPB} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(5)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{SLPB}, \mathcal{L}^{(5)}_{SLPB}, \boldsymbol{t}^{(5)}_{SLPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.121}$$

を抽出する．ここで，$\mathcal{L}^{(5)}_{SLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(5)}_{SLPB} = (1,2,1)$ である．この段階で 4 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 2 になる．次に，第 6 畳み込み層が $\mathbf{Y}^{(5)}_{SLPB} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}^{(6)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{SLPB}, \mathcal{L}^{(6)}_{SLPB}, \boldsymbol{t}^{(6)}_{SLPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.122}$$

を抽出する．ここで，$\mathcal{L}^{(6)}_{SLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(6)}_{SLPB} = (1,2,1)$ である．この段階で 2 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 1 になる．Strings Polyphonicity Block は処理をこれで終了する．Strings Low Polyphonicity Block のパラメータをまとめて表 3.35 に示す.

表 3.35: Strings Low Polyphonicity Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{SLPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$	$\boldsymbol{t}^{(1)}_{SLPB} = (1,3,1)$
第 2 層	$\mathcal{L}^{(2)}_{SLPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$	$\boldsymbol{t}^{(2)}_{SLPB} = (1,2,1)$
第 3 層	$\mathcal{L}^{(3)}_{SLPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$	$\boldsymbol{t}^{(3)}_{SLPB} = (1,2,1)$
第 4 層	$\mathcal{L}^{(4)}_{SLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(4)}_{SLPB} = (1,2,1)$
第 5 層	$\mathcal{L}^{(5)}_{SLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(5)}_{SLPB} = (1,2,1)$
第 6 層	$\mathcal{L}^{(6)}_{SLPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(6)}_{SLPB} = (1,2,1)$

Strings High Block は，ストリングス高域部の同時発音数テンソル $\mathbf{Y}_{StringsHPoly}$ から特徴抽出を行う．$\mathbf{Y}_{StringsHPoly}$ は $\mathbf{X}_{Strings}$ の第 4 モードのインデックスを i とし，以下のように求める．

$$\boldsymbol{Y}_{StringsHPoly} = \frac{1}{36}\sum_{i=37}^{72}\mathbf{X}_{Strings}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{3.123}$$

$$\mathbf{Y}_{StringsHPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{StringsHPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{3.124}$$

ここで $\mathrm{Reshape}_{Poly}$ は入力テンソルを $1\times4\times96\times1$ の 4 階テンソルに並び替える関数である．続いて，6 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}_{StringsHPoly} \in \mathbb{R}^{1\times4\times96\times1}$ から

$$\mathbf{Y}^{(1)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{StringsHPoly}, \mathcal{L}^{(1)}_{SHPB}, \boldsymbol{t}^{(1)}_{SHPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{3.125}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{SHPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$ であり，$\boldsymbol{t}^{(1)}_{SHPB} = (1,3,1)$ である．この段階で 32 分音符の 3 連符のリズムの特徴抽出を開始し，第 3 モードのサイズが 32 になる．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{SHPB} \in \mathbb{R}^{32\times4\times32\times1}$ から

$$\mathbf{Y}^{(2)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SHPB}, \mathcal{L}^{(2)}_{SHPB}, \boldsymbol{t}^{(2)}_{SHPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{3.126}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{SHPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(2)}_{SHPB} = (1,2,1)$ である．この段階で 32 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 16 になる．次に，第 3 畳み込み層が $\mathbf{Y}^{(2)}_{SHPB} \in \mathbb{R}^{32\times4\times16\times1}$ から

$$\mathbf{Y}^{(3)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{SHPB}, \mathcal{L}^{(3)}_{SHPB}, \boldsymbol{t}^{(3)}_{SHPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{3.127}$$

を抽出する．ここで，$\mathcal{L}^{(3)}_{SHPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$ であり，$\boldsymbol{t}^{(3)}_{SHPB} = (1,2,1)$ である．この段階で 16 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 8 になる．次に，第 4 畳み込み層が $\mathbf{Y}^{(3)}_{SHPB} \in \mathbb{R}^{64\times4\times8\times1}$ から

$$\mathbf{Y}^{(4)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{SHPB}, \mathcal{L}^{(4)}_{SHPB}, \boldsymbol{t}^{(4)}_{SHPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.128}$$

を抽出する．ここで，$\mathcal{L}^{(4)}_{SHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(4)}_{SHPB} = (1,2,1)$ である．この段階で 8 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 4 になる．次に，第 5 畳み込み層が $\mathbf{Y}^{(4)}_{SHPB} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(5)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{SHPB}, \mathcal{L}^{(5)}_{SHPB}, \boldsymbol{t}^{(5)}_{SHPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{3.129}$$

を抽出する．ここで，$\mathcal{L}^{(5)}_{SHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(5)}_{SHPB} = (1,2,1)$ である．この段階で 4 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 2 になる．次に，第 6 畳み込み層が $\mathbf{Y}^{(5)}_{SHPB} \in \mathbb{R}^{64\times4\times2\times1}$ から

$$\mathbf{Y}^{(6)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{SHPB}, \mathcal{L}^{(6)}_{SHPB}, \boldsymbol{t}^{(6)}_{SHPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{3.130}$$

を抽出する．ここで，$\mathcal{L}^{(6)}_{SHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$ であり，$\boldsymbol{t}^{(6)}_{SHPB} = (1,2,1)$ である．この段階で 2 分音符単位のリズムの特徴抽出を開始し，第 3 モードのサイズが 1 になる．Strings Polyphonicity Block は処理をこれで終了する．Strings High Polyphonicity Block のパラメータをまとめて表 3.36 に示す.

表 3.36: Strings High Polyphonicity Block の転置畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{SHPB} \in \mathbb{R}^{32\times1\times1\times3\times1}$	$\boldsymbol{t}^{(1)}_{SHPB} = (1,3,1)$
第 2 層	$\mathcal{L}^{(2)}_{SHPB} \in \mathbb{R}^{32\times32\times1\times2\times1}$	$\boldsymbol{t}^{(2)}_{SHPB} = (1,2,1)$
第 3 層	$\mathcal{L}^{(3)}_{SHPB} \in \mathbb{R}^{64\times32\times1\times2\times1}$	$\boldsymbol{t}^{(3)}_{SHPB} = (1,2,1)$
第 4 層	$\mathcal{L}^{(4)}_{SHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(4)}_{SHPB} = (1,2,1)$
第 5 層	$\mathcal{L}^{(5)}_{SHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(5)}_{SHPB} = (1,2,1)$
第 6 層	$\mathcal{L}^{(6)}_{SHPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(6)}_{SHPB} = (1,2,1)$

最後に低域 $\mathbf{Y}^{(1)}_{SLPB}$ と高域 $\mathbf{Y}^{(1)}_{SHPB}$ を

$$\mathbf{Y}_{SPB} = \mathrm{Concat}(4, \mathbf{Y}^{(1)}_{SLPB}, \mathbf{Y}^{(1)}_{SHPB}) \in \mathbb{R}^{128\times4\times1\times1} \tag{3.131}$$

によって結合する．これが，ストリングストラックの同時発音数特徴量であり，Merged Block 2 に入力される．

3.8 Merged Block 2

Merged Block 2 では，Polyphonicity Blocks で抽出した各トラックの同時発音数特徴量を一つにまとめ，特徴抽出を行う．まず，入力された特徴 $\mathbf{Y}_{PiPB} \in \mathbb{R}^{128\times4\times1\times1}$，

$\mathbf{Y}_{GPB} \in \mathbb{R}^{64\times4\times1\times1}$，$\mathbf{Y}_{BaPB} \in \mathbb{R}^{64\times4\times1\times1}$，$\mathbf{Y}_{SPB} \in \mathbb{R}^{128\times4\times1\times1}$ を

$$\mathbf{Y}^{(0)}_{MB2} = \mathrm{Concat}(1, \mathbf{Y}_{PiPB}, \mathbf{Y}_{GPB}, \mathbf{Y}_{BaPB}, \mathbf{Y}_{SPB}) \in \mathbb{R}^{384\times4\times1\times1} \tag{3.132}$$

によって結合する．続いて，2 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}^{(0)}_{MB2} \in \mathbb{R}^{384\times4\times1\times1}$ から

$$\mathbf{Y}^{(1)}_{MB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{MB2}, \mathcal{L}^{(1)}_{MB2}, \boldsymbol{t}^{(1)}_{MB2})) \in \mathbb{R}^{128\times4\times1\times1} \tag{3.133}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{MB2} \in \mathbb{R}^{128\times384\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{MB2} = (1, 1, 1)$ である．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{MB2} \in \mathbb{R}^{128\times4\times1\times1}$ から

$$\mathbf{Y}^{(2)}_{MB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{MB2}, \mathcal{L}^{(2)}_{MB2}, \boldsymbol{t}^{(2)}_{MB2})) \in \mathbb{R}^{128\times1\times1\times1} \tag{3.134}$$

を生成する．ここで，$\mathcal{L}^{(2)}_{MB2} \in \mathbb{R}^{128\times128\times4\times1\times1}$ であり，$\boldsymbol{t}^{(2)}_{MB2} = (4, 1, 1)$ である．この段階で小節ごとのリズム特徴抽出を開始し，第 2 モードのサイズが 1 になる．Merged Block 2 は処理をこれで終了し，$\mathbf{Y}^{(2)}_{MB2}$ を Merged Block 2 の出力 $\mathbf{Y}_{MB2}$ として出力する：$\mathbf{Y}_{MB2} = \mathbf{Y}^{(2)}_{MB2}$．これが，有音程楽器全体の同時発音数から抽出した特徴量であり，Merged Block 3 に入力される．Merged Block 2 のパラメータをまとめて表 3.37 に示す．

表 3.37: Merged Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{MB2} \in \mathbb{R}^{128\times384\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{MB2} = (1, 1, 1)$
第 2 層	$\mathcal{L}^{(2)}_{MB2} \in \mathbb{R}^{128\times128\times4\times1\times1}$	$\boldsymbol{t}^{(2)}_{MB2} = (4, 1, 1)$

3.9 Individual Percussion Block

提案手法では，打楽器トラック $\mathbf{X}_{Percussion} \in \mathbb{R}^{1\times4\times96\times84}$ をドラム打楽器 $\mathbf{X}_{Drums} \in \mathbb{R}^{1\times4\times96\times15}$ とその他の打楽器 $\mathbf{X}_{OtherPercussion} \in \mathbb{R}^{1\times4\times96\times15}$ に分離して処理を行う．この処理を，ドラム打楽器 $\mathbf{X}_{Drums}$ を抽出する関数 $\mathrm{Exploit}_D$ とその他の打楽器 $\mathbf{X}_{OtherPercussion}$ を抽出する関数 $\mathrm{Exploit}_{OP}$ を用いて，

$$\mathbf{X}_{Drums} = \mathrm{Exploit}_D(\mathbf{X}_{Percussion}) \in \mathbb{R}^{1\times4\times96\times15} \tag{3.135}$$

$$\mathbf{X}_{OtherPercussion} = \mathrm{Exploit}_{OP}(\mathbf{X}_{Percussion}) \in \mathbb{R}^{1\times4\times96\times15} \tag{3.136}$$

のように表す．ここで，

$$\mathbf{X}_{Percussion} = (\boldsymbol{X}^{1}_{Pe}, \boldsymbol{X}^{2}_{Pe}, \ldots, \boldsymbol{X}^{84}_{Pe}),$$

$$\mathbf{X}_{Drums} = (\boldsymbol{X}^{1}_{Drums}, \boldsymbol{X}^{2}_{Drums}, \ldots, \boldsymbol{X}^{15}_{Drums}),$$

$$\mathbf{X}_{OtherPercussion} = (\boldsymbol{X}^{1}_{OtherPercussion}, \boldsymbol{X}^{2}_{OtherPercussion}, \ldots, \boldsymbol{X}^{15}_{OtherPercussion})$$

であり，$\boldsymbol{X}_{Pe}, \boldsymbol{X}_{Drums}, \boldsymbol{X}_{OtherPercussion}$ の上付きの添え字は音高方向のインデックスを表す．表 3.38 が示す通りに，関数 $\mathrm{Exploit}_{D}$ は $\mathbf{X}_{Percussion}$ を $\mathbf{X}_{Drums}$ に変換し，関数 $\mathrm{Exploit}_{OP}$ は $\mathbf{X}_{Percussion}$ を $\mathbf{X}_{OtherPercussion}$ に変換する．なお，判別器の入力が学習データの際，表 3.38 の変換前に複数の $\boldsymbol{X}_{Pe}$ が存在する場合がある．この時は，変換前の複数の $\boldsymbol{X}_{Pe}$ の和を算出する．こうして得られた $\mathbf{X}_{Drums}$ は Drums Block に入力され，$\mathbf{X}_{OtherPercussion}$ は Other Percussion Block に入力される．

表 3.38: 判別器による打楽器変換

変換前	変換後
$\boldsymbol{X}^{35}_{Pe}, \boldsymbol{X}^{36}_{Pe}$	$\boldsymbol{X}^{1}_{Drum}$
$\boldsymbol{X}^{37}_{Pe}$	$\boldsymbol{X}^{2}_{Drum}$
$\boldsymbol{X}^{38}_{Pe}, \boldsymbol{X}^{40}_{Pe}$	$\boldsymbol{X}^{3}_{Drum}$
$\boldsymbol{X}^{41}_{Pe}, \boldsymbol{X}^{43}_{Pe}$	$\boldsymbol{X}^{4}_{Drum}$
$\boldsymbol{X}^{45}_{Pe}, \boldsymbol{X}^{47}_{Pe}$	$\boldsymbol{X}^{5}_{Drum}$
$\boldsymbol{X}^{48}_{Pe}, \boldsymbol{X}^{50}_{Pe}$	$\boldsymbol{X}^{6}_{Drum}$
$\boldsymbol{X}^{44}_{Pe}$	$\boldsymbol{X}^{7}_{Drum}$
$\boldsymbol{X}^{42}_{Pe}$	$\boldsymbol{X}^{8}_{Drum}$
$\boldsymbol{X}^{46}_{Pe}$	$\boldsymbol{X}^{9}_{Drum}$
$\boldsymbol{X}^{49}_{Pe}, \boldsymbol{X}^{57}_{Pe}$	$\boldsymbol{X}^{10}_{Drum}$
$\boldsymbol{X}^{51}_{Pe}, \boldsymbol{X}^{59}_{Pe}$	$\boldsymbol{X}^{11}_{Drum}$
$\boldsymbol{X}^{52}_{Pe}$	$\boldsymbol{X}^{12}_{Drum}$
$\boldsymbol{X}^{53}_{Pe}$	$\boldsymbol{X}^{13}_{Drum}$
$\boldsymbol{X}^{55}_{Pe}$	$\boldsymbol{X}^{14}_{Drum}$
$\boldsymbol{X}^{56}_{Pe}$	$\boldsymbol{X}^{15}_{Drum}$

変換前	変換後
$\boldsymbol{X}^{39}_{Pe}$	$\boldsymbol{X}^{1}_{OtherPercussion}$
$\boldsymbol{X}^{54}_{Pe}$	$\boldsymbol{X}^{2}_{OtherPercussion}$
$\boldsymbol{X}^{58}_{Pe}$	$\boldsymbol{X}^{3}_{OtherPercussion}$
$\boldsymbol{X}^{60}_{Pe}, \boldsymbol{X}^{61}_{Pe}$	$\boldsymbol{X}^{4}_{OtherPercussion}$
$\boldsymbol{X}^{62}_{Pe}, \boldsymbol{X}^{63}_{Pe}, \boldsymbol{X}^{64}_{Pe}$	$\boldsymbol{X}^{5}_{OtherPercussion}$
$\boldsymbol{X}^{65}_{Pe}, \boldsymbol{X}^{66}_{Pe}$	$\boldsymbol{X}^{6}_{OtherPercussion}$
$\boldsymbol{X}^{67}_{Pe}, \boldsymbol{X}^{68}_{Pe}$	$\boldsymbol{X}^{7}_{OtherPercussion}$
$\boldsymbol{X}^{69}_{Pe}$	$\boldsymbol{X}^{8}_{OtherPercussion}$
$\boldsymbol{X}^{70}_{Pe}$	$\boldsymbol{X}^{9}_{OtherPercussion}$
$\boldsymbol{X}^{71}_{Pe}, \boldsymbol{X}^{72}_{Pe}$	$\boldsymbol{X}^{10}_{OtherPercussion}$
$\boldsymbol{X}^{73}_{Pe}, \boldsymbol{X}^{74}_{Pe}$	$\boldsymbol{X}^{11}_{OtherPercussion}$
$\boldsymbol{X}^{75}_{Pe}$	$\boldsymbol{X}^{12}_{OtherPercussion}$
$\boldsymbol{X}^{76}_{Pe}, \boldsymbol{X}^{77}_{Pe}$	$\boldsymbol{X}^{13}_{OtherPercussion}$
$\boldsymbol{X}^{78}_{Pe}, \boldsymbol{X}^{79}_{Pe}$	$\boldsymbol{X}^{14}_{OtherPercussion}$
$\boldsymbol{X}^{80}_{Pe}, \boldsymbol{X}^{81}_{Pe}$	$\boldsymbol{X}^{15}_{OtherPercussion}$

3.9.1 Drums Block

Drums Block は 3 個のサブブロックで構成される．それらを Drums Block 1，Drums Block 2，および Drums Block 3 と呼ぶことにする．Drums Block 1 は，表 3.39 に示す 2 層の畳み込み層で構成される．第 1 畳み込み層は $\mathbf{X}_{Drums} \in \mathbb{R}^{1\times4\times96\times15}$ から

$$\mathbf{Y}^{(1)}_{DB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Drums}, \mathcal{L}^{(1)}_{DB1}, \boldsymbol{t}^{(1)}_{DB1})) \in \mathbb{R}^{32\times4\times96\times1} \tag{3.137}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{DB1} \in \mathbb{R}^{32\times1\times1\times1\times15}$ であり，$\boldsymbol{t}^{(1)}_{DB1} = (1, 1, 15)$ である．この段階で第 4 モードのサイズが 1 になり，各音色の特徴抽出を開始する．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{DB1} \in \mathbb{R}^{32\times4\times4\times15}$ から

$$\mathbf{Y}^{(2)}_{DB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{DB1}, \mathcal{L}^{(2)}_{DB1}, \boldsymbol{t}^{(2)}_{DB1})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.138}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{DB1} \in \mathbb{R}^{64\times32\times1\times24\times1}$ であり，$\boldsymbol{t}^{(2)}_{DB1} = (1, 24, 1)$ 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抽出を開始する．以上のように，Drums Block 1 では，第 4 モードの特徴抽出を行ってから第 3 モードの特徴抽出を行う．

表 3.39: Drums Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{DB1} \in \mathbb{R}^{32\times1\times1\times1\times15}$	$\boldsymbol{t}^{(1)}_{DB1} = (1, 1, 15)$
第 2 層	$\mathcal{L}^{(2)}_{DB1} \in \mathbb{R}^{64\times32\times1\times24\times1}$	$\boldsymbol{t}^{(2)}_{DB1} = (1, 24, 1)$

一方，Drums Block 2 は同じサイズの特徴量テンソルを逆の手順で抽出する．すなわち，第 3 モードの特徴量を抽出してから第 4 モードの特徴抽出を行う．第 1 畳み込み層は $\mathbf{X}_{Drums} \in \mathbb{R}^{1\times4\times96\times15}$ から

$$\mathbf{Y}^{(1)}_{DB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Drums}, \mathcal{L}^{(1)}_{DB2}, \boldsymbol{t}^{(1)}_{DB2})) \in \mathbb{R}^{32\times4\times4\times15} \tag{3.139}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{DB2} \in \mathbb{R}^{32\times1\times1\times24\times1}$ であり，$\boldsymbol{t}^{(1)}_{DB2} = (1, 24, 1)$ 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抽出を開始する．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{DB2} \in \mathbb{R}^{32\times4\times4\times15}$ から

$$\mathbf{Y}^{(2)}_{DB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{DB2}, \mathcal{L}^{(2)}_{DB2}, \boldsymbol{t}^{(2)}_{DB2})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.140}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{DB2} \in \mathbb{R}^{64\times32\times1\times1\times15}$ であり，$\boldsymbol{t}^{(2)}_{DB2} = (1, 1, 15)$ である．この段階で第 4 モードのサイズが 1 になり，各音色の特徴抽出を開始する．

表 3.40: Drums Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{DB2} \in \mathbb{R}^{32\times1\times1\times24\times1}$	$\boldsymbol{t}^{(1)}_{DB2} = (1, 24, 1)$
第 2 層	$\mathcal{L}^{(2)}_{DB2} \in \mathbb{R}^{64\times32\times1\times1\times15}$	$\boldsymbol{t}^{(2)}_{DB2} = (1, 1, 15)$

Drums Block 3 では，Drums Block 1 と Drums Block 2 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(2)}_{DB1} \in \mathbb{R}^{64\times4\times4\times1}$，$\mathbf{Y}^{(2)}_{DB2} \in \mathbb{R}^{64\times4\times4\times1}$ を

$$\mathbf{Y}^{(0)}_{DB3} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{DB1}, \mathbf{Y}^{(2)}_{DB2}) \in \mathbb{R}^{128\times4\times4\times1} \tag{3.141}$$

によって結合する．続いて，2 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}^{(0)}_{DB3} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{Y}^{(1)}_{DB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{DB3}, \mathcal{L}^{(1)}_{DB3}, \boldsymbol{t}^{(1)}_{DB3})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.142}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{DB3} \in \mathbb{R}^{64\times128\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{DB3} = (1, 1, 1)$ である．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{DB3} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(2)}_{DB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{DB3}, \mathcal{L}^{(2)}_{DB3}, \boldsymbol{t}^{(2)}_{DB3})) \in \mathbb{R}^{128\times4\times1\times1} \tag{3.143}$$

を生成する．ここで，$\mathcal{L}^{(2)}_{DB3} \in \mathbb{R}^{128\times64\times1\times4\times1}$ であり，$\boldsymbol{t}^{(2)}_{DB3} = (1, 4, 1)$ である．この段階で全音符単位の特徴抽出を開始し，第 3 モードのサイズが 1 になる．Drums Block 3 は処理をこれで終了し，$\mathbf{Y}^{(2)}_{DB3}$ を Drums Block の出力 $\mathbf{Y}_{DB}$ として出力する : $\mathbf{Y}_{DB} = \mathbf{Y}^{(2)}_{DB3}$. これが，ドラム打楽器から抽出した特徴量であり，Percussion Block に入力される．Drums Block 3 のパラメータをまとめて表 3.41 に示す．

表 3.41: Drums Block 3 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{DB3} \in \mathbb{R}^{64\times128\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{DB3} = (1, 1, 1)$
第 2 層	$\mathcal{L}^{(2)}_{DB3} \in \mathbb{R}^{128\times64\times1\times4\times4}$	$\boldsymbol{t}^{(2)}_{DB3} = (1, 4, 1)$

3.9.2 Other Percussion Block

Other Percussion Block も Drums Block と同様に，3 個のサブブロックで構成する．それらを Other Percussion Block 1，Other Percussion Block 2，および Other Percussion

Block 3 と呼ぶことにする．Other Percussion Block 1 は，表 3.42 に示す 2 層の畳み込み層で構成される．第 1 畳み込み層は $\mathbf{X}_{OtherPercussion} \in \mathbb{R}^{1\times4\times96\times15}$ から

$$\mathbf{Y}^{(1)}_{OPB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{OtherPercussion}, \mathcal{L}^{(1)}_{OPB1}, \boldsymbol{t}^{(1)}_{OPB1})) \in \mathbb{R}^{32\times4\times96\times1} \tag{3.144}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{OPB1} \in \mathbb{R}^{32\times1\times1\times1\times15}$ であり，$\boldsymbol{t}^{(1)}_{OPB1} = (1, 1, 15)$ である．この段階で第 4 モードのサイズが 1 になり，各音色の特徴抽出を開始する．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{OPB1} \in \mathbb{R}^{32\times4\times4\times15}$ から

$$\mathbf{Y}^{(2)}_{OPB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{OPB1}, \mathcal{L}^{(2)}_{OPB1}, \boldsymbol{t}^{(2)}_{OPB1})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.145}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{OPB1} \in \mathbb{R}^{64\times32\times1\times24\times1}$ であり，$\boldsymbol{t}^{(2)}_{OPB1} = (1, 24, 1)$ 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抽出を開始する．以上のように，Other Percussion 1 では，第 4 モードの特徴抽出を行ってから第 3 モードの特徴抽出を行う．

表 3.42: Other Percussion Block 1 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{OPB1} \in \mathbb{R}^{32\times1\times1\times1\times15}$	$\boldsymbol{t}^{(1)}_{OPB1} = (1, 1, 15)$
第 2 層	$\mathcal{L}^{(2)}_{OPB1} \in \mathbb{R}^{64\times32\times1\times24\times1}$	$\boldsymbol{t}^{(2)}_{OPB1} = (1, 24, 1)$

一方，Other Percussion Block 2 は同じサイズの特徴量テンソルを逆の手順で抽出する．すなわち，第 3 モードの特徴量を抽出してから第 4 モードの特徴抽出を行う．第 1 畳み込み層は $\mathbf{X}_{OtherPercussion} \in \mathbb{R}^{1\times4\times96\times15}$ から

$$\mathbf{Y}^{(1)}_{OPB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{OtherPercussion}, \mathcal{L}^{(1)}_{OPB2}, \boldsymbol{t}^{(1)}_{OPB2})) \in \mathbb{R}^{32\times4\times4\times15} \tag{3.146}$$

を抽出する．ここで，$\mathcal{L}^{(1)}_{OPB2} \in \mathbb{R}^{32\times1\times1\times24\times1}$ であり，$\boldsymbol{t}^{(1)}_{OPB2} = (1, 24, 1)$ 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抽出を開始する．次に，第 2 畳み込み層は $\mathbf{Y}^{(1)}_{OPB2} \in \mathbb{R}^{32\times4\times4\times15}$ から

$$\mathbf{Y}^{(2)}_{OPB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{OPB2}, \mathcal{L}^{(2)}_{OPB2}, \boldsymbol{t}^{(2)}_{OPB2})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.147}$$

を抽出する．ここで，$\mathcal{L}^{(2)}_{OPB2} \in \mathbb{R}^{64\times32\times1\times1\times15}$ であり，$\boldsymbol{t}^{(2)}_{OPB2} = (1, 1, 15)$ である．この段階で第 4 モードのサイズが 1 になり，各音色の特徴抽出を開始する．

表 3.43: Other Percussion Block 2 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{OPB2} \in \mathbb{R}^{32\times1\times1\times24\times1}$	$\boldsymbol{t}^{(1)}_{OPB2} = (1, 24, 1)$
第 2 層	$\mathcal{L}^{(2)}_{OPB2} \in \mathbb{R}^{64\times32\times1\times1\times15}$	$\boldsymbol{t}^{(2)}_{OPB2} = (1, 1, 15)$

Other Percussion Block 3 では，Other Percussion Block 1 と Other Percussion Block 2 の出力から特徴抽出を以下のように行う．まず，入力された特徴 $\mathbf{Y}^{(2)}_{OPB1} \in \mathbb{R}^{64\times4\times4\times1}$，$\mathbf{Y}^{(2)}_{OPB2} \in \mathbb{R}^{64\times4\times4\times1}$ を

$$\mathbf{Y}^{(0)}_{OPB3} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{OPB1}, \mathbf{Y}^{(2)}_{OPB2}) \in \mathbb{R}^{128\times4\times4\times1} \tag{3.148}$$

によって結合する．続いて，2 段の畳み込み層によって処理を継続する．まず，第 1 畳み込み層が $\mathbf{Y}^{(0)}_{OPB3} \in \mathbb{R}^{128\times4\times4\times1}$ から

$$\mathbf{Y}^{(1)}_{OPB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{OPB3}, \mathcal{L}^{(1)}_{OPB3}, \boldsymbol{t}^{(1)}_{OPB3})) \in \mathbb{R}^{64\times4\times4\times1} \tag{3.149}$$

を生成する．ここで，$\mathcal{L}^{(1)}_{OPB3} \in \mathbb{R}^{64\times128\times1\times1\times1}$ であり，$\boldsymbol{t}^{(1)}_{OPB3} = (1, 1, 1)$ である．次に，第 2 畳み込み層が $\mathbf{Y}^{(1)}_{OPB3} \in \mathbb{R}^{64\times4\times4\times1}$ から

$$\mathbf{Y}^{(2)}_{OPB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{OPB3}, \mathcal{L}^{(2)}_{OPB3}, \boldsymbol{t}^{(2)}_{OPB3})) \in \mathbb{R}^{128\times4\times1\times1} \tag{3.150}$$

を生成する．ここで，$\mathcal{L}^{(2)}_{OPB3} \in \mathbb{R}^{128\times64\times1\times4\times1}$ であり，$\boldsymbol{t}^{(2)}_{OPB3} = (1, 4, 1)$ である．この段階で全音符単位の特徴抽出を開始し，第 3 モードのサイズが 1 になる．Other Percussion Block 3 は処理をこれで終了し，$\mathbf{Y}^{(2)}_{OPB3}$ を Other Percussion Block の出力 $\mathbf{Y}_{OPB}$ として出力する：$\mathbf{Y}_{OPB} = \mathbf{Y}^{(2)}_{OPB3}$．これが，ドラム打楽器以外の打楽器から抽出した特徴量であり，Percussion Block に入力される．Other Percussion Block 3 のパラメータをまとめて表 3.44 に示す．

表 3.44: Other Percussion Block 3 の畳み込み層における各種パラメータ

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	$\mathcal{L}^{(1)}_{OPB3} \in \mathbb{R}^{64\times128\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{OPB3} = (1, 1, 1)$
第 2 層	$\mathcal{L}^{(2)}_{OPB3} \in \mathbb{R}^{128\times64\times1\times4\times4}$	$\boldsymbol{t}^{(2)}_{OPB3} = (1, 4, 1)$

3.10 Percussion Block

Percussion Block では，打楽器全体に共通する特徴量の抽出を行う．まず，入力された特徴 $\mathbf{Y}_{DB} \in \mathbb{R}^{128\times4\times1\times1}$，$\mathbf{Y}_{OPB} \in \mathbb{R}^{128\times4\times1\times1}$ を

$$\mathbf{Y}_{PB}^{(0)} = \mathrm{Concat}(4, \mathbf{Y}_{DPB}, \mathbf{Y}_{OPB}) \in \mathbb{R}^{128\times4\times1\times2} \tag{3.151}$$

によって結合する．続いて，畳み込み層が $\mathbf{Y}_{PB}^{(0)} \in \mathbb{R}^{128\times4\times1\times2}$ から

$$\mathbf{Y}_{PB}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{PB}^{(0)}, \mathcal{L}_{PB}^{(1)}, \boldsymbol{t}_{PB}^{(1)})) \in \mathbb{R}^{128\times1\times1\times1} \tag{3.152}$$

を生成する．ここで，$\mathcal{L}_{PB}^{(1)} \in \mathbb{R}^{128\times128\times4\times1\times2}$ であり，$\boldsymbol{t}_{PB}^{(1)} = (4, 1, 2)$ である．この段階で，第 1 モードのサイズが 1，第 2 モードのサイズが 1 となり，打楽器全体の特徴抽出を開始する．Percussion Block は処理をこれで終了し，$\mathbf{Y}_{PB}^{(1)}$ を Percussion Block の出力 $\mathbf{Y}_{PB}$ として出力する：$\mathbf{Y}_{PB} = \mathbf{Y}_{PB}^{(1)}$．これが，打楽器全体から抽出した特徴量であり，Merged Block 3 に入力される．Percussion Block のパラメータをまとめて表 3.45 に示す．

表 3.45: Percussion Block の畳み込み層における各種パラメータ

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	$\mathcal{L}_{PB3}^{(1)} \in \mathbb{R}^{128\times128\times4\times1\times2}$	$\boldsymbol{t}_{PB3}^{(1)} = (4, 1, 2)$

3.11 Merged Block 3

Merged Block 3 では楽曲全体に対する評価値を出力する．まず，入力された特徴 $\mathbf{Y}_{MB1} \in \mathbb{R}^{128\times1\times1\times1}$，$\mathbf{Y}_{MB2} \in \mathbb{R}^{128\times1\times1\times1}$，$\mathbf{Y}_{PB} \in \mathbb{R}^{128\times1\times1\times1}$ を

$$\boldsymbol{y}_{MB3}^{(0)} = \mathrm{Reshape}_{MB3}(\mathrm{Concat}(1, \mathbf{Y}_{MB1}, \mathbf{Y}_{MB2}, \mathbf{Y}_{PB})) \in \mathbb{R}^{384} \tag{3.153}$$

によって結合する．ここで，$\mathrm{Reshape}_{MB3}$ は $384\times1\times1\times1$ の 4 階テンソルを 384 次元ベクトルへと並び替える関数である．続いて，出力ベクトル $\boldsymbol{y}_{MB3}^{(0)} \in \mathbb{R}^{384}$ から，

$$\boldsymbol{y}_{MB3}^{(1)} = W_{MB3}\boldsymbol{y}_{MB3}^{(0)} + \boldsymbol{b}_{MB3} \in \mathbb{R}^{128} \tag{3.154}$$

を生成する．ここで，$W_{MB3} \in \mathbb{R}^{128\times384}$ は矩形行列であり，$\boldsymbol{b}_{MB3} \in \mathbb{R}^{128}$ はバイアスベクトルである．

最後に，$\boldsymbol{y}_{MB3}^{(1)}$ の各成分の平均値を算出し， Merged Block 3 の出力 y_{TP} として出力する：

$$y_{TP} = \frac{1}{128} \sum_{i=1}^{128} \boldsymbol{y}_{MB3}^{(1)}[i] \tag{3.155}$$

ここで，i はベクトル $\boldsymbol{y}_{MB3}^{(1)}$ の成分を示すインデックスである．また，これが楽曲に対する評価値であり，判別器の最終的な出力 $\mathcal{D}_{TP}(\boldsymbol{X})$ となる．

3.12 結び

本章は，判別器を構成する各種ブロックの詳細を述べた．