第1章　判別器の構成ブロック

1.1　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

$$\mathbf{X}_{Piano} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Pi}) \in \mathbb{R}^{1\times4\times96\times84} \tag{1.1}$$

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

$$\mathbf{X}_{Guitar} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Gu}) \in \mathbb{R}^{1\times4\times96\times84} \tag{1.2}$$

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

$$\mathbf{X}_{Bass} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Ba}) \in \mathbb{R}^{1\times4\times96\times84} \tag{1.3}$$

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

$$\mathbf{X}_{Strings} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{St}) \in \mathbb{R}^{1\times4\times96\times84} \tag{1.4}$$

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

$$\mathbf{X}_{Percussion} = \mathrm{Reshape}_{expand}(\boldsymbol{X}_{Pe}) \in \mathbb{R}^{1\times4\times96\times84} \tag{1.5}$$

へと変換する．

1.2 Individual Blocks

1.2.1 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 1.1 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

$$\mathbf{Y}^{(1)}_{PiB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB1}, \boldsymbol{t}^{(1)}_{PiB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{1.6}$$

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

$$\mathbf{Y}^{(2)}_{PiB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB1}, \mathcal{L}^{(2)}_{PiB1}, \boldsymbol{t}^{(2)}_{PiB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.7}$$

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 1.1: 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

$$\mathbf{Y}^{(1)}_{PiB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB2}, \boldsymbol{t}^{(1)}_{PiB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.8}$$

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

$$\mathbf{Y}^{(2)}_{PiB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB2}, \mathcal{L}^{(2)}_{PiB2}, \boldsymbol{t}^{(2)}_{PiB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.9}$$

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 1.2 にまとめて示す．

表 1.2: 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 1.3 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

$$\mathbf{Y}^{(1)}_{PiB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB3}, \boldsymbol{t}^{(1)}_{PiB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.10}$$

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

$$\mathbf{Y}^{(2)}_{PiB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB3}, \mathcal{L}^{(2)}_{PiB3}, \boldsymbol{t}^{(2)}_{PiB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.11}$$

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

$$\mathbf{Y}^{(3)}_{PiB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiB3}, \mathcal{L}^{(3)}_{PiB3}, \boldsymbol{t}^{(3)}_{PiB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.12}$$

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 1.3 にまとめて示す．

表 1.3: 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 1.3 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

$$\mathbf{Y}^{(1)}_{PiB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Piano}, \mathcal{L}^{(1)}_{PiB4}, \boldsymbol{t}^{(1)}_{PiB4})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.13}$$

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

$$\mathbf{Y}^{(2)}_{PiB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiB4}, \mathcal{L}^{(2)}_{PiB4}, \boldsymbol{t}^{(2)}_{PiB4})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.14}$$

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

$$\mathbf{Y}^{(3)}_{PiB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiB4}, \mathcal{L}^{(3)}_{PiB4}, \boldsymbol{t}^{(3)}_{PiB4})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.15}$$

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 1.4 にまとめて示す．

表 1.4: 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

$$\mathbf{Y}^{(0)}_{PiB5} = \mathrm{Concat}(1, \mathbf{Y}^{(3)}_{PiB3}, \mathbf{Y}^{(3)}_{PiB4}) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.16}$$

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

$$\mathbf{Y}^{(1)}_{PiB5} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{PiB5}, \mathcal{L}^{(1)}_{PiB5}, \boldsymbol{t}^{(1)}_{PiB5})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.17}$$

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 1.5 にまとめて示す．

表 1.5: 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

$$\mathbf{Y}^{(0)}_{PiB6} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{PiB1}, \mathbf{Y}^{(2)}_{PiB2}, \mathbf{Y}^{(1)}_{PiB5}) \in \mathbb{R}^{160\times4\times16\times7} \tag{1.18}$$

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

$$\mathbf{Y}^{(1)}_{PiB6} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{PiB6}, \mathcal{L}^{(1)}_{PiB6}, \boldsymbol{t}^{(1)}_{PiB6})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.19}$$

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 1.6 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 1.6: 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	$\mathcal{L}^{(1)}_{PiB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{PiB6} = (1,1,1)$

1.2.2 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 1.7 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

$$\mathbf{Y}^{(1)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB1}, \boldsymbol{t}^{(1)}_{GB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{1.20}$$

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

$$\mathbf{Y}^{(2)}_{GB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB1}, \mathcal{L}^{(2)}_{GB1}, \boldsymbol{t}^{(2)}_{GB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.21}$$

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 1.7: 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

$$\mathbf{Y}^{(1)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB2}, \boldsymbol{t}^{(1)}_{GB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.22}$$

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

$$\mathbf{Y}^{(2)}_{GB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB2}, \mathcal{L}^{(2)}_{GB2}, \boldsymbol{t}^{(2)}_{GB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.23}$$

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 1.8 にまとめて示す．

表 1.8: 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 1.9 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

$$\mathbf{Y}^{(1)}_{GB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB3}, \boldsymbol{t}^{(1)}_{GB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.24}$$

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

$$\mathbf{Y}^{(2)}_{GB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB3}, \mathcal{L}^{(2)}_{GB3}, \boldsymbol{t}^{(2)}_{GB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.25}$$

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

$$\mathbf{Y}^{(3)}_{GB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{GB3}, \mathcal{L}^{(3)}_{GB3}, \boldsymbol{t}^{(3)}_{GB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.26}$$

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 1.9 にまとめて示す.

表 1.9: 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 1.9 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

$$\mathbf{Y}^{(1)}_{GB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Guitar}, \mathcal{L}^{(1)}_{GB4}, \boldsymbol{t}^{(1)}_{GB4})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.27}$$

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

$$\mathbf{Y}^{(2)}_{GB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GB4}, \mathcal{L}^{(2)}_{GB4}, \boldsymbol{t}^{(2)}_{GB4})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.28}$$

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

$$\mathbf{Y}^{(3)}_{GB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{GB4}, \mathcal{L}^{(3)}_{GB4}, \boldsymbol{t}^{(3)}_{GB4})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.29}$$

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 1.10 にまとめて示す．

表 1.10: 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

$$\mathbf{Y}^{(0)}_{GB5} = \mathrm{Concat}(1, \mathbf{Y}^{(3)}_{GB3}, \mathbf{Y}^{(3)}_{GB4}) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.30}$$

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

$$\mathbf{Y}^{(1)}_{GB5} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{GB5}, \mathcal{L}^{(1)}_{GB5}, \boldsymbol{t}^{(1)}_{GB5})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.31}$$

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 1.11 にまとめて示す．

表 1.11: 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

$$\mathbf{Y}^{(0)}_{GB6} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{GB1}, \mathbf{Y}^{(2)}_{GB2}, \mathbf{Y}^{(1)}_{GB5}) \in \mathbb{R}^{160\times4\times16\times7} \tag{1.32}$$

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

$$\mathbf{Y}^{(1)}_{GB6} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{GB6}, \mathcal{L}^{(1)}_{GB6}, \boldsymbol{t}^{(1)}_{GB6})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.33}$$

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 1.12 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 1.12: 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	$\mathcal{L}^{(1)}_{GB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{GB6} = (1,1,1)$

1.2.3 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 1.13 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

$$\mathbf{Y}^{(1)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB1}, \boldsymbol{t}^{(1)}_{BaB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{1.34}$$

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

$$\mathbf{Y}^{(2)}_{BaB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB1}, \mathcal{L}^{(2)}_{BaB1}, \boldsymbol{t}^{(2)}_{BaB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.35}$$

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 1.13: 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

$$\mathbf{Y}^{(1)}_{BaB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB2}, \boldsymbol{t}^{(1)}_{BaB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.36}$$

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

$$\mathbf{Y}^{(2)}_{BaB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB2}, \mathcal{L}^{(2)}_{BaB2}, \boldsymbol{t}^{(2)}_{BaB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.37}$$

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 1.14 にまとめて示す．

表 1.14: 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 1.15 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

$$\mathbf{Y}^{(1)}_{BaB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB3}, \boldsymbol{t}^{(1)}_{BaB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.38}$$

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

$$\mathbf{Y}^{(2)}_{BaB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB3}, \mathcal{L}^{(2)}_{BaB3}, \boldsymbol{t}^{(2)}_{BaB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.39}$$

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

$$\mathbf{Y}^{(3)}_{BaB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{BaB3}, \mathcal{L}^{(3)}_{BaB3}, \boldsymbol{t}^{(3)}_{BaB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.40}$$

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 1.15 にまとめて示す.

表 1.15: 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 1.15 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

$$\mathbf{Y}^{(1)}_{BaB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Bass}, \mathcal{L}^{(1)}_{BaB4}, \boldsymbol{t}^{(1)}_{BaB4})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.41}$$

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

$$\mathbf{Y}^{(2)}_{BaB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaB4}, \mathcal{L}^{(2)}_{BaB4}, \boldsymbol{t}^{(2)}_{BaB4})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.42}$$

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

$$\mathbf{Y}^{(3)}_{BaB4} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{BaB4}, \mathcal{L}^{(3)}_{BaB4}, \boldsymbol{t}^{(3)}_{BaB4})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.43}$$

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 1.16 にまとめて示す.

表 1.16: 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

$$\mathbf{Y}^{(0)}_{BaB5} = \mathrm{Concat}(1, \mathbf{Y}^{(3)}_{BaB3}, \mathbf{Y}^{(3)}_{BaB4}) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.44}$$

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

$$\mathbf{Y}^{(1)}_{BaB5} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{BaB5}, \mathcal{L}^{(1)}_{BaB5}, \boldsymbol{t}^{(1)}_{BaB5})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.45}$$

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 1.17 にまとめて示す.

表 1.17: 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

$$\mathbf{Y}^{(0)}_{BaB6} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{BaB1}, \mathbf{Y}^{(2)}_{BaB2}, \mathbf{Y}^{(1)}_{BaB5}) \in \mathbb{R}^{160\times4\times16\times7} \tag{1.46}$$

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

$$\mathbf{Y}^{(1)}_{BaB6} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{BaB6}, \mathcal{L}^{(1)}_{BaB6}, \boldsymbol{t}^{(1)}_{BaB6})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.47}$$

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 1.18 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 1.18: 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	$\mathcal{L}^{(1)}_{BaB6} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}^{(1)}_{BaB6} = (1,1,1)$

1.2.4 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 1.19 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

$$\mathbf{Y}^{(1)}_{SB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}^{(1)}_{SB1}, \boldsymbol{t}^{(1)}_{SB1})) \in \mathbb{R}^{32\times4\times96\times7} \tag{1.48}$$

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

$$\mathbf{Y}^{(2)}_{SB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SB1}, \mathcal{L}^{(2)}_{SB1}, \boldsymbol{t}^{(2)}_{SB1})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.49}$$

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 1.19: 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

$$\mathbf{Y}^{(1)}_{SB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}^{(1)}_{SB2}, \boldsymbol{t}^{(1)}_{SB2})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.50}$$

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

$$\mathbf{Y}^{(2)}_{SB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SB2}, \mathcal{L}^{(2)}_{SB2}, \boldsymbol{t}^{(2)}_{SB2})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.51}$$

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 1.20 にまとめて示す.

表 1.20: 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 1.21 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

$$\mathbf{Y}^{(1)}_{SB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}^{(1)}_{SB3}, \boldsymbol{t}^{(1)}_{SB3})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.52}$$

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

$$\mathbf{Y}^{(2)}_{SB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SB3}, \mathcal{L}^{(2)}_{SB3}, \boldsymbol{t}^{(2)}_{SB3})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.53}$$

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

$$\mathbf{Y}^{(3)}_{SB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{SB3}, \mathcal{L}^{(3)}_{SB3}, \boldsymbol{t}^{(3)}_{SB3})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.54}$$

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 1.21 にまとめて示す.

表 1.21: 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 1.21 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

$$\mathbf{Y}_{SB4}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Strings}, \mathcal{L}_{SB4}^{(1)}, \boldsymbol{t}_{SB4}^{(1)})) \in \mathbb{R}^{32\times4\times96\times84} \tag{1.55}$$

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

$$\mathbf{Y}_{SB4}^{(2)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB4}^{(1)}, \mathcal{L}_{SB4}^{(2)}, \boldsymbol{t}_{SB4}^{(2)})) \in \mathbb{R}^{32\times4\times16\times84} \tag{1.56}$$

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

$$\mathbf{Y}_{SB4}^{(3)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB4}^{(2)}, \mathcal{L}_{SB4}^{(3)}, \boldsymbol{t}_{SB4}^{(3)})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.57}$$

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 1.22 にまとめて示す.

表 1.22: 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

$$\mathbf{Y}_{SB5}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{SB3}^{(3)}, \mathbf{Y}_{SB4}^{(3)}) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.58}$$

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

$$\mathbf{Y}_{SB5}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB5}^{(0)}, \mathcal{L}_{SB5}^{(1)}, \boldsymbol{t}_{SB5}^{(1)})) \in \mathbb{R}^{32\times4\times16\times7} \tag{1.59}$$

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 1.23 にまとめて示す.

表 1.23: 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

$$\mathbf{Y}_{SB6}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{SB1}^{(2)}, \mathbf{Y}_{SB2}^{(2)}, \mathbf{Y}_{SB5}^{(1)}) \in \mathbb{R}^{160\times4\times16\times7} \tag{1.60}$$

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

$$\mathbf{Y}_{SB6}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{SB6}^{(0)}, \mathcal{L}_{SB6}^{(1)}, \boldsymbol{t}_{SB6}^{(1)})) \in \mathbb{R}^{64\times4\times16\times7} \tag{1.61}$$

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 1.24 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 1.24: 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	$\mathcal{L}_{SB6}^{(1)} \in \mathbb{R}^{64\times160\times1\times1\times1}$	$\boldsymbol{t}_{SB6}^{(1)} = (1, 1, 1)$

1.3 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

$$\mathbf{Y}_{BB} = \mathrm{Concat}(1, \mathbf{Y}_{PiB}, \mathbf{Y}_{GB}, \mathbf{Y}_{BaB}, \mathbf{Y}_{SB}) \in \mathbb{R}^{256\times4\times16\times7} \tag{1.62}$$

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．

1.3.1 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

$$\mathbf{Y}_{LB}^{(0)} = \mathrm{Slice}(\mathbf{Y}_{BB}, 4, 1, 2) \in \mathbb{R}^{256\times4\times16\times2} \tag{1.63}$$

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

$$\mathbf{Y}_{LB}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{LB}^{(0)}, \mathcal{L}_{LB}^{(1)}, \boldsymbol{t}_{LB}^{(1)})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.64}$$

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 1.25 に示す．

表 1.25: 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	$\mathcal{L}_{LB}^{(1)} \in \mathbb{R}^{64\times256\times1\times4\times2}$	$\boldsymbol{t}_{LB}^{(1)} = (1, 4, 2)$

1.3.2 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

$$\mathbf{Y}_{MiB}^{(0)} = \mathrm{Slice}(\mathbf{Y}_{BB}, 4, 3, 4) \in \mathbb{R}^{256\times4\times16\times2} \tag{1.65}$$

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

$$\mathbf{Y}^{(1)}_{MiB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{MiB}, \mathcal{L}^{(1)}_{MiB}, \boldsymbol{t}^{(1)}_{MiB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.66}$$

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 1.26 に示す．

表 1.26: 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	$\mathcal{L}^{(1)}_{MiB} \in \mathbb{R}^{64\times256\times1\times4\times2}$	$\boldsymbol{t}^{(1)}_{MiB} = (1, 4, 2)$

1.3.3 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

$$\mathbf{Y}^{(0)}_{HB} = \mathrm{Slice}(\mathbf{Y}_{BB}, 4, 5, 7) \in \mathbb{R}^{256\times4\times16\times3} \tag{1.67}$$

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

$$\mathbf{Y}^{(1)}_{HB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{HB}, \mathcal{L}^{(1)}_{HB}, \boldsymbol{t}^{(1)}_{HB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.68}$$

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 1.27 に示す．

表 1.27: 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	$\mathcal{L}^{(1)}_{HB} \in \mathbb{R}^{64\times256\times1\times4\times3}$	$\boldsymbol{t}^{(1)}_{HB} = (1, 4, 3)$

1.4 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

$$\mathbf{Y}_{TB}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{LB}, \mathbf{Y}_{MiB}, \mathbf{Y}_{HB}) \in \mathbb{R}^{192\times4\times4\times1} \tag{1.69}$$

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

$$\mathbf{Y}_{TB}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{TB}^{(0)}, \mathcal{L}_{TB}^{(1)}, \boldsymbol{t}_{TB}^{(1)})) \in \mathbb{R}^{128\times4\times4\times1} \tag{1.70}$$

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

$$\mathbf{Y}_{TB}^{(2)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{TB}^{(1)}, \mathcal{L}_{TB}^{(2)}, \boldsymbol{t}_{TB}^{(2)})) \in \mathbb{R}^{128\times4\times1\times1} \tag{1.71}$$

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 1.28 に示す．

表 1.28: 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	$\mathcal{L}_{TB}^{(1)} \in \mathbb{R}^{128\times128\times1\times4\times1}$	$\boldsymbol{t}_{TB}^{(1)} = (1,4,1)$

1.5 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

$$\mathbf{Y}_{Chroma}^{(0)} = \mathrm{Concat}(1, \mathbf{X}_{Piano}, \mathbf{X}_{Guitar}, \mathbf{X}_{Bass}, \mathbf{X}_{Strings}) \in \mathbb{R}^{4\times4\times96\times84} \tag{1.72}$$

によって結合する．続いて，

$$\mathscr{Y}_{Chroma}^{(1)} = \mathrm{Reshape}_{Chroma}(\mathbf{Y}_{Chroma}^{(0)}) \in \mathbb{R}^{4\times4\times8\times12\times12\times7} \tag{1.73}$$

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

$$\boldsymbol{Y}_{Chroma} = \frac{1}{4\cdot12\cdot7}\sum_{i=1}^{4}\sum_{j=1}^{12}\sum_{k=1}^{7}\mathscr{Y}^{(1)}_{Chroma}[i,:,:,j,:,k] \in \mathbb{R}^{4\times8\times12} \tag{1.74}$$

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

$$\mathbf{Y}^{(0)}_{CB} = \mathrm{Reshape}_{CB}(\boldsymbol{Y}_{Chroma}) \in \mathbb{R}^{1\times4\times8\times12} \tag{1.75}$$

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

$$\mathbf{Y}^{(1)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{CB}, \mathcal{L}^{(1)}_{CB}, \boldsymbol{t}^{(1)}_{CB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.76}$$

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

$$\mathbf{Y}^{(2)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{CB}, \mathcal{L}^{(2)}_{CB}, \boldsymbol{t}^{(2)}_{CB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.77}$$

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

$$\mathbf{Y}^{(3)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{CB}, \mathcal{L}^{(3)}_{CB}, \boldsymbol{t}^{(3)}_{CB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.78}$$

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

$$\mathbf{Y}^{(4)}_{CB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{CB}, \mathcal{L}^{(4)}_{CB}, \boldsymbol{t}^{(4)}_{CB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.79}$$

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 1.29 に示す．

表 1.29: 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	$\mathcal{L}_{CB}^{(4)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{CB}^{(4)} = (1, 2, 1)$

1.6 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

$$\mathbf{Y}_{M1}^{(0)} = \mathrm{Concat}(1, \mathbf{Y}_{TB}, \mathbf{Y}_{CB}) \in \mathbb{R}^{192\times4\times1\times1} \tag{1.80}$$

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

$$\mathbf{Y}_{MB1}^{(1)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{MB1}^{(0)}, \mathcal{L}_{MB1}^{(1)}, \boldsymbol{t}_{MB1}^{(1)})) \in \mathbb{R}^{128\times1\times1\times1} \tag{1.81}$$

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 1.30 に示す．

表 1.30: 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	$\mathcal{L}_{MB1}^{(1)} \in \mathbb{R}^{128\times192\times4\times1\times1}$	$\boldsymbol{t}_{MB1}^{(1)} = (4, 1, 1)$

1.7 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．

1.7.1 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．

$$\boldsymbol{Y}_{PianoLPoly} = \frac{1}{36}\sum_{i=1}^{36}\mathbf{X}_{Piano}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{1.82}$$

$$\mathbf{Y}_{PianoLPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{PianoLPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{1.83}$$

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

$$\mathbf{Y}^{(1)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{PianoLPoly}, \mathcal{L}^{(1)}_{PiLPB}, \boldsymbol{t}^{(1)}_{PiLPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{1.84}$$

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

$$\mathbf{Y}^{(2)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiLPB}, \mathcal{L}^{(2)}_{PiLPB}, \boldsymbol{t}^{(2)}_{PiLPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{1.85}$$

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

$$\mathbf{Y}^{(3)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiLPB}, \mathcal{L}^{(3)}_{PiLPB}, \boldsymbol{t}^{(3)}_{PiLPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.86}$$

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

$$\mathbf{Y}^{(4)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{PiLPB}, \mathcal{L}^{(4)}_{PiLPB}, \boldsymbol{t}^{(4)}_{PiLPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.87}$$

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

$$\mathbf{Y}^{(5)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{PiLPB}, \mathcal{L}^{(5)}_{PiLPB}, \boldsymbol{t}^{(5)}_{PiLPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.88}$$

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

$$\mathbf{Y}^{(6)}_{PiLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{PiLPB}, \mathcal{L}^{(6)}_{PiLPB}, \boldsymbol{t}^{(6)}_{PiLPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.89}$$

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 1.31 に示す．

表 1.31: 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．

$$\boldsymbol{Y}_{PianoHPoly} = \frac{1}{48}\sum_{i=37}^{84} \mathbf{X}_{Piano}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{1.90}$$

$$\mathbf{Y}_{PianoHPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{PianoHPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{1.91}$$

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

$$\mathbf{Y}^{(1)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{PianoHPoly}, \mathcal{L}^{(1)}_{PiHPB}, \boldsymbol{t}^{(1)}_{PiHPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{1.92}$$

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

$$\mathbf{Y}^{(2)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{PiHPB}, \mathcal{L}^{(2)}_{PiHPB}, \boldsymbol{t}^{(2)}_{PiHPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{1.93}$$

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

$$\mathbf{Y}^{(3)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{PiHPB}, \mathcal{L}^{(3)}_{PiHPB}, \boldsymbol{t}^{(3)}_{PiHPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.94}$$

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

$$\mathbf{Y}^{(4)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{PiHPB}, \mathcal{L}^{(4)}_{PiHPB}, \boldsymbol{t}^{(4)}_{PiHPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.95}$$

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

$$\mathbf{Y}^{(5)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{PiHPB}, \mathcal{L}^{(5)}_{PiHPB}, \boldsymbol{t}^{(5)}_{PiHPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.96}$$

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

$$\mathbf{Y}^{(6)}_{PiHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{PiHPB}, \mathcal{L}^{(6)}_{PiHPB}, \boldsymbol{t}^{(6)}_{PiHPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.97}$$

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 1.32 に示す.

表 1.32: 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

$$\mathbf{Y}_{PiPB} = \mathrm{Concat}(4, \mathbf{Y}^{(6)}_{PiLPB}, \mathbf{Y}^{(6)}_{PiHPB}) \in \mathbb{R}^{128\times4\times1\times1} \tag{1.98}$$

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．

1.7.2 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．

$$\boldsymbol{Y}_{GuitarPoly} = \frac{1}{84}\sum_{i=1}^{84} \mathbf{X}_{Guitar}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{1.99}$$

$$\mathbf{Y}_{GuitarPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{GuitarPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{1.100}$$

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

$$\mathbf{Y}^{(1)}_{GPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GuitarPoly}, \mathcal{L}^{(1)}_{GPB}, \boldsymbol{t}^{(1)}_{GPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{1.101}$$

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

$$\mathbf{Y}^{(2)}_{GPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{GPB}, \mathcal{L}^{(2)}_{GPB}, \boldsymbol{t}^{(2)}_{GPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{1.102}$$

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

$$\mathbf{Y}_{GPB}^{(3)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(2)}, \mathcal{L}_{GPB}^{(3)}, \boldsymbol{t}_{GPB}^{(3)})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.103}$$

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

$$\mathbf{Y}_{GPB}^{(4)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(3)}, \mathcal{L}_{GPB}^{(4)}, \boldsymbol{t}_{GPB}^{(4)})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.104}$$

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

$$\mathbf{Y}_{GPB}^{(5)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(4)}, \mathcal{L}_{GPB}^{(5)}, \boldsymbol{t}_{GPB}^{(5)})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.105}$$

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

$$\mathbf{Y}_{GPB}^{(6)} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{GPB}^{(5)}, \mathcal{L}_{GPB}^{(6)}, \boldsymbol{t}_{GPB}^{(6)})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.106}$$

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 1.33 に示す.

表 1.33: 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	$\mathcal{L}_{GPB}^{(6)} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}_{GPB}^{(6)} = (1,2,1)$

1.7.3 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．

$$\boldsymbol{Y}_{BassPoly} = \frac{1}{84}\sum_{i=1}^{84}\mathbf{X}_{Bass}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{1.107}$$

$$\mathbf{Y}_{BassPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{BassPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{1.108}$$

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

$$\mathbf{Y}^{(1)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{BassPoly}, \mathcal{L}^{(1)}_{BaPB}, \boldsymbol{t}^{(1)}_{BaPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{1.109}$$

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

$$\mathbf{Y}^{(2)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{BaPB}, \mathcal{L}^{(2)}_{BaPB}, \boldsymbol{t}^{(2)}_{BaPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{1.110}$$

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

$$\mathbf{Y}^{(3)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{BaPB}, \mathcal{L}^{(3)}_{BaPB}, \boldsymbol{t}^{(3)}_{BaPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.111}$$

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

$$\mathbf{Y}^{(4)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{BaPB}, \mathcal{L}^{(4)}_{BaPB}, \boldsymbol{t}^{(4)}_{BaPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.112}$$

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

$$\mathbf{Y}^{(5)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{BaPB}, \mathcal{L}^{(5)}_{BaPB}, \boldsymbol{t}^{(5)}_{BaPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.113}$$

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

$$\mathbf{Y}^{(6)}_{BaPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{BaPB}, \mathcal{L}^{(6)}_{BaPB}, \boldsymbol{t}^{(6)}_{BaPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.114}$$

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 1.34 に示す．

表 1.34: 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	$\mathcal{L}^{(6)}_{BaPB} \in \mathbb{R}^{64\times64\times1\times2\times1}$	$\boldsymbol{t}^{(6)}_{BaPB} = (1,2,1)$

1.7.4 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．

$$\boldsymbol{Y}_{StringsLPoly} = \frac{1}{36}\sum_{i=1}^{36} \mathbf{X}_{Strings}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{1.115}$$

$$\mathbf{Y}_{StringsLPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{StringsLPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{1.116}$$

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

$$\mathbf{Y}^{(1)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{StringsLPoly}, \mathcal{L}^{(1)}_{SLPB}, \boldsymbol{t}^{(1)}_{SLPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{1.117}$$

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

$$\mathbf{Y}^{(2)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SLPB}, \mathcal{L}^{(2)}_{SLPB}, \boldsymbol{t}^{(2)}_{SLPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{1.118}$$

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

$$\mathbf{Y}^{(3)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{SLPB}, \mathcal{L}^{(3)}_{SLPB}, \boldsymbol{t}^{(3)}_{SLPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.119}$$

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

$$\mathbf{Y}^{(4)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{SLPB}, \mathcal{L}^{(4)}_{SLPB}, \boldsymbol{t}^{(4)}_{SLPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.120}$$

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

$$\mathbf{Y}^{(5)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{SLPB}, \mathcal{L}^{(5)}_{SLPB}, \boldsymbol{t}^{(5)}_{SLPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.121}$$

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

$$\mathbf{Y}^{(6)}_{SLPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{SLPB}, \mathcal{L}^{(6)}_{SLPB}, \boldsymbol{t}^{(6)}_{SLPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.122}$$

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 1.35 に示す.

表 1.35: 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．

$$\boldsymbol{Y}_{StringsHPoly} = \frac{1}{36}\sum_{i=37}^{72}\mathbf{X}_{Strings}[:,:,:,i] \in \mathbb{R}^{1\times4\times96} \tag{1.123}$$

$$\mathbf{Y}_{StringsHPoly} = \mathrm{Reshape}_{Poly}(\boldsymbol{Y}_{StringsHPoly}) \in \mathbb{R}^{1\times4\times96\times1} \tag{1.124}$$

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

$$\mathbf{Y}^{(1)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}_{StringsHPoly}, \mathcal{L}^{(1)}_{SHPB}, \boldsymbol{t}^{(1)}_{SHPB})) \in \mathbb{R}^{32\times4\times32\times1} \tag{1.125}$$

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

$$\mathbf{Y}^{(2)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{SHPB}, \mathcal{L}^{(2)}_{SHPB}, \boldsymbol{t}^{(2)}_{SHPB})) \in \mathbb{R}^{32\times4\times16\times1} \tag{1.126}$$

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

$$\mathbf{Y}^{(3)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(2)}_{SHPB}, \mathcal{L}^{(3)}_{SHPB}, \boldsymbol{t}^{(3)}_{SHPB})) \in \mathbb{R}^{64\times4\times8\times1} \tag{1.127}$$

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

$$\mathbf{Y}^{(4)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(3)}_{SHPB}, \mathcal{L}^{(4)}_{SHPB}, \boldsymbol{t}^{(4)}_{SHPB})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.128}$$

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

$$\mathbf{Y}^{(5)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(4)}_{SHPB}, \mathcal{L}^{(5)}_{SHPB}, \boldsymbol{t}^{(5)}_{SHPB})) \in \mathbb{R}^{64\times4\times2\times1} \tag{1.129}$$

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

$$\mathbf{Y}^{(6)}_{SHPB} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(5)}_{SHPB}, \mathcal{L}^{(6)}_{SHPB}, \boldsymbol{t}^{(6)}_{SHPB})) \in \mathbb{R}^{64\times4\times1\times1} \tag{1.130}$$

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 1.36 に示す.

表 1.36: 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

$$\mathbf{Y}_{SPB} = \mathrm{Concat}(4, \mathbf{Y}^{(1)}_{SLPB}, \mathbf{Y}^{(1)}_{SHPB}) \in \mathbb{R}^{128\times4\times1\times1} \tag{1.131}$$

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．

1.8 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

$$\mathbf{Y}^{(0)}_{MB2} = \mathrm{Concat}(1, \mathbf{Y}_{PiPB}, \mathbf{Y}_{GPB}, \mathbf{Y}_{BaPB}, \mathbf{Y}_{SPB}) \in \mathbb{R}^{384\times4\times1\times1} \tag{1.132}$$

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

$$\mathbf{Y}^{(1)}_{MB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{MB2}, \mathcal{L}^{(1)}_{MB2}, \boldsymbol{t}^{(1)}_{MB2})) \in \mathbb{R}^{128\times4\times1\times1} \tag{1.133}$$

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

$$\mathbf{Y}^{(2)}_{MB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{MB2}, \mathcal{L}^{(2)}_{MB2}, \boldsymbol{t}^{(2)}_{MB2})) \in \mathbb{R}^{128\times1\times1\times1} \tag{1.134}$$

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 1.37 に示す．

表 1.37: 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	$\mathcal{L}^{(2)}_{MB2} \in \mathbb{R}^{128\times128\times4\times1\times1}$	$\boldsymbol{t}^{(2)}_{MB2} = (4, 1, 1)$

1.9 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，

$$\mathbf{X}_{Drums} = \mathrm{Exploit}_D(\mathbf{X}_{Percussion}) \in \mathbb{R}^{1\times4\times96\times15} \tag{1.135}$$

$$\mathbf{X}_{OtherPercussion} = \mathrm{Exploit}_{OP}(\mathbf{X}_{Percussion}) \in \mathbb{R}^{1\times4\times96\times15} \tag{1.136}$$

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 1.38 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 1.38 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 1.38: 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
$\boldsymbol{X}^{39}_{Pe}$	$\boldsymbol{X}^{1}_{OtherPercussion}$
$\boldsymbol{X}^{54}_{Pe}$	$\boldsymbol{X}^{2}_{OtherPercussion}$
$\boldsymbol{X}^{58}_{Pe}$	$\boldsymbol{X}^{3}_{OtherPercussion}$
$\boldsymbol{X}^{60}_{Pe}, \boldsymbol{X}^{61}_{Pe}$	$\boldsymbol{X}^{4}_{OtherPercussion}$
$\boldsymbol{X}^{62}_{Pe}, \boldsymbol{X}^{63}_{Pe}, \boldsymbol{X}^{64}_{Pe}$	$\boldsymbol{X}^{5}_{OtherPercussion}$
$\boldsymbol{X}^{65}_{Pe}, \boldsymbol{X}^{66}_{Pe}$	$\boldsymbol{X}^{6}_{OtherPercussion}$
$\boldsymbol{X}^{67}_{Pe}, \boldsymbol{X}^{68}_{Pe}$	$\boldsymbol{X}^{7}_{OtherPercussion}$
$\boldsymbol{X}^{69}_{Pe}$	$\boldsymbol{X}^{8}_{OtherPercussion}$
$\boldsymbol{X}^{70}_{Pe}$	$\boldsymbol{X}^{9}_{OtherPercussion}$
$\boldsymbol{X}^{71}_{Pe}, \boldsymbol{X}^{72}_{Pe}$	$\boldsymbol{X}^{10}_{OtherPercussion}$
$\boldsymbol{X}^{73}_{Pe}, \boldsymbol{X}^{74}_{Pe}$	$\boldsymbol{X}^{11}_{OtherPercussion}$
$\boldsymbol{X}^{75}_{Pe}$	$\boldsymbol{X}^{12}_{OtherPercussion}$
$\boldsymbol{X}^{76}_{Pe}, \boldsymbol{X}^{77}_{Pe}$	$\boldsymbol{X}^{13}_{OtherPercussion}$
$\boldsymbol{X}^{78}_{Pe}, \boldsymbol{X}^{79}_{Pe}$	$\boldsymbol{X}^{14}_{OtherPercussion}$
$\boldsymbol{X}^{80}_{Pe}, \boldsymbol{X}^{81}_{Pe}$	$\boldsymbol{X}^{15}_{OtherPercussion}$

1.9.1 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 1.39 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

$$\mathbf{Y}^{(1)}_{DB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Drums}, \mathcal{L}^{(1)}_{DB1}, \boldsymbol{t}^{(1)}_{DB1})) \in \mathbb{R}^{32\times4\times96\times1} \tag{1.137}$$

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

$$\mathbf{Y}^{(2)}_{DB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{DB1}, \mathcal{L}^{(2)}_{DB1}, \boldsymbol{t}^{(2)}_{DB1})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.138}$$

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 1.39: 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

$$\mathbf{Y}^{(1)}_{DB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{Drums}, \mathcal{L}^{(1)}_{DB2}, \boldsymbol{t}^{(1)}_{DB2})) \in \mathbb{R}^{32\times4\times4\times15} \tag{1.139}$$

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

$$\mathbf{Y}^{(2)}_{DB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{DB2}, \mathcal{L}^{(2)}_{DB2}, \boldsymbol{t}^{(2)}_{DB2})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.140}$$

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 1.40: 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

$$\mathbf{Y}^{(0)}_{DB3} = \mathrm{Concat}(1, \mathbf{Y}^{(2)}_{DB1}, \mathbf{Y}^{(2)}_{DB2}) \in \mathbb{R}^{128\times4\times4\times1} \tag{1.141}$$

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

$$\mathbf{Y}^{(1)}_{DB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(0)}_{DB3}, \mathcal{L}^{(1)}_{DB3}, \boldsymbol{t}^{(1)}_{DB3})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.142}$$

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

$$\mathbf{Y}^{(2)}_{DB3} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{DB3}, \mathcal{L}^{(2)}_{DB3}, \boldsymbol{t}^{(2)}_{DB3})) \in \mathbb{R}^{128\times4\times1\times1} \tag{1.143}$$

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 1.41 に示す．

表 1.41: 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	$\mathcal{L}^{(2)}_{DB3} \in \mathbb{R}^{128\times64\times1\times4\times4}$	$\boldsymbol{t}^{(2)}_{DB3} = (1, 4, 1)$

1.9.2 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 1.42 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

$$\mathbf{Y}^{(1)}_{OPB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{OtherPercussion}, \mathcal{L}^{(1)}_{OPB1}, \boldsymbol{t}^{(1)}_{OPB1})) \in \mathbb{R}^{32\times4\times96\times1} \tag{1.144}$$

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

$$\mathbf{Y}^{(2)}_{OPB1} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{OPB1}, \mathcal{L}^{(2)}_{OPB1}, \boldsymbol{t}^{(2)}_{OPB1})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.145}$$

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 1.42: 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

$$\mathbf{Y}^{(1)}_{OPB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{X}_{OtherPercussion}, \mathcal{L}^{(1)}_{OPB2}, \boldsymbol{t}^{(1)}_{OPB2})) \in \mathbb{R}^{32\times4\times4\times15} \tag{1.146}$$

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

$$\mathbf{Y}^{(2)}_{OPB2} = \mathrm{LeakyReLU}(\mathrm{Conv}(\mathbf{Y}^{(1)}_{OPB2}, \mathcal{L}^{(2)}_{OPB2}, \boldsymbol{t}^{(2)}_{OPB2})) \in \mathbb{R}^{64\times4\times4\times1} \tag{1.147}$$

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．