

Bab 2

TELAAH PUSTAKA

2.1 Tinjauan Pustaka Penelitian

Penelitian ini dilakukan dengan terlebih dahulu membahas sejumlah teori dan konsep yang berhubungan dengan metode CNN. Landasan teori yang dibahas terdiri dari audio, akor, *bass*, *Beat Tracking*, *Chromagram*, dan Metode CNN.

2.1.1 Audio

Audio adalah sebuah deret waktu, dimana sumbu y adalah amplitudo arus yang sesuai dengan sebuah membran loudspeaker dan sumbu x adalah sumbu yang sesuai dengan satuan waktu data tersebut. Telinga manusia hanya dapat mendengar bunyi dengan rentang frekuensi antara 20 Hz hingga 20 KHz (20.000Hz). Angka 20 Hz sebagai frekuensi suara terendah yang dapat didengar, sedangkan 20 KHz merupakan frekuensi tertinggi yang dapat didengar. Gelombang suara mengandung sejumlah komponen penting, seperti amplitudo, panjang gelombang, dan frekuensi. Komponen-komponen tersebut mampu membuat suara yang satu berbeda dengan suara yang lain.

Amplitudo sendiri merupakan kekuatan atau daya gelombang sebuah sinyal. Nilai amplitudo diinterpretasikan sebagai volume. Semakin besar nilai dari sebuah amplitudo, maka semakin keras suara yang dihasilkan. Sebaliknya, jika nilai dari suatu amplitudo semakin kecil, maka suara yang dihasilkan semakin lemah. Frekuensi adalah jumlah dari siklus yang terjadi dalam satu detik dan memiliki satuan Hertz (Hz). Getaran gelombang suara yang cepat membuat frekuensi semakin tinggi. Misalnya, menyanyi nada tinggi membuat tali suara pada pita suara bergetar secara cepat.

2.1.2 Akor

Akor adalah sebuah kombinasi tiga nada atau lebih yang dibunyikan secara bersamaan. Akor tiga not yang berjarak tiga scale pada root sampai not ketiga disebut triad. Untuk mempermudah pemusik dalam menentukan apakah suatu kombinasi dari tiga nada tertentu merupakan sebuah triad atau bukan, maka pemusik dapat membentuk diagram circle of thirds. Scale atau skala nada merupakan jarak yang menandakan lokasi suatu nada dengan titik mulai tertentu. Terdapat empat jenis triad yang dibedakan berdasarkan skala nadanya, diantaranya :

1. *Major Triad*

Major triad merupakan tiga nada yang dibunyikan dengan skala nada 1-2- 1,5. Akor major menunjukkan identitas suatu bagian lagu lebih dekat dengan root karena ada satu buah akor major yang menempati Scale I.

2. *Minor triad*

Minor triad merupakan tiga nada yang dibunyikan dengan skala 1-1,5-2.

3. *Diminished triad*

Diminished triad merupakan tiga nada yang dibunyikan dengan skala 1- 1,5-1,5.

4. *Augmented triad*

Augmented triad merupakan tiga nada yang dibunyikan dengan skala 1- 1,5-2,5.

Akor *seventh* merupakan akor empat not yang kelas *pitch*nya dapat diatur sebagai sepertiga. Skala nada akor major *seventh* adalah 1-2-1,5-2. Sedangkan skala nada untuk minor *seventh* adalah 1-1,5-2-2. Sama halnya dengan triad, kelas nada yang dimiliki oleh akor *seventh* menempati posisi yang berdekatan (rumpun empat kelas) pada lingkaran pertiga. Keempat anggota akor ketujuh adalah *root*, *third*, *fifth*, dan *seventh*. Terdapat suatu cara untuk melakukan teknik *blocking* pada akor minor *seventh*. Dengan memanfaatkan teknik *blocking*, maka *root* dapat berperan sebagai nada *bass*. Sementara *third*, *fifth*, dan *seventh* sebagai major triad.

2.1.3 *Beat Tracking*

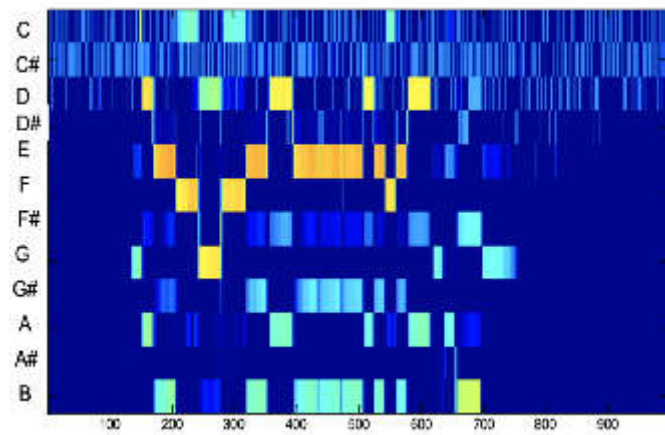
Beat tracking digunakan untuk menentukan contoh waktu dalam rekaman audio, di mana pendengar manusia cenderung mengetuk kakinya ke musik. *Beat tracking* pada rancangan menggunakan metode *dynamic programming*. *Dynamic programming* atau *dynamic optimization* bukan merupakan rumus matematika yang mampu memberikan jawaban dengan hanya sekedar memberikan input. Sebaliknya, *dynamic programming* adalah kombinasi pemikiran terstruktur dan pola pikir analitis untuk menyelesaikan sebuah permasalahan.

Untuk melakukan *beat tracking*, hal yang harus dilakukan terlebih dahulu adalah mencari sinyal yang berupa onset. Onset merupakan lokasi dimana sinyal berada pada nilai yang tinggi secara tiba-tiba. Onset yang ditemukan kemudian dilakukan penandaan. Onset yang ditemukan bisa saja merupakan *false positive*. Untuk mengurangi *false positive* tersebut, maka harus dilakukan perhitungan untuk menemukan urutan umum terpanjang dari onset.

2.1.4 *Chromagram*

Chromagram adalah transformasi properti frekuensi-waktu sinyal menjadi prekursor *pitch* yang berubah-ubah untuk sementara waktu. Transformasi ini didasarkan pada pengamatan persepsi tentang sistem pendengaran dan telah terbukti memiliki beberapa sifat matematika yang menarik. *Chromagram* memperluas konsep *chroma* untuk memasukkan dimensi waktu. Seperti halnya kita menggunakan *spektrogram* untuk menyimpulkan properti tentang distribusi energi sinyal dari frekuensi dan waktu, *chromagram* dapat digunakan untuk menyimpulkan properti tentang distribusi energi sinyal terhadap kroma dan waktu.

Terdapat dua masalah yang muncul dalam mengembangkan konsep *chromagram* ini, yakni definisi dan bagaimana cara menghitung *chromagram*. *Chroma* adalah pemetaan banyak-ke-satu (*many-to-one*) frekuensi, sementara *chromagram* adalah ukuran kekuatan sinyal sebagai fungsi dari *chroma* dan waktu. Oleh karena itu, *chromagram* dapat didefinisikan sebagai pemetaan banyak-ke-satu (*many-to-one*) kekuatan sinyal pada frekuensi milik kelas *chroma* yang sama. Pembagian antara transformasi langsung dari sinyal asli dan transformasi dari gambar frekuensi waktu dapat digunakan untuk menghitung *chromagram*. Contoh *chromagram* dapat dilihat dengan jelas pada Gambar 2.1.



Gambar 2.1: Chromagram dari sebuah alat musik Clarinet

2.1.5 Windowing

Windowing berfungsi untuk meminimalisir sinyal yang tak kontinu pada awal dan akhir masing-masing *frame*. Nilai panjang filter didapatkan dengan menggunakan rumus di bawah. Nilai panjang tersebut akan dipakai untuk membuat respon *impulse* di simulasi dan implementasi. Nilai panjang *filter* juga akan berpengaruh kepada faktor *roll-off* atau faktor kelandaian dari suatu *filter*.

Rumus untuk menghitung *window* dapat dilihat pada persamaan

1.

$$Yt(n) = xt(n)w(n), 0 \leq n \leq N - 1 \quad (1)$$

Rumus untuk menghitung *Hamming Window* dapat dilihat pada persamaan

2.

$$w(n) = 0.54 - 0.46\cos(2N\pi - n1), 0 \leq n \leq N - 1 \quad (2)$$

Keterangan:

N = Panjang setiap frame

$xt(n)$ = Sinyal masukan ke- n pada frame ke- t

$w(n)$ = Fungsi Hamming Window

$Yt(n)$ = Nilai hasil windowing untuk sinyal masukan ke- n pada frame ke- t

Data mentah dari masukan yang pertama kali dibaca oleh bahasa pemrograman Python adalah sinyal dalam domain waktu. Data sinyal dalam domain waktu tersebut perlu diubah ke dalam satuan frekuensi. Proses untuk mengubah data mentah yang berbentuk sinyal tersebut diubah dengan menggunakan metode *Discrete Fourier Transform* (DFT). Pada tahap ini, setiap *frame* yang terdiri dari N sampel dikonversi dari domain waktu ke domain frekuensi. Rumus dari DFT dapat dilihat di dalam persamaan 3.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk}, k = 0, 1, \dots, N-1 \quad (3)$$

Keterangan:

$X(k)$ = Nilai transformasi Fourier (FT) ke- k

$x(n)$ = Input signal ke- n

N = Panjang input signal

Hasil proses windowing pada persamaan 1 kemudian dimasukkan sebagai input signal ke- n , sehingga menghasilkan persamaan 4.

$$Zt(k) = \sum_{n=0}^{N-1} Yt(n) e^{j2\pi nk}, k = 0, 1, \dots, N-1 \quad (4)$$

$$e^{i\theta} = \cos \cos\theta + \sin \sin\theta \quad (5)$$

Berdasarkan rumus Euler pada persamaan 5, dengan e adalah bilangan eksponen, i adalah bilangan imajiner, dan \cos dengan \sin adalah fungsi trigonometri. Jika persamaan 5 disubstitusikan ke persamaan 4 maka rumus DFT menjadi persamaan 6 di bawah ini.

$$Zt(k) = \sum_{n=0}^{N-1} Yt(n) [\cos \cos(2\pi nkN) - j \sin \sin(2\pi nkN)] \quad (6)$$

$$k = 0, 1, \dots, N-1$$

Keterangan:

$Zt(k)$ = Nilai FT yang mengandung bilangan kompleks

$Yt(n)$ = Nilai hamming window ke- n pada frame ke- t

j = Bilangan imajiner

Hasil perhitungan pada persamaan 6 mengandung bilangan riil dan imajiner. Oleh karena itu, nilai Fourier spektrum didapat dari nilai *magnitude* yang dibentuk kedua bilangan riil dan imajiner. Besar nilai *magnitude* dari persamaan $z = x + y$ tersebut dihitung di dalam persamaan 7.

$$r = M(n) = \sqrt{x^2 + y^2} \quad (7)$$

Keterangan:

z = Nilai FT yang mengandung bilangan kompleks

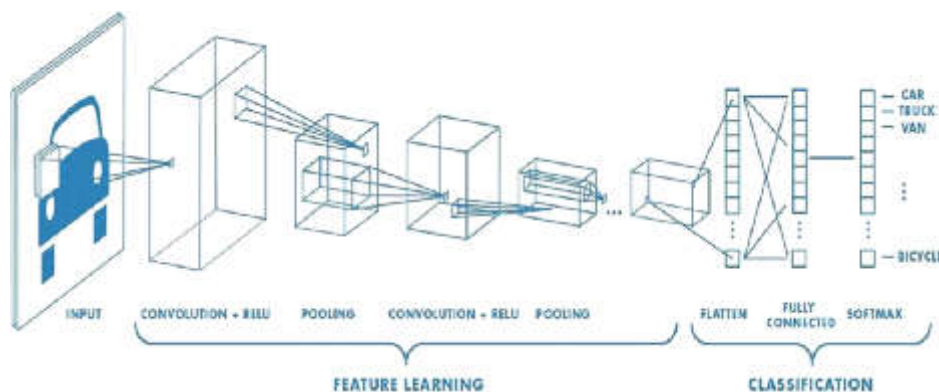
$M(n)$ = Magnitude

x = koefisien bagian riil

y = Koefisien bagian imajiner

2.1.6 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah salah satu jenis *neural network* yang biasa digunakan pada data *image*. CNN bisa digunakan untuk mendeteksi dan mengenali objek pada sebuah *image*. CNN terdiri dari *neuron* yang memiliki *weight*, bias, dan fungsi aktivasi. Cara kerja CNN memiliki kesamaan pada MLP, namun dalam CNN setiap *neuron* dipresentasikan dalam bentuk dua dimensi, tidak seperti MLP yang setiap *neuron* hanya berukuran satu dimensi. Pada CNN, data yang dipropagasikan oleh jaringan adalah data dua dimensi, sehingga operasi linear dan parameter bobot pada CNN berbeda. Pada CNN operasi linear menggunakan operasi konvolusi, sedangkan bobot tidak lagi satu dimensi saja, namun berbentuk empat dimensi yang merupakan kumpulan kernel konvolusi. Karena sifat proses konvolusi, maka CNN hanya dapat digunakan pada data yang memiliki struktur dua dimensi seperti citra dan suara. Arsitektur dari CNN dibagi menjadi dua bagian besar, *Feature Extraction Layer* dan *Fully-Connected Layer* (MLP). Arsitektur CNN dapat dilihat pada Gambar 2.2.



Gambar 2.2: Arsitektur *Convolutional Neural Network*

I Wayan Suartika Eka Putra (2016) dalam penelitiannya yang berjudul *Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101* mengimplementasikan salah satu metode *machine learning* yang dapat digunakan untuk klasifikasi citra objek yaitu CNN.

Pada penelitian yang dilakukan, peneliti mengimplementasikan metode CNN yang terdiri dari dua tahap. Tahap pertama adalah klasifikasi citra menggunakan *feedforward*. Tahap kedua merupakan tahap pembelajaran dengan metode *backpropagation*. Hasil uji coba dari klasifikasi citra objek dengan tingkat *confusion* yang berbeda pada basis data *Caltech 101* menghasilkan rata-rata nilai akurasi tergolong tinggi, sehingga dapat disimpulkan bahwa metode CNN yang digunakan pada penelitian tersebut mampu melakukan klasifikasi dengan baik.

Berbeda dengan Dzikry Maulana Hakim dan Ednawati Rainarli(2019) dalam penelitiannya yang berjudul *Convolutional Neural Network* untuk Pengenalan Citra Notasi Musik. Peneliti melakukan pengujian untuk melakukan pengenalan citra notasi musik dengan dua cara. Pertama, peneliti menguji performa CNN menggunakan data notasi musik yang telah dipotong dan yang kedua peneliti melakukan pengujian menggunakan sebaris notasi musik. Nilai akurasi yang didapatkan untuk pengenalan sebaris notasi musik tidak terlalu besar, yaitu 26,19%. Walaupun untuk proses segmentasi masih belum maksimal dalam memotong setiap notasi, namun metode CNN bekerja sangat baik untuk mengenali setiap notasi musik yang telah dipotong dengan benar. Hal ini ditunjukkan dari nilai akurasi yang mencapai 95,56%.

2.1.7 Music Information Retrieval

Music Information Retrieval (MIR) terutama berkaitan dengan ekstraksi dan inferensi fitur yang berarti dari musik seperti sinyal *audio*, representasi simbolik atau sumber eksternal seperti halaman web. Berikut merupakan contoh tugas MIR, antara lain *fingerprinting*, *cover song detection*, *genre recognition*, *key detection*, *beat tracking*, *symbolic melodic similarity*, *source separation*, *pitch tracking*, *tempo estimation*, dan masih banyak lagi. MIR digunakan pada sistem untuk menyiapkan data mentah menjadi bentuk data yang dapat dilakukan observasi oleh model. Berikut merupakan beberapa fitur pada MIR yang digunakan oleh sistem :

1. Beat Tracking

Sistem memanfaatkan fitur *beat tracking* yang disediakan oleh pustaka *librosa* yang menghasilkan keluaran dalam bentuk list. Nilai yang dihasilkan dalam satuan detik.

2. Harmonic Separation

Fitur *harmonic separation* berfungsi untuk memisahkan data masukan berupa *harmonic* dan *percussive*. Contoh dari *harmonic* adalah nada yang dibunyikan oleh piano dan suara vokal, sedangkan *percussive* merupakan kumpulan data suara yang dimainkan dengan cara dipukul, digesek, maupun dikocok.

2.2 Penelitian Terdahulu

I Wayan Suartika Eka Putra (2016) dalam penelitiannya yang berjudul Klasifikasi Citra Menggunakan *Convolutional Neural Network* (CNN) pada *Caltech 101* mengimplementasikan salah satu metode *machine learning* yang dapat digunakan untuk klasifikasi citra objek yaitu CNN. Pada penelitian yang dilakukan, peneliti mengimplementasikan metode CNN yang terdiri dari dua tahap. Tahap pertama adalah klasifikasi citra menggunakan *feedforward*. Tahap kedua merupakan tahap pembelajaran dengan metode *backpropagation*. Sebelum melakukan klasifikasi, terlebih dahulu dilakukan praproses dengan metode *wrapping* dan *cropping* untuk memfokuskan objek yang akan diklasifikasi. Selanjutnya dilakukan *training* menggunakan metode *feedforward* dan *backpropagation*. Terakhir adalah tahap klasifikasi menggunakan metode *feedforward* dengan bobot dan bias yang diperbarui. Hasil uji coba dari klasifikasi citra objek dengan tingkat *confusion* yang berbeda pada basis data *Caltech 101* menghasilkan rata-rata nilai akurasi tergolong tinggi, sehingga dapat disimpulkan bahwa metode CNN yang digunakan pada penelitian tersebut mampu melakukan klasifikasi dengan baik.

Danny Lionel, Rudy Adipranata dan Endang Setyati (2016) dalam penelitian yang berjudul Klasifikasi Genre Musik Menggunakan Metode *Deep Learning Convolutional Neural Network* dan *MelSpektrogram* mengimplementasikan metode *Mel-spectrogram*. Dimana *Mel spektrogram* merupakan hasil pemetaan fitur yang telah diambil oleh metode MFCC, yang akan diklasifikasikan dan dimasukkan ke dalam *Convolutional Neural Network*. Yang akan dibedakan *activation function* nya yaitu ReLU dan ELU. Penelitian ini menunjukkan bahwa pengambilan fitur dari *audio* dengan menggunakan MFCC merupakan metode yang benar dan dalam hasil pengujian, banyaknya dataset, iterasi *training*, dan spesifikasi komputer sangat mempengaruhi tingkat akurasi dan lama pembuatan *neural network model* yang optimal. Dalam hasil penelitian ini telah diuji beberapa kali didapatkan hasil akurasi yang paling

optimal yaitu 99%.

Dzikry Maulana Hakim dan Ednawati Rainarli (2019) pada penelitiannya yang berjudul *Convolutional Neural Network* untuk Pengenalan Citra Notasi Musik melakukan penelitian untuk mengenali citra notasi musik dengan metode CNN. Pada penelitian ini, untuk pengenalan notasi musik digunakan *Convolutional Neural Network* (CNN). Arsitektur CNN yang dipakai adalah *kernel 3x3*, jumlah layer pada *feature learning* sebanyak 3 *convolutional layer*, 3 *pooling layer*, *filter* pada *convolutional layer* 64, 128, 256, dan jumlah *neuron* pada *hidden layer* sebanyak 7168. Pengujian dilakukan dengan dua cara, yang pertama menguji performa CNN menggunakan data notasi musik yang telah dipotong dan yang kedua adalah melakukan pengujian menggunakan sebaris notasi musik. Nilai akurasi yang didapatkan untuk pengenalan sebaris notasi musik tidak terlalu besar, yaitu 26,19%. Walaupun untuk proses segmentasi masih belum maksimal dalam memotong setiap notasi, namun metode CNN bekerja sangat baik untuk mengenali setiap notasi musik yang telah dipotong dengan benar. Hal ini ditunjukkan dari nilai akurasi yang mencapai 95,56%.

Nick Collins melakukan penelitian dengan membuat *library SuperCollider Music Information Retrieval* (SCMIR). SCMIR adalah set ekstensi untuk bahasa pemrograman *audio* yang berfungsi untuk melakukan analisis otomatis sebuah *file audio*. *Library* ini mendukung teknologi pencarian informasi musik umum, termasuk untuk pemrosesan *batch* di seluruh *file* suara. *Library* memiliki kelebihan yang diambil dari mode *Non-Real-time scsynth* dan *plug-in* untuk ekstraksi fitur cepat, serta fitur tambahan untuk perhitungan intensif. Hasil dari penelitian ini adalah Nick Collins dapat meningkatkan hubungan antara sintesis pencocokan fitur, mendengarkan mesin dalam sistem interaktif, dan tugas MIR, melintasi *real time task* dan *non-real time* dengan memprioritaskan adopsi teknologi MIR dalam sistem interaktif. Selain itu, *library* SCMIR juga membawa teknologi pencarian informasi musik ke bahasa pemrograman komputer musik yang sudah terkenal. Rincian mengenai penelitian terdahulu dapat dilihat pada tabel 2.1.

Tabel 2.1: Penelitian Terdahulu.

No	Peneliti	Judul Penelitian	Metode	Hasil
1	I Wayan Suartika Eka Putra	Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101	Mengimplementasikan salah satu metode machine learning yang dapat digunakan untuk klasifikasi citra objek yaitu CNN	Hasil uji coba dari klasifikasi citra objek dengan tingkat confusion yang berbeda pada basis data Caltech 101 menghasilkan rata-rata nilai akurasi tergolong tinggi, sehingga dapat disimpulkan bahwa metode CNN yang digunakan pada penelitian tersebut mampu melakukan klasifikasi dengan baik
2	Danny Lionel, Rudy Adipranata dan Endang Setyati	Klasifikasi Genre Musik Menggunakan Metode Deep Learning Convolutional Neural Network dan MelSpektrogram	Mengimplementasikan metode agar suatu audio file dapat dikenali secara otomatis dari fitur-fitur yang telah diekstrak sebelumnya dengan bantuan MFCC (Mel Frequency Cepstral Coefficients) dan ANN (Artificial Neural Network).	Penelitian ini menunjukkan bahwa pengambilan fitur dari audio dengan menggunakan MFCC merupakan metode yang benar dan dalam hasil pengujian, banyaknya dataset, iterasi training, dan spesifikasi komputer sangat mempengaruhi tingkat akurasi dan lama pembuatan neural network model yang optimal. Dalam hasil penelitian ini telah diuji beberapa kali didapatkan hasil akurasi yang paling optimal yaitu 99%

Tabel 2.1: Penelitian Terdahulu. (Lanjutan)

No	Peneliti	Judul Penelitian	Metode	Hasil
3	Dzikry Maulana Hakim dan Ednawati Rainarli	Convolutional Neural Network untuk Pengenalan Citra Notasi Musik	Mengetahui cara agar sistem dapat mendeteksi sebuah notasi musik dan kemudian mengenali notasi musik tersebut	Pengujian dilakukan dengan dua cara, yang pertama menguji performa CNN menggunakan data notasi musik yang telah dipotong dan yang kedua adalah melakukan pengujian menggunakan sebaris notasi musik. Nilai akurasi yang didapatkan untuk pengenalan sebaris notasi musik tidak terlalu besar, yaitu 26,19%. Walaupun untuk proses segmentasi masih belum maksimal dalam memotong setiap notasi, namun metode CNN bekerja sangat baik untuk mengenali setiap notasi musik yang telah dipotong dengan benar. Hal ini ditunjukkan dari nilai akurasi yang mencapai 95,56%.

Tabel 2.1: Penelitian Terdahulu. (Lanjutan)

4	Nick Collins	SCMIR: A SUPERCOLLIDER MUSIC INFORMATION RETRIEVAL LIBRARY	Membuat sebuah library extension set untuk menganalisis file audio secara otomatis	Peneliti berhasil mengembangkan library SCMIR sehingga dapat meningkatkan hubungan antara sintesis pencocokan fitur, mendengarkan mesin dalam sistem interaktif, dan mampu melintasi pekerjaan realtime dan non-realtime dengan memprioritaskan adopsi teknologi MIR dalam sistem interaktif.
---	--------------	--	--	---

Tabel 2.1: Penelitian Terdahulu. (Lanjutan)

5	Florian Krebs, Sebastian Boock, and Gerhard Widmer	RHYTHMIC PATTERN MODELING FOR BEAT AND DOWNBEAT TRACKING IN MUSICAL AUDIO	Menyelidiki penggunaan pemodelan pola ritmis untuk menyimpulkan struktur metrik dalam rekaman audio musikal	<p>Penelitian yang dilakukan menunjukkan bahwa menghitung fitur onset setidaknya untuk dua band frekuensi yang berbeda meningkatkan kinerja pelacakan suram secara signifikan dibandingkan dengan fitur tunggal yang mencakup seluruh rentang frekuensi. Dalam perbandingan dengan enam sistem referensi, pemodelan gaya tarian secara eksplisit sebagai pola ritmik terbukti mengurangi kesalahan oktaf (mendeteksi setengah atau tempo ganda) dalam pelacakan ketukan. Selain itu, pelacakan suram ditingkatkan secara substansial dibandingkan dengan varian yang hanya memodelkan meter dan dua sistem referensi.</p>
---	--	---	---	---

Tabel 2.1: Penelitian Terdahulu. (Lanjutan)

6	Albert Parlys, Ajub Ajulian Zahra, dan Achmad Hidayatno	Penggolongan Lagu Berdasarkan Spektogram dengan Convolutional Neural Network	Penelitian ini akan membahas perancangan sebuah sistem untuk menggolongkan lagu berdasarkan spektogram. Masukan sistem berupa lagu dengan format audio MP3 yang diubah ke dalam bentuk spektogram kemudian dilatih menggunakan Convolutional Neural Network. Ciri lagu akan diperoleh kemudian diklasifikan ke dalam lima genre berbeda yaitu pop, rock, classic, dubstep, dan reggae	Berdasarkan hasil pelatihan dan pengujian dengan filter 3x3 didapat nilai akurasi penggolongan lagu sebesar 100% pada 750 data latih dan 98% pada 50 lagu data uji. Algoritme pembelajaran terbaik pada pelatihan dengan filter yang sama adalah algoritme Adam yang lebih cepat dibandingkan dengan Adadelata, Adagrad, dan SGD
---	---	--	---	--