

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang digunakan dalam dokumen skripsi yang bertema Penggunaan *Fused Neighborhood Attention* untuk Kompresi Analisis Struktur Musikal Audio yang Dipisahkan.

1.1 Latar Belakang

Music Information Retrieval (MIR) berkaitan dengan ekstraksi dan penarikan fitur-fitur bermakna dalam musik, pengindeksan musik menggunakan fitur-fitur tersebut, serta pengembangan berbagai skema pencarian dan pengambilan informasi. Musik memainkan peran penting sebagai medium ekspresi universal yang menghubungkan setiap generasi secara kultural dan emosional (Naveed dkk., 2017). Perkembangan teknologi informasi mendorong pelaku industri musik untuk melakukan distribusi musik menggunakan teknologi digital. Digitalisasi yang dialami industri musik memudahkan pencarian produk kreatif musik bagi setiap individu. Proses pencarian informasi musik dapat digunakan dalam tujuan pencarian hiburan bagi satu orang atau kelompok. Lebih dari itu, pencarian informasi mengenai musik juga membantu proses kreatif dalam produksi musik, baik secara profesional maupun amatir (Kostagiolas dkk., 2015). Dengan demikian, MIR bertujuan untuk menyediakan keseluruhan musik yang pernah diciptakan dapat didengarkan oleh setiap orang, baik konsumen, musisi amatir maupun profesional lain di bidang musik.

Analisis struktur pada musik bertujuan untuk membagi lagu menjadi segmen-segmen tidak tumpang tindih yang diberi label sesuai jenis segmen (Nieto dkk., 2020). Praktik ini berakar dari teori musik Barat yang menganalisis bentuk musik berdasarkan bagian-bagian penting dalam skala pendek seperti motif maupun dalam skala panjang seperti *verse* dan *chorus*. Meskipun bersifat subjektif, pendengar dapat menyetujui beberapa batas segmen yang penting. Dalam ranah MIR, analisis struktur musik penting digunakan untuk menguji model persepsi musik. Selain itu, analisis struktur pada musik memiliki aplikasi yang lebih luas seperti navigasi katalog musik berskala besar, rekomendasi, produksi, dan generasi musik (Goto, 2006).

Salah satu tantangan utama dalam analisis struktur musik adalah menangkap hubungan kontekstual dalam rentang waktu yang panjang sekaligus mempertahankan efisiensi komputasi. Algoritma pembelajaran mesin yang baru seperti *deep learning* membantu menghadapi tantangan tersebut (Humphrey dkk., 2012). Pendekatan *deep learning* berbasis *attention*, seperti yang digunakan dalam *transformer*, telah banyak digunakan dalam berbagai domain karena fleksibilitas dan efektivitasnya dalam mengenali pola. Penelitian (Chen dkk., 2023) menunjukkan bahwa jaringan *transformers* dapat digunakan dalam analisis struktur musik dengan menyatukan berbagai fitur akustik ke dalam operasi *multi-headed self attention*. Penelitian tersebut memberikan hasil *state-of-the-art* dalam tugas analisis struktur musik dengan metrik akurasi *pairwise F-measure* (PwF) sebesar 0,592 dan *Boundary Hit Rate F-measure* dengan jendela 0.5 detik (HR.5F) sebesar 0,621. Penelitian tersebut tidak menggunakan mekanisme *attention* yang memiliki kompleksitas waktu yang belum optimal.

Neighborhood Attention (NA) diajukan untuk mengurangi kompleksitas waktu yang ada pada jaringan *attention* dengan memanfaatkan jendela bergeser (Hassani dkk., 2023). Metode ini memiliki beberapa keunggulan yang ada dibanding pendahulunya yaitu adanya kesamaan ekuivarians translasional. NA juga mencoba mengatasi masalah yang dialami oleh *attention* lokal dengan menghilangkan pergeseran piksel dan *masked attention*. NA yang digunakan dalam jaringan *Neighborhood Attention Transformers* (NAT) memberikan hasil yang memuaskan pada dataset klasifikasi citra ImageNet dengan akurasi tertinggi sebesar 83,2% (Hassani dkk., 2023).

NA dapat digunakan untuk menggantikan mekanisme *attention* tidak hanya pada kasus-kasus yang melibatkan data citra. Penelitian Kim & Nam (2023) menunjukkan bahwa NA berhasil memberikan akurasi yang cukup tinggi dalam analisis struktur musik tanpa membebani komputer dengan komputasi yang berat. Analisis tersebut dilakukan dalam audio musik yang telah terpisah berdasarkan instrumen. Rekaman musik yang diolah dalam penelitian tersebut dipisahkan menggunakan model pembelajaran mesin menjadi beberapa rekaman instrumen yang terpisah. Model dalam penelitian tersebut menghasilkan metrik akurasi *pairwise F-measure* (PwF) tertinggi 0,738 dan *Boundary Hit Rate F-measure* dengan jendela 0,5 detik (HR.5F) tertinggi 0,660. Seluruh hasil tersebut dicapai dengan jumlah parameter yang relatif sedikit yaitu sekitar 300 ribu parameter.

Fused Neighborhood Attention (FNA) memberikan pengurangan waktu yang semula kuadratik dari mekanisme Neighborhood Attention menjadi linier dengan memanfaatkan perkalian titik tergabung pada kernel *attention* (Hassani dkk., 2024). FNA melakukan peningkatan kecepatan terhadap *runtime* pelatihan maupun pengujian model NA. FNA menunjukkan peningkatan *throughput* sebesar 28% pada tugas ImageNet. Meskipun demikian, penelitian tersebut belum menyelidiki dampak implementasi FNA pada data lain seperti data musik. Untuk itu, diperlukan penyelidikan terhadap kinerja FNA dalam analisis struktur musik dalam audio terpisahkan.

Penelitian ini mengajukan modifikasi dari model NA dalam penelitian Kim & Nam (2023) menggunakan FNA. Untuk membandingkan kinerja FNA dan NA, dibangun dua berbeda yang menggunakan FNA dan NA. Kedua model tersebut kemudian dibandingkan berdasarkan metrik evaluasi analisis struktur musik serta metrik komputasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana *Fused Neighborhood Attention* (FNA) dapat diadaptasi dan diimplementasikan untuk meningkatkan kinerja model analisis struktur musik terpisahkan?
2. Bagaimana perbandingan nilai akurasi antara model *Neighborhood Attention* (NA) dengan *Fused Neighborhood Attention* (FNA) dalam analisis struktur musik?

1.3 Tujuan dan Manfaat

Penelitian ini bertujuan untuk melakukan adaptasi *Fused Neighborhood Attention* (FNA) untuk model analisis struktur musik. Adaptasi yang dimaksud meliputi implementasinya dalam model analisis struktur musik dalam penelitian Kim & Nam (2023). Selain itu, penelitian yang dilakukan juga bertujuan untuk membandingkan akurasi model *Neighborhood Attention* (NA) dengan FNA.

Manfaat dari penelitian ini adalah memberikan pemahaman yang lebih mendalam mengenai *Fused Neighborhood Attention* (FNA) bagi penelitian berikutnya. Dengan pemahaman mendalam tersebut, inovasi dapat dilakukan oleh penelitian berikutnya.

1.4 Ruang Lingkup

Dalam penelitian ini, terdapat ruang lingkup yang dideskripsikan untuk membatasi cakupan pembahasan penelitian “Analisis Struktur Musik pada Audio Terpisah Menggunakan Metode Fused Neighborhood Attention”. Berikut adalah ruang lingkup yang ada dalam penelitian ini:

1. Penelitian ini melibatkan dua model berbeda Fused Neighborhood Attention (FNA) dan Neighborhood Attention (NA). Kedua model tersebut menggunakan parameter terlatih yang didapat dari pelatihan dalam penelitian Kim & Nam (2023).
2. Penelitian ini membahas tentang performa model analisis struktur musik pada tingkat segmen. Pelabelan segmen yang dilakukan disesuaikan dengan segmen yang terdapat dalam model pada penelitian Kim & Nam (2023).
3. Genre dipilih sesuai dengan genre yang terdapat dalam data latih model pada penelitian Kim & Nam (2023).
4. Data yang digunakan untuk pengujian diakses dari dataset Kaggle yang berupa 487 file audio yang terdapat dalam dataset SALAMI versi 2.0. Dataset tersebut terakhir diperbaharui pada 17 Maret 2015.
5. Representasi musik yang digunakan dalam penelitian ini ialah file audio berformat MP3 dengan *bit rate* 32 - 128 kbps.

1.5 Sistematika Penelitian

Penelitian ini tersusun atas lima bab yang membahas “*Analisis Struktur Musik pada Audio Terpisah menggunakan Metode Fused Neighborhood Attention*”. Pembagian tersebut dilakukan supaya penulisan dalam skripsi ini menjadi lebih teratur dan runtut. Berikut adalah ringkasan singkat dari penelitian yang akan dibahas:

BAB I Bagian ini menyajikan latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi yang berkaitan dengan evaluasi performa model Fused Neighborhood Attention dalam tugas temu balik informasi musik, khususnya analisis struktur musikal pada data audio yang telah dipisahkan.

- BAB II Bab ini menyajikan studi pustaka mengenai dasar teori yang digunakan dalam penelitian ini, termasuk *state-of-the-art* dalam temu balik informasi musik, teori terkait jaringan syaraf tiruan dan teknik Neighborhood Attention (NA) dan Fused Neighborhood Attention (FNA). Selain itu, dijelaskan juga konsep *source separation* pada data audio, representasi fitur musik, evaluasi dalam analisis struktur musik, serta berbagai *tools* dan *library* yang digunakan dalam implementasi model.
- BAB III Bab ini menjelaskan metodologi yang digunakan dalam penelitian ini, meliputi alur penelitian, pengumpulan dan pra-pemrosesan data audio, serta pemisahan sumber suara (*source separation*) untuk mendapatkan *stem* audio yang terpisah. Selanjutnya, dilakukan ekstraksi fitur dari data audio, diikuti dengan penerapan Fused Neighborhood Attention (FNA) dalam model Transformer untuk menganalisis struktur musikal. Untuk evaluasi, model dibandingkan menggunakan metrik yang umum dalam temu balik informasi musik, seperti *F-measure pairwise* dan *boundary hit rate*.
- BAB IV Bab ini menyajikan hasil eksperimen dan analisis performa model Fused Neighborhood Attention dibandingkan dengan model Neighborhood Attention dalam tugas analisis struktur musikal. Pembahasan mencakup perbandingan akurasi segmentasi, efisiensi komputasi, serta kecepatan inferensi antara kedua metode. Selain itu, dilakukan analisis mendalam mengenai keunggulan dan keterbatasan FNA dalam konteks data audio, serta implikasi hasil terhadap pengembangan sistem temu balik informasi musik di masa depan.
- BAB V Bab ini berisi kesimpulan dari penelitian mengenai efektivitas Fused Neighborhood Attention dalam tugas temu balik informasi musik, khususnya analisis struktur musikal. Selain itu, diberikan saran untuk pengembangan penelitian lebih lanjut, seperti eksplorasi metode atensi yang lebih efisien serta penerapan model pada dataset yang lebih luas dan beragam guna meningkatkan generalisasi model dalam aplikasi dunia nyata.