p-ISSN : 2301-8402, e-ISSN : 2685-368X, available at : https://ejournal.unsrat.ac.id/index.php/elekdankom

An End-to-End Framework for Optical Music Recognition and Jianpu Transliteration

Transliterasi Partitur Musik ke Not Angka Menggunakan Framework *End-to-end*

Christian Cleavy Sunkudon, Arthur Mourits Rumagit, Vecky Canisius Poekoel
Dept. of Electrical Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia
e-mails : christiansunkudon0261@student.unsrat.ac.id [1)], arthur_rumagit@unsrat.ac.id [2)], vecky_poekoel@unsrat.ac.id [3)]
Received: 16 July 2026; revised: 18 September 2026; accepted: 11 October 2026

Abstract — Optical Music Recognition **aims to automatically convert music notation into digital representations. Recent deep learning-based OMR approaches have achieved significant improvements in music symbol detection; however, most studies remain focused on object detection without fully addressing musical interpretation and end-to-end score transliteration. This research proposes an end-to-end hybrid framework for converting Western staff notation into numbered musical notation using a Dual-Model YOLO11 architecture that separates structural and semantic symbol detection. The detected symbols are subsequently processed through a Geometric Projection Layer to estimate note positions and transform geometric information into musical representations, followed by Musical Context Reconstruction to establish spatial relationships and musical structure based on music theory rules. Experimental evaluation on six test scores yielded transliteration accuracies ranging from 18.81% to 100%, with performance varying according to the characteristics and structural complexity of each musical score. These results demonstrate the feasibility of integrating deep learning-based symbol detection with geometric processing and rule-based musical reconstruction for end-to-end music score transliteration.**

Keywords— Geometric Projection Layer; Jianpu; Musical Context Reconstruction; Optical Music Recognition; Symbol Detection; Transliteration.

Abstrak — Optical Music Recognition **(OMR) merupakan bidang penelitian yang bertujuan mengonversi notasi musik ke dalam bentuk digital secara otomatis. Sebagian besar penelitian OMR berbasis deep learning masih berfokus pada deteksi simbol musik, sementara proses interpretasi hubungan antar simbol dan transliterasi notasi secara utuh masih menjadi tantangan. Penelitian ini mengusulkan sebuah framework hybrid *end-to-end* untuk mentransliterasikan notasi balok menjadi not angka menggunakan pendekatan Dual-Model YOLO11 yang memisahkan proses deteksi menjadi model struktural dan model semantik. Hasil deteksi kemudian diproses melalui Geometric Projection Layer untuk memperoleh estimasi posisi nada dan mentransformasikan informasi geometris menjadi representasi musikal, dilanjutkan dengan Musical Context Reconstruction untuk membangun hubungan spasial dan struktur musikal berdasarkan aturan teori musik. Hasil pengujian pada enam partitur menunjukkan akurasi transliterasi antara 18.81% hingga 100%, dengan variasi performa yang dipengaruhi oleh karakteristik dan kompleksitas struktur musikal pada masing-masing partitur. Hasil tersebut menunjukkan bahwa integrasi deteksi simbol berbasis deep learning, pemrosesan geometris, dan rekonstruksi musikal berbasis aturan dapat digunakan untuk mendukung proses transliterasi partitur musik secara *end-to-end***

Kata kunci — Geometric Projection Layer; Jianpu; Musical Context Reconstruction; Optical Music Recognition; Symbol Detection; Transliteration.

I. PENDAHULUAN

Optical Music Recognition (OMR) merupakan salah satu bidang penelitian dalam Music Information Retrieval yang bertujuan mengonversi notasi musik menjadi representasi digital secara otomatis [1], [2], [3], [4]. Tantangan utama OMR terletak pada kompleksitas notasi balok yang terdiri atas berbagai simbol dengan karakteristik visual, hubungan spasial, dan fungsi musikal yang saling bergantung. Interpretasi notasi musik tidak hanya ditentukan oleh bentuk simbol, tetapi juga oleh posisi relatif terhadap garis paranada dan hubungan antar simbol berdasarkan aturan teori musik.

Perkembangan deep learning telah meningkatkan kemampuan OMR dalam mengenali simbol musik melalui pendekatan CNN, deteksi objek, dan Transformer pada partitur cetak maupun tulisan tangan [5] [6] [7], [8], [9], [10], [11]. Penelitian OMR juga mulai berkembang menuju pendekatan *end-to-end* berbasis Transformer yang melakukan transkripsi langsung dari citra partitur [12], [13], [14]. Selain itu, ketersediaan dataset seperti DeepScoresV2 dan MUSCIMA++ turut mendukung perkembangan penelitian OMR melalui anotasi dan evaluasi yang lebih terstandar [15], [16].

Meskipun kemampuan pengenalan simbol telah meningkat, keluaran deteksi objek berupa kelas dan bounding box belum secara langsung merepresentasikan informasi musikal. Interpretasi memerlukan hubungan simbol terhadap paranada, urutan waktu, durasi, dan keterkaitan musikal lainnya. Pendekatan notation graph, layout-aware representation, dan graph learning telah digunakan untuk membangun hubungan antar simbol [17], [18], [19]. Namun, proses deteksi, interpretasi, dan transliterasi masih dapat berlangsung secara terpisah. Oleh karena itu, diperlukan pendekatan yang mengintegrasikan pengenalan simbol, rekonstruksi struktur musikal, dan transliterasi dalam satu alur.

Selain hubungan antar simbol, karakteristik visual simbol musik juga beragam. Elemen seperti garis paranada, kepala not dan beam berperan dalam membentuk struktur geometris, sedangkan accidental, rest, titik augmentasi, dan simbol pendukung lainnya memberikan informasi musikal. Oleh karena itu, penelitian ini menerapkan dekomposisi tugas melalui dua model deteksi, yaitu Model Struktural dan Model Semantik, dengan cakupan kelas yang disesuaikan dengan fungsi masing-masing.

Not angka atau Jianpu merupakan bentuk notasi yang menggunakan angka untuk merepresentasikan derajat nada serta simbol tambahan untuk menunjukkan register dan durasi [20], [21], [22]. Bentuk notasi ini banyak digunakan dalam pembelajaran dan praktik musik di Indonesia.

Dengan demikian, diperlukan mekanisme yang menjembatani informasi geometris hasil deteksi dengan representasi musikal sebelum rendering. Penelitian sebelumnya telah mengeksplorasi transliterasi notasi balok ke notasi angka berbasis MusicXML [20], tetapi pendekatan tersebut menggunakan representasi musik terstruktur sebagai masukan dan belum mencakup interpretasi langsung dari citra partitur.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan framework hybrid *end-to-end* untuk Optical Music Recognition dan transliterasi notasi balok ke not angka (Jianpu). Framework mengintegrasikan Dual-Model YOLO11 untuk deteksi simbol struktural dan semantik, Geometric Projection Layer untuk mentransformasikan informasi geometris hasil deteksi menjadi representasi musikal berdasarkan struktur paranada, serta Musical Context Reconstruction berbasis aturan untuk membangun hubungan antar simbol dan struktur musical events. Rangkaian tersebut kemudian digunakan untuk menghasilkan transliterasi notasi balok ke not angka. Kontribusi penelitian terletak pada integrasi proses deteksi simbol, transformasi geometris, rekonstruksi konteks musikal, dan transliterasi dalam satu framework, bukan pada pengembangan algoritma deteksi YOLO11 itu sendiri.

Berdasarkan kontribusi tersebut, penelitian ini bertujuan mengembangkan dan mengevaluasi framework OMR yang mampu melakukan transliterasi partitur notasi balok menjadi not angka secara *end-to-end*. Evaluasi dilakukan pada dua tingkat, yaitu performa deteksi simbol pada Model Struktural dan Model Semantik serta kemampuan framework dalam mempertahankan dan merekonstruksi musical events hingga menghasilkan transliterasi not angka yang tervalidasi.

II. METODE

A. Framework

Penelitian ini mengembangkan sebuah framework Optical Music Recognition (OMR) untuk mentransliterasikan notasi balok menjadi not angka (Jianpu) secara *end-to-end*. Framework yang diusulkan tidak hanya melakukan deteksi simbol musik, tetapi juga mengolah informasi geometris dan hubungan semantik antar simbol untuk membentuk representasi musikal yang diperlukan dalam proses transliterasi.

Framework yang diusulkan terdiri atas enam tahapan utama, yaitu persiapan dataset, deteksi simbol menggunakan Dual-Model YOLO11, transformasi informasi geometris melalui Geometric Projection Layer, rekonstruksi konteks musikal menggunakan Musical Context Reconstruction, rendering representasi musik (not angka), dan evaluasi sistem. Keenam tahapan tersebut saling terhubung, sehingga keluaran dari setiap tahap menjadi masukan bagi tahap berikutnya dalam proses transliterasi notasi balok menjadi not angka (Jianpu).

Diagram alur framework penelitian yang diusulkan ditunjukkan pada Gambar 1. Berdasarkan Gambar 1, framework yang diusulkan terdiri atas enam tahapan utama, yaitu persiapan dataset, deteksi simbol menggunakan Dual-Model YOLO11, Geometric Projection Layer, Musical Context Reconstruction, rendering keluaran, dan evaluasi sistem. Setiap tahapan saling terhubung untuk menghasilkan transliterasi notasi balok menjadi not angka (Jianpu) dan MIDI secara *end-to-end*.

Secara konseptual, framework yang diusulkan terdiri atas tiga lapisan utama, yaitu Detection Layer, Interpretation Layer, dan Representation Layer. Ketiga lapisan tersebut bekerja secara berurutan untuk mentransformasikan citra partitur menjadi representasi musik digital yang utuh.

B. Persiapan Dataset

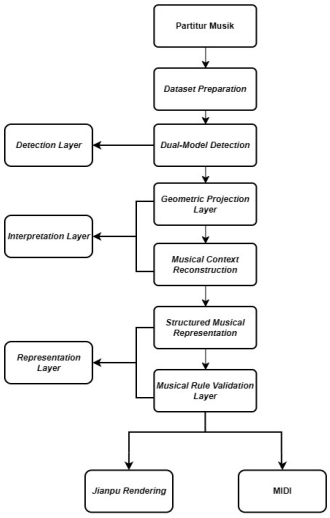
Penelitian ini menggunakan DeepScoresV2 Dense sebagai dataset utama untuk proses pelatihan model deteksi simbol musik [15]. Dataset tersebut dipilih karena menyediakan anotasi simbol musik dengan kepadatan objek yang tinggi sehingga mampu merepresentasikan berbagai simbol pada partitur musik secara komprehensif. Setelah proses penyaringan dan penyesuaian dataset, diperoleh 1.714 citra dengan 45 kelas simbol yang digunakan dalam penelitian.

Sebelum pelatihan, dataset diproses melalui Dataset Preparation Pipeline sebagaimana ditunjukkan pada Gambar 2. Tahapan tersebut meliputi parsing anotasi, pemetaan kelas, penyaringan kelas, reduksi anotasi, konversi anotasi ke format YOLO, pembagian data pelatihan dan validasi, serta penyeimbangan data validasi. Penyaringan kelas dilakukan untuk mempertahankan simbol yang berkontribusi terhadap proses transliterasi, sedangkan reduksi anotasi digunakan untuk mengendalikan jumlah anotasi pada kelas dengan frekuensi tinggi. Reduksi dilakukan setelah penyaringan kelas dengan mengurangi jumlah anotasi pada kelas yang memiliki frekuensi sangat tinggi, sehingga distribusi anotasi menjadi lebih seimbang dan ukuran dataset lebih sesuai untuk proses pelatihan. Anotasi yang telah melalui proses penyaringan dan pengurangan jumlah tersebut kemudian dikonversi ke dalam format YOLO.

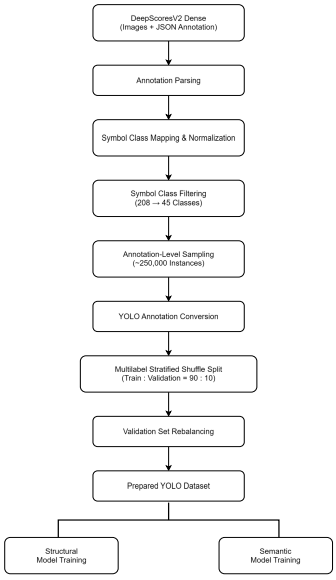
Setelah proses tersebut, diperoleh 45 kelas, yang selanjutnya digunakan dalam Dual-Model YOLO11. Model Struktural menggunakan 23 kelas untuk merepresentasikan elemen dasar dan geometris partitur, sedangkan Model Semantik mencakup keseluruhan 45 kelas. Rincian pembagian kelas disajikan pada Tabel I.

Dataset awal dibagi menggunakan rasio 90:10 dengan pendekatan Multilabel Stratified Shuffle Split. Selanjutnya dilakukan penyeimbangan data validasi dengan memindahkan sejumlah citra dari data pelatihan ke data validasi untuk meningkatkan keterwakilan kelas. Hasil akhir pembagian data adalah 1.362 citra pelatihan dan 352 citra validasi. Setelah penyaringan kelas dan reduksi anotasi, diperoleh 402.800 anotasi, terdiri atas 250.000 anotasi pada data pelatihan dan 152.800 anotasi pada data validasi.

Dengan demikian, 402.800 anotasi yang digunakan dalam dataset akhir terdiri atas 250.000 anotasi pelatihan dan 152.800 anotasi validasi, sedangkan perbedaan jumlah anotasi antar



Gambar 1. Diagram Alur Penelitian



Gambar 2. Dataset Preparation Pipeline

TABEL I. CAKUPAN KELAS PADA DUAL-MODEL YOLO11

Model	Jumlah	Keterangan	Kelas
Struktural	23	Digunakan pada Model Struktural dan Model Semantik	noteheadBlack, noteheadHalf, noteheadWhole, staff, beam, flag8thUp, flag8thDown, flag16thUp, flag16thDown, augmentationDot, accidentalSharp, accidentalFlat, accidentalNatural, restWhole, restHalf, restQuarter, rest8th, rest16th, clefG, clefF, keySharp, keyFlat, keyNatural
Semantik	22	Hanya digunakan pada Model Semantik	tie, slur, fermataAbove, fermataBelow, brace, coda, segno, repeatDot, flag32ndUp, flag32ndDown, flag64thUp, flag64thDown, flag128thUp, flag128thDown, rest32nd, rest64th, rest128th, dynamicP, dynamicM, dynamicF, dynamicCrescendoHairpin, dynamicDiminuendoHairpin
Total	45	Keseluruhan kelas yang digunakan dalam framework	

subset merupakan konsekuensi dari distribusi simbol pada citra setelah proses reduksi anotasi dan penyeimbangan data validasi, bukan pembagian anotasi secara langsung dengan rasio 90:10.

C. Dual-Model YOLO11 for Symbol Detection

Pendekatan Dual-Model YOLO11 menjadi framework yang diusulkan dalam penelitian ini. Dual-Model YOLO11 terdiri atas dua model deteksi yang dilatih secara terpisah dengan cakupan kelas yang berbeda. Pendekatan ini dipilih karena simbol pada partitur memiliki karakteristik visual dan fungsi musikal yang beragam, sehingga pemisahan tugas deteksi memungkinkan setiap model difokuskan pada informasi yang diperlukan dalam tahapan transliterasi. Pendekatan ini diilustrasikan pada Gambar 3.

Berdasarkan Gambar 3, pendekatan ini melakukan pemisahan tugas deteksi menjadi Model Struktural dan Model Semantik. Model Struktural difokuskan pada pendeteksian elemen dasar partitur yang diperlukan untuk memperoleh informasi geometris, sedangkan Model Semantik digunakan untuk melengkapi hasil deteksi dengan simbol-simbol yang diperlukan dalam interpretasi musikal.

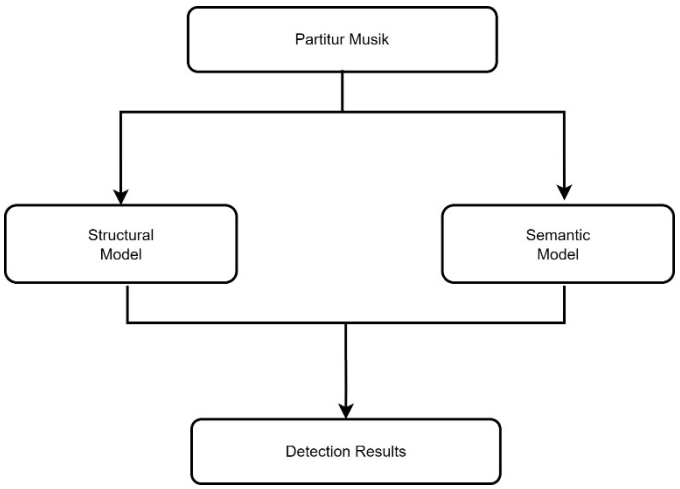
TABEL II. KONFIGURASI PELATIHAN DUAL-MODEL YOLO11

Parameter	Konfigurasi
Model	YOLO11m
Pretrained	Ya
Resolusi input	1792 × 1792 piksel
Epoch	200
Batch size	8
Optimizer	SGD
Learning rate	0.01
Jumlah kelas	Struktural: 23; Semantik: 45

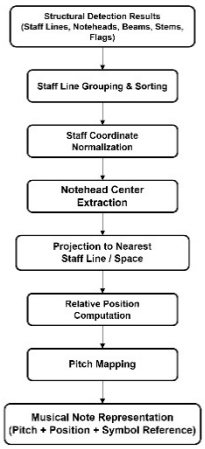
Model Struktural menggunakan 23 kelas untuk merepresentasikan elemen dasar dan struktur geometris partitur, meliputi kepala not, staff, beam, flag, augmentation dot, accidental, rest, clef, dan key signature. Informasi posisi dari kelas-kelas tersebut digunakan untuk memperoleh representasi spasial partitur, khususnya dalam menentukan posisi simbol terhadap paranada dan hubungan geometris antar objek.

Model Semantik menggunakan 45 kelas yang terdiri atas 23 kelas bersama dengan Model Struktural dan 22 kelas tambahan yang diperlukan untuk melengkapi representasi musikal. Kelas tambahan tersebut mencakup simbol seperti tie, slur, fermata, brace, coda, segno, repeat dot, flag dan rest dengan durasi yang lebih pendek, serta dynamic marking. Hasil deteksi Model Semantik digunakan untuk melengkapi informasi yang diperoleh dari Model Struktural sehingga konteks simbol dan hubungan musikal dapat direkonstruksi pada tahap berikutnya. Konfigurasi pelatihan utama disajikan pada Tabel II.

Kedua model menggunakan arsitektur YOLO11m dengan konfigurasi pelatihan yang sama, sedangkan perbedaannya terletak pada cakupan kelas yang digunakan. Model pertama dilatih menggunakan bobot YOLO11m pre-trained, kemudian bobot terbaik dari model tersebut digunakan sebagai model awal untuk pelatihan Model Semantik.



Gambar 3. Dual-Model Symbol Detection



Gambar 4. Geometric Projection Layer

Implementasi menggunakan Python 3.13.7, Ultralytics 8.4.21, PyTorch 2.10.0+cu128, OpenCV 4.13.0.92, dan Music21 9.9.1 pada GPU NVIDIA A100 80 GB. Augmentasi yang digunakan meliputi mosaic 0,5, mixup 0,1, dan scale 0,5, dengan SGD, learning rate 0,01, momentum 0,937, dan weight decay 0,0005. Pada inferensi digunakan confidence threshold 0,001, NMS IoU 0,50, ukuran input 1280 × 1280 piksel, dan maksimum 3.000 deteksi per citra.

Hasil deteksi dari kedua model selanjutnya digunakan secara bersama dalam tahapan selanjutnya, yaitu Geometric Projection Layer dan Musical Context Reconstruction.

D. Geometric Projection Layer

Hasil deteksi dari Model Struktural dan Model Semantik berupa kelas simbol beserta koordinat bounding box belum secara langsung merepresentasikan posisi nada (tinggi nada). Dalam notasi balok, posisi nada ditentukan oleh hubungan vertikal kepala not terhadap garis dan spasi pada paranada. Oleh karena itu, diperlukan mekanisme yang menghubungkan informasi spasial hasil deteksi dengan struktur geometris partitur agar posisi kepala not dapat diinterpretasikan menjadi informasi musikal.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan Geometric Projection Layer (GPL) yang berfungsi mengintegrasikan hasil deteksi simbol dengan struktur paranada. Proses GPL diawali dengan mengelompokkan simbol yang terdeteksi berdasarkan kedekatannya terhadap paranada yang sesuai. Pengelompokan ini digunakan untuk menentukan konteks geometris setiap simbol sehingga posisi kepala not dapat diinterpretasikan relatif terhadap struktur kepala nada yang menaunginya.

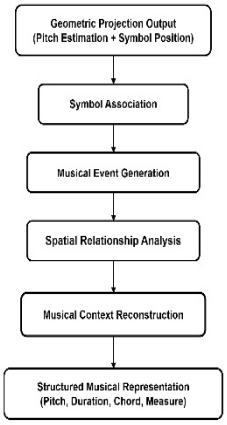
Selanjutnya, sistem melakukan penyempurnaan garis paranada pada area paranada hasil deteksi. Tahap ini dilakukan untuk memperoleh posisi lima garis paranada secara lebih akurat daripada hanya menggunakan koordinat bounding box hasil deteksi. Area paranada dipotong dengan tambahan padding, kemudian dikonversi ke citra grayscale dan dibinerisasi menggunakan Otsu thresholding. Proyeksi horizontal (horizontal projection) digunakan untuk memperoleh distribusi piksel pada arah vertikal dan mengidentifikasi kandidat posisi garis paranada. Kandidat garis yang terlalu berdekatan disaring, sedangkan garis yang hilang dapat diestimasi berdasarkan pola jarak antar garis yang terdeteksi. Dengan proses tersebut, sistem memperoleh representasi lima garis paranada yang digunakan sebagai acuan geometris untuk pemetaan posisi nada.

Setelah posisi garis paranada diperoleh, posisi vertikal kepala not juga diperbaiki melalui proses penyempurnaan kepala not. Proses ini menggunakan area bounding box kepala not, kemudian menerapkan grayscale, Otsu thresholding, dan proyeksi horizontal. Distribusi piksel hasil proyeksi diberi bobot sehingga posisi vertikal kepala not diperoleh dari pusat massa distribusi tersebut. Posisi ini digunakan sebagai representasi vertikal kepala not sebelum dilakukan pemetaan ke grid paranada.

Posisi vertikal kepala not kemudian diproyeksikan terhadap grid paranada berdasarkan jarak antara garis pertama dan garis kelima. Jarak tersebut dibagi menjadi delapan interval sehingga posisi garis dan spasi pada paranada dapat direpresentasikan dalam satuan langkah nada. Dengan pendekatan ini, posisi kepala not tidak hanya dibandingkan terhadap garis paranada terdekat, tetapi dipetakan ke posisi diskrit pada grid yang mencakup posisi nada pada garis, spasi, serta posisi yang berada di luar lima garis utama.

Selanjutnya, posisi relatif kepala not dipetakan menjadi nilai tinggi nada berdasarkan clef yang digunakan pada garis paranada. Sistem menggunakan clefG sebagai referensi treble clef dan clefF sebagai referensi bass clef. Berdasarkan posisi kepala not pada grid paranada dan clef yang sesuai, sistem menentukan nama nada dan oktaf. Informasi accidental dan key signature yang terdeteksi kemudian digunakan untuk menyesuaikan nilai tinggi nada sebelum dibentuk menjadi musical event.

Selain informasi tinggi nada, koordinat horizontal kepala not tetap dipertahankan karena diperlukan untuk menentukan urutan temporal dan hubungan spasial antar simbol pada tahap Musical Context Reconstruction. Dengan demikian, keluaran Geometric Projection Layer berupa simbol musik yang telah dilengkapi informasi posisi geometris dan estimasi tinggi nada, yang selanjutnya digunakan sebagai masukan pada tahap



Gambar 5. Musical Context Reconstruction

rekonstruksi konteks musikal.

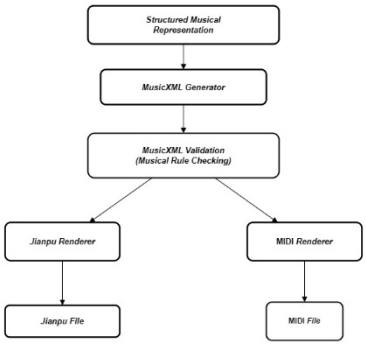
E. Musical Context Reconstruction

Hasil keluaran Geometric Projection Layer masih berupa kumpulan simbol musik yang telah memiliki informasi posisi geometris dan estimasi tinggi nada. Informasi tersebut belum membentuk representasi musik yang utuh karena hubungan antar simbol, seperti urutan nada, durasi, akor (chord), maupun keterkaitan simbol pendukung belum ditentukan. Oleh karena itu, diperlukan proses rekonstruksi konteks musikal sebelum partitur dapat ditransliterasikan.

Musical Context Reconstruction merupakan tahap yang mengintegrasikan informasi hasil Geometric Projection Layer untuk membangun representasi musikal yang lebih utuh. Pada tahap ini, simbol-simbol yang telah memiliki informasi posisi geometris dan estimasi tinggi nada dianalisis berdasarkan hubungan spasial dan aturan teori musik untuk membentuk struktur musikal yang koheren. Proses ini mencakup pembentukan hubungan antar simbol, identifikasi kelompok nada atau chord, serta penyusunan informasi ritmis dan struktural sebelum ditransformasikan ke dalam representasi notasi angka.

Tahapan ini diilustrasikan pada Gambar 5. Proses diawali dengan mengelompokkan kepala not yang memiliki hubungan dengan tangkai nada maupun beam untuk mengidentifikasi not tunggal atau akor. Selanjutnya, hubungan antar nada dibangun berdasarkan urutan posisi horizontal untuk memperoleh urutan pembacaan musik. Simbol pendukung seperti accidental, titik augmentasi, rest, dan dynamic marking kemudian diintegrasikan ke dalam setiap musical event sehingga informasi musikal yang diperlukan dapat dibentuk secara lengkap.

Hasil proses ini berupa representasi musik yang terdiri atas kumpulan musical event beserta atribut tinggi nada, durasi, posisi birama, dan simbol pendukung lainnya. Representasi tersebut selanjutnya dikonversi ke dalam struktur MusicXML sebagai representasi perantara (intermediate representation) untuk validasi struktur musik sebelum dilakukan rendering menjadi not angka.



Gambar 6. Musical Representation Validation and Rendering

Secara ringkas, MCR mengurutkan simbol berdasarkan posisi horizontal, mengelompokkan kepala not dengan tangkai nada atau beam, mengasosiasikan simbol pendukung berdasarkan hubungan spasial, membentuk note/rest event, kemudian memvalidasi struktur temporal menggunakan Music21.

F. Musical Representation Validation and Rendering

Hasil Musical Context Reconstruction membentuk representasi musik terstruktur yang memuat informasi tinggi nada, durasi, posisi birama, dan hubungan antar simbol. Representasi tersebut dikonversi ke MusicXML sebagai representasi perantara untuk memvalidasi konsistensi struktur musik sebelum proses transliterasi ke not angka. MusicXML juga digunakan untuk mendukung pembangkitan MIDI dan pemeriksaan struktur musikal, sedangkan score matrix tetap menjadi representasi utama pada proses transliterasi dan evaluasi.

Pada penelitian ini, MusicXML tidak digunakan sebagai keluaran utama, melainkan sebagai representasi perantara untuk mempertahankan struktur musikal selama proses validasi. Struktur tersebut memungkinkan hubungan antar nada, durasi, tanda istirahat, dan pembagian birama diperiksa sebelum dilakukan proses transliterasi ke dalam bentuk format notasi angka.

Pada tahap validasi, MusicXML digunakan sebagai representasi struktural untuk pemeriksaan konsistensi keluaran musik dan pembangkitan MIDI. Hasil validasi tersebut tidak digunakan untuk menggantikan score matrix yang dihasilkan oleh proses deteksi dan rekonstruksi. Dengan demikian, keluaran score matrix dari pipeline tetap menjadi representasi utama yang digunakan dalam proses transliterasi dan evaluasi *end-to-end*.

G. Evaluation Strategy

Untuk mengevaluasi kinerja framework yang diusulkan, penelitian ini melakukan evaluasi pada dua aspek utama, yaitu performa deteksi simbol dan transliterasi *end-to-end*. Evaluasi deteksi pada Model Struktural dan Model Semantik menggunakan Precision, Recall, mAP@50, dan mAP@50–95. Evaluasi transliterasi membandingkan representasi not angka hasil pipeline terhadap reference partitur pada hierarki Part–

TABEL III. DATASET STATISTIC

Dataset	**Images**	**Annotations**
Train	1.362	250.000
Validation	352	152.800
Total	1.714	402.800

Bar–Event. Alignment event dilakukan pada setiap birama menggunakan dynamic programming dengan mempertimbangkan kesamaan simbol dan perbedaan durasi. Akurasi simbol dihitung berdasarkan jumlah simbol yang benar terhadap total simbol pada reference.

Kesesuaian simbol Jianpu dievaluasi secara terpisah dari durasi. Suatu event dinyatakan benar (correct) apabila simbol Jianpu hasil transliterasi sesuai dengan simbol pada reference, dengan persamaan Akurasi = (Jumlah Simbol Benar / Total Simbol Reference) × 100%.

Pengujian transliterasi dilakukan pada enam partitur dengan karakteristik yang berbeda, yang mencakup partitur monofonik, himne, dan partitur polifonik SATB. Setiap partitur dievaluasi berdasarkan jumlah simbol yang berhasil ditransliterasikan dengan benar dan jumlah simbol yang mengalami kesalahan.

III. Hasil dan Pembahasan

Pada bagian ini disajikan hasil implementasi framework yang diusulkan beserta analisis pada setiap tahapan proses. Evaluasi dilakukan secara bertahap mulai dari hasil persiapan dataset, performa deteksi simbol menggunakan Dual-Model YOLO11, analisis Geometric Projection Layer, rekonstruksi konteks musikal, hingga kemampuan sistem dalam mentransliterasikan notasi balok menjadi not angka. Pendekatan ini bertujuan untuk menunjukkan kontribusi setiap komponen terhadap kinerja framework secara keseluruhan.

A. Dataset Preparation Results

Tahap persiapan dataset menghasilkan kumpulan data pelatihan yang telah disesuaikan dengan kebutuhan framework Optical Music Recognition yang diusulkan. Seluruh citra partitur berasal dari dataset DeepScoresV2 Dense, kemudian dilakukan proses validasi anotasi, pemetaan kelas simbol, penyaringan simbol yang tidak relevan terhadap proses transliterasi, konversi anotasi ke format YOLO, serta pembagian data pelatihan dan validasi menggunakan pendekatan Multilabel Stratified Shuffle Split. Proses tersebut menghasilkan distribusi data yang lebih representatif sehingga setiap kelas simbol tetap muncul pada kedua subset data. Tabel III menunjukkan distribusi dataset akhir yang digunakan untuk pelatihan dan evaluasi Dual-Model YOLO11.

B. Dual-Model Detection Performance

Setelah proses persiapan dataset selesai, tahap berikutnya bertujuan untuk mengevaluasi kemampuan Dual-Model YOLO11 dalam mendeteksi simbol-simbol musik yang digunakan pada proses transliterasi. Berbeda dengan sebagian besar penelitian OMR yang menggunakan satu model untuk mendeteksi seluruh simbol secara bersamaan, framework yang diusulkan memisahkan proses deteksi menjadi Model Struktural dan Model Semantik. Pemisahan ini dilakukan karena kedua kelompok simbol memiliki karakteristik visual serta fungsi musikal yang berbeda sehingga diharapkan dapat mengurangi kompleksitas pembelajaran dan meningkatkan kualitas deteksi.

TABEL IV. DUAL-MODEL DETECTION PERFORMANCE

Model	Precision	Recall	**mAP@50**	**mAP@50–95**
Struktural	0,968	0,763	0,866	0,693
Semantik	0,917	0,862	0,917	0,748

Model Struktural digunakan untuk mengenali simbol-simbol yang berperan dalam pembentukan struktur geometris partitur, sedangkan Model Semantik mencakup kelas-kelas simbol yang diperlukan untuk melengkapi informasi musikal pada tahap rekonstruksi. Kedua model menggunakan konfigurasi pelatihan yang sama, dengan perbedaan utama pada cakupan kelas yang digunakan. Pembagian tersebut memungkinkan hasil deteksi disesuaikan dengan kebutuhan tahapan Geometric Projection Layer dan Musical Context Reconstruction.

Evaluasi dilakukan menggunakan metrik Precision, Recall, mAP@50, dan mAP@50–95 pada data validasi. Keempat metrik tersebut dipilih karena mampu memberikan gambaran mengenai kemampuan model dalam mengenali simbol musik secara tepat sekaligus mengukur kualitas lokalisasi objek.

Selain menggunakan metrik evaluasi kuantitatif, hasil deteksi pada masing-masing model dianalisis untuk melihat karakteristik pengenalan simbol dan kesesuaiannya dengan kebutuhan tahapan Geometric Projection Layer dan Musical Context Reconstruction. Analisis dilakukan melalui hasil inferensi pada beberapa partitur uji dengan tingkat kompleksitas yang berbeda.

Penggunaan dua model deteksi juga memungkinkan setiap model mempelajari karakteristik simbol yang lebih homogen dibandingkan apabila seluruh simbol dipelajari secara bersamaan. Dengan demikian, proses evaluasi tidak hanya berfokus pada nilai metrik deteksi, tetapi juga mempertimbangkan kualitas representasi simbol yang akan digunakan pada tahap Geometric Projection Layer dan Musical Context Reconstruction.

Hasil evaluasi kedua model disajikan pada Tabel IV. Berdasarkan Tabel IV, kedua model menunjukkan performa deteksi yang baik. Model Struktural memperoleh Precision lebih tinggi, yaitu 96,8%, sedangkan Model Semantik menghasilkan Recall dan mAP yang lebih tinggi, masing-masing 86,2%, 91,7% pada mAP@50 dan 74,8% pada mAP@50–95. Perbedaan tersebut menunjukkan bahwa kedua model memberikan karakteristik deteksi yang saling melengkapi sesuai cakupan kelasnya.

Sebaliknya, nilai Precision yang tinggi pada Model Struktural mengindikasikan bahwa prediksi yang dihasilkan cenderung lebih selektif. Salah satu karakteristik yang dapat meningkatkan kompleksitas lokalisasi pada Model Struktural adalah keberadaan simbol yang berdekatan secara spasial, seperti kepala not, tangkai nada, beam, dan garis paranada.

TABEL V. SINGLE-MODEL AND DUAL-MODEL DETECTION COMPARISON

Model	Precision	Recall	F1	mAP@50	mAP@50–95
Single-Model	64,91%	75,21%	69,68%	69,84%	47,30%
Dual Model	**92,31%**	**89,35%**	**90,80%**	**92,36%**	**68,98%**

Kondisi tersebut dapat menjadi tantangan dalam membedakan objek yang memiliki posisi berdekatan. Untuk mengevaluasi lebih lanjut kontribusi dekomposisi tugas yang diusulkan, kinerja deteksi kerangka Dual-Model dibandingkan dengan baseline Single-Model YOLO11 menggunakan ruang evaluasi 45-kelas yang sama dan 352 citra validasi. Perbandingan disajikan dalam Tabel V.

Model Tunggal mencapai Precision 64,91%, Penarikan Kembali 75,21%, dan skor F1 69,68%, dengan mAP@50 dan mAP@50–95 masing-masing 69,84% dan 47,30%. Sebagai perbandingan, Model Ganda mencapai Precision 92,31%, Recall 89,35%, dan skor F1 90,80%, serta memperoleh skor mAP@50 92,36% dan 68,98% mAP@50-95. Peningkatan tersebut adalah 27,40, 14,14, dan 21,12 poin persentase untuk skor Precision, Recall, dan F1, masing-masing. Skor mAP@50 dan mAP@50–95 juga meningkat sebesar 22,52 dan 21,69 poin persentase.

Peningkatan pada kedua metrik mAP menunjukkan bahwa perbedaan antara kedua pendekatan tidak terbatas pada confidence threshold, tetapi juga tercermin dalam perilaku precision keseluruhan objek yang terdeteksi di ambang IoU. Hal ini sangat relevan untuk notasi musik, di mana beberapa simbol dapat muncul dalam jarak spasial yang dekat dan di mana kesalahan lokalisasi dapat memengaruhi interpretasi geometris berikutnya. Recall yang lebih tinggi yang diperoleh oleh Dual-Model juga menunjukkan bahwa dekomposisi tugas memberikan cakupan yang lebih luas untuk kelas simbol target sambil mempertahankan Precision yang lebih tinggi dibandingkan Single-Model.

Hasil menunjukkan bahwa pipeline Dual-Model yang dievaluasi memberikan kinerja deteksi yang lebih tinggi dibandingkan baseline Single-Model yang dievaluasi berdasarkan konfigurasi eksperimental yang digunakan dalam studi ini. Namun, perbandingan ini harus diartikan sebagai perbandingan tingkat pipeline, bukan arsitektur ablation *yang* dikontrol ketat, karena kedua konfigurasi dilatih pada resolusi input yang berbeda.

Untuk memberikan gambaran yang lebih rinci mengenai perbedaan kinerja kedua konfigurasi, analisis per kelas dilakukan pada sejumlah simbol yang merepresentasikan elemen notasi dengan karakteristik berbeda, meliputi simbol yang sering muncul, elemen struktural, serta simbol berukuran kecil dan memiliki karakteristik spasial yang lebih kompleks. Tabel VI menunjukkan bahwa Dual-Model secara umum memberikan peningkatan kinerja pada sebagian besar kelas yang dianalisis.

Peningkatan terlihat jelas pada noteheadBlack, yang merupakan salah satu kelas dengan jumlah anotasi terbesar. Nilai AP@50 meningkat dari 69,35% pada Single-Model menjadi 87,03% pada Dual-Model, sedangkan AP@50-95 meningkat dari 16,66% menjadi 49,99%. Hal serupa terlihat pada elemen struktural seperti paranada dan beam. AP@50 untuk paranada meningkat dari 80,62% menjadi 98,29%, sementara beam meningkat dari 48,01% menjadi 89,24%. Hasil tersebut menunjukkan peningkatan kemampuan deteksi pada objek yang memiliki karakteristik struktural dalam citra notasi musik.

TABEL VI. ANALISIS *PER-CLASS SINGLE* DAN DUAL-MODEL

Kelas	Single AP@50	Single AP@50-95	Dual AP@50	Dual AP@50-95
noteheadBlack	69,35%	16,66%	87,03%	49,99%
staff	80,62%	77,00%	98,29%	86,38%
beam	48,01%	39,50%	89,24%	61,41%
clefG	79,48%	31,20%	99,95%	97,67%
accidentalSharp	78,76%	61,28%	99,49%	80,75%
restQuarter	70,38%	50,70%	99,61%	72,58%
flag16thDown	20,41%	13,91%	54,41%	36,52%
augmentationDot	25,55%	9,80%	45,80%	12,01%
repeatDot	37,85%	13,96%	54,05%	13,91%

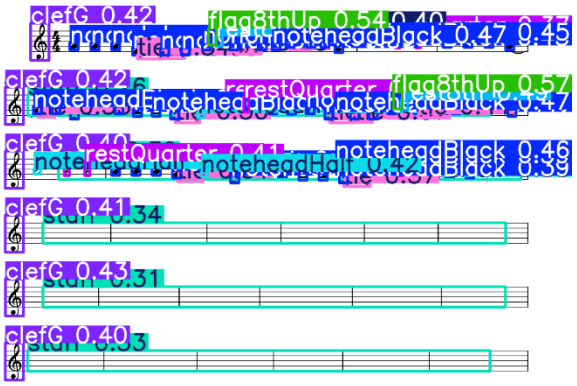
Peningkatan yang cukup besar juga terlihat pada clefG, accidentalSharp, dan restQuarter. Pada clefG, AP@50 meningkat dari 79,48% menjadi 99,95%, sedangkan AP@50-95 meningkat dari 31,20% menjadi 97,67%. Pada accidentalSharp, kedua metrik meningkat dari 78,76% dan 61,28% menjadi 99,49% dan 80,75%. Sementara itu, restQuarter meningkat dari 70,38% menjadi 99,61% pada AP@50.

Meskipun demikian, beberapa simbol tetap menunjukkan kinerja yang relatif lebih rendah. flag16thDown meningkat dari 20,41% menjadi 54,41% pada AP@50, tetapi nilainya masih lebih rendah dibandingkan kelas lain. Demikian pula augmentationDot dan repeatDot masih memiliki AP@50-95 yang rendah, masing-masing 12,01% dan 13,91% pada Dual-Model. Hal ini menunjukkan bahwa simbol berukuran kecil atau memiliki bentuk dan posisi yang lebih sulit dibedakan masih menjadi sumber kesalahan dalam proses deteksi.

Menariknya, keyFlat menunjukkan pola yang berbeda. Single-Model memperoleh AP@50 sebesar 92,72% dan AP50-95 sebesar 76,40%, sedangkan Dual-Model memperoleh 82,10% dan 64,64%. Dengan demikian, peningkatan kinerja Dual-Model tidak terjadi secara seragam pada seluruh kelas. Beberapa kelas yang telah memiliki performa tinggi pada Single-Model dapat mengalami penurunan setelah penggunaan konfigurasi Dual-Model. Temuan ini menunjukkan bahwa pemisahan deteksi berdasarkan karakteristik struktural dan semantik memberikan keuntungan yang berbeda bergantung pada karakteristik masing-masing simbol.

Secara keseluruhan, Dual-Model menunjukkan peningkatan kinerja deteksi pada sebagian besar kelas dibandingkan Single-Model. Peningkatan tersebut mencakup berbagai jenis simbol, baik elemen struktural maupun simbol semantik. Namun, beberapa objek berukuran kecil dan memiliki karakteristik spasial yang kompleks masih menunjukkan nilai AP yang relatif rendah.

Temuan ini menunjukkan bahwa Dual-Model memberikan peningkatan kinerja secara umum, sementara deteksi simbol



Gambar 7. Hasil Inferensi Dual-Model YOLO11

berukuran kecil dan ambigu tetap menjadi tantangan dalam proses Optical Music Recognition (OMR).

C. Geometric Projection Analysis

Setelah simbol musik berhasil dideteksi menggunakan Dual-Model YOLO11, informasi yang dihasilkan masih berupa kelas simbol beserta koordinat bounding box. Informasi tersebut belum dapat digunakan secara langsung untuk menentukan *tinggi nada* karena posisi nada pada notasi balok ditentukan oleh hubungan geometris antara kepala not dan struktur paranada, sehingga koordinat bounding box perlu diinterpretasikan relatif terhadap posisi paranada. Oleh karena itu, penelitian ini menerapkan Geometric Projection Layer sebagai tahap interpretasi geometris sebelum proses rekonstruksi musikal.

Geometric Projection Layer bekerja dengan memanfaatkan hasil deteksi simbol struktural untuk membangun kembali struktur paranada, kemudian memproyeksikan posisi setiap kepala not terhadap garis maupun spasi pada paranada. Melalui proses ini, setiap kepala not memperoleh informasi tinggi nada sesuai posisi relatifnya terhadap sistem paranada sehingga dapat direpresentasikan sebagai nada musikal yang valid. Ilustrasi proses tersebut ditunjukkan pada Gambar 7.

Berdasarkan Gambar 7, hasil inferensi menyediakan lokasi setiap simbol musik yang akan digunakan sebagai masukan pada Geometric Projection Layer. Meskipun seluruh simbol telah berhasil dideteksi, informasi tersebut masih berupa kelas simbol beserta koordinat bounding box. Oleh karena itu, diperlukan proses interpretasi geometris untuk menentukan posisi nada berdasarkan hubungan kepala not terhadap garis paranada.

Untuk memeriksa kontribusi proses penyempurnaan geometris, dilakukan analisis ablasi dengan membandingkan pemetaan geometris baseline dengan pipeline yang menggunakan GPL. Analisis mencakup konfigurasi baseline, dan full-GPL. Setelah dilakukan penyelarasan structure-aware antara urutan prediksi dan referensi MusicXML, diperoleh 58 pasangan prediksi ground truth yang dapat dievaluasi. Percobaan dilakukan dengan menggunakan partitur Hujan Berkat ‘Kan Tercurah yang berupa partitur monofonik. Partitur dipilih karena struktur notasinya singkat dan tidak memiliki bagian yang rumit. Hasil ini tertuang dalam Tabel VII.

TABEL VII. ABLASI GPL

Metric	Baseline	GPL
Matched prediction–GT notes	58	58
Exact tinggi nada accuracy	100,00%	100,00%
MAE (semitone)	0,00	0,00
Mean absolute vertical refinement	—	0,345 px
Median absolute vertical refinement	—	0,241 px
Cases moving toward GT-derived target	—	36/58 (62,07%)

Berdasarkan Tabel VII, data menunjukkan bahwa kedua konfigurasi menghasilkan penetapan *tinggi nada* diskrit yang identik, dengan akurasi tinggi nada 100% dan MAE sebesar 0.00 semitone. Dengan demikian, pada partitur yang dievaluasi, GPL tidak menghasilkan peningkatan tambahan pada akurasi tinggi nada. Hal ini menunjukkan bahwa posisi kepala not yang dihasilkan pada konfigurasi baseline telah berada pada wilayah grid tinggi nada yang sama dengan referensi, sehingga perubahan geometris yang dilakukan GPL belum cukup untuk mengubah keputusan tinggi nada diskrit.

Meskipun tidak menghasilkan perubahan pada akurasi tinggi nada, GPL tetap melakukan proses penyempurnaan posisi geometris kepala not. Analisis visual terhadap 58 prediksi menunjukkan rata-rata perubahan vertikal absolut sebesar 0,345 piksel dan median sebesar 0,241 piksel. Sebanyak 36 dari 58 prediksi (62,07%) bergerak menuju target geometris yang diturunkan dari referensi tinggi nada. Hasil ini menunjukkan bahwa GPL berfungsi sebagai mekanisme penyempurnaan posisi geometris sebelum proses pemetaan ke tinggi nada diskrit, tetapi efek penyempurnaan tersebut belum menghasilkan perubahan pada keputusan tinggi nada pada partitur yang diuji.

Perlu dicatat bahwa target geometris yang digunakan dalam analisis ini diturunkan dari informasi tinggi nada MusicXML dan geometri paranada yang terdeteksi, bukan merupakan anotasi koordinat piksel independen. Oleh karena itu, hasil pengukuran geometris tersebut digunakan sebagai analisis diagnostik, bukan sebagai metrik akurasi lokalisasi piksel secara independen.

Hasil dari Geometric Projection Layer berupa informasi tinggi nada untuk setiap kepala not yang selanjutnya digunakan pada tahap Musical Context Reconstruction.

D. Musical Context Reconstruction

Setelah proses Geometric Projection Layer menghasilkan informasi tinggi nada untuk setiap kepala not, tahap berikutnya adalah Musical Context Reconstruction. Tahap ini bertujuan membangun representasi musik dari hasil deteksi simbol dengan mengintegrasikan informasi geometris, atribut simbol, urutan pembacaan, dan struktur musikal. Pada tahap ini, kepala not dihubungkan dengan atribut musikal yang berkaitan, seperti accidental, titik augmentasi, beam, tie, dan rest, sehingga terbentuk music event yang merepresentasikan kejadian musikal secara terstruktur.

TABEL VIII. HASIL MUSIC EVENTS

Start	End	Tinggi nada
500	525	F4
550	575	F4
600	625	F4
650	675	F4
700	725	G4
750	775	A4
800	875	B-4
850	875	F4
900	950	F4
950	975	A4

Berbeda dengan proses deteksi objek yang menghasilkan kelas simbol dan koordinat bounding box, MCR menyusun hubungan antar simbol berdasarkan posisi spasial, urutan pembacaan, serta aturan notasi musik. Informasi *tinggi nada* dari GPL menjadi atribut utama note event, sedangkan informasi durasi diperoleh dari karakteristik kepala not, flag, beam, dan titik augmentasi. Event *rest* juga dipertahankan sebagai bagian dari representasi temporal. Dengan demikian, keluaran MCR tidak lagi berupa kumpulan simbol yang berdiri sendiri, tetapi menjadi rangkaian music event yang memiliki informasi *tinggi nada*, durasi, dan posisi temporal.

Pada proses rekonstruksi, sistem menghasilkan 63 raw music events pada partitur yang diuji, yang terdiri atas 58 note events dan 5 rest events. Jumlah tersebut tetap sama ketika konfigurasi time signature diuji pada 4/4 dan 6/8. Hal ini menunjukkan bahwa time signature tidak memengaruhi pembentukan event awal dari hasil deteksi dan GPL. Perbedaan akibat time signature muncul pada tahap validasi struktur musik, ketika event ditempatkan ke dalam measure dan event yang melintasi batas birama dapat direpresentasikan melalui pemisahan dan tie reconstruction. Pada konfigurasi 4/4 terbentuk 20 measure, sedangkan konfigurasi 6/8 membentuk 21 measure, sementara jumlah event pada *score matrix* tetap 63.

Untuk menjaga konsistensi struktur temporal, hasil event selanjutnya divalidasi menggunakan struktur musik berbasis Music21. Time signature diterapkan sebelum pembentukan measure, kemudian sistem melakukan rekonstruksi tie untuk event yang melintasi batas birama. Tahap ini menghasilkan representasi tervalidasi yang dapat memiliki jumlah objek berbeda dari raw music events, karena satu kejadian musikal dapat direpresentasikan sebagai beberapa bagian setelah disesuaikan dengan struktur birama. Oleh karena itu, jumlah event setelah validasi digunakan sebagai representasi struktural, bukan sebagai ukuran akurasi deteksi.

Hasil rekonstruksi konteks musikal ditunjukkan pada Tabel VIII. Setiap music event telah memiliki atribut utama berupa waktu mulai (start), waktu berakhir (end), informasi *tinggi nada*. Representasi ini menunjukkan bahwa hasil deteksi telah berhasil dikonversi menjadi data musik simbolik yang siap digunakan pada proses transliterasi.

Berdasarkan Tabel VIII, music event memiliki posisi temporal yang memungkinkan event yang berlangsung berurutan maupun secara bersamaan direpresentasikan secara terpisah. Informasi tersebut kemudian digunakan sebagai masukan pada tahap transliterasi ke not angka serta pembentukan representasi musik lainnya.

Informasi start dan end digunakan sebagai representasi posisi temporal setiap music event. Informasi tersebut diperlukan untuk menjaga urutan pembacaan nada serta sinkronisasi antar simbol ketika proses rendering dilakukan ke dalam format not angka maupun MusicXML. Dengan adanya representasi temporal, sistem dapat merepresentasikan urutan dan hubungan waktu antar music event sebelum proses transliterasi dilakukan.

Tahap Musical Context Reconstruction merupakan pembeda utama framework yang diusulkan dibandingkan pendekatan Optical Music Recognition berbasis deteksi objek. Sebagian besar penelitian menghasilkan keluaran berupa simbol beserta koordinatnya, sedangkan framework ini mengubah hasil deteksi menjadi music events yang telah memiliki atribut musikal dan temporal. Representasi tersebut menjadi dasar pembentukan struktur musik yang konsisten sebelum proses transliterasi dilakukan.

Selain digunakan sebagai masukan proses transliterasi, representasi music event juga memudahkan proses analisis terhadap hasil Optical Music Recognition. Setiap event dapat ditelusuri kembali terhadap simbol asalnya sehingga mempermudah proses validasi apabila terjadi kesalahan interpretasi. Pendekatan ini juga memungkinkan pengembangan lebih lanjut, seperti analisis harmoni, ekstraksi melodi, maupun konversi ke format musik simbolik lainnya tanpa harus melakukan proses deteksi ulang.

E. End-to-End Transliterasi ke Not Angka

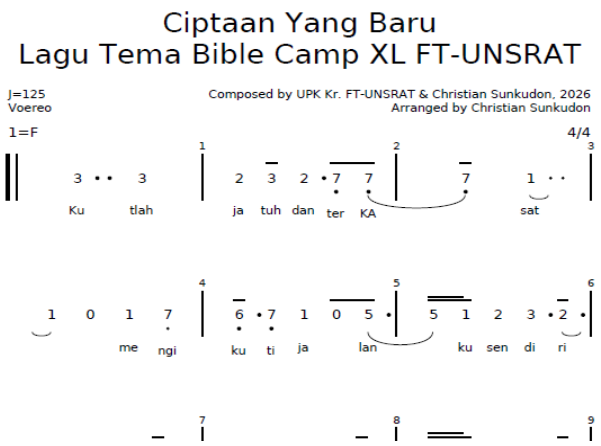
Tahap akhir pada framework yang diusulkan adalah proses rendering hasil Musical Context Reconstruction menjadi not angka (Jianpu). Pada tahap ini, representasi music event yang telah memiliki informasi tinggi nada, durasi, urutan pembacaan, serta hubungan antar simbol dikonversi menjadi not angka sesuai aturan transliterasi. Sebelum proses rendering dilakukan, representasi musik terlebih dahulu divalidasi melalui struktur MusicXML untuk memastikan konsistensi hubungan antar simbol, sehingga keluaran not angka yang dihasilkan tetap sesuai dengan struktur partitur asli.

Hasil transliterasi ditunjukkan pada Gambar 8. Berdasarkan hasil tersebut, sistem mampu menghasilkan representasi not angka secara otomatis dari citra partitur tanpa memerlukan proses anotasi atau penyuntingan manual. Informasi nada, tanda istirahat, serta urutan pembacaan berhasil dipertahankan sehingga not angka yang dihasilkan memiliki kesesuaian dengan struktur musik pada partitur masukan.

F. Hasil Evaluasi Framework *Transliterasi Pada Partitur Uji*

Untuk mengevaluasi kemampuan framework secara menyeluruh, pengujian dilakukan pada beberapa jenis partitur dengan karakteristik yang berbeda, meliputi partitur monofonik, SATB, serta partitur yang mengandung variasi simbol musikal seperti accidental, beam, tie, slur, dan rest.

Evaluasi ini bertujuan untuk mengetahui konsistensi proses transliterasi pada berbagai kondisi partitur yang umum



Gambar 8. Hasil Notasi Angka

dijumpai tanpa melakukan penyesuaian parameter ataupun proses anotasi tambahan.

Hasil evaluasi dirangkum pada Tabel IX. Setiap partitur diproses menggunakan framework yang sama, dimulai dari tahap deteksi simbol menggunakan Dual-Model YOLO11, dilanjutkan dengan Geometric Projection Layer, Musical Context Reconstruction, hingga proses rendering menjadi not angka (Jianpu). Selain menghasilkan keluaran utama berupa not angka, sistem juga membentuk representasi MusicXML sebagai media validasi struktur musik sebelum proses *rendering* akhir dilakukan.

Enam partitur dipilih untuk merepresentasikan variasi karakteristik musik, meliputi partitur monofonik, himne, dan polifonik SATB, serta variasi simbol seperti accidental, beam, tie, slur, dan rest. Setiap partitur diproses menggunakan framework yang sama, mulai dari deteksi simbol dengan Dual-Model YOLO11, Geometric Projection Layer, Musical Context Reconstruction, hingga rendering menjadi not angka (Jianpu). Evaluasi dilakukan dengan membandingkan hasil transliterasi terhadap representasi referensi sehingga metrik yang diperoleh merepresentasikan performa transliterasi secara *end-to-end*.

Hasil evaluasi menunjukkan adanya variasi tingkat akurasi transliterasi pada enam partitur yang diuji. Dua partitur, yaitu Fair Phyllis I Saw dan Ciptaan Yang Baru, mencapai akurasi 100%. Bila Kulihat Bintang Gemerlapan mencapai akurasi 95,8%, sedangkan Hujan Berkat ’Kan Tercurah mencapai 95,24%. Pada partitur Hymne INKINDO diperoleh akurasi sebesar 71,55%, sementara Mars INKINDO memperoleh akurasi sebesar 18,81%. Hasil tersebut menunjukkan bahwa performa framework dipengaruhi oleh kompleksitas struktur dan hubungan antar simbol pada masing-masing partitur.

Pada Bila Kulihat Bintang Gemerlapan, dari total 119 simbol yang dievaluasi, sebanyak 114 simbol berhasil ditransliterasikan dengan benar dan 5 simbol mengalami kesalahan interpretasi. Kesalahan tersebut berkaitan dengan garis tie dan posisi antar not yang relatif rapat. Sementara itu, penurunan akurasi pada Hymne INKINDO dan terutama Mars INKINDO menunjukkan bahwa sistem masih menghadapi kesulitan dalam mempertahankan hubungan antar simbol pada struktur musik yang lebih kompleks, khususnya pada partitur polifonik. Kondisi ini menunjukkan bahwa kesalahan transliterasi tidak hanya berkaitan dengan pengenalan simbol

TABEL IX. EVALUASI HASIL TRANSLITERASI

No.	Partitur	Total Simbol	Benar	Salah	Akurasi
1.	Bila Kulihat Bintang Gemerlapan	119	114	5	95,8%
2.	Fair Phyllis I saw	297	297	0	100%
3.	Mars INKINDO	638	120	518	18,81%
4.	Hymne INKINDO	348	249	99	71,55%
5.	Hujan Berkat ‘Kan Tercurah	63	60	3	95,24%
6.	Ciptaan Yang Baru	453	453	0	100%

secara individual, tetapi juga dengan proses rekonstruksi hubungan musikal setelah tahap deteksi.

Hasil evaluasi tersebut menunjukkan bahwa performa transliterasi tidak hanya ditentukan oleh kemampuan deteksi simbol, tetapi juga oleh proses interpretasi hubungan musikal pada tahap Geometric Projection Layer dan Musical Context Reconstruction. Dengan demikian, keberhasilan framework bergantung pada integrasi seluruh komponen yang bekerja secara berurutan untuk menghasilkan representasi musik simbolik yang konsisten sebelum proses rendering ke not angka dilakukan.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa framework mampu melakukan proses transliterasi secara *end-to-end* pada berbagai jenis partitur, tetapi tingkat keberhasilannya berbeda sesuai dengan karakteristik dan kompleksitas partitur. Hasil yang tinggi pada beberapa partitur menunjukkan bahwa rangkaian deteksi simbol, pemrosesan geometris, dan rekonstruksi konteks musikal dapat menghasilkan transliterasi yang sesuai. Sebaliknya, hasil yang lebih rendah pada partitur tertentu menunjukkan bahwa hubungan antar simbol dan struktur musikal yang lebih kompleks masih menjadi tantangan bagi framework yang diusulkan.

G. Discussion

Hasil pengujian menunjukkan bahwa pemisahan deteksi menjadi Model Struktural dan Model Semantik memberikan informasi yang saling melengkapi dalam proses transliterasi. Model Struktural mengenali elemen pembentuk struktur notasi, seperti staff, brace, beam, dan kepala not, sedangkan Model Semantik mengenali simbol yang berkaitan dengan makna musikal, seperti accidental, rest, dynamic marking, dan simbol pendukung lainnya. Informasi dari kedua model selanjutnya digunakan oleh Geometric Projection Layer dan Musical

Context Reconstruction untuk membentuk representasi musik sebelum proses transliterasi.

Penggunaan MusicXML sebagai representasi internal juga memberikan mekanisme validasi terhadap struktur musik sebelum proses transliterasi akhir. Pada penelitian ini, MusicXML tidak digunakan sebagai keluaran utama, tetapi sebagai media untuk memastikan konsistensi hubungan antar simbol dan struktur musikal sebelum direnderer menjadi not angka (Jianpu). Pendekatan tersebut memungkinkan hasil deteksi tidak langsung dikonversi menjadi not angka, tetapi terlebih dahulu direkonstruksi menjadi representasi music event yang mempertahankan informasi tinggi nada, durasi, dan hubungan temporal.

Hasil evaluasi menunjukkan bahwa akurasi transliterasi tidak hanya ditentukan oleh performa deteksi objek. Meskipun model YOLO menghasilkan kinerja deteksi yang tinggi, kesalahan masih dapat muncul pada tahap pengelompokan simbol, penentuan hubungan antar nada, pembentukan music event, dan transliterasi. Hal ini terlihat pada partitur Mars INKINDO dan Hymne INKINDO yang memiliki struktur musikal lebih kompleks dan menghasilkan akurasi lebih rendah dibandingkan beberapa partitur lainnya. Framework juga masih dievaluasi pada partitur cetak dengan kualitas citra yang baik, sehingga kemampuan terhadap partitur hasil pemindaian berkualitas rendah dan tulisan tangan belum dapat disimpulkan. Selain itu, proses Musical Context Reconstruction masih menggunakan pendekatan berbasis aturan, sehingga fleksibilitas terhadap variasi hubungan antar simbol masih menjadi keterbatasan.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Framework Optical Music Recognition yang diusulkan berhasil mengintegrasikan Dual-Model YOLO11, Geometric Projection Layer, dan Musical Context Reconstruction untuk melakukan transliterasi notasi balok menjadi not angka (Jianpu) secara *end-to-end*. Dual-Model memperoleh Precision 92,31%, Recall 89,35%, F1 90,80%, mAP@50 sebesar 92,36%, dan mAP50–95 sebesar 68,98%. Hal ini menunjukkan peningkatan kinerja dibandingkan Single-Model pada konfigurasi eksperimen yang digunakan.

Geometric Projection Layer berhasil menginterpretasikan posisi kepala not menjadi informasi tinggi nada. Pada ablasi terhadap 58 pasangan prediction–ground truth, baseline dan GPL menghasilkan akurasi tinggi nada 100% dengan MAE 0 semitone. GPL tidak mengubah keputusan *tinggi nada* pada partitur yang diuji, tetapi melakukan penyempurnaan geometris dengan rata-rata perubahan vertikal absolut 0,345 piksel. Musical Context Reconstruction selanjutnya mengintegrasikan informasi tinggi nada, durasi, urutan pembacaan, dan hubungan antar simbol menjadi music events sebagai dasar transliterasi.

Evaluasi enam partitur menghasilkan akurasi transliterasi 18,81%–100%. Performa lebih rendah ditemukan pada partitur dengan struktur musikal yang lebih kompleks, khususnya polifonik. Hasil tersebut menunjukkan bahwa keberhasilan transliterasi tidak hanya bergantung pada deteksi simbol, tetapi juga pada kemampuan rekonstruksi hubungan musikal secara keseluruhan.

B. Saran

Pengembangan selanjutnya dapat memperluas evaluasi pada partitur hasil pemindaian berkualitas rendah, tulisan tangan, serta tata letak musikal yang lebih kompleks. Musical Context Reconstruction yang masih menggunakan pendekatan berbasis aturan juga dapat dikembangkan menggunakan Graph Neural Network (GNN) atau Transformer agar hubungan antar simbol dapat dipelajari secara lebih adaptif. Selain itu, penggunaan dataset dan partitur uji yang lebih beragam diperlukan untuk mengevaluasi kemampuan generalisasi framework secara lebih menyeluruh.

V. KUTIPAN

[1] R. L. Abdulazeez and F. Alizadeh, “Deep Learning-Based Optical Music Recognition for Semantic Representation of Non-overlap and Overlap Music Notes,” *ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY*, vol. 12, no. 1, pp. 79–87, Mar. 2024, doi: 10.14500/aro.11402.

[2] M. Alfaro-Contreras, J. M. Iñesta, and J. Calvo-Zaragoza, “Optical Music Recognition for homophonic scores with neural networks and synthetic music generation,” *Int. J. Multimed. Inf. Retr.*, vol. 12, no. 1, Jun. 2023, doi: 10.1007/s13735-023-00278-5.

[3] Y. Liu, R. Wu, Y. Wu, L. Luo, and W. Xu, “A Stave-Aware Optical Music Recognition on Monophonic Scores for Camera-Based Scenarios,” *Applied Sciences (Switzerland)*, vol. 13, no. 16, Aug. 2023, doi: 10.3390/app13169360.

[4] J. Calvo-Zaragoza, J. Hajic, and A. Pacha, “Understanding Optical Music Recognition,” *ACM Comput. Surv.*, vol. 53, no. 4, Sep. 2020, doi: 10.1145/3397499.

[5] A. Pacha, J. Hajič, and J. Calvo-Zaragoza, “A baseline for general music object detection with deep learning,” *Applied Sciences (Switzerland)*, vol. 8, no. 9, Aug. 2018, doi: 10.3390/app8091488.

[6] F. Luo, Y. Dai, J. Fuentes, W. Ding, and X. Zhang, “M-DETR: Multi-scale DETR for Optical Music Recognition,” *Expert Syst. Appl.*, vol. 249, 2024, doi: 10.1016/j.eswa.2024.123664.

[7] A. Ríos-Vila, D. Rizo, J. M. Iñesta, and J. Calvo-Zaragoza, “End-to-end Optical Music Recognition for pianoform sheet music,” in *International Journal on Document Analysis and Recognition*, Springer Science and Business Media Deutschland GmbH, Sep. 2023, pp. 347–362. doi: 10.1007/s10032-023-00432-z.

[8] M. Alfaro-Contreras, A. Ríos-Vila, J. J. Valero-Mas, J. M. Iñesta, and J. Calvo-Zaragoza, “Decoupling music notation to improve end-to-end Optical Music Recognition,” *Pattern Recognit. Lett.*, vol. 158, pp. 157–163, Jun. 2022, doi: 10.1016/j.patrec.2022.04.032.

[9] Y. Zhang, Z. Huang, Y. Zhang, and K. Ren, “A detector for page-level handwritten music object recognition based on deep learning,” *Neural Comput. Appl.*, vol. 35, no. 13, 2023, doi: 10.1007/s00521-023-08216-6.

[10] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-End Object Detection with Transformers,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020. doi: 10.1007/978-3-030-58452-8_13.

[11] C.-Y. Wang and H.-Y. M. Liao, “YOLOv1 to YOLOv10: The Fastest and Most Accurate Real-time Object Detection Systems,” *APSIPA Trans. Signal Inf. Process.*, vol. 13, no. 1, 2024, doi: 10.1561/116.20240058.

[12] Y. Li, H. Liu, Q. Jin, M. Cai, and P. Li, “TrOMR:Transformer-Based Polyphonic Optical Music Recognition,” Aug. 2023, [Online]. Available: http://arxiv.org/abs/2308.09370

[13] A. Ríos-Vila, J. Calvo-Zaragoza, and T. Paquet, “Sheet Music Transformer: End-To-End Optical Music Recognition Beyond Monophonic Transcription,” Apr. 2024, [Online]. Available: http://arxiv.org/abs/2402.07596

[14] A. Rios-Vila, E. Fuentes-Martinez, and F. J. Castellanos, “An implicit layout-aware transformer for full-page end-to-end Optical Music Recognition,” *Int. J. Multimed. Inf. Retr.*, vol. 14, no. 4, Dec. 2025, doi: 10.1007/s13735-025-00385-5.

[15] L. Tuggener, Y. Putra Satyawan, A. Pacha, J. Schmidhuber, and T. Stadelmann, “The DeepScoresV2 Dataset and Benchmark for Music

Object Detection," Milan, Italy, Sep. 2020. doi: https://doi.org/10.5281/zenodo.4012193.
[16] J. Hajic and P. Pecina, "The MUSCIMA++ Dataset for Handwritten Optical Music Recognition," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2017. doi: 10.1109/ICDAR.2017.16.
[17] C. Garrido-Munoz, A. Rios-Vila, and J. Calvo-Zaragoza, "A holistic approach for image-to-graph: application to Optical Music Recognition," *International Journal on Document Analysis and Recognition*, vol. 25, no. 4, 2022, doi: 10.1007/s10032-022-00417-4.
[18] G. Yang, M. Zhang, L. Qiu, Y. Wan, and N. A. Smith, "Toward a More Complete OMR Solution," Aug. 2024, [Online]. Available: http://arxiv.org/abs/2409.00316
[19] A. Pacha, J. Calvo-Zaragoza, and J. Hajič, "Learning notation graph construction for full-pipeline Optical Music Recognition," in *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, 2019.
[20] J. Dimpudus, A. Sambul, and A. Lumenta, "Transliteration Block Notation Application Into Number Notation Using The MusicXML Format," *Jurnal Teknik Informatika*, vol. 7, no. 1, pp. 75–82, Jan. 2022, Accessed: Oct. 16, 2024. [Online]. Available: https://ejournal.unsrat.ac.id/v3/index.php/informatika/article/view/36298
[21] R. Yang, M. Giraud, and F. Levé, "Music with Numbers: Jianpu Number-Based Notation in Cultural Heritage and Digital Humanities," London, Nov. 2025. [Online]. Available: https://hal.science/hal-05029666v1
[22] F. Bu, R. Li, Z. Li, Y. Li, L. Fan, and P. Huang, "The Renaissance of Expert Systems: Optical Recognition of Printed Chinese Jianpu Musical Scores with Lyrics," Dec. 2025, [Online]. Available: http://arxiv.org/abs/2512.14758



Christian Cleavy Sunkudon lahir di Manado, tanggal 24 Desember 2001, pada pasangan Alfie Gustaf Paul Sunkudon dan Stely Susan Tiwow. Penulis mulai menempuh pendidikan pada TK Katolik St. Theresia Malalayang, pada tahun 2005 sampai tahun 2006, untuk Tingkat A. Kemudian melanjutkan pendidikan Tingkat B mulai tahun 2006 sampai tahun 2007.

Setelah menamatkan pendidikan pada Taman Kanak-kanak, kemudian melanjutkan pendidikan Sekolah Dasar pada SD Katolik St. Theresia Malalayang, mulai tahun 2007 sampai 2013. Kemudian melanjutkan pendidikan menengah pertama pada SMP Katolik St. Theresia mulai tahun 2013 sampai tahun 2016.

Setelah menamatkan pendidikan menengah pertama, penulis melanjutkan pendidikan tingkat atas pada SMA Katolik Rex Mundi Manado dari tahun 2016 sampai tahun 2019. Penulis terdaftar sebagai mahasiswa Strata 1 (S1) pada Program Studi Teknik Informatika, Jurusan Teknik Elektro, Fakultas Teknik Universitas Sam Ratulangi mulai tanggal 1 Agustus 2019, hingga 7 Desember 2023 saat Penulis menamatkan status sebagai Sarjana.

Penulis aktif sebagai pekerja lepas, dan juga melatih paduan suara dan sebagai pianis dalam perlombaan paduan suara.