

BAB II

TINJAUAN PUSTAKA

Optical Music Recognition (OMR) adalah proses mengonversi notasi musik dari media cetak menjadi file dengan format yang dapat dibaca oleh mesin, misalnya dengan MEI atau MusicXML. OMR memiliki peranan yang sangat penting dalam upaya melestarikan dan mengarsipkan partitur-partitur cetak secara digital. Tujuannya adalah untuk memudahkan akses dan analisis musik secara digital. Tantangan utama dalam proses OMR terletak pada deteksi elemen-elemen berukuran kecil, seperti tanda mula, khususnya pada partitur cetak yang terdistorsi. Sistem OMR tradisional dirasa tidak mampu mendeteksi elemen-elemen secara akurat. Hal ini menunjukkan bahwa perlu ada pendekatan baru, seperti *deep learning* yang dapat melakukan deteksi dengan lebih adaptif [11].

CNN merupakan arsitektur jaringan saraf tiruan (JST) yang telah membantu banyak bidang dalam deteksi citra, misalnya pengenalan tulisan tangan dan deteksi objek. CNN menggunakan lapisan konvolusi untuk mempelajari fitur-fitur spasial pada gambar sehingga cocok digunakan untuk mendeteksi elemen-elemen kompleks seperti tanda mula pada partitur bernotasi balok [12].

Meskipun CNN menunjukkan hasil yang baik dalam melakukan klasifikasi, algoritma ini memiliki keterbatasan dalam deteksi objek, sehingga diperlukan algoritma untuk melakukan *object detection*. YOLO adalah salah satu metode deteksi objek yang populer karena kecepatan dan akurasinya. Berbeda dengan Faster R-CNN yang sedikit lambat, YOLO dapat memproses citra dengan satu kali proses (*single-stage*) untuk memprediksi *bounding box* dan probabilitas kelas secara bersamaan [13]. Pada algoritma yang terbaru, YOLOv8 menunjukkan peningkatan signifikan dalam *mean average precision* dan kecepatan inferensinya. Karena itu, algoritma ini menjadi pilihan yang cocok untuk digunakan dalam deteksi tanda mula pada partitur bernotasi balok.

Penelitian yang dilakukan oleh Calvo-Zaragoza, dkk. [14] pada tahun 2017 tidak menerapkan model deteksi dan klasifikasi pada halaman partitur musik yang utuh. Klasifikasi simbol/objek musik dilakukan pada gambar yang sudah yang

sudah terpotong (*isolated*) dengan menggunakan CNN sebagai *feature extractor* dengan SVM, KNN, atau *Random Forest* sebagai klasifikator. Penelitian ini menghasilkan *error rate* yang rendah. Akan tetapi, keterbatasan utama penelitian ini adalah metode tersebut hanya bekerja pada objek yang sudah dipotong (*segmented*) dan tidak menerapkan deteksi pada halaman partitur utuh. Celah ini menyebabkan metode tersebut tidak dapat diterapkan secara langsung pada kasus nyata seperti pada citra hasil *scan*.

Pacha, dkk. [15] pada tahun 2018 melakukan penelitian untuk membandingkan model deteksi (Faster R-CNN, RetinaNet) dan segmentasi piksel (U-Net) pada tiga jenis dataset. Ketiga model menghasilkan akurasi (mAP) yang relatif rendah karena menggunakan model generik untuk berbagai jenis partitur tanpa spesialisasi yang mendalam. Hal ini mengakibatkan rendahnya nilai mAP saat melakukan deteksi pada partitur dengan variasi simbol yang kompleks, sehingga diperlukan pendekatan yang lebih terfokus untuk meningkatkan akurasi deteksi pada komponen krusial seperti tanda mula.

Pacha, dkk. [16] pada penelitian yang lain kembali membandingkan tiga model deteksi, yaitu Faster R-CNN, R-FCN, dan SSD. Faster R-CNN memberikan hasil terbaik dengan nilai mAP sebesar 80% pada potongan kecil partitur dan 76% pada satu halaman partitur. Pada penelitian ini, model masih kesulitan mendeteksi objek yang berukuran kecil dan tumpang tindih, sehingga menghasilkan nilai mAP sebesar 77,80% pada tanda mula. Selain itu, tanda mula hanya dideteksi dan tidak diklasifikasikan lebih lanjut ke dalam satu dari lima belas jenis tanda mula.

Nawade, dkk. [17] pada tahun 2021 melakukan penelitian yang berfokus pada klasifikasi simbol musik dengan tulisan tangan. Model utama yang digunakan adalah MobileNetV2 dengan pendekatan *transfer learning* dan pelatihan dari awal. Selain itu, penulis juga mengombinasikan CNN dengan *classifier* tradisional seperti SVM, KNN, dan *Random Forest*. Hasil eksperimen menunjukkan akurasi sangat tinggi, mencapai nilai 99,75% pada skenario pelatihan dari awal. Penelitian ini menunjukkan bahwa CNN ringan seperti MobileNetV2 sangat efektif untuk OMR. Namun, ruang lingkup penelitian ini terbatas pada simbol musik terisolasi dan belum mencakup deteksi simbol pada partitur utuh.

Setyo dan Kusuma [18] pada tahun 2024 menggunakan pendekatan *deep learning* untuk melakukan OMR, seperti dengan *Vision Transformer* (ViT), CoAtNet-0, dan ConvNeXt-Tiny. Fokus utama penelitian ini adalah untuk klasifikasi simbol musik dan membandingkan performa arsitektur CNN konvensional dan CNN-*Transformer hybrid*. Hasil penelitian menunjukkan bahwa ConvNeXt-Tiny memberikan hasil terbaik dengan akurasi 98,15% dan 98,43% pada dua dataset berbeda. Penelitian ini menegaskan bahwa *deep learning* sangat efektif untuk klasifikasi simbol musik. Namun, pendekatan yang digunakan masih untuk potongan simbol musik dan belum mengakomodasi skenario deteksi pada partitur asli.

Nugroho dan Zahra [19] pada penelitiannya di tahun 2024 mengusulkan sistem OMR non *end-to-end* untuk notasi dengan memisahkan proses pengenalan posisi (*pitch*) dan durasi not ke dalam dua model CNN yang berbeda. Pendekatan ini dipilih untuk mengurangi kompleksitas sistem dan meningkatkan akurasi masing-masing klasifikasi (*pitch* dan durasi). Hasil pengujian menunjukkan akurasi 97,88% untuk klasifikasi posisi dan 99,23% untuk klasifikasi durasi. Namun, penelitian ini masih terbatas pada pengenalan not dan belum membahas *accidentals* dan tanda mula yang juga memengaruhi *pitch* dari suatu notasi.

Terdapat beberapa penelitian yang berhasil dalam mendeteksi simbol notasi balok, tetapi hanya terdapat sedikit yang berfokus pada deteksi dan klasifikasi tanda mula, padahal komponen ini penting untuk menentukan konteks notasi dalam suatu partitur. Hal ini menjadi peluang untuk mengembangkan model klasifikasi tanda mula yang dapat mengatasi variasi penulisan, derau, dan gangguan lainnya. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan mengembangkan sistem deteksi klasifikasi tanda mula berbasis YOLOv8 dan CNN dengan akurasi tinggi pada dataset musik yang beragam.

Tabel 2.1 Perbandingan Kajian Penelitian

No.	Peneliti (Tahun)	Metode	Fokus Objek	Kelebihan	Kekurangan / Celah Riset	Metrik Evaluasi
1.	Calvo-Zaragoza dkk. (2017) [14]	CNN (<i>feature extractor</i>) + KNN, SVM, <i>Random Forest</i>	Klasifikasi simbol musik tulis tangan	Model sangat efisien dan akurat dengan <i>error rate</i> di bawah 3%.	Fokus hanya pada klasifikasi simbol yang sudah terpotong, tidak mendeteksi dari partitur.	<i>Accuracy</i>
2.	Pacha dkk. (2018) [15]	Faster R-CNN, RetinaNet, U-Net	Deteksi simbol musik cetak dan tulis tangan	Memberikan <i>baseline</i> pada dataset yang heterogen, mengevaluasi deteksi vs segmentasi piksel.	Hasil akurasi keseluruhan masih rendah (mAP < 25%).	mAP, w-mAP
3.	Pacha dkk. (2018) [16]	Faster R-CNN, SSD, R-FCN (detektor), Inception-ResNet-v2 & ResNet101	Deteksi simbol tulis tangan	Mencapai akurasi tinggi (mAP $\pm 80\%$) pada simbol tulis tangan dan mampu mendeteksi banyak simbol musik secara lengkap.	Kesulitan mendeteksi objek kecil dan simbol tumpang tindih, menghasilkan nilai mAP 77,80% pada objek tanda mula.	mAP, w-mAP

4.	Nawade dkk. (2021)	MobileNetV2 + SVM/KNN/RF	Klasifikasi simbol musik tulis tangan	Akurasi sangat tinggi ($\pm 99,75\%$), model ringan dan efisien	Fokus hanya pada potongan objek, tidak melakukan deteksi pada partitur utuh	<i>Accuracy, Precision, Recall, F1-score</i>
5.	Setyo & Kusuma (2024)	ViT, CoAtNet, ConvNeXt	Klasifikasi simbol musik	Menggunakan arsitektur <i>deep learning</i> modern, performa sangat baik (akurasi > 98%)	Hanya mengklasifikasi potongan objek, belum mencerminkan kondisi partitur asli, belum menerapkan deteksi objek	<i>Accuracy, Weighted F1-score</i>
6.	Nugroho & Zahra (2024)	CNN	Posisi & durasi not	Akurasi sangat tinggi ($\pm 97\%$), pendekatan modular mempermudah optimasi tiap tugas	Masih bergantung pada <i>template matching</i> pada tahap awal, belum menggunakan deteksi objek	<i>Accuracy, Precision, Recall, F1-score</i>
7.	Penelitian Ini (2025)	YOLOv8 (detektor) + custom CNN (klasifikator)	Tanda mula (<i>key signature</i>)	Pendekatan <i>two-stage pipeline</i> yang memisahkan deteksi dan klasifikasi, menggunakan <i>negative samples</i> , ringan untuk <i>real-time</i> .	Fokus terbatas pada tanda mula, tidak membaca not dan tanda dinamika, performa klasifikasi bergantung pada hasil deteksi.	<i>Accuracy, Precision, Recall, F1-score, mAP50, mAP50-95</i>