Mars: Jurnal Teknik Mesin, Industri, Elektro dan Ilmu Komputer
Volume. 2 Nomor. 5 Tahun 2024
e-ISSN : 3031-8742, dan p-ISSN : 3031-8750, Hal. 164-170
DOI :https://doi.org/10.61132/mars.v2i5.1213
Available online at: https://journal.arteii.or.id/index.php/Mars

Klasifikasi Instrumen Musik dari Sinyal Audio menggunakan *ResNet*

Fathoni Dwi Atmoko
Program Studi Teknik Informatika, Universitas Nahdlatul Ulama, Indonesia
**Penulis Korespondensi: fathonidwiatmoko@gmail.com*

Abstract. *This study presents the implementation of Transfer learning using the ResNet-18 architecture for classifying 10 musical instrument categories based on visual representations of audio signals. The audio waveform is transformed into image-like inputs appropriate for CNN processing, accompanied by data augmentation and ImageNet-standard normalization. ResNet-18 is utilized due to its efficient feature extraction capability enabled by residual blocks, which help overcome vanishing gradient issues. The model was trained for 10 Epochs using the AdamW optimizer and Cross-Entropy Loss. Experimental results show that the model achieved a maximum validation accuracy of 77.35%, with a stable downward trend in training loss, indicating effective feature learning. However, several misclassification cases were observed, particularly among instruments with similar spectral characteristics, such as drum–violin and tabla–sitar. These findings demonstrate that while ResNet-18 performs reliably for musical instrument classification, further improvements remain possible through deeper architectures like ResNet-50, more comprehensive hyperparameter optimization, and the use of richer audio representations such as Mel-Spectrograms. This research provides an essential foundation for developing automated music analysis systems powered by Deep Learning.*

Keywords: *Musical Instrument Classification, ResNet-18, Transfer learning, Convolutional Neural Network, Deep Learning.*

Abstrak. Penelitian ini membahas implementasi *Transfer learning* menggunakan arsitektur *ResNet*-18 untuk melakukan klasifikasi 10 jenis instrumen musik berdasarkan representasi visual sinyal audio. Transformasi sinyal audio dilakukan dengan mengonversinya menjadi citra yang sesuai sebagai masukan CNN, disertai proses augmentasi dan normalisasi standar *ImageNet*. *ResNet*-18 dipilih karena kemampuannya mempelajari fitur visual secara efektif melalui mekanisme residual block yang mampu mengatasi permasalahan vanishing gradient. Model dilatih selama 10 *Epoch* menggunakan optimizer AdamW dan fungsi loss Cross-Entropy. Hasil pengujian menunjukkan bahwa model mampu mencapai akurasi validasi maksimum sebesar 77,35%, dengan tren konvergensi yang stabil pada loss pelatihan. Meskipun demikian, masih dijumpai beberapa kesalahan klasifikasi pada instrumen dengan karakteristik spektral yang mirip, seperti drum–violin dan tabla–sitar. Temuan ini menegaskan bahwa *ResNet*-18 efektif untuk tugas klasifikasi instrumen musik, namun masih terdapat ruang peningkatan melalui penggunaan arsitektur yang lebih dalam seperti *ResNet*-50, optimasi hyperparameter lebih lanjut, serta eksplorasi representasi audio yang lebih kaya seperti Mel-Spektrogram. Penelitian ini menjadi landasan penting bagi pengembangan sistem analisis musik otomatis berbasis Deep Learning.

Kata kunci: Klasifikasi instrumen musik, *ResNet*-18, *Transfer learning*, Convolutional Neural Network, deep learning.

1. LATAR BELAKANG

Indonesia adalah negara dengan beragam etnis serta kekayaan seni dan budaya yang unik. Kesenian sebagai bagian dari kebudayaan merupakan hasil cipta dan karsa manusia yang diwujudkan dalam berbagai bentuk, seperti seni rupa, seni lukis, tari, grafis, maupun musik. Seni sebagai karya manusia sering kali memiliki perjalanan yang sejalan dengan kehidupan penciptanya. Sementara itu, budaya merupakan hasil dari budi dan daya manusia yang tercermin melalui cipta, rasa, dan karsa, serta memuat kebiasaan yang berkembang dalam masyarakat. Seiring perkembangan zaman, perubahan turut memengaruhi sejarah dan budaya Indonesia, termasuk di dalamnya keberadaan senjata tradisional (Setyawati, 2018).

Received Agustus 10, 2024; Revised September 13, 2024; Accepted Oktober 21, 2024; Published Oktober 29, 2024

Sistem informasi musik modern sangat bergantung pada kemampuan untuk secara otomatis mengidentifikasi komponen inti musik, salah satunya adalah instrumen yang dimainkan. Klasifikasi instrumen musik otomatis (AMIC) menjadi krusial untuk pengindeksan, pencarian berbasis konten, dan analisis musik secara struktural. Secara historis, AMIC dilakukan menggunakan fitur-fitur berbasis domain waktu dan frekuensi seperti *Mel-Frequency Cepstral Coefficients* (MFCC), namun metode ini memerlukan proses rekayasa fitur yang intensif.

Perkembangan Deep Learning telah menggeser paradigma, memungkinkan penggunaan Convolutional Neural Network (CNN) yang mampu mempelajari representasi fitur secara hierarkis langsung dari data. Untuk sinyal audio, ini dilakukan dengan mengubah sinyal deret waktu satu dimensi menjadi representasi visual dua dimensi, seperti Spektrogram atau Mel-Spektrogram, yang kemudian dapat diperlakukan sebagai citra (Yang et al., 2023).

Di antara berbagai arsitektur CNN, *Residual Network (ResNet)* menonjol karena kemampuannya melatih jaringan yang sangat dalam melalui mekanisme shortcut connections (koneksi pintas) yang mengatasi masalah vanishing gradient. Arsitektur ini telah terbukti efektif dalam berbagai tugas klasifikasi citra, termasuk dalam domain audio dan lingkungan yang bising (Yang et al., 2023).

Penelitian ini bertujuan untuk menerapkan dan mengevaluasi kinerja arsitektur *ResNet*-18, varian *ResNet* yang efisien untuk mengklasifikasikan 10 jenis instrumen musik. Strategi *Transfer learning* dengan bobot *Pretrained* dari *ImageNet* diaplikasikan untuk memastikan kinerja yang optimal dan efisiensi komputasi, seperti yang juga diterapkan dalam penelitian klasifikasi lainnya (Nainggolan et al., 2024).

2. KAJIAN TEORITIS

Klasifikasi Instrumen Musik dan Representasi Spektral

Dalam konteks *deep learning* untuk klasifikasi audio, sinyal mentah sering diubah menjadi citra yang menangkap karakteristik unik instrumen. Mel-Spektrogram adalah representasi yang populer karena skala frekuensinya lebih selaras dengan pendengaran manusia. Dengan mengubah sinyal audio menjadi citra spektral, tugas klasifikasi dapat memanfaatkan kekuatan model CNN yang awalnya dirancang untuk tugas visi komputer. Keberhasilan metode ini terlihat dari penerapan *ResNet* dalam klasifikasi suara di lingkungan bising (Yang et al., 2023).

Convolutional Neural Network (CNN) dan *ResNet*

CNN adalah jaringan saraf tiruan yang dirancang untuk memproses data dalam bentuk *grid* seperti citra. *ResNet* (He et al., 2016) memperluas kemampuan CNN dengan memperkenalkan *residual block*. Blok ini memungkinkan gradien untuk mengalir langsung melalui koneksi pintas ke lapisan sebelumnya, sehingga memungkinkan kedalaman jaringan yang signifikan tanpa degradasi kinerja.

ResNet, yang diperkenalkan oleh He et al. (2016), menghadirkan terobosan penting dalam arsitektur deep learning melalui konsep *identity mapping* pada jaringan yang sangat dalam. Inovasi ini diwujudkan melalui penggunaan *residual block*, yaitu struktur yang menambahkan langsung input blok ke output transformasi non-linear di dalam blok tersebut. Secara matematis, mekanisme ini dinyatakan sebagai berikut:

$$y=F(x,\{Wi\})+x$$

di mana x merupakan input blok, F(x,{Wi}) adalah fungsi transformasi yang dipelajari (misalnya operasi konvolusi), dan y adalah output blok residual. Pendekatan ini memungkinkan jaringan mempelajari fungsi identitas dengan lebih mudah, sehingga jaringan yang semakin dalam tidak mengalami penurunan kinerja dibandingkan arsitektur yang lebih dangkal.

Model *ResNet*-18 memiliki 18 lapisan yang dapat dilatih dan sering menjadi pilihan dalam penelitian yang membandingkan arsitektur CNN, termasuk dengan *ResNet*-50 yang lebih dalam (Alberto & Hermanto, 2023). *ResNet* juga telah berhasil diterapkan dalam konteks yang berhubungan dengan musik, seperti analisis struktur musik, di mana *ResNet*-50 mencapai akurasi hingga 87% dan dalam sistem rekomendasi musik berbasis emosi (Saputra & Nudin, 2021)

Transfer learning

Transfer learning merupakan metode yang mengekstraksi bobot dari jaringan yang sudah dilatih sebelumnya (*Pretrained* model) dan mentransfernya ke jaringan target lain yang tidak terlatih (Wani et al., 2019). *Transfer learning* membutuhkan model yang sudah dilatih sebelumnya atau yang dikenal sebagai *Pretrained* model untuk mempelajari tugas yang baru (Wijaya et al., 2021).

Transfer learning memanfaatkan pengetahuan yang diperoleh dari pelatihan model pada tugas yang besar (misalnya *ImageNet*) ke tugas yang baru. Dalam penelitian ini, bobot *Pretrained ResNet*-18 digunakan. Keuntungan utama dari pendekatan ini adalah pengurangan kebutuhan data yang besar dan waktu pelatihan, serta peningkatan generalisasi, karena model

telah mempelajari fitur-fitur visual dasar seperti garis dan tekstur Training dengan menggunakan metode *Transfer learning* dapat dilakukan dengan mudah dan cepat dikarenakan menggunakan dataset yang sedikit (Tsiakmaki et al., 2020).

3. METODE PENELITIAN

Dataset dan Pra-pemrosesan Data

Penelitian ini menggunakan dataset citra instrumen musik yang terdiri dari 10 kelas berbeda, termasuk instrumen perkusi (*drum*, *tabla*), tiup (*flute*, *saxophone*), dan senar (*guitar*, *violin*, *sitar*). Seperti yang terlihat pada gambar dibawah ini

	class	image_count	avg_width	avg_height	min_width	min_height	max_width	max_height	formats	corrupt_files
0	drum	197	145	127	87	60	162	140	jpeg, png	0
1	violin	196	136	125	42	47	162	140	jpeg, png	0
2	sitar	195	135	123	46	43	162	140	jpeg, png	0
3	banjo	194	131	128	51	52	162	140	jpeg, png	0
4	guitar	194	138	124	56	49	162	140	jpeg, png	0

Gambar 1. Citra Instrumen Musik

Prosedur pra-pemrosesan data diterapkan untuk menyesuaikan input dengan kebutuhan *ResNet*-18:

1. Pembagian Data: Dataset dibagi menjadi 80% data pelatihan dan 20% data validasi melalui pembagian acak.
2. Transformasi Standar: Citra diubah ukurannya menjadi 256x256 piksel, diikuti dengan *Center Crop* menjadi ukuran input standar *ResNet*, yaitu 224x224 piksel.
3. Augmentasi Data: Untuk set pelatihan, diterapkan teknik augmentasi berupa *Random Horizontal Flip* dan *Random Rotation* (10 derajat) untuk meningkatkan keragaman dan mencegah *overfitting*.
4. Normalisasi: Citra dinormalisasi menggunakan nilai rata-rata (μ) = [0.485, 0.456, 0.406] dan deviasi standar (σ) = [0.229, 0.224, 0.225] *ImageNet*.

Konfigurasi Model dan Pelatihan

Model yang digunakan adalah *ResNet*-18 dengan bobot *Pretrained*. Lapisan *Fully Connected* (FC) terakhir diganti untuk menghasilkan 10 *output* yang sesuai dengan jumlah kelas instrumen.

Tabel 1. Model Parameter

Parameter	**Nilai**
Arsitektur Model	*ResNet*-18 (*Pretrained*)
Jumlah Kelas (*NUM_CLASSES*)	10
Ukuran Batch (BATCH_SIZE)	32
Jumlah *Epoch* (NUM_*EPOCH*S)	10
Fungsi Loss	*Cross-Entropy Loss*
Optimizer	AdamW
Learning Rate (lr)	0.001

Pelatihan dilakukan selama 10 *Epoch*. *Cross-Entropy Loss* digunakan sebagai fungsi kerugian untuk mengukur kinerja klasifikasi, dan *optimizer* AdamW digunakan untuk meminimalkan *loss* tersebut

4. HASIL DAN PEMBAHASAN

Kinerja Model *ResNet*-18

Menyajikan metrik kinerja utama model *ResNet*-18 selama proses pelatihan.

Tabel 2. Kinerja Pelatihan dan Validasi *ResNet*-18

Epoch	***Train Loss***	***Val Accuracy***
1	1.1911	0.4724
2	0.9544	0.6575
3	0.7498	0.6713
4	0.6582	0.7265
5	0.5658	0.6713
6	0.5141	0.6989
7	0.3623	0.7265
8	0.4234	**0.7735**
9	0.4037	0.7238
10	**0.2818**	0.8094

Hasil menunjukkan bahwa model mengalami peningkatan akurasi validasi yang cepat pada *Epoch* awal, dari 47,24% menjadi 77,35%, menegaskan manfaat signifikan dari *Transfer learning*. *Loss* pelatihan menunjukkan tren penurunan yang stabil, berakhir pada 0,2818 pada *Epoch* terakhir, yang mengindikasikan bahwa model secara efektif mempelajari fitur dari data pelatihan. Akurasi validasi tertinggi tercapai pada 77,35% pada *Epoch* ke-8. Peningkatan yang signifikan pada akurasi validasi menunjukkan kemampuan generalisasi model yang baik.

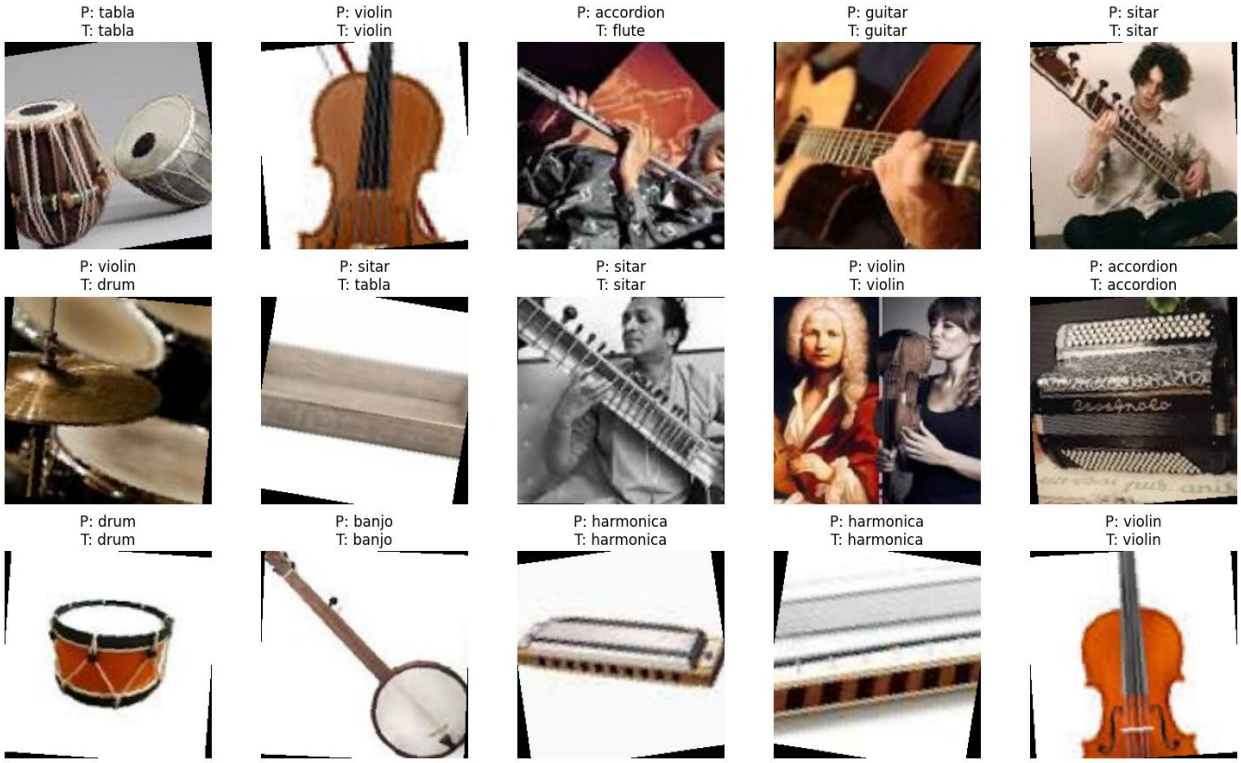
Analisis Kesalahan dan Komparasi

Meskipun akurasi validasi mencapai hampir 78%, terdapat indikasi fluktuasi kinerja (penurunan akurasi pada *Epoch* 9) dan kasus-kasus kesalahan klasifikasi. Analisis visual menunjukkan ambiguitas klasifikasi pada pasangan instrumen, misalnya:

- Model salah memprediksi *drum* sebagai *violin*.
- Model salah memprediksi *tabla* sebagai *sitar*.
- Model salah memprediksi *flute* sebagai *accordion*.

Kesalahan ini kemungkinan disebabkan oleh dua faktor: 1) Kemiripan fitur spektral antara instrumen yang memiliki *timbre* yang serupa, dan 2) Keterbatasan *ResNet*-18 yang lebih ringan dibandingkan dengan *ResNet*-50. Penelitian lain menunjukkan bahwa *ResNet*-50 dapat memberikan akurasi yang lebih tinggi, seperti pada klasifikasi struktur musik yang mencapai 87%, atau pada klasifikasi citra lainnya (Alberto & Hermanto, 2023).

Hasil ini konsisten dengan penelitian implementasi *ResNet*-18 pada domain lain, di mana model menunjukkan kinerja yang solid namun memiliki potensi peningkatan melalui eksplorasi arsitektur yang lebih dalam dan optimasi *hyperparameter*. Seperti yang terlihat pada gambar dibawah ini.



Gambar 2. Klasifikasi Alat Musik

5. KESIMPULAN DAN SARAN

Penelitian ini telah berhasil mengimplementasikan *Transfer learning* menggunakan arsitektur *ResNet*-18 untuk mengklasifikasikan 10 kelas instrumen musik dari representasi visual sinyal audio. Model menunjukkan konvergensi yang cepat dan mencapai akurasi validasi maksimum sebesar 77,35%. Hasil ini memvalidasi efektivitas *ResNet* dan strategi *Transfer learning* dalam mengatasi kompleksitas tugas klasifikasi instrumen musik, meskipun terdapat

tantangan dalam membedakan instrumen dengan fitur spektral yang mirip. Penelitian selanjutnya dapat diarahkan pada beberapa pengembangan penting, antara lain mengganti arsitektur *ResNet*-18 dengan varian yang lebih dalam seperti *ResNet*-50 untuk menguji apakah peningkatan kedalaman jaringan mampu menangkap fitur yang lebih diskriminatif dan meningkatkan akurasi hingga melampaui 80%. Selain itu, optimasi hyperparameter yang lebih komprehensif—termasuk penerapan skema *learning rate scheduling* dapat dilakukan untuk mempertahankan akurasi puncak dan menghindari fluktuasi kinerja setelah *Epoch* ke-8. Analisis kuantitatif terhadap *confusion matrix* juga menjadi langkah penting untuk mengidentifikasi pasangan kelas yang paling sering salah diklasifikasikan, sehingga dapat digunakan sebagai dasar dalam rekayasa fitur atau strategi augmentasi yang lebih spesifik. Terakhir, penggunaan input berupa Mel-Spektrogram yang diekstraksi langsung dari sinyal audio mentah perlu dipertimbangkan agar pendekatan yang digunakan lebih selaras dengan fokus penelitian pada pemrosesan sinyal audio.

DAFTAR REFERENSI

Alberto, J., & Hermanto, D. (2023). Klasifikasi jenis burung menggunakan metode CNN dan arsitektur ResNet-50. Jurnal Teknik Informatika dan Sistem Informasi, 10(3), 34–46.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).

Nainggolan, C. E. S., Nasir, M., Fatoni, & Udariansyah, D. (2024). Perbandingan klasifikasi jenis sampah menggunakan Convolutional Neural Network dengan arsitektur ResNet18 dan ResNet50. CSRID Journal, 16(1), 76–90.

Saputra, F. A., & Nudin, S. R. (2021). Pengembangan sistem rekomendasi pada pemutar musik menggunakan face emotion detection dan ResNet berbasis website. Artikel ilmiah, detail publikasi tidak tersedia.

Setyawati, E. (2018). Aplikasi pengenalan jenis keris tradisional dengan menggunakan augmented reality berbasis Android (pp. 590–595).

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. Applied Sciences, 10(6). https://doi.org/10.3390/app10062145

Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2019). Advances in deep learning (Vol. 57).

Wijaya, A. E., Swastika, W., & Kelana, O. H. (2021). Implementasi transfer learning pada Convolutional Neural Network untuk diagnosis Covid-19 dan pneumonia pada citra X-ray. Sainsbertek: Jurnal Ilmiah Sains dan Teknologi, 2(1), 10–15. https://doi.org/10.33479/sb.v2i1.125

Yang, C., Gan, X., Peng, A., & Yuan, X. (2023). ResNet based on multi-feature attention mechanism for sound classification in noisy environments. Sustainability, 15(14), 10762. https://doi.org/10.3390/su151410762