

Deteksi Aksesoris Tarian Bali dengan Metode YOLO11

Ni Luh Wiwik Sri Rahayu Ginantra^{1*}, Theresia Hendrawati², Sang Made Diva Wikananda³

^{1*, 2, 3} Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia

^{1*} wiwik@instiki.ac.id, ² theresia.hendrawati@instiki.ac.id, ³ sangmadedivawikananda08@gmail.com

INFO ARTIKEL

Article history:
Received Juni 2026
Published Juli 2026

ABSTRAK

Tarian Bali merupakan warisan budaya yang sarat nilai estetika, terutama pada aksesoris yang dikenakan penari, sehingga dokumentasi dan identifikasinya penting bagi pelestarian budaya. Identifikasi manual atas aksesoris yang beragam dan rumit menyita waktu serta rentan kesalahan. Penelitian ini merancang sistem deteksi otomatis aksesoris pada lima jenis tari Bali, yaitu Legong, Puspanjali, Jauk, Baris, dan Topeng, menggunakan algoritma YOLOv11n. Dataset sebanyak 1.896 gambar dikumpulkan melalui scraping Google dan Instagram, dibersihkan, dilabeli dengan 28 kelas aksesoris, lalu diaugmentasi menjadi 4.736 gambar. Model dilatih selama 100 epoch pada ukuran citra 640 piksel. Setelah pelatihan, kinerja dievaluasi memakai *precision*, *recall*, dan *mAP*. Hasil pelatihan model YOLOv11n menunjukkan performa deteksi dengan *precision* sebesar 86,5%, *mAP50* sebesar 82,9%, dan *recall* sebesar 75%.

Kata Kunci: deteksi aksesoris, deteksi objek, pelestarian budaya, tarian Bali, YOLOv11.

ABSTRAK

Balinese dance is a cultural heritage rich in aesthetic value, particularly in the accessories worn by the dancers, making their documentation important for cultural preservation. Manual identification of these diverse and intricate accessories is time-consuming and error-prone. This study designed an automatic detection system for accessories in five Balinese dances, namely Legong, Puspanjali, Jauk, Baris, and Topeng, using the YOLOv11n algorithm. A dataset of 1,896 images was collected through scraping from Google and Instagram, cleaned, labeled into 28 accessory classes, and augmented to 4,736 images. The model was trained for 100 epochs at an image size of 640 pixels. The training results of the YOLOv11n model showed detection performance with a precision of 86.5%, mAP50 of 82.9%, and recall of 75%. Large accessories were detected very accurately, whereas small objects such as senteng and tombak remained difficult to recognize.

Keywords: accessories detection, object detection, cultural preservation, Balinese dance, YOLOv11

©2026 Authors. Licensed Under [CC-BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

1. Pendahuluan

Tarian Bali merupakan salah satu bentuk seni budaya yang sangat dihargai, baik di tingkat nasional maupun internasional. Bali dikenal sebagai provinsi yang masih kental dengan budayanya, sehingga lahir beragam jenis tarian seperti tari Kecak, tari Pendet, dan tari Legong (Yosiana & Wulandari, 2023). Tari Bali memiliki hubungan yang erat dengan kehidupan masyarakatnya dan terus berkembang dengan ide-ide kreatif para senimannya, sambil tetap menjadi bagian integral dari kehidupan komunitas. Keunikan tari

Bali yang memadukan kepercayaan lokal, Hinduisme, dan unsur kebudayaan luar menjadikannya subjek penelitian yang menarik sekaligus pengalaman estetis yang memikat (Dewa dkk., 2024).

Keindahan tarian ini terletak pada harmoni antara gerakan penari, ekspresi wajah, musik pengiring, serta pakaian dan aksesoris yang dikenakan. Aksesoris tersebut bukan sekadar pelengkap estetika, melainkan mengandung makna simbolis yang mencerminkan nilai-nilai tradisional masyarakat Bali. Lima jenis tari yang menjadi fokus penelitian ini, yaitu tari Legong, Puspanjali, Jauk, Baris, dan Topeng, masing-masing menggunakan aksesoris khas yang rumit, seperti mahkota (*gelungan*), hiasan kepala, topeng, hingga ornamen kostum. Oleh karena itu, pelestarian budaya Bali, termasuk dokumentasi aksesoris tarian, menjadi tugas penting di tengah arus modernisasi.

Desain aksesoris yang beragam dan detail memerlukan waktu serta tenaga besar bila diidentifikasi secara manual. Kompleksitas bentuk, ukuran, dan ornamen menjadikan aksesoris tari Bali sebagai objek menarik untuk diselesaikan menggunakan teknologi *computer vision*. Perkembangan kecerdasan buatan, khususnya pada bidang *computer vision*, memungkinkan ekstraksi informasi dari media visual seperti gambar dan video secara otomatis dan efisien (Nuresa Qodri dkk., 2024). Dengan demikian, tantangan identifikasi manual aksesoris mulai dapat diatasi melalui pendekatan deteksi objek berbasis pembelajaran mesin.

You Only Look Once (YOLO) merupakan salah satu metode *state-of-the-art* untuk deteksi objek secara *real-time*. YOLO termasuk dalam keluarga *Convolutional Neural Network* (CNN) yang memprediksi *bounding box* dan probabilitas kelas secara langsung pada satu gambar penuh dalam satu kali evaluasi (Sri Wisna dkk., 2020). Algoritma ini terbukti efektif menghasilkan deteksi berakurasi tinggi (Redmon & Farhadi, 2018) dan telah banyak diterapkan, mulai dari pemantauan jarak fisik (Hasani dkk., 2022) hingga deteksi pelat kendaraan (Nugroho dkk., 2024). Versi terbaru, YOLOv11, memperkenalkan arsitektur yang lebih efisien melalui blok C3K2, *Spatial Pyramid Pooling Fast* (SPPF), dan mekanisme perhatian C2PSA, sehingga mampu meningkatkan deteksi objek kecil tanpa mengorbankan kecepatan inferensi.

Meskipun penelitian deteksi objek berbasis YOLO telah banyak dilakukan pada domain umum, penerapannya untuk aksesoris tari Bali masih sangat terbatas. Sebagian besar studi sebelumnya berfokus pada objek generik seperti kendaraan, makanan, atau manusia (Rahma dkk., 2021), bukan pada artefak budaya dengan ornamen yang padat dan saling tumpang tindih. Kesenjangan inilah yang menjadi celah penelitian. Berdasarkan uraian tersebut, penelitian ini bertujuan merancang dan mengimplementasikan model YOLOv11n untuk mendeteksi aksesoris pakaian pada lima jenis tari Bali dengan tingkat akurasi yang tinggi, sekaligus mengintegrasikannya dengan Segment Anything Model 2 (SAM 2) untuk memperkaya keluaran berupa segmentasi piksel. Hasil penelitian diharapkan mempermudah identifikasi aksesoris dan mendukung dokumentasi digital bagi pelestarian budaya.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental dengan data sekunder berupa citra lima jenis tari Bali, yaitu tari Legong, Puspanjali, Jauk, Baris, dan Topeng. Kegiatan penelitian dimulai sejak September 2024. Secara umum, tahapan penelitian terdiri atas pengumpulan data, pembersihan data, penggabungan dan pelabelan dataset, augmentasi, pembagian data, pelatihan model YOLOv11n, evaluasi. Alur kerja tersebut dirancang agar setiap tahap menghasilkan masukan yang valid bagi tahap berikutnya.

Pengumpulan data dilakukan melalui dua cara, yaitu *scraping* otomatis dan pengunduhan manual. *Scraping* memanfaatkan skrip Python dengan pustaka Selenium untuk mengekstraksi gambar dari Google secara otomatis, sedangkan pengunduhan manual dilakukan dari Instagram dengan menelusuri akun dan tagar yang relevan. Gabungan kedua metode ini bertujuan memastikan keberagaman dataset. Tahap *cleaning* kemudian menghapus gambar yang tidak relevan, beresolusi rendah, atau buram, serta menormalkan format citra ke standar yang konsisten. Dari sekitar 3.000 gambar hasil *scraping*, tersisa 1.896 gambar berkualitas setelah proses pembersihan.

Gambar yang telah bersih digabungkan, lalu dilabeli menggunakan *tools* Roboflow. Pelabelan dilakukan dengan menggambar *bounding box* di sekitar objek dan memberi label kelas aksesoris, misalnya *gelungan legong*, *badong*, *kepet*, *lamak*, hingga *ampok-ampok*, dengan total 28 kelas. Untuk memperkaya variasi data, diterapkan augmentasi berupa rotasi, konversi *grayscale*, pembalikan horizontal dan vertikal, serta penyesuaian kecerahan, sehingga jumlah dataset meningkat menjadi 4.736 gambar. Dataset selanjutnya dibagi secara sistematis dengan proporsi 70% data latih, 20% data validasi, dan 10% data uji.

Pelatihan model menggunakan kerangka kerja YOLOv11n dari pustaka Ultralytics yang dijalankan di atas PyTorch. Bobot pralatih *yolo11n.pt* dimuat, kemudian dilatih ulang (*fine-tuning*) pada dataset kustom

selama 100 *epoch* dengan ukuran citra masukan 640 piksel dan akselerasi GPU. Proses pelatihan memanfaatkan data latih untuk mengoptimalkan parameter dan data validasi untuk memantau indikasi *overfitting* melalui metrik *loss*, *precision*, *recall*, dan *mean Average Precision* (mAP). Model terbaik disimpan dalam format *.pt* agar dapat digunakan kembali untuk inferensi.

Evaluasi kuantitatif dilakukan pada data uji menggunakan empat metrik utama: *precision*, yang mengukur ketepatan prediksi positif; *recall*, yang mengukur kemampuan mendeteksi seluruh objek relevan; serta mAP@0.5 dan mAP@0.5:0.95 yang menilai kualitas deteksi pada berbagai nilai *Intersection over Union* (IoU). *Precision* dan *recall* dihitung dengan persamaan (1) dan (2), dengan TP, FP, dan FN berturut-turut menyatakan *true positive*, *false positive*, dan *false negative*.

$$Precision = TP / (TP + FP) \quad (1)$$

$$Recall = TP / (TP + FN) \quad (2)$$

Sebagai tahap akhir, keluaran *bounding box* dari YOLOv11n digunakan sebagai *prompt* bagi Segment Anything Model 2 (SAM 2). SAM 2 merupakan model segmentasi citra yang dikembangkan Meta dan mampu menghasilkan *mask* segmentasi berdasarkan masukan kotak pembatas atau titik acuan (Ravi dkk., 2024). Dengan memproses area di dalam *bounding box*, SAM 2 memisahkan objek aksesoris dari latar belakang berbasis piksel, sehingga visualisasi menjadi lebih kaya dan mendukung analisis lanjutan seperti pengukuran dimensi dan identifikasi pola ornamen.

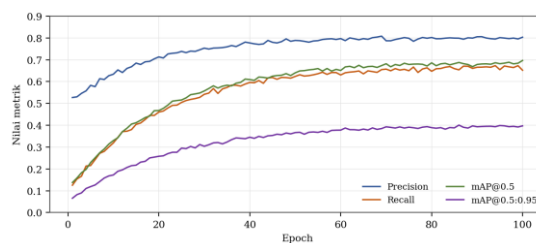
3. Hasil dan Pembahasan

Tahap *preprocessing* menghasilkan dataset akhir sebanyak 4.736 gambar yang siap dilatih. Komposisi data yang seimbang antarkelas tari menjadi penting untuk mencegah bias model terhadap kelas yang dominan. Tabel 1 merangkum transformasi jumlah data dari hasil *scraping* mentah hingga dataset siap latih. Penyusutan tajam dari 3.000 menjadi 1.896 gambar pada tahap *cleaning* menegaskan bahwa data hasil *scraping* mengandung banyak derau, sedangkan lonjakan menjadi 4.736 gambar setelah augmentasi memperlihatkan kontribusi augmentasi dalam memperluas variasi data tanpa pengumpulan citra baru.

Tabel 1. Transformasi jumlah data pada tahap preprocessing

Tahap	Jumlah Gambar
Hasil scraping mentah	3.000
Setelah cleaning data	1.896
Setelah augmentasi	4.736
Data latih (70%)	± 3.315
Data validasi (20%)	± 947
Data uji (10%)	± 474

Proses pelatihan YOLOv11n selama 100 *epoch* menunjukkan konvergensi yang sehat. Sebagaimana terlihat pada Gambar 1, nilai *box loss*, *cls loss*, dan *dfl loss* menurun secara konsisten, sementara *precision*, *recall*, dan mAP meningkat tajam pada *epoch* awal lalu menstabil. Pola ini mengindikasikan bahwa model belajar mengenali fitur aksesoris secara efektif tanpa gejala *overfitting* yang berarti, karena kurva validasi mengikuti tren kurva pelatihan.



Gambar 1. Konvergensi metrik selama pelatihan YOLOv11n

Sumber: Hasil pelatihan, 2025

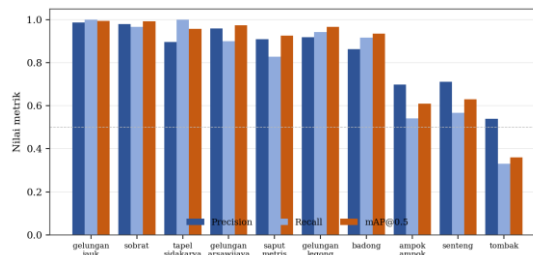
Evaluasi pada data uji memberikan *precision* keseluruhan sebesar 0,865, *recall* sebesar 0,750, mAP@0.5 sebesar 0,829, dan mAP@0.5:0.95 sebesar 0,504. Nilai *precision* yang lebih tinggi daripada *recall* menunjukkan bahwa model cenderung konservatif: prediksi yang dikeluarkan umumnya benar, namun sebagian objek masih terlewat. Selisih yang lebar antara mAP@0.5 dan mAP@0.5:0.95 menandakan

bahwa lokalisasi *bounding box* sudah baik pada ambang IoU longgar, tetapi belum sepenuhnya presisi pada ambang yang ketat.

Analisis per kelas memperlihatkan variasi performa yang signifikan, sebagaimana dirangkum pada Tabel 2 dan divisualisasikan pada Gambar 2. Kelas aksesoris berukuran besar dan berciri visual khas, seperti *gelungan jauk* ($mAP@0.5 = 0,995$), *sobrat* (0,992), dan *gelungan arsanijaya* (0,974), terdeteksi dengan akurasi sangat tinggi. Sebaliknya, objek kecil atau bertekstur mirip latar, seperti *tombak* (0,360), *senteng* (0,630), dan *ampok-ampok* (0,609), menunjukkan performa rendah. Temuan ini konsisten dengan tantangan klasik deteksi objek kecil dan menegaskan pengaruh keterbatasan jumlah serta keberagaman data pada kelas tertentu.

Tabel 2. Performa deteksi YOLOv11n pada kelas terpilih

Kelas	Precision	Recall	mAP@0.5
Semua kelas	0,865	0,750	0,829
gelungan jauk	0,988	1,000	0,995
sobrat	0,980	0,966	0,992
gelungan arsanijaya	0,959	0,900	0,974
gelungan legong	0,918	0,942	0,967
tapel sidakarya	0,897	1,000	0,958
saput metris	0,910	0,828	0,926
ampok ampok	0,698	0,541	0,609
senteng	0,711	0,567	0,630
tombak	0,540	0,330	0,360



Gambar 2. Perbandingan precision, recall, dan mAP@0.5 antarkelas

Sumber: Hasil evaluasi, 2025

Kesalahan klasifikasi yang teramati pada *confusion matrix* sebagian besar terjadi antara kelas berbentuk serupa, misalnya *tapel tupeng tua* dengan *tapel tupeng keras*, serta antara objek kecil dengan latar belakang (*background*). Pola ini menunjukkan bahwa kemiripan visual antaraksesoris, bukan semata kelemahan arsitektur, menjadi sumber kesalahan utama. Penambahan data pada kelas minoritas dan penajaman anotasi berpotensi memperbaiki *recall* tanpa harus mengganti model.

4. Kesimpulan

Penelitian ini berhasil merancang dan mengimplementasikan model YOLOv11n untuk mendeteksi aksesoris pada lima jenis tari Bali dengan performa yang baik, yaitu *precision* 0,865, *recall* 0,750, $mAP@0.5$ 0,829, dan $mAP@0.5:0.95$ 0,504. Kelas aksesoris berukuran besar terdeteksi dengan akurasi tinggi, sementara objek kecil seperti tombak dan senteng masih sulit dikenali akibat keterbatasan data dan kemiripan visual antarkelas. Integrasi dengan Segment Anything Model 2 memperkaya keluaran deteksi melalui *mask* segmentasi, dengan varian *tiny* dan *small* terbukti paling efektif sekaligus efisien. Dengan demikian, tujuan penelitian untuk membangun sistem deteksi aksesoris tari Bali yang akurat telah tercapai. Penelitian lanjutan disarankan menambah data pada kelas minoritas, memperbaiki kualitas anotasi, serta menyertakan *ground truth mask* yang lebih lengkap agar evaluasi segmentasi menjadi lebih objektif dan terukur.

5. Daftar Pustaka

- Dewa, G. S., Sukrawati, N. M., & Untara, I. M. (2024). Nilai pendidikan karakter dalam pewarisan tari Baris Jojol Lutung Jenuk di Desa Selulung Kintamani. *Jurnal Ilmiah Research Student*, 1(5), 1–7. <https://doi.org/10.61722/jirs.v1i5.1170>
- Giannakis, I., Bhardwaj, A., Sam, L., & Leontidis, G. (2024). A flexible deep learning crater detection scheme using Segment Anything Model (SAM). *Icarus*, 408, 115797. <https://doi.org/10.1016/j.icarus.2023.115797>
- Hasani, M. C., Milenasari, F., & Prasetyo, A. (2022). Pemantauan physical distance pada area umum menggunakan YOLO Tiny V3. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), 146–152. <https://doi.org/10.29207/resti.v6i1.3808>
- Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access*, 12, 94250–94295. <https://doi.org/10.1109/ACCESS.2024.3425166>
- Ma, J., Kim, S., Li, F., & Wang, B. (2024). *Segment anything in medical images and videos: Benchmark and deployment*. arXiv. <https://arxiv.org/abs/2408.03322>
- Nugroho, B. P., Prihati, Y., & Saputra, R. (2024). Implementation of YOLO V5 algorithm in vehicle license plate detection application design. *Journal of Information Technology and Computer Science (INTECOMS)*, 7(3), 1–10.
- Nuresa Qodri, K., Al Banna, D., & Pratama, R. (2024). Pemanfaatan SAM dan YOLOv8 untuk deteksi dan segmentasi pada citra MRI tumor otak. *Jurnal Teknologi Informasi*, 5(2), 1–12.
- Rahma, L., Syaputra, H., & Akbar, M. (2021). Objek deteksi makanan khas Palembang menggunakan algoritma YOLO (You Only Look Once). *Jurnal Nasional Ilmu Komputer*, 2(3), 1–10.
- Ravi, N., Gabeur, V., Hu, Y.-T., & Feichtenhofer, C. (2024). *SAM 2: Segment anything in images and videos*. arXiv. <https://arxiv.org/abs/2408.00714>
- Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. arXiv. <https://arxiv.org/abs/1804.02767>
- Salamah, I., Said, M. R. A., & Lindawati. (2022). Perancangan alat identifikasi wajah dengan algoritma You Only Look Once (YOLO) untuk presensi mahasiswa. *Jurnal Media Informatika Budidarma*, 6(3), 1492–1500. <https://doi.org/10.30865/mib.v6i3.4399>
- Sri Wisna, J. H., Matulatan, T., & Hayaty, M. (2020). Implementasi metode YOLO untuk deteksi objek secara real-time. *Jurnal Sustainable: Jurnal Hasil Penelitian dan Industri Terapan*, 9(1), 8–14.
- Yosiana, M., & Wulandari, R. (2023). Komodifikasi tubuh perempuan pada tarian Joged Bumbung Bali di YouTube. *Jurnal Komunikasi dan Kajian Media*, 2(4), 1–12. <https://doi.org/10.56127/juki>