

Deteksi Lesi Serviks Berdasarkan Citra *Colposcopy* Menggunakan YOLO

SITI SIROTUL AZHAR¹, ELLY MATUL IMAH^{2,*}

¹ Program Studi S1 Sains Data, Universitas Negeri Surabaya, Indonesia

² Program Studi S1 Kecerdasan Artifisial, Universitas Negeri Surabaya, Indonesia

Abstrak

Kanker serviks merupakan penyakit dengan tingkat mortalitas tinggi yang memerlukan deteksi dini berbasis citra medis, khususnya citra *colposcopy*. Penelitian ini menerapkan YOLO berbasis *instance segmentation*, yaitu arsitektur yang mampu mendeteksi objek sekaligus menghasilkan *mask* segmentasi secara *end-to-end*, untuk melakukan segmentasi area serviks dan lesi secara terintegrasi. Dataset yang digunakan berupa citra *colposcopy* beranotasi dua kelas, yaitu serviks dan lesi, dengan anotasi segmentasi yang disusun berdasarkan metadata klinis sebagai *ground truth*. Tiga arsitektur YOLO, yaitu YOLOv8, YOLO11, dan YOLO26, dievaluasi secara komparatif menggunakan beberapa variasi ukuran model. Evaluasi dilakukan menggunakan metrik *Intersection over Union* (IoU) untuk mengukur tingkat kesesuaian hasil segmentasi model terhadap *ground truth*. Hasil eksperimen menunjukkan bahwa dari seluruh varian yang dievaluasi, YOLOv8m, YOLO11l, dan YOLO26s merupakan model dengan performa terbaik pada masing-masing generasi. Di antara ketiganya, YOLO26s memperoleh performa tertinggi dengan nilai IoU sebesar 0,97 pada kelas serviks dan 0,90 pada kelas lesi. Nilai tersebut menunjukkan bahwa YOLO26s mampu menghasilkan segmentasi yang sangat akurat pada area serviks serta mempertahankan performa yang baik pada area lesi yang memiliki bentuk dan batas lebih kompleks. Temuan ini memberikan evaluasi komparatif terhadap tiga generasi arsitektur YOLO untuk tugas *instance segmentation* pada citra *colposcopy* sekaligus menunjukkan potensi YOLO26s sebagai model yang paling efektif untuk segmentasi serviks dan lesi pada dataset yang digunakan. Pengembangan lebih lanjut diperlukan melalui perluasan variasi dataset dan evaluasi pada kondisi data yang lebih beragam untuk meningkatkan *robustness* model.

Kata Kunci kanker serviks; *colposcopy*; deep learning; YOLO; *instance segmentation*

Abstract

Cervical cancer is a disease with a high mortality rate that requires early detection using medical imaging, particularly colposcopy images. This study employs YOLO-based instance segmentation, an architecture capable of simultaneously detecting objects and generating segmentation masks in an end-to-end manner, to perform integrated segmentation of the cervix and cervical lesions. The dataset consists of colposcopy images annotated with two classes, namely the cervix and lesions, where the segmentation annotations were constructed based on clinical metadata as the ground truth. Three YOLO architectures, namely YOLOv8, YOLO11, and YOLO26, were comparatively evaluated using multiple model size variants. Model performance was assessed using the Intersection over Union (IoU) metric to measure the agreement between the predicted

*Corresponding author. Email: ellymatul@unesa.ac.id.

segmentation masks and the ground truth. Experimental results show that, among all evaluated variants, YOLOv8m, YOLO11l, and YOLO26s achieved the best performance within their respective generations. Among these models, YOLO26s obtained the highest performance, achieving IoU scores of 0.97 for the cervix class and 0.90 for the lesion class. These results indicate that YOLO26s is capable of producing highly accurate segmentation of the cervix while maintaining strong performance in segmenting lesions, which typically exhibit more complex shapes and boundaries. The findings provide a comparative evaluation of three generations of YOLO architectures for instance segmentation on colposcopy images and demonstrate the potential of YOLO26s as the most effective model for cervix and lesion segmentation on the dataset used in this study. Further research is needed by expanding dataset diversity and evaluating the model under more varied imaging conditions to improve its robustness.

Keywords *cervical cancer; colposcopy; deep learning; YOLO; instance segmentation*

1 Pendahuluan

Kanker serviks merupakan salah satu penyebab utama kematian pada wanita di dunia dengan tingkat mortalitas yang masih tinggi menurut World Health Organization (WHO, 2024). Di Indonesia, prevalensi kanker serviks dilaporkan sekitar 0.8% atau 98,692 kasus, dengan kelompok usia 55–64 tahun mencatat kasus tertinggi sebesar 4.6% (Asrina dkk., 2025). *Infeksi Human Papillomavirus* (HPV) menjadi faktor utama yang memicu perubahan abnormal pada jaringan serviks dan dapat berkembang menjadi lesi pra-kanker maupun kanker serviks (Wolf dkk., 2024). Deteksi dini menjadi faktor penting karena kanker serviks pada tahap awal sering tidak menunjukkan gejala yang spesifik. *Colposcopy* digunakan sebagai prosedur standar untuk mengidentifikasi area lesi melalui pengamatan visual. Lesi serviks ditandai oleh perubahan warna, tekstur, pola vaskularisasi, dan batas jaringan yang dapat mengindikasikan terjadinya displasia atau *Cervical Intraepithelial Neoplasia* (CIN). Variasi karakteristik visual tersebut menyebabkan proses interpretasi sangat bergantung pada pengalaman dokter sehingga berpotensi menimbulkan subjektivitas antar pengamat (Himabindu dkk., 2025). Kondisi tersebut mendorong pemanfaatan *Artificial Intelligence* (AI), khususnya *deep learning*, untuk meningkatkan konsistensi dan akurasi analisis citra medis.

Pendekatan berbasis deteksi seperti *Region-Based Convolutional Neural Network* (R-CNN) mampu mengidentifikasi lokasi lesi, namun masih menghadapi keterbatasan dalam menangkap batas area secara presisi (Kiptoo dkk., 2021). Pada kasus lesi serviks, informasi batas lesi menjadi penting karena berkaitan dengan ukuran dan distribusi area abnormal yang diamati selama pemeriksaan. Metode segmentasi seperti U-Net menunjukkan kemampuan spasial yang lebih baik, tetapi masih mengalami kesulitan dalam membedakan jaringan normal dan abnormal pada citra dengan karakteristik visual kompleks (Shinohara dkk., 2023). Pendekatan multi-tahap seperti *Improved Faster R-CNN* dan *Cervical Lesion Segmentation Network* (CLS-Net) berhasil meningkatkan performa deteksi dan segmentasi, namun kompleksitas arsitektur yang tinggi menyebabkan efisiensi komputasi menjadi terbatas, terutama pada lesi berukuran kecil dan dengan batas tidak tegas (Yu dkk., 2022). Sebagian besar penelitian sebelumnya masih memisahkan proses deteksi dan segmentasi atau menggunakan pendekatan multi-tahap yang kompleks sehingga meningkatkan kebutuhan komputasi dan waktu pemrosesan.

Perkembangan model YOLO memungkinkan implementasi *instance segmentation* yang mampu melakukan deteksi objek dan segmentasi secara simultan dalam satu arsitektur *end-to-end*. Pendekatan ini berpotensi menghasilkan representasi spasial yang lebih akurat dengan

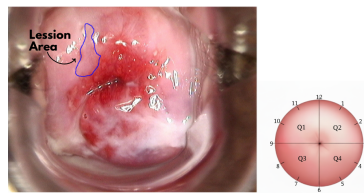
efisiensi komputasi yang lebih baik dibandingkan metode multi-tahap. Analisis terhadap berbagai varian YOLO diperlukan untuk memahami pengaruh perbedaan desain arsitektur terhadap kualitas segmentasi lesi serviks pada citra *colposcopy*. Penelitian ini mengevaluasi beberapa varian YOLO berbasis *instance segmentation* untuk mendeteksi dan melakukan segmentasi area serviks serta lesi pada citra *colposcopy*. Evaluasi dilakukan menggunakan beberapa metrik, dengan fokus pada *Intersection over Union* (IoU) sebagai metrik utama untuk menilai ketepatan segmentasi. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan sistem diagnosis berbasis AI yang lebih akurat, konsisten, dan efisien pada analisis citra serviks.

2 Metode

2.1 Dataset

Data yang digunakan dalam penelitian ini diperoleh dari International Agency for Research on Cancer (IARC) di bawah naungan World Health Organization (WHO) dan bersifat terbatas (*confidential*). Dataset awal terdiri atas 25 data metadata pemeriksaan *colposcopy*. Citra *colposcopy* merupakan citra digital yang diperoleh melalui pemeriksaan serviks menggunakan *colposcope*, yaitu alat pembesar dengan pencahayaan kuat untuk menampilkan jaringan serviks secara lebih jelas (ACOG, 2024; Cooper dan Dunton, 2025). Citra ini umumnya digunakan untuk menilai area serviks yang dicurigai mengalami kelainan setelah pemeriksaan skrining awal.

Dalam pemeriksaan *colposcopy*, struktur serviks normal mencakup *Squamocolumnar Junction* (SCJ), *Columnar Area* (CA), dan *Transformation Zone* (TZ), yaitu area yang sering menjadi lokasi munculnya perubahan abnormal. Lesi serviks merujuk pada perubahan abnormal pada epitel serviks yang dapat berkaitan dengan kondisi prakanker seperti *Cervical Intraepithelial Neoplasia* (CIN). Secara visual, lesi dapat dikenali melalui perubahan warna, tekstur, batas tepi, ketebalan permukaan, serta pola vaskularisasi jaringan (Alec, 2024). Karena interpretasi visual dapat bervariasi antar pemeriksa, dataset *colposcopy* yang terstruktur dan teranotasi diperlukan untuk mendukung deteksi dan segmentasi lesi secara lebih objektif. Contoh area lesi ditunjukkan pada Gambar 1.

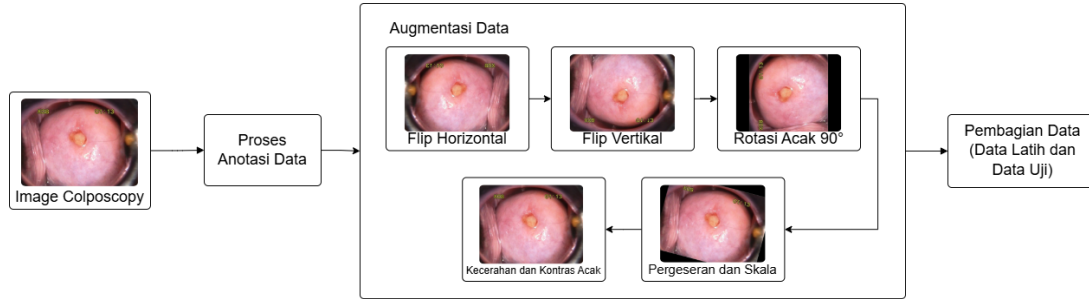


Gambar 1: Contoh area lesi serviks pada citra *colposcopy*

2.2 Pemrosesan Data

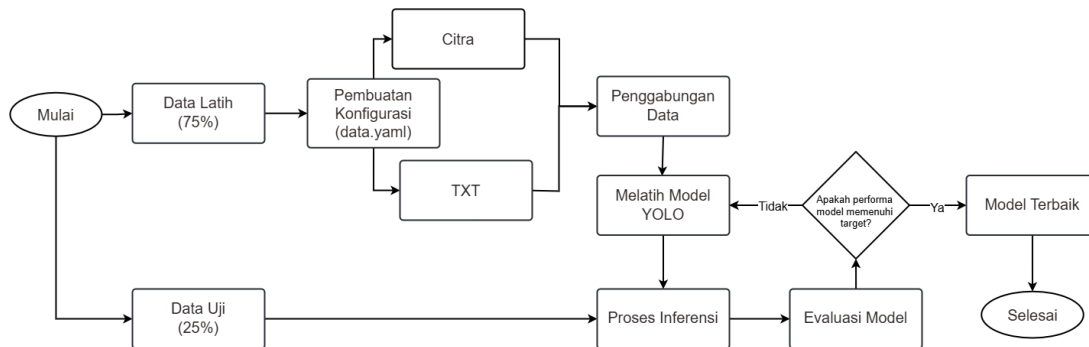
Pipeline pemrosesan data mencakup anotasi, augmentasi, serta pembagian dataset. Rangkaian proses ini dirancang untuk memastikan kualitas data yang optimal serta mendukung proses pembelajaran model secara efektif. Alur keseluruhan ditunjukkan pada Gambar 2. Anotasi data dilakukan untuk memberikan label pada area penting dalam citra sebagai referensi pembelajaran model, sehingga mendukung proses deteksi dan segmentasi secara lebih akurat (Saranyadkk., 2023). Pada penelitian ini, anotasi dilakukan menggunakan Roboflow, yaitu platform yang

banyak digunakan untuk pengelolaan dataset dan anotasi citra, termasuk pada bidang medis (Roboflow, 2025). Proses anotasi mencakup pembuatan *mask* segmentasi pada area serviks dan lesi berdasarkan metadata klinis yang tersedia. Metadata digunakan sebagai panduan untuk menjaga konsistensi pelabelan, terutama ketika batas lesi tidak terlihat secara jelas. Dari keselu-



Gambar 2: Diagram Pemrosesan Data

ruhan data yang tersedia, hanya delapan citra *colposcopy* yang dilengkapi informasi lokasi area lesi. Informasi tersebut kemudian digunakan sebagai dasar dalam pembuatan *mask* segmentasi. Untuk meningkatkan jumlah dan variasi data, augmentasi manual diterapkan menggunakan pustaka Albumentations (Mumuni dan Mumuni, 2022). Setiap citra menghasilkan enam belas variasi melalui transformasi rotasi, *flipping*, dan translasi posisi, sehingga total dataset menjadi 136 citra, terdiri atas delapan citra asli dan 128 citra hasil augmentasi. Transformasi tersebut bertujuan untuk meningkatkan variasi orientasi dan posisi objek agar model lebih *robust* terhadap perubahan visual pada citra *colposcopy*. Dataset kemudian dibagi menjadi data pelatihan dan pengujian menggunakan metode *random splitting* dengan rasio 75:25 (Oktafiani dkk., 2023). Pemisahan dilakukan pada pasangan citra dan anotasi untuk menjaga konsistensi antara data masukan dan label segmentasi. Selain itu, penggunaan *random seed* tetap diterapkan untuk memastikan reproduktibilitas eksperimen. Pendekatan ini digunakan agar evaluasi model dilakukan secara objektif terhadap data yang tidak digunakan selama proses pelatihan.



Gambar 3: Diagram Deteksi Lesi Serviks

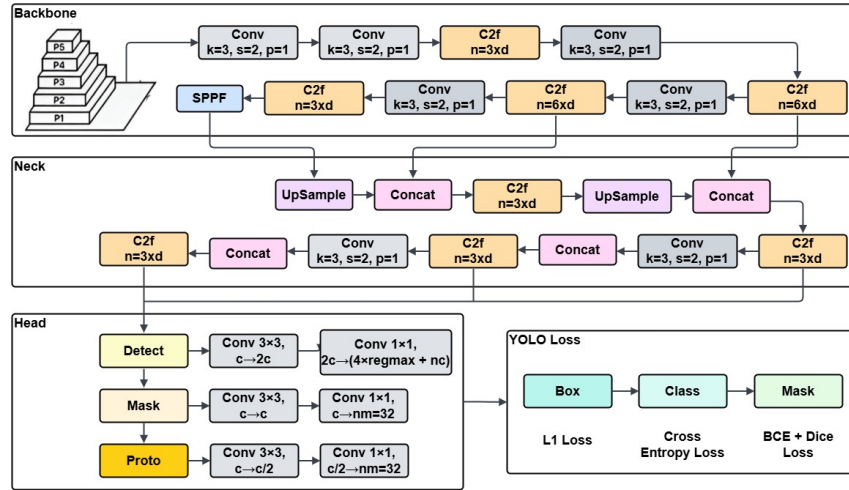
2.3 Deteksi Lesi Serviks Menggunakan YOLO

You Only Look Once (YOLO) merupakan algoritma *single-stage object detection* yang mampu melakukan deteksi objek dalam satu kali proses inferensi. Pada perkembangannya, YOLO ti-

dak hanya digunakan untuk deteksi objek, tetapi juga mendukung tugas *instance segmentation*, yaitu identifikasi lokasi objek sekaligus segmentasi batas objek pada tingkat piksel (Baldovino dkk., 2024). Dalam penelitian ini, YOLO digunakan untuk mendeteksi dan melakukan segmentasi lesi serviks pada citra *colposcopy*. Model yang dievaluasi meliputi YOLOv8, YOLO11, dan YOLO26. Alur keseluruhan proses deteksi mulai dari *input* citra hingga evaluasi model ditunjukkan pada Gambar 3.

2.3.1 YOLOv8

YOLOv8 merupakan model deteksi objek yang diperkenalkan oleh Ultralytics pada tahun 2023. Model ini menggunakan pendekatan *anchor-free detection* untuk meningkatkan efisiensi inferensi dan menyederhanakan proses prediksi objek (Jocher dkk., 2023). Arsitekturnya terdiri atas *backbone*, *neck*, dan *head*, sebagaimana ditunjukkan pada Gambar 4.



Gambar 4: Arsitektur YOLOv8

- Backbone** Backbone berfungsi mengekstraksi fitur dari citra masukan menggunakan modul C2f. Proses ekstraksi fitur dilakukan melalui operasi konvolusi yang dapat dinyatakan pada persamaan (1).

$$s(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \quad (1)$$

Dengan I menyatakan citra masukan, K kernel konvolusi, dan $s(i, j)$ nilai pada *feature map* hasil ekstraksi fitur. Melalui proses ini, model dapat mengenali pola spasial seperti tepi, tekstur, dan bentuk lesi yang menjadi dasar proses deteksi dan segmentasi. Pada tahap akhir *backbone*, modul *Spatial Pyramid Pooling Fast* (SPPF) digunakan untuk menggabungkan informasi fitur dari berbagai skala sehingga meningkatkan kemampuan deteksi objek dengan ukuran yang bervariasi.

- Neck** YOLOv8 menggunakan kombinasi *Feature Pyramid Network* (FPN) dan *Path Aggregation Network* (PAN) untuk mengintegrasikan fitur multi-skala. FPN menyebarkan informasi semantik dari lapisan dalam ke lapisan yang lebih dangkal, sedangkan PAN memperkuat informasi spasial dari lapisan dangkal ke lapisan yang lebih dalam. Mekanisme ini membantu model mendeteksi objek kecil maupun besar secara lebih efektif.

- c. **Head** YOLOv8 menerapkan *decoupled head* yang memisahkan proses klasifikasi dan regresi *bounding box*. Pada tahap deteksi, model menghasilkan prediksi objek yang direpresentasikan pada Persamaan (2)

$$B = (x, y, w, h, c) \quad (2)$$

dengan (x) dan (y) menyatakan koordinat pusat *bounding box*, (w) dan (h) menyatakan lebar dan tinggi *bounding box*, sedangkan (c) merupakan *confidence score* yang menunjukkan tingkat keyakinan model terhadap keberadaan objek pada lokasi yang diprediksi.

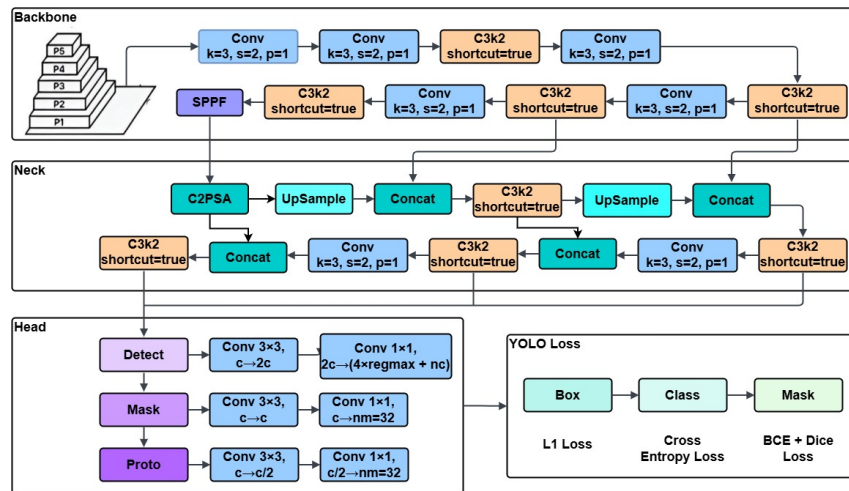
Pada varian segmentasi, *head* dilengkapi cabang segmentasi (*mask branch*) untuk menghasilkan *mask* objek pada tingkat piksel. Prediksi segmentasi dapat dinyatakan pada Persamaan (3)

$$M = f(I) \quad (3)$$

dengan (I) menyatakan citra masukan, (f) merepresentasikan fungsi pemetaan yang dipelajari oleh model YOLO selama proses pelatihan, dan (M) merupakan *mask* segmentasi yang dihasilkan model. Setiap piksel pada *mask* menunjukkan probabilitas keberadaan objek sehingga batas lesi dapat direpresentasikan secara lebih presisi dibandingkan pendekatan berbasis *bounding box*. Selama pelatihan, model mengoptimalkan prediksi segmentasi menggunakan kombinasi *Binary Cross Entropy* (BCE) dan *Dice Loss*. BCE digunakan untuk mengevaluasi kesalahan klasifikasi piksel, sedangkan Dice Loss mengukur kesesuaian area prediksi dengan *ground truth* sehingga membantu menghasilkan segmentasi yang lebih akurat (Hakim, 2019).

2.3.2 YOLO11

YOLO11 merupakan pengembangan dari YOLOv8 yang mempertahankan struktur dasar *backbone*, *neck*, dan *head*, tetapi menghadirkan beberapa peningkatan pada representasi fitur dan mekanisme perhatian (*attention*) (Jocher dan Qiu, 2024). Arsitektur YOLO11 ditunjukkan pada Gambar 5.



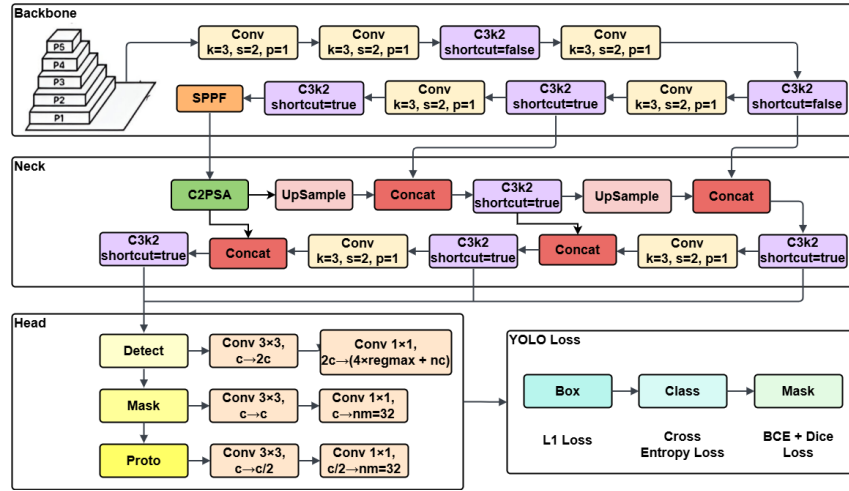
Gambar 5: Arsitektur YOLO11

- a. **Backbone** Perbedaan utama YOLO11 terletak pada penggunaan blok C3k2 yang menggantikan C2f pada YOLOv8. Blok ini dirancang untuk meningkatkan efisiensi ekstraksi fitur dengan kompleksitas komputasi yang tetap terjaga. Selain itu, YOLO11 menambahkan modul *Cross Stage Partial with Partial Spatial Attention* (C2PSA) setelah SPPF untuk memperkuat representasi fitur melalui mekanisme perhatian spasial. Modul ini memungkinkan model memberikan fokus yang lebih besar pada area yang mengandung informasi penting sehingga berpotensi meningkatkan kemampuan deteksi objek dengan karakteristik visual yang kompleks.
- b. **Neck** YOLO11 tetap menggunakan mekanisme fusi fitur multi-skala seperti pada YOLOv8, namun proses agregasi fitur didukung oleh blok C3k2 dan modul C2PSA. Integrasi tersebut membantu meningkatkan kualitas representasi fitur pada berbagai tingkat resolusi sehingga informasi spasial dapat dipertahankan secara lebih efektif.
- c. **Head** Bagian *head* masih menggunakan pendekatan *decoupled head* untuk memisahkan proses klasifikasi dan regresi objek. Peningkatan pada tahap ini difokuskan pada penyempurnaan representasi fitur yang diteruskan dari *backbone* dan *neck*, sehingga model dapat menghasilkan prediksi lokasi dan segmentasi yang lebih akurat. Kombinasi blok C3k2 dan mekanisme perhatian memungkinkan YOLO11 menangkap detail objek secara lebih baik dibandingkan generasi sebelumnya (Khanam dan Hussain, 2024).

2.3.3 YOLO26

YOLO26 merupakan generasi terbaru dalam keluarga YOLO yang dirancang dengan fokus pada efisiensi komputasi dan optimalisasi proses inferensi (Jocher dan Qiu, 2026). Arsitektur YOLO26 ditunjukkan pada Gambar 6. Meskipun tetap mempertahankan struktur dasar YOLO, model ini menghadirkan beberapa penyempurnaan untuk mengurangi kompleksitas komputasi tanpa mengorbankan performa deteksi dan segmentasi.

- a. **Backbone** YOLO26 menggunakan konfigurasi *backbone* yang lebih ringan dibandingkan YOLO11 dengan jumlah parameter dan operasi komputasi yang lebih rendah. Selain itu, penambahan *shortcut connection* membantu memperlancar aliran gradien selama pelatihan sehingga proses pembelajaran menjadi lebih stabil dan efisien.
- b. **Neck** Pada bagian *neck*, YOLO26 mempertahankan mekanisme fusi fitur multi-skala dengan desain yang lebih ringkas. Integrasi modul perhatian tetap digunakan untuk mempertahankan kualitas representasi fitur, namun dengan konfigurasi yang dioptimalkan agar lebih efisien dalam memodelkan hubungan spasial dan semantik antar fitur.
- c. **Head** Pembaruan utama YOLO26 terletak pada mekanisme prediksi objek. Model ini menerapkan strategi *dual assignment* yang mengombinasikan pendekatan *one-to-many* dan *one-to-one assignment* selama proses pelatihan. Pendekatan tersebut memungkinkan model memperoleh representasi objek yang lebih baik sekaligus menghasilkan prediksi yang lebih konsisten. Selain itu, YOLO26 menggantikan proses *Non-Maximum Suppression* (NMS) dengan mekanisme seleksi berbasis Top-K, yaitu memilih sejumlah prediksi dengan skor kepercayaan tertinggi secara langsung. Strategi ini membantu mengurangi kompleksitas inferensi dan meningkatkan efisiensi komputasi tanpa mengurangi kualitas deteksi (Hidayatullah dan Tubagus, 2026).



Gambar 6: Arsitektur YOLO26

2.3.4 Konfigurasi dan Pelatihan Model YOLO

Dataset yang telah melalui proses anotasi selanjutnya diekspor menggunakan format *YOLO Segmentation*. Pada format ini, setiap citra (.jpg) menghasilkan satu berkas anotasi (.txt) dengan nama file yang sama. Berkas anotasi tersebut berisi informasi kelas objek beserta koordinat titik-titik *polygon* yang merepresentasikan batas objek. Pasangan citra dan berkas anotasi kemudian disusun ke dalam struktur direktori dataset yang terdiri atas folder *train* dan *test*. Masing-masing folder memiliki subfolder *images* untuk menyimpan citra (.jpg) dan subfolder *labels* untuk menyimpan berkas anotasi (.txt). Penyusunan tersebut bertujuan agar setiap citra dapat dipasangkan secara otomatis dengan berkas anotasinya berdasarkan nama file yang sama selama proses pelatihan menggunakan *framework* Ultralytics. Dataset yang telah tersusun dikonfigurasi menggunakan berkas *data.yaml*. Berkas ini berfungsi sebagai konfigurasi utama yang berisi lokasi dataset serta daftar kelas yang digunakan selama proses pelatihan. Direktori *train/images* digunakan sebagai data pelatihan, sedangkan direktori *test/images* digunakan sebagai data pengujian. Konfigurasi tersebut memungkinkan *framework* Ultralytics secara otomatis mencari berkas anotasi yang bersesuaian pada folder *labels*. Contoh konfigurasi *data.yaml* ditunjukkan pada Kode Program 1.

Kode Program 1: Contoh konfigurasi *data.yaml*

```
path: dataset
train: train/images
test: test/images

names:
0: Lesi
1: Serviks
```

Setiap citra memiliki satu berkas anotasi dengan nama yang sama dan berekstensi (.txt). Berkas tersebut dibuat menggunakan format *YOLO Segmentation*, di mana setiap baris merepresentasikan satu objek yang terdiri atas indeks kelas diikuti pasangan koordinat titik-titik *polygon*

yang telah dinormalisasi terhadap lebar dan tinggi citra. Representasi berbasis *polygon* memungkinkan batas objek digambarkan secara lebih presisi dibandingkan *bounding box* (Perreault dkk., 2021). Contoh format berkas anotasi ditunjukkan pada Kode Program 2.

Kode Program 2: Contoh berkas anotasi YOLO *Segmentation*

```
0 0.24250 0.40333 0.23500 0.42333 0.23500 0.45500
0.24875 0.50500 0.25875 0.51500 0.28250 0.52333
0.38750 0.28500 0.35375 0.28000 0.31375 0.31500

1 0.16375 0.36167 0.18125 0.55667 0.19375 0.61333
0.20000 0.68333 0.21125 0.72167 0.21125 0.74500
0.27250 0.19667 0.23625 0.23667 0.18750 0.31000
```

Pada Kode Program 2, angka pertama pada setiap baris menunjukkan indeks kelas objek, yaitu kelas Lesi (0) dan Serviks (1). Pasangan angka yang mengikuti merupakan koordinat titik-titik *polygon* (x, y) yang telah dinormalisasi ke rentang 0–1. Koordinat tersebut digunakan untuk membentuk *mask* objek sehingga model dapat mempelajari bentuk dan batas objek secara lebih akurat. Sebelum proses pelatihan, seluruh citra diubah ukurannya menjadi 640×640 piksel. Ketiga model YOLO dikonfigurasi menggunakan parameter pelatihan yang sama untuk memastikan perbandingan performa dilakukan secara adil. Pelatihan dilakukan menggunakan *framework* Ultralytics dengan dukungan GPU NVIDIA Tesla T4. Seluruh model diinisialisasi menggunakan *pretrained weights* dari dataset COCO melalui pendekatan *transfer learning* (Sapkota dan Karkee, 2026). Setiap model dilatih selama 250 *epoch* menggunakan *batch size* sebesar 8, *optimizer* AdamW, *learning rate* sebesar 0,001667, dan *momentum* sebesar 0,9. Mekanisme *early stopping* dengan nilai *patience* sebesar 100 diterapkan untuk mencegah terjadinya *overfitting*. Augmentasi manual dan augmentasi otomatis juga diterapkan untuk meningkatkan kemampuan generalisasi model terhadap variasi data sehingga model menjadi lebih *robust* pada kondisi citra yang berbeda.

2.3.5 Komparasi Arsitektur YOLO

Perbedaan utama antara YOLOv8, YOLO11, dan YOLO26 dirangkum pada Tabel 1. YOLOv8 digunakan sebagai model dasar dengan *backbone* C2f dan mekanisme *anchor-free detection*. YOLO11 memperkenalkan blok C3k2 dan modul perhatian C2PSA untuk meningkatkan representasi fitur. YOLO26 mempertahankan mekanisme perhatian tersebut dengan arsitektur yang lebih ringan serta menerapkan strategi seleksi prediksi berbasis Top-K untuk meningkatkan efisiensi inferensi. Perbedaan karakteristik tersebut menjadi dasar evaluasi komparatif pada tugas deteksi dan segmentasi lesi serviks. Arsitektur YOLO memiliki beberapa varian ukuran model, yaitu *nano* (n), *small* (s), *medium* (m), *large* (l), dan *extra large* (x). Perbedaan setiap varian terletak pada jumlah parameter, kapasitas representasi, dan kebutuhan komputasi, sedangkan arsitektur dasarnya tetap sama. Pada penelitian ini, seluruh varian ukuran model dari masing-masing generasi YOLO dievaluasi untuk memperoleh konfigurasi model yang paling sesuai pada tugas deteksi dan segmentasi lesi serviks. Untuk mendukung analisis komparatif antar generasi, Tabel 2 menyajikan kompleksitas model yang menjadi fokus pembahasan, yaitu YOLOv8m-seg, YOLO11l-seg, dan YOLO26s-seg. Kompleksitas model ditinjau berdasarkan jumlah parameter dan GFLOPs sebagai indikator kapasitas representasi serta kebutuhan komputasi. Informasi tersebut digunakan sebagai dasar dalam menganalisis hubungan antara kompleksitas model dan performa yang diperoleh pada tahap evaluasi.

Tabel 1: Perbandingan Arsitektur YOLO

Komponen	YOLOv8	YOLO11	YOLO26
Backbone	C2f	C3k2	C3k2
Modul Pasca-Backbone	SPPF	SPPF	SPPF
Mekanisme Attention	Tidak digu- nakan	C2PSA	C2PSA + shor- tcut
Neck	PAN-FPN	PAN-FPN de- ngan optimasi	PAN-FPN de- ngan optimasi
Head	Decoupled he- ad	Improved decou- pled head	Decoupled head
Seleksi Prediksi	NMS	NMS	Top-K

Tabel 2: Kompleksitas Model YOLO

Model	Parameter (M)	GFLOPs
YOLOv8m-seg	27.2	104.7
YOLO11l-seg	27.6	132.6
YOLO26s-seg	11.4	37.1

2.3.6 Metrik Evaluasi

Selama proses pelatihan, performa model dipantau menggunakan nilai *loss* yang dihitung pada setiap *epoch*. Pada model YOLO *Segmentation*, komponen *loss* terdiri atas *box loss*, *classification loss* (*cls_loss*), dan *segmentation loss* (*seg_loss*). *Box loss* digunakan untuk mengukur kesalahan model dalam memprediksi lokasi objek, *classification loss* mengukur kesalahan dalam mengklasifikasikan objek, sedangkan *segmentation loss* mengukur kesesuaian antara *mask* hasil prediksi dengan *ground truth*. Penurunan nilai ketiga komponen *loss* selama proses pelatihan menunjukkan bahwa model semakin mampu mempelajari karakteristik objek dan menghasilkan prediksi yang lebih akurat. Evaluasi performa model dilakukan menggunakan metrik *Intersection over Union* (IoU), *Precision*, *Recall*, *F1-Score*, dan *mean Average Precision* (mAP). IoU digunakan sebagai metrik utama karena mengukur tingkat kesesuaian antara hasil segmentasi yang dihasilkan model dengan *ground truth* (Li dkk., 2020). Nilai IoU dihitung dengan membandingkan luas area tumpang tindih antara hasil prediksi dan *ground truth* terhadap luas gabungan keduanya, sebagaimana dirumuskan pada Persamaan (4). *Precision* dan *Recall* digunakan untuk mengevaluasi ketepatan serta kelengkapan deteksi. *F1-Score* digunakan untuk mengukur keseimbangan antara keduanya. mAP digunakan untuk memberikan gambaran performa model secara keseluruhan pada berbagai ambang IoU. Pendekatan evaluasi ini memungkinkan analisis komprehensif terhadap performa model, khususnya dalam mendeteksi dan melakukan segmentasi lesi serviks dengan karakteristik visual yang kompleks.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{A \cap B}{A \cup B} \quad (4)$$

3 Hasil dan Pembahasan

3.1 Hasil Eksperimen dan Evaluasi Kinerja Model

Dataset penelitian terdiri dari 136 citra *colposcopy* yang diperoleh dari 8 citra asli dan 128 citra hasil augmentasi. Data dibagi menggunakan rasio 75:25 menjadi 102 data latih dan 34 data uji. Eksperimen dilakukan pada beberapa variasi ukuran model dari arsitektur YOLOv8, YOLO11, dan YOLO26 untuk menganalisis pengaruh kompleksitas model terhadap performa segmentasi lesi serviks. Evaluasi dilakukan menggunakan metrik IoU, *precision*, *recall*, *F1-score*, AP. Hasil pengujian menunjukkan bahwa YOLOv8m, YOLO11l, dan YOLO26s merupakan varian dengan performa terbaik pada masing-masing generasi, sehingga ketiga model tersebut menjadi fokus analisis komparatif pada pembahasan selanjutnya. Berdasarkan Tabel 3, seluruh model menghasilkan performa tinggi pada objek serviks. Perbedaan performa lebih terlihat pada objek lesi yang memiliki bentuk dan batas area lebih kompleks dibandingkan serviks. Di antara ketiga model tersebut, YOLO26s menghasilkan performa terbaik secara keseluruhan dengan IoU sebesar 0.90 pada objek lesi dan 0.97 pada objek serviks, serta mAP50 sebesar 1.00 dan mAP50-95 sebesar 0.94. Nilai IoU pada objek lesi setara dengan YOLO11l, sedangkan IoU pada objek serviks menjadi yang tertinggi dibandingkan model lainnya. Sementara itu, YOLOv8m menunjukkan performa yang relatif stabil pada seluruh metrik evaluasi, meskipun nilai IoU masih berada di bawah YOLO26s.

Tabel 3: Hasil Evaluasi Model YOLO

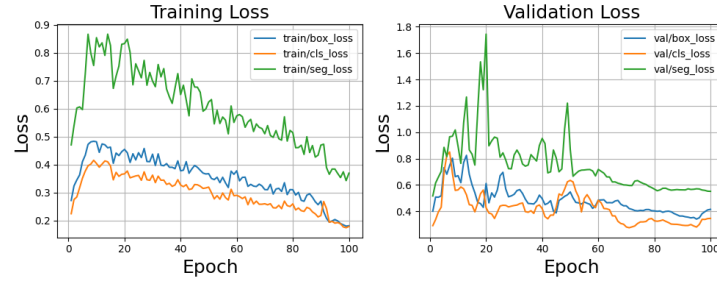
Model	Objek	IoU	BBox				Mask			
			Prec.	Rec.	F1	AP	Prec.	Rec.	F1	AP
YOLOv8m	Lesi	0.88	1.00	0.98	0.99	0.94	1.00	0.98	0.99	0.87
	Serviks	0.95	0.99	1.00	1.00	0.99	0.99	1.00	1.00	0.97
YOLO11l	Lesi	0.90	1.00	0.98	0.99	0.94	1.00	0.98	0.99	0.88
	Serviks	0.94	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.98
YOLO26s	Lesi	0.90	1.00	1.00	1.00	0.95	1.00	1.00	1.00	0.89
	Serviks	0.97	0.99	1.00	0.99	0.99	0.99	1.00	0.99	0.98

Perbedaan hasil segmentasi menunjukkan bahwa desain arsitektur memengaruhi kemampuan model dalam merepresentasikan area lesi. YOLO26s menghasilkan kualitas segmentasi terbaik, yang mengindikasikan kemampuan representasi spasial yang lebih baik pada area lesi dengan bentuk tidak beraturan. Peningkatan performa pada YOLO26s dipengaruhi oleh penggunaan mekanisme *attention* dan strategi seleksi prediksi berbasis Top-K yang membantu model lebih fokus pada area dengan tingkat kepercayaan tinggi. YOLO11l menghasilkan performa yang kompetitif meskipun memiliki kompleksitas model yang lebih tinggi. Hasil tersebut menunjukkan bahwa peningkatan jumlah parameter tidak selalu menghasilkan peningkatan performa yang signifikan pada dataset terbatas. Performa tinggi pada objek serviks menunjukkan bahwa area serviks memiliki karakteristik visual yang relatif konsisten sehingga lebih mudah dipelajari oleh seluruh model. Segmentasi lesi menjadi lebih menantang karena variasi bentuk, tekstur, dan batas area yang lebih kompleks.

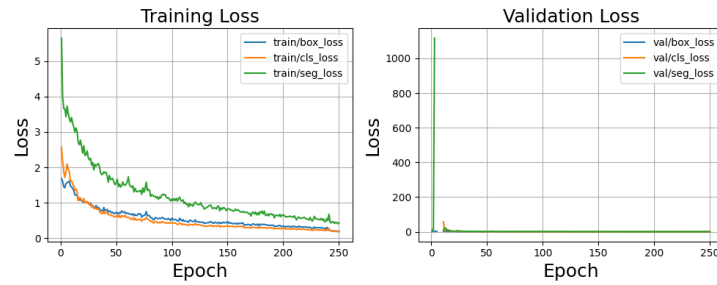
3.2 Analisis Stabilitas Pelatihan

Stabilitas pelatihan dianalisis berdasarkan perubahan *bbox_loss*, *cls_loss*, dan *seg_loss* pada data latih dan validasi. Perbandingan kurva pelatihan menunjukkan bahwa perbedaan antar model lebih terlihat pada komponen segmentasi, terutama karena objek lesi memiliki variasi bentuk,

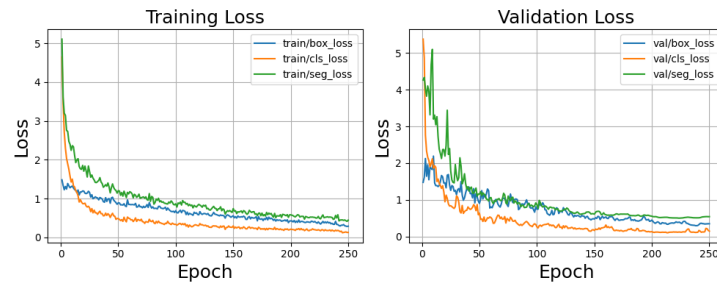
ukuran, dan tekstur yang lebih kompleks. Berdasarkan Gambar 7, perbedaan stabilitas pelatihan antar model terlihat pada pola perubahan *loss* masing-masing model. Pada Gambar 7(a), YOLOv8m menunjukkan penurunan *loss*, tetapi masih disertai fluktuasi yang cukup tinggi, khususnya pada *seg_loss*. Gambar 7(b) menunjukkan bahwa YOLO11l mengalami ketidakstabilan pada fase awal pelatihan, namun kemudian mencapai pola konvergensi yang lebih stabil. Gambar 7(c) memperlihatkan bahwa YOLO26s memiliki penurunan *training loss* dan *validation loss* yang paling konsisten, dengan kurva yang relatif beriringan hingga akhir pelatihan.



(a) Kurva pelatihan YOLOv8m



(b) Kurva pelatihan YOLO11l



(c) Kurva pelatihan YOLO26s

Gambar 7: Perbandingan kurva pelatihan model (a) YOLOv8m, (b) YOLO11l, dan (c) YOLO26s

Pola pelatihan tersebut sejalan dengan hasil evaluasi, di mana YOLO26s memperoleh performa tertinggi pada sebagian besar metrik. Penurunan *seg_loss* yang konsisten menunjukkan bahwa model lebih mampu menyesuaikan prediksi *mask* terhadap *ground truth*. Namun, performa tinggi pada dataset terbatas tetap perlu diinterpretasikan secara hati-hati karena dapat dipengaruhi oleh kecenderungan model mempelajari pola spesifik pada data latih. Dengan demikian, YOLO26s menunjukkan kombinasi terbaik antara performa evaluasi dan stabilitas pelatihan, sedangkan YOLO11l menunjukkan konvergensi yang cukup stabil setelah fase awal, dan YOLOv8m memiliki fluktuasi pembelajaran yang lebih tinggi.

3.3 Analisis Visualisasi Hasil Segmentasi

Analisis visual dilakukan untuk mengevaluasi kemampuan model dalam membedakan citra normal dan abnormal serta melihat kualitas batas segmentasi yang dihasilkan. Visualisasi ini penting karena metrik kuantitatif belum sepenuhnya menggambarkan ketepatan spasial deteksi, terutama pada area lesi dengan batas yang tidak tegas. Berdasarkan Gambar 8, YOLOv8m tidak menghasilkan deteksi positif pada citra normal, sehingga model mampu menekan kemunculan *false positive* pada sampel tersebut. Pada citra abnormal, model mampu mendeteksi area lesi dengan nilai *confidence* sebesar 0,83. Namun, batas segmentasi masih menunjukkan ketidakteraturan pada beberapa bagian, terutama pada area dengan transisi warna dan tekstur yang tidak jelas.



Gambar 8: Hasil segmentasi YOLOv8m pada (a) citra normal dan (b) citra abnormal

Gambar 9 menunjukkan bahwa YOLO11l menghasilkan satu deteksi tidak relevan pada citra normal, yang mengindikasikan adanya *false positive*. Pada citra abnormal, model dapat mendeteksi area lesi dengan nilai *confidence* sebesar 0,70. Meskipun demikian, batas segmentasi belum sepenuhnya konsisten, sehingga masih terdapat bagian prediksi yang belum mengikuti kontur lesi secara optimal.



Gambar 9: Hasil segmentasi YOLO11l pada (a) citra normal dan (b) citra abnormal

Gambar 10 menunjukkan bahwa YOLO26s tidak menghasilkan deteksi lesi pada citra normal, sehingga model mampu menghindari *false positive* pada sampel tersebut. Pada citra abnormal, model mampu mendeteksi area lesi dengan nilai *confidence* sebesar 0,62. Meskipun nilai *confidence* pada sampel ini lebih rendah dibandingkan YOLOv8m dan YOLO11l, hasil evaluasi keseluruhan menunjukkan bahwa YOLO26s tetap memiliki performa terbaik secara kuantitatif. Hal ini menunjukkan bahwa *confidence* pada satu sampel visual tidak dapat dijadikan satu-satunya dasar untuk menilai performa model secara menyeluruh. Secara umum, hasil visualisasi menunjukkan bahwa ketiga model mampu mengenali area lesi pada citra abnormal, tetapi masih

menghadapi tantangan dalam menghasilkan batas segmentasi yang halus dan konsisten. Kesalahan segmentasi terutama muncul pada area dengan kontras rendah, tekstur jaringan yang bervariasi, dan batas lesi yang tidak jelas. YOLO26s menunjukkan keseimbangan terbaik antara kemampuan menghindari *false positive* pada citra normal dan performa evaluasi kuantitatif secara keseluruhan. Namun, hasil visual ini juga menegaskan bahwa validasi pada dataset yang lebih besar dan lebih beragam tetap diperlukan untuk memastikan kemampuan generalisasi model.



Gambar 10: Hasil segmentasi YOLO26s pada (a) citra normal dan (b) citra abnormal

3.4 Perbandingan dengan Penelitian Terdahulu

Perbandingan dengan penelitian terdahulu dilakukan untuk melihat posisi performa model terhadap pendekatan *deep learning* pada tugas deteksi dan segmentasi lesi serviks. Tabel 4 bersifat tidak langsung karena setiap penelitian menggunakan metode anotasi dan karakteristik citra yang berbeda. Penelitian ini menghasilkan nilai IoU sebesar 0.90 dan mAP50 sebesar 1.00 menggunakan model YOLO26s. Nilai tersebut lebih tinggi dibandingkan beberapa penelitian segmentasi sebelumnya seperti SD-U-Net, RUC-U2Net, dan Light-MDDNet (Liu dkk., 2021; He dkk., 2024; Yang dkk., 2025).

Tabel 4: Perbandingan Penelitian Terdahulu

Peneliti	Metode	Model	IoU	mAP50
Liu dkk. (2021)	Segmentasi	SD-U-Net	0.63	-
Nurqolbiah dkk. (2023)	Deteksi	YOLOv7	-	0.92
He dkk. (2024)	Segmentasi	RUC-U2Net	0.85	-
Yang dkk. (2025)	Segmentasi	Light-MDDNet	0.84	-
Penelitian ini	Segmentasi	YOLO26s	0.90	1.00

Penelitian ini juga menghasilkan nilai mAP50 yang kompetitif dibandingkan pendekatan deteksi berbasis YOLOv7 (Nurqolbiah dkk., 2023). Hasil tersebut menunjukkan bahwa pendekatan *instance segmentation* berbasis YOLO mampu menghasilkan segmentasi area lesi yang presisi pada dataset penelitian ini. Integrasi deteksi dan segmentasi dalam satu arsitektur memungkinkan model mempelajari lokasi dan bentuk objek secara simultan sehingga meningkatkan ketepatan segmentasi spasial. Nilai mAP50 yang sangat tinggi perlu diinterpretasikan secara hati-hati karena kemungkinan dipengaruhi oleh ukuran dataset yang terbatas dan karakteristik data yang relatif homogen. Jumlah data yang kecil berpotensi meningkatkan risiko *overfitting* serta membatasi kemampuan generalisasi model terhadap citra *colposcopy* dengan variasi kondisi klinis yang lebih luas. Penggunaan dataset yang lebih besar, lebih beragam, dan berasal dari berbagai sumber masih diperlukan untuk mengevaluasi *robustness* dan kemampuan generalisasi model pada

implementasi klinis nyata.

4 Kesimpulan

Penelitian ini mengevaluasi kemampuan model YOLO berbasis *instance segmentation* untuk mendeteksi dan melakukan segmentasi lesi serviks pada citra *colposcopy*. Pendekatan *single-stage* memungkinkan proses deteksi lokasi objek dan segmentasi area lesi dilakukan secara simultan dalam satu kali inferensi, sehingga meningkatkan efisiensi dibandingkan pendekatan yang memisahkan kedua proses tersebut. Pada skema ini, *bounding box* dan *mask* segmentasi dihasilkan secara bersamaan, dengan *mask* merepresentasikan area lesi secara lebih rinci di dalam objek yang terdeteksi. Hasil eksperimen menunjukkan bahwa seluruh model mampu mendeteksi area serviks dan lesi dengan performa yang baik. YOLO26s menunjukkan performa tertinggi pada segmentasi area lesi, yang ditunjukkan oleh nilai IoU sebesar 0.90 dan mAP50 sebesar 1.00, sedangkan YOLO11l dan YOLOv8m menunjukkan performa yang lebih konsisten dalam menjaga keseimbangan antara *precision* dan *recall*. Temuan ini menunjukkan bahwa pendekatan YOLO berbasis *instance segmentation* memiliki potensi untuk mendukung analisis citra medis secara otomatis, khususnya pada deteksi dan segmentasi lesi serviks, meskipun peningkatan kemampuan generalisasi masih diperlukan untuk penerapan pada data yang lebih beragam.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi S1 Sains Data Universitas Negeri Surabaya serta seluruh pihak yang telah memberikan dukungan dan masukan selama pelaksanaan penelitian ini.

Pustaka

- ACOG (2024). Colposcopy *acog*. *American College of Obstetricians & Gynecologists*.
- Alec, P. V. M. (2024). Cervical cancer information. *Geneva Foundation for Medical Education and Research*.
- Asrina, A., Aliya, N. A., Pasira, I., Magfira, N., Salsadila, A. P., Fadillah, N., dan Undaryati, Y. M. (2025). Update terbaru kanker serviks di indonesia. *OBAT: Jurnal Riset Ilmu Farmasi dan Kesehatan*, 3:212–221.
- Baldovino, R. G., Vidad, A. J. P., Abastillas, R. P. B., Bugtai, N. T., Dadios, E. P., Vicerra, R. R. P., Bandala, A. A., See, A. R., dan Roxas, N. R. (2024). Comprehensive analysis on ultralytics-supported yolo models for detection and recognition of large office objects for indoor navigation. *Procedia Computer Science*, 246:3851–3858.
- Cooper, D. B. dan Dunton, C. J. (2025). *Colposcopy*. StatPearls Publishing. [Updated 2023 Nov 12].
- Hakim, D. M. (2019). *Optical Music Recognition Pada Citra Notasi Musik Menggunakan Convolutional Neural Network*. PhD thesis, Universitas Komputer Indonesia.
- He, Y., Liu, L., Wang, J., Zhao, N., dan He, H. (2024). Colposcopic image segmentation based on feature refinement and attention. *IEEE Access*, 12:40856–40870.
- Hidayatullah, P. dan Tubagus, R. (2026). Yolo26: A comprehensive architecture overview and key improvements a preprint. Technical report.

- Himabindu, D. D., Lydia, E. L., Rajesh, M. V., Ahmed, M. A., dan Ishak, M. K. (2025). Leveraging swin transformer with ensemble of deep learning model for cervical cancer screening using colposcopy images. *Scientific Reports*, 15.
- Jocher, G., Chaurasia, A., dan Qiu, J. (2023). Ultralytics yolov8.
- Jocher, G. dan Qiu, J. (2024). Ultralytics yolo11.
- Jocher, G. dan Qiu, J. (2026). Ultralytics yolo26.
- Khanam, R. dan Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements.
- Kiptoo, S., Nderu, L., dan Mutanu, L. (2021). Automated detection of cervical pre-cancerous lesions using regional-based convolutional neural network. *International Journal of Sciences: Basic and Applied Research (IJSBAR) International Journal of Sciences: Basic and Applied Research*, 56:104–123.
- Li, Z., Yu, L., dan Zhang, J. (2020). Distribution-aware margin calibration for medical image segmentation. *ArXiv*, abs/2011.01462.
- Liu, J., Chen, Q., Fan, J., dan Wu, Y. (2021). Hsil colposcopy image segmentation using improved u-net. In *Proceedings - 2021 36th Youth Academic Annual Conference of Chinese Association of Automation, YAC 2021*, pages 891–897. Institute of Electrical and Electronics Engineers Inc.
- Mumuni, A. dan Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, 16:100258.
- Nurqolbiah, F., Nurmaini, S., dan Saputra, T. (2023). Deteksi lesi pra-kanker serviks pada citra kolposkopi menggunakan convolutional neural network dengan arsitektur yolov7. *Jurnal Sistem Komputer dan Informatika (JSON)*, 5:376.
- Oktafiani, R., Hermawan, A., dan Avianto, D. (2023). Pengaruh komposisi split data terhadap performa klasifikasi penyakit kanker payudara menggunakan algoritma machine learning. *Jurnal Sains dan Informatika*, pages 19–28.
- Perreault, H., Bilodeau, G.-A., Montréal, N. S. P. M., dan Montréal, M. H. G. (2021). Centerpoly: real-time instance segmentation using bounding polygons. Technical report.
- Roboflow (2025). Computer vision annotation formats.
- Sapkota, R. dan Karkee, M. (2026). Ultralytics yolo evolution: An overview of yolo26, yolo11, yolov8 and yolov5 object detectors for computer vision and pattern recognition.
- Saranya, T., Deisy, C., Sridevi, S., dan Anbananthen, K. S. M. (2023). A comparative study of deep learning and internet of things for precision agriculture.
- Shinohara, T., Murakami, K., dan Matsumura, N. (2023). Diagnosis assistance in colposcopy by segmenting acetowhite epithelium using u-net with images before and after acetic acid solution application. *Diagnostics*, 13.
- WHO (2024). Cervical cancer.
- Wolf, J., Kist, L. F., Pereira, S. B., Quessada, M. A., Petek, H., Pille, A., Maccari, J. G., Mutlaq, M. P., dan Nasi, L. A. (2024). Human papillomavirus infection: Epidemiology, biology, host interactions, cancer development, prevention, and therapeutics. *Reviews in medical virology*, 34:e2537.
- Yang, J., Zhang, Y., Fan, W., Wang, J., Zhang, X., Liu, C., Liu, S., dan Xue, L. (2025). A novel lightweight multi-scale feature fusion segmentation algorithm for real-time cervical lesion screening. *Scientific Reports*, 15.
- Yu, H., Fan, Y., Ma, H., Zhang, H., Cao, C., Yu, X., Sun, J., Cao, Y., dan Liu, Y. (2022). Segmentation of the cervical lesion region in colposcopic images based on deep learning. *Frontiers in Oncology*, 12.