Multi Object Tracking pada Citra Jantung Anak menggunakan *Deep Learning*

Muhammad Hanif Sya'bani
Jurusan Sistem Komputer
Fakultas Ilmu Komputer
Universitas Sriwijaya

Annisa Darmawahyuni
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
riset.annisadarmawahyuni@gmail.com

Ade Iriani Sapitri
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
adeirianisapitri13@gmail.com

Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
mh070940@gmail.com

Anggun Islami
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
anggunislami2@gmail.com

***Abstract*—Penyakit jantung bawaan (PJB) pada anak, seperti kasus *atrial septal defect* (ASD) dan *atrioventricular septal defect* (AVSD), memerlukan diagnosis yang akurat melalui analisis dinamis. Namun, metode analisis video ekokardiografi yang ada seringkali terbatas pada analisis per *frame* dan belum mampu melacak perubahan temporal secara konsisten. Penelitian ini mengusulkan pendekatan pelacakan multi-objek berbasis segmentasi menggunakan *deep learning* untuk mengatasi keterbatasan tersebut. Metode yang diusulkan mengintegrasikan arsitektur *You Only Look Once Versi 8* (YOLOv8) untuk melakukan segmentasi, yang kemudian diikuti oleh algoritma *ByteTrack* untuk pelacakan objek yang konsisten antar *frame* pada video ekokardiografi. Pendekatan ini diterapkan pada video kasus ASD, AVSD, dan kondisi jantung normal. Sebagai hasil dari penelitian, hasil kinerja didapatkan dengan matriks *Intersection Over Union* (IoU) sebesar 68% dan *Mean Average Precision Segmentation Mask* (mAP_{mask}) 85% untuk segementasi. Adapun evaluasi kinerja pelacakan mencapai hasil terbaik pada kondisi normal dengan *Multiple Object Tracking Accuracy* (MOTA) sebesar 81,98%, *ID F1 Score* (IDF1) 80,79%, dan *Localization Accuracy* (LocA) 80,22%.**

Keywords — Atrial Septal Defect, Atrioventricular Septal Defect, ByteTrack, Segmentasi, Pelacakan Objek, YOLOv8

I. PENDAHULUAN

Penyakit jantung bawaan (PJB) adalah kondisi yang melibatkan masalah struktural pada struktur jantung yang sudah ada sejak lahir, dengan insidensi sekitar 8 per 1.000 kelahiran hidup secara global [1]. Beberapa kasus penyakit jantung pada anak adalah *atrial septal defect* (ASD), dan *atrioventricular septal defect* (AVSD) yang dapat mengganggu fungsi jantung dan memerlukan penanganan segera [2][3][4][5]. Oleh karena itu, pemantauan berkelanjutan terhadap penyakit jantung pada anak sangat penting untuk meningkatkan tingkat kelangsungan hidup dan mengurangi morbiditas jangka panjang [6]. Salah satu solusi untuk menjawab tantangan ini, kecerdasan buatan terutama *deep learning* telah banyak digunakan untuk melakukan otomatisasi tugas termasuk dalam analisis video ekokardiogram jantung [7] .

Beberapa studi telah membahas mengenai klasifikasi gambar, deteksi objek dan segmentasi statis. Eksplorasi mengenai karakteristik dari defek septum jantung dalam video ekokardiografi masih terbatas, terutama terkait penerapan segmentasi instans yang dipadukan dengan pelacakan multi-objek. Pada penelitian [8], model segmentasi universal (UniverSeg) dikembangkan pada data medis yang besar dan memperkenalkan teknik konvolusi silang yang memungkinkan model untuk menghubungkan informasi antara gambar yang ingin diprediksi dengan gambar contoh yang sudah memiliki hasil segmentasi sehingga menunjukkan hasil yang baik tetapi terbatas pada data dua dimensi dengan satu label dan tanpa menangani analisis video. Sementara pada penelitian [9] arsitektur YOLOv5 dan YOLOv7 dimanfaatkan untuk melakukan deteksi dan segmentasi tumor otak dari citra *Magnetic Resonance Image* (MRI) statis. Model tersebut menunjukkan performa yang sangat baik namun, pendekatan ini masih berfokus pada analisis citra statis dan belum diaplikasikan untuk tugas segmentasi dinamis pada data video. Terlepas dari perkembangan yang ada, metode-metode tersebut masih berfokus pada analisis tingkat *frame* dan tidak mencakup pelacakan objek dalam video ekokardiografi.

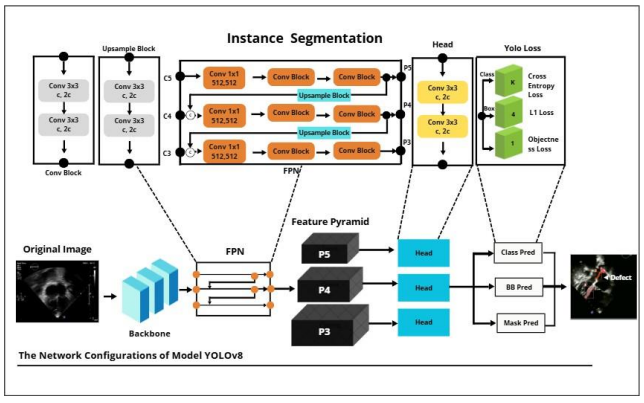
Pelacakan multi-objek pada video ekokardiografi menjadi fokus penelitian karena potensinya dalam melacak defek septum jantung. Proses ini tidak hanya menuntut lokalisasi yang tepat, tetapi juga konsistensi temporal sepanjang siklus jantung pada anak [10][11]. Beberapa metode telah dikembangkan menggunakan *deep learning*, seperti YOLOv5 dengam *deep-SORT* untuk pelacakan jantung janin [12] namun pendekatan tersebut berfokus hanya melacak objek anatomi besar dan tidak dilatih untuk deteksi objek yang lebih kecil. Pada [13], model *Segment Anything Model Tracking* yang dikembangkan khusus untuk modalitas USG (SonoSAMTrack) berfokus pada pelacakan kontur secara umum, bukan pada objek kecil seperti defek. Namun, model ini juga belum terintegrasi dengan algoritma pelacakan multi-objek serta belum diuji pada data ekokardiografi pediatrik. Keterbatasan tersebut menunjukkan adanya celah klinis yang cukup berdampak. Segmentasi saja tidak memadai karena hanya mengenali defek septum dari satu *frame* ke *frame* berikutnya tanpa mampu menangkap perubahan temporalnya [14]. Sebaliknya, pelacakan temporal menawarkan konsistensi berkelanjutan serta memungkinkan dokter memantau perubahan morfologi defek septum secara dinamis seiring waktu [15].

Sehingga penelitian ini mengusulkan pendekatan dengan algoritma pembelajaran berbasis *deep learning* dengan arsitektur YOLOv8 untuk melakukan segmentasi serta *ByteTrack* untuk pelacakan objek defek septum dalam video ekokardiografi. Pada penelitian ini, proses pelacakan berbasis segmentasi difokuskan pada kasus ASD, AVSD, serta kondisi jantung normal yang kemudian akan dievaluasi kinerjanya. Dengan demikian, hasil penelitian ini dapat menjadi landasan bagi pengembangan metode yang lebih efektif dalam membantu proses diagnosis penyakit jantung bawaan pada anak.

II. METODOLOGI

A. *Arsitektur YOLOv8*

YOLOv8 merupakan salah satu model deteksi objek yang telah berkembang pesat selama beberapa tahun terakhir, dengan setiap iterasi baru memperkenalkan peningkatan dalam akurasi dan/atau kecepatan [16]. Ilustrasi arsitektur dari YOLOv8 ditunjukkan pada Gambar 1.



Gambar 1. Arsitektur YOLOv8

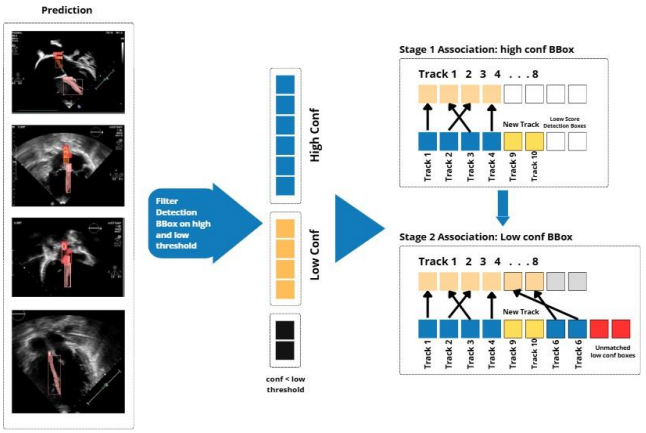
YOLO terdiri dari 2 komponen utama yaitu backbone dan head yang dapat dibagi lebih lanjut menjadi neck dan segment [17]. Backbone berfungsi sebagai *feature extractor* yang mengambil representasi dari citra input. Pada YOLOv8, backbone menggunakan kombinasi lapisan konvolusi 3×3, modul C2f sebagai pengganti C3 pada YOLOv5, serta Spatial Pyramid Pooling-Fast (SPPF) untuk menangkap informasi multi-skala [18].

Neck berfungsi menghubungkan *backbone* dengan *head*, dengan tujuan menggabungkan fitur dari berbagai tingkat resolusi agar model dapat mendeteksi objek secara akurat. Hasilnya, model memiliki representasi fitur yang lengkap untuk klasifikasi dan deteksi lokasi objek [19][20].

Bagian *head* berfungsi untuk melakukan prediksi akhir berdasarkan fitur yang telah diperkaya oleh *neck*. Hasil segmentasi akhir untuk setiap instans objek diperoleh dengan melakukan kombinasi linear antara *prototype mask* dan *mask coefficients* yang bersangkutan, kemudian hasilnya di-*crop* menggunakan *bounding box* yang telah diprediksi [21][22].

B. *ByteTrack*

Algoritma *ByteTrack* dipilih karena memiliki keunggulan dalam menjaga konsistensi identitas objek, meskipun terjadi oklusi atau perubahan gerakan cepat pada citra [23]. Berikut ilustrasi algoritma dari ByteTrack ditunjukkan pada Gambar 2.



Gambar 2. Algoritma ByteTrack

Mekanisme kerja *ByteTrack* didasarkan pada paradigma *Tracking-by-Detection* (TBD), sebuah pendekatan umum dalam pelacakan objek. Objek dideteksi di setiap *frame*, lalu hasil deteksi tersebut dihubungkan untuk membentuk jejak (*trajectory*) [23][24]. Berbeda dengan algoritma TBD pada umumnya yang langsung membuang kotak deteksi dengan skor kepercayaan rendah, *ByteTrack* memanfaatkan informasi ini. Metode asosiasi data yang digunakan untuk melakukan proses pencocokan dua tahap. Tahap pertama mengasosiasikan kotak deteksi dengan skor kepercayaan tinggi ke jejak (*trajectory*) yang sudah ada menggunakan metrik IoU dan algoritma Hungarian [25].

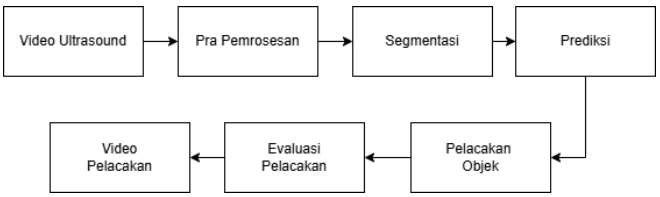
Pada tahap kedua, kotak deteksi dengan skor kepercayaan rendah tidak langsung dibuang. Sebaliknya, kotak-kotak ini dicocokkan kembali dengan jejak yang belum berhasil terasosiasi pada tahap pertama. Algoritma ini mengelola pembuatan jejak baru untuk objek yang baru terdeteksi dan menghapus jejak yang telah hilang dalam durasi tertentu untuk menjaga efisiensi pelacakan [23][25].

III. EKSPERIMEN DAN HASIL

Pada penelitian ini berfokus pada data kasus penyakit jantung bawaan yang teridentifikasi ASD, AVSD serta kondisi Normal. Penelitian ini melakukan proses segmentasi dengan menggunakan YOLOv8 untuk melakukan proses segmentasi gambar dan algoritma *ByteTrack* untuk proses pelacakan pada data yang diusulkan.

A. Material dan Metode

Melakukan proses pelacakan pada citra jantung anak dimulai dari tahapan pengumpulan data berupa video ekokardiogram, baik dengan kondisi normal maupun yang terindentifikasi ASD, dan AVSD. Data yang diperoleh kemudian melalui proses pengolahan awal agar siap digunakan. Penelitian ini mengusulkan metode segmentasi menggunakan YOLOv8 yang diintegrasikan dengan algoritma *ByteTrack* untuk melakukan pelacakan objek secara konsisten pada setiap *frame* video. Adapun tahapan yang dilakukan ditunjukkan pada Gambar 3.



Gambar 3. Alur Proses Segmentasi dan Pelacakan

Gambar 3 merupakan proses segmentasi dan pelacakan yang akan dilakukan pada penelitian ini. Ada beberapa tahapan yang akan dilakukan yaitu persiapan data video, pra-pemrosesan, segmentasi gambar, prediksi, pelacakan objek, evaluasi hasil pelacakan, sampai jadi video hasil pelacakan. Semua tahapan ini dilakukan agar objek yang menjadi target dapat diidentifikasi dan dilacak pergerakannya secara akurat dan konsisten dari satu *frame* ke *frame* berikutnya.

Data Persiapan

Data yang digunakan pada penelitian ini adalah video ekokardiografi jantung anak pada kasus ASD, AVSD serta kondisi normal. Setiap kasus direpresentasikan melalui video dari pandangan jantung *four-chamber* (4CH). Gambar 4 merupakan video yang diusulkan pada penelitian ini.



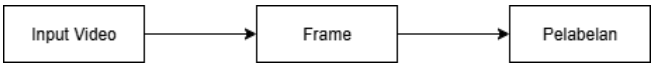
(a) (b) (c)

Gambar 4. (a) ASD (b) AVSD (c) Normal

Video pada penelitian ini diolah kembali sehingga menjadi data yang siap digunakan untuk tahap selanjutnya. Proses selanjutnya melakukan pra-pemrosesan data yang sudah didapatkan.

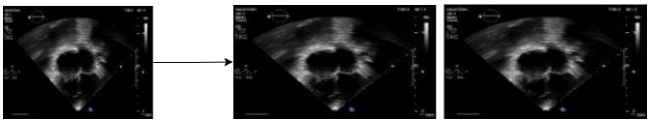
Pra-Pemrosesan

Tahap ini mengolah data menjadi siap dipakai untuk di implementasi sebagai input dari algoritma yang akan digunakan nantinya. Berikut adalah proses awal dalam melakukan pra-pemrosesan data.



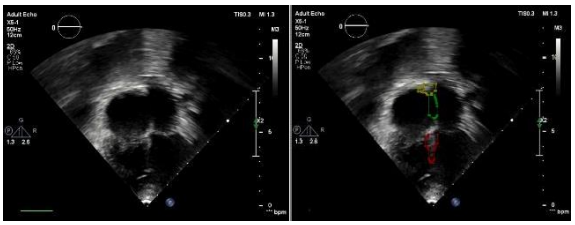
Gambar 5. Alur Pemrosesan Data

Pada Gambar 5 ditunjukkan alur pemrosesan dimulai dari video yang didapatkan diubah menjadi *frame* atau beberapa gambar. Tahap ini dilakukan dengan library python OpenCV. Berikut hasil *frame* yang didapatkan ditujukkan pada Gambar 6.



Gambar 6. Video ke *Frame*

Hasil yang didapatkan dari proses pengubahan video ke *frame* akan menghasilkan banyak *frame* tergantung dari durasi dan *frame per second* (FPS). Berdasarkan output yang didapatkan, *frame* memiliki ukuran 800 x 600. Tahap berikutnya adalah pelabelan data agar proses segmentasi berjalan dengan baik dan akurat. Proses label gambar atau *ground truth* dilakukan menggunakan bantuan model prediksi segmentasi sebelumnya. Pada tahap ini akan menandai bagian defek septum, interatrial septum, dan interventricular septum. Berikut hasil yang didapatkan dari proses penglabelan ditunjukkan pada gambar 7.



Gambar 7. Penglabelan Gambar

B. Training dan Testing Segmentasi

Proses data latih untuk segmentasi instans yang dilakukan dengan menggunakan dataset ASD, AVSD, dan Normal. Dataset ini dibagi menjadi tiga bagian utama, data latih, data validasi, dan data *unseen*. Setiap model akan dilatih menggunakan pasangan data gambar asli dan *ground truth* yang sesuai. Rincian lengkap mengenai konfigurasi data latih, validasi dan uji disajikan pada Tabel I.

Tabel I Split Data

Total Frame		
Train	Val	Unseen
24.490	6.306	11.491

C. Evaluasi Hasil Segmentasi

Pada proses evaluasi, gambar memberikan hasil yang diperoleh dengan menggunakan metode yang di usulkan. Kinerja algoritma dievaluasi menggunakan data validasi dan data unseen sebagai data pengujian. Evaluasi ini menggunakan dua metrik evaluasi, yaitu *Intersection over Union* (IoU) dan *Mean Average Precision Segmentation Mask* (mAP_{mask}). IoU, seperti dijelaskan pada Persamaan (1), digunakan untuk mengukur seberapa akurat tumpang tindih antara area prediksi dengan area sebenarnya. Sementara itu, mAP_{mask}, yang dirumuskan pada Persamaan (2), memberikan evaluasi yang lebih komprehensif dengan mengukur presisi rata-rata model pada ambang batas IoU.

$$IoU = \frac{\left|\hat{M} \cap M\right|}{\left|\hat{M} \cup M\right|} \quad (1)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{n} APi \quad (2)$$

Tabel II hasil evaluasi segmentasi

Split	mAP_{mask} (%)	IoU (%)
Validasi	82.00	65.30
Unseen	84.00	69.25

Berdasarkan hasil yang ditunjukkan pada Gambar 8 dan Tabel II, model YOLOv8 mampu melakukan segmentasi dengan performa yang baik. Pada data validasi, model mencapai nilai mAP_{mask} sebesar 82,00% dan IoU sebesar 65,30%, sedangkan pada data *unseen*, performa meningkat dengan mAP_{mask} sebesar 84,00% dan IoU sebesar 69.25%. Hasil ini menunjukkan bahwa model tidak hanya mampu mengenali pola segmentasi pada data pelatihan, tetapi juga memiliki kemampuan generalisasi yang kuat terhadap data baru yang belum pernah dilihat sebelumnya.

ASD **AVSD** **NORMAL**



Gambar 8. Hasil segmentasi menggunakan YOLOv8

Pada Gambar 8 setiap objek ditandai dengan bounding box dan mask. Pada kasus ASD terdapat tiga objek yang tersegmentasi, defek septum ditandai berwarna orange, interatrial septum ditandai dengan warna merah dan interventricular septum ditandai dengan warna merah muda. Kasus AVSD menunjukkan struktur serupa, namun dengan perbedaan spesifik di mana dua objek defek septum berhasil tersegmentasi. Sementara pada kondisi Normal hanya ada interatrial septum ditandai dengan warna merah dan interventricular septum ditandai dengan warna merah muda. Hasil ini mengindikasikan bahwa model segmentasi yang dikembangkan memiliki kinerja yang konsisten dalam mengidentifikasi serta membedakan struktur anatomi jantung pada berbagai kondisi.

D. Pelacakan Objek

Setelah model segmentasi dari *view* 4CH diperoleh, proses pelacakan objek diterapkan pada setiap data video. Adapun sebaran data dibagi dalam kasus ASD, AVSD, dan Normal ke *train* dan *unseen* seperti pada Tabel III.

Tabel III Sebaran data video ekokardiografi

Kasus	Total Video	
	Train	Test Unseen
ASD	55	21
AVSD	3	2
NORMAL	66	24

Tahap ini hanya difokuskan untuk melacak objek yang telah diklasifikasikan sebagai defek septum. Setiap *frame* pertama yang terdeteksi sebuah defek, objek tersebut diinisialisasi dengan ID pelacakan yang bersifat unik dan informasi lokasinya disimpan. Pada *frame* berikutnya, setiap deteksi defek yang baru akan dicocokkan dengan informasi defek yang tersimpan dari *frame* sebelumnya. Jika sebuah deteksi baru cocok, ia akan mewarisi ID yang sama, sehingga kontinuitas jejaknya terjaga. Hasil akhir dari tahap ini adalah video hasil pelacakan, di mana setiap defek septum secara konsisten ditandai dengan visualisasi dengan *bounding box* dan ID pada setiap *frame* kemunculannya.

E. Evaluasi Pelacakan Objek

Setelah proses pelacakan selesai, evaluasi dilakukan untuk mengukur efektivitas metode yang diusulkan. Evaluasi ini bertujuan untuk mengukur seberapa baik model dapat secara konsisten mengidentifikasi dan mengikuti defek septum di sepanjang video. Pengujian kinerja menggunakan beberapa metrik seperti pada persamaan (3)(4)(5)

$$LocA = \frac{1}{|TP|}\sum_{I=1}^{|TP|} IoU(b_i^t, b_j^{t+1}) \quad (3)$$

$$IDF1 = \frac{2 \cdot (IDTP)}{2 \cdot (IDTP + IDFN + IDFP)} \quad (4)$$

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t (GT_t)} \quad (5)$$

Tabel IV Performa YOLOv8

Kasus	YOLOv8 (%)		
	LocA	IDF1	MOTA
ASD	76.77	63.50	59.10
AVSD	77.76	74.82	51.93
Normal	80.22	80.79	81.98

Tabel IV merupakan hasil pelacakan menggunakan YOLOv8 dengan *ByteTrack* pada data *unseen*

IV. DISKUSI

Dalam penelitian ini, penerapan *deep learning* digunakan untuk menghasilkan sistem pelacakan multi-objek pada video ekokardiografi jantung anak melalui integrasi model segmentasi YOLOv8 dan algoritma pelacakan *ByteTrack*. Pendekatan ini memanfaatkan deteksi spasial dari YOLOv8 serta kemampuan asosiasi temporal dari *ByteTrack* dalam mempertahankan identitas objek sepanjang rangkaian *frame*.

Model menunjukkan kinerja yang baik, terutama pada data yang belum pernah dilihat sebelumnya, yang tercermin dari hasil evaluasi menggunakan metrik mAP_{mask} dan IoU pada tahap segmentasi. Model mampu mengenali dan melokalisasi struktur septum jantung dengan baik, termasuk area defek. Pada tahap pelacakan, performa yang diukur melalui metrik LocA, IDF1, dan MOTA menunjukkan hasil yang konsisten pada berbagai kondisi jantung. Nilai tertinggi diperoleh pada kondisi jantung normal, sedangkan performa relatif lebih rendah pada kasus AVSD disebabkan oleh kompleksitas morfologi defek dan keterbatasan data. Secara keseluruhan, pendekatan ini berpotensi meningkatkan efisiensi analisis citra ekokardiografi untuk pemantauan struktur jantung anak secara berkelanjutan.

V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan metode pelacakan multi-objek integrasi YOLOv8 untuk melakukan proses segmentasi dan *ByteTrack* untuk proses pelacakan pada video ekokardiografi jantung anak. Hasil penelitian pada proses segmentasi digunakan proses evaluasi dengan mAP_{mask} dan IoU. Hasil penelitian menunjukkan bahwa model segmentasi mencapai kinerja yang baik, terutama pada data *unseen* dengan mAP_{mask} 85,20% dan IoU 67,73%. Pada proses pelacakan, metode yang diusulkan menunjukkan hasil yang konsisten, dengan kinerja terbaik dicapai pada kondisi jantung normal, di mana nilai MOTA mencapai 81,98%, IDF1 80,79%, dan LocA 80,22%. Hasil yang didapatkan selama proses pelacakan dapat dilihat pada Tabel IV.

Temuan ini menunjukkan potensi penerapan metode pelacakan otomatis dalam mendukung analisis klinis ekokardiografi, meningkatkan efisiensi interpretasi video jantung, serta memperkuat konsistensi pelaporan hasil pemeriksaan.

REFERENSI

[1] L. Song *et al.*, "Clinical profile of congenital heart diseases detected in a tertiary hospital in China: a retrospective analysis," *Front Cardiovasc Med*, vol. 10, 2023, doi: 10.3389/fcvm.2023.1131383.

[2] R. Nova, S. Nurmaini, R. U. Partan, and S. T. Putra, "Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: A preliminary study," *Inform Med Unlocked*, vol. 24, Jan. 2021, doi: 10.1016/j.imu.2021.100601.

[3] A. Dozen *et al.*, "Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information," *Biomolecules*, vol. 10, no. 11, pp. 1–17, Nov. 2020, doi: 10.3390/biom10111526.

[4] K. Miura *et al.*, "Deep learning-based model detects atrial septal defects from electrocardiography: a cross-sectional multicenter hospital-based study," *EClinicalMedicine*, vol. 63, Sep. 2023, doi: 10.1016/j.eclinm.2023.102141.

[5] Y. LIU *et al.*, "Intelligent diagnosis of atrial septal defect in children using echocardiography with deep learning," *Virtual Reality and Intelligent Hardware*, vol. 6, no. 3, pp. 217–225, Jun. 2024, doi: 10.1016/j.vrih.2023.05.002.

[6] R. Vullings, "Fetal Electrocardiography and Deep Learning for Prenatal Detection of Congenital Heart Disease," in *2019 Computing in Cardiology (CinC)*, 2019, p. Page 1-Page 4. doi: 10.22489/CinC.2019.072.

[7] E. Cinteza *et al.*, "Can Artificial Intelligence Revolutionize the Diagnosis and Management of the Atrial Septal Defect in Children?," Jan. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics14020132.

[8] V. I. Butoi, J. J. Gonzalez Ortiz, T. Ma, M. R. Sabuncu, J. Guttag, and A. V Dalca, "UniverSeg: Universal Medical Image Segmentation." [Online]. Available: https://universeg.csail.mit.edu

[9] M. F. Almufareh, M. Imran, A. Khan, M. Humayun, and M. Asim, "Automated Brain Tumor

Segmentation and Classification in MRI Using YOLO-Based Deep Learning," *IEEE Access*, vol. 12, pp. 16189–16207, 2024, doi: 10.1109/ACCESS.2024.3359418.

[10] Y. S. Yoo, S. H. Lee, and S. H. Bae, "Effective Multi-Object Tracking via Global Object Models and Object Constraint Learning," *Sensors (Basel)*, vol. 22, no. 20, Oct. 2022, doi: 10.3390/s22207943.

[11] J. Wang *et al.*, "Automated interpretation of congenital heart disease from multi-view echocardiograms," *Med Image Anal*, vol. 69, Apr. 2021, doi: 10.1016/j.media.2020.101942.

[12] E. DANDIL, M. TURKAN, F. E. URFALI, İ. BİYİK, and M. KORKMAZ, "Fetal Movement Detection and Anatomical Plane Recognition using YOLOv5 Network in Ultrasound Scans," *European Journal of Science and Technology*, Jun. 2021, doi: 10.31590/ejosat.951786.

[13] H. Ravishankar *et al.*, "SonoSAMTrack -- Segment and Track Anything on Ultrasound Images," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2310.16872

[14] R. K. Baghbaderani, Y. Li, S. Wang, and H. Qi, "Temporally-Consistent Video Semantic Segmentation with Bidirectional Occlusion-guided Feature Propagation," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 674–684. doi: 10.1109/WACV57701.2024.00074.

[15] Z. Zhang, Q. Wu, S. Ding, X. Wang, and J. Ye, "Echo-Vision-FM: A Pre-training and Fine-tuning Framework for Echocardiogram Video Vision Foundation Model," Oct. 10, 2024. doi: 10.1101/2024.10.09.24315195.

[16] R. Sapkota, D. Ahmed, and M. Karkee, "Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments."

[17] A. A. Alsuwaylimi, "Enhanced YOLOv8-Seg Instance Segmentation for Real-Time Submerged Debris Detection," *IEEE Access*, vol. 12, pp. 117833–117849, 2024, doi: 10.1109/ACCESS.2024.3448258.

[18] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.02696

[19] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1612.03144

[20] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," Sep. 2018, [Online]. Available: http://arxiv.org/abs/1803.01534

[21] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time Instance Segmentation," Oct. 2019, [Online]. Available: http://arxiv.org/abs/1904.02689

[22] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++: Better Real-time Instance Segmentation," Sep. 2020, doi: 10.1109/TPAMI.2020.3014297.

[23] Y. Zhang *et al.*, "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2110.06864

[24] Y. Zhang *et al.*, "ByteTrackV2: 2D and 3D Multi-Object Tracking by Associating Every Detection Box," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.15334

[25] Y. Wang and V. Y. Mariano, "A Multi Object Tracking Framework Based on YOLOv8s and Bytetrack Algorithm," *IEEE Access*, vol. 12, pp. 120711–120719, 2024, doi: 10.1109/ACCESS.2024.3450370.