Model Transformer Berbasis Deep Learning Untuk Segmentasi Citra Jantung Janin Dengan Kardiomegali

1st Gracya Nurmala Sinurat
Jurusan Sistem Komputer
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia

2nd Siti Nurmaini
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
sitinurmaini@gmail.com

3rd Annisa Darmawahyuni
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
riset.annisadarmawahyuni@gmail.com

Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
gracyasinurat469@gmail.com

4th Ade Iriani Sapitri
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
adeirianisapitri13@gmail.com

5th Anggun Islami
Intelligent System Research Group
Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
anggunislami2@gmail.com

Ringkasan*—Peran segmentasi ruang jantung janin yang akurat sangat krusial untuk diagnosis dini penyakit jantung bawaan. Namun, segmentasi manual pada citra ultrasonografi (USG) masih menghadapi tantangan signifikan, seperti subjektivitas operator serta kualitas citra yang rendah akibat noise dan kontras yang lemah. Penelitian ini mengusulkan penggunaan arsitektur SegFormer-B3, yaitu model pembelajaran dalam berbasis Transformer, untuk melakukan segmentasi otomatis pada area *Heart Circumference* (HC) dan *Chest Circumference* (CC) citra ultrasonograpi jantung janin dengan indikasi kardiomegali. Model ini menggabungkan *hierarchical encoder* yang mampu mengekstraksi fitur multi-skala dan *MLP decoder* ringan untuk menghasilkan peta segmentasi yang efisien. Dataset penelitian terdiri dari 509 citra ultrasonografi yang digunakan untuk pelatihan dan pengujian model. Evaluasi dilakukan menggunakan metrik *Intersection over Union (IoU)* dan *Dice Coefficient* untuk menilai kesesuaian hasil prediksi terhadap *ground truth*. Hasil menunjukkan nilai *IoU* dan *Dice Coefficient* masing-masing sebesar 0.6859 dan 0.8137 untuk kelas HC dan untuk kelas CC sebesar 0.5834 dan 0.7369. Temuan ini menunjukkan bahwa SegFormer-B3 mampu melakukan segmentasi citra jantung janin dengan tingkat konsistensi yang memadai, serta berpotensi digunakan sebagai alat bantu analisis morfologi awal dalam skrining kardiomegali prenatal.

***Keywords*—Segmentasi Citra Medis, Kardiomegali Janin, SegFormer, Transformer, Deep Learning.**

Abstract*— *Accurate segmentation of fetal heart chambers plays a crucial role in the early diagnosis of congenital heart disease. However, manual segmentation in fetal ultrasonography (USG) images remains challenging due to operator subjectivity and low image quality affected by noise and weak contrast. This study proposes the implementation of the SegFormer-B3 architecture, a Transformer-based deep learning model, to automatically segment the Heart Circumference (HC) and Chest Circumference (CC) regions in fetal cardiac ultrasound images with cardiomegaly indications. The model combines a hierarchical encoder for multi-scale feature extraction and a lightweight MLP decoder to generate efficient segmentation maps. A total of 509 fetal ultrasound images were used for model training and evaluation. The segmentation performance was assessed using Intersection over Union (IoU) and Dice Coefficient metrics to measure the agreement between the predicted masks and the ground truth. Experimental results achieved IoU and Dice Scores of 0.6859 and 0.8137 for HC, and for CC of 0.5834 and 0.7369, demonstrating that SegFormer-B3 provides consistent segmentation performance. These findings suggest that the proposed approach has strong potential as a reliable tool for early morphological analysis in prenatal cardiomegaly screening.

***Keywords*—Medical Image Segmentation, Fetal Cardiomegaly, SegFormer, Transformer, Deep Learning**

I. PENDAHULUAN

Kemampuan untuk melakukan segmentasi ruang jantung janin secara akurat memegang peranan vital dalam mendeteksi penyakit jantung bawaan sejak dini [1]. Penerapan algoritma deep learning dalam domain analisis citra medis telah menjadi subjek dalam berbagai literatur ilmiah [2]. Salah satu kelainan spesifik yang menjadi fokus adalah kardiomegali, atau pembesaran jantung. Ada berbagai macam metode deteksi kardiomegali, salah satunya menggunakan arsitektur Convolutional Neural Network (CNN) seperti U-Net [3]. Pengembangan sistem *computer-aided diagnose* (CAD) sangat penting untuk meningkatkan objektivitas dan akurasi [4]. Pada penelitian sebelumnya, CNN merupakan arsitektur dominan untuk segmentasi *four-chamber view* [5]. Namun, segmentasi manual masih banyak digunakan dalam praktik klinis. Segmentasi manual memiliki keterbatasan. Proses ini bergantung pada operator dan lama [6]. Citra ultrasonografi janin juga memiliki tantangan. Citra ultrasonografi sering terpengaruh oleh *noise* dan artefak, sehingga memerlukan metode deep learning yang *robust* [7]. Keberhasilan deep learning tidak terbatas pada jantung.

ISSN : xxxx-xxxx
Vol.1

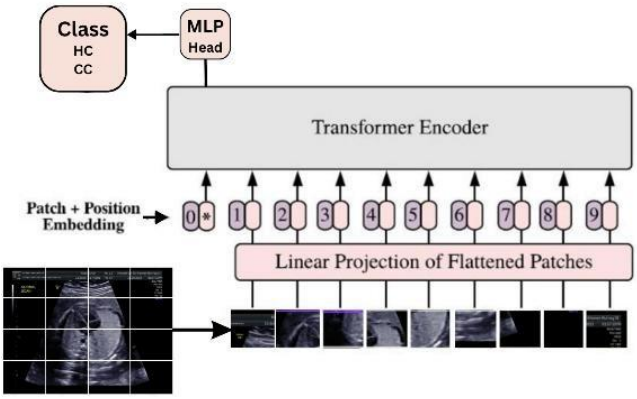
Segmentasi multiclass berbasis deep learning juga telah berhasil diterapkan pada struktur janin lain, seperti jaringan otak [8]. Meskipun CNN berhasil, CNN memiliki keterbatasan. CNN terbatas dalam menangkap hubungan spasial jarak jauh [9]. Pentingnya deteksi kardiomegali juga ditekankan pada modalitas Sinar-X [10]. Oleh karena itu, beberapa tahun terakhir banyak algoritma diusulkan berbasis Transformer [9]. Di antara arsitektur Transformer, SegFormer telah diidentifikasi sebagai kandidat kuat. Model ini telah dieksplorasi dan dimodifikasi untuk meningkatkan kinerjanya pada segmentasi ekokardiografi ventrikel kiri [11]. Keberhasilan SegFormer dalam domain janin telah divalidasi. Teknik *fine-tuning* terbukti efektif untuk segmentasi organ janin lainnya, seperti kepala janin [12]. Transisi menuju Transformer didorong karena CNN murni kesulitan memodelkan konteks global [13]. Arsitektur Transformer murni juga terbukti efektif untuk segmentasi ventrikel kiri [14]. Selain itu, efisiensi *backbone* SegFormer telah dimanfaatkan dalam arsitektur hibrida lain. Seperti pada penelitian yang dilakukan oleh [15] menggunakan MiTU-Net. Arsitektur Transformer murni lainnya juga menunjukkan hasil yang baik [16].

Berdasarkan perkembangan tersebut, penelitian ini mengusulkan penerapan model SegFormer. Model ini adalah arsitektur deep learning berbasis Transformer yang efisien. Penelitian ini difokuskan untuk melakukan segmentasi otomatis pada area *Heart Circumference* (HC) dan *Chest Circumference* (CC). Hasil penelitian ini dapat menjadi acuan atau landasan bagi eksplorasi lebih lanjut dalam domain terkait.

II. METODOLOGI

A. Vision Transformer (ViT)

Vision Transformer (ViT) merupakan sebuah arsitektur deep learning yang mengadopsi mekanisme *self-attention*, awalnya populer di pemrosesan bahasa alami untuk tugas pengolahan data citra [17]. Model ini bekerja dengan cara membagi citra masukan menjadi potongan kecil yang disebut sebagai *patch*, kemudian setiap *patch* dikonversi menjadi vektor *embedding* yang selanjutnya diproses oleh blok Transformer untuk menghasilkan representasi fitur global [17][18]. Pendekatan ini memungkinkan model menangkap hubungan spasial antar-piksel secara menyeluruh tanpa menggunakan operasi konvolusi seperti pada CNN, sehingga meningkatkan kemampuan model dalam memahami konteks global pada citra medis [18][19]. Penerapan ViT dalam analisis citra ultrasonografi janin mampu meningkatkan deteksi kelainan struktural secara signifikan karena arsitektur Transformer dapat mempelajari representasi spasial yang kompleks dengan stabilitas tinggi terhadap *noise* dan variasi intensitas citra [17]. Hal ini menjadikan ViT sangat potensial untuk mendukung proses diagnosis dini melalui analisis berbasis segmentasi otomatis pada citra prenatal.



Gambar 1. Ilustrasi Arsitektur ViT

Tahapan awal pada ViT adalah memecah citra berukuran $H \times W \times C$ menjadi N potongan *patch* berukuran $P \times P$. Setiap *patch* direpresentasikan sebagai vektor *embedding* menggunakan persamaan berikut:

$$x_p = x_1 E; x_1\ E; \ldots; x_N E + E_{pos} \quad (1)$$

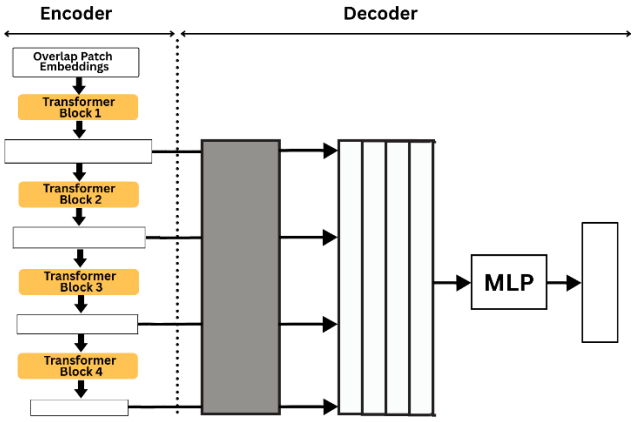
di mana E adalah matriks bobot *embedding* dan E_{pos} merupakan *positional encoding* untuk mempertahankan informasi spasial dati citra [17]. Setiap vektor kemudian diproses oleh mekanisme *multi-head self-attention* (MHSA) seperti pada persamaan berikut:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d^K}}\right)V \quad (2)$$

Mekanisme ini memungkinkan model memahami hubungan antarbagian citra tanpa operasi konvolusi seperti pada CNN [18][19]. Pendekatan ini dalam menangkap konteks global pada citra medis telah terbukti. Hal ini menjadikannya diadopsi secara luas dalam beragam aplikasi analisis citra, yang mencakup deteksi kelainan jantung janin pada citra ultrasonografi [17][20]. Selain itu, arsitektur ViT juga menjadi dasar bagi pengembangan model lanjutan seperti Swin Transformer [21] dan SegFormer [22].

B. Arsitektur SegFormer

SegFormer merupakan pengembangan dari Vision Transformer yang dirancang khusus untuk tugas semantik segmentasi citra secara efisien dan akurat [22]. Arsitektur ini menggabungkan kemampuan global *attention* dari Transformer dengan efisiensi komputasi melalui desain *encoder–decoder* yang sederhana namun efektif [22][23]. Ilustrasi arsitektur SegFormer secara umum ditunjukkan pada Gambar 2.



Gambar 2. Arsitektur SegFormer

Bagian *encoder* bertugas mengekstraksi fitur citra pada beberapa tingkat resolusi menggunakan *overlapping patch embedding* dan blok Transformer bertingkat. Proses ekstraksi fitur dilakukan secara hierarkis, sehingga menghasilkan empat tingkat representasi fitur $F1$, $F2$, $F3$, dan $F4$. Tahapan tersebut dapat dinyatakan dengan persamaan berikut:

$$Fi = Transformer(Ei(Fi-1)) \quad (3)$$

di mana Ei merupakan *embedding layer* pada tahap ke–i dan Fi merupakan keluaran fitur pada tahap tersebut [22][23]. Pendekatan hierarkis ini memungkinkan model menangkap informasi lokal dan global secara bersamaan, yang sangat penting dalam menganalisis struktur anatomi kompleks seperti jantung janin pada citra ultrasonografi [17][22].

Selanjutnya, proses *decoding* dilakukan menggunakan *MLP decoder* untuk menggabungkan hasil dari seluruh lapisan *encoder* tanpa memerlukan proses interpolasi yang kompleks. Penggabungan fitur dari berbagai tingkat resolusi dinyatakan sebagai:

$$Y = MLP(Concat(F1,F2,F3,F4)) \quad (4)$$

Pendekatan ini memungkinkan segmentasi dilakukan secara efisien dan presisi tinggi, bahkan pada citra dengan batas anatomi yang samar seperti jantung janin [22][24]. Untuk meningkatkan akurasi segmentasi, penelitian ini juga mengadaptasi konsep *multi-scale feature fusion* sebagaimana diterapkan pada *Dual-Path Decoder* [23], serta mekanisme *feedback attention* untuk memperbaiki ketepatan batas hasil segmentasi sebagaimana pada penelitian *Efficient Multistage Feedback Attention*[24]. Hasil keluaran SegFormer kemudian diproyeksikan menjadi peta segmentasi dua kelas, yaitu HC dan CC. Untuk mengoptimalkan model selama pelatihan, digunakan fungsi kerugian *Dice Loss* (L_{Dice}) . Fungsi ini dihitung berdasarkan *Dice Coefficient*, sebagaimana ditunjukkan pada persamaan berikut:

$$Dice\ Coefficient = \frac{2\sum_i p_i g_i}{2\sum_i p_i^2 + \sum_i g_i^2} \quad (5)$$

$$L_{Dice} = 1 - Dice\ Coefficient \quad (6)$$

di mana pi merupakan nilai prediksi probabilitas model pada piksel ke–i dan gi merupakan nilai sebenarnya dari *ground truth* [22][24]. Selama pelatihan, model dioptimalkan untuk meminimalkan L_{Dice} (mendekati 0), yang secara efektif memaksimalkan *Dice Coefficient* (mendekati 1) atau kesamaan tinggi antara segmentasi dan data sebenarnya. Selain itu, kinerja segmentasi model juga dievaluasi menggunakan metrik *IoU* untuk setiap kelas HC dan CC guna menilai akurasi spasial per kelas secara spesifik. *IoU* dihitung berdasarkan perbandingan jumlah piksel yang benar-benar tumpang tindih antara hasil prediksi dan *ground truth*, sebagaimana ditunjukkan pada persamaan berikut:

$$IoU = \frac{\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i - \sum_i p_i g_i} \quad (7)$$

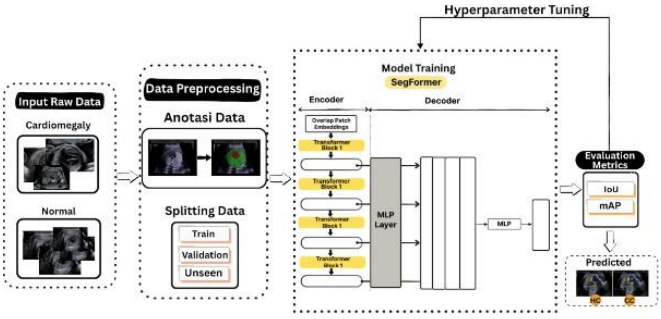
di mana pi dan gi masing mewakili nilai piksel hasil prediksi dan *ground truth* (1 untuk area objek, 0 untuk latar belakang) [23] [25]. Nilai *IoU* yang tinggi mengindikasikan bahwa hasil segmentasi model memiliki tumpang tindih area yang besar dengan *ground truth*, menunjukkan performa segmentasi yang baik dan konsisten terhadap batas anatomi jantung janin.

III. EKSPERIMEN

Proses penelitian ini dijelaskan melalui serangkaian eksperimen dan hasil yang diperoleh. Data citra ultrasonografi jantung janin dikumpulkan dan diproses terlebih dahulu sebelum dimasukkan ke dalam sistem segmentasi. Studi ini berfokus pada penerapan model SegFormer-B3, yaitu arsitektur berbasis Transformer yang digunakan untuk melakukan segmentasi pada citra jantung janin dengan kondisi kardiomegali. Dengan menggunakan model ini, area HC dan CC dapat dipisahkan secara akurat dari struktur sekitarnya. Peta segmentasi yang dihasilkan kemudian dievaluasi menggunakan metrik akurasi, yaitu *Dice Coefficient* dan *IoU*.

A. Material dan Metode

Proses segmentasi, khususnya pada citra jantung janin yang menunjukkan kardiomegali, dilakukan melalui beberapa tahapan yang saling berurutan. Tahapan ini diawali dengan pengumpulan data citra ultranosografi jantung janin yang digunakan sebagai dasar proses pemodelan. Setelah data terkumpul, dilakukan tahap pra-pemrosesan untuk menyiapkan data agar dapat digunakan dalam pelatihan model. Tujuan utama dari rangkaian proses ini adalah untuk membantu peneliti merancang dan mengevaluasi kinerja arsitektur SegFormer dalam menangani tugas segmentasi citra medis. Sebagaimana telah dijelaskan sebelumnya, penelitian ini berfokus pada pembagian dua area utama pada citra jantung janin, yaitu HC dan CC. Gambar 3 menunjukkan alur proses segmentasi yang diterapkan pada penelitian ini.



Gambar 3. Proses Segmentasi

Data Persiapan

Data yang dianalisis dalam penelitian ini adalah citra ultrasonografi jantung janin dengan kondisi terindikasi kardiomegali. Proses segmentasi difokuskan pada dua area utama, yaitu HC dan CC, yang berfungsi untuk membedakan batas anatomi jantung janin. Gambar 4 menunjukkan citra ultrasonografi janin dengan indikasi kardiomegali yang digunakan dalam penelitian ini.



Gambar 4. Citra ultrasonografi kardiomegali

Untuk menghasilkan data yang siap digunakan pada tahap selanjutnya, citra ultrasonografi terlebih dahulu diproses ulang melalui tahap pra-pemrosesan guna meningkatkan kualitas dan konsistensi data, meliputi proses normalisasi intensitas, penghilangan *noise*, serta penyesuaian ukuran citra agar seragam sebelum digunakan dalam pelatihan model.

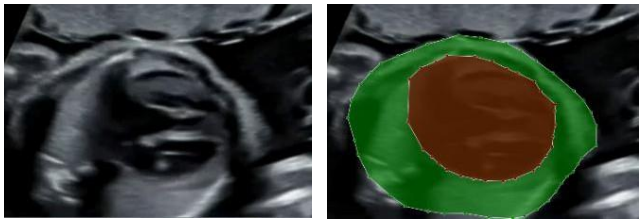
Pra-Pemrosesan

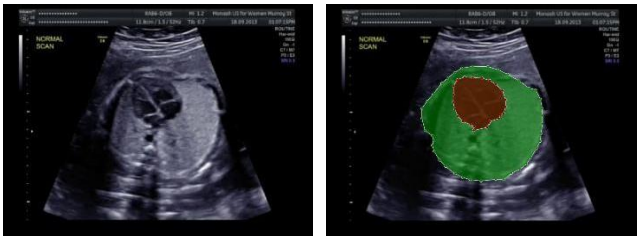
Pada tahap ini, data dipersiapkan agar dapat digunakan sebagai input pada sistem pemodelan berbasis Transformer yang akan digunakan pada proses segmentasi citra. Tahapan awal dalam proses pra-pemrosesan data ditunjukkan pada Gambar 5 yang menggambarkan alur pra-pemrosesan data secara keseluruhan.



Gambar 5. Alur Pra-pemrosesan Data

Dilihat dari diagram alur di atas, citra ultrasonografi yang telah diperoleh dalam bentuk beberapa *frame* selanjutnya melalui tahap pemberian label pada area yang akan disegmentasi. Tahap ini dilakukan dengan bantuan perangkat lunak *LabelMe* dan divalidasi oleh dokter untuk memastikan ketepatan area segmentasi. Proses pelabelan ini berperan penting dalam mempersiapkan data untuk tahap pelatihan model segmentasi. Proses ini menghasilkan *mask biner* (hitam-putih) untuk setiap kelas HC dan CC, yang berfungsi sebagai *ground truth* atau target yang harus dipelajari oleh model. Gambar 6 menunjukkan contoh perbandingan visual antara citra ultrasonografi asli dan *ground truth* yang telah dihasilkan. Setelah seluruh data citra dan *mask ground truth* selesai disiapkan, langkah selanjutnya adalah pembagian dataset untuk pelatihan, validasi, dan pengujian.





Original Citra *Ground Truth*

Gambar 6. Pelabelan Citra

B. Pembagian Data

Dataset ultrasonografi jantung janin dengan kardiomegali terdiri dari 509 citra yang berasal dari 12 pasien. Data dibagi menjadi tiga *subset*, yaitu citra untuk pelatihan (*train*), citra untuk validasi (*validation*), dan citra sebagai data pengujian (*unseen*) guna menilai kemampuan generalisasi model. Tabel I menunjukkan pembagian *dataset* yang digunakan dalam penelitian ini.

Tabel II. Pembagian *Dataset*

	Original Citra	*Ground Truth*
Data *Train*	350	350
Data *Validation*	50	50
Data *Unseen*	109	109

C. Evaluasi Metrik

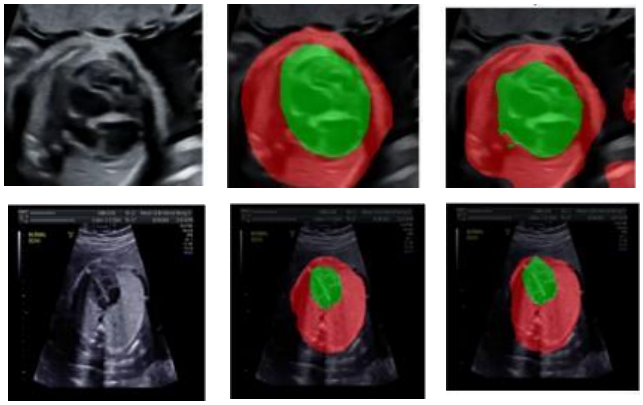
Pada tahap evaluasi, hasil segmentasi yang diperoleh dari model berbasis Transformer, SegFormer-B3, dianalisis untuk menilai tingkat akurasi terhadap citra ultrasonografi jantung janin yang telah diberi label. Citra *ground truth* yang disegmentasi secara manual digunakan sebagai acuan standar dalam proses validasi. Hasil prediksi model dikonversi menjadi citra *biner* dengan dimensi yang sama guna memudahkan proses perbandingan dengan *ground truth*. Kinerja model divalidasi menggunakan data pengujian melalui penghitungan metrik kuantitatif, yaitu *Dice Coefficient* dan *IoU*. Perhitungan nilai metrik ini dilakukan berdasarkan persamaan (5) dan (7), yang menggambarkan tingkat kesesuaian antara hasil prediksi dan citra *ground truth*.

Tabel III. Hasil Evaluasi

	IoU	*Dice Coefficient*
HC	0.6859	0.8137
CC	0.5834	0.7369

Ilustrasi dan hasil segmentasi yang diperoleh menggunakan metode SegFormer disajikan pada Gambar 7 dan Tabel II.





(a) (b) (c)

Gambar 7. (a) *Original Citra* (b) *Ground Truth* (c) Prediksi (*SegFormer-B3*)

IV. DISKUSI

Penelitian ini telah mengimplementasikan arsitektur SegFormer-B3 untuk tugas segmentasi pada citra ultrasonografi jantung janin yang terindikasi kardiomegali. Model dilatih menggunakan data yang telah diberi label untuk mengenali area HC dan CC. Hasil evaluasi menunjukkan bahwa model memperoleh performa optimum pada kelas HC, model mencapai *IoU* sebesar 0.6859 dan *Dice Coefficient* sebesar 0.8137, sedangkan pada kelas CC nilai yang diperoleh masing-masing 0.5834 dan 0.7369. Nilai tersebut menunjukkan bahwa model memberikan kinerja segmentasi yang memadai, meskipun masih terdapat tantangan pada area jantung akibat kontras citra yang rendah.

V. KESIMPULAN

Kesimpulannya, penelitian ini mengusulkan penggunaan arsitektur SegFormer-B3, model pembelajaran dalam berbasis Transformer, untuk melakukan segmentasi citra ultrasonografi jantung janin dengan indikasi kardiomegali. Arsitektur ini mampu mempelajari fitur lokal dan global melalui mekanisme *self-attention*, sehingga menghasilkan segmentasi yang cukup konsisten dan representatif pada area anatomi jantung janin.. Pendekatan ini berpotensi mendukung analisis morfologi awal dalam skrining kardiomegali prenatal. Penelitian selanjutnya akan difokuskan pada peningkatan akurasi dan kemampuan generalisasi model melalui penyesuaian *hyperparameter*, perluasan data pelatihan, serta eksplorasi varian arsitektur Transformer lainnya.

REFERENCES

[1] S. An *et al.*, "Computerized Medical Imaging and Graphics A category attention instance segmentation network for four cardiac chambers segmentation in fetal echocardiography," *Comput. Med. Imaging Graph.*, vol. 93, no. September, p. 101983, 2021, doi: 10.1016/j.compmedimag.2021.101983.

[2] X. Liu, L. Song, S. Liu, and Y. Zhang, "A Review of Deep-Learning-Based Medical Image Segmentation Methods," 2021.

[3] S. S. Sarpotdar, "Cardiomegaly Detection using Deep Convolutional Neural Network with U-Net".

[4] S. Nurmaini, A. I. Sapitri, and M. T. Roseno, "iScience," *ISCIENCE*, vol. 28, no. 5, p. 112288, doi: 10.1016/j.isci.2025.112288.

[5] M. N. Rachmatullah, S. Nurmaini, A. I. Sapitri, A. Darmawahyuni, and B. Tutuko, "Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view," vol. 10, no. 4, pp. 1987–1996, 2021, doi: 10.11591/eei.v10i4.3060.

[6] B. Tutuko, F. Firdaus, R. U. Partan, and N. Bernolian, "Congenital Heart Defects Detection," 2021.

[7] H. G. A. Van Der Pol, L. M. Van Karnenbeek, M. Wijkhuizen, F. Geldof, and B. Dashtbozorg, "applied sciences Deep Learning for Point-of-Care Ultrasound Image Quality Enhancement : A Review," 2024.

[8] F. Mris, "Deep Learning-Based Multiclass Brain Tissue Segmentation in," 2023.

[9] S. Qiao *et al.*, "DPC-MSGATNet : dual-path chain multi-scale gated axial-transformer network for four-chamber view segmentation in fetal echocardiography," *Complex Intell. Syst.*, vol. 9, no. 4, pp. 4503–4519, 2023, doi: 10.1007/s40747-023-00968-x.

[10] A. Melese, B. Enyew, and Y. Agegnehu, "Intelligent Systems with Applications Early-stage cardiomegaly detection and classification from X-ray images using convolutional neural networks and transfer learning," *Intell. Syst. with Appl.*, vol. 24, no. October, p. 200453, 2024, doi: 10.1016/j.iswa.2024.200453.

[11] H. Wu, G. Qu, Z. Xiao, and F. Chunyu, "Heliyon Enhancing left ventricular segmentation in echocardiography with a modified mixed attention mechanism in SegFormer architecture," *Heliyon*, vol. 10, no. 15, p. e34845, 2024, doi: 10.1016/j.heliyon.2024.e34845.

[12] H. Allaoui, "ScienceDirect Fine-tuned Fine-tuned SegFormer SegFormer for for enhanced enhanced fetal fetal head head segmentation segmentation," *Procedia Comput. Sci.*, vol. 251, no. 2018, pp. 350–357, 2024, doi: 10.1016/j.procs.2024.11.120.

[13] Q. He, Q. Yang, and M. Xie, "HCTNet : A hybrid CNN-transformer network for breast ultrasound image segmentation," *Comput. Biol. Med.*, vol. 155, no. January, p. 106629, 2023, doi: 10.1016/j.compbiomed.2023.106629.

[14] M. Liao *et al.*, "Left Ventricle Segmentation in Echocardiography with Transformer," pp. 1–14, 2023.

[15] K. Curran, "BACKBONE FOR SEGMENTING PUBIC SYMPHYSIS - FETAL HEAD," no. Dl, 2018.

[16] P. Cai, L. Jiang, Y. Li, and L. Lan, "Pubic Symphysis-Fetal Head Segmentation Using Pure

Transformer with Bi-level Routing Attention".

[17] E. Jain, P. Kaushik, V. Kukreja, A. Dogra, and B. Goyal, "Fetal Diagnostics using Vision Transformer for Enhanced Health and Severity Abstract :," pp. 1–30, 2025, doi: 10.2174/0115734056360199250227053012.

[18] A. K. Mondal, A. Bhattacharjee, and P. Singla, "xViTCOS : Explainable Vision Transformer Based COVID-19 Screening Using Radiography," vol. 10, no. December 2021, 2022.

[19] L. Zhang, W. Huang, and B. Fan, "Neurocomputing SARFormer : Segmenting Anything Guided Transformer for semantic segmentation," *Neurocomputing*, vol. 635, no. February, p. 129915, 2025, doi: 10.1016/j.neucom.2025.129915.

[20] X. Zhu, X. Dong, W. Yu, H. Liang, and B. Kong, "Refactored Maskformer : Refactor localization and classification for improved universal image segmentation ☆," vol. 87, no. February, 2025.

[21] X. Zhang and X. Wang, "Domain-Adaptive Organ Segmentation Architecture in Clinical Imaging through SegFormer," no. 7, 2025.

[22] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer : Simple and Efficient Design for Semantic Segmentation with Transformers," no. NeurIPS, pp. 1–14, 2021.

[23] H. Zhang, S. Wang, and Y. Zhang, "ScienceDirect A Dual-Path Multi-Scale Feature Feature Fusion Fusion Decoder Decoder for for SegFormer SegFormer," *Procedia Comput. Sci.*, vol. 222, pp. 147–156, 2023, doi: 10.1016/j.procs.2023.08.152.

[24] D. Made, S. Arsa, T. Ilyas, S. Park, L. Chua, and H. Kim, "Computerized Medical Imaging and Graphics Efficient multi-stage feedback attention for diverse lesion in cancer image segmentation," *Comput. Med. Imaging Graph.*, vol. 116, no. August 2023, p. 102417, 2024, doi: 10.1016/j.compmedimag.2024.102417.

[25] R. Rekha, P. Shruti, M. Deekshitha, and J. Akash, "Biomedical Signal Processing and Control DCSwin-UNet : Dual Encoder U-Net based on CNN and Swin Transformer with Trainable Multiplication Layer for brain tumor segmentation from MRI images," *Biomed. Signal Process. Control*, vol. 110, no. PA, p. 108325, 2025, doi: 10.1016/j.bspc.2025.108325.