

Ghost U-Net dan EfficientNetV2: Segmentasi dan Klasifikasi Citra Retina Untuk Pendeteksian Glaukoma

Willy Wijaya¹, Dhimas Fauzan Al Asri², Hergi Alfizar³, Fauzan Asrin⁴

¹Fakultas Teknik Universitas Tanjungpura, Pontianak, 78124 Indonesia, D1041221028@student.untan.ac.id

²Fakultas Teknik Universitas Tanjungpura, Pontianak, 78124 Indonesia, D1041221036@student.untan.ac.id

³Fakultas Teknik Universitas Tanjungpura, Pontianak, 78124 Indonesia, D1041221018@student.untan.ac.id

⁴Fakultas Teknik Universitas Tanjungpura, Pontianak, 78124 Indonesia, asrin@informatika.untan.ac.id

Corresponding Author: Willy Wijaya

INTISARI — Glaukoma, kondisi mata yang menyerang saraf optik secara diam-diam, merupakan penyebab kebutaan permanen kedua di dunia. Diagnosis dan pemantauan glaukoma yang rumit dan memakan waktu seringkali menjadi penghambat dalam upaya penanggulangan penyakit ini. Namun, penelitian ini menghadirkan solusi cerdas berbasis *deep learning* yang menghapus kendala tersebut dengan elegan. Mengombinasikan kehebatan *Ghost U-Net* untuk segmentasi citra retina glaukoma dan *EfficientNetV2B2* untuk klasifikasi, pendekatan ini memungkinkan proses deteksi yang akurat dengan beban komputasi yang sangat ringan. *Ghost U-Net* mengekstraksi fitur semantik retina, sementara *EfficientNetV2B2* mengklasifikasikannya dengan presisi tinggi. Keunggulan komputasi rendah membuat solusi cerdas ini sangat menjanjikan untuk diterapkan di lingkungan dengan keterbatasan sumber daya, seperti klinik atau pusat layanan kesehatan di daerah terpencil. Dengan kemampuannya mendeteksi glaukoma pada tahap awal, solusi ini berpotensi besar untuk membantu mencegah kebutaan permanen dan mengubah nasib jutaan penderita glaukoma di seluruh dunia. Tidak lagi terhalang oleh kendala teknis dan komputasi berat, dunia kedokteran dapat memanfaatkan kecerdasan buatan untuk melakukan deteksi glaukoma. Hasil uji dari dataset *ORIGA*, *REFUGE*, dan *G1020* dengan total 2.870 gambar retina, *Ghost U-Net* dengan parameter sekitar 1,8 juta mencapai akurasi segmentasi sebesar 90,78%, sensitivitas 90,70%, dan *dice coefficient* sebesar 86,23%. Sementara itu dalam tugas klasifikasi *EfficientNetV2B2* dengan jumlah parameter sekitar 8,9 juta mencapai akurasi *training* sebesar 91,60% dan akurasi *test* sebesar 92,75%. Hasil penelitian menunjukkan keseimbangan antara kedua model dengan parameter yang lebih kecil, tetapi tetap mempertahankan akurasi dengan baik

KATA KUNCI — Glaukoma, *Deep Learning*, *Ghost U-Net*, *EfficientNetV2*, Segmentasi, Klasifikasi

I. PENDAHULUAN

Glaukoma adalah suatu kondisi di mana terjadi kerusakan saraf mata dikarenakan meningkatnya tekanan pada bola mata [1]. Diperkirakan ada 60,5 juta penderita glaukoma di dunia, dan jumlah ini diprediksi akan bertambah menjadi 79,6 juta. Di Asia, hampir setengah dari kasus glaukoma ditemukan, dengan sebagian besar berupa Glaukoma Primer Sudut Tertutup (GPSTp). Sekitar 4,5 juta orang menjadi buta akibat Glaukoma Primer Sudut Terbuka (GPSTa) dan 3,9 juta lainnya karena GPSTp di tahun 2010. Diperkirakan, jumlah ini akan meningkat menjadi 5,9 juta untuk GPSTa dan 5,3 juta untuk GPSTp pada tahun 2020.. Mengembangkan perangkat otomatis untuk deteksi dini glaukoma sangat penting karena dampak positifnya bagi umat manusia dan perekonomian global [3].

Dalam segmentasi citra untuk analisis *Optic Cup* (OC) dan *Disk Cup* (DC), metode seperti *U-Net* [4], *DeepLabV3* [5], dan *Mask R-CNN* [6] telah digunakan. Namun, jumlah parameter yang tinggi dapat menghambat performa karena keterbatasan sumber daya komputasi. Untuk mengatasi ini, kami mengusulkan *Ghost U-Net*, arsitektur *deep learning* yang efisien dan ringan, serta *EfficientNetV2-B2*, metode *deep learning* yang dirancang untuk performa efisien..

Dalam penelitian ini, kami menggunakan *Ghost-UNet*, arsitektur *encoder-decoder* asimetris untuk segmentasi semantik akurat. Model ini menggabungkan *Ghost-Net* dan *U-Net*, memungkinkan segmentasi dengan jumlah parameter yang lebih sedikit [7]. Untuk tugas klasifikasi, kami menggunakan *EfficientNetV2* berbasis arsitektur CNN yang efektif. Arsitektur ini menskalakan lebar, resolusi, dan kedalaman secara konstan tanpa mengorbankan efisiensi model [8].

Berbeda dengan penelitian sebelumnya, studi ini fokus pada optimalisasi efisiensi komputasi tanpa mengorbankan akurasi

dalam segmentasi dan klasifikasi citra retina untuk deteksi glaukoma. Sementara penelitian terdahulu mencapai akurasi tinggi, tetapi dengan arsitektur jaringan yang kompleks dan memerlukan sumber daya besar, pendekatan kami yaitu merancang model lebih kecil dengan kinerja setara atau lebih baik dari model kompleks tersebut.

Tujuan penelitian ini adalah mengembangkan metode efisien dan akurat untuk segmentasi dan klasifikasi citra retina glaukoma. Kami menggunakan *Ghost-UNet* karena memiliki sedikit parameter dengan akurasi tinggi, memudahkan implementasi pada infrastruktur sederhana. Untuk klasifikasi, kami memilih *EfficientNet* karena mampu menskalakan dimensi tanpa mengorbankan efisiensi, menawarkan akurasi lebih baik dibandingkan model seperti *VGG-16* atau *ResNet50* [8]. Kombinasi *Ghost-UNet* dan *EfficientNet* memungkinkan segmentasi dan klasifikasi akurat dengan beban komputasi lebih ringan, memfasilitasi penerapan di lingkungan dengan keterbatasan sumber daya.

II. METODOLOGI

Penelitian ini memiliki lima tahap metodologi penelitian, yaitu mengumpulkan dan melakukan *preprocess* dataset, membuat model segmentasi, membuat model klasifikasi, melatih neural network, dan mengevaluasi model.

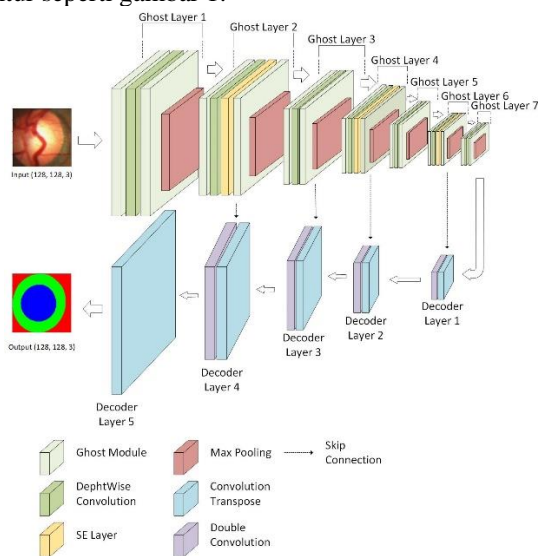
A. DATASET

Penelitian menggunakan dataset glaukoma dari *Kaggle*, <https://www.kaggle.com/datasets/arnavjain1/glaucoma-datasets> terdiri dari *ORIGA(-light)* (650 gambar retina, anotasi *Singapore Eye Research Institute* untuk segmentasi cakram optik dan cawan), *REFUGE* (1.200 gambar retina dari *Retinal Fundus Glaucoma Challenge*), dan *G1020* (1.020 gambar fundus

beresolusi tinggi, anotasi untuk diagnosis glaukoma). Semua gambar diubah ukuran menjadi 128 x 128 piksel.

B. METODE SEGMENTASI GHOST U-NET

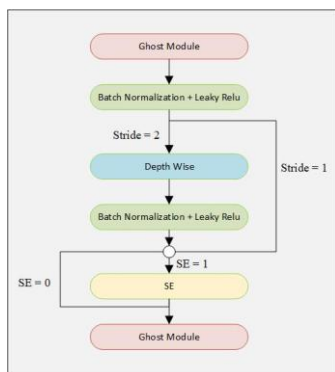
Ghost U-net [7] adalah varian dari arsitektur *U-Net* yang menggunakan *Ghost Module* untuk meningkatkan efisiensi komputasi tanpa mengorbankan akurasi. Arsitektur ini membagi *channels* menjadi *sub-channels* kecil yang diproses secara paralel sebelum digabungkan kembali. Terdiri dari tujuh blok *encoder* dan lima blok *decoder* yang asimetris. Ilustrasi arsitektur seperti gambar 1.



Gambar 1. Arsitektur Ghost U-Net

1) GHOST LAYER

Jantung dari kecerdikan *encoder Ghost U-Net* adalah *Ghost Layer*, sebuah lapisan yang menerapkan konsep "mengerjakan lebih banyak dengan lebih sedikit" dalam arsitektur *deep learning*. Lapisan ini terdiri dari rangkaian *Ghost Module*, *DepthWise Convolution*, *Batch Normalization* + *Leaky ReLU*, dan *SE Layer*.

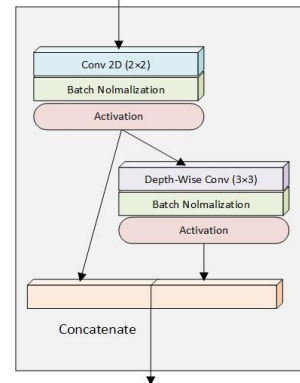


Gambar 2. Ghost Layer

Dari Gambar 2, Arsitektur ini terdiri dari dua modul *Ghost Module* di awal dan akhir, serta lapisan-lapisan *Batch Normalization*, *Depth-Wise*, dan *SE*. *Stride* yang digunakan dapat bernilai 1 atau 2, dan *SE* dapat bernilai 0 atau 1. Lapisan-lapisan ini digunakan untuk menghasilkan berbagai peta fitur. Semua lapisan kecuali *Ghost Module* terakhir menggunakan *Batch Normalization* dan *Leaky ReLU* untuk efisiensi yang lebih baik.

2) GHOST MODULE

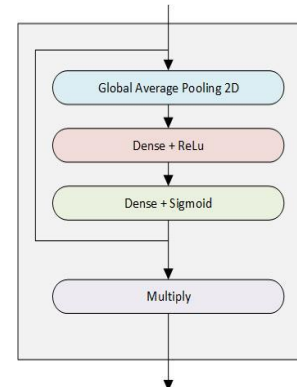
Ghost module [9] pada *Ghost U-Net* menggunakan tiga tahap: *Primary Convolution* (2D convolution), *cheap operation* (depth-wise convolution), dan *concatenation* untuk menggabungkan hasil dari kedua tahap sebelumnya. Setiap tahap melibatkan *Batch Normalization* dan *Activation function*. Gambaran arsitekturnya terlihat pada Gambar 3.



Gambar 3. Ghost Module

3) SE LAYER

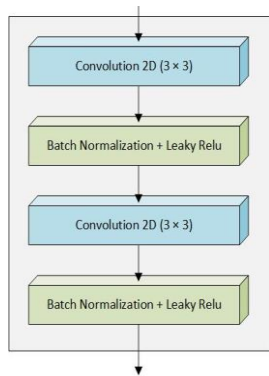
Ghost U-Net menggunakan *SE Layer*[10] dalam arsitekturnya. *Ghost U-Net* menggunakan *global average pooling 2D* untuk mengompres informasi dari setiap saluran menjadi statistik global pada tahap *squeeze*. Selanjutnya, pada tahap *expand*, terdapat dua *fully connected layer*: satu dengan aktivasi *ReLU* untuk memperluas dimensi, dan satu lagi dengan aktivasi *sigmoid* untuk menghasilkan vektor kalibrasi. Vektor ini digunakan untuk menyesuaikan perhatian pada setiap saluran fitur asli. Dapat dilihat pada gambar 4.



Gambar 4. SE Layer

4) DOUBLE CONVOLUTION BLOCK

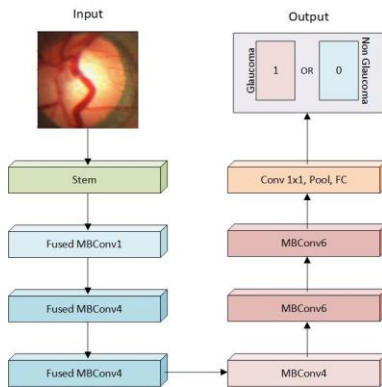
Ghost U-Net menggunakan blok *double convolution* [11] yang terdiri dari dua lapisan konvolusi dengan ukuran (3x3). Setiap lapisan konvolusi ini diikuti oleh *batch normalization* dan aktivasi *Leaky ReLU*. Hal ini memastikan bahwa fitur-fitur yang dipulihkan dalam jalur *decoder* memiliki representasi yang kuat dan mengurangi risiko *overfitting*. Arsitekturnya dapat dilihat pada gambar 5



Gambar 5. Double Convolution

C. METODE KLASIFIKASI EFFICIENT-NETV2

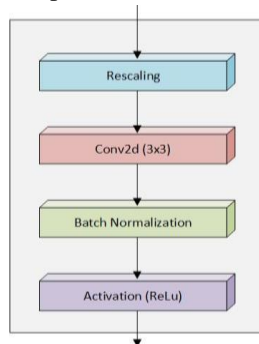
EfficientNetV2[12] dengan varian *EfficientNetV2-B2* yang digunakan dalam penelitian ini, merupakan kemajuan dalam klasifikasi gambar dibandingkan dengan *EfficientNetV1* [13]. *FusedMBConv* adalah inovasi utama yang mengurangi redundansi komputasi untuk mempercepat pelatihan dan efisiensi parameter. *EfficientNetV2-B2* menawarkan keseimbangan yang baik antara kedalaman jaringan dan kompleksitas parameter, cocok untuk tugas klasifikasi gambar dengan modifikasi untuk meningkatkan kinerja dan efisiensi dari *EfficientNetV1-B2*. Arsitektur dapat dilihat pada gambar 6



Gambar 6. Arsitektur EfficientNetV2-B2

1) STEM BLOCK

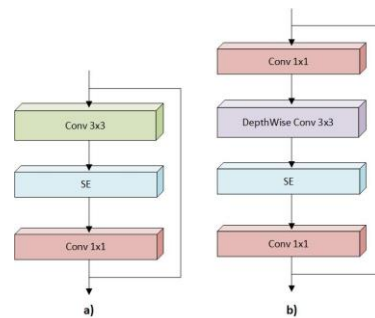
Stem block merupakan lapisan fundamental paling awal dalam *EfficientNetV2-B2*. Lapisan ini memiliki tujuan penting untuk mengekstraksi fitur awal dari *input*. Lapisan stem terdiri dari beberapa lapisan yang saling terkait, yaitu dimulai dengan *rescaling*, *Conv2d* (3x3), *BatchNormalization*, dan *Activation*. Arsitektur dapat dilihat pada Gambar 7.



Gambar 7. Stem Block

2) MBCONV DAN FUSED-MBCONV

MBconv atau *Mobile Inverted Bottleneck Convolution*, dipekenalkan dalam *EfficientNetV1* sebagai modul dengan parameter yang efisien. Blok ini bertujuan untuk meningkatkan kapasitas representasi sambil mengurangi kompleksitas komputasi. *MBConv* terdiri dari empat lapisan yaitu dimulai dengan *pointwise convolution*, *depth wise convolution*, *SE layer*, dan diakhiri lagi dengan *pointwise convolution*. *Fused MBConv* yang dipekenalkan pada *EfficientNetV2* merupakan evolusi dari *MBConv*. Blok ini bertujuan untuk meningkatkan efisiensi jaringan dengan mengurangi redundansi komputasi. *FusedMBConv* mencapai hal ini dengan menggabungkan beberapa lapisan, yaitu dimulai dengan lapisan *Conv* (3x3), *SE layer*, dan diakhiri dengan *pointwise convolution*. Seperti pada gambar 8



Gambar 8. a) Fused MBConv b) MBConv

III. HASIL DAN DISKUSI

A. MODEL SEGMENTASI GHOST U-NET

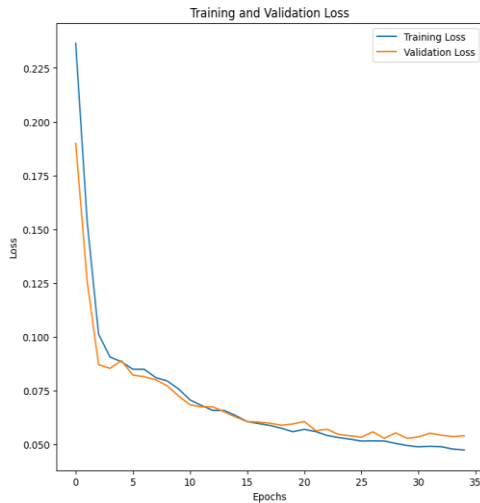
Ghost U-Net memiliki sekitar 1.8 juta parameter dan dilatih dengan 1.552 data train dan 518 data validasi menggunakan optimizer Adam (learning rate 0.001) dengan Dice loss function. Proses pelatihan dipantau dengan validation loss dan metrik seperti *Accuracy*, *Val-Dice-Coef*, dan *sensitivity* digunakan untuk evaluasi kinerja model. Hyperparameter yang digunakan dapat dilihat pada tabel 1

Tabel I

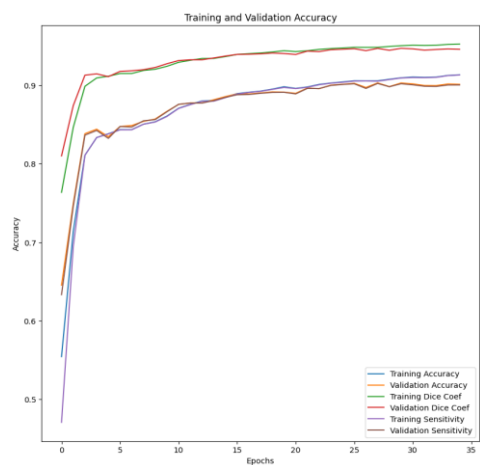
Hyperparameter Ghost-U-Net

Hyper Parameter	Nilai
<i>Batch Size</i>	32
<i>Learning Rate</i>	0.001
<i>Jumlah Epoch</i>	40
<i>Optimizer</i>	<i>Adam</i>
<i>Loss Function</i>	<i>Dice Loss</i>

Proses *training* mencapai titik optimal pada epoch ke 32, seperti yang ditunjukkan pada grafik *loss* untuk *data train* dan *data test* (Gambar 9). Hasil evaluasi model dengan metrik *accuracy*, *val-dice-coef*, dan *sensitivity*, *Ghost U-Net* mendapat hasil akurasi sebesar 90.78%, dengan *sensitivity* sebesar 90.70% dan *val-dice-coef* sebesar 86.23%.



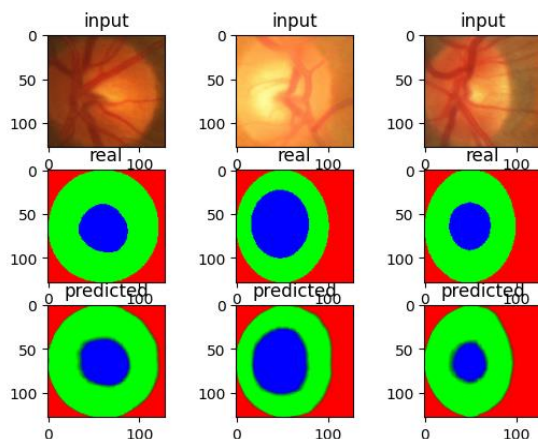
(a)



(b)

Gambar 9. Hasil training loss dan validation loss a), hasil training accuracy dan validation accuracy b).

Lebih lanjut, contoh visual hasil prediksi menunjukkan kemampuan model yang dengan cukup baik mensegmentasikan *optic cup* dan *disc cup* pada citra retina. Seperti yang dapat dilihat pada gambar 10, hasil segmentasi sangat cocok dengan anotasi *ground truth*, yang menunjukkan efektivitas model dalam menangkap struktur retina.



Gambar 10. Hasil Prediksi dibandingkan dengan *ground truth image*

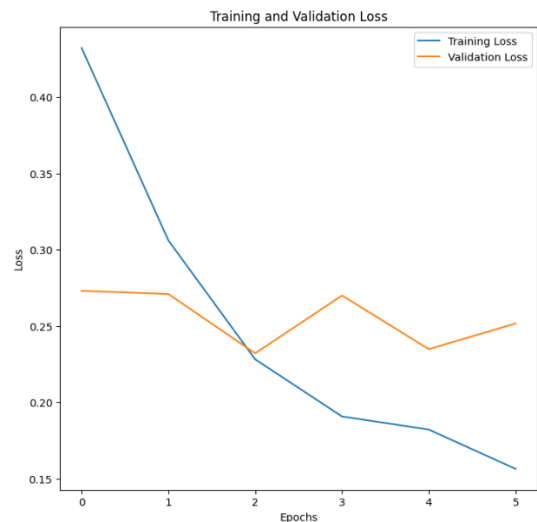
B. MODEL KLASIFIKASI EFFICIENTNETV2-B2

Dalam penelitian ini, tugas klasifikasi pada citra retina dilakukan dengan metode *EfficientNetV2B2*. Model ini memiliki 8,9 juta parameter dan dilatih dengan data yang diproses dari model segmentasi *Ghost U-net*. Kemudian data dibagi menjadi set pelatihan dan pengujian dengan proporsi jumlah data latih 1.485 dan data tes 371 dengan digunakannya *optimizer Adam* dengan learning rate 0,001. *Binary crossentropy loss function* digunakan untuk menghitung nilai loss pada model. Untuk mencegah *overfitting*, proses *training* dimonitor dengan *early stopping* menggunakan *validation loss*. Metrik *accuracy* digunakan untuk mengukur kinerja model *EfficientNetV2B2* ini.

Tabel II
Hyperparameter *EfficientNetV2-B2*

Hyper Parameter	Nilai
<i>Batch Size</i>	32
<i>Learning Rate</i>	0.001
<i>Jumlah Epoch</i>	6
<i>Optimizer</i>	<i>Adam</i>
<i>Loss Function</i>	<i>Binary crossentropy</i>

Proses *training* mencapai titik optimal pada *epoch* ke 3, seperti yang ditunjukkan pada grafik *loss* untuk *data train* dan *data validation*. Hasil evaluasi model dengan metrik *accuracy*, *EfficientNetV2-B2* mendapat hasil akurasi train sebesar 91,60% dan akurasi test sebesar 92.75%. grafik akurasi dan loss dapat dilihat pada gambar 11.



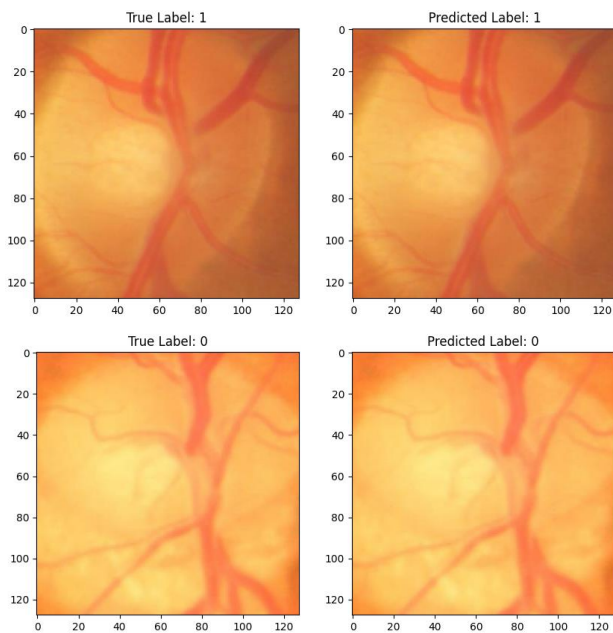
(a)



(b)

Gambar 11 Hasil training loss dan validation loss a), hasil training accuracy dan validation accuracy b).

Berikut adalah contoh visual dari hasil klasifikasi *EfficientNetV2B2* menunjukkan kemampuan yang baik dari model dalam klasifikasi, yang dapat dilihat dengan kecocokan model hasil prediksi model dengan *ground truth*. contoh dapat dilihat pada gambar 12



Gambar 12 Hasil Prediksi *EfficientNetV2-B2*

C. ANALISIS HASIL DAN MANFAAT

Penelitian ini menunjukkan *Ghost U-Net* dengan parameter yang jauh lebih kecil dibanding metode lainnya dapat mencapai akurasi segmentasi citra retina sebesar 90.78%. Model ini menggunakan modul *Ghost* untuk mengurangi redundansi parameter dan mempercepat pelatihan, ideal untuk analisis detail dan segmentasi struktur mikroskopis seperti *optic cup* (OC) dan *disk cup* (DC) pada gambar retina.

Untuk tugas klasifikasi gambar retina, *EfficientNetV2-B2* dapat mencapai akurasi 92.75% pada data uji. Model ini memanfaatkan *FusedMBConv* untuk efisiensi parameter dan waktu pelatihan optimal. Kombinasi *Ghost U-Net* untuk

segmentasi mendalam dan *EfficientNetV2-B2* untuk klasifikasi menawarkan solusi efisien untuk deteksi dini glaukoma di lingkungan rumah sakit dengan sumber daya komputasi terbatas. Langkah berikutnya dapat fokus pada integrasi dua model ini untuk meningkatkan pelayanan deteksi glaukoma dan adaptasi terhadap variasi gambar retina yang kompleks.

IV. KESIMPULAN

Penelitian kami mengadopsi strategi seimbang dengan merancang model berparameter lebih kecil namun tetap mempertahankan atau bahkan meningkatkan performa dibandingkan dengan model kompleks. Kami mengakui keterbatasan dalam studi ini terkait ukuran dan keragaman dataset, yang dapat ditingkatkan untuk meningkatkan generalisasi model. Kami juga menyarankan eksplorasi lebih lanjut terhadap penyetelan *hyperparameter* dan teknik augmentasi data untuk meningkatkan kinerja segmentasi dan klasifikasi. Langkah selanjutnya akan fokus pada memperluas dataset dan memanfaatkan pembelajaran transfer dari model yang sudah dilatih, dengan tujuan meningkatkan deteksi dini glaukoma melalui pengembangan model deep learning yang efisien.

REFERENSI

- [1] W. R. Perdani, R. Magdalena, and N. K. Caecar Pratiwi, "Deep Learning untuk Klasifikasi Glaukoma dengan menggunakan Arsitektur EfficientNet," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 10, no. 2, p. 322, Apr. 2022, doi: 10.26760/elkomika.v10i2.322.
- [2] "Keputusan Menteri Kesehatan Republik Indonesia Nomor hk.01.07/menkes/1488/2023 Tentang Pedoman Nasional Pelayanan Kedokteran Tata Laksana Glaukoma."
- [3] A. Sarhan, J. Rokne, and R. Alhaji, "Glaucoma detection using image processing techniques: A literature review," *Comput. Med. Imaging Graph.*, vol. 78, p. 101657, Dec. 2019, doi: 10.1016/j.compmedimag.2019.101657.
- [4] M. B. Sudhan *et al.*, "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model," *J. Healthc. Eng.*, vol. 2022, pp. 1–10, Feb. 2022, doi: 10.1155/2022/1601354.
- [5] S. Sreng, N. Maneerat, K. Hamamoto, and K. Y. Win, "Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images," *Appl. Sci.*, vol. 10, no. 14, p. 4916, Jul. 2020, doi: 10.3390/app10144916.
- [6] J. Kim, L. Tran, T. Peto, and E. Y. Chew, "Identifying Those at Risk of Glaucoma: A Deep Learning Approach for Optic Disc and Cup Segmentation and Their Boundary Analysis," *Diagnostics*, vol. 12, no. 5, p. 1063, Apr. 2022, doi: 10.3390/diagnostics12051063.
- [7] I. A. Kazerouni, G. Dooley, and D. Toal, "Ghost-UNet: An Asymmetric Encoder-Decoder Architecture for Semantic Segmentation From Scratch," *IEEE Access*, vol. 9, pp. 97457–97465, 2021, doi: 10.1109/ACCESS.2021.3094925.
- [8] S. Tripathy, R. Singh, and M. Ray, "Automation of Brain Tumor Identification using EfficientNet on Magnetic Resonance Images," *Procedia Comput. Sci.*, vol. 218, pp. 1551–1560, 2023, doi: 10.1016/j.procs.2023.01.133.
- [9] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features from Cheap Operations." arXiv, 2019, doi: 10.48550/ARXIV.1911.11907.
- [10] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," 2017, doi: 10.48550/ARXIV.1709.01507.
- [11] W. Liu, F. Su, X. Jin, H. Li, and R. Qin, "Bispace Domain Adaptation Network for Remotely Sensed Semantic Segmentation," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–11, 2020, doi: 10.1109/TGRS.2020.3035561.
- [12] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," 2021, doi: 10.48550/ARXIV.2104.00298.
- [13] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2019, doi: 10.48550/ARXIV.1905.11946.