

Segmentasi dan Klasifikasi Sel pada Citra Histologi dengan Menggunakan Jaringan Konvolusional Encoder-Decoder

Laura Husain Olii¹, Novie Theresia Br. Pasaribu², Meilan Jimmy Hasugian³, Audyati Gany⁴

^{1,2,3,4} Program Studi Teknik Elektro Universitas Kristen Maranatha
Jl. Surya Sumantri No.65 Bandung 40164 INDONESIA

1822025@eng.maranatha.edu¹, novie.theresia@eng.maranatha.edu²,
jhasugian@maranatha.edu³, audyati.gany@eng.maranatha.edu⁴

Intisari—Analisis citra histologi merupakan hal yang esensial dalam penelitian medis untuk membantu diagnosis penyakit. Sel pada citra histologi dapat memiliki variasi bentuk dan ukuran yang beragam. Segmentasi dan klasifikasi sel pada citra histologi adalah langkah awal yang penting untuk dilakukan dalam analisis citra histologi untuk mengetahui kondisi dari sel pada jaringan agar dapat dilakukan analisis lebih lanjut. Dataset yang digunakan pada penelitian ini yaitu dataset Lizard yang disediakan secara publik pada kompetisi Conic 2022. Model jaringan konvolusional encoder-decoder arsitektur LinkNet dengan modifikasi pada bagian encoder dibuat untuk melakukan segmentasi dan klasifikasi sel pada citra histologi menjadi enam kelas yaitu sel neutrofil, sel epitel, sel limfosit, sel plasma, sel eosinofil, dan jaringan ikat. Model pretrained Convolutional Neural Network yang digunakan untuk menjadi encoder LinkNet adalah VGG16, DenseNet121, dan EfficientNet-B2. Berdasarkan hasil pengujian dengan menggunakan metric Mean Intersection over Union (mIoU) pada arsitektur LinkNet dengan masing-masing arsitektur encoder didapatkan hasil performansi terbaik dengan menggunakan arsitektur LinkNet dengan encoder DenseNet121 sebesar 0.5791 pada kelas sel epitel, 0.3935 pada kelas sel limfosit, 0.0444 pada kelas sel plasma, dan 0.4342 pada kelas jaringan ikat.

Keywords—Segmentasi, Klasifikasi, Jaringan Konvolusional Encoder-Decoder

I. PENDAHULUAN

Sel merupakan unit struktural dan fungsional terkecil pada organisme makhluk hidup [1]. Setiap komponen dalam suatu jaringan seperti sel, kerapatan sel, dan morfologi sel merupakan informasi yang dapat digunakan untuk menganalisis ciri atau karakteristik dari suatu penyakit. Penilaian terhadap jaringan dan morfologi sel dapat dilakukan secara manual oleh ahli patologi melalui citra histologi dengan pewarna Haematoxylin & Eosin (*HE stained images*) [2]. Penilaian citra histologi secara manual masih banyak dilakukan hingga saat ini namun subjektivitas dari ahli patologi dapat memengaruhi hasil analisis. Pada analisis citra histologi untuk mengetahui fitur dari sebuah sel untuk analisis lebih lanjut dilakukan segmentasi sel yang merupakan langkah awal dalam analisis citra histologi.

Hal yang menjadi hambatan dalam melakukan segmentasi sel yaitu sel memiliki variasi yang sangat beragam salah

satunya dilihat dari bentuk dan ukuran yang berbeda pada jenis sel, jenis penyakit, dan area yang berbeda meskipun dalam satu *sample* jaringan yang sama. *Deep Learning* merupakan salah satu metode yang dapat digunakan untuk melakukan segmentasi dan klasifikasi yang berkaitan dengan analisis citra medis. *Convolutional Neural Networks* (CNN) atau jaringan konvolusional merupakan model *deep learning* yang banyak diaplikasikan pada bidang *computer vision*. Arsitektur pada model *deep learning* yang banyak digunakan untuk melakukan segmentasi semantik atau klasifikasi per-piksel yaitu jaringan konvolusional *encoder-decoder*. Jaringan *encoder* digunakan untuk menghasilkan *feature maps* tetapi masih dengan resolusi yang rendah [3]. Jaringan *decoder* digunakan untuk mengembalikan *feature maps* dengan resolusi rendah dari *encoder* ke resolusi input untuk kemudian dilakukan klasifikasi setiap piksel.

Suresh Kallam, dkk melakukan segmentasi dan klasifikasi sel citra histologi dari dataset Science Bowl 2018 menggunakan arsitektur LinkNet dengan nilai akurasi validasi 97.2 % dan nilai IoU 89.8% [4]. Pada penelitian ini, permasalahan segmentasi dan klasifikasi jenis sel pada citra histologi dilakukan dengan mengeksplorasi jaringan konvolusional *encoder-decoder* arsitektur LinkNet [5] serta memodifikasi *encoder* menggunakan *pre-trained* model yaitu VGG16, DenseNet-121, dan EfficientNet-B2.

II. MATERIAL DAN METODE







A. Dataset

Citra histologi dengan pewarna Haematoxykin & Eosin (HE) yang digunakan dalam penelitian ini diambil dari kompetisi Conic 2022 Challenge [6] yang berasal dari *Lizard dataset*. Citra histologi HE ini memiliki perbesaran 20x (~0,5 mikron / piksel) dari 6 sumber data yang berbeda.

Pada dataset ini, beberapa jenis inti sel yang akan disegmentasi dan diklasifikasikan terdiri dari: sel neutrofil, sel epitel, sel limfosit, sel plasma, sel eosinofil, dan jaringan ikat (*connective tissue*). Sebagai gambaran, contoh dari bentuk beberapa sel ini disajikan dalam Tabel I.

Citra histologi yang diperoleh memiliki ukuran 256x256 piksel. Dan secara total jumlah sel yang terdapat dari keseluruhan 4981 citra dari dataset disajikan pada Tabel II.

TABEL I
JENIS SEL

Gambar	Nama Sel	Keterangan
	Neutrofil	Sel darah putih yang melawan infeksi sekaligus melindungi dari berbagai ancaman penyakit.
	Epitel	Sel yang menutupi dan melindungi permukaan organ.
	Limfosit	Sel darah putih yang menghasilkan antibodi.
	Plasma	Sel darah putih yang menghasilkan antibodi.
	Eosinofil	Sel darah putih yang melawan parasit dan infeksi.
	Jaringan Ikat (<i>Connective</i>)	Sekumpulan sel yang melindungi dan memberikan struktur pada jaringan dan organ, menyimpan lemak, dan memperbaiki kerusakan pada jaringan.

TABEL II
DISTRIBUSI SEL PADA DATASET LIZARD

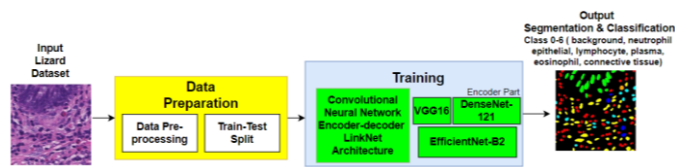
Kelas	Jenis Sel	Jumlah Sel
Kelas 1	Neutrofil	4012
Kelas 2	Epitel	221017
Kelas 3	Limfosit	93612
Kelas 4	Plasma	24793
Kelas 5	Eosinofil	2999
Kelas 6	Jaringan ikat	98783

Tampak pada Tabel II salah satu yang menjadi tantangan dalam permasalahan ini adalah adanya ketidakseimbangan data jumlah sel (*imbalanced data*).

B. Diagram Blok Sistem

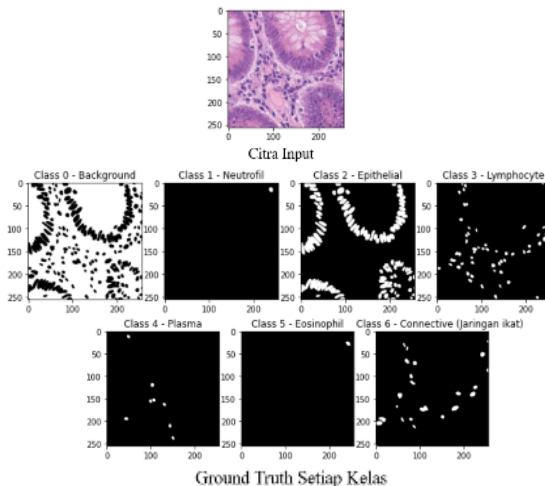
Diagram blok sistem untuk segmentasi semantik dan klasifikasi sel pada citra histologi yang dilakukan dalam penelitian ini ditunjukkan pada Gbr 1. Pada tahap *data preparation* data citra akan disesuaikan dengan kebutuhan masing-masing jaringan terlatih (*pre-trained network*). *Data preparation* terdiri data *preprocessing* dan *train-test split*. Setelah tahap *data preparation* dilakukan proses *training* sesuai dengan *pretrained network* yang diujicoba.

Setiap citra memiliki *ground truth* yang terdiri dari tujuh *channel* yang merepresentasikan jenis kelas yaitu *channel 0* untuk kelas 0 (*background*), *channel 1* untuk kelas 1 (neutrofil), *channel 2* untuk kelas 2 (epitel), *channel 3* untuk kelas 3 (limfosit), *channel 4* untuk kelas 4 (plasma), *channel 5* untuk kelas 5 (eosinofil), dan *channel 6* untuk kelas 6 (jaringan ikat/*connective*)



Gbr. 1 Diagram Blok Proses Segmentasi Semantik dan Klasifikasi Sel pada Citra Histologi

Contoh tampilan citra sel dan tiap-tiap *channel* tampak pada gambar berikut ini.



Gbr. 2 Citra Input dengan Pasangan *Ground Truth* pada Setiap Kelas

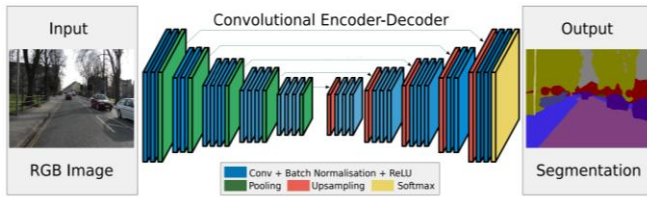
Proses *resize* dilakukan untuk mengubah ukuran citra sesuai dengan yang dibutuhkan oleh *pre-trained model*. Seluruh citra pada dataset yang digunakan sebelumnya berukuran 256x256 piksel dan akan di-*resize* menjadi berukuran 224x224 piksel.

Kemudian dilanjutkan proses normalisasi. Untuk *pre-trained model* VGG16 dilakukan dengan menggunakan *mean image subtraction* yaitu dengan melakukan *zero-centering* atau nilai piksel dikurangi nilai mean pada setiap *channel* R, G, B [11]. Untuk *pre-trained model* DenseNet-121 dan EfficientNet-B2 nilai piksel diskalakan 0-1 kemudian untuk setiap *channel* dinormalisasi [12].

Train-Test Split dilakukan untuk memisahkan data *training*, *validation*, dan *test*. Rasio pembagian data *training*, *validation*, dan *test* pada penelitian ini yaitu 80% data *training*, 10% data *validation*, dan 10% data *test*.

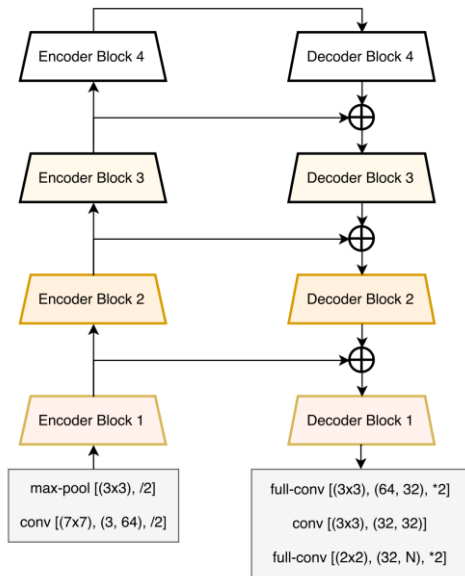
C. Arsitektur Encoder-Decoder

Isu yang akan dipecahkan dalam penelitian ini dapat dikategorikan sebagai isu segmentasi semantik. Segmentasi semantik adalah salah satu tema penting dalam bidang *computer vision* yang secara khusus melakukan proses klasifikasi tiap piksel ke dalam kategori kelas yang telah ditentukan. Salah satu metode yang digunakan dalam isu ini adalah *deep learning* [7] yang memiliki arsitektur *encoder-decoder*. Skema umum arsitektur ini dapat dilihat pada Gbr 3 [8].



Gbr. 3 Arsitektur Encoder-Decoder [8]

Pada penelitian ini struktur utama dalam arsitektur *encoder-decoder* yang digunakan adalah arsitektur LinkNet [5] seperti tampak pada Gbr 4. Struktur LinkNet terdiri dari blok inisial, blok *encoder*, blok *decoder* dan blok final. Blok inisial terdiri dari *layer* konvolusi dengan kernel ukuran 7x7 dan stride 2, *layer* *max-pooling* dengan kernel ukuran 3x3. LinkNet menghubungkan *encoder* ke *decoder* dengan melakukan *summation* atau setiap output dari *encoder* ditambahkan ke input dari *decoder* pasangannya untuk mendapatkan informasi yang telah dipelajari oleh *encoder* yang bisa digunakan oleh *decoder* dan kemudian dilakukan operasi *upsampling*.



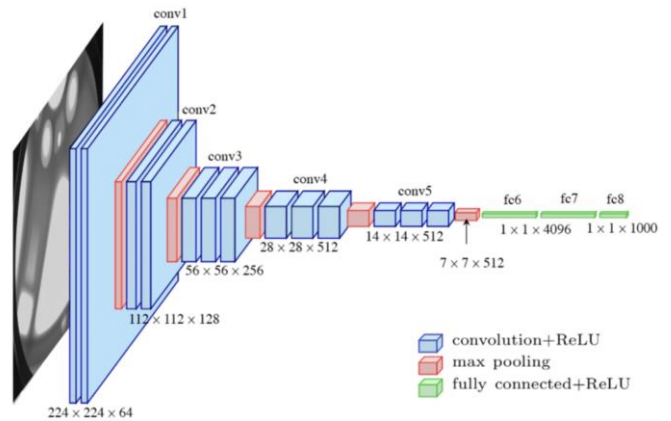
Gbr. 4 Arsitektur Utama LinkNet [5]

D. Pretrained Network pada Lapisan Encoder

Pada struktur aslinya seperti yang digagas oleh Abhishek haurasia dan Eugenio Culurciello [5] LinkNet memanfaatkan jaringan terlatih (*pretrained network*) ResNet18. Namun dalam uji coba awal dalam penelitian ini, model tersebut tidak memberikan hasil yang maksimal. Pada penelitian ini jaringan terlatih VGG16, DenseNet121, dan EfficientNetB2 digunakan pada lapisan *Encoder*. Ketiga model CNN tersebut telah dilatih dengan *dataset* ImageNet sehingga dapat berfungsi sebagai *feature extractor*. Sebelum dilakukan training

dilakukan pengaturan beberapa parameter seperti *loss function* yang digunakan adalah *categorical cross entropy* dan *optimizer* yang digunakan yaitu *optimizer Adam*.

Model VGG16 adalah model CNN yang diusulkan oleh Karen Simonyan dan Andrew Zisserman [9]. Struktur VGG16 ditunjukkan pada Gbr 5 yang terdiri dari 5 blok konvolusi dengan ukuran kernel 3x3, *max pooling*, dan 3 *fully connected layer*. Blok konvolusi terdiri dari operasi konvolusi 3x3 dengan setiap *layer* konvolusi dilengkapi dengan fungsi aktivasi ReLU. Pada setiap akhir blok hasil konvolusi diikuti dengan *layer* *max-pooling* untuk melakukan *downsampling*. Bagian *fully connected layer* digunakan untuk melakukan klasifikasi.



Gbr. 5 Model CNN VGG16 [9]

DenseNet-121 adalah model CNN yang diperkenalkan oleh Gao Huang, dkk. Pada Gbr 6 ditunjukkan struktur DenseNet-121 yang terdiri dari *layer* konvolusi dengan ukuran kernel 7x7, *max pooling* ukuran 3x3, 4 *dense* blok dan 3 *transition layer*. *Dense* blok terdiri dari *batch normalization*, fungsi aktivasi ReLU dan *layer* konvolusi. *Transition layer* terdiri dari *layer* konvolusi dan *average pooling*. Setelah melalui *dense* blok dan *transition layer* terdapat *global average pooling* ukuran 7x7 dan *fully connected layer*.

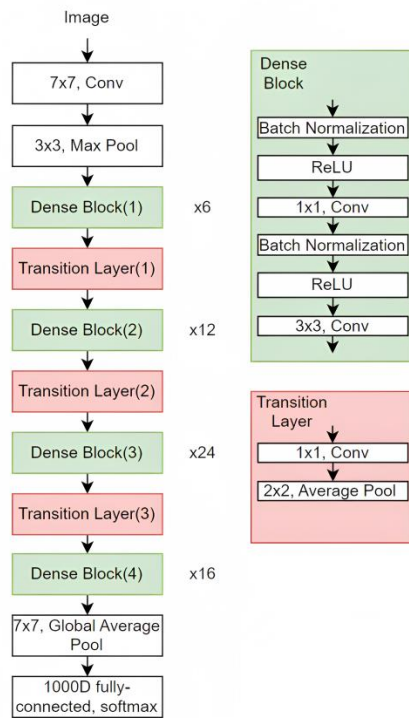
E. Penilaian Kinerja

Intersection over Union (IoU) merupakan *metric* yang dapat digunakan untuk mengukur seberapa besar *overlap* antara *bounding box* hasil prediksi dengan *bounding box* target atau *ground truth* [7]. Persamaan (1) menunjukkan persamaan IoU yaitu area irisan atau *intersection* dibagi area gabungan atau *union* [11].

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{\text{area of intersection}}{\text{area of union}} \quad (1)$$

Pada permasalahan segmentasi semantik nilai IoU diperoleh dengan menghitung jumlah piksel yang menjadi irisan antara nilai benar (*ground truth*) dan hasil prediksi kemudian dibagi dengan total piksel yang menjadi gabungan dari keduanya. Hal ini dilakukan karena yang diutamakan bukanlah berupa bentuk *bounding box* melainkan seberapa

banyak piksel dari hasil prediksi yang dapat sesuai dengan piksel dari sel yang akan disegmentasi.



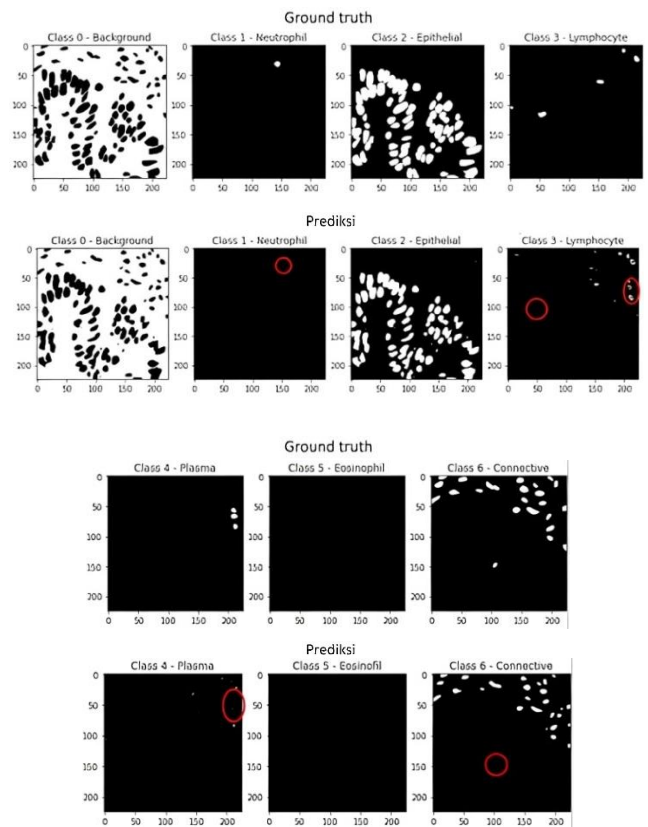
Gbr. 6 Arsitektur DenseNet-121 [10]

Selanjutnya nilai *Mean Intersection over Union* (mIoU) untuk setiap kelas dihitung dengan mengambil nilai rata-rata IoU dari seluruh citra yaitu dengan menjumlahkan seluruh nilai IoU pada suatu kelas dari seluruh citra dibagi dengan jumlah citra yang memuat kelas tersebut.

III. HASIL

Hasil segmentasi dan klasifikasi dari salah satu sampel pada data test menggunakan model LinkNet-DenseNet121 ditunjukkan pada Gbr 7. Pada gambar tersebut masih terdapat beberapa kesalahan prediksi. Ukuran sel pada hasil prediksi lebih kecil dibandingkan dengan sel pada *ground truth*.

Perbandingan hasil kinerja seluruh model dalam melakukan prediksi pada data *test* ditampilkan pada Tabel III. Berdasarkan Tabel III model dengan hasil pengujian dengan *metric mean IoU* (mIoU) tertinggi yaitu arsitektur LinkNet dengan *encoder* DenseNet-121. Dengan arsitektur LinkNet-DenseNet121 diperoleh nilai mIoU untuk kelas 2 (epitel) sebesar 0.5791, kelas 3 (limfosit) sebesar 0.3935, kelas 4 (plasma) sebesar 0.0444, dan kelas 6 (jaringan ikat/*connective*) sebesar 0.4342. Diantara seluruh arsitektur terdapat dua arsitektur yang dapat melakukan prediksi untuk kelas 4 (plasma) yaitu arsitektur LinkNet-Densenet121 dan LinkNet-EfficientNetB2. Untuk seluruh arsitektur dalam memprediksi kelas 1 (neutrofil) dan kelas 5 (eosinofil) diperoleh nilai 0.0 yang menunjukkan bahwa tidak ada irisan antara *ground truth* dengan prediksi.



Gbr. 7 Hasil Prediksi dan *Ground Truth* menggunakan Model LinkNet-DenseNet121

TABEL III
PERBANDINGAN PERFORMA SETIAP MODEL

mIoU	Arsitektur	Arsitektur		
		LinkNet-DenseNet121	LinkNet-VGG16	LinkNet-EfficientNetB2
Kelas 0 Back-ground		0.9643	0.9693	0.9628
Kelas 1 Neutrofil		0.0	0.0	0.0
Kelas 2 Epitel		0.5791	0.5734	0.5367
Kelas 3 Limfosit		0.3935	0.3661	0.2500
Kelas 4 Plasma		0.0444	0.0	0.0256
Kelas 5 Eosinofil		0.0	0.0	0.0
Kelas 6 Jaringan ikat		0.4342	0.4278	0.3798

Nilai mIoU sebesar 0.0 menunjukkan model tidak berhasil dalam memprediksi kelas 1 dan kelas 5. Hal ini dapat dirasionalisasikan karena distribusi sel pada kelas 1 dan kelas 5 sangat minoritas dibandingkan dengan kelas-kelas yang lain (lihat Tabel II). Perbandingan jumlah sel antara kelas 2 (sel epitel) dengan kelas 1 (sel neutrofil) yaitu 1:55. Perbandingan jumlah sel antara kelas 2 (sel epitel) dengan kelas 5 (sel eosinofil) yaitu 1:73. Berdasarkan perbandingan distribusi jumlah sel tiap kelas tersebut menjadikan kelas 1 (sel neutrofil) dan kelas 5 (sel eosinofil), dengan jumlah sel minoritas (*imbalanced data*) dibandingkan dengan kelas lainnya sehingga memiliki mIoU yang rendah. Masalah *imbalanced data* ini mempengaruhi hasil segmentasi dan klasifikasi pada

kelas minoritas yaitu kelas 1 (sel neutrofil) dan kelas 5 (sel eosinofil).

IV. SIMPULAN

Segmentasi dan klasifikasi sel pada citra histologi berhasil diimplementasikan dengan menggunakan jaringan konvolusional *encoder-decoder*. Model dengan nilai mIoU tertinggi yaitu LinkNet dengan *encoder* DenseNet-121 dengan nilai mIoU sebesar 0.5791 pada kelas 2 (sel epitel), 0.3935 pada kelas 3 (sel limfosit), 0.0444 pada kelas 4 (sel plasma), dan 0.4342 pada kelas 6 (jaringan ikat). Dataset dengan distribusi sel yang *imbalanced* mempengaruhi hasil prediksi dalam melakukan segmentasi dan klasifikasi pada kelas minoritas.

Pengembangan selanjutnya, dengan memperbanyak data khususnya untuk kelas minoritas yaitu dengan melakukan *cropping* pada bagian yang terdapat sel dari kelas minoritas dan membuang bagian sel dari kelas yang mendominasi sehingga rasio antara sel pada kelas yang minoritas dan sel pada kelas yang mendominasi relatif tidak berbeda jauh.

REFERENSI

- [1] S. R. Bolsover, J. S. Hyams, E. A. Shephard, and H. A. White, *CELL BIOLOGY A Short Course*. 2004.
- [2] S. Graham *et al.*, "Lizard: A Large-Scale Dataset for Colonic Nuclear Instance Segmentation and Classification," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-Octob, pp. 684–693, 2021.
- [3] A. A. Adeyinka, M. O. Adebisi, N. O. Akande, R. O. Ogundokun, A. A. Kayode, and T. O. Oladele, *A Deep Convolutional Encoder-Decoder Architecture for Retinal Blood Vessels Segmentation*, vol. 11623 LNCS, no. August 2020. Springer International Publishing, 2019.
- [4] V. A. Natarajan, M. Sunil Kumar, R. Patan, S. Kallam, and M. Y. Noor Mohamed, "Segmentation of Nuclei in Histopathology images using Fully Convolutional Deep Neural Architecture," *2020 Int. Conf. Comput. Inf. Technol. ICCIT 2020*, no. March 2021, 2020.
- [5] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," *2017 IEEE Vis. Commun. Image Process. VCIP 2017*, vol. 2018-Janua, pp. 1–4, 2018.
- [6] CoNIC: Colon Nuclei Identification and Counting Challenge, <http://https://conic-challenge.grand-challenge.org> diakses Januari 2023
- [7] A. Gulli, S. Pal, *Deep Learning with Keras*. 2017.
- [8] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.
- [10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017.
- [11] Ghiani, L., dkk. In-Field Automatic Detection of Grape Bunches under a Totally Uncontrolled Environment. *Sensors*. Vol 21 no 11, 2021.
- [12] https://www.tensorflow.org/api_docs/python/tf/keras/applications/vgg16 (accessed Jan. 30, 2023).
- [13] https://www.tensorflow.org/api_docs/python/tf/keras/applications/densenet/preprocess_input (accessed Jan. 30, 2023)