

# IMPLEMENTASI ARSITEKTUR VGG-UNET DALAM MELAKUKAN SEGMENTASI KERETAKAN PADA CITRA BANGUNAN

Jimmy Aprilyanto<sup>1\*</sup>, Yohannes<sup>2</sup>

<sup>1,2</sup>Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang  
<sup>1</sup>jimmyaprl1004@gmail.com, <sup>2</sup>yohannesmasterous@mdp.ac.id

---

## Kata kunci:

keretakan; optimizer; UNet; VGG-16

---

**Abstract:** Cracks are surface symptom caused by a force or combination of forces acting beyond the ability of a building or its constituent materials to respond. The indoor repeat detection process is usually done manually, but manual detection has many drawbacks such as safety and time. In this study, the author uses his VGG-UNet convolutional neural network (CNN) architecture and three optimizers Adam, RMSprop, and his SGD to detect cracks in building images. The dataset used in this crack detection study is a crack segmentation dataset consisting of 11,298 images composed of original and ground truth images. The dataset is split into 9,603 training data and his 1,695 testing data. Optimizer Adam has the highest score for VGG-UNet architecture with a Mean Intersection over Union (MIoU) score of 70.35%. On the other hand, SGD Optimizer achieved the lowest score with MIoU value of 61.24%.

**Abstrak:** Retak adalah manifestasi permukaan yang disebabkan oleh gaya atau kombinasi gaya yang bekerja melebihi kapasitas yang dapat ditampung oleh suatu bangunan atau komponen materialnya. Proses pendeteksian retakan pada bangunan biasanya dilakukan secara manual, namun pendeteksian secara manual memiliki banyak kekurangan seperti keamanan, waktu dan sebagainya. Dalam studi ini, penulis menggunakan *Convolutional Neural Network* (CNN) arsitektur VGG-UNet dan *Optimizer* Adam, RMSprop, dan SGD untuk mendeteksi retakan pada citra bangunan. Kumpulan data yang digunakan dalam penelitian pendeteksian retakan ini adalah kumpulan data segmentasi retakan yang terdiri dari 11.298 citra masing-masing citra asli dan citra *ground truth*. Dataset dibagi menjadi 9.603 data latih dan 1.695 data uji. *Optimizer* Adam memiliki skor tertinggi untuk arsitektur VGG-UNet dengan skor *Mean Intersection over Union* (MIoU) sebesar 70,35%. Di sisi lain, optimizer SGD mencapai skor terendah dengan nilai MIoU 61,24%.

---

Aprilyanto & Yohannes. (2023). Implementasi Arsitektur VGG-Unet Dalam Melakukan Segmentasi Keretakan pada Citra Bangunan. *MDP Student Conference 2023*

---

## PENDAHULUAN

### Latar Belakang

Setelah bangunan selesai, bangunan itu memburuk seiring waktu karena erosi dan faktor lainnya [1]. Salah satu cara untuk mengetahui kualitas suatu bangunan adalah dengan menemukan retakan pada bangunan tersebut. Mendeteksi retakan pada permukaan bangunan penting dilakukan untuk mengetahui

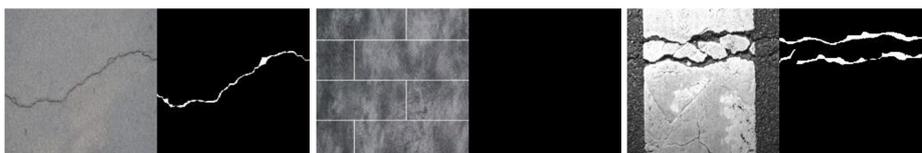
kondisi bangunan [2]. Keretakan merupakan suatu gejala yang timbul pada permukaan yang diakibatkan oleh suatu gaya atau kombinasi gaya yang bekerja dengan melebihi kapasitas yang dapat diterima oleh suatu bangunan atau komponen material dari bangunan tersebut [3]. Deteksi keretakan bangunan biasanya dilakukan secara manual, yang memiliki banyak kelemahan seperti masalah keamanan, waktu, biaya, dan sebagainya. Selain mendeteksi keretakan pada bangunan secara manual, ada juga deteksi keretakan menggunakan *computer vision* yang telah dikembangkan selama beberapa dekade yang salah satunya adalah *deep learning* yang digunakan untuk melakukan segmentasi bagian tertentu dalam citra [4]. Citra dalam bahasa latin *imago* merupakan representasi, kemiripan atau inialisasi dari suatu objek atau benda yang dapat dikelompokkan menjadi citra tampak dan citra tak tampak, yang dapat didefinisikan sebagai fungsi dua variabel  $f(x,y)$  dengan  $x$  dan  $y$  merupakan koordinat spesial dan nilai dari  $f(x,y)$  itu sendiri adalah intensitas dari citra pada koordinat tersebut [5].

Segmentasi menggunakan *deep learning* sudah banyak dilakukan dan diujii dalam membantu manusia dalam berbagai bidang salah satunya adalah penelitian [6] yang menggunakan *deep learning* CNN arsitektur Dense-Unet 569 untuk melakukan segmentasi secara semantik 3 dimensi dari hati dan tumor. Arsitektur ini memanfaatkan koneksi dari arsitektur DenseNet dan *links* dari arsitektur UNet yang mempertahankan ekstraksi fitur tingkat rendah akan tetapi memberikan hasil yang efektif. Untuk segmentasi hati, arsitektur Dense-Unet 569 menghasilkan nilai DICE sebesar 96,2% dan DICE global sebesar 96,7%. Sedangkan untuk segmentasi tumor, arsitektur DenseUNet 569 menghasilkan DICE sebesar 69,6% dan untuk DICE global sebesar 80,7%. Adapun penelitian lain yang menggunakan *deep learning* adalah penelitian [7] yang melakukan segmentasi tumor pada otak. Penelitian [7] menggunakan arsitektur UNet yang telah dikembangkan dengan arsitektur VGG-16. Penelitian [7] juga membandingkan arsitektur VGG-UNet dengan arsitektur UNet yang biasa dimana arsitektur VGG-UNet memberikan hasil yang lebih baik daripada UNet biasa dengan nilai *pixel accuracy* untuk VGG-UNet sebesar 0,998. Adapun penelitian lainnya yakni [8] yang menggunakan *deep learning* CNN untuk melakukan segmentasi keretakan pada struktur besi seperti jembatan. Arsitektur UNet dan model FCN dengan mengimplementasikan arsitektur ResNet 34 sebagai *backbone* digunakan pada pada penelitian [8] dengan membandingkan hasil evaluasi IoU. Penelitian [8] membuktikan bahwa arsitektur UNet memberikan performa hasil segmentasi yang lebih bagus dengan *epoch* yang lebih sedikit dibandingkan model FCN dengan arsitektur ResNet 34 sebagai *backbone* dengan nilai IoU pada UNet sebesar 0,6506 pada *epoch* ke-90.

Dari penelitian-penelitian terdahulu tersebut, maka dilakukan penelitian yang melakukan segmentasi keretakan pada citra bangunan menggunakan arsitektur VGG-Unet yang bertujuan untuk dapat menerapkan metode CNN arsitektur VGG-UNet dalam melakukan segmentasi bagian keretakan pada permukaan bangunan menggunakan tiga *optimizer* yakni Adam, RMSprop, dan SGD.

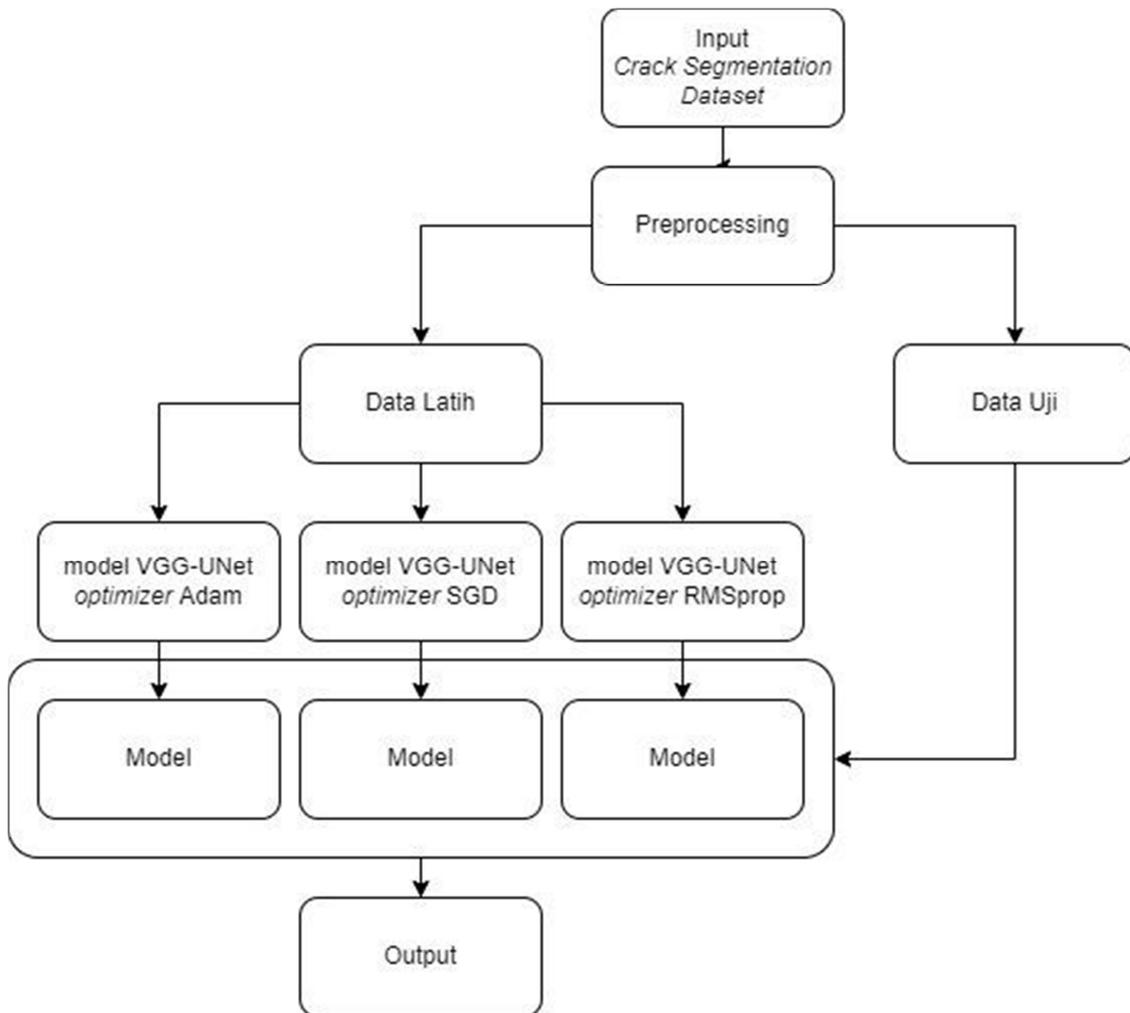
## METODE

Penelitian ini menggunakan dataset publik yang didapat pada *website Kaggle.com* dengan nama *Crack Segmentation Dataset* [9]. Dataset ini memiliki ukuran awal sebesar  $448 \times 448$  piksel dengan total citra sebanyak 11.298 citra masing-masing untuk citra original dan citra *mask* yang nantinya akan digunakan sebagai *ground truth* dari citra original. Dataset ini telah dibagi menjadi data latih dan data uji dengan total untuk data latih sebanyak 9.603 citra sedangkan untuk data uji sebanyak 1.695 citra. Untuk contoh dataset dapat dilihat pada Gambar 1.



Gambar 1. Crack Segmentation Dataset

Tahapan pertama pada metode dimulai dengan *preprocessing* citra yakni me-*resize* ukuran citra yang awalnya sebesar  $448 \times 448$  menjadi  $224 \times 224$ . Selain *resize* ukuran citra, nilai piksel pada *mask* citra baik untuk data latih dan data uji diubah yang awalnya memiliki nilai 0 dan 255 menjadi 0 dan 1. Ekstensi pada citra *mask* untuk data latih dan data uji juga diubah yang awalnya memiliki ekstensi .jpg menjadi .png.



**Gambar 2. Rancangan Sistem**

Tahapan selanjutnya yakni data latih digunakan untuk melatih model CNN arsitektur VGG-Unet dengan tiga *optimizer* yakni Adam, SGD, dan RMSprop sehingga mendapatkan tiga buah model dengan *optimizer* yang berbeda. Setelah mendapatkan model yang telah dilatih, dilakukan tahap pengujian menggunakan data uji sehingga mendapatkan citra *mask* prediksi dari model yang telah dilatih. Tahapan terakhir yakni evaluasi dimana hasil *mask* segmentasi yang telah diprediksi oleh model yang telah dilatih dibandingkan dengan *mask ground truth* citra yang didapat dari dataset menggunakan evaluasi *Mean Intersection over Union* (MioU).

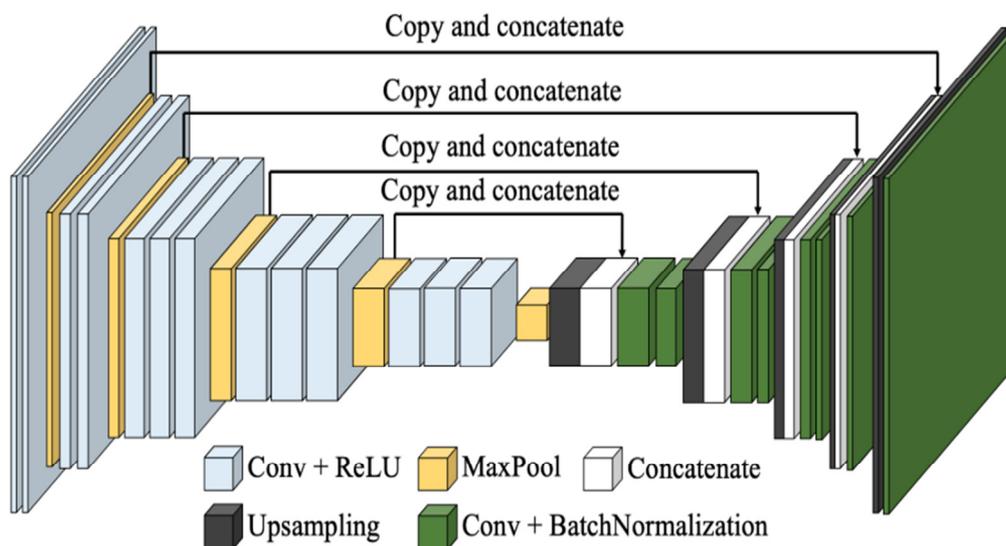
### **Convolutional Neural Network (CNN)**

*Convolutional Neural Network* (CNN) merupakan varian dari *Multi-Layer Perceptron* (MLP) yang terinspirasi dari jaringan saraf manusia terdiri dari tiga susunan dimensi neuron yakni lebar, tinggi dan

kedalaman. Lebar dan tinggi merupakan ukuran dari lapisan, sedangkan kedalaman merupakan total dari lapisan [10].

### VGG-UNet

CNN dengan arsitektur VGG-UNet hampir sama dengan model CNN arsitektur UNet biasa, dimana memiliki 2 bagian yakni *contracting path* yang berfungsi untuk menangkap konteks pada citra. Sedangkan bagian *expanding path* terdiri dari 5 *upsampling block* dan menggunakan *expanding path* yang simetris berfungsi untuk membuat lokalisasi secara presisi dari konteks yang didapat pada bagian *contracting path* [11]. Akan tetapi bagian *contracting path* pada UNet diubah dengan arsitektur VGG-16 dimana jaringan VGG memiliki layer yang sangat dalam dan filter konvolusi yang kecil yang dapat membantu model mengekstrak fitur yang cacat lebih bagus [12]. Arsitektur VGG-UNet dapat dilihat pada Gambar 3.



Gambar 3. Rancangan Sistem

Dapat dilihat pada Gambar 3, terdapat dua bagian pada arsitektur VGG-UNet. *Down-sampling block* terdiri dari layer konvolusi yang digabungkan dengan fungsi aktivasi ReLU dan layer *max pooling*. Untuk *up-sampling block* terdiri dari *upsampling layer* dan layer konvolusi dan juga diikuti oleh layer tambahan yakni *batch normalization*. Pada proses *up-sampling block*, beberapa layer disalin dan digabungkan/di-*concatenated* dengan output dari layer konvolusi. Pada proses *down-sampling block*, lebar dan tinggi citra menjadi setengah dari ukuran awal, sedangkan pada proses *up-sampling block*, lebar dan tinggi citra menjadi dua kali lebih besar dari ukuran semula [11].

### Optimizer Adam

*Optimizer Adaptive Moment Estimation*, atau sering disingkat *optimizer Adam*, ditemukan oleh Diederik P. Kingma dan Jimmy Lei Ba menggabungkan keunggulan dari dua metode terkenal lainnya, AdaGrad dan RMSProp. Beberapa keuntungan dari pengoptimal Adam adalah bahwa ukuran pembaruan parameter tidak berubah ketika gradien diskalakan ulang, ukuran langkah secara kasar dibatasi oleh hyperparameter ukuran langkah. Pengoptimal Adam juga tidak memerlukan lensa tetap, bekerja dengan gradien tipis, dan melakukan pencerahan bertahap secara alami [13]. *Optimizer Adam* dapat dilihat pada Persamaan 1.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \quad (1)$$

$\theta_{t+1}$  merupakan parameter hasil pembaruan,  $\theta_t$  merupakan parameter yang belum diperbarui,  $\eta$  merupakan *learning rate*,  $\hat{m}$  adalah gradien kuadrat momen orde pertama, dan  $\hat{v}$  merupakan gradien kuadrat momen orde kedua.

### Optimizer SGD

SGD adalah varian dari *Gradient Decent Optimization* yang selalu memperbarui semua parameter data yang dilatih. SGD tidak berulang saat memperbarui parameter, menghasilkan kinerja yang lebih cepat pada kumpulan data besar [14]. Untuk *optimizer* SGD dapat dilihat pada Persamaan 2.

$$\theta = \theta - \eta * \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)}) \quad (2)$$

Dimana  $\theta$  merupakan parameter hasil pembaruan,  $\eta$  adalah *learning rate*,  $x^{(i)}$  dan  $y^{(i)}$  merupakan data yang sedang dilatih [14].

### Optimizer RMSprop

RMSprop adalah varian AdaGrad yang bekerja lebih baik di lingkungan non-cembung dengan mengganti gradien berkerumun dengan rata-rata pergerakan berbobot eksponensial [14]. *Optimizer* RMSprop dapat dilihat pada Persamaan 3, 4, dan 5.

$$r = \rho r + (1 - \rho) g \odot g \quad (3)$$

$$\Delta\theta = -\frac{\alpha}{\delta + \sqrt{r}} \odot g \quad (4)$$

$$\theta = \theta + \Delta\theta \quad (5)$$

$r$  adalah *acumulated squared gradient*,  $\rho$  adalah *decay grade*,  $\Delta\theta$  adalah *computer update*,  $\alpha$  adalah *learning rate*, dan  $\delta$  adalah konstanta yang memiliki nilai  $10^{-7}$ , dan  $\theta$  merupakan parameter inisial [14].

### Mean Intersection over Union (MIoU)

*Mean Intersection Over Union* (MIoU) merupakan standar pengukuran evaluasi yang digunakan untuk mengukur hasil segmentasi dengan membagi hasil irisan dari citra *mask* prediksi dan *citra mask ground truth* dengan gabungan antara kedua citra tersebut sehingga memberikan hasil evaluasi dengan nilai terkecil sebesar 0 dan nilai evaluasi terbesar sebesar 1. Rasio ini bisa diformulasikan menjadi banyaknya nilai *true positive* (*intersection*) dibagi dengan *true positive*, *false positive*, dan *false negative* (*Union*) [15]. IoU dapat dilihat pada persamaan 6 dan untuk MIoU dapat dilihat pada Persamaan 7

$$IoU = \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (6)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (7)$$

Dimana :

$k + 1$  Total kelas

$p_{ij}$  Jumlah piksel kelas ke-i yang diprediksi sebagai kelas ke-j

$p_{ji}$  Jumlah piksel kelas ke-j yang diprediksi sebagai kelas ke-i

$p_{ii}$  Jumlah piksel kelas ke-i yang diprediksi sebagai kelas ke-i

$p_{ii}$  mewakili nilai dari *true positive*, sedangkan  $p_{ij}$  dan  $p_{ji}$  biasanya diinterpretasikan masing - masing sebagai *false positive* dan *false negative* [15].

## HASIL DAN PEMBAHASAN

Pengujian pada penelitian ini dilakukan dengan CNN arsitektur VGG-UNet dengan membandingkan tiga *optimizer* yakni Adam, RMSprop, dan SGD. Parameter tetap yang digunakan pada penelitian segmentasi keretakan pada citra bangunan ini adalah *epoch* yang digunakan sebesar 20, *batch size* sebesar 2, dan parameter *learning rate* sebesar 0,001.

Hasil *training* dari ketiga skenario yakni menggunakan tiga *optimizer* yaitu Adam, SGD dan RMSprop dapat dilihat pada Tabel 1

**Tabel 1. Akurasi dan Loss Training Arsitektur VGG-Unet Optimizer Adam**

Epoch	Adam		SGD		RMSprop	
	Akurasi Training	Loss Training	Akurasi Training	Loss Training	Akurasi Training	Loss Training
1	0,9264	0,2743	0,931	0,2295	0,9252	0,2596
2	0,9684	0,0967	0,963	0,1284	0,9662	0,0957
3	0,9706	0,0863	0,9643	0,1217	0,9688	0,0895
4	0,9695	0,086	0,9651	0,1158	0,9686	0,087
5	0,9708	0,0833	0,9646	0,1159	0,9708	0,0827
6	0,9724	0,0791	0,9657	0,112	0,9719	0,0807
7	0,9724	0,0775	0,9661	0,1114	0,9719	0,0773
8	0,9738	0,0732	0,9658	0,1117	0,9736	0,0743
9	0,9743	0,0712	0,9667	0,1085	0,9741	0,0725
10	0,9742	0,0716	0,9665	0,1076	0,9739	0,0731
11	0,9749	0,069	0,9671	0,1052	0,975	0,0696
12	0,9738	0,0734	0,9673	0,105	0,974	0,0732
13	0,9749	0,0691	0,9682	0,1038	0,9747	0,0704
14	0,9742	0,0719	0,9674	0,1041	0,9741	0,0728
15	0,975	0,0689	0,9684	0,1025	0,9752	0,0691
16	0,9759	0,0661	0,9672	0,1047	0,9755	0,067
17	0,9756	0,0664	0,9675	0,1044	0,9751	0,0681
18	0,976	0,065	0,9693	0,0991	0,9756	0,0674
19	0,977	0,0615	0,9677	0,103	0,977	0,0625
20	0,9768	0,0636	0,9692	0,0968	0,9765	0,0653

Adapun hasil evaluasi *Mean Intersection over Union* (MIoU) dari ketiga *optimizer* dalam melakukan segmentasi keretakan pada citra bangunan dapat dilihat pada Tabel 2.

**Tabel 2. Perbandingan Hasil Evaluasi dari Arsitektur VGG-Unet Menggunakan Ketiga Optimizer**

Optimizer	Adam		MIoU
	Akurasi Training	Loss Training	
Adam	0,9706	0,4364	0,7035
SGD	0,9630	0,2418	0,6024
RMSprop	0,9658	0,4195	0,6927

Dari Tabel 2 dapat dilihat bahwa penggunaan *optimizer* Adam dalam model CNN arsitektur VGG-UNet dalam melakukan segmentasi keretakan pada citra bangunan mendapatkan hasil evaluasi terbaik dengan skor MIoU sebesar 0,7035 atau sebesar 70,35%. Sedangkan untuk *optimizer* SGD mendapatkan hasil evaluasi MIoU terkecil yakni sebesar 0,6024 atau sebesar 60,24%. Penggunaan *optimizer* RMSprop berada di posisi kedua dengan skor MIoU sebesar 0,6927 atau sebesar 69,27%. Hal ini sesuai dengan penelitian [16] yang menyatakan bahwa penggunaan *optimizer* Adam dan RMSprop mendapatkan hasil yang lebih baik daripada penggunaan *optimizer* SGD.

## SIMPULAN

Kesimpulan dari penelitian ini adalah model CNN arsitektur VGG-UNet dapat melakukan segmentasi keretakan pada citra bangunan dengan cukup baik dengan hasil evaluasi MIoU tertinggi didapat menggunakan optimizer Adam yakni sebesar 70,35%. Sedangkan untuk hasil evaluasi MIoU terkecil didapat pada saat menggunakan optimizer SGD yakni sebesar 60,24%. Dari hasil tersebut sehingga juga dapat disimpulkan bahwa optimizer mempengaruhi baik itu hasil evaluasi dari model tersebut ataupun akurasi dan loss pada saat melatih model.

## DAFTAR PUSTAKA

- [1] Z. Li, H. Zhu, dan M. Huang, "A Deep Learning-Based Fine Crack Segmentation Network on Full-Scale Steel Bridge Images with Complicated Backgrounds," *IEEE Access*, Vol. 9, Hal. 114989–114997, 2021, doi: 10.1109/ACCESS.2021.3105279.
- [2] P. N. Hadinata, D. Simanta, dan L. Eddy, "Deep Convolutional Neural Network untuk Mendeteksi Retak pada Permukaan Beton yang Memiliki Void," *J. Sustain. Constr.*, Vol. 1, No. 1, Hal. 45–55, 2021, doi: 10.26593/josc.v1i1.5151.
- [3] A. S. Ariyanto, "Analisis Jenis Kerusakan pada Bangunan Gedung Bertingkat (Studi Kasus pada Gedung Apartemen dan Hotel Candiland Semarang)," *Bangun Rekaprima*, Vol. 06, No. 1, Hal. 45–57, 2020, [Daring]. Tersedia pada: file:///C:/Users/user/Downloads/1929-108084-1-SM (1).pdf
- [4] C. Han, T. Ma, J. Huyan, X. Huang, dan Y. Zhang, "CrackW-Net: A Novel Pavement Crack Image Segmentation Convolutional Neural Network," *IEEE Trans. Intell. Transp. Syst.*, No. August, Hal. 1–10, 2021, doi: 10.1109/tits.2021.3095507.
- [5] U. Karolina, S. Hartanto, M. Robi, dan I. Ferdian, "Operasi Konvolusi pada Pengolahan Citra Digital Untuk Deteksi Tepi Menggunakan Matlab," *MDP STUDENT Conf.*, hal. 95–103, 2022.
- [6] N. Alalwan, A. Abozeid, A. A. A. ElHabshy, dan A. Alzahrani, "Efficient 3D Deep Learning Model for Medical Image Semantic Segmentation," *Alexandria Eng. J.*, Vol. 60, No. 1, Hal. 1231–1239, 2021, doi: 10.1016/j.aej.2020.10.046.
- [7] S. Ghosh, A. Chaki, dan K. Santosh, "Improved U-Net Architecture with VGG-16 for Brain Tumor Segmentation," *Phys. Eng. Sci. Med.*, Vol. 44, No. 3, Hal. 703–712, 2021, doi: 10.1007/s13246-021-01019-w.
- [8] C. Dong, L. Li, J. Yan, Z. Zhang, H. Pan, dan F. N. Catbas, "Pixel-Level Fatigue Crack Segmentation In Large-Scale Images of Steel Structures Using An Encoder–Decoder Network," *Sensors*, Vol. 21, No. 12, Hal. 1–17, 2021, doi: 10.3390/s21124135.
- [9] L. Middha, "Crack Segmentation Dataset," *kaggle.com*, 2020. <https://www.kaggle.com/datasets/lakshaymiddha/crack-segmentation-dataset> (diakses 4 Oktober 2022).
- [10] Micheal, "Klasifikasi Spesies Kupu Kupu Menggunakan Metode Convolutional Neural Network," *MDP Student Conf.*, Vol. 1, No. 1, Hal. 569–577, 2022, [Daring]. Tersedia pada: <https://jurnal.mdp.ac.id/index.php/msc/article/view/1928>

- [11] J. Shi dkk., "Improvement of Damage Segmentation Based on Pixel-Level Data Balance Using VGG-UNet," *Appl. Sci.*, Vol. 11, No. 2, Hal. 518, Jan 2021, doi: 10.3390/app11020518.
- [12] Y. Mei, H. Jin, B. Yu, E. Wu, dan K. Yang, "Visual Geometry Group-UNet: Deep Learning Ultrasonic Image Reconstruction For Curved Parts," *J. Acoust. Soc. Am.*, Vol. 149, No. 5, Hal. 2997–3009, 2021, doi: 10.1121/10.0004827.
- [13] D. P. Kingma dan J. L. Ba, "Adam: A Method For Stochastic Optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Hal. 1–15, 2015.
- [14] N. D. Miranda, L. Novamizanti, S. Rizal, F. T. Elektro, dan U. Telkom, "Convolutional Neural Network pada Klasifikasi Sidik Jari Menggunakan Resnet-50 Classification of Fingerprint Pattern Using Convolutional Neural Network in Clahe Image," *J. Tek. Inform.*, Vol. 1, No. 2, Hal. 61–68, 2020.
- [15] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, dan J. Garcia-Rodriguez, "A Review on Deep Learning Techniques Applied to Semantic Segmentation," Hal. 1–23, Apr 2017, [Daring]. Tersedia pada: <http://arxiv.org/abs/1704.06857>
- [16] A. Willyanto, D. Alamsyah, dan H. Irsyad, "Identifikasi Tulisan Tangan Aksara Jepang Hiragana Menggunakan Metode CNN Arsitektur VGG-16," *J. Algoritm.*, Vol. 2, No. 1, Hal. 1–11, 2021.