

# Pengenalan Varietas Ikan Koi Berdasarkan Foto Menggunakan Simple Linear Iterative Clustering Superpixel Segmentation dan Convolutional Neural

Arya Tandy Hermawan<sup>1</sup>, Ilham Ari Elbaith Zaeni<sup>1\*</sup>, Aji Prasetya Wibawa<sup>1</sup>, Gunawan<sup>2</sup>,  
Yosi Kristian<sup>2</sup>, Shandy Pranata Darmawan<sup>2</sup>

<sup>1</sup>Universitas Negeri Malang, Jl. Semarang No. 5 Malang, Jawa Timur, Indonesia

<sup>2</sup>Institut Sains dan Teknologi Terpadu Surabaya, Jawa Timur, Indonesia

\*Penulis korespondensi, Surel: pembimbing123@gmail.com

Paper received: Paper received: 04-11-2021; revised: 13-11-2021; accepted: 19-11-2021

## Abstract

### Abstract

Object segmentation and image recognition are two computer vision tasks which are still being developed until today. Simple Linear Iterative Clustering is an algorithm which is very popular to help with object segmentation tasks because it is the best in terms of result and speed. In image recognition, Convolutional Neural Networks are also one of the best approaches for any kind of recognition tasks because of their efficiency and the ability to recognize objects like animals do. Koi fish have become a very interesting object to be researched because they are difficult to segment and distinguished between their varieties. The dataset consists of 600 images of Koi fish from 10 different varieties. The Koi fish's recognition process begins with generating super pixels for the input image. The next step is to merge all neighborhood super pixels by their color similarities. After this step, almost all the background pixels should be detected so that the actual object, the Koi fish, can be segmented. The segmented image is then given to a Convolutional Neural Networks, to learn any important features which distinguish every Koi fish variety from one another. A trained Convolutional Neural Networks can then give a Koi fish variety prediction for an input image. Based on a series of segmentation and model tests performed, it is proven that the segmentation technique, which uses Simple Linear Iterative Clustering in this project, performs exceptionally well across almost all the images in the dataset. The model produced from this project is also able to classify a wide range of Koi fish varieties accurately at 90% accuracy with segmentation and 87% without segmentation.

**Keywords:** superpixel; SLIC; graph; machine learning; neural network

## Abstrak

Segmentasi dan pengenalan objek pada gambar masih merupakan dua buah masalah pada *computer vision* yang masih terus diteliti dan dikembangkan hingga saat ini. *Simple Linear Iterative Clustering* merupakan salah satu algoritma segmentasi superpixel yang cukup populer untuk membantu melakukan segmentasi objek karena memiliki hasil superpixel yang baik dan dapat berjalan dengan cepat. Untuk pengenalan objek, *Convolutional Neural Networks* masih merupakan salah satu yang terbaik untuk berbagai masalah karena efisien dan mampu mengenali objek pada gambar layaknya hewan mengenali objek yang dilihatnya. Ikan koi menjadi sebuah objek yang menarik untuk diteliti karena sulit untuk disegmentasi dan dikenali jenisnya bahkan oleh manusia. Dataset yang digunakan berisi 600 gambar yang terdiri dari 10 varietas ikan koi. Pengenalan ikan koi diawali dengan melakukan *generate superpixel* pada gambar input, kemudian menggabungkan *superpixel-superpixel* terdekat yang memiliki warna yang mirip. Dengan cara ini, maka hampir seluruh *pixel background* dapat dideteksi sehingga objek ikan koi dapat disegmentasi. Gambar hasil segmentasi kemudian dilatihkan ke *Convolutional Neural Networks* yang akan mempelajari fitur-fitur penting pada setiap jenis ikan koi yang diteliti. *Convolutional Neural Networks* yang telah dilatih kemudian dapat memberikan prediksi varietas ikan koi dari sebuah input gambar. Berdasarkan hasil uji coba segmentasi dan model yang digunakan, dibuktikan bahwa teknik segmentasi yang memanfaatkan *Simple Linear Iterative Clustering* yang dilakukan berhasil untuk hampir seluruh gambar pada

dataset. Model yang dibuat mampu mengklasifikasikan varietas ikan koi dengan akurasi 90% dengan segmentasi dan 87% tanpa segmentasi.

**Kata kunci:** superpixel; SLIC; graph; machine learning; neural network

## 1. Pendahuluan

Ikan koi merupakan jenis ikan hias yang banyak dipelihara oleh banyak kalangan masyarakat, terutama di wilayah Asia. Beberapa menjadikannya sebagai sebuah koleksi karena keindahan dan keunikan dari warna dan corak ikan koi. Berdasarkan keunikan warna dan corak ikan koi, kemudian orang mencoba untuk mengelompokkan ikan koi ke dalam suatu kelompok varietas berdasarkan warna dan coraknya. Namun seiring dengan berjalannya waktu, jumlah varietas ikan koi semakin bertambah dan tidak terkendali karena orang-orang mencoba untuk mengawinsilangkan ikan-ikan dengan varietas yang berbeda untuk menghasilkan sebuah varietas ikan koi yang baru. Hal ini tentunya dapat menjadi membingungkan karena beberapa varietas dari ikan koi tampak sangat mirip dan susah dibedakan bahkan oleh seorang yang mengenal ikan koi cukup lama. Melalui penelitian ini diharapkan dapat menghasilkan sebuah program yang dapat mengklasifikasikan varietas ikan koi secara otomatis sehingga dapat membantu orang-orang untuk mengenali varietas dari macam-macam ikan koi berdasarkan sebuah foto.



**Gambar 1. Varietas Ikan Koi yang Digunakan**

Salah satu teknik yang digunakan pada pengerjaan penelitian ini adalah segmentasi *superpixel*. Berbeda dengan sebuah *pixel* pada gambar, *superpixel* merupakan sebuah daerah pada gambar yang mencakup beberapa *pixel* yang berada di dalam areanya. Salah satu algoritma yang paling banyak digunakan untuk melakukan segmentasi *superpixel* adalah Simple Linear Iterative Clustering (SLIC) (Achanta dkk., 2010)(Achanta dkk., 2012) karena mudah dipahami, cepat, dan memiliki hasil output yang sangat baik. Melihat sebuah gambar dari sisi *superpixel* dapat membantu untuk menyederhanakan informasi pada gambar. Sebuah ikan koi selalu difoto di dalam air. Pengambilan foto dari luar air terhadap objek di dalam air dapat menimbulkan efek-efek cahaya yang tidak diinginkan pada hasil foto. Oleh karena itu, *superpixel* diharapkan dapat membantu menyederhanakan informasi-informasi yang tidak dibutuhkan seperti efek-efek cahaya.

Pada penelitian ini juga digunakan sebuah model neural network yaitu Convolutional Neural Networks (CNN)(LeCun dkk., 1991). CNN sudah banyak digunakan untuk menyelesaikan berbagai masalah seperti klasifikasi(Krizhevsky dkk., 2012)(Rastegari, dkk., 2016)(Simonyan & Zisserman, 2014)(Springenberg dkk., 2014) (Purnomo, 2016), *object detection*(He dkk., 2015)( Redmon & Farhadi, 2016)(Redmon dkk., 2016). CNN juga dikenal sangat baik dan efisien dalam mempelajari sekumpulan gambar untuk kemudian dicari fitur-fitur penting pada sekumpulan gambar tersebut tanpa perlu diprogram secara eksplisit . Hal ini tentunya sangat bermanfaat pada pembuatan penelitian ini karena ada berjuta-juta corak ikan koi yang berbeda untuk setiap varietas ikan koi, sehingga menyebabkan sulitnya untuk

merumuskan fitur untuk sebuah varietas ikan koi ke dalam sebuah persamaan matematis. Dengan memberi komputer kemampuan untuk mempelajari data secara otomatis, diharapkan komputer mampu mendeteksi fitur-fitur dari varietas ikan koi yang selama ini sulit untuk dirumuskan.

## 2. Metode

### 2.1. Segmentasi

Proses segmentasi (Shi & Malik, 2000) (Trémeau & Colantoni, 2000) bertujuan untuk menghilangkan *background* dari setiap gambar ikan koi yang bervariasi. Segmentasi dilakukan untuk menyederhanakan informasi pada gambar yang nantinya dipelajari oleh CNN. *Background* pada gambar tidak menentukan apa varietas dari ikan koi pada gambar. Sehingga dengan dihilangkannya *background*, diharapkan CNN nantinya hanya perlu melihat informasi yang penting saja, yaitu objek ikan koi.

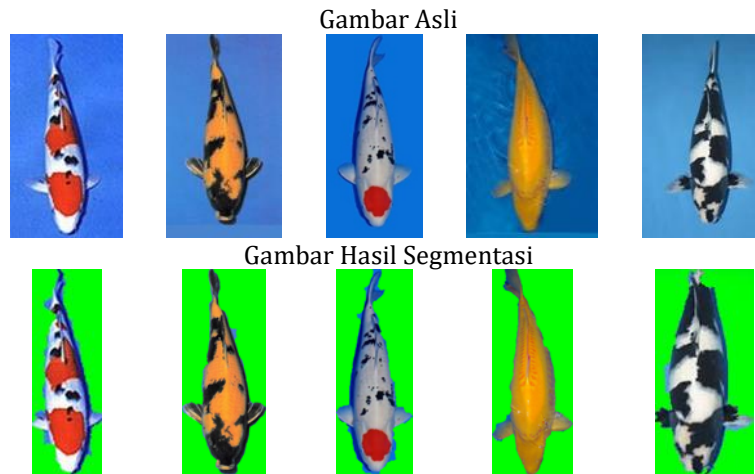
Proses segmentasi diawali dengan sebuah input berupa gambar RGB dimana terdapat sebuah ikan koi dengan posisi ekor diatas dan kepala di bawah. Contoh gambar ikan koi untuk setiap varietas pada dataset dapat dilihat pada Gambar 1.

Kemudian digenerate superpixel-superpixel pada gambar input menggunakan algoritma SLIC (Achanta dkk., 2010) (Achanta dkk., 2012). Pada penelitian ini digunakan SLIC Zero (SLICO), yang merupakan varian dari SLIC yang dapat secara adaptif menyesuaikan kompleksitas dari setiap superpixel yang dihasilkan. Setelah superpixel berhasil digenerate, dilakukan pembuatan graf yang menghubungkan setiap superpixel dengan seluruh superpixel yang merupakan tetangganya dengan sebuah edge yang memiliki bobot berupa jarak warna kedua superpixel. Perhitungan jarak warna dapat menggunakan rumus manhattan distance.

$$d = |C_2 - C_1|$$

Dimana  $d$  adalah jarak warna antar kedua superpixel, dan  $C_i$  adalah nilai rata-rata warna untuk superpixel  $i$ . Edge-edge yang terbentuk kemudian dipotong apabila memiliki bobot yang lebih kecil dari sebuah *threshold*  $t$ . Pada tahap ini, seluruh superpixel milik background diharapkan telah tergabung menjadi satu kesatuan sehingga warna background dapat direplace dengan sebuah warna solid. Langkah terakhir adalah mencari bounding box dari ikan koi pada gambar untuk kemudian dicrop. Hasil crop ikan koi dengan warna background yang telah direplace ini merupakan output dari proses segmentasi.

Dalam proses segmentasi terdapat dua buah parameter yaitu  $k$ , yang merupakan jumlah superpixel yang diinginkan, dan  $t$ , yang merupakan threshold graph cut. Seluruh gambar pada dataset dapat disegmentasi dengan baik dengan cara menyetting kedua parameter dengan nilai yang optimal. Pada dataset yang digunakan,  $k = 300$  dan  $t = 35$  merupakan nilai parameter yang dapat melakukan segmentasi dengan baik pada hampir seluruh gambar dengan berbagai macam resolusi.



Gambar 2. Contoh Hasil Segmentasi

Pada Gambar 2 dapat dilihat contoh hasil segmentasi yang dilakukan untuk beberapa gambar ikan koi pada dataset. Dapat dilihat bahwa pada seluruh gambar, masih terdapat beberapa pixel background yang ikut tersegmentasi seperti pada contoh hasil segmentasi ke 1, 2, dan 3. Sedangkan pada gambar ke 4 dan 5, dapat dilihat bahwa ada beberapa pixel milik ikan koi yang hilang pada bagian sirip dan ekor. Hal ini disebabkan karena warna pixel pada bagian sirip dan ekor ikan koi cenderung transparan sehingga sangat mirip dengan background. Namun hilangnya pixel-pixel pada daerah sirip dan ekor ini tidak menjadi masalah karena bukan merupakan pembeda utama antar varietas ikan koi.

## 2.2. Arsitektur CNN

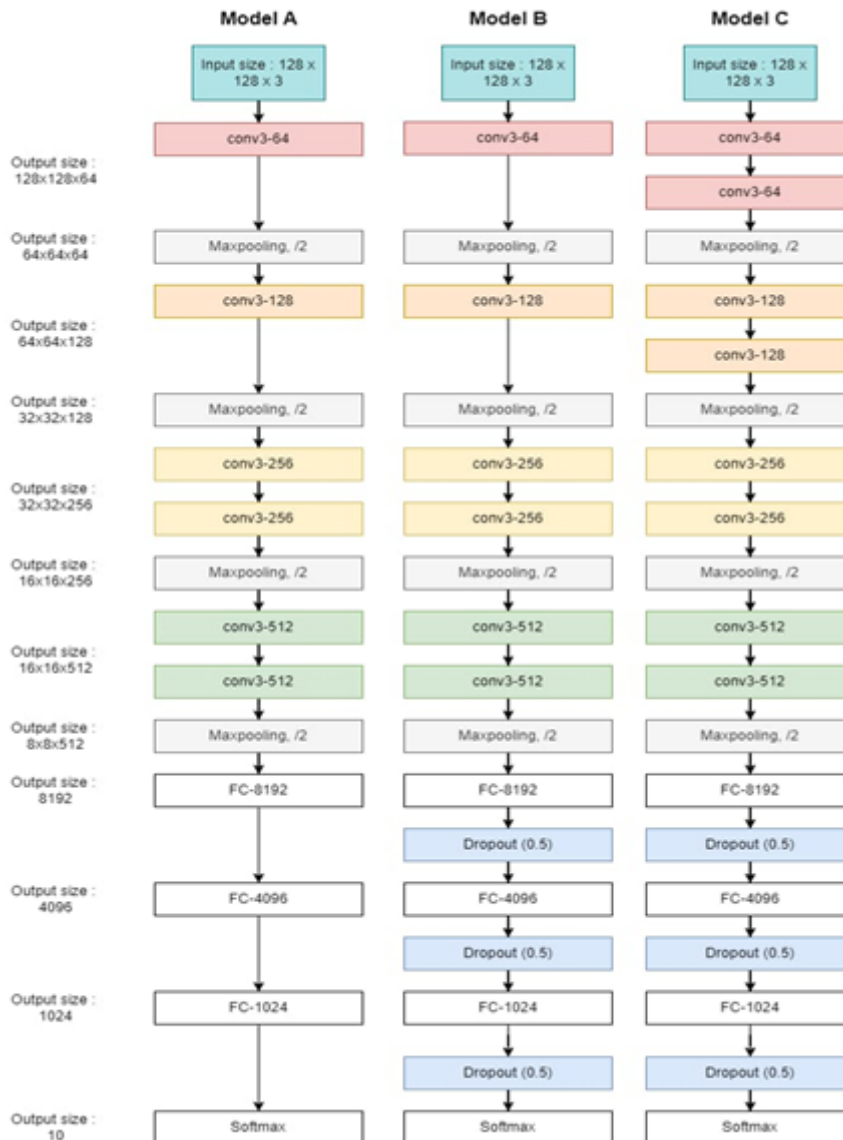
Pada bagian ini dijelaskan mengenai arsitektur Convolutional Neural Network (LeCun dkk., 1998) yang digunakan (Krizhevsky dkk., 2012). Terdapat tiga buah arsitektur yang digunakan sehingga dapat dianalisa performa masing-masing arsitektur terhadap kemampuan menyelesaikan masalah pengenalan ikan koi. Arsitektur CNN dibuat berdasarkan model VGG16 (Simonyan & Zisserman, 2014) dan (Limantoro, 2017) dengan beberapa modifikasi seperti pengurangan layer, penambahan layer dropout, dan lain sebagainya. Tabel arsitektur CNN yang digunakan dapat dilihat pada Gambar 3.

Seluruh model CNN yang digunakan merupakan kombinasi dari convolutional layer, maxpooling layer, fully connected layer, dan dropout layer. Convolutional layer pada seluruh CNN menggunakan kernel berukuran 3x3, stride sebesar 1, dan dilakukan penambahan zero-padding sebesar 1. Dengan konfigurasi seperti ini, maka dimensi panjang dan lebar volume output convolutional layer akan sama dengan panjang dan lebar input, sedangkan kedalaman volume akan mengikuti jumlah kernel yang digunakan. Sebagai contoh CONV3-64 menandakan sebuah convolutional layer dengan 64 kernel yang masing-masing berukuran 3x3.

Maxpooling layer yang digunakan adalah berukuran 2x2 dengan stride sebesar 2. Dengan konfigurasi seperti ini, maka dimensi panjang dan lebar dari volume output maxpooling layer adalah 50% dari panjang dan lebar volume input. Dengan kata lain, konfigurasi maxpooling ini menyebabkan jumlah aktivasi neuron pada volume input tereduksi sebanyak 75%.

Sebelum Fully connected layer, akan dilakukan flatten yang mengubah gambar 2D menjadi data 1 dimensi. Fully connected layer diberi nama FC- $n$ , dimana  $n$  adalah jumlah neuron pada sebuah fully connected layer. Pada beberapa arsitektur CNN pada Gambar 3, beberapa model menggunakan dropout layer yang berfungsi untuk melakukan regularisasi agar mengurangi kemungkinan model untuk mengalami overfitting. Probabilitas dropout yang digunakan adalah sebesar 0.5 atau 50%.

Gambar 3. Arsitektur CNN



Pada tiga buah model yang terdapat pada Gambar 3, dapat dilihat bahwa Model A merupakan arsitektur CNN tanpa regularisasi. Model ini digunakan dengan tujuan untuk mengetahui pengaruh regularisasi yang dilakukan terhadap performa CNN. Kemudian pada Model B, dapat dilihat bahwa convolutional layer pada awal CNN dikurangi dua buah dari Model C. Hal ini dilakukan untuk menguji apakah fitur-fitur dasar dari ikan koi berpengaruh terhadap varietasnya. Pada umumnya, layer-layer awal CNN akan mendeteksi fitur-fitur dasar seperti garis dan bentuk. Sedangkan pada masalah pengenalan ikan koi ini, pada seluruh

gambar pada dataset, seluruh ikan koi memiliki bentuk yang relatif sama. Sehingga dilakukan pengurangan parameter pada layer-layer awal CNN.

### 2.2.1. Training

Proses training dilakukan dengan menggunakan CPU Intel® Core™ i5-6500 (up to 3.60GHz) dan 32GB RAM. Setiap model ditraining dengan dataset training yang terdiri dari 400 gambar untuk 10 class (40 gambar setiap kelas), dan dilakukan validasi setiap 10 epoch terhadap 100 gambar dari dataset validasi (10 gambar setiap kelas). Perbandingan waktu training untuk masing-masing model dapat dilihat pada Tabel 1.

**Tabel 1. Perbandingan Waktu Training**

Nama Model	Waktu
Model A @91 epoch	1h 41m 8s
Model B @142 epoch	2h 42m 13s
Model C @151 epoch	4h 16m 21s

Pada Tabel 1 dapat dilihat pada setiap model berhenti training pada epoch yang berbeda-beda. Hal ini dikarenakan training akan secara otomatis berhenti ketika nilai cost function pada saat validasi tidak menurun selama 50 epoch. Setelah proses training selesai, selanjutnya dipilih checkpoint terbaik dari setiap model sehingga dapat dibandingkan performa dari setiap model.

## 3. Hasil dan Pembahasan

Uji coba pengenalan ikan koi dilakukan untuk setiap model pada Tabel 2 terhadap 100 gambar dari dataset testing. Berdasarkan hasil uji coba yang dilakukan, diketahui bahwa Model B merupakan model dengan performa yang paling baik. Diikuti dengan Model C yang memiliki performa dibawah Model B hanya sekitar 1% pada pengukuran akurasi dan rata-rata F1 score.

**Tabel 2. Perbandingan Akurasi**

Nama Model	Akurasi
Model A	84.0%
Model B	90.0%
Model C	89.0%

Pada Tabel 3 dapat dilihat bahwa Model B memiliki nilai akurasi terbaik, yaitu 90.0%, sedikit diatas Model C yang memiliki akurasi 89.0%. Pada perbandingan nilai rata-rata F1 score, Model B juga merupakan model dengan performa terbaik dengan nilai 0.901.

**Tabel 3. Perbandingan Rata-rata F1 Score**

Nama Model	F1 Score
Model A	0.831
Model B	0.901
Model C	0.888

Model yang digunakan cukup cepat untuk membuat prediksi dari sebuah input gambar. Mula-mula sebuah gambar disegmentasi terlebih dahulu menggunakan teknik yang telah dijelaskan, kecepatan segmentasi bervariasi tergantung dari resolusi gambar yang digunakan. Perbandingan kecepatan segmentasi beberapa resolusi gambar dapat dilihat pada Tabel 4.

**Tabel 4. Perbandingan Kecepatan Segmentasi**

Resolusi Gambar (w x h) k=300, t=35	Kecepatan (ms)
183 x 275	236
188 x 268	248
383 x 596	1,025
597 x 897	2,361

Setelah gambar disegmentasi, barulah gambar hasil segmentasi diberikan ke CNN untuk dicari class-nya. Perbandingan kecepatan setiap model CNN yang digunakan dapat dilihat pada Tabel 5.

**Tabel 5. Perbandingan Kecepatan Testing CNN**

Nama Model	Kecepatan (ms)
Model A	611
Model B	613
Model C	652

Berdasarkan uji coba kecepatan testing yang tertera pada Tabel 5, dapat dilihat bahwa ketiga model memiliki kecepatan berkisar di 600ms, yang berarti cukup cepat untuk digunakan pada dunia nyata. Berikutnya dilakukan uji coba pengaruh segmentasi objek dari gambar terhadap performa CNN. Pada uji coba ini, model yang digunakan adalah Model B yang merupakan model terbaik yang diujicobakan.

**Tabel 6. Perbandingan Akurasi Model B Tanpa dan Dengan Segmentasi**

Tanpa Segmentasi	Dengan Segmentasi
87.0%	90.0%

**Tabel 8. Perbandingan Rata-rata F1 Score Model B Tanpa dan Dengan Segmentasi**

Tanpa Segmentasi	Dengan Segmentasi
0.870	0.901

Pada Tabel 7 dapat dilihat bahwa tanpa segmentasi, performa Model B turun sebesar 3.0%. Hal yang sama juga ditemui pada nilai rata-rata F1 score. Berdasarkan hasil uji coba pada Tabel 7 dan Tabel 8 dapat dilihat bahwa teknik segmentasi objek yang digunakan mampu untuk meningkatkan performa model sebesar 3.0%. Sedangkan pada Gambar 4 dapat dilihat beberapa contoh hasil klasifikasi ikan koi dari data test. Dua label berwarna orange menandakan hasil klasifikasi yang salah. Pada pojok kiri bawah seharusnya kelasnya Tancho, namun ada varian Koi Tancho yang juga memiliki spot hitam seperti Sanke, sehingga pada sistem juga terjadi kesalahan. Sedangkan pada pojok kanan seharusnya varian Koi Showa namun dikenali sebagai Sanke. Dua varian ini memang sangat mirip dan sulit dibedakan orang awam.



Gambar 4. Contoh Hasil Klasifikasi Model B Pada Data Test

#### 4. Simpulan

Adapun beberapa kesimpulan yang didapatkan antara lain: Teknik segmentasi yang digunakan pada penelitian ini mampu untuk mensegmentasi seluruh gambar ikan koi pada dataset dengan hasil yang relatif baik tanpa menghilangkan informasi penting yang menjadi pembeda antar varietas ikan koi. Karena teknik segmentasi yang digunakan berbasis kemiripan antar superpixel background, maka untuk mendapatkan hasil yang baik background dari gambar input harus memiliki gradient yang relatif stabil dengan warna objek yang lebih mencolok dibanding background. Pada teknik segmentasi yang digunakan, tidak ditemukan sebuah nilai jumlah superpixel  $k$  dan threshold  $t$  yang dapat memberikan hasil segmentasi yang baik untuk gambar input dengan kualitas baik dan buruk, sehingga untuk beberapa gambar, nilai kedua parameter harus di-set secara manual. Pada uji coba model CNN yang dilakukan, dibuktikan bahwa penggunaan layer dropout yang berfungsi untuk melakukan regularisasi memiliki pengaruh yang signifikan terhadap performa dari CNN yang dilatih dengan menggunakan dataset yang memiliki ukuran relatif kecil. Penambahan jumlah parameter pada CNN, seperti yang diujicobakan yaitu antara model B dan model C tidak meningkatkan performa dari CNN dalam mengklasifikasikan varietas ikan koi. Teknik segmentasi objek yang dilakukan memiliki pengaruh terhadap peningkatan performa CNN sebesar 3.0%.

#### Daftar Rujukan

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). Slic superpixels (No. EPFL-REPORT-149300).
- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2274-2282.
- He, S., Lau, R. W., Liu, W., Huang, Z., & Yang, Q. (2015). Supercnn: A superpixelwise convolutional neural network for salient object detection. *International Journal of Computer Vision*, 115(3), 330-344.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.



- Limantoro, S. E., Kristian, Y., & Purwanto, D. D. (2017). Deteksi Pengendara Sepeda Motor Menggunakan Deep Convolutional Neural Networks. In *Seminar Nasional Teknologi Informasi dan Komunikasi* (pp. 79-86).
- Purnomo, M. H., Kristian, Y., Setyati, E., Rosiani, U. D., & Setiawan, E. I. (2016, October). Limitless possibilities of pervasive biomedical engineering: Directing the implementation of affective computing on automatic health monitoring system. In *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 1-4). IEEE.
- Rastegari, M., Ordonez, V., Redmon, J., & Farhadi, A. (2016, September). Xnor-net: Imagenet classification using binary convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV* (pp. 525-542). Cham: Springer International Publishing.
- Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788)..
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), 888-905.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: All convolutional net. *arXiv preprint arXiv:1412.6806*.
- Trémeau, A., & Colantoni, P. (2000). Region's adjacency graph applied to color image segmentation. *IEEE Transactions on image processing*, 9(4), 735-744.