

PENGENALAN TULISAN KARAKTER MANDARIN (HANZI) DENGAN MENGGUNAKAN METODE EFFICIENTDET

Roberto Davin¹⁾, Chairisni Lubis²⁾, Tony³⁾

^{1), 2)}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara

³⁾Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara

Jl. Letjen S. Parman No. 1, Jakarta

email : roberto.535210022@stu.untar.ac.id¹⁾, chairisnil@fti.untar.ac.id²⁾, tony@fti.untar.ac.id³⁾

Corresponding Author: chairisnil@fti.untar.ac.id

ABSTRACT

In today's era of globalization and digitalization, the need to understand and recognize language differences is becoming increasingly important, especially with the increasing cross-cultural interactions. Mandarin, as the second international language, one of the languages of the United Nations (UN), and one of the languages with the largest number of native speakers in the world, has a complex writing system consisting of thousands of characters called Hanzi. Each of these characters not only represents a sound but also a specific meaning, making the recognition of Mandarin characters a challenge. In this study, a system was developed to recognize 200 Chinese characters using the EfficientDet method, an object detection model developed by researchers from Google Research. EfficientDet is known for its efficiency in detecting objects with high accuracy while maintaining fast processing speed. In conclusion, the system obtained overall mAP results of 58%, mAR 42.64%, IoU 49.18%, precision 75%, recall 64%, and F1-score 65%.

Key words

EfficientDet, Mandarin, object detection, recognition

1. Pendahuluan

Memahami dan mengidentifikasi perbedaan bahasa telah menjadi semakin penting di dunia yang mengglobal dan digital saat ini, terutama karena hubungan lintas budaya telah meningkat. *Oxford Learner's Dictionary* mendefinisikan bahasa sebagai sistem komunikasi lisan dan tulisan yang digunakan oleh anggota suatu negara atau wilayah tertentu [1]. Sekitar 8.324 bahasa lisan dan isyarat telah didokumentasikan hingga saat ini oleh pemerintah, organisasi publik, dan kelompok akademis, menurut metodologi *World Atlas of Languages*. Sekitar 7.000 dari bahasa-bahasa ini masih digunakan hingga saat ini [2].

Seiring dengan pesatnya perkembangan ekonomi dan teknologi dunia, Tiongkok secara aktif mempromosikan daya tarik bahasa Mandarin. Kini, banyak negara menganggap pembelajaran bahasa Mandarin penting

untuk memahami berbagai aspek budaya Tiongkok dan membina hubungan yang lebih kuat dengan Tiongkok, seperti halnya di Indonesia. Bahasa Mandarin, sebagai bahasa internasional kedua, merupakan bahasa resmi Perserikatan Bangsa-Bangsa (PBB) dan salah satu bahasa yang paling banyak digunakan di dunia. Bahasa Mandarin memiliki sistem penulisan yang rumit yang terdiri dari ribuan karakter, yang dikenal sebagai *Hanzi* [3]. Setiap karakter tidak hanya mewakili bunyi tertentu, tetapi juga makna tertentu, sehingga pengenalan karakter Mandarin menjadi tantangan tersendiri.

Kompleksitas bahasa Mandarin melampaui jumlah karakternya yang sangat banyak hingga mencakup keberadaan dua sistem penulisan utama, yaitu karakter tradisional (繁体字, diucapkan Fántǐ zì) dan karakter yang disederhanakan (简体字, diucapkan Jiǎntǐ zì). Sistem penulisan ini telah berkembang seiring dengan sejarah panjang dan geografi yang luas dari penutur bahasa Mandarin. Karakter tradisional, yang telah digunakan selama ribuan tahun, masih dipertahankan di Taiwan, Hong Kong, dan Makau [4]. Versi ini sering kali memiliki goresan yang lebih rumit dan dilihat oleh banyak pengguna sebagai lebih kaya dalam nilai estetika dan budaya. Di sisi lain, karakter yang disederhanakan dihasilkan dari reformasi bahasa yang dilakukan oleh pemerintah Tiongkok pada pertengahan abad ke-20 [5].

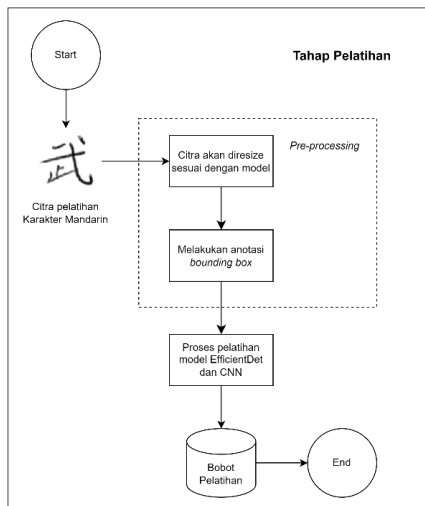
Berdasarkan hasil penelitian terdahulu, dapat terlihat bahwa metode yang digunakan sangat mempengaruhi tingkat keakuratan dalam pengenalan karakter Mandarin. Pada penelitian Ardy Kuncoro dengan metode *Backpropagation Neural Network*, akurasi yang diperoleh sebesar 59.2% [6]. Sedangkan pada penelitian Tiffany dengan metode *Convolutional Neural Network* menggunakan model VGG-19, akurasi mencapai hingga 99.2% ketika diuji dengan data Kaggle dan 62.8% dengan dokumentasi pribadi [7]. Perbedaan signifikan ini menunjukkan bahwa pendekatan berbasis CNN, khususnya dengan model VGG-19, dapat memberikan performa yang jauh lebih baik dibandingkan metode tradisional *Backpropagation Neural Network* dalam tugas pengenalan karakter Mandarin.

Penelitian ini menggunakan model deteksi objek EfficientDet dari Google Research untuk membuat

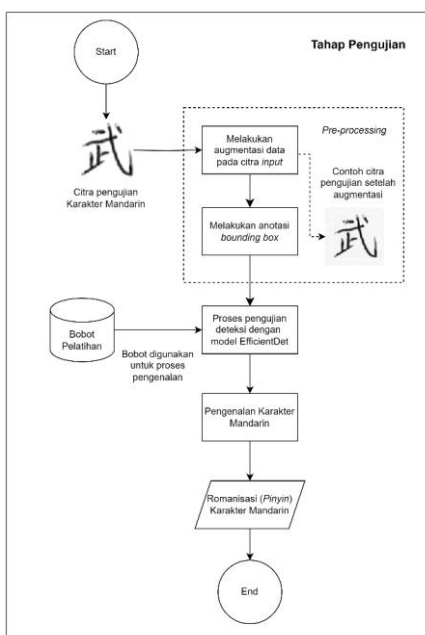
sistem pengenalan karakter Mandarin. EfficientDet terkenal karena kemampuannya mendeteksi objek dengan presisi tinggi sambil memproses data dengan cepat [8]. Dengan penggunaan EfficientDet, sistem ini berupaya mempermudah pembelajaran dan penggunaan bahasa Mandarin sekaligus menawarkan dasar yang kuat untuk penelitian di masa mendatang dalam teknologi deteksi objek dan pengenalan karakter.

2. Metode Penelitian

Sistem pengenalan tulisan karakter Mandarin yang dirancang bertujuan untuk mengidentifikasi dan mengelompokkan karakter Mandarin (*Hanzi*) berdasarkan jumlah goresan pada karakter tersebut yang dapat digunakan sebagai media pembelajaran bahasa Mandarin. **Gambar 1** dan **Gambar 2** menunjukkan tahapan alur pelatihan dan pengujian.



Gambar 1. Flowchart Tahap Pelatihan

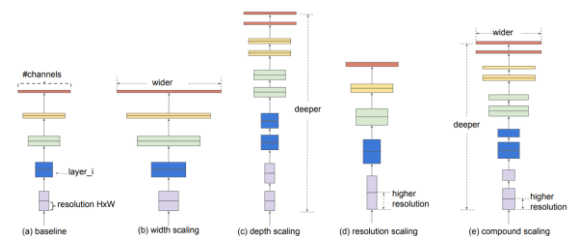


Gambar 2. Flowchart Tahap Pengujian

Rancangan sistem yang dibuat merupakan sebuah program desktop yang dapat melakukan pengenalan tulisan karakter Mandarin dengan menggunakan metode EfficientDet.

2.1 EfficientNet

EfficientNet merupakan arsitektur jaringan saraf tiruan (*neural network*) yang dirancang dengan tujuan untuk mencapai keseimbangan ideal antara akurasi dan efisiensi komputasi [9]. Keunggulan utama EfficientNet terletak pada kemampuannya menyeimbangkan skala jaringan secara proporsional pada tiga aspek utama, kedalaman (*depth*), lebar (*width*), dan resolusi (*resolution*). Pendekatan ini dikenal dengan istilah “*compound scaling*” yang memungkinkan peningkatan kapasitas model secara optimal tanpa mengorbankan efisiensi. **Gambar 3** menunjukkan tipe-tipe *model scaling*.

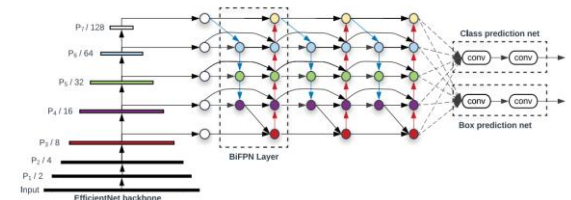


Gambar 3. Tipe-tipe Model Scaling [9]

2.2 EfficientDet

EfficientDet yang awalnya diperkenalkan oleh peneliti Google pada tahun 2020 merupakan model deteksi objek canggih yang berupaya mencapai keseimbangan terbaik antara efisiensi dan akurasi komputasi [10]. EfficientDet dengan cepat menarik perhatian karena kemampuannya untuk meningkatkan kompleksitas model tanpa mengorbankan efisiensi.

Model ini menggunakan teknik *compound scaling*, yang terinspirasi dari EfficientNet, sehingga memungkinkan penskalaan proporsional pada tiga aspek utama yaitu kedalaman, lebar, dan resolusi. Pendekatan ini memungkinkan EfficientDet untuk menjaga kinerja yang optimal dalam hal akurasi dan kecepatan, menjadikannya pilihan unggul untuk aplikasi deteksi objek modern [8]. **Gambar 4** menunjukkan arsitektur dari model EfficientDet.



Gambar 4. Arsitektur EfficientDet [8]

Arsitektur EfficientDet terdiri atas beberapa komponen penting, EfficientNet *backbone*, BiFPN *layer*, dan *class & box prediction networks*.

EfficientNet *backbone* digunakan sebagai fondasi dan mengekstrak peta fitur (*feature map*) pada resolusi

yang berbeda dari citra *input*, yaitu $P_1/2, P_2/4, P_3/8, P_4/16, P_5/32, P_6/64$, dan $P_7/128$.

BiFPN *layer* merupakan versi FPN (*Feature Pyramid Network*) yang disempurnakan dan bertanggung jawab atas penggabungan karakteristik dari skala yang berbeda untuk meningkatkan kinerja deteksi. BiFPN menggunakan pendekatan dua arah berbobot, yaitu metode *top-down* dan *bottom-up* yang menggunakan pendekatan *upsampling* atau *downsampling* untuk memodifikasi secara dinamis pentingnya fitur-fitur yang berbeda.

Dalam BiFPN, digunakan konsep *weighted feature fusion* untuk menggabungkan peta fitur dari setiap saluran dengan menerapkan bobot yang dipelajari selama pelatihan. Berikut merupakan rumusnya:

$$F_{fused} = \frac{w_1 * F_1 - w_2 * F_2}{w_1 + w_2 + \epsilon} \quad (1)$$

F_1, F_2 = input feature map yang digabungkan
 w_1, w_2 = bobot dari feature map yang digabungkan
 ϵ = nilai konstan untuk menghindari pembagian dengan 0 ($\epsilon = 10^{-4}$)

Setelah pemrosesan penggabungan fitur yang dilakukan oleh BiFPN *layer* telah selesai, fitur yang telah digabungkan akan diteruskan ke jaringan prediksi kelas dan prediksi kotak (*class and box prediction networks*). Jaringan prediksi kelas bertanggung jawab atas probabilitas setiap *anchor box* yang mencakup objek dari sebuah kelas. Jaringan prediksi kotak digunakan untuk melakukan regresi koordinat *bounding box* yang mencakup objek yang telah dideteksi.

Selama proses deteksi, model EfficientDet akan melakukan modifikasi terhadap *anchor box* untuk mendeteksi objek dengan lebih baik. Pada akhirnya, model akan menghasilkan empat nilai regresi untuk setiap *anchor box*, yaitu t_x, t_y, t_w , dan t_h yang akan digunakan untuk menentukan prediksi akhir berupa *bounding box*.

$$t_x = \frac{x_g - x_a}{w_a} \quad (2)$$

$$t_y = \frac{y_g - y_a}{h_a} \quad (3)$$

$$t_w = \ln\left(\frac{w_g}{w_a}\right) \quad (4)$$

$$t_h = \ln\left(\frac{h_g}{h_a}\right) \quad (5)$$

t_x = *offset* koordinat x dari koordinat pusat
 t_y = *offset* koordinat y dari koordinat pusat
 t_w = faktor skala lebar
 t_h = faktor skala tinggi
 x_a, y_a = koordinat pusat x dan y *anchor box*
 w_a, h_a = lebar dan tinggi *anchor box*
 x_g, y_g = koordinat pusat x dan y *ground-truth box*
 w_g, h_g = lebar dan tinggi *ground-truth box*

3. Hasil Percobaan

Percobaan model EfficientDet dilakukan dengan total data citra pelatihan dan pengujian sebanyak 2626 citra dengan perbandingan rasio 70:30 yang mencakup 200 kelas. Pada data pelatihan, tulisan karakter Mandarin diperoleh dan dikumpulkan dari tulisan beberapa teman. Untuk data pengujian, digunakan tulisan karakter Mandarin yang ditulis sendiri, dilengkapi dengan augmentasi data berupa penambahan *noise*.

Pemilihan model terbaik akan didasarkan pada nilai *mean average precision* (mAP) dan total *loss* yang diperoleh model. **Tabel 1** menampilkan perbandingan hasil pelatihan beberapa model dengan variasi *batch size* dan *epoch* yang berbeda.

Tabel 1 Perbandingan pelatihan model EfficientDet



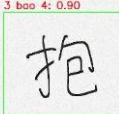

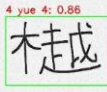
Epoch	Batch Size	mAP	mAR	IoU	Total Loss
20	8	5.05%	36.43%	3.66%	0.486
50	4	11.78%	33.16%	9.94%	0.437
100	8	58%	42.64%	49.18%	0.056

Berdasarkan hasil dari **Tabel 1**, dapat disimpulkan bahwa model dengan 100 *epoch* dan *batch size* sebanyak 8 merupakan model terbaik, yang memperoleh hasil mAP 58%, mAR 42.64%, IoU 49.18%, dan total *loss* 0.056.



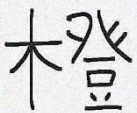
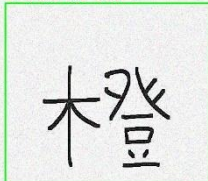
Adapun ketika dilakukan pengujian (terlihat pada **Tabel 2**) terhadap data pengujian citra tanpa *noise*, model memperoleh akurasi mAP yang cukup tinggi, yaitu sebesar 91.585% pada 200 kelas karakter Mandarin. Namun, ketika diuji dengan data dengan *noise*, akurasi mAP menurun menjadi 21.42%.

Tabel 2 Hasil Skenario Data Pengujian

Skenario	Contoh Data berdasarkan Jumlah Goresan	Akurasi (Confidence Score)	Mean Average Precision (mAP)
Karakter Mandarin Tanpa Noise			95.46%
			92.26%
			90.32%

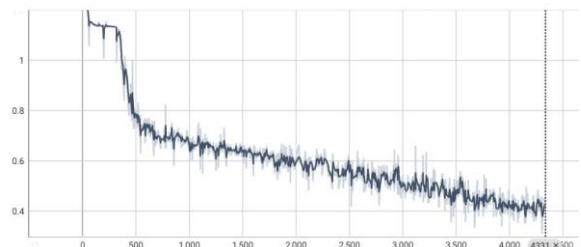
	魔		88.3%
Total Mean Average Precision (mAP)			91.585%
Karakter Mandarin Dengan Noise	八		22.54%
	抱		14.27%
	萌		16.3%
	越		32.57%
Total Mean Average Precision (mAP)			21.42%

Tabel 3 Hasil Skenario Data Pengujian

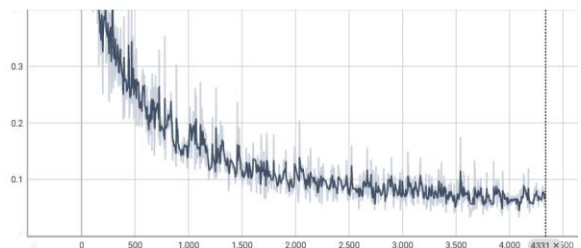
	
Tulisan Karakter chéng Tanpa Noise	Hasil Prediksi Tulisan Karakter chéng Tanpa Noise (TP)
	
Tulisan Karakter chéng Dengan Noise	Hasil Prediksi Tulisan Karakter chéng Dengan Noise (FP)

Tabel 3 menunjukkan tulisan karakter Mandarin yang dibaca sebagai chéng, berhasil terdeteksi dan dikenali dengan tepat (*true positive*) ketika digunakan *input* gambar tanpa *noise*. Namun, ketika digunakan gambar dengan *noise*, akan diperoleh hasil prediksi yang salah (*false positive*).

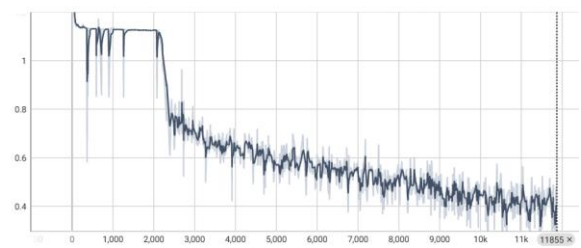
Hasil pelatihan ketiga model menunjukkan bahwa seiring bertambahnya jumlah *steps*, nilai *classification* dan *regression loss* pada masing-masing model menurun. Grafik pada Gambar 5 hingga Gambar 10 memperlihatkan penurunan ini secara bertahap, hingga akhirnya model 3 mencapai nilai *loss* terendah sebesar 0.056. Penurunan *loss* yang konsisten ini menunjukkan bahwa model semakin mampu meminimalkan kesalahan prediksi, membuktikan proses pembelajaran yang berjalan efektif dalam meningkatkan akurasi pengenalan karakter Mandarin.



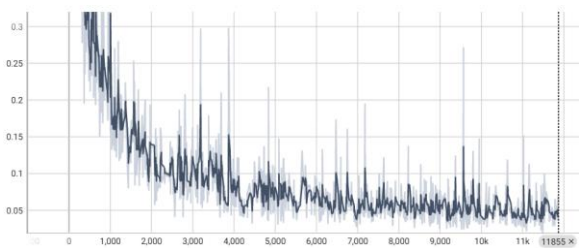
Gambar 5. Grafik Classification Loss Model 1



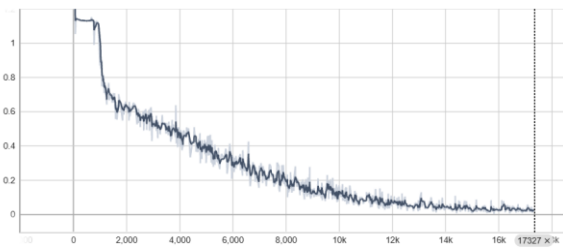
Gambar 6. Grafik Regression Loss Model 1



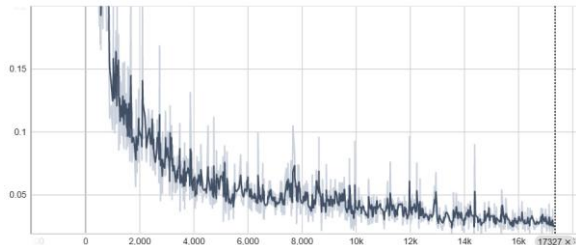
Gambar 7. Grafik Classification Loss Model 2



Gambar 8. Grafik Regression Loss Model 2



Gambar 9. Grafik Classification Loss Model 3



Gambar 10. Grafik Regression Loss Model 3

4. Kesimpulan

Dari hasil pengujian tulisan karakter Mandarin yang dilakukan sebelumnya, dapat disimpulkan beberapa hal, yakni:

1. Hasil pengujian keseluruhan modul yang dikembangkan telah berfungsi dengan baik, sesuai dengan rancangan.
2. Sistem telah berhasil mendeteksi dan mengenali *input* gambar yang tidak memiliki *noise* dengan tingkat akurasi mAP keseluruhan sebesar 91.585%. Namun, untuk *input* gambar yang memiliki *noise* memperoleh akurasi mAP keseluruhan sebesar 21.42%.
3. Sistem pengenalan tulisan karakter Mandarin ini dapat dikembangkan dengan bantuan metode lain seperti *Optical Character Recognition (OCR)*.

REFERENSI

- [1] Oxford University Press, "Oxford Learner's Dictionaries," [Online]. Available: <https://www.oxfordlearnersdictionaries.com/definition/english/language?q=language>. [Accessed 3 August 2024].
- [2] UNESCO, "World Atlas of Languages," [Online]. Available: <https://en.wal.unesco.org/discover/languages>. [Accessed 3 August 2024].
- [3] Sartono, "Badan Pengembangan dan Pembinaan Bahasa," 6 July 2020. [Online]. Available: <https://badanbahasa.kemdikbud.go.id/artikel-detail/864/bahasa-internasional>. [Accessed 3 August 2024].
- [4] A. Meredith, "Chinese Language Institute," 13 December 2023. [Online]. Available: <https://studycli.org/chinese-characters/traditional/>. [Accessed 3 August 2024].
- [5] O. Black, "Chinese Language Institute," 6 December 2023. [Online]. Available: <https://studycli.org/chinese-characters/simplified/>. [Accessed 3 August 2024].
- [6] C. Lubis, Tony and A. Kuncoro, "Pengenalan Karakter Mandarin dengan Backpropagation Neural Network,"

Seminar Nasional Teknologi Informasi dan Multimedia, pp. 25-30, 2016.

- [7] Tiffany, C. Lubis and Tony, "Pengenalan Karakter Mandarin Menggunakan Metode Convolutional Neural Network dengan Model VGG-19," Archive Skripsi FTI Untar, 2023.
- [8] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10781-10790, 2020.
- [9] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning (ICML)*, California, 2019.
- [10] J. Jia, M. Fu, X. Liu and B. Zheng, "Underwater Object Detection Based on Improved EfficientDet," *Remote Sensing*, 2022.

Roberto Davin, mahasiswa S1 yang sedang menjalankan studi pada program studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara.

Chairisni Lubis, memperoleh gelar Dra. Matematika dan Ilmu Pengetahuan Alam dari Universitas Indonesia dan gelar M.Kom dari Universitas Indonesia. Saat ini sebagai staf pengajar di Fakultas Teknologi Informasi, Universitas Tarumanagara.

Tony, memperoleh gelar S.Kom. pada tahun 2005 dari Universitas Tarumanagara, M.Kom. pada tahun 2010 dari Universitas Indonesia, dan Ph.D. pada tahun 2021 dari Curtin University. Saat ini sebagai staf pengajar di Fakultas Teknologi Informasi, Universitas Tarumanagara.