

PENGENALAN OBJEK MAKANAN CEPAT SAJI PADA VIDEO DAN REAL TIME WEBCAM MENGGUNAKAN METODE YOU LOOK ONLY ONCE (YOLO)

¹Oktaviani Ella Karlina, ²Dina Indarti

^{1,2}Fakultas Teknologi Industri Universitas Gunadarma

Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

¹oek40506@gmail.com, ²dina_indarti@staff.gunadarma.ac.id

Abstrak

Makanan cepat saji banyak diminati masyarakat di Indonesia saat ini. Makanan cepat saji mengandung lemak dan kalori yang tinggi tetapi kurang mengandung gizi yang sangat dibutuhkan tubuh sehingga dapat menimbulkan berbagai penyakit. Salah satu cara untuk mengontrol konsumsi makanan cepat saji yaitu penggunaan metode pengenalan objek makanan cepat saji dari video dan real time webcam. Metode yang dapat digunakan untuk pengenalan objek pada citra makanan cepat saji adalah deep learning. You Look Only Once (YOLO) merupakan salah satu model deep learning yang dapat digunakan untuk pengenalan objek. Penelitian ini bertujuan untuk pengenalan objek pada citra makanan cepat saji menggunakan YOLO. Penelitian ini terdiri dari beberapa tahap yaitu pengumpulan data, pra-proses data, konfigurasi jaringan YOLO, pelatihan model YOLO dan pengujian. Jumlah data citra yang digunakan dalam pelatihan yaitu 468 citra yang terdiri dari tiga jenis makanan cepat saji. Nilai avg loss pada model akhir yang dibangun dengan YOLO yaitu 4.6% dan nilai validasi mAP 100%. Berdasarkan hasil pengujian yang dilakukan menggunakan video dan real time webcam, objek pada citra makanan cepat saji berhasil dikenali dengan akurasi 63% sampai 100%.

Kata Kunci: Deep learning, makanan cepat saji, pengenalan objek, You Only Look Once

Abstract

Fast food is much in demand by people in Indonesia today. Fast food contains high fat and calories but lacks nutrients that are needed by the body so that it can cause various diseases. One way to control fast food consumption is the use of fast food object recognition methods from video and real time webcam. The method that can be used for object recognition in fast food images is deep learning. You Look Only Once (YOLO) is a deep learning model that can be used for object recognition. This study aims to recognize objects in the image of fast food using YOLO. This research consists of several stages, namely data collection, data preprocessing, YOLO network configuration, YOLO model training and testing. The amount of image data used in the training is 468 images consisting of three types of fast food. The avg loss value in the final model built with YOLO is 4.6% and the mAP validation value is 100%. Based on the results of tests conducted using video and real time webcam, objects in the image of fast food have been detected properly.

Keywords: Deep learning, fast food, object recognition, You Only Look Once

PENDAHULUAN

Manusia memiliki kebutuhan paling mendasar yang merupakan prioritas diban-

dingkan kebutuhan lainnya yaitu kebutuhan fisiologis. Kebutuhan fisiologis merupakan kebutuhan yang harus terpenuhi sehingga manusia dapat bertahan hidup. Salah satu

kebutuhan fisiologis yaitu makanan atau pangan [1]. Definisi pangan menurut Badan Pengawas Obat dan Makanan Republik Indonesia (BPOM RI) adalah segala sesuatu yang berasal dari sumber hayati baik yang diolah maupun tidak diolah yang digunakan untuk konsumsi manusia [2].

Indonesia dengan jumlah penduduk mencapai 261,9 juta pada tahun 2017 menjadi negara tujuan pemasaran produk pangan atau makanan dari berbagai negara [3]. Berbagai produk makanan masuk ke Indonesia sehingga terdapat banyak alternatif pilihan makanan mulai dari jenis, harga, maupun kualitas makanan yang beraneka ragam. Pengaruh masuknya berbagai produk makanan dari berbagai negara berdampak pada pergeseran konsep makanan masyarakat di Indonesia. Salah satu bentuk pergeseran konsep makanan di Indonesia yaitu adanya makanan cepat saji (*fast food*) dalam pilihan konsumsi masyarakat di Indonesia [4].

Makanan cepat saji merupakan makanan yang dapat dikonsumsi dalam waktu seminimal mungkin [5]. Makanan cepat saji sangat mudah ditemui dan hampir semua masyarakat pernah mengkonsumsi. Para ahli gizi dan dokter mengatakan bahwa makanan cepat saji mengandung lemak dan kalori yang tinggi tetapi kurang mengandung gizi yang sangat dibutuhkan tubuh. Oleh karena itu, jika terlalu sering mengkonsumsi makanan cepat saji dapat membahayakan kesehatan tubuh [6]. Banyak jenis makanan cepat saji yang diminati masyarakat di Indonesia antara lain

pizza, *hamburger*, dan *hotdog*. Salah satu cara untuk mengontrol konsumsi makanan cepat saji diperlukan suatu metode pengenalan objek makanan cepat saji dari video dan *realtime webcam*.

Pengenalan objek merupakan salah satu permasalahan dalam *computer vision* yang berhubungan dengan pengenalan suatu objek dalam citra dan video digital [7]. *Computer vision* merupakan salah satu cabang dari kecerdasan buatan (*artificial intelligence*) yang difokuskan pada pengembangan algoritma untuk menganalisis informasi dari suatu citra ke dalam bentuk informasi yang sebenarnya di dunia nyata. Saat ini *deep learning* banyak digunakan pada penelitian mengenai *computer vision*. Terdapat dua pendekatan dalam pengenalan objek menggunakan *deep learning* yaitu pendekatan berdasarkan *region proposal* atau metode dua tahap dan pendekatan berdasarkan metode regresi atau metode satu tahap. Metode pengenalan objek yang menggunakan dua tahap antara lain *region-based convolutional neural networks* (R-CNN) [8], *Spatial Pyramid Pooling Network* (SPP-Net) [9], *Fast R-CNN* [10], and *Faster R-CNN* [11]. Metode tersebut menggunakan dua tahap karena mengklasifikasikan ulang dengan menerapkan model ke citra di beberapa lokasi dan skala lalu memberi nilai pada citra sebagai bahan evaluasi untuk pengenalan objek. Metode pengenalan objek dalam satu tahap tanpa melakukan klasifikasi ulang sehingga menjadi lebih cepat yaitu *You Only*

Look Once (YOLO) [12], modifikasi YOLO (YOLOv2) [13], dan YOLO-LITE [14]. YOLO menerapkan *neural network* pada sebuah citra, kemudian membagi citra menjadi daerah dan memprediksi *bounding box* serta probabilitas untuk masing-masing daerah. Probabilitas untuk setiap *bounding box* kemudian dihitung untuk mengklasifikasikan sebagai objek atau bukan. YOLO dapat melakukan pengenalan objek secara *real-time* dengan kecepatan 45 *frame per second* [12]. YOLO-LITE merupakan pengembangan dari YOLO yang dapat dijalankan pada komputer *portable* seperti laptop atau *cellphone* dengan keterbatasan *Graphics Processing Unit* (GPU) [14].

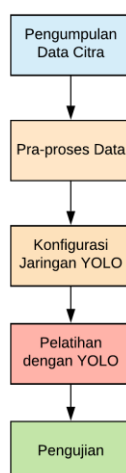
Beberapa penelitian pengenalan objek menggunakan YOLO telah dilakukan. Penelitian yang dilakukan oleh Zhang, Huang, Jin, dan Li membahas mengenai algoritma pengenalan rambu lalu lintas secara *real-time* menggunakan modifikasi YOLO. Hasil penelitian tersebut menunjukkan bahwa metode modifikasi YOLOv2 dapat mendeteksi rambu lalu lintas dengan lebih cepat dan lebih baik [15]. Liu, Chen, Li, dan Hu melakukan penelitian mengenai metode pendeteksian citra pejalan kaki berdasarkan pada YOLOv2. Hasil penelitian tersebut menunjukkan bahwa metode YOLOv2 memiliki akurasi dan kecepatan yang baik dalam mendeteksi objek

pada citra pejalan kaki [16]. Hasil penelitian yang dilakukan oleh Jupiyandi, Saniputra, Pratama, Dharmawan, dan Cholissodin menunjukkan bahwa algoritma modifikasi YOLO dapat mendeteksi jumlah mobil dengan tepat sehingga dapat mengetahui jumlah- tempat parkir mobil yang masih tersedia [17].

Oleh karena itu, pada penelitian ini membahas mengenai pengenalan objek pada citra makanan cepat saji menggunakan *You Only Look Once* (YOLO). Model YOLO yang diperoleh selanjutnya digunakan untuk mengenali objek dari video dan *real-time webcam*.

METODE PENELITIAN

Tahapan-tahapan penelitian secara umum dapat dilihat pada Gambar 1. Berdasarkan Gambar 1, penelitian dimulai dengan pengumpulan data citra lalu pra-proses data citra yang terdiri dari pelabelan dan perubahan ukuran citra. Kemudian dilakukan konfigurasi jaringan YOLO sesuai dengan data citra dan dilatih sehingga model baru YOLO terbentuk. Tahapan terakhir penelitian yaitu melakukan pengujian model menggunakan data video dan *real-time webcam*.



Gambar 1. Skema Penelitian

Pengumpulan Data Citra

Citra makanan cepat saji yang digunakan dalam penelitian ini diambil dari *dataset* pada penelitian yang dilakukan oleh Matsuda, Hoashi, dan Yanai [18]. Jumlah citra sebagai *dataset* adalah 468 citra dengan *format* jpg, yang terdiri dari tiga jenis makanan yaitu *hamburger* berjumlah 233 citra, *pizza* berjumlah 133 citra dan *hot dog* berjumlah 102 citra.

Pra-proses Data Citra

Selanjutnya dilakukan Pra-proses data yang terdiri dari pelabelan dan perubahan ukuran citra. Pelabelan citra merupakan tahap awal dimana setiap citra pada *dataset* diberikan label dengan tujuan untuk menyimpan informasi citra. Proses label dilakukan dengan cara memberikan *bounding box* beserta nama kelas pada setiap objek citra. Selanjutnya dilakukan perubahan ukuran citra untuk meningkatkan performa model YOLO dalam pengenalan objek.

Konfigurasi Jaringan YOLO

Konfigurasi jaringan diperlukan sebagai jaringan model untuk memuat data yang akan dilatih. *Batch* atau iterasi yang digunakan yaitu 64 yang berarti setiap iterasi memuat 64 citra. *Subdivision* yaitu 8 sehingga proses pemodelan membagi *batch* menjadi 8 bagian. Oleh karena itu, terdapat $64/8 = 8$ citra per *minibatch* yang dikirim ke GPU untuk diproses dan diulang sebanyak 8 kali hingga selesai. Pada iterasi selanjutnya dimulai dengan 64 citra baru.

Pelatihan Model YOLO

Model YOLO yang digunakan pada penelitian ini memiliki jaringan saraf yang terdiri dari lapisan konvolusi dengan *kernel* 3×3 dan lapisan *max-pooling* dengan 2×2 *kernel* seperti terlihat pada Gambar 2. Lapisan konvolusional terakhir memiliki 1×1 *kernel* digunakan untuk mengecilkan data ke bentuk $13 \times 13 \times 40$. 13×13 adalah ukuran *grid*, 40 didapat dari penjumlahan rumus *filter*.

layer	filters	size/strd(dil)	input	output
0 conv	16	3 x 3/ 1	416 x 416 x 3 ->	416 x 416 x 16 0.150 BF
1 max		2 x 2/ 2	416 x 416 x 16 ->	208 x 208 x 16 0.003 BF
2 conv	32	3 x 3/ 1	208 x 208 x 16 ->	208 x 208 x 32 0.399 BF
3 max		2 x 2/ 2	208 x 208 x 32 ->	104 x 104 x 32 0.001 BF
4 conv	64	3 x 3/ 1	104 x 104 x 32 ->	104 x 104 x 64 0.399 BF
5 max		2 x 2/ 2	104 x 104 x 64 ->	52 x 52 x 64 0.001 BF
6 conv	128	3 x 3/ 1	52 x 52 x 64 ->	52 x 52 x 128 0.399 BF
7 max		2 x 2/ 2	52 x 52 x 128 ->	26 x 26 x 128 0.000 BF
8 conv	256	3 x 3/ 1	26 x 26 x 128 ->	26 x 26 x 256 0.399 BF
9 max		2 x 2/ 2	26 x 26 x 256 ->	13 x 13 x 256 0.000 BF
10 conv	512	3 x 3/ 1	13 x 13 x 256 ->	13 x 13 x 512 0.399 BF
11 max		2 x 2/ 1	13 x 13 x 512 ->	13 x 13 x 512 0.000 BF
12 conv	1024	3 x 3/ 1	13 x 13 x 512 ->	13 x 13 x1024 1.595 BF
13 conv	512	3 x 3/ 1	13 x 13 x1024 ->	13 x 13 x 512 1.595 BF
14 conv	40	1 x 1/ 1	13 x 13 x 512 ->	13 x 13 x 40 0.007 BF
15 detection				

Gambar 2. Model YOLO

Pengujian Model YOLO

Pengujian dilakukan dengan dua cara yaitu video dan *real-time* menggunakan *webcam* laptop. Video didapat dari situs *youtube* menggunakan kata kunci makanan *hamburger, pizza, dan hot dog*. Pengujian dilakukan untuk mengetahui tingkat akurasi objek dengan model baru yang sudah dilatih sebelumnya.

HASIL DAN PEMBAHASAN

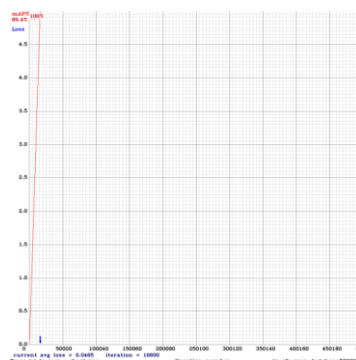
Hasil Pelatihan

Proses pelatihan dilakukan sebanyak 16.600 iterasi yang menghasilkan sebuah model baru. Model akhir yang dibangun memiliki nilai *avg loss* 4.6% dan nilai validasi

mAP (mean Average Precision) 100% seperti terlihat pada Gambar 3. Nilai *mAP* 100% menunjukkan bahwa tingkat akurasi pengenalan objek pada citra makanan cepat saji menggunakan model YOLO yang telah dilatih sebesar 100%.

Hasil Pengujian Pada Video

Salah satu hasil pengujian pada sebuah video dengan objek *pizza* dapat dilihat pada Gambar 4. Berdasarkan Gambar 4, akurasi pengenalan objek *pizza* pada video sebesar 99%. Pengenalan objek pada video dapat memberikan hasil akurasi yang berbeda-beda pada saat objek yang dikenali berpindah posisi.



Gambar 3. Nilai Avg Loss dan mAP Hasil Pelatihan



Gambar 4. Hasil Pengenalan Objek *Pizza* Pada Video

Pengujian model YOLO dalam pengenalan objek *hamburger*, *hotdog* dan *pizza* pada video dilakukan dengan menghitung akurasi *frame per second*. Video secara keseluruhan memiliki durasi 11 menit 7 detik. Hasil pengujian *frame per second* pada video dengan objek *hamburger* dapat dilihat pada Tabel 1.

Berdasarkan Tabel 1, objek *hamburger* dapat dikenali dengan akurasi antara 80% sampai 91%. Hasil pengujian untuk objek *hotdog* pada video dapat dilihat pada Tabel 2.

Objek *hotdog* dapat dikenali dengan akurasi antara 81% sampai 96% seperti yang dapat dilihat pada Tabel 2. Hasil pengujian untuk objek *pizza* pada video dapat dilihat pada Tabel 3.

Berdasarkan Tabel 3, objek *pizza* dapat dikenali dengan akurasi antara 81% sampai 100%. Hasil pengujian menggunakan sebuah video yang terdapat tiga objek yaitu *hamburger*, *hotdog*, dan *pizza* menunjukkan bahwa model YOLO dapat mengenali ketiga objek dengan baik.

Tabel 1. Hasil Pengujian Pengenalan Objek *Hamburger* Pada Video

Waktu (menit:detik)	Akurasi
00:00 – 00:03	83%
00:03 – 00:06	89%
00:07 – 00:08	83%
00:09 – 00:11	88%
00:12 – 00:13	82%
00:14 – 00:16	88%
00:17 – 00:18	82%
00:19 – 00:21	84%
00:22 – 00:23	81%
00:24 – 01:01	83%
01:02 – 01:03	80%
01:04 – 01:05	91%

Tabel 2. Hasil Pengujian Pengenalan Objek *Hotdog* Pada Video

Waktu (menit:detik)	Akurasi	Waktu (menit:detik)	Akurasi
01:06 – 01:08	85%	03:14 – 03:16	93%
01:09 – 01:13	83%	03:17 – 03:18	94%
01:14 – 01:16	81%	03:19 – 03:21	90%
01:17 – 01:18	86%	03:22 – 03:23	92%
01:19 – 01:21	83%	03:24 – 04:01	90%
01:22 – 01:23	80%	04:02 – 04:06	87%
01:24 – 02:01	87%	04:07 – 04:08	91%
02:02 – 02:03	88%	04:09 – 04:11	94%
02:02 – 02:06	87%	04:12 – 04:13	96%
02:07 – 02:08	85%	04:14 – 04:16	95%
02:09 – 02:11	88%	04:17 – 04:18	94%
02:12 – 02:13	87%	04:19 – 04:21	96%
02:14 – 02:16	86%	04:22 – 05:03	95%
02:17 – 02:18	89%	05:04 – 05:08	94%
02:19 – 02:21	87%	05:09 – 05:11	93%
02:22 – 02:23	93%	05:12 – 05:16	94%
02:24 – 03:01	92%	05:17 – 05:18	81%
03:02 – 03:03	94%	05:19 – 05:21	84%
03:04 – 03:06	91%	05:22 – 05:23	91%
03:07 – 03:08	90%	05:24 – 06:00	87%
03:09 – 03:13	92%		

Tabel 3. Hasil Pengujian Pengenalan Objek Hamburger Pada Video

Waktu (menit:detik)	Akurasi	Waktu (menit:detik)	Akurasi
06:01 – 06:02	90%	09:10 – 09:12	96%
06:02 – 06:09	97%	09:13 – 09:14	99%
06:10 – 06:12	98%	09:15 – 09:17	100%
06:13 – 06:14	96%	09:18 – 09:19	99%
06:15 – 06:17	97%	09:20 – 09:22	97%
06:18 – 07:02	98%	09:23 – 10:02	91%
07:03 – 07:12	99%	10:03 – 10:04	92%
07:13 – 07:17	100%	10:05 – 10:07	83%
07:18 – 07:24	99%	10:08 – 10:09	86%
08:00 – 08:12	100%	10:10 – 10:12	82%
08:13 – 08:19	99%	10:13 – 10:14	88%
08:20 – 08:22	89%	10:15 – 10:22	95%
08:23 – 09:04	99%	10:23 – 10:24	85%
09:05 – 09:07	100%	11:00 – 11:02	87%
09:08 – 09:09	84%	11:03 – 11:07	81%

Hasil Pengujian Secara *Real Time*

Menggunakan *Webcam*

Pengujian model YOLO mengenali objek yaitu *hamburger*, *pizza* dan *hotdog*

menggunakan *webcam* dapat dilihat pada Gambar 5. Akurasi pengenalan objek *hamburger* sebesar 87%, 79% untuk objek *hotdog* dan 63% untuk objek *pizza*.



Gambar 5. Hasil Pengenalan Objek *Hamburger*, *Pizza*, dan *Hotdog* Menggunakan *Webcam*

KESIMPULAN DAN SARAN

Implementasi algoritma *You Look Only Once* (YOLO) dalam pengenalan objek makanan cepat saji memiliki akurasi validasi mAP sebesar 100% dan *avg loss* sebesar 4.6%. Oleh karena itu dapat disimpulkan bahwa implementasi *deep learning* menggunakan YOLO mampu melakukan pengenalan objek pada citra makanan cepat saji dengan baik. Hasil dari pengujian model dalam pengenalan objek tiga jenis makanan cepat saji menggunakan video dan *real time webcam* memiliki akurasi 63% sampai 100%. Pengenalan objek memberikan hasil akurasi yang berbeda-beda pada saat objek yang dikenali berpindah posisi.

Pada penelitian lebih lanjut dapat menambahkan jumlah citra dalam tahap pelatihan sehingga saat objek berpindah posisi, pengenalan objek bisa lebih cepat dan akurat. Selain itu, spesifikasi komputer diharapkan memiliki kapasitas RAM dan kecepatan GPU yang tinggi sehingga mampu mengurangi waktu dalam pelatihan model.

DAFTAR PUSTAKA

- [1] N. S. Kasiati dan N. W. D. Rosmalawati, *Kebutuhan dasar manusia I*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2016.
- [2] Badan Pengawas Obat dan Makanan Republik Indonesia, *Peraturan Kepala Badan Pengawas Obat dan Makanan Republik Indonesia Nomor 1 Tahun 2017 Tentang Pengawasan Pangan Olahan Organik*. Jakarta: Badan Pengawas Obat dan Makanan Republik Indonesia, 2017.
- [3] Biro Pusat Statistik, *Statistik Indonesia, Statistical Yearbook of Indonesia*. Jakarta: Biro Pusat Statistik, 2018.
- [4] Sudyarto, "Daya saing ketahanan pangan melalui identifikasi sikap kepercayaan konsumen remaja terhadap produk makanan cepat saji (*fast food*) dan makanan lokal (tradisional)", *Jurnal Riset Ekonomi dan Bisnis*, vol. 9, no. 2, hal. 91 – 97, 2009.

- [5] F. Hayati, “Faktor-faktor yang berhubungan dengan konsumsi *fast food* waralaba modern dan tradisional pada remaja siswa SMU Negeri di Jakarta Selatan”, *Skripsi*, Institut Pertanian Bogor, Bogor, 2000.
- [6] J. Young-seon, *Why? Food and nutrition*. Jakarta: PT. Elex Media Komputindo, 2010.
- [7] R. Munir, *Pengolahan citra digital*. Bandung: Informatika, 2007.
- [8] R. Girshick, J. Donahue, T. Darrell, dan J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, dalam *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, hal. 580 – 587.
- [9] K. He, X. Zhang, S. Ren, dan J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, hal. 1904 – 1916, 2015.
- [10] R. Girshick, “Fast R-CNN”, dalam *Proceedings of the IEEE International Conference on Computer Vision*, 2015, hal. 1440 – 1448.
- [11] S. Ren, K. He, R. Girshick, dan J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, hal. 1137 – 1149, 2017.
- [12] J. Redmon, S. Divvala, R. Girshick, dan A. Farhadi, “You only look once: Unified, real-time object detection”, dalam *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, hal. 779 – 788.
- [13] J. Redmon dan A. Farhadi, “YOLO9000: Better, faster, stronger”, dalam *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, hal. 6517 – 6525.
- [14] R. Huang, J. Pedoeem, dan C. Chen, “YOLO-LITE: A real-time object detection algorithm optimized for Non-GPU computers”, dalam *Proceedings of the IEEE International Conference on Big Data*, 2018, hal. 2503 – 2510.
- [15] J. Zhang, M. Huang, X. Jin, dan A. Li, “A real-time Chinese traffic sign detection algorithm based on modified YOLOv2”, *Algorithms*, vol. 10, no. 4, 127, 2017.
- [16] Z. Liu, Z. Chen, Z. Li, dan W. Hu, “An efficient pedestrian detection method based on YOLOv2”, *Mathematical Problems in Engineering*, vol. 2018, 2018.
- [17] S. Jupiyandi, F. R. Saniputra, Y. Pratama, M. R. Dharmawan, dan I. Cholissodin, “Pengembangan deteksi citra mobil untuk mengetahui jumlah

tempat parkir menggunakan CUDA dan modified YOLO”, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 6, no. 4, hal. 413 – 419, 2019.

[18] Y. Matsuda, H. Hoashi, dan K. Yanai, “Recognition of multiple-food images by detecting candidate regions”, dalam *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, 2012, hal. 25 – 30.