

ANALISIS KINERJA *CONTENT-BASED IMAGE RETRIEVAL* BERBASIS *AUGMENTED REALITY* MENGGUNAKAN METODE *YOU ONLY LOOK ONCE* UNTUK MAKANAN NASIONAL INDONESIA

PERFORMANCE ANALYSIS OF *CONTENT-BASED IMAGE RETRIEVAL* BASED ON *AUGMENTED REALITY* USING *YOU ONLY LOOK ONCE* FOR NATIONAL FOOD OF INDONESIA

Yudhistira Hadi Firdaus¹, Suryo Adhi Wibowo², Gelar Budiman³

^{1,2,3}Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

¹yudhistirahf@telkomuniversity.ac.id, ²suryoadhiwibowo@telkomuniversity.ac.id,

³gelarbudiman@telkomuniversity.ac.id

Abstrak

Teknologi terus dikembangkan oleh para peneliti dari seluruh belahan dunia. *Content-Based Image Retrieval* (CBIR) menjadi salah satu teknologi yang terus dikembangkan. CBIR merupakan teknologi yang dapat menampilkan koleksi gambar yang serupa berdasarkan masukan sebuah gambar kueri. Pada Tugas Akhir ini dirancang sebuah sistem CBIR yang digabungkan dengan teknologi *Augmented Reality* (AR) dan menggunakan metode *You Only Look Once* (YOLO) sebagai algoritma pendeteksi objek. Sebelumnya pencarian koleksi gambar dapat dilakukan dengan masukan teks atau bisa disebut *Text-Based Image Retrieval*. Namun pencarian gambar dengan teks tidak dapat dilakukan apabila nama serta definisi objek pada gambar tidak diketahui. Image retrieval dapat dilakukan secara *real-time* menggunakan input video dengan penggabungan kinerja CBIR, AR, dan YOLO. Tugas Akhir ini menerapkan CBIR dengan metode YOLO dengan harapan kinerja sistem dapat bekerja dengan optimal. Parameter kinerja terdiri dari *recall*, *presisi*, dan *F1 score*. Output yang didapatkan pada Tugas Akhir ini adalah sistem berhasil bekerja dengan nilai rata-rata tertinggi untuk *recall* sebesar 97,1%, untuk *presisi* sebesar 89%, dan untuk *F1 score* sebesar 92% dengan *dataset* berisikan lima kelas makanan nasional Indonesia.

Kata kunci : *Content-Based Image Retrieval*, *Augmented Reality*, YOLO, *Image Retrieval*, Makanan Nasional Indonesia.

Abstract

Technology continues to be developed by researchers from all over the world. *Content-Based Image Retrieval* (CBIR) is one of the technologies that continue to be developed. CBIR is a technology that can retrieve image collections that similar based on query image input. In this final project, the CBIR system that combined with *Augmented Reality* (AR) technology and uses *You Only Look Once* (YOLO) method as an object detection algorithm will be made. Previously, image retrieval can be done by text input or we can call it *Text-Based Image Retrieval*. However, image retrieval by text can not be done if the object's name or definition in the image is unknown. Image retrieval can be done in *real-time* using video input with the combined performance of CBIR, AR, and YOLO. This final project applying CBIR with the YOLO method in the hope that the performance of the system can optimally work. Performance parameters consist of *recall*, *precision*, and *F1 score*. The output obtained in this research is the system working with the result of the highest average value for a *recall* is 97.1%, for *precision* is 89%, and for *F1 score* is 92% with a *dataset* containing five Indonesian national foods.

Keywords: *Content-Based Image Retrieval*, *Augmented Reality*, YOLO, *Image Retrieval*, Makanan Nasional Indonesia.

1. Pendahuluan

Indonesia merupakan negara kepulauan yang memiliki banyak suku dan budaya. Setiap suku dan budaya di Indonesia memiliki berbagai macam ciri khasnya masing-masing, salah satunya dalam bidang makanan. Tidak hanya menarik perhatian wisatawan dalam negeri, makanan khas Indonesia juga mulai banyak diminati oleh wisatawan mancanegara. Pada 2019 sektor makanan dan minuman diyakini terus menjadi hal paling utama, terutama pada sektor industri pariwisata yang mengembangkan wisata kuliner. Pada 2018 Kementerian Perindustrian (Kemenperin) menyebut sektor makanan dan minuman berhasil menyumbang produk domestik bruto (PDB) nasional sebesar 6,34% [1]. Untuk memperkenalkan makanan khas yang ada di Indonesia kepada wisatawan mancanegara, bisa dilakukan dengan menampilkan sebuah koleksi gambar dari makanan yang ingin diperkenalkan. Sulit untuk wisatawan mancanegara yang hanya ingin melihat koleksi gambar makanan di

hadapannya namun tidak mengetahui apa nama makanan tersebut. Maka dari itu perlu dibuat sebuah sistem untuk menampilkan koleksi gambar berdasarkan masukan suatu gambar lain yang serupa. Pencarian melalui gambar telah banyak dilakukan oleh beberapa perusahaan teknologi. Contohnya adalah Google dan Microsoft [2]. Pencarian sekumpulan gambar yang serupa, diambil dari beberapa sumber, atau dengan sudut pandang yang berbeda-beda, berdasarkan satu gambar serupa lain yang diunggah. Proses ini dinamakan Content-Based Image Retrieval (CBIR) [3]. Dalam sistem CBIR, teknik pengolahan gambar digunakan untuk mengekstraksi fitur visual seperti warna, tekstur dan bentuk dari sebuah gambar. Sistem menggunakan model kueri untuk mengubah gambar menjadi representasi internal kueri, berdasarkan fitur yang diekstraksi dari gambar masukan. Model pencarian melakukan pencarian gambar dengan menghitung kemiripan antara gambar dalam objek dan gambar kueri [4]. Pengembangan CBIR sudah pernah dilakukan sebelumnya dengan menggunakan teknik *K-Nearest Neighbor* (KNN) [5] namun pada penelitian tersebut hanya menggunakan metode ekstraksi warna saja yang menyebabkan objek yang berbeda namun dengan warna dan bentuk yang mirip tetap akan ditampilkan. Penelitian menggunakan teknik *deep-learning* juga pernah dilakukan, yaitu dengan menggunakan teknik *Convolutional Neural Network* (CNN) [2], pada penelitian tersebut hasilnya sudah presisi sempurna, hanya saja kecepatannya yang perlu dikembangkan. *You Only Look Once* (YOLO) merupakan algoritma deep learning yang dikembangkan oleh Joseph Redmon dan Ali Farhadi [6]. Penulis akan merancang sistem CBIR menggunakan metode deep learning YOLO. Kinerja YOLO melampaui kecepatan maksimal dari Faster R-CNN dan merealisasikan keseimbangan antara kecepatan dan akurasi. YOLOv2 telah mencapai 76.8 mAP pada 67 FPS dan 78.6 mAP pada 40 FPS, melampaui metode Faster R-CNN dengan ResNet yang mencapai 76.4 mAP namun hanya pada 8 sampai 15 FPS saja [7]. YOLO diklaim sangat cocok dalam deteksi objek secara *real-time*. Maka dari itu yang diharapkan dari penggunaan YOLO dalam CBIR adalah terhubungnya antara dunia nyata dengan informasi dalam bentuk digital secara langsung pada saat itu juga. Hal tersebut dapat diwujudkan dengan adanya teknologi *Augmented Reality* (AR). AR menggunakan teknologi untuk membuat sebuah layer informasi yang dapat di akses orang-orang dalam satu persepsi antara dunia nyata dengan konten digital yang dihasilkan oleh *software* komputer [8]. Maka dari itu, dalam penelitian ini akan diuji analisis kinerja *Content-Based Image Retrieval* berbasis *Augmented Reality* menggunakan metode YOLO.

2. Dasar Teori dan Metodologi

2.1 Content-Based Image Retrieval

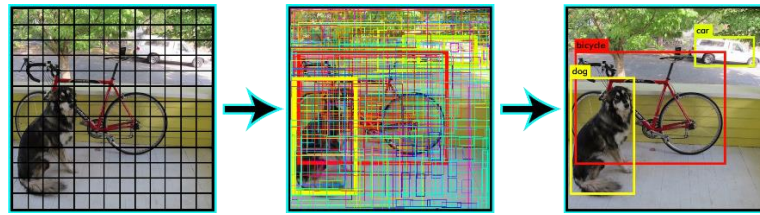
Content-Based Image Retrieval adalah sebuah teknologi untuk mencari koleksi gambar yang serupa dari sebuah tempat penyimpanan gambar yang isinya sangat banyak dan bervariasi berdasarkan konten visual gambar yang dicari. kata *content* dalam *Content-Based Image Retrieval* (CBIR) merepresentasikan deskripsi dari gambar menggunakan fitur gambar yang bervariasi [2]. Ekstraksi fitur visual primer dari gambar menjadi kunci utama dalam *Content-Based Image Retrieval* [9]. *Content-Based Image Retrieval* bekerja dengan cara mengekstraksi fitur visual dari gambar kueri. Kemudian fitur dari gambar kueri yang didapatkan dibandingkan dengan seluruh fitur dari gambar yang berada dalam tempat penyimpanan. Lalu seluruh gambar yang fitur visualnya mendekati gambar kueri akan ditampilkan sebagai keluaran.

2.2 Augmented Reality

Augmented Reality merupakan teknologi yang menggabungkan dunia nyata dengan dunia virtual. *Augmented Reality* diciptakan agar manusia dapat memandang dunia nyata dan virtual dengan satu persepsi. dengan adanya *Augmented Reality* pengguna dapat berinteraksi dengan dunia nyata dan virtual secara langsung. Sistem *Augmented Reality* memiliki tiga langkah simpel: *Recognition*, *Tracking*, dan *Mix*. Pada saat *recognition*, setiap gambar, objek, wajah, tubuh atau ruang dikenali untuk memilih objek virtual apa yang akan ditumpangkan. Ketika *tracking* secara real-time pada ruang dari gambar, objek, wajah, atau tubuh berlangsung, media dalam bentuk video, objek 3D, objek 2D, teks, dan lain lain ditumpangkan ke ruang tersebut [10].

2.3 You Only Look Once

YOLO atau *You Only Look Once* sesuai namanya secara teknik adalah sebuah metode object-recognition yang melakukan CNN secara penuh hanya sekali. YOLO kemudian setelah melewati seluruh gambar melalui jaringan memprediksi boundary box lalu menghitung probabilitas dari label objek. deteksi objek dilakukan dengan regresi [11]. Semua proses diambil oleh convolutional neural network memprediksi banyak bounding boxes dan probabilitas kelas sengan serentak sekali jalan. Pada YOLO, gambar dilihat secara keseluruhan. Seperti saat melihat konteks yang lebih besar, banyaknya latar belakang kesasalahan dikurangi [11].



Gambar 1. Model *You Only Look Once* [6].

2.4 Batch Size

Batch size adalah *hyperparameter* yang dapat ditentukan oleh pengguna untuk mengatur banyaknya sample *dataset* yang akan dimasukkan dalam sekali *training*. Misalkan terdapat 512 *dataset*, lalu *batch size* diatur sebesar 64, maka sistem akan mengambil 64 sample dari *dataset* untuk di-*train*. Lalu setelah selesai, sistem akan mengambil lagi 64 *dataset* lainnya untuk *training* berikutnya. Semakin besar ukuran *batch size* maka semakin besar memori yang dibutuhkan dan waktu yang diperlukan lebih lama untuk sekali *training* tergantung dari perangkat yang digunakan.

2.5 Learning Rate

Learning rate merupakan salah satu *hyperparameter* yang digunakan untuk mengatur *weight* dari *neural network* untuk mengurangi *loss gradient*. Hal ini menentukan seberapa cepat atau lambatnya proses *training* demi mendapat *weight* yang optimal. *Learning rate* yang terlalu besar mengakibatkan proses *training* yang lebih cepat namun ada kemungkinan hasil yang optimal akan terlewatkan. *Learning rate* yang terlalu kecil akan sangat memperlambat proses *training* untuk mendapatkan hasil yang optimal.

2.6 Step Training

Step training bisa disebut juga iterasi. Satu iterasi berarti 1 *batch dataset* telah melalui proses *training* dalam *neural network*. Contohnya apabila terdapat 512 citra *dataset*, lalu diatur *batch size*-nya sebesar 64, maka perlu 8 iterasi untuk melatih seluruh citra pada *dataset*. Apabila *step* terlalu banyak maka akan terjadi *over-fitting* yang berarti sistem akan menghafal data bukan mengenali data. Sebagai contoh ketika system mempelajari fitur kucing dan terjadi *over-fitting*, maka ada kemungkinan sistem hanya mengenali kucing dengan warna tertentu saja tidak dengan warna lainnya. Apabila *step* terlalu sedikit maka sistem belum memiliki cukup pengetahuan tentang fitur dari object yang dilatih.

2.7 Perhitungan Parameter Performansi

Parameter performansi digunakan sebagai tolak ukur kinerja suatu sistem. Adapun parameter performansi yang digunakan sebagai acuan dalam tugas akhir ini adalah akurasi, presisi dan waktu komputasi. Secara umum performansi dalam tugas akhir ini diukur berdasarkan hasil yang didapatkan dari proses *training* dan prediksi *dataset* uji dengan *confusion matrix* digunakan sebagai acuan untuk menghitung akurasi dan presisi.

a. Recall

Recall berhubungan dengan kemampuan sistem untuk menampilkan jenis makanan yang relevan pada basisdata dengan perhitungan,

$$R = \frac{A}{B} \quad (1)$$

dimana R adalah recall, A adalah jumlah gambar relevan yang ditampilkan, dan B adalah jumlah gambar relevan pada basisdata.

b. Presisi

Presisi dalam CBIR merupakan kecocokan antara gambar relevan yang sistem tampilkan dengan data yang sistem tampilkan dengan perhitungan,

$$P = \frac{A}{C} \quad (2)$$

dimana P adalah presisi, A adalah jumlah gambar relevan yang ditampilkan, dan C adalah jumlah gambar yang ditampilkan.

c. **F1 Score**

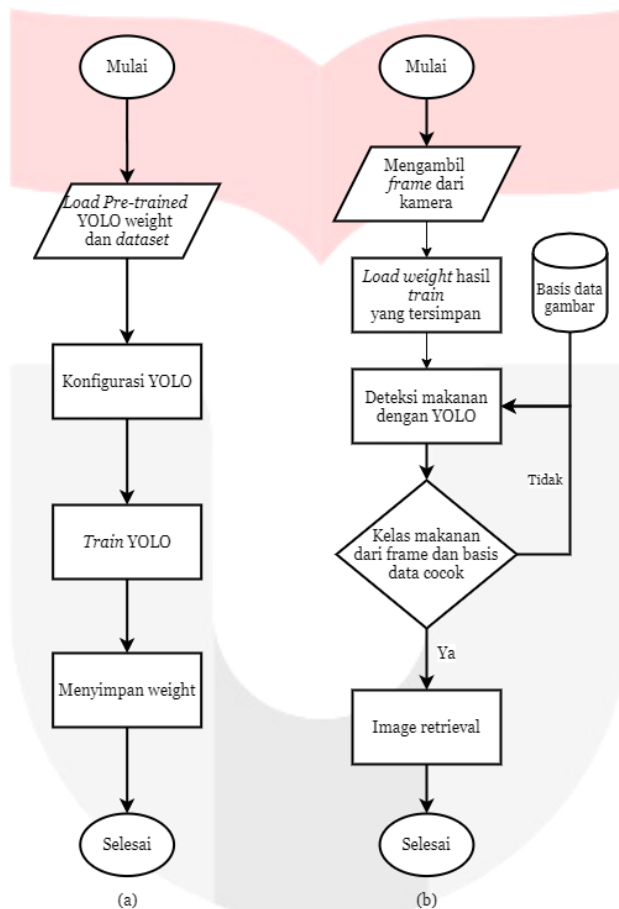
F1 Score merupakan rata-rata harmonis dari *recall* dan presisi, pada umumnya recall yang baik belum tentu presisinya juga baik, begitu juga sebaliknya. Maka dari itu akan dihitung juga nilai F1 score pada penelitian kali ini.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{3}$$

dimana F1 adalah *F1 score*, P adalah presisi, dan R adalah *recall*.

3. **Perancangan dan Pembentukan Model Sistem**

Dalam penelitian kali ini dibuat desain sistem *Content-Based Image Retrieval* dengan masukan berupa video. Skema dari penelitian ini menggunakan algoritma YOLO yang arsitekturnya terinspirasi dari arsitektur CNN GoogLeNet. YOLO digunakan untuk mendeteksi makanan pada citra. Alur kerja sistem dapat dilihat pada Gambar 2.



Gambar 2. Diagram alir perancangan sistem *Content-Based Image Retrieval* makanan menggunakan YOLO (a) Pelatihan dan (b) Pengujian.

3.1 **Load dan Resize Dataset**

Pre-Trained weight pada YOLO merupakan *weight* yang sudah dilatih dan memiliki 80 kelas. Pada penelitian kali ini digunakan *dataset* yang memiliki 5 kelas yaitu makanan nasional Indonesia. *Pre-trained weight* dari YOLO digunakan untuk *transfer learning* dengan tujuan mempermudah proses pelatihan menggunakan pengetahuan dari proses *learning network* sebelumnya untuk melatih 5 kelas baru dari makanan nasional Indonesia. *Dataset* yang digunakan adalah citra dari makanan nasional Indonesia yang telah ditentukan Kementerian Pariwisata Indonesia yaitu rendang, soto, sate, nasi goreng, gado-gado seperti pada gambar 3. *Dataset* diambil dari campuran antara tangkapan video dari penulis sendiri yang diambil per-frame-nya dan citra dari internet dipilih secara acak. Rincian *dataset* dapat dilihat pada tabel 1.

Tabel 1. Sistematika Pembagian Data

Kelas	Data latih	Data Validasi	Data Uji
Rendang	240	60	100

Soto	240	60	100
Sate	240	60	100
Nasi Goreng	240	60	100
Gado-gado	240	60	100



Gambar 3. Citra makanan : (a) Gado-gado, (b) Nasi Goreng, (c) Rendang, (d) Sate, (e) Soto.

3.2 Konfigurasi Hyperparameter

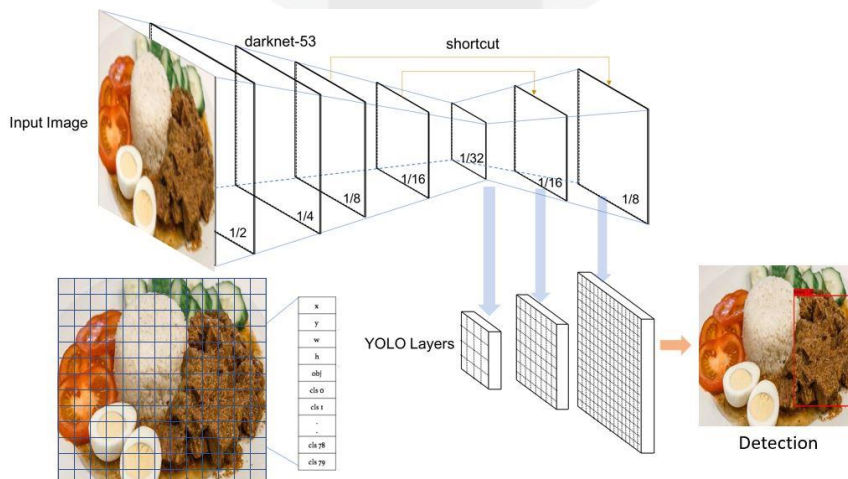
Konfigurasi *hyperparameter* yang diterapkan meliputi pengaturan pada jumlah *step*, *batch size* (bs), serta *learning rate* (lr). Terdapat sejumlah 9 skenario konfigurasi model yang dirincikan pada tabel 2.

Tabel 2. Skema konfigurasi *hyperparameter*.

Model	lr	bs	Step
A	0.001	16	4000, 5000, 6000
B	0.002	16	4000, 5000, 6000
C	0.003	16	4000, 5000, 6000
D	0.001	32	4000, 5000, 6000
E	0.002	32	4000, 5000, 6000
F	0.003	32	4000, 5000, 6000
G	0.001	64	4000, 5000, 6000
H	0.002	64	4000, 5000, 6000
I	0.003	64	4000, 5000, 6000

3.3 Proses YOLO

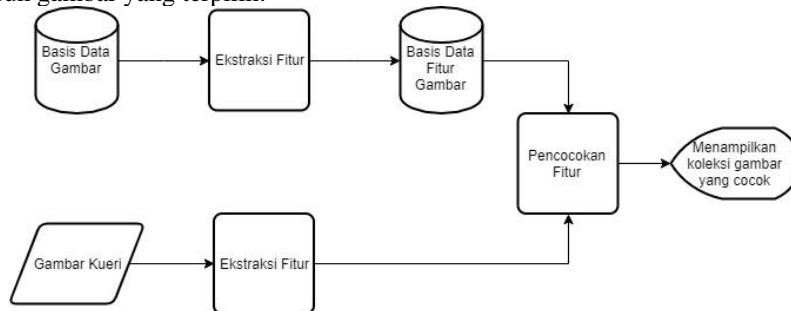
YOLO bekerja dengan awalan mengubah citra masukan menjadi 416×416 . Lalu dalam sistem ini menggunakan Darknet-53 sebagai pengekstraksi fitur dan pendeteksi objek pada citra. Kemudian Non-Max Suppression digunakan agar munculnya *bounding box* dibatasi dengan hanya memunculkan *bounding box* sesuai dengan *threshold* yang sudah ditentukan.



Gambar 4. Hasil deteksi YOLO (a) Gado-gado, (b) Nasi Goreng, (c) Rendang, (d) Sate, (e) Soto.

3.4 Image Retrieval

YOLO akan melakukan deteksi makanan pada hasil *capture frame-to-frame* dari kamera dan citra pada basisdata, setelah itu akan dilakukan *image retrieval*. Pada hasil *capture frame-to-frame* dari kamera akan terdeteksi makanan dengan prediksi salah satu kelas dari tabel 1 dengan *final score* diatas 20%. Setelah itu sistem akan memilih citra pada basisdata yang terprediksi memiliki kelas yang sama dengan citra pada *frame* kamera. Lalu sistem akan menampilkan seluruh gambar yang terpilih.



Gambar 5. Flowchart Content-Based Image Retrieval.

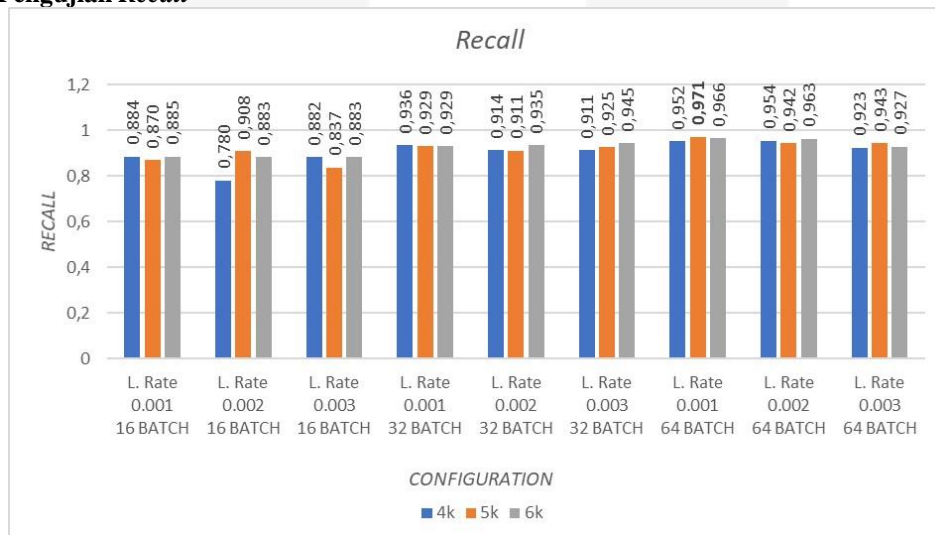
4. Pengujian Model Sistem

Pengujian dilakukan terhadap 500 citra uji yang dibagi untuk 5 kelas menjadikan 100 citra untuk tiap kelasnya. Skenario pengujian diperlukan untuk dianalisa dan kemudian dicari konfigurasi yang terbaik. Berikut ini adalah tiga skenario yang dilakukan :

1. Skenario 1: Pengujian terhadap parameter *recall*.
2. Skenario 2: Pengujian terhadap parameter presisi.
3. Skenario 3: Pengujian terhadap parameter F1 score.

Pada bagian ini diperlihatkan hasil keseluruhan dari pengujian sistem menggunakan konfigurasi batch size 16, 32, dan 64, dengan masing-masing *learning rate* 0.001, 0.002, dan 0.003. Setiap konfigurasi telah dilatih hingga step 4000, 5000, dan 6000. Hasil yang dipaparkan merupakan hasil rata-rata dari keseluruhan 5 kelas makanan.

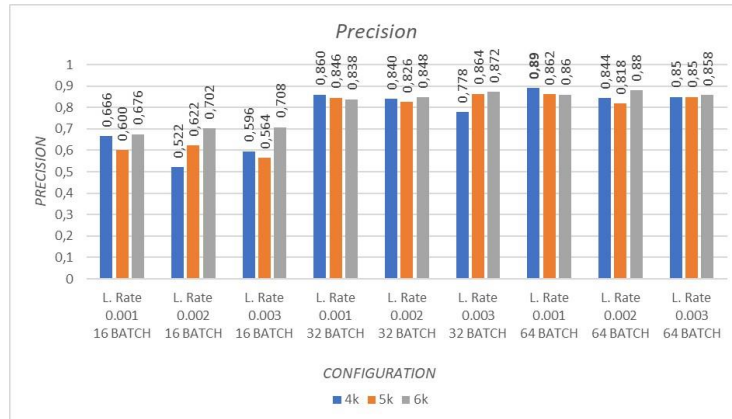
4.1 Hasil Pengujian Recall



Gambar 6. Hasil skenario 1.

Pada Gambar 6 memperlihatkan grafik hasil pengujian sistem pada parameter recall. Dari grafik diketahui bahwa nilai recall tertinggi adalah 97,1%. Nilai recall tertinggi terdapat pada konfigurasi *learning rate* 0.001, *batch size* 64, pada step 5000. Hal tersebut menunjukkan bahwa citra tidak relevan yang terambil dengan konfigurasi tersebut paling sedikit yang menjadikan konfigurasi tersebut paling efisien. Dari grafik juga kita dapat mengetahui bahwa nilai recall semakin meningkat dari batch size 16 ke batch size 64. Hal ini dikarenakan semakin besar *batch size*, citra yang dilatih ke dalam sekali *neural network* semakin banyak sehingga fitur yang didapat lebih rinci.

4.2 Hasil Pengujian Presisi



Gambar 7. Hasil skenario 2.

Hasil pengujian sistem pada parameter presisi dapat dilihat pada Gambar 7 dalam bentuk grafik. Hasil pengujian dengan nilai presisi tertinggi berada pada konfigurasi *learning rate* 0.001, *batch size* 64, pada *step* 4000. Nilai terbaik pada parameter presisi memiliki konfigurasi yang berbeda dari konfigurasi terbaik pada parameter *recall*. Hal tersebut menunjukkan bahwa nilai *recall* tidak memiliki hubungan langsung dengan nilai presisi. Dapat diperhatikan bahwa pengaruh *learning rate* terhadap nilai *recall* fluktuatif atau tidak stabil. *Learning rate* yang terlalu tinggi menyebabkan sistem akan lebih mengalami *loss* yang tidak stabil, namun apabila terlalu rendah, sistem perlu waktu yang jauh lebih banyak agar mendapatkan hasil yang optimal dan akan muncul kemungkinan memori perangkat bisa saja *overload* sebelum hasil optimal didapatkan.

4.3 Hasil Pengujian F1 Score



Gambar 8. Hasil skenario 3.

Grafik hasil pengujian sistem pada parameter F1 score dapat dilihat pada Gambar 8. Pada grafik terlihat konfigurasi dengan *F1 score* terbaik berada pada konfigurasi *learning rate* 0.001, *batch size* 64, pada *step* 4000. Hal ini menunjukkan bahwa semakin besar nilai presisi dan *recall*. Semakin besar juga *F1 score*. Semakin *F1 score* mendekati 1 maka kinerja sistem semakin baik. Parameter *F1 score* ini merupakan rata-rata harmonis dari dua parameter sebelumnya. Hal ini menunjukkan konfigurasi dengan nilai *F1 score* tertinggi sudah pasti memiliki nilai *recall* dan presisi yang tinggi dan seimbang. Maka konfigurasi terbaik pada penelitian ini diambil dari konfigurasi yang memiliki nilai *F1 score* tertinggi yaitu pada konfigurasi *batch size* 64, *learning rate* 0.001 pada *step* 4000

5. Kesimpulan

Analisis kinerja *Content-Based Image Retrieval* dengan metode *You Only Look Once* berbasis *Augmented Reality* telah berhasil dilakukan pada lima kelas makanan nasional Indonesia dengan hasil rata-rata nilai yang tertinggi untuk *recall* sebesar 97,1%, untuk presisi sebesar 89%, dan untuk *F1 score* sebesar 92%. Dari hasil pengujian didapatkan konfigurasi terbaik dengan pengaturan *batch size* 64, *learning rate* 0.001, pada *step* 4000. Semakin besar konfigurasi *batch size*, maka semakin tinggi nilai *recall*, presisi, dan *F1 score*. Untuk konfigurasi *learning rate* hasilnya terlihat meningkat bersamaan dengan *epoch* hanya pada *batch* 32 saja, pada *batch* lain hasilnya cenderung tidak stabil. Hal ini dikarenakan tiap makanan sebenarnya memiliki konfigurasi terbaiknya

masing-masing untuk tiap parameter, namun untuk memilih konfigurasi yang terbaik untuk semuanya sekaligus maka diambil rata-rata dari keseluruhan kelas.

6. Daftar Pustaka:

- [1] "Industri Kuliner Topang Ekonomi Nasional," Sindonews.com, 2019. [Online]. Available: <https://ekbis.sindonews.com/read/1368165/34/industrikuliner-topang-ekonomi-nasional-1546736372>. [Accessed 24 Oktober 2019].
- [2] Z. Rian, V. Christianti and J. Hendryli, "Content-Based Image Retrieval using Convolutional Neural Networks," p. 1, 2019.
- [3] R. Ashraf, M. Ahmed, S. Jabbar, S. Khalid, A. Ahmad, S. Din and G. Jeon, "Content Based Image Retrieval by Using Color Descriptor and Discrete Wavelet Transform," *J Med Syst*, p. 1, 2018.
- [4] J. WAN, D. WANG, S. C. H. HOI, P. WU and J. ZHU, "Deep learning for content-based image retrieval: A comprehensive study," p. 158, 2014.
- [5] D. A and R. Longadge, "Content Based Image Retrieval using Color Feature Extraction with KNN Classification," *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 5, p. 1275, 2014.
- [6] Redmon, Joshep and F. Ali, "YOLO9000: Better, Faster, Stronger," *arXiv:1612.08242*, 2016.
- [7] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," 2019.
- [8] M. Farshid, J. Paschen, T. Eriksson and J. Kietzmann, "Go boldly!: Explore augmentedreality(AR),virtualreality(VR),andmixedreality(MR)forbusiness," *Business Horizons*, vol. 61, no. 5, p. 659, 2018.
- [9] G. -H. Liu, J. -Y. Yang and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recognit.* 48, pp. 2554-2566, 2015.
- [10] A. Dihraj and G. Sharvari, "COMPARATIVE STUDY OF AUGMENTED REALITY SDK'S," *International Journal on Computational Sciences & Applications*, vol. 5, no. 1, 2015.
- [11] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, " You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.