COMPLETE
Journal of Computer, Electronic, and Telecommunication
ISSN 2723-4371, E-ISSN 2723-5912
doi.org/10.52435/complete.v1i1.51

Investigasi Parameter *Epoch* Pada Arsitektur ResNet-50 Untuk Klasifikasi Pornografi

Faiz Nashrullah [1], Suryo Adhi Wibowo [2]* dan Gelar Budiman [3]

[1] Telkom University; faissme@student.telkomuniversity.ac.id
[2] Telkom University; suryoadhiwibowo@telkomuniversity.ac.id
[3] Telkom University; gelarbudiman@telkomuniversity.ac.id
* Korespondensi: suryoadhiwibowo@telkomuniversity.ac.id; Tel.: +6282145143639

Abstrak: Kemajuan teknologi informasi yang cepat dan tak terkontrol membuat berbagai konten negatif seperti pornografi dapat dengan mudah diakses. Konten pornografi terbukti dapat menyebabkan berbagai permasalahan terutama pada generasi muda. Beberapa pengembangan metode untuk pendeteksian pornografi telah dilakukan namun masih terkendala pada keterbatasan karakteristik data masukan. Pada penelitian kali ini dikembangkan sistem pendeteksi konten pornografi berbasis klasifikasi menggunakan *Convolutional Neural Network* (CNN) dengan arsitektur ResNet-50 untuk mengatasi permasalahan sebelumnya. Dalam proses perancangan model sistem, diterapkan berbagai konfigurasi *epoch* dan didapatkan bahwa performa dari sistem memiliki kecenderungan untuk meningkat seiring dengan pertambahan *epoch*. Performa terbaik diraih oleh sistem pada konfigurasi *epoch* 60 dengan akurasi 91,033%.

Kata kunci: CNN; ResNet-50; ResNet; Klasifikasi; *Hyperparameter*; *Epoch*

1. Pendahuluan

Kemajuan teknologi yang sangat pesat terutama dalam teknologi penyebaran informasi terkhusus internet, memberikan aksesibilitas yang mudah dan cepat kepada semua orang. Keuntungan tersebut dapat memberikan dampak yang buruk jika terjadi penyalahgunaan yang dilakukan oleh beberapa pihak dengan cara menyebarkan konten negatif, salah satunya pornografi. Tidak dapat dipungkiri bahwa konten pornografi memiliki persebaran yang sangat masif dan cepat di internet. Definisi dari pornografi adalah materi yang menciptakan pikiran seksual, berisi gambar eksplisit atau deskripsi tindakan seksual yang melibatkan alat kelamin [1]. Pornografi dapat menimbulkan kecanduan dan terdapat ulasan dari berbagai penelitian [2] yang menunjukkan bahwa pornografi dapat mengganggu kesehatan perkembangan seksual pada anak modern.

Menanggapi permasalahan tersebut, diperlukan metode deteksi atau *filtering* terhadap konten pornografi, yang kemudian dapat dikembangkan menjadi piranti lunak penangkal pornografi, serta bisa diterapkan di berbagai *platform*. Dalam masa awal penelitian tentang konten pornografi, Basilio *et al* [3] menggunakan metode *thresholding* untuk membuat metode pendeteksi citra pornografi dengan memperhatikan banyaknya komposisi kulit dalam suatu citra. Tetapi penentuan konten pornografi yang mengacu pada komposisi warna kulit memiliki tingkat akurasi yang rendah jika *background* dari objek dalam citra mempunyai komposisi warna yang serupa dengan kulit asli manusia, bisa disebut juga dengan *noise*. Tidak semua warna kulit juga dapat terdeteksi dikarenakan hanya warna kulit yang berada di antara ketentuan *threshold* saja yang terdeteksi. Karena itu metode yang dirancang [3] tidak terlalu cocok untuk menangani data dengan *noise* yang banyak sehingga terbatas dalam segi karakteristik citra pornografi yang dapat terdeteksi.

Berdasar penelitian yang telah dilakukan sebelumnya, pada penelitian ini dikembangkan sebuah metode klasifikasi untuk mendeteksi suatu citra pornografi dengan menggunakan metode *Convolutional Neural Network* (CNN) yang merupakan salah satu contoh *deep learning*. Penerapan

berbagai metode *deep learning* mulai digunakan untuk melakukan klasifikasi citra ataupun berbagai pengolahan video berbasis *computer vision* [4]–[7]. CNN sendiri mengadopsi cara kerja dan memiliki struktur tiruan dari neuron otak manusia. Dengan model struktur tersebut, CNN mampu melakukan ekstraksi fitur dari citra secara lebih mendetail, serta CNN dapat menyimpan dan menjadikan acuan hasil pembelajaran yang telah dilakukan dengan adanya parameter *weight* dan *bias* di dalam arsitekturnya. Dengan konsep tersebut CNN dapat menerima *input* citra dengan karakteristik lebih umum atau dengan banyak *noise* sekalipun.

Struktur CNN secara umum disusun oleh tiga komponen utama *layer*, yaitu *convolutional layer*, *pooling layer* dan *full connection layer* [8]. Ketiga *layer* tersebut dapat dikombinasikan sesuai dengan tujuan dan kebutuhan, sehingga tercipta berbagai arsitektur dari CNN dengan keunggulannya masing-masing. Pada penelitian ini, sistem klasifikasi dibuat dengan menggunakan ResNet-50 [9] sebagai model arsitektur dari CNN. ResNet-50 memiliki konsep *shortcut connection* yang mencegah sistem dari kehilangan banyak informasi selama *training* dilakukan. Sistem klasifikasi yang dibuat dalam penelitian ini ditujukan untuk dapat mendeteksi apakah suatu gambar yang menjadi *input* dalam sistem merupakan pornografi atau bukan.

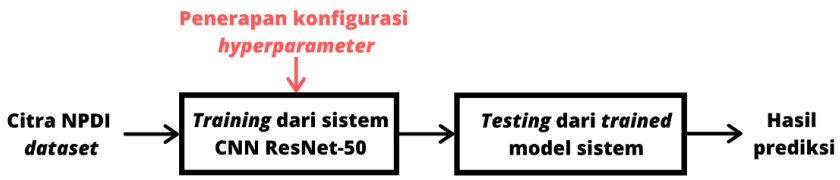
Dalam penelitian ini dilakukan investigasi dari pengaruh penerapan konfigurasi *hypeparameter epoch* ke dalam arsitektur yang dibuat untuk sistem klasifikasi pornografi. *Epoch* adalah suatu kondisi ketika seluruh *dataset* yang menjadi masukan model sistem telah selesai melewati seluruh neural network secara *forward propagation* dan *back propagation* dalam 1 kali putaran [10]. Diharapkan berdasarkan hasil percobaan yang dilakukan, dapat diketahui pengaruh penerapan konfigurasi *epoch* terhadap akurasi yang didapatkan sistem klasifikasi.

2. Bahan dan Metode

Penelitian yang dilakukan diawali dengan merancang sebuah sistem klasifikasi gambar pornografi menggunakan CNN dengan ResNet-50 sebagai arsitekturnya. Dalam perancangan yang dilakukan, NPDI *dataset* [11] dipilih menjadi data citra masukan untuk proses *training* dan *testing* dari percobaan yang dilakukan. Sebelum dilakukan proses *training*, *hyperparameter* seperti *epoch*, *learning rate*, dan *batch size* diatur terlebih dahulu. Secara umum garis besar dari perancangan sistem ditunjukkan oleh Gambar 1.

2.1 NPDI *dataset*

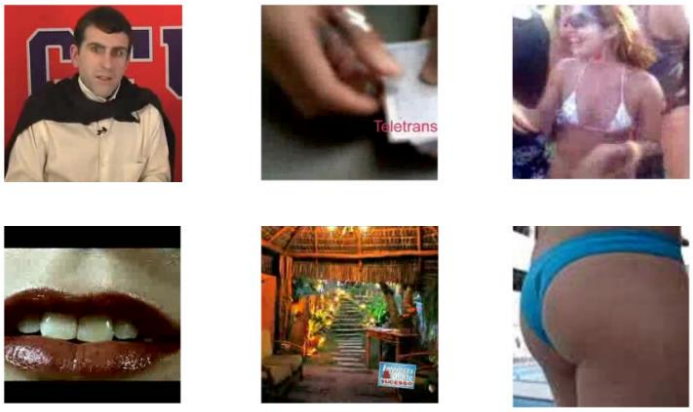
Pada penelitian kali ini digunakan NPDI *dataset* [11] sebagai data masukan untuk sistem yang dirancang. NPDI *dataset* berisikan video yang berdurasi hampir 80 jam yang terdiri dari 400 video pornografi dan 400 video non-pornografi. Untuk keperluan *data preprocessing*, video dalam *dataset* tersebut diekstrak menggunakan STOIK *Video Converter* dan menghasilkan 16727 *frames* yang sudah dipilih dari total semua frames ekstraksi. Konten dalam *dataset* ini juga terdiri dari variasi berbagai macam etnis di dalamnya. Dari total keseluruhan *frames* yang diekstrak, hanya digunakan sebagian *dataset* dan kemudian dilakukan augmentasi pada *dataset* tersebut sehingga total didapatkan 15000 *frames* atau citra untuk keperluan penelitian ini. *Dataset* NPDI diorganisir sesuai yang tertera pada Tabel 1 dan Gambar 2 merupakan *preview* dari citra NPDI *dataset*.



Gambar 1. Diagram garis besar perancangan sistem.

Tabel 1. Sistematika pembagian data.

Kelas	Data Latih	Data Validasi	Data Uji
Pornografi	6000	600	1500
Non-Pornografi	6000	600	1500



Gambar 2. *Preview* dari NPDI *dataset* [11].

2.2 *Convolutional Neural Network*

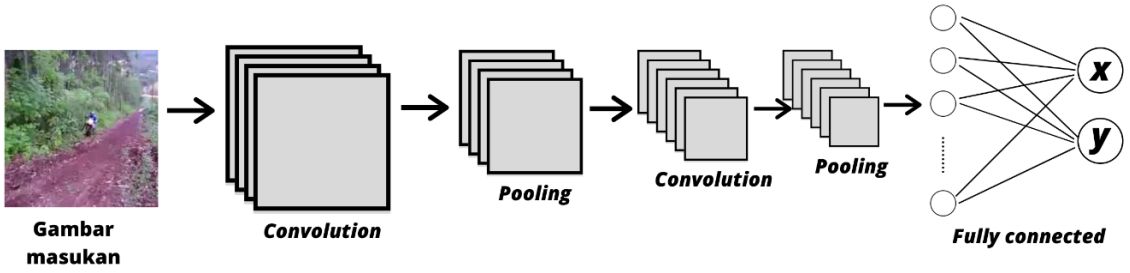
Secara garis besar, struktur dari CNN terdiri dari 3 macam *layer* penyusun utama yaitu, *convolutional layer*, *pooling layer*, dan *fully connection* atau *fully connected* layer. Struktur pada arsitektur CNN biasanya terdiri dari berbagai tingkatan kedalaman *network* yang tiap-tiap struktur merepresentasikan fitur tersendiri [8]. Pada *convolutional layer* dilakukan proses konvolusi untuk mengekstrak fitur-fitur yang terdapat pada citra, proses konvolusi secara sederhana dirumuskan dalam persamaan 1,

$$h(x) = f(x) * g(x), \tag{1}$$

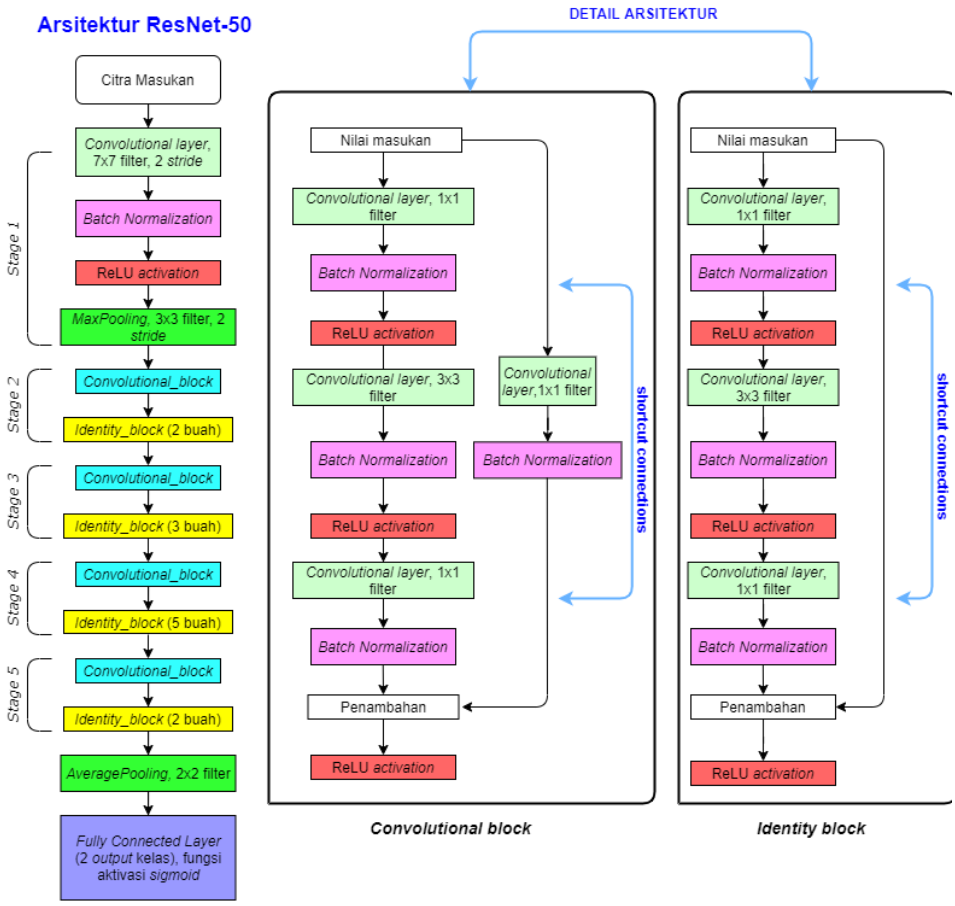
dari persamaan 1, $h(x)$ adalah *output* matriks hasil konvolusi atau dapat disebut *feature map*, $f(x)$ adalah matriks dari citra *input* dan $g(x)$ adalah *kernel* atau filter. Setelah konvolusi selesai dilakukan, kemudian *feature map* keluaran tersebut direduksi dengan menggunakan *pooling layer*. Terakhir setelah melewati proses konvolusi dan *pooling*, nilai-nilai hasil perhitungan *layer* sebelumnya diteruskan ke *fully connected layer* untuk dilakukan proses prediksi guna menghasilkan *output* kelas. Arsitektur CNN secara umum ditampilkan pada Gambar 3.

2.3 ResNet-50

ResNet-50 [9] merupakan salah satu arsitektur dari CNN yang memperkenalkan sebuah konsep baru yaitu *shortcut connections*. Munculnya konsep *shortcut connections* yang ada pada arsitektur ResNet-50 memiliki keterkaitan dengan *vanishing gradient problem* yang terjadi ketika usaha memperdalam struktur suatu *network* dilakukan. Bagaimanapun memperdalam suatu *network* dengan tujuan meningkatkan performansinya tidak bisa dilakukan hanya dengan cara menumpuk *layer*. Semakin dalam suatu *network* dapat memunculkan *vanishing gradient problem* yang bisa membuat *gradient* menjadi sangat kecil yang berakibat pada menurunnya performansi atau akurasi [9].



Gambar 3. Arsitektur CNN secara umum.



Gambar 4. Diagram blok arsitektur ResNet-50.

Karena itu ResNet memperkenalkan konsep *shortcut connections* dan dalam konsep ini fitur yang merupakan input dari layer sebelumnya juga dijadikan sebagai input terhadap output dari layer tersebut. Cara ini dilakukan sebagai solusi untuk meminimalisir hilangnya fitur-fitur penting pada saat proses konvolusi. Secara keseluruhan ResNet-50 terdiri dari 5 *stage* proses konvolusi yang kemudian dilanjutkan *average pooling* dan diakhiri dengan *fully connected layer* sebagai *layer* prediksi. Gambar 4 merupakan diagram blok susunan dari arsitektur ResNet-50 dalam penelitian ini.

Terdapat perbedaan antara arsitektur ResNet-50 yang digunakan dalam penelitian ini dengan arsitektur aslinya [9]. Hal ini dikarenakan terdapat sedikit modifikasi yang dilakukan untuk menyesuaikan dengan kebutuhan sistem klasifikasi. Modifikasi yang dilakukan antara lain:

- Mengganti jumlah *output* pada *fully connected layer* dari 1000 kelas menjadi 2 kelas.
- Mengganti fungsi aktivasi pada *fully connected layer* dari *softmax* menjadi *sigmoid*.

2.3.1 Proses *Training* ResNet-50

Training dilakukan untuk membentuk model dari ResNet-50 yang kemudian digunakan sebagai sistem klasifikasi. Sebelum *training* dijalankan, terdapat beberapa konfigurasi *hyperparameter* yang harus diatur terlebih dahulu. *Hyperparameter* yang dikonfigurasi antara lain, *learning rate*, *batch size* dan 5 macam konfigurasi *epoch* untuk keperluan investigasi, tertera pada Tabel 2.

Tabel 2. Konfigurasi *hyperparameter*.

Learning rate	*Batch Size*	*Epoch*
0.001	32	20, 40, 60, 80, 100

Sesuai Tabel 2, proses *training* dibagi menjadi 5 tahapan *epoch* dimulai dari 20 sampai dengan 100. Tahap pertama dilakukan proses *load dataset* NPDI sebanyak 12000 citra ke dalam sistem sebagai data masukan untuk *training*. Dikarenakan jumlah *sample* data cukup besar, selanjutnya citra 3 dimensi tersebut diperkecil menjadi resolusi 64×64 agar komputasi dapat berjalan lebih cepat. Terdapat dua kelas dari *dataset* yang dijadikan *input* dalam proses *training*, yaitu pornografi dan non-pornografi. Selanjutnya proses ekstraksi fitur dilakukan dari *stage* 1 sampai dengan 5.

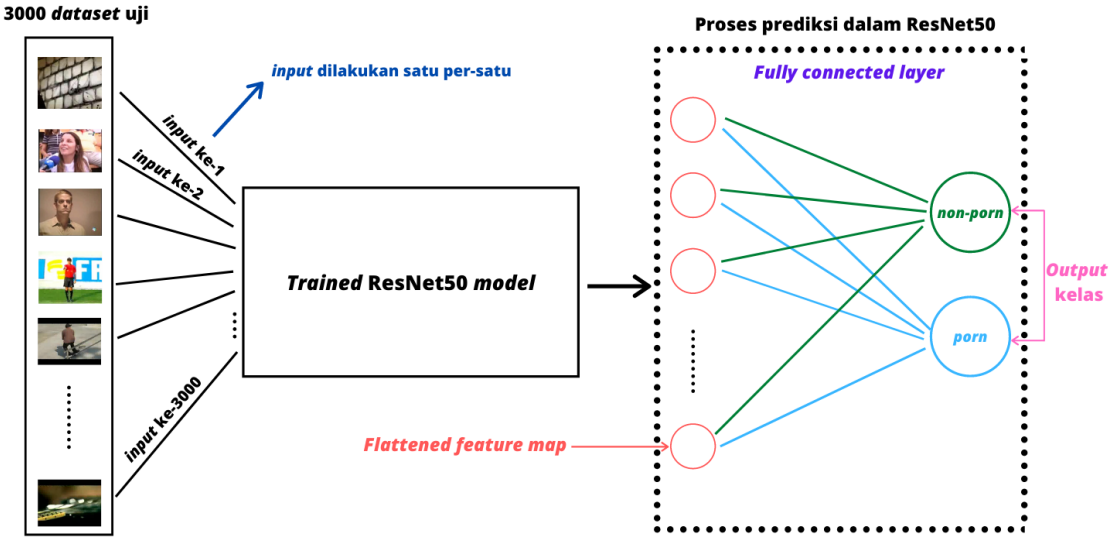
Pada *stage 1* citra dikonvolusi oleh *convolutional layer* dengan ukuran filter 7×7 dan *stride* 2. Konvolusi tersebut menghasilkan *feature map* yang kemudian dinormalisasi oleh *Batch Normalization*. Normalisasi yang dilakukan bertujuan untuk membuat nilai-nilai *output* dari *layer* sebelumnya yang sangat bervariasi menjadi 1 skala yang sama, hal ini dapat mempercepat proses *training* karena beban komputasi berkurang. Hasil dari normalisasi selanjutnya masuk ke *layer* aktivasi, didalamnya terdapat ReLU atau *Rectified Linear Unit activation function* yang digunakan untuk membuat hasil dari ekstraksi fitur menjadi *non-linear*. Selanjutnya nilai keluaran dari fungsi aktivasi direduksi oleh *maxpooling layer* sebelum diteruskan pada *stage* 2.

Diantara *stage* 2 sampai dengan *stage* 5, proses ekstraksi fitur dilakukan oleh kombinasi susunan antara *convolutional block* dan *identity block*. Pada blok-blok tersebut, citra mengalami pereduksian dan peningkatan dimensi. Hal ini disebabkan oleh tipe arsitektur pada kedua blok tersebut terdiri dari tiga tahapan konvolusi dengan ukuran filter 1×1, 3×3 dan 1×1. Tipe arsitektur blok seperti ini dikenal dengan sebutan *'bottleneck'* dan ditujukan untuk menghemat waktu komputasi. Setelah proses ekstraksi fitur selesai kemudian *feature map* diproses ke dalam *fully connected layer* untuk melakukan proses prediksi dibantu dengan fungsi aktivasi *sigmoid*. Total keseluruhan parameter *training* yang dihasilkan ketika melakukan *training* adalah 23,589,761 parameter.

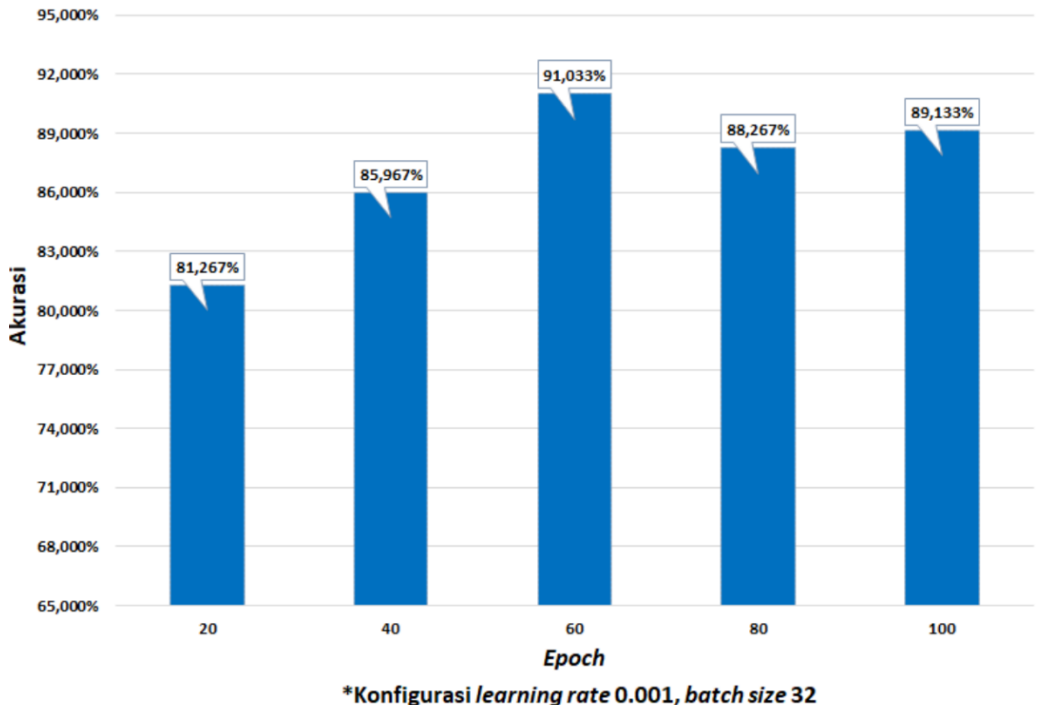
3. Hasil

Untuk menguji model sistem yang sudah dilatih pada tahap sebelumnya, dilakukan eksperimen pengetesan sistem terhadap 3000 citra *dataset* uji. Keseluruhan citra *dataset* uji dipilih secara acak agar mampu merepresentasikan seluruh karakteristik citra yang terdapat dalam NPDI *dataset*. Jika digambarkan, proses pengujian yang dilakukan untuk mendapatkan hasil performa sistem dapat dilihat pada Gambar 5.

Sebanyak 3000 *dataset* yang menjadi masukkan untuk sistem secara satu-persatu di *load* ke sistem. Lalu dilakukan proses prediksi dengan menggunakan model sistem ResNet-50 yang telah dilatih. Setelah mengalami proses ekstraksi fitur dengan bantuan *convolutional layer* dan *pooling layer*, nilai-nilai ekstraksi tersebut kemudian dirubah dimensinya menggunakan fungsi *flattening* agar bisa di proses ke *fully connected layer*. Dalam *fully connected layer* dilakukan proses prediksi yang menempatkan suatu citra masukan ke dalam kategori kelas tertentu. Hasil keseluruhan dari proses prediksi yang dijalankan tertera pada Gambar 6.



Gambar 5. Skema pengujian sistem.



Gambar 6. Hasil pengujian model dengan penerapan konfigurasi *epoch*.

Dari grafik hasil pengujian dapat dilihat jika keseluruhan dari 5 konfigurasi *epoch* menghasilkan akurasi diatas 80%. Seiring bertambahnya jumlah *epoch* dimulai dari *epoch* ke-20, 40 dan 60, akurasi klasifikasi yang dihasilkan oleh sistem cenderung mengalami kenaikan yang signifikan mulai dari 81,267%, 85,967% dan 91,033%. Tetapi ketika konfigurasi jumlah *epoch* bertambah dari 60 ke-80, akurasi klasifikasi yang dihasilkan turun sebesar 2,766% menjadi 88,267%. Pada konfigurasi *epoch* ke-100 atau yang terakhir, akurasi dari model mengalami sedikit peningkatan dibandingkan akurasi pada *epoch* ke-80, yaitu menjadi 89,133%. Dari keseluruhan hasil yang didapatkan dari eksperimen, akurasi terbaik diraih pada konfigurasi *epoch* 60 dengan hasil 91,033% dan terburuk pada *epoch* 20 dengan hasil 81,267%.

4. Diskusi

Mengacu pada hasil eksperimen yang didapat pada eksperimen yang tertera dalam Gambar 5. Akurasi dari sistem klasifikasi yang dibuat mengalami tren peningkatan seiring dengan pertambahan jumlah *epoch*. Tetapi terjadi sedikit fluktuasi akurasi, tepatnya pada *epoch* 60 performa akurasi dari sistem meningkat drastis yang kemudian turun lagi pada pertambahan *epoch* selanjutnya. Jika menilik secara teori, seiring bertambahnya nilai *epoch* maka performa dari sistem yang dirancang juga semakin meningkat dikarenakan sistem lebih mampu menggeneralisir data berdasarkan *update* dari *weight* dan *bias* atau pembelajaran yang sudah dilakukan pada *epoch* sebelumnya.

Fluktuasi nilai seperti ini terjadi karena sistem mengalami *overfitting* yang menyebabkan pada beberapa kondisi tertentu sistem tidak dapat menggeneralisir atau memprediksi data dengan baik. *Overfitting* yang terjadi dikarenakan jumlah dari *dataset* yang digunakan masih kurang atau belum mencapai jumlah yang sesuai. *Overfitting* juga dimungkinkan terjadi dikarenakan karakteristik *dataset* untuk kelas non-pornografi yang terdapat dalam NPDI *dataset* terlalu mudah untuk di prediksi, sehingga pendistribusian nilai saat proses *training* menjadi tidak seimbang.

Tidak ada acuan pasti dalam menentukan nilai *epoch* yang ingin diterapkan, hal ini terkait dengan konvergensi dari *neural network* yang dilatih. Suatu model dapat dikatakan konvergen jika dalam proses *training* yang dilakukan sudah tidak terdapat lagi peningkatan performa yang signifikan, hal ini bisa dilihat dari parameter *loss* saat *training*. Jika nilai *loss* selama *training* sudah tidak mengalami penurunan secara signifikan seiring pertambahan *epoch*, maka hal tersebut dapat menjadi pertanda bahwa model telah mencapai konvergensi dan sebaiknya penambahan jumlah *epoch* atau proses *training* dihentikan. Tetapi selama performa dari sistem masih mengalami tren kenaikan yang signifikan, maka penambahan jumlah *epoch* masih dimungkinkan sehingga proses *training* masih bisa dilanjutkan.

Kesimpulan

Berdasarkan hasil yang telah didapat dari eksperimen, dapat disimpulkan bahwa peran *epoch* sangat penting dalam mempengaruhi performa dari sistem klasifikasi CNN dengan ResNet-50 yang dirancang. Seiring meningkatnya jumlah *epoch* didapatkan tren naik dari performansi sistem, dikarenakan sistem mampu menggeneralisir data lebih baik berkat pembelajaran yang dilakukan pada *epoch* sebelumnya. Tetapi fluktuasi nilai masih mungkin terjadi dikarenakan *overfitting* yang dialami sistem. Sangat disarankan untuk penelitian selanjutnya agar dapat meningkatkan jumlah *epoch* latih dan menambah dari segi kuantitas maupun karakteristik *dataset* masukan sistem. Diharapkan, penelitian ini dapat dijadikan acuan untuk pengembangan metode ataupun piranti lunak yang dapat meminimalisir dampak negatif dari pornografi.

Terkait dengan riset kedepan yang dikerjakan akan berfokus mengenai pengaruh konfigurasi *learning rate* terhadap performansi klasifikasi citra pornografi dengan menggunakan ResNet-50. Serta terdapat rencana riset *optional* untuk mencari pengaruh modifikasi arsitektur, terutama dalam susunan dan jumlah *layer* ResNet-50 terhadap performa klasifikasi pornografi.

Kontribusi Penulis: konseptualisasi, F.N. and S.A.O.; metodologi, F.N.,G.B.; piranti lunak, F.N.; validasi, F.N., S.A.O. dan G.B.; analisis formal, F.N.; investigasi, F.N., S.A.O.; sumber, S.A.O., G.B.; kurasi data, F.N.; penulisan— persiapan *draft* asli, F.N.; penulisan—peninjauan dan penyuntingan, S.A.O., G.B.; visualisasi, F.N.; pimpinan, F.N., S.A.O.; administrasi proyek, S.A.O., G.B.; akuisisi dana, F.N.

Referensi

1. R. C. Reid, D. S. Li, R. Gilliland, J. A. Stein, and T. Fong, “Reliability, validity, and psychometric

development of the pornography consumption inventory in a sample of hypersexual men," *J. Sex Marital Ther.*, vol. 37, no. 5, pp. 359–385, 2011, doi: 10.1080/0092623X.2011.607047.

2. E. W. Owens, R. J. Behun, J. C. Manning, and R. C. Reid, "The Impact of Internet Pornography on Adolescents: A Review of the Research," *Sexual Addiction and Compulsivity*. 2012, doi: 10.1080/10720162.2012.660431.
3. J. A. M. Basilio, G. A. Torres, G. S. Pérez, L. K. T. Medina, and H. M. P. Meana, "Explicit image detection using YCbCr space color model as skin detection," 2011. https://www.scopus.com/inward/record.uri?partnerID=HzOxMe3b&scp=79958723400&origin=inward
4. S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Convolutional Shallow Features for Performance Improvement of Histogram of Oriented Gradients in Visual Object Tracking," *Math. Probl. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/6329864.
5. S. A. Wibowo, H. Lee, E. K. Kim, and S. Kim, "Collaborative Learning based on Convolutional Features and Correlation Filter for Visual Tracking," *Int. J. Control. Autom. Syst.*, vol. 16, no. 1, pp. 335–349, 2018, doi: 10.1007/s12555-017-0062-x.
6. A. Wiranata, S. A. Wibowo, R. Patmasari, F. T. Elektro, and U. Telkom, "Region-Based Convolutional Neural Network Untuk Deteksi Kendaraan Region-Based Convolutional Neural Networks." https://openlibrary.telkomuniversity.ac.id/home/catalog/id/149659/slug/analisis-performansi-region-based-convolutional-neural-network-untuk-deteksi-kendaraan.html
7. A. Y. Nasirudin, S. A. Wibowo, and R. Purnamasari, "Analisis Parameter Performansi Pada Pengenalan Objek Berbasis Region-Convolutional Neural Network Performance Parameter Analysis For Region-Convolutional Neural Network Based Object Recognition", PP.28.https://openlibrary.telkomuniversity.ac.id/home /catalog/id/149655/slug/analisis parameter performansi-pada-pengenalan-objek-berbasis-region-convolutional-neural- network.html
8. K. Zhou, L. Zhuo, Z. Geng, J. Zhang, and X. G. Li, "Convolutional neural networks based pornographic image classification," *Proc. - 2016 IEEE 2nd Int. Conf. Multimed. Big Data, BigMM 2016*, pp. 206–209, 2016, doi: 10.1109/BigMM.2016.29.
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
10. S. Sharma, "Epoch vs Batch Size vs Iterations - Towards Data Science," *Medium*, 2017. https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9 (accessed Jun. 26, 2020).
11. S. Avila, N. Thome, M. Cord, E. Valle, and A. De A. Araújo, "Pooling in image representation: The visual codeword point of view," *Comput. Vis. Image Underst.*, 2013, doi: 10.1016/j.cviu.2012.09.007.

© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).