Identifikasi Jumlah Manusia Dalam Kerumunan Menggunakan Convolutional Neural Network

Fernando [1)] Lina [2)]

[1) 2)] Teknik Informatika Universitas Tarumanaraga
Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia
[1)] fernandolie7@gmail.com [2)] lina@untar.ac.id

ABSTRACT

Public space is a place that generally used by the community in order to meet their needs and where crowds are usually formed. In a crowd, the number of people can be the first indicator in an anomaly in where the more number of people exists in a crowd, the more supervision is needed on the crowd to prevent chaos or other things that are not desirable in public spaces. The need for a crowd-counting is certainly needed to facilitate supervision and also the awareness of people in crowds. This research meant to develop a system that can identifies a crowd based on the number of people that exist in the crowd and also give the number of people as an output. The system applied the Convolutional Neural Network (CNN) algorithm. The CNN model is trained using a labeled crowd dataset with a total of 4372 crowd photos. The CNN works as a regression model that will count the number of people from the feature extracted from the image. The evalution shows the Mean Absolute Error value achieved is 55.1176 in the test data.

Key words

Convolutional Neural Network, Crowd, Crowd Counting, Public Space, Regression

1. Pendahuluan

Ruang publik adalah suatu ruang dimana seluruh masyarakat umum mempunyai akses untuk menggunakan tempat tersebut [3]. Ruang publik bersifat terbuka dan bebas digunakan oleh masyarakat sesuai dengan peraturan yang berlaku di ruang publik tersebut. Ruang publik juga menjadi tempat bagi masyarakat untuk melakukan aktifitasnya baik secara individu maupun berkelompok. Ketika seseorang menjalankan aktivitasnya dalam sebuah ruang publik, seseorang dapat berjumpa dengan masyarakat lain dalam waktu dan tempat yang sama dan menciptakan kerumunan. Dikarenakan sifatnya yang terbuka untuk masyarakat umum, maka kerumunan menjadi hal yang lumrah ada dalam suatu ruang publik.

Kerumunan adalah kumpulan orang dan sebagainya yang tidak teratur dan bersifat sementara. Kerumunan yang tercipta di dalam ruang publik memiliki beberapa bentuk seperti antrian, orasi, penonton, kepanikan, dan berbagai bentuk lainnya sesuai tujuan dan kondisi dari masyarakat yang ada di dalam kerumunan tersebut. Kerumunan yang terjadi di sebuah ruang publik dapat diukur lewat jumlah orang yang terdapat dalam kerumunan tersebut. Kerumunan dapat diklasifikasikan berdasarkan tingkat kepadatan orang yang ada di dalam kerumunan tersebut menjadi kategori padat dan renggang. Dikarenakan ruang publik terbuka bagi masyarakat, kerumunan padat dapat kita jumpai di ruang publik secara lumrah. Namun untuk menjaga agar lingkungan ruang publik dapat tetap kondusif digunakan oleh masyarakat, perlu ada penanganan terhadap kerumunan yang tercipta dalam suatu ruang publik terutama apabila kerumunan tersebut tercipta dalam skala padat. Pengetahuan tepat akan jumlah kerumunan di ruang publik dapat menyediakan wawasan berharga untuk tugas-tugas seperti perencanaan kota, analisis pola belanja konsumen, dan menjaga keamanan kerumunan umum [4].

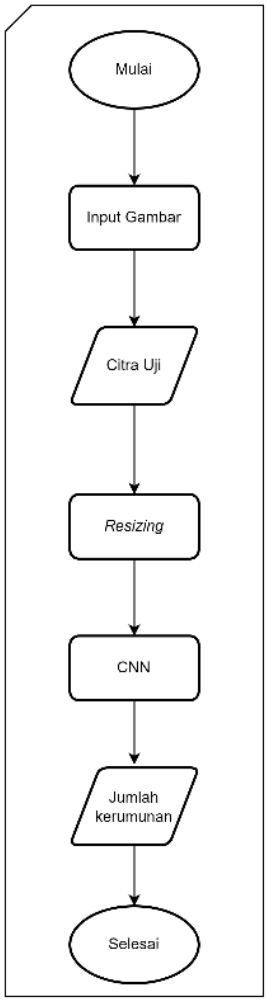
Terdapat berbagai pendekatan metode seperti metode *supervised* dan *unsupervised* berdasarkan label dari data yang digunakan untuk melakukan penghitungan jumlah orang dalam kerumunan. Penghitungan kerumunan dengan pendekatan *supervised* dapat dibagi menjadi penghitungan dengan regresi, estimasi kepadatan, deteksi, dan *Convolutional Neural Network* (CNN) [2]. Pada penelitian yang pernah dilakukan sebelumnya dibuat sebuah sistem deteksi kerumunan yang mendeteksi kerumunan dari gambar menggunakan *Fully Convolutional Network* (FCN) yang diambil pada perspektif *drone* untuk menentukan apakah *drone* menemukan kerumunan atau tidak dalam kebutuhan *social distancing* dengan akurasi sebesar 0.978 [1].

Penelitian ini bertujuan menciptakan sistem identifikasi kerumunan lewat jumlah orang yang ditemukan kerumunan dengan menggunakan metode CNN. Penelitian ini juga akan mengimplementasikan

model yang mampu mengolah data dengan beberapa halangan berupa objek bukan manusia dan gangguan cuaca.

2. Metode Penelitian

Pada penelitian ini, sistem dibangun menggunakan metode CNN. Sistem dibangun untuk melakukan regresi perhitungan jumlah orang dalam kerumunan. Model yang dihasilkan akan diimplementasikan pada program Alur kerja pada sistem dapat dilihat pada Gambar 1.



Gambar 1. *Flowchart* program

2.1 Dataset

Pada percobaan ini, dataset yang digunakan untuk algoritma CNN adalah dataset JHU-CROWD++ yang memiliki 4372 data dengan rata-rata resolusi sebesar 1430×910 piksel. Jumlah orang yang terdata dalam keseluruhan gambar memiliki rata-rata 346 orang dengan rentang dari 0 hingga 25791 orang dalam satu gambar. Dengan berkembangkan kebutuhan akan *crowd counting* maka kebutuhan data untuk menciptakan model *crowd counting* juga meningkat. Namun beberapa dataset yang ada memiliki kecenderungan berupa data yang kurang bervariasi dan juga minimnya kondisi cuaca yang mempengaruhi gambar. Dataset JHU-CROWD++ merupakan dataset *crowd counting* dengan variasi tempat dan kondisi yang sangat luas dari tiap gambarnya dengan meliputi halangan dan juga kondisi cuaca yang mempengaruhi gambar kerumunan [5]. Dataset JHU-CROWD++ juga dilengkapi dengan anotasi dan juga label dari jumlah orang yang ada pada tiap gambar dalam dataset. Contoh data citra dapat dilihat pada Gambar 2.

Gambar 2. Contoh gambar kerumunan pada dataset

2.2 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah salah satu metode dalam *deep learning* yang populer dipakai dalam pengembangan sistem pengenalan gambar dan ucapan. Metode CNN berpusat pada lapisan konvolusi yang digunakan untuk mengurangi dimensi gambar yang tinggi dengan menjaga ciri penting yang terdapat dalam gambar. Itulah mengapa CNN sangat cocok digunakan pada kasus pengenalan gambar.

Meskipun CNN umum digunakan dalam klasifikasi gambar, CNN juga dapat digunakan untuk menciptakan model untuk melakukan perhitungan jumlah orang dalam kerumunan. Seperti pada penelitian [6] yang menjadi salah satu pionir penggunaan model CNN dalam bentuk regresi untuk menghitung jumlah orang dalam kerumunan. Model CNN dapat digunakan dengan baik untuk mendeteksi orang dalam kerumunan padat dikarenakan kemampuannya untuk mengambil ciri penting baik dari piksel berukuran rendah yang dimana pendekatan tradisional akan mendapati kesulitan untuk melakukan hal serupa pada data kerumunan padat.

CNN secara garis besar terdiri dari dua bagian yaitu ekstraksi ciri dan *fully connected*. Proses ekstraksi ciri dilakukan untuk mendapatkan informasi dari suatu gambar dengan mengurangi dimensi dari gambar tersebut lewat lapisan konvolusi dan lapisan *pooling*. Sedangkan *fully-connected* bertujuan untuk melakukan transformasi data yang sudah dilakukan proses ekstraksi ciri menjadi keluaran yang diinginkan seperti klasifikasi atau dalam penelitian kali ini adalah regresi jumlah orang dalam kerumunan.

Lapisan konvolusi melakukan operasi antara gambar input dan juga parameter filter yang dapat dinotasikan sebagai berikut:

$$X_i = W_i \otimes [X_{i-1}, 1]^T \quad (1)$$

Dengan X_i adalah peta fitur input maupun output dari layer ke-i dan W_i adalah parameter filter, serta $\otimes$ melambangkan operasi konvolusi.

Lapisan *pooling* melakukan reduksi dimensi dari peta fitur hasil konvolusi baik dengan operasi *pooling* maksimum, minimum, dan juga rata-rata.

Setelah dilakukan *pooling*, proses akan dilanjutkan pada lapisan *fully connected*. Lapisan *fully connected* dapat dihitung sebagai :

$$y_{im} = W_{im}X_{(i-1)} + B_m \quad (2)$$

B mendenotasikan bias, lapisan ini membutuhkan angka tetap dari input dan output untuk mengubah peta fitur menjadi nilai kebenaran.

Lapisan neuron melakukan proses aktivasi non-linear oleh fungsi seperti *Rectified Linear Unit* (ReLU) sebelum diproses oleh lapisan berikutnya.

Lapisan *loss* yang digunakan pada model CNN ini termasuk pada regresi yang menggunakan *Mean Squared Error* (MSE) sebagai denotasi nilai *loss* dimana Y menunjukkan nilai prediksi dan Ya menunjukkan nilai kebenaran pada proses pembelajaran model yang dirumuskan sebagai berikut :

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(Y_i - Ya_i)^2 \quad (3)$$

Evaluasi lalu akan diujikan pada model dengan melakukan perhitungan nilai *Mean Absolute Error* (MAE) dan *Mean Deviation Error* (MDE) dengan rumus sebagai berikut :

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|Y_i - Ya_i| \quad (4)$$

$$MDE = \frac{1}{N}\sum_{i=1}^{n}\frac{|Y_i - Ya_i|}{Ya} \quad (5)$$

Pelatihan model dilakukan dengan jumlah *epoch* sebanyak 50 *epoch* dan menggunakan arsitektur Resnet50V2 yang memiliki kedalaman 50 lapisan yang digunakan. Arsitektur Pelatihan dapat dilihat pada Tabel 1.

Tabel 1 Konfigurasi Arsitektur Sistem

Lapisan	Parameter
Resnet50V2	
Dense	1024, ReLU
Dense	1, Linear

Parameter yang diproses kedalam model adalah input gambar dari data latih yang seluruhnya akan dilakukan proses *resizing* ke dalam ukuran 224×224 piksel dalam format warna RGB. Model CNN akan mengeluarkan vektor ukuran 1×1 dengan tipe data *float* yang merepresentasikan jumlah orang yang dideteksi dari gambar yang diproses oleh model. Tahap terakhir proses perhitungan akan mengubah nilai keluaran menjadi sebuah bilangan bulat dikarenakan sifat orang yang sebaiknya dihitung sebagai kesatuan utuh dan bulat. Pelatihan model dilakukan terhadap 4 macam konfigurasi dataset yang dibedakan berdasarkan nilai maksimal jumlah orang yang terdapat pada gambar yang akan menghasilkan 4 macam model dengan jumlah data latih yang berbeda. Konfigurasi dilakukan sehingga model memiliki variasi proses peta fitur berdasarkan karakteristik dataset yang cukup berbeda antara tiap konfigurasinya dimana dataset dengan konfigurasi jumlah orang yang semakin besar cenderung memiliki halangan yang relatif besar dibanding dataset dengan konfigurasi jumlah orang yang relatif kecil. Konfigurasi dataset untuk pelatihan dapat dilihat pada Tabel 2.

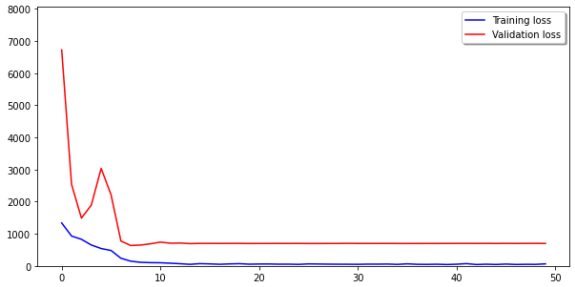
Tabel 2 Konfigurasi Dataset Pelatihan Model

Nama Model	Jumlah Maksimal orang dalam kerumunan	Jumlah Data
Model A	150	2240
Model B	200	2567
Model C	250	2775
Model D	500	3298

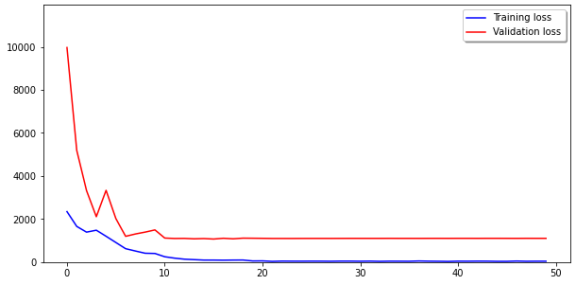
Untuk konfigurasi *learning rate* pada setiap model menggunakan konfigurasi yang sama yaitu nilai dinamis menggunakan *ReduceLROnPlateau* yang dimulai dengan nilai 0.001. Konfigurasi *learning rate* ini dimaksudkan agar model bisa mencapai pembelajaran optimal lewat nilai *learning rate* yang dinamis.

Pelatihan dilanjutkan berdasarkan konfigurasi yang sudah dibuat sebelumnya dimana nilai *loss* yang didapat dari nilai MSE data validasi model dapat ditelusuri pada grafik *loss* dan *epoch* pada setiap

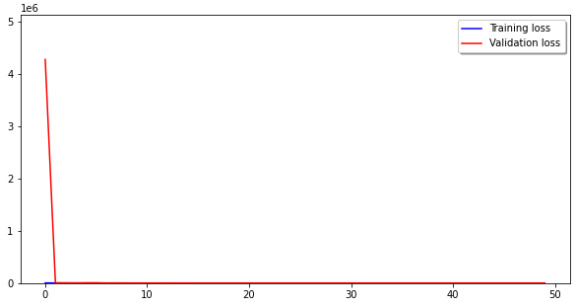
model di Gambar 3, Gambar 4, Gambar 5, dan Gambar 6.



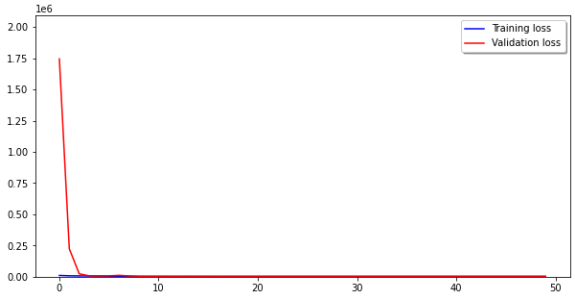
Gambar 3. Grafik *Loss* CNN Model A



Gambar 4. Grafik *Loss* CNN Model B



Gambar 5. Grafik *Loss* CNN Model C



Gambar 6. Grafik *Loss* CNN Model D

Dari setiap grafik *loss* milik seluruh model, dapat dilihat bahwa nilai *loss* cenderung sudah mencapai titik minimum pada pelatihan sebanyak 50 *epoch*. Konfigurasi tambahan berupa nilai *epoch* dirasa tidak dibutuhkan berdasarkan hasil *loss* yang sudah tercapai pada 4 macam konfigurasi model.

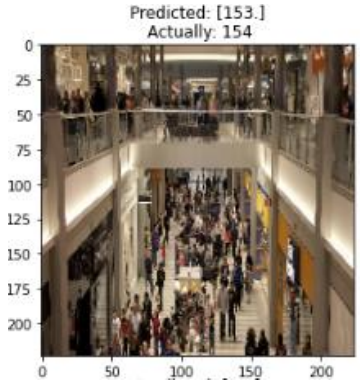
Pelatihan kemudian dilanjutkan dengan melakukan evaluasi metrik MAE pada setiap model yang sudah dilatih. Data yang digunakan pada evaluasi ini adalah data latih dengan konfigurasi data maksimal 250 jumlah orang dalam gambar dengan total data sebanyak 2775 data. Data evaluasi ini memiliki irisan terhadap data yang digunakan dalam pelatihan dari seluruh model. Evaluasi menggunakan data latih dapat dilihat pada Tabel 3.

Tabel 3 Evaluasi Data Latih Pada Model

Nama Model	*Mean Absolute Error* (MAE)
Model A	60.4578
Model B	65.7432
Model C	69.7664
Model D	71.9776

3. Hasil Percobaan

Model yang telah dilatih kemudian diberi pengujian untuk mengevaluasi hasil evaluasi dari proses pelatihan untuk menemukan model terbaik. Proses pengujian akan mengeluarkan bilangan bulat yang merepresentasikan jumlah orang yang ada di dalam input gambar. Contoh output pada program dengan input salah satu gambar dapat dilihat pada gambar 6.



Gambar 6. Hasil pengujian model CNN

Label *predicted* merupakan hasil keluaran dari model CNN sedangkan label *actually* merupakan label dari nilai *ground truth* dari data yang digunakan dalam pengujian. Untuk mendapatkan model terbaik dari keseluruhan model yang telah dilatih, pengujian model CNN akan menggunakan dataset dari JHU-CROWD++ yang tidak digunakan pada proses pelatihan dengan konfigurasi maksimal jumlah orang sebanyak 250 orang dengan jumlah data evaluasi sebanyak 380 data. Pengujian akan dilakukan dengan membandingkan performa tiap model dalam metrik evaluasi MAE, MSE, dan MDE dari data uji. Keseluruhan data akan diproses tanpa adanya proses selain *resizing* saat diproses oleh semua model. Hasil evaluasi dari keempat model yang telah dilatih dapat dilihat pada Tabel 4.

Tabel 4 Evaluasi Data Tes Pada Model

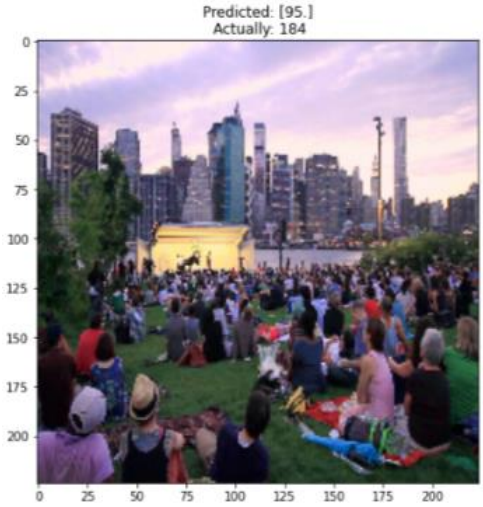
Epoch	Model	MAE	MSE	MDE
50	Model A	55.6320	5203.8388	1.4911
50	Model B	59.7251	5742.0856	1.7951
50	Model C	62.2148	6197.4965	1.9202
50	Model D	72.7027	8773.2049	2.3681

Berdasarkan 3 metrik evaluasi yang diterapkan pada 4 model yang sudah dilatih dengan konfigurasi dataset yang berbeda, terlihat secara jelas bahwa Model A dengan konfigurasi dataset dengan jumlah maksimal 150 orang memiliki performa terbaik dibandingkan dengan 3 Model dengan konfigurasi lainnya. Model A memiliki keunggulan pada 3 metrik evaluasi yang diukur dari data tes yang diujikan. Dengan demikian, Model A menjadi model CNN paling optimal yang dapat digunakan pada sistem identifikasi jumlah orang dalam kerumunan dalam penelitian ini dengan nilai MAE sebanyak 55.1176 pada data tes JHU-CROWD++.

Dalam pengujian model A ke dalam data uji JHU-CROWD++ ditemukan 3 jenis perilaku model dalam melakukan prediksi jumlah orang dalam kerumunan. Dari 3 perilaku model yang telah disebutkan sebelumnya, 2 perilaku model memiliki kecenderungan melakukan prediksi yang tidak akurat yang akan disebut dengan perilaku negatif dan 1 perilaku yang mampu melakukan prediksi jumlah orang dalam kerumunan dalam tingkat akurasi yang memuaskan yang akan disebut sebagai perilaku positif.

Gambar 7. Contoh hasil perilaku positif pengujian model CNN

Dapat dilihat pada Gambar 6 bahwa model berhasil melakukan prediksi jumlah orang dengan akurasi yang cukup baik yaitu dengan nilai MAE sebesar 3. Namun ditemukan perilaku negatif performa model dalam beberapa kasus yang menyebabkan keluaran dari model mempunyai nilai MAE yang sangat besar, mulai dari prediksi model yang cenderung lebih sedikit maupun lebih banyak dari yang semestinya.

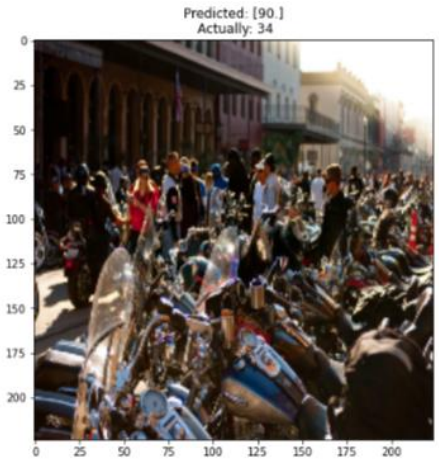


Gambar 8. Contoh pertama hasil perilaku negatif pengujian model CNN

Gambar 8 merupakan salah satu contoh dari perilaku negatif pada pengujian model dimana keluaran yang dihasilkan model mempunyai nilai yang sangat rendah dibandingkan label *ground truth* dari data uji dengan nilai MAE sebesar 89.

Gambar 9. Contoh kedua hasil perilaku negatif pengujian model CNN

Gambar 9 merupakan contoh hasil pengujian lain dengan perilaku negatif yang sama dengan Gambar 8 yang mengeluarkan nilai prediksi yang sangat kecil dibanding nilai *ground truth* dari data uji. Namun yang membedakan Gambar 9 dengan Gambar 8 adalah cuaca salju yang muncul pada Gambar 9 yang memiliki karakteristik yang sama dengan sebuah *noise* pada gambar.



Gambar 10. Contoh ketiga hasil perilaku negatif pengujian model CNN

Gambar 10 adalah salah satu contoh perilaku negatif kedua dari model dimana model mengembalikan output yang cenderung lebih banyak dibandingkan nilai *ground truth* dari data uji. Ciri utama yang bisa dilihat dari Gambar 10 adalah adanya keberadaan objek bukan manusia yang ada di dalam gambar yaitu motor yang cukup mendominasi gambar uji.

4. Kesimpulan

Kesimpulan yang didapatkan dari penelitian identifikasi jumlah orang dalam kerumunan menggunakan CNN adalah sebagai berikut :

- Metode CNN dengan menggunakan arsitektur Resnet50V2 dapat melakukan perhitungan jumlah orang dalam citra kerumunan tanpa adanya kecenderungan *overfitting*.
- Pelatihan model CNN dengan arsitektur Resnet50V2 dinilai sudah mencapai performa optimal dengan menggunakan epoch sebesar 50 *epoch*.
- *Noise* dan juga objek non-manusia dalam foto memengaruhi hasil penghitungan jumlah orang dalam kerumunan, sehingga hasil performa terbaik akan dicapai pada foto dengan *noise* yang rendah dan juga minimnya keberadaan objek bukan manusia yang terdapat dalam foto.
- Implementasi model CNN pada data uji mencapai tingkat akurasi MAE sebesar 55.63 dimana pada beberapa pengujian, model dapat melakukan prediksi dengan nilai MAE yang minim.
- Resolusi gambar memiliki pengaruh pada performa model dikarenakan adanya kecenderungan ciri orang yang hilang saat proses *resizing*.

Penelitian ini memiliki beberapa saran untuk penelitian berikutnya, dimana diharapkan dapat menambah data uji yang memiliki nilai halangan, *noise*, distribusi objek iregular yang diperbanyak untuk meningkatkan performa model. Penelitian berikutnya juga memasukkan sampel negatif berupa gambar objek non-manusia dengan jumlah orang bernilai 0 untuk dapat mempelajari fitur objek non-manusia untuk meningkatkan performa model.

REFERENSI

[1] Fatih, Muhammad., Suciati, Nanik., Navastara, Dini A., 2021 “Deteksi Kerumunan Menggunakan Metode Fully-Convolutional Network pada kamera”,JURNAL TEKNIK ITS,Vol.10, Nomor 2, Surabaya.

[2] Ilyas, Naveed., Shahzad, Ahsan., & Kim, Kiseon., 2019, “Convolutional-Neural Network-Based Image Crowd Counting: Review, Categorization, Analysis, and Performance Evaluation”. Sensors (Basel), Vol. 20(1), Nomor 43, Basel, Switzerland.

[3] Kustianingrum, Dwi., Sukarya, Angga K., Nugraha, Rifan A., Rachadi, Franderdi., 2013, “Fungsi dan Aktifitas Taman Ganesha Sebagai Ruang Publik di Kota Bandung”, Jurnal Reka Karsa. Vol. 1, Nomor 2, Bandung.

[4] Marsden, Mark., McGuinness, Kevin., Little, Suzanne., O’Conner, Noel., 2017, “Fully Convolutional Crowd Counting on Highly Congested Scenes”, Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Vol. 5, Porto, Portugal.

[5] Sindagi, Vishwanath A., Yasarla, Rajeev., & Patel, Vishal M., 2020, “JHU-CROWD++: Large-Scale Crowd Counting Dataset and A Benchmark Method”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 44.

[6] Wang, Chuan., Zhang, Hua., Yang, Liang., Liu, Si., Cao, Xiaochun., 2015, “Deep People Counting in Extremely Dense Crowds”, In Proceedings of the 23rd ACM international conference on Multimedia (MM '15) - Association for Computing Machinery, New York, NY, USA.

Fernando, saat ini sebagai mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara

Lina S.T., M.Kom., Ph.D., memperoleh gelar magister dari Universitas Indonesia. Kemudian memperoleh gelar Doktor dari Nagoya University. Saat ini sebagai Dosen program studi Teknik Informatika, Falkutas Teknologi Informasi, Universitas Tarumanagara