

# ANALISIS SISTEM DETEKSI PENGGUNAAN HARD HAT PADA PEKERJA KONSTRUKSI MENGGUNAKAN METODE JARINGAN SYARAF TIRUAN KONVOLUSI

## ANALYSIS OF DETECTION SYSTEM ON HARD HAT USE BY CONSTRUCTION WORKERS USING CONVOLUTIONAL NEURAL NETWORK METHOD

Faris Fadhur Rachman<sup>1</sup>, Hertiana Bethaningtyas<sup>2</sup>, Reza Fauzi Iskandar<sup>3</sup>

<sup>1,2,3</sup> Prodi S1 Teknik Fisika, Fakultas Teknik Elektro, Universitas Telkom

<sup>1</sup>farisfadhlur@student.telkomuniversity.ac.id, <sup>2</sup>hertiana@telkomuniversity.ac.id, <sup>3</sup>rezafauzi@telkomuniversity.ac.id

### Abstrak

Keselamatan kerja merupakan sebuah tanggung jawab perusahaan dalam menjaga asset agar tidak mengalami kerugian, salah satu faktor yang dapat mengancam keselamatan kerja adalah kedisiplinan penggunaan *hard hat*. Oleh karena itu penggunaan *hard hat* perlu di monitor untuk menjaga nilai K3 perusahaan dan juga keselamatan pekerja. Metode penelitian yang diusulkan oleh penulis adalah menggunakan metode deteksi objek melalui kamera menggunakan *Convolutional Neural Network* (CNN) atau Jaringan Syaraf Konvolusi dengan beberapa arsitektur yang mendukung deteksi objek secara waktu-nyata yaitu *Faster Region-based CNN* dan *Single Shot Multibox Detector*. Sehingga dalam aplikasinya CNN tersebut digunakan untuk mendeteksi objek berupa alat keselamatan yang digunakan oleh seseorang. Hasil akhir dari penelitian ini adalah sebuah sistem keselamatan yang ditampilkan melalui GUI menggunakan arsitektur CNN dengan performansi presisi terbaik pada model *Faster R-CNN Inception V2* yang mencapai nilai presisi 86,841% dan performansi rasio *frame rate* terbaik pada model SSD *Mobilenet V1* dengan nilai 10,364 kali lebih besar dibandingkan dengan model *Faster R-CNN Inception V2*. Selain kedua metrik tersebut dilakukan perhitungan nilai presisi berdasarkan lokalisasi deteksi menggunakan metrik COCO yang menghasilkan nilai terbaik pada model *Faster R-CNN Inception V2* dengan nilai 47.120%.

**Kata Kunci:** Deteksi Objek, Jaringan Syaraf, dan Keselamatan

### Abstract

Work safety is a company's responsibility in protecting assets so that the company doesn't suffer losses, one of the threat that may risk work safety is the discipline of hard hat use. Because of that the use of hard hat needs to be monitored to not only to maintain company's OSH value, but also worker's safety. The research method proposed by the author is to use the object detection method through the camera using Convolutional Neural Network (CNN) with some architecture supports that real-time object detection, namely Faster Region-based CNN and Single Shot Multibox Detector. So that the CNN may be used to detect objects in the form of safety equipment used by someone. The final result of this research is a safety system displayed through a GUI using CNN architecture with the best precision performance in the Faster R-CNN Inception V2 model which reaches a precision value of 86.841% and the best frame rate performance in the SSD Mobilenet V1 with a value of 10,364 times higher than the Faster R-CNN Inception V2 model. In addition to these two metrics, the calculation of the value of localization precision is carried out using the COCO metric which produces the best value in the Faster R-CNN Inception V2 model with a value of 47,120%.

**Keywords:** Object Detection, Neural Network, and Safety

### 1. Pendahuluan

Keselamatan selalu dikaitkan dengan kesehatan karena sistem keselamatan di Indonesia terkait dengan program K3 (Keselamatan dan Kesehatan Kerja), dengan filosofi dasar K3 adalah menjamin keutuhan dan kesempurnaan melalui perlindungan atas keselamatan dan kesehatan para pekerja dalam menjalankan pekerjaannya [1]. Helm merupakan APD (Alat Pelindung Diri) yang disediakan secara cuma-cuma oleh perusahaan dan digunakan untuk pekerjaan pada ketinggian dan wajib diterapkan dan digunakan sebagai pada peraturan K3 dijelaskan[2]. Pada peringatan bulan K3 tahun 2019 Menteri Ketenagakerjaan, Hanif Dhakiri, beserta Dirpel. BPJS Ketenagakerjaan, Krishna Syarif, menyatakan bahwa pada tahun 2018 angka kecelakaan kerja meningkat dari angka 123.041 pada tahun 2017 menjadi 173.105 [3].

Pemantauan pelaksanaan K3 pada umumnya dilakukan secara manual namun beberapa automasi dari sistem K3 memiliki potensial untuk diaplikasikan salah satunya adalah, deteksi penggunaan helm berbasis sensor namun deteksi tersebut memerlukan sensor pada setiap helm yang akan dikenakan sehingga dapat mengurangi nilai praktisnya baik secara biaya maupun pengaplikasiannya [4]. Dengan pernyataan tersebut Qi

Fang, et. al. melakukan penelitian dengan melakukan pendeteksian kelalaian penggunaan hard hat menggunakan Faster RCNN untuk membuktikan adaptibilitas dari penggunaan CNN terhadap metode deteksi kelalaian penggunaan hard hat. Salah satu kelebihan dari penelitian yang dilakukan oleh Qi Fang, et. al. yaitu dilakukannya pengukuran presisi dari model terhadap faktor-faktor lingkungan yang dapat mempengaruhi gambar dengan hasil yang dominan di atas 90%. Dan beberapa kekurangan dari penelitian tersebut yaitu: deteksi terbatas hanya pada pekerja yang tidak menggunakan hard hat, selain itu metode evaluasi yang dilakukan dengan identifikasi *true positive*, *false positive*, dan *false negative* tidak disertai parameter IoU, dan model yang digunakan pada penelitian tersebut terbatas hanya dengan model Faster R-CNN.

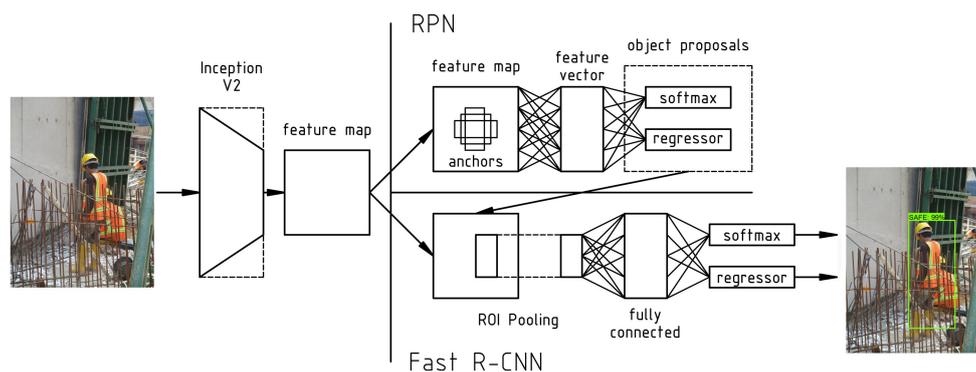
Oleh karena itu, terdapat beberapa solusi yang dapat diusulkan untuk permasalahan tersebut diantaranya yaitu menyertakan kelas penggunaan hard-hat, selain itu pada penelitian ini juga dilakukan metode evaluasi yang menghasilkan metrik berdasarkan IoU sehingga metrik tersebut dapat melambungkan tingkat lokalisasi dari deteksi, dan menyertakan model lain seperti *Single Shot multibox Detector* atau SSD untuk melihat perbedaan performansi model *Faster R-CNN* dengan model SSD tersebut dan bagaimana kedua model tersebut dapat digunakan berdasarkan nilai performansi yang dihasilkan.

## 2. Metodologi dan Perancangan Sistem

Dalam perancangan sistem deteksi penggunaan *hard hat*, sebuah arsitektur *Convolutional Neural Network* atau CNN digunakan untuk melakukan pengenalan citra *hard hat* kepada sistem deteksi tersebut. *Convolutional Neural Network* atau Jaringan Syaraf Konvolusi merupakan sebuah Jaringan Syaraf Tiruan yang identik dengan operasi konvolusi sebelum dilanjutkan ke jaringan syaraf terhubung, hal tersebut dilakukan guna mengkompresi data besar menjadi fitur sederhana yang dapat diolah oleh sejumlah neuron yang tersedia. *Convolution Neural Network* atau CNN bekerja dengan mengolah data matriks gambar kedalam layer konvolusi dan aktifasi, sehingga sistem dapat mengklasifikasi suatu bagian dari citra yang sesuai dengan kelas yang dilatihkan pada sistem tersebut. Dalam penelitian ini arsitektur CNN yang digunakan adalah *Faster R-CNN* dengan *backbone Inception V2* dan *Single Shot Multibox Detector* dengan *backbone Mobilenet V1*.

### 2.1 Faster Region-based Convolutional Neural Network

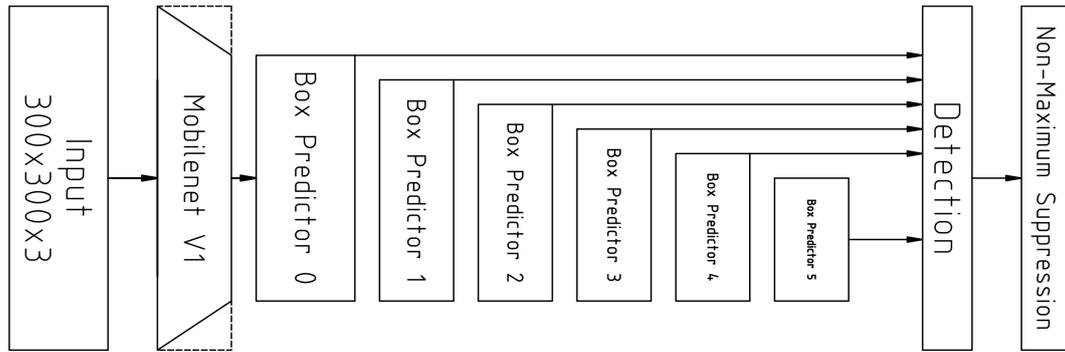
*Faster R-CNN* adalah sebuah arsitektur CNN yang diusulkan oleh Shaoqing Ren dkk. *Faster R-CNN* yang juga disebut meta-arsitektur merupakan pengembangan dari *Fast R-CNN* dengan komputasi yang lebih cepat [5]. *Faster R-CNN* melibatkan 3 proses yaitu ekstraksi fitur, RPN, dan *Fast R-CNN*. Dalam penelitian ini *Faster R-CNN* menggunakan *feature extractor Inception V2*[6]. Arsitektur *Faster R-CNN* ditunjukkan oleh Gambar 1 dibawah ini.



Gambar 1. Arsitektur *Faster R-CNN* [7]

### 2.2 Single Shot Multibox Detector

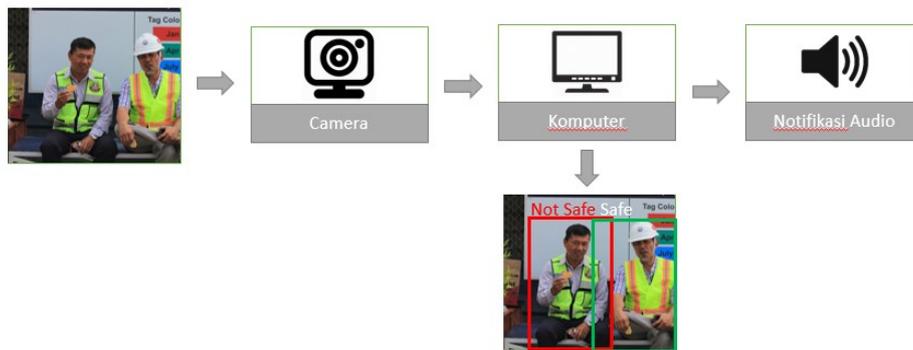
SSD adalah sebuah arsitektur CNN yang diusulkan oleh Wei Liu dkk[8]. SSD merupakan salah satu metode *object detection* yang ditujukan untuk memiliki tingkat akurasi tinggi dengan melakukan komputasi sederhana yaitu melakukan prediksi dalam sekali tembak gambar. SSD melakukan deteksi dengan menerapkan 9 tingkat *bounding box* pada setiap *pixel* pada fitur map dengan skala berbeda, lalu melakukan prediksi pada setiap *bounding box* tersebut. Dalam penelitian ini SSD menggunakan *feature extractor Mobilenet V1* [9]. Dibawah ini pada Gambar 2 menggambarkan susunan Arsitektur *Single Shot Multibox Detector*.



Gambar 2. Arsitektur Single Shot Multibox Detector [8]

2.3 Diagram Blok

Sistem terdiri atas kamera, monitor, dan speaker yang saling terhubung. Kamera digunakan untuk mengambil citra latih dan citra waktu nyata yang ditampilkan untuk identifikasi citra. Kamera terhubung dengan monitor PC agar dapat memproses gambar latih, melatih sistem untuk mengenali kelas hingga menampilkan hasil identifikasi citra melalui GUI. Hasil identifikasi tersebut diproses oleh komputer sehingga apabila sistem mendeteksi tubuh manusia tanpa helm sistem akan membunyikan speaker sebagai notifikasi pengguna.



Gambar 3. Diagram Blok Sistem

2.4 Metode Evaluasi

Pengujian atau evaluasi dilakukan dengan memberikan *input* model dengan data evaluasi agar sistem dapat dilakukan penilaian tingkat presisi sistem dalam mengklasifikasi gambar selain gambar yang telah dilatih. Setiap hasil prediksi yang dilakukan oleh model dapat dibedakan sesuai dengan nilai IoU hasil prediksi dengan *ground truth* yang dilambangkan oleh persamaan 1 dan seperti yang ditunjukkan pada Gambar 4 dibawah.



Threshold IoU = 0.5  
 GT = Ground Truth  
 TP = True Positive  
 FP = False Positive  
 FN = False Negative

Gambar 4. Klasifikasi Hasil Prediksi Terhadap *Ground Truth*

Dengan memperoleh jumlah prediksi *true positive*, *false positive*, dan *false negative* selanjutnya diukur menggunakan metrik salah satunya yaitu nilai *Average Precision* (AP) yang membandingkan sebagaimana baik

hasil prediksi untuk kedua model tersebut dalam memprediksi *ground truth* pada data evaluasi dengan tepat. Selain itu nilai *Average Recall* (AR) menunjukkan persentase dari jumlah objek yang berhasil terdeteksi terhadap seluruh objek pada data evaluasi. Persamaan 2 dan 3 dibawah ini merupakan metrik utama dalam mengukur performansi dari kedua model yang sudah dibuat.

$$\text{Intersect over Union} = \frac{\text{Ground Truth} \cap \text{Bounding Box Prediksi}}{\text{Ground Truth} \cup \text{Bounding Box Prediksi}} \quad (1)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

### 3. Pembahasan

#### 3.1 Pengolahan Dataset

Klasifikasi untuk sistem keselamatan dibagi menjadi dua dengan fokus kepada penggunaan *hard hat*. Pada setiap pekerja yaitu kelas *Safe* dan *Not Safe*. Kedua ini dibedakan oleh penggunaan *hard hat* pada tubuh pekerja, sehingga gambar pekerja yang menggunakan *hard hat* dan pekerja tidak menggunakan *hard hat* dijadikan sebagai *region of interest* seperti yang ditunjukkan oleh Gambar 4 di bawah ini.



**Gambar 5.** ROI untuk Setiap Kelas yang dideteksi

Menggunakan data berupa foto yang dikumpulkan dari sebuah proyek pembangunan kantor Shell di Jakarta. Data tersebut selanjutnya dibagi menjadi data *training* dan data evaluasi. Data *training* adalah data yang digunakan untuk melatih *neural network*, sementara data evaluasi adalah data yang nantinya digunakan untuk mengevaluasi kinerja dari model *neural network* yang dibangun. Distribusi pembagian dataset pekerja ditunjukkan oleh Tabel 1 dibawah ini.

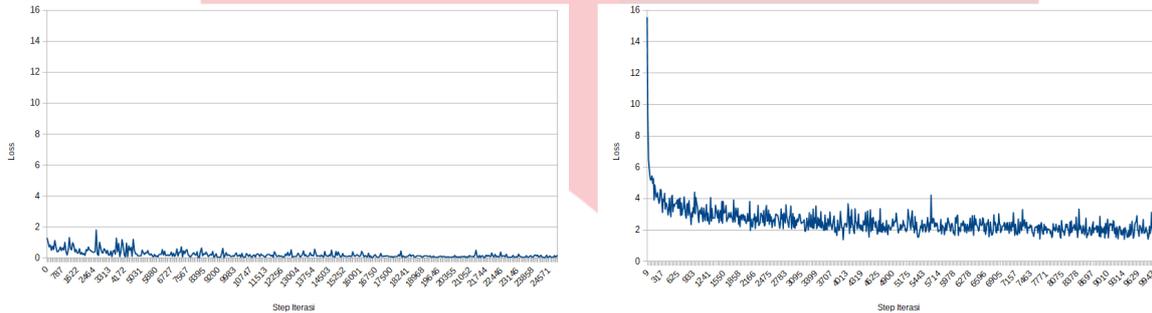
**Tabel 1.** Pembagian Jumlah *Dataset*

Objek Citra	<i>Dataset</i>	Data <i>Training</i>	Data Evaluasi
<i>Safe</i>	916	734	182
<i>Not Safe</i>	672	537	135

Seluruh data tersebut selanjutnya dilakukan diberikan label pada gambar yang terkait dengan kriteria kelas yaitu *safe* dan *not safe*. Pelabelan dilakukan untuk memberikan *ground truth* yang menjadi input pada proses *training* dan metrik perbandingan *ground truth* dengan prediksi model pada proses evaluasi.

### 3.2 Model Neural Network Sistem Keselamatan

Model *neural network* sistem keselamatan memerlukan kumpulan data dengan *ground truth* seperti yang dilakukan pada proses *pre-processing* sebelumnya untuk mengenali gambar. Oleh karena itu data *training* tersebut dilatih kedalam model *neural network* sistem keselamatan.



**Gambar 6.** Grafik *Total Loss* dalam proses training model *Faster R-CNN Inception V2* dan *SSD Mobilenet V1*

Visualisasi grafik *loss* terhadap setiap step pelatihan model *neural network* sistem keselamatan yang dibedakan dalam jumlah step dalam proses *training* ditunjukkan oleh Gambar 6 menunjukkan proses pembelajaran sistem dalam mengenali dataset yang sudah diberikan *ground truth* terlebih dahulu. Kedua model *Faster R-CNN* dan *SSD* yang sudah dilatih selanjutnya dilakukan pengujian kedua model dengan input gambar pada data evaluasi yang ditunjukkan oleh Tabel 2.

**Tabel 2.** Contoh hasil klasifikasi penggunaan helm

No.	Gambar	Hasil <i>Faster R-CNN Inception V2</i>	Hasil <i>SSD Mobilenet V1</i>
1			
2			
3			

### 3.3 Evaluasi Model

Dalam penelitian ini digunakan format metrik COCO [10] untuk memberikan penilaian lokalisasi sesuai dengan perubahan nilai *Intersect over Union* atau IoU, IoU dimaknai sebagai persamaan 3. Format metrik COCO mengutamakan perhitungan *mean precision* dan *recall* dengan nilai IoU pada *range* 0,5 sampai dengan 0,9 dengan interval 0,05 untuk menggambarkan performa lokalisasi model yang di evaluasi. Selain itu juga format metrik COCO memberikan hasil *mean precision* pada *threshold* IoU 0,5 yang merupakan metrik PASCAL VOC dan IoU *strict* yaitu IoU 0,75 untuk menggambarkan perubahan nilai presisi ketika penilaian presisi dibatasi dengan *threshold* IoU yang lebih tinggi.

**Tabel 3.** Nilai *Average Precision* yang dicapai model

Metrik Ukuran Objek <i>Threshold</i> IoU	AP <i>All Size</i> 0,5	AP <i>All Size</i> 0,75	AP <i>All Size</i> 0,5;0,05;0,9	AP <i>Small</i> 0,5;0,05;0,9	AP <i>Medium</i> 0,5;0,05;0,9	AP <i>Large</i> 0,5;0,05;0,9
<i>Faster R-CNN Inception V2</i>	86,841 %	44,585 %	47,120 %	70,000 %	30,431 %	51,578 %
SSD <i>Mobilenet V1</i>	68,894 %	22,051 %	31,424 %	0 %	7,784 %	35,590 %

Hasil evaluasi nilai presisi dari sistem dengan diberikan data evaluasi ditunjukkan oleh tabel 3 diatas. Pada *threshold* IoU PASCAL VOC yaitu IoU 0,5 sistem dapat menemukan *true positive* lebih banyak sehingga memiliki nilai presisi yang tinggi, terutama pada model *Faster R-CNN Inception V2* dengan nilai presisi 86,841%. Namun ketika diberikan *threshold* IoU yang ketat kedua model mengalami penurunan nilai presisi hingga setengah dari presisi *threshold* IoU standar, hal ini menunjukkan nilai lokalisasi dari kedua model yang cenderung menurun seiring dengan semakin tinggi nilai IoU. Dalam tabel 3 juga ditunjukkan nilai presisi sesuai pada ukuran objek dimana objek *small* ukuran  $< 32 \text{ px}^2$ , objek *medium*  $96 \text{ px}^2 >$  ukuran  $> 32 \text{ px}^2$ , dan objek *large* ukuran  $> 96 \text{ px}^2$

**Tabel 4.** Nilai *Average Recall* yang dicapai model

Metrik Ukuran Objek <i>Max. Detection</i>	AR <i>All Size</i> det = 1	AR <i>All Size</i> det = 10	AR <i>All Size</i> det = 100	AR <i>Small</i> det = 100	AR <i>Medium</i> det = 100	AR <i>Large</i> det = 100
<i>Faster R-CNN Inception V2</i>	34,404 %	61,280 %	62,728 %	70,000 %	49,048 %	65,714 %
SSD <i>Mobilenet V1</i>	26,219 %	46,807 %	49,524 %	0 %	28,410 %	53,738 %

Selain presisi nilai *recall* juga merupakan hasil yang didapat dengan memberikan kedua model input berupa data evaluasi. Nilai yang ditunjukkan pada tabel 4 diatas adalah nilai rata-rata *recall* maksimum dalam range IoU 0,5;0,05;0,9 terhadap seluruh objek pada data evaluasi. Pada deteksi maksimum yaitu 100 objek per gambar model *Faster R-CNN Inception V2* mengungguli model SSD *Mobilenet V1* dengan nilai *Average Recall* sebesar 62,728%. Selain perhitungan *recall* sesuai dengan deteksi maksimum, *recall* juga dihitung sesuai dengan ukuran objek seperti yang telah didefinisikan pada bagian presisi sesuai dengan ukuran objek.

**Tabel 5.** Nilai *Frame Rate* yang dicapai model

Metrik	<i>Local Frame Rate</i>	<i>Frame Rate Ratio</i>
<i>Faster R-CNN Inception V2</i>	1,004 FPS	x1
SSD <i>Mobilenet V1</i>	10,405 FPS	x10,364

Tabel 5 menunjukkan nilai *Frame Rate* rata-rata yang dapat diperoleh oleh kedua model. Walaupun model *Faster R-CNN Inception V2* mengungguli model *SSD Mobilenet V1* dalam presisi dan *recall* namun model *SSD Mobilenet V1* dapat melakukan komputasi lebih cepat sehingga memungkinkan untuk model tersebut mencapai *frame rate* sepuluh kali lebih cepat dibandingkan model *Faster R-CNN Inception V2*, oleh karena itu faktor *frame rate* ini dapat menjadi *trade-off* dari model *SSD Mobilenet V1* untuk penggunaan yang lebih dibutuhkan *frame rate* yang bagus melainkan nilai presisi yang tinggi.

#### 4. Kesimpulan

Dengan melatih model *Faster R-CNN* dan *SSD* menggunakan *dataset* berupa gambar pada tempat konstruksi kedua model dapat mengidentifikasi objek dengan nilai presisi terbaik pada model *Faster R-CNN Inception V2* yang mencapai nilai presisi 86,841% namun dalam performansi *frame rate* model *SSD Mobilenet V1* memiliki kemampuan 10,364 kali lebih baik dibandingkan dengan model *Faster R-CNN Inception V2*, oleh karena itu hasil dari kedua nilai performansi dapat membuktikan *trade-off* antara presisi dan *frame rate* dari kedua model tersebut yang dapat dimanfaatkan untuk penggunaan sistem keselamatan. Selain dari presisi dan *frame rate* perhitungan nilai presisi dengan *COCO metrics* menghasilkan nilai performansi lokalisasi dari kedua model, pada model *Faster R-CNN Inception V2* didapatkan nilai 47,120% sementara untuk model *SSD Mobilenet V1* didapatkan nilai 31,424%.

#### Reference

- [1] Kemnaker RI, "Keputusan Menteri Ketenagakerjaan Republik Indonesia, Nomor 386 Tahun 2014," 2014.
- [2] Kementerian Ketenagakerjaan Republik Indonesia, "Peraturan Menteri Ketenagakerjaan No. 9 tahun 2016 tentang Keselamatan dan kesehatan kerja dalam pekerjaan di ketinggian," no. 386, pp. 1–36, 2016.
- [3] Tri, "Tahun 2018 Angka Kecelakaan Kerja Cenderung Meningkat, BPJS Ketenagakerjaan Bayar Santunan Rp1,2 Triliun," 2019. [Online]. Available: <https://poskotanews.com/2019/01/15/angka-kecelakaan-kerjacenderung-meningkat-bpjs-ketenagakerjaan-bayar-santunan-rp12-triliun/> %0D.
- [4] Q. Fang et al., "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," *Autom. Constr.*, vol. 85, no. September 2017, pp. 1–9, 2018.
- [5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016.
- [7] A. KHAZRI, "Faster RCNN Object detection," 2019. <https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4>.
- [8] W. Liu, "SSD: Single Shot Multibox Detector," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 39, pp. 1137-1149, 2016.
- [9] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. <http://arxiv.org/abs/1704.04861>
- [10] COCO Consortium, "COCO Dataset Detection Evaluation." <https://cocodataset.org/#detection-eval>.