2. LANDASAN TEORI

2.1. Tinjauan Pustaka

Pada bagian ini akan dijelaskan mengenai teori-teori yang dipakai dalam Penerapan *Depth Estimation* untuk Deteksi Keadaan Sekitar Forklift menggunakan Metode YOLOv3.

2.1.1. Forklift

Pada Erick (2021) tertulis, Forklift adalah salah satu jenis truk yang dapat digunakan untuk mengangkat, menurunkan, serta memindahkan barang yang berat dari tempat pertama ke tempat lainnya. Kendaraan ini digunakan untuk mengangkat benda yang terlalu berat atau sulit diangkat sendiri oleh manusia, bisa digunakan baik di luar maupun di dalam ruangan. Bongkar muat barang di gudang, ekspedisi, pabrik, supermarket, dan pelabuhan biasanya membutuhkan bantuan kendaraan pengangkut ini. Forklift sendiri terbagi menjadi beberapa jenis, diantaranya ada forklift diesel, *gasoline*, elektrik, dan *Reach Truck*. Menurut SaferLifts (n. d.) kecepatan yang bisa ditempuh oleh forklift dapat mencapai 20km/jam dalam keadaan di luar ruangan, dan hanya 5km/jam jika dalam ruangan terutama untuk gudang yang barang-barang kimia atau bahan-bahan produksi.

Gambar 2.1 Forklift

Sumber : BigJoe. (2023). *Forklift*. https://bigjoeforklifts.com/cdn/shop/products/DSC_2031_1024x1024.png?v=1610332885

2.1.2. Opencv

Opencv (*Open Source Computer Vision Library*) adalah sebuah *open source* untuk memproses *computer vision* disertai dengan *software library* berisi *machine learning*. Opencv dibangun untuk menyediakan infrastruktur umum untuk pengaplikasian *computer vision* dan mempercepat penggunaan persepsi mesin dalam produk komersial. Dengan menjadi produk berlisensi Apache 2, Opencv memudahkan bisnis untuk memanfaatkan dan memodifikasi kode.

2.1.3. *Image Processing*

Dalam Mulyawan (2021), *image processing* dapat diartikan sebagai metode untuk mengubah citra menjadi bentuk digital dan melakukan beberapa operasi padanya, untuk mendapatkan citra yang disempurnakan atau untuk mengekstrak beberapa informasi berguna darinya. Inputnya berupa gambar, seperti *frame*, video, atau foto dan outputnya dapat berupa gambar atau karakteristik yang terkait dengan gambar tersebut. Ada beberapa fase atau tahapan dalam melakukan image processing, diantaranya :

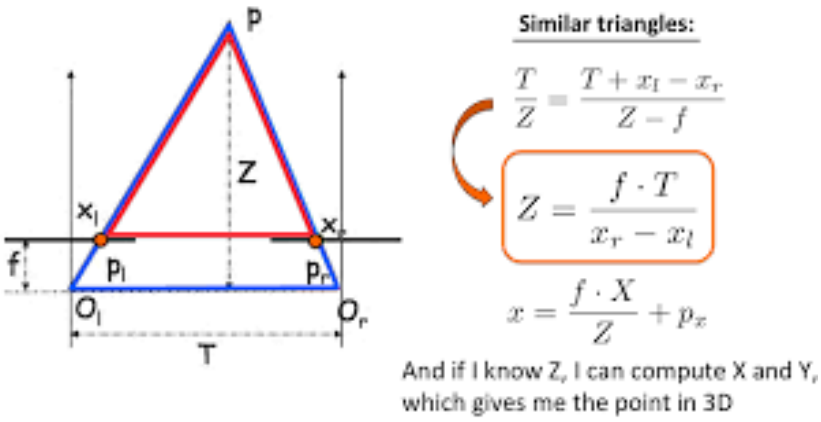
a. *Acquisition*: Dalam tahapan ini, pekerjaan utamanya meliputi *scaling* dan *color conversion*, sebagai contoh misalnya RGB ke gray atau sebaliknya
b. *Enhancement*: Digunakan untuk mengekstrak beberapa detail tersembunyi dari sebuah gambar dan bersifat subjektif
c. *Restoration*: Berhubungan dengan daya tarik gambar, tetapi bersifat objektif
d. *Color image processing*: Berkaitan dengan *Pseudocolor* dan model warna pemrosesan gambar penuh warna berlaku untuk pemrosesan gambar digital
e. *Wavelets* dan *multi-resolution processing*: Merupakan dasar untuk mempresentasikan gambar dalam berbagai derajat
f. *Compression*: Mengembangkan beberapa fungsi, terutama berkaitan dengan ukuran atau resolusi gambar
g. *Morphological processing*: Berkaitan dengan alat untuk mengekstraksi komponen gambar yang berguna dalam representasi dan deskripsi bentuk
h. *Segmentation procedure*: Mempartisi gambar menjadi bagian atau objek penyusunnya. Perlu kalian ketahui bahwa segmentasi otonom, biasanya adalah tugas tersulit dalam *image processing*.

i. *Representation*: Mengikuti tahap *output* (keluaran) segmentasi. Pemilihan representasi hanyalah bagian dari solusi untuk mengubah data *raw* (mentah) menjadi data yang sudah diproses.

j. *Object detection* dan *recognition*: Tahapan ini adalah proses yang akan memberikan label ke suatu objek berdasarkan deskriptornya.

2.1.4. Depth Estimation

Menurut Vasiljevic, et al. (2019), depth estimation adalah tugas untuk mengukur jarak setiap piksel relatif terhadap kamera. Kedalaman diekstraksi dari gambar monokuler (Tunggal) atau stereo (beberapa tampilan suatu pandangan). Metode tradisional menggunakan geometri multi-tampilan untuk menemukan hubungan antar gambar. Metode yang lebih baru dapat memperkirakan kedalaman secara langsung dengan meminimalkan kerugian regresi, atau dengan belajar menghasilkan tampilan baru dari suatu rangkaian. Gambar 2.2 berikut merupakan rumus yang dapat digunakan dalam *depth estimation*



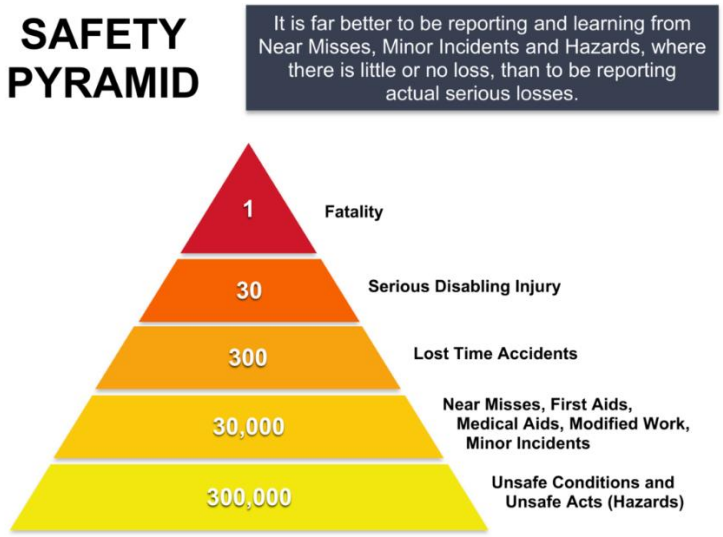
Gambar 2.2 Rumus *depth estimation*

Sumber : Fidler, S. (n. d.). *Depth From Stereo*.
http://www.cs.toronto.edu/~fidler/slides/2015/CSC420/lecture12_hres.pdf, p. 18

2.1.5. Safety Pyramid

Dalam Kohler (2019) tertulis, *Safety pyramid* adalah teori bahwa ada hubungan langsung antara kecelakaan serius, kecelakaan kecil, dan nyaris celaka. Herbert Heinrich, pelopor dalam kesehatan dan keselamatan tempat kerja, pertama kali mengusulkan hubungan tersebut pada tahun 1931 dengan menentukan bahwa jika kecelakaan kecil

dikurangi maka akan terjadi penurunan yang sesuai dalam jumlah kecelakaan serius. Heinrich mengusulkan hubungan satu kecelakaan cedera berat dengan 29 kecelakaan cedera ringan, dengan 300 kecelakaan tanpa cedera. Hubungan tersebut sering ditampilkan secara piktorial dalam bentuk segitiga atau piramida. Piramida keselamatan telah disebut sebagai landasan kesehatan dan keselamatan selama 80 tahun terakhir atau lebih. Banyak sistem keselamatan memasukkan premis bahwa pelaporan dan penanganan insiden nyaris celaka dan penyebab perilakunya hampir dapat menghilangkan kecelakaan besar. Gambar 2.3 berikut merupakan gambaran safety pyramid



Gambar 2.3 *Safety Pyramid*

Sumber: InUnison. (2023). *Safety Pyramid*. https://inunison.io/wpcontent/uploads/2019/10/HS-Pyramid-1024x747.png, p. 1

2.1.6. Convolutional Layer

Menurut Goodfellow (2016), Layer konvolusi (convolutional layer) adalah komponen penting dalam jaringan saraf tiruan konvolusi (convolutional neural network atau CNN). Layer ini digunakan untuk mengekstraksi fitur-fitur penting dari data input, seperti gambar. Konvolusi adalah operasi matematis yang memungkinkan model untuk memahami hubungan spasial antara elemen-elemen dalam input, seperti piksel dalam gambar.

Berikut adalah beberapa konsep penting tentang convolutional layer:

1. Filter (Kernel): Filter adalah matriks kecil yang digunakan untuk melakukan operasi konvolusi pada data input. Filter ini bergerak melintasi seluruh input, dan

pada setiap langkah, mengalikan elemen-elemen input dengan elemen-elemen dalam filter dan kemudian menjumlahkannya. Ini membantu mengidentifikasi pola-pola dalam input.

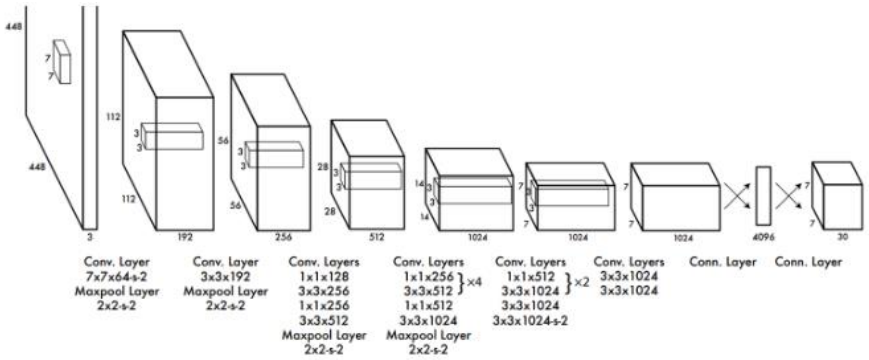
2. Stride : Stride adalah langkah yang digunakan oleh filter saat bergerak melintasi input. Jika stride adalah 1, filter akan bergerak satu langkah piksel pada setiap iterasi. Jika stride lebih besar dari 1, filter akan melompati beberapa piksel pada setiap iterasi. Stride ini mempengaruhi ukuran output dari layer konvolusi.
3. Padding : Padding adalah proses menambahkan piksel-piksel nol di sekitar input sebelum menjalankan operasi konvolusi. Ini berguna untuk mempertahankan ukuran output yang sama dengan input asli, sehingga menghindari penurunan ukuran yang signifikan.
4. *Depth (Channel)* : Data input dan filter biasanya memiliki tiga dimensi: tinggi (*height*), lebar (*width*), dan kedalaman (*depth*). Kedalaman ini mengacu pada jumlah channel atau *feature maps* dalam data input. Setiap channel mewakili informasi khusus dalam data, seperti warna dalam gambar RGB.
5. *Activation Function*: Setelah operasi konvolusi selesai, hasilnya akan diteruskan melalui fungsi aktivasi seperti ReLU (Rectified Linear Unit) untuk memperkenalkan non-linearitas ke dalam jaringan.
6. *Pooling*: Setelah beberapa layer konvolusi, biasanya dilakukan operasi pooling seperti *max-pooling* atau *average-pooling* untuk mengurangi ukuran spatial data, sehingga mengurangi jumlah parameter dan mempertahankan fitur-fitur yang paling penting.

2.1.7. YOLO (You Only Look Once)

Menurut Aggarwal (2020), YOLO atau *You Only Look Once*, adalah algoritma deteksi objek real-time yang populer. YOLO menggabungkan proses multi-langkah, menggunakan jaringan syaraf tunggal untuk melakukan klasifikasi dan prediksi kotak pembatas untuk objek yang terdeteksi. Oleh karena itu, ini sangat dioptimalkan untuk kinerja deteksi dan dapat berjalan lebih cepat daripada menjalankan dua jaringan saraf terpisah untuk mendeteksi dan mengklasifikasikan objek secara terpisah. Hal ini dilakukan dengan menggunakan kembali pengklasifikasi gambar tradisional untuk digunakan dalam tugas regresi dalam mengidentifikasi kotak pembatas objek. Selain itu, YOLO dapat

menggeneralisasi representasi berbagai objek, sehingga lebih dapat diterapkan di berbagai lingkungan baru.

Arsitektur dari algoritma YOLO menggunakan Convolutional Neural Network yang memiliki 24 convolutional layers dan diikuti oleh 2 fully connected layers. Convolutional layer berfungsi untuk mengekstraksi fitur dari input gambar, sedangkan fully connected layer berperan dalam memprediksi probabilitas output dan koordinat (Nissa, 2023). Gambar 2.4 berikut merupakan arsitektur dari algoritma YOLO.



Gambar 2.4 Arsitektur Algoritma YOLO

Sumber : Nissa, N. K. (2023). *Cara Kerja Object Detection dengan YOLO (You Only Look Once)*. https://pacmann.io/blog/cara-kerja-object-detection-dengan-yolo, p. 1

2.1.8 Roboflow

Dalam (Roboflow, 2024) Roboflow adalah aplikasi berbasis web yang digunakan untuk membuat model computer vision. Roboflow memberdayakan developer untuk membangun aplikasi computer vision mereka sendiri, terlepas dari keahlian atau pengalaman. Roboflow menyediakan semua alat yang dibutuhkan mulai dari ide hingga model computer vision yang tangguh untuk diterapkan. Beberapa fitur yang disediakan pada roboflow antara lain adalah Roboflow Train, Roboflow Datasets, Roboflow Deploy, Workspaces, API Key, dll.

2.2 Tinjauan Studi

Dalam sub bab ini berisi pembahasan mengenai beberapa penelitian yang telah dilakukan sebelumnya.

2.2.1 Estimasi Posisi dengan Menggunakan Kamera Monokular (Afrisal, 2019)

Permasalahan yang diangkat dalam penelitian ini adalah pendekatan tradisional menggunakan dua kamera atau kamera 3D dengan sensor kedalaman, yang memerlukan sensor tambahan dan perangkat keras yang mahal. Dalam konteks ini, permasalahan yang dihadapi adalah bagaimana melakukan estimasi posisi dengan cara yang lebih sederhana dan terjangkau menggunakan kamera tunggal (monokular) tanpa mengorbankan akurasi.

Metode yang digunakan dalam penelitian ini adalah digunakan pendekatan monokular (kamera tunggal) dengan metode sequential multiple view geometry. Teknik ini memiliki tujuan untuk mengestimasi posisi objek berdasarkan transformasi citra dari sudut pandang yang berbeda.

Hasil dari penelitian ini adalah pendekatan monokular dengan menggunakan model kamera pinhole dan metode sequential multiple view geometry mampu melakukan estimasi pose dan posisi dengan cukup akurat. Dalam pengujian terhadap citra papan catur (checkerboard), estimasi jarak dan sudut memiliki galat yang relatif kecil, yaitu sebesar 2,9 mm untuk estimasi jarak dan 0,3° untuk estimasi sudut.
Dalam skripsi ini dilakukan pengembangan pada kegunaan sistem untuk depth estimation, bukan hanya untuk estimasi posisi suatu objek terhadap kamera yang terletak pada bagian forklift.

2.2.2 Monocular Depth Estimation pada Scene dalam Ruangan menggunakan U-net dengan Resnet (Pinasthika, 2022)

Permasalahan yang diangkat dalam penelitian ini adalah bagaimana mengembangkan solusi yang lebih terjangkau dan efisien dalam mengestimasi jarak terhadap objek, dengan menggunakan kamera monokular sebagai alternatif.

Metode yang digunakan dalam penelitian ini adalah pendekatan menggunakan kamera monokular untuk mengestimasi nilai kedalaman pada citra dengan menggunakan metode *Deep Neural Networks* (DNN). Spesifiknya, arsitektur DNN U-Net dengan penggunaan *Residual Network* (ResNet) pada bagian encoder digunakan.

Hasil dari penelitian ini adalah model yang mampu mengestimasi kedalaman citra dengan akurat. Model yang dihasilkan mampu bersaing dengan penelitian sebelumnya, dengan nilai *Root Mean Squared Error* (RMSE) sebesar 0.2272, *Relative Error in Depth* (REL) sebesar 1.3676, akurasi dengan threshold 1.25 sebesar 56.22%, akurasi dengan threshold 1.252 sebesar 78.97%, dan akurasi dengan threshold 1.253 sebesar 89.29%.

Pengujian inferensi model juga menunjukkan kinerja yang memadai, dengan jumlah frames per second (FPS) yang berkisar antara 5-12 FPS.

Dalam skripsi ini dilakukan pengaplikasian depth estimation akan dilakukan terhadap barang-barang yang ada pada pabrik PT Henkel untuk mendeteksi area sekitar forklift, tentunya dengan sudut tertentu untuk estimasi jarak.

2.2.3 Model Deteksi Wajah (Face Tracking) dan Pengukuran Jarak Wajah (Distance Estimation) Secara Realtime menggunakan 3D Stereo Vision Camera untuk Face Robotic System (Winarno, 2014)

Permasalahan yang diangkat dalam penelitian ini adalah deteksi wajah dan estimasi jarak terhadap wajah manusia dalam gambar atau video, baik dalam konteks diam maupun secara realtime. Penelitian ini fokus pada pengembangan sistem pendeteksian wajah yang dapat digunakan dalam visi komputer, navigasi, dan robotika. Kendala yang ingin diatasi adalah bagaimana mendeteksi wajah manusia dalam gambar dan mengestimasi jaraknya menggunakan algoritma yang efisien dan akurat.

Metode yang digunakan dalam penelitian ini adalah stereo vision camera untuk mendeteksi dan mengukur jarak objek wajah dalam gambar. Dalam hal ini, estimasi jarak dihitung berdasarkan perhitungan jarak proyeksi dari gambar 2 dimensi menjadi gambar 3 dimensi menggunakan 2 titik lensa pada stereo vision camera. Penggabungan dua framework, yaitu face tracking dan distance estimation, digunakan untuk mengembangkan sistem deteksi dan estimasi jarak obyek wajah.

Hasil dari penelitian ini adalah pengembangan sebuah sistem yang dapat melakukan tracking terhadap wajah manusia dan secara simultan melakukan estimasi jarak terhadap wajah tersebut menggunakan kamera stereo vision 3D. Sistem ini memberikan dasar bagi pengembangan sistem robotika yang berkaitan dengan wajah manusia, seperti robot penyaji, robot pencari wajah, dan sistem robot lainnya yang beroperasi berdasarkan deteksi dan pengenalan wajah manusia.

Dalam skripsi ini dilakukan pengaplikasian depth estimation akan dilakukan terhadap barang-barang yang ada pada pabrik PT Henkel untuk mendeteksi area sekitar forklift. Metode yang digunakan juga berbeda yakni menggunakan monocular