

**UNIVERSITAS GUNADARMA**  
**FAKULTAS TEKNOLOGI INDUSTRI**



**ESTIMASI POSE TIGA DIMENSI DARI GAMBAR  
MONOKULER MENGGUNAKAN DEEP NEURAL  
NETWORK**

**Disusun oleh:**

<b>Nama</b>	<b>: Denilson</b>
<b>NPM</b>	<b>: 51416815</b>
<b>Jurusan</b>	<b>: Teknik Informatika</b>
<b>Pembimbing</b>	<b>: Dr. Dharmayanti, ST., MMSI.</b>

**Diajukan Guna Melengkapi Sebagian Syarat  
Dalam Mencapai Gelar Sarjana Strata Satu (S1)**

**Depok**

**2020**

## **LEMBAR ORIGINALITAS DAN PUBLIKASI**

Saya yang bertanda tangan di bawah ini,

Nama : Denilson  
NPM : 51416815  
Judul Penulisan Ilmiah : Estimasi Pose Tiga Dimensi dari Gambar Monokuler  
Menggunakan Deep Neural Network  
Tanggal Sidang : tanggal  
Tanggal Lulus : tanggal

menyatakan bahwa tulisan ini adalah merupakan hasil karya saya sendiri dan dapat dipublikasikan sepenuhnya oleh Universitas Gunadarma. Segala kutipan dalam bentuk apa pun telah mengikuti kaidah, etika yang berlaku. Mengenai isi dan tulisan adalah merupakan tanggung jawab Penulis, bukan Universitas Gunadarma.

Demikian pernyataan ini dibuat dengan sebenarnya dan dengan penuh kesadaran.

Depok, April 2020

Denilson

## **LEMBAR PENGESAHAN**

### **Komisi Pembimbing**

No	Nama	Kedudukan
1	Dr. Dharmayanti, ST., MMSI.	Ketua
2	DIGANTI NAMA PENGUJI 2	DIGANTI JABATAN PENGUJI 2
3	DIGANTI NAMA PENGUJI 3	DIGANTI JABATAN PENGUJI 3

Tanggal Sidang : tgl bln thn

### **Panitia Ujian**

No	Nama	Kedudukan
1	DIGANTI NAMA PENGUJI 1	DIGANTI JABATAN PENGUJI 1
2	DIGANTI NAMA PENGUJI 2	DIGANTI JABATAN PENGUJI 2
3	DIGANTI NAMA PENGUJI 3	DIGANTI JABATAN PENGUJI 3
4	DIGANTI NAMA PENGUJI 4	DIGANTI JABATAN PENGUJI 4
5	DIGANTI NAMA PENGUJI 5	DIGANTI JABATAN PENGUJI 5

Tanggal Lulus : tgl bln thn

## **MENGETAHUI**

Pembimbing

Bagian Sidang Sarjana

(Dr. Dharmayanti, ST., MMSI.)

(NAMA BAGIAN SARJANA)

## **ABSTRAKSI**

Denilson, 51416815

ESTIMASI POSE TIGA DIMENSI DARI GAMBAR MONOKULER  
MENGUNAKAN DEEP NEURAL NETWORK

Tugas Akhir. Jurusan Teknik Informatika, Fakultas Teknologi Industri,  
Universitas Gunadarma, 2020

Kata Kunci : dibuaturut abjad sekitar 3-5 kata kunci  
(jml hlm romawi + jml hlm arab + Lampiran)

Abstraksi.

Daftar Pustaka (thn terlama-thn terbaru)

## **ABSTRACT**

Denilson, 51416815

THREE DIMENSIONAL POSE ESTIMATION FROM MONOCULAR IMAGE  
USING DEEP NEURAL NETWORK

Thesis. Informatics Engineering, Faculty of Industrial Technology,  
Gunadarma University, 2020

Keywords: dibuaturut abjad sekitar 3-5 kata kunci  
(jml hlm romawi + jml hlm arab + Lampiran)

Abstract.

Bibliography (thn terlama-thn terbaru)

## KATA PENGANTAR

Segala puji dan syukur penulis ucapkan ke hadirat Tuhan Yang Maha Esa yang telah memberikan berkat, anugerah dan karunia yang melimpah, sehingga penulis dapat menyelesaikan Tugas Akhir ini pada waktu yang telah ditentukan.

Tugas Akhir ini disusun guna melengkapi sebagian syarat untuk memperoleh gelar Sarjana Teknik Informatika Universitas Gunadarma. Adapun judul Tugas Akhir ini adalah "Estimasi Pose Tiga Dimensi Dari Gambar Monokuler Menggunakan Deep Neural Network".

Walaupun banyak kesulitan yang penulis harus hadapi ketika menyusun Tugas Akhir ini, namun berkat bantuan dan dorongan dari berbagai pihak, akhirnya Tugas Akhir ini dapat diselesaikan dengan baik. Untuk itu penulis tidak lupa mengucapkan terima kasih kepada:

1. Ibu Prof. E. S. Margianti, SE, MM selaku rektor Universitas Gunadarma
2. .... selaku Dekan Fakultas ..... Universitas Gunadarma
3. .... selaku Ketua Jurusan .....
4. .... selaku Bagian Sidang Sarjana
5. Ibu Dr. Dharmayanti, ST., MMSI sebagai pembimbing penulis yang ditengah-tengah kesibukannya telah membimbing penulis sehingga penulisan ini dapat diselesaikan.
6. Keluarga yang selalu mendukung dan terus memberikan motivasi.
7. Semua pihak yang terlibat dalam membantu penyelesaian Tugas Akhir ini.

Sebagai manusia biasa yang tak luput dari kesalahan, maka penulis meminta maaf atas segala kekurangan dan keterbatasan dalam penyusunan Tugas Akhir ini. Penulis sadari bahwa penulisan ini masih jauh dari sempurna, disebabkan karena berbagai keterbatasan yang penulis miliki. Untuk itu penulis mengharapkan kritik

dan saran yang bersifat membangun untuk menjadi perbaikan di masa yang akan datang.

Akhir kata, penulis berharap penulisan ini dapat bermanfaat bagi kita semua dan bagi penulis pribadi khususnya, serta dapat digunakan sebagaimana mestinya.

Depok, April 2020

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	i
<b>LEMBAR ORIGINALITAS DAN PUBLIKASI</b>	ii
<b>LEMBAR PENGESAHAN</b>	iii
<b>ABSTRAKSI</b>	iv
<b>ABSTRACT</b>	v
<b>KATA PENGANTAR</b>	vi
<b>DAFTAR ISI</b>	ix
<b>DAFTAR TABEL</b>	x
<b>DAFTAR GAMBAR</b>	xi
<b>DAFTAR LAMPIRAN</b>	xii
<b>BAB I : PENDAHULUAN</b>	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Metode Penelitian	2
1.6 Sistematika Penulisan	4
<b>BAB II : TINJAUAN PUSTAKA</b>	5
2.1 Teorema Penaksiran Universal	5
2.2 Jaringan Saraf	5
2.3 Jaringan Saraf Tiruan	6
2.4 Fungsi Aktivasi	8
2.5 <i>Residual Network</i>	9
2.6 Optimisasi Model	9
2.6.1 <i>Backpropagation</i>	10
2.6.2 <i>Learning Rate</i>	11
2.6.3 <i>Mean Squared Error</i>	11



2.7	Estimasi Pose Dua Dimensi . . . . .	11
2.8	Estimasi Pose Tiga Dimensi . . . . .	13
2.9	PyTorch . . . . .	14
<b>BAB III</b>	<b>: PENDEKATAN . . . . .</b>	<b>15</b>
3.1	Langkah-Langkah yang dilakukan . . . . .	15
3.2	Alur Pelatihan Model . . . . .	15
3.3	Pra Pemrosesan Data . . . . .	15
3.4	Arsitektur . . . . .	15
3.5	Pelatihan Model . . . . .	15
<b>BAB IV</b>	<b>: HASIL DAN ANALISIS . . . . .</b>	<b>16</b>
4.1	Plot Grafik . . . . .	16
4.2	Matplotlib . . . . .	16
<b>BAB V</b>	<b>: PENUTUP . . . . .</b>	<b>17</b>
5.1	Kesimpulan . . . . .	17
5.2	Saran . . . . .	17
<b>DAFTAR PUSTAKA</b>	<b>. . . . .</b>	<b>18</b>
<b>LAMPIRAN</b>	<b>. . . . .</b>	<b>L1</b>

## **DAFTAR TABEL**

3.1	Sebuah tabel . . . . .	15
-----	------------------------	----

## DAFTAR GAMBAR

2.1	Ilustrasi Jaringan Saraf Manusia . . . . .	6
2.2	Ilustrasi Sel Saraf Manusia . . . . .	6
2.3	Ilustrasi Jaringan Saraf Tiruan . . . . .	7
2.4	Ilustrasi Sel Saraf Tiruan . . . . .	7
2.5	<i>Rectified Linear Unit</i> . . . . .	8
2.6	<i>Residual Block</i> . . . . .	9
2.7	<i>Gradient Descent</i> . . . . .	10
2.8	Skema <i>Backpropagation</i> . . . . .	10
2.9	Ilustrasi Perbedaan <i>Learning Rate</i> . . . . .	11
2.10	Pendekatan <i>Top-Down</i> dan <i>Bottom-Up</i> . . . . .	12
2.11	Pencarian Pose Tiga Dimensi Lokal . . . . .	13
2.12	Pencarian Pose Tiga Dimensi Global . . . . .	14

## **DAFTAR LAMPIRAN**

Lampiran 1 Listing Program .....	10
----------------------------------	----

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Pemanfaatan teknologi yang terkomputerisasi oleh manusia selalu meninggalkan rekam jejak yang tersimpan dalam bentuk data digital. Data digital yang umumnya dimanfaatkan oleh manusia meliputi teks, citra audio, citra visual, dan citra audio visual yang dimampatkan kedalam suatu media penyimpanan. Pengaksesan data ini tergolong mudah karena data dikonstruksi dengan struktur dan format yang tertentu sehingga dapat diakses dimana saja. Jumlah rekam jejak digital yang tersedia semakin bertambah seiring dengan mudahnya akses terhadap teknologi secara skala besar.

Rekam jejak digital yang terkumpul bersifat laten yang berarti data yang tersedia tidak memiliki arti semantik. Data ini hanya akan berguna apabila ada suatu aturan yang ditentukan oleh pengguna. Proses pembuatan citra visual terbentuk dari hasil perekaman kamera dengan mengubah gelombang elektromagnetik yang dipantulkan oleh benda dari posisi tertentu menuju lensa kamera menjadi tiga lapis deretan angka numerik yang mewakili warna merah, hijau, dan biru. Citra visual yang tercipta hanya memiliki informasi numerik mengenai warna dalam bentuk gambar, sedangkan informasi posisi benda saat perekaman sudah hilang. Otak manusia dapat mengartikan kondisi suatu benda dalam suatu gambar tanpa kita ketahui cara kerjanya secara pasti. Hal yang sama dapat dilakukan oleh komputer dengan membuat pemetaan dari suatu gambar ke suatu kondisi yang diinginkan.

Permodelan pembelajaran dalam atau *deep learning* dapat memetakan suatu domain ke domain lainnya secara otomatis menggunakan pelatihan jaringan saraf tiruan dalam atau *deep neural network*. Pembelajaran dalam menggunakan jaringan saraf tiruan dalam dapat digunakan untuk memajukan teknologi khususnya pada bidang visi komputer seperti melakukan estimasi pose tiga dimensi tubuh manusia yang terdapat dalam suatu gambar.

## 1.2 Rumusan Masalah

Permasalahan yang ingin diselesaikan adalah mengembalikan informasi posisi pose tubuh manusia yang telah hilang dengan membuat pemetaan dari gambar yang ditangkap menggunakan kamera monokuler ke koordinat setiap titik kunci dari pose. Model yang telah dilatih harus mampu melakukan proses pemetaan terhadap gambar baru.

## 1.3 Batasan Masalah

Penelitian ini menganggap setiap pose dua dimensi maupun pose tiga dimensi berada dalam koordinat lokal. Setiap pose ditransformasi ke dalam observasi kamera dengan titik kunci pinggang sebagai posisi tengah. Hal ini dilakukan karena pemetaan hanya menggunakan grafik datar tanpa informasi kedalaman titik kunci sehingga dapat menghindari masalah kedalaman yang ambigu.

## 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membuat sebuah aplikasi yang dapat membaca titik kunci dari sebuah citra visual datar dan melakukan transformasi ke pose lokal dua dimensi sehingga dapat dipetakan oleh jaringan saraf tiruan yang dimodelkan ke bentuk pose lokal tiga dimensi. Pose hasil juga divisualisasikan secara interaktif sehingga dapat dipergunakan untuk kepentingan yang sesuai.

## 1.5 Metode Penelitian

Penelitian dibagi menjadi tiga tahap besar terurut yang terdiri dari *data preprocessing*, pelatihan model jaringan saraf tiruan, dan visualisasi. Tahap pertama dan tahap ketiga tidak melibatkan pemelajaran mesin. Masalah utama dari penelitian ini berada pada tahap kedua tentang pelatihan jaringan saraf tiruan untuk melakukan pemetaan pose dua dimensi ke pose tiga dimensi.

*Dataset* yang digunakan dalam penelitian ini adalah *Human3.6M* yang berisi 3,6 juta pose unik yang dilakukan oleh sebelas aktor profesional dan

direkam menggunakan empat sudut kamera yang berbeda beserta dengan koordinat setiap titik kunci dari hasil penangkapan alat *motion capture* [11].

Pose dua dimensi dan tiga dimensi merupakan variabel yang relevan untuk masalah. Pada tahap *data preprocessing* akan dilakukan ekstraksi variabel ini kedalam bentuk numerik sehingga mudah untuk digunakan saat melakukan pelatihan jaringan saraf tiruan. Tahap selanjutnya akan dilakukan pembuatan, permodelan, pelatihan, dan evaluasi jaringan saraf tiruan dalam untuk pemetaan titik kunci yang kemudian dilanjutkan dengan percobaan model dengan video. Tahap terakhir menampilkan visualisasi gambar, pose dua dimensi, dan pose tiga dimensi.

Perangkat keras yang digunakan dalam penelitian ini adalah satu unit laptop dengan spesifikasi:

- CPU Intel Core I7 7700HQ
- Memori 24 GB DDR4
- GPU NVIDIA GTX 1060 6GB
- SSD NVME SAMSUNG 120 GB
- HDD SATA 1 TB

Perangkat lunak yang digunakan dalam penelitian ini adalah satu unit laptop dengan spesifikasi:

- Python 3.7
- Jupyter Lab
- Git
- GitHub
- LaTeX

## **1.6 Sistematika Penulisan**

Adapun sistematika penulisan yang digunakan dalam penulisan ini adalah sebagai berikut:

PENDAHULUAN, mengemukakan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

TINJAUAN PUSTAKA, menjelaskan kumpulan teori yang digunakan dalam mendukung proses penyelesaian program.

PENDEKATAN, mengemukakan langkah-langkah yang dicapai untuk membuat program.

HASIL DAN ANALISIS, mendalami hasil yang tercapai dengan analisis secara mendalam.

PENUTUP, mengulas lebih lanjut mengenai kesimpulan yang dapat ditarik dari hasil disertai dengan saran yang dapat menyempurnakan penelitian selanjutnya.



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Teorema Penaksiran Universal**

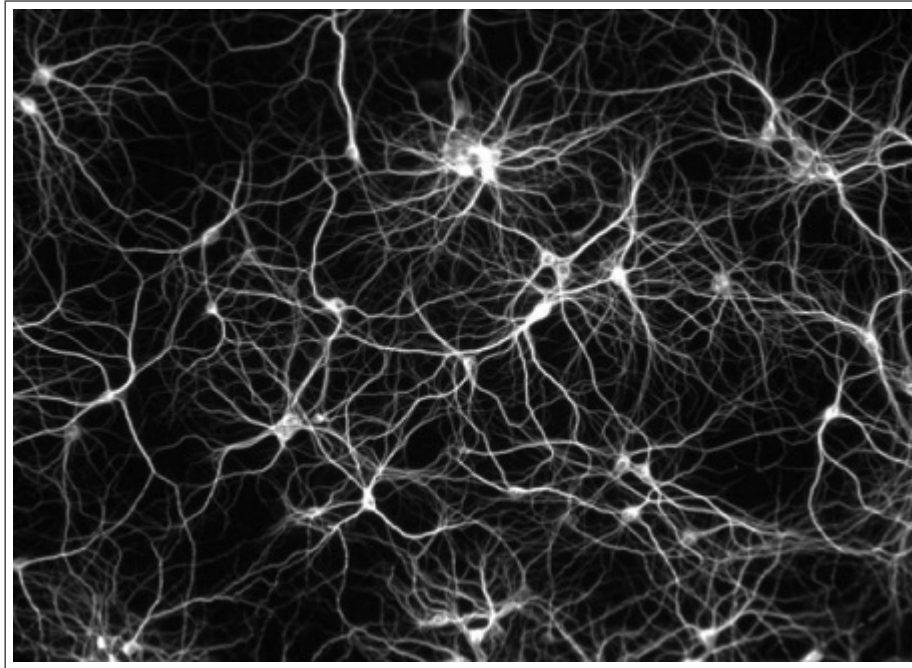
Teorema penaksiran universal atau *universal approximation theorem* menyatakan bahwa sebuah model jaringan *feed-forward* dapat membentuk fungsi apapun secara subjektif. Sebuah model jaringan saraf tiruan dibentuk dari serangkaian lapisan yang didalamnya terdapat deretan sel saraf atau *neuron* dengan kuantitas tertentu. Semakin panjang rangkaian lapisan yang tersedia, maka semakin banyak saraf yang tersedia sehingga dapat memetakan fungsi yang sulit. Model jaringan yang memiliki banyak saraf dapat mempelajari pola-pola yang ada dari satu domain ke domain lainnya [7].

Teorema penaksiran universal memiliki dua sifat yang dikategorikan berdasarkan pemanfaatannya dalam melakukan pembelajaran mesin. Sifat pertama adalah suatu model jaringan saraf tiruan dapat memperkirakan suatu fungsi dengan batasan-batasan tertentu sesuai dengan fungsi aktivasi pada lapisan terakhir. Sifat kedua adalah sebuah fungsi kontinu dengan jumlah variabel sembarang dapat ditiru oleh sebuah jaringan saraf tiruan dengan jumlah yang sembarang [13].

#### **2.2 Jaringan Saraf**

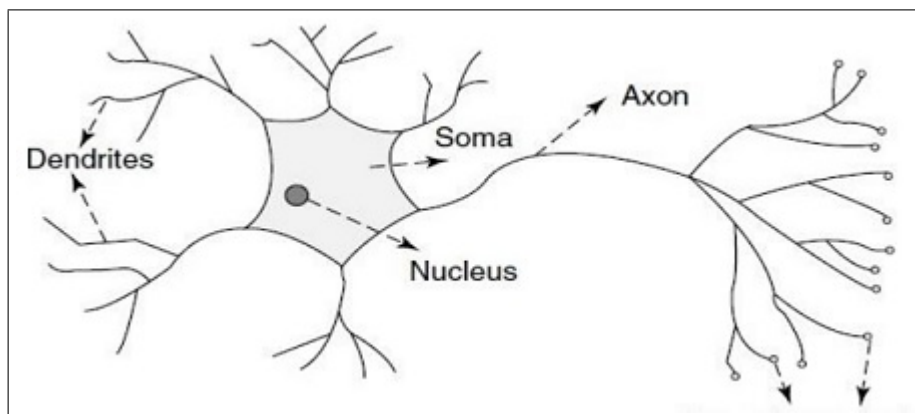
Otak manusia terdiri dari kumpulan sel saraf yang saling terkoneksi satu sama lain. Sebuah sel saraf adalah sel yang dapat memproses dan mengantarkan informasi apabila dirangsang dengan tegangan elektrokimia. Sel-sel saraf tidak pernah memperbanyak dirinya dan tidak digantikan apabila ada yang rusak. Jumlah sel saraf yang terdapat dalam otak manusia diperkirakan sebanyak satu miliar. Setiap sel saraf diperkirakan berkoneksi dengan sepuluh ribu sel saraf lainnya melalui sinapsis yang berarti otak manusia beroperasi seperti prosesor dengan kecepatan satu triliun bit per detik [9].

Bentuk sel saraf sangat bervariasi dengan berbagai ukuran, bentuk, dan sifat elektrokimianya. Sebuah sel saraf memiliki badan yang terdiri dari beberapa



**Gambar 2.1: Ilustrasi Jaringan Saraf Manusia**

struktur penting meliputi *soma*, *dendrites*, *axon*, dan *synapses* seperti pada gambar 2.2. Sebuah sel saraf akan menerima beberapa masukan melalui *synapses*, memproses inputan tersebut melewati *dendrites*, kemudian diteruskan melalui *soma*, dan diberikan kepada sel saraf lainnya melalui *axon* [23].

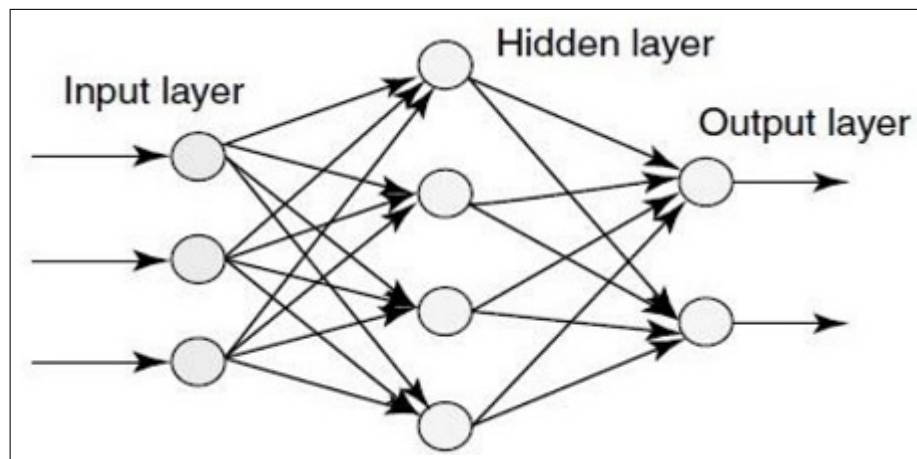


**Gambar 2.2: Ilustrasi Sel Saraf Manusia**

### 2.3 Jaringan Saraf Tiruan

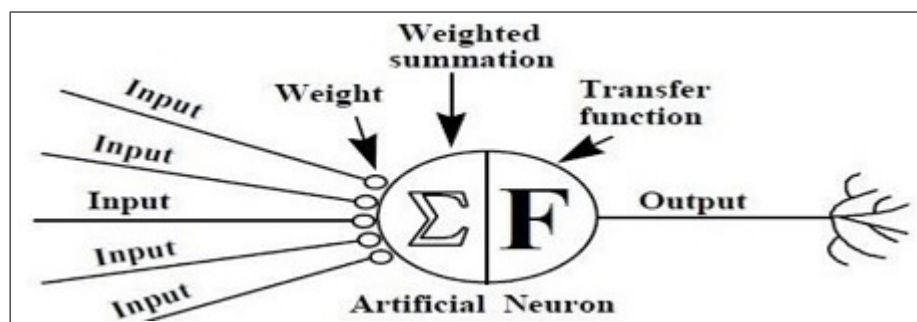
Jaringan Saraf Tiruan adalah sistem komputasi yang cara kerjanya menyerupai jaringan saraf pada otak makhluk hidup. Sebuah jaringan saraf tiruan

dapat dengan otomatis memodelkan fungsi sembarang yang tingkat kesulitannya sesuai dengan jumlah koneksi yang tersedia. Jaringan ini menyerupai jaringan saraf asli dimana sebuah saraf tiruan menerima banyak masukan dari saraf lainnya kemudian dioperasikan dengan bobot yang terkandung pada sel tersebut dan akhirnya diteruskan ke sel berikutnya. Jaringan saraf umumnya dibentuk kedalam beberapa lapisan seperti pada gambar 2.3 [1].



**Gambar 2.3: Ilustrasi Jaringan Saraf Tiruan**

Sebuah sel saraf dapat dibagi menjadi empat bagian yang meliputi masukan, bobot, fungsi transfer atau aktivasi, dan keluaran. Jumlah masukan pada suatu sel saraf tiruan berjumlah sebanyak output dari sel-sel yang berada pada layer sebelumnya. Bobot sel adalah nilai numerik yang menjadi identitas dari sel yang merupakan hasil penyesuaian dari proses latihan. Fungsi aktivasi adalah sebuah fungsi menerima hasil operasi antara bobot dan masukan. Keluaran merupakan hasil dari sel yang diteruskan ke lapisan selanjutnya. Ilustrasi sebuah sel saraf tiruan dapat dilihat pada gambar 2.4 [20].



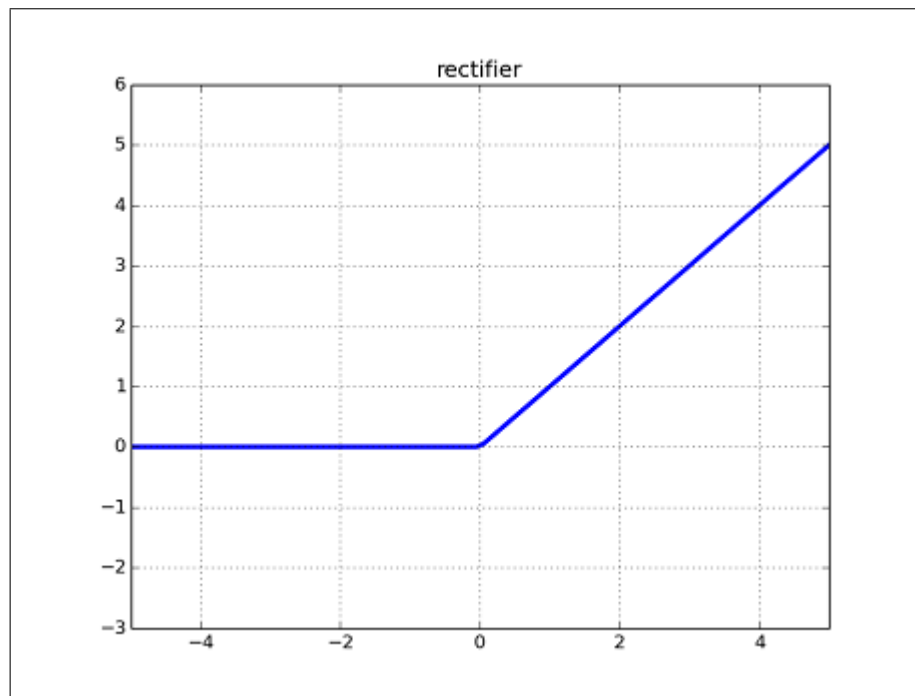
**Gambar 2.4: Ilustrasi Sel Saraf Tiruan**

## 2.4 Fungsi Aktivasi

Fungsi aktivasi pada sel saraf tiruan berfungsi untuk mengkonversi hasil operasi matriks antara masukan dan bobot sebuah sel dari sistem linier ke sistem nonlinier. Konversi nilai ini dilakukan agar setiap sel mengambil perannya pada saat proses pelatihan model. Beberapa fungsi aktivasi yang dipakai pada umumnya adalah *Sigmoid*, *ReLU*, *Tanh*, dan *Softmax* [2].

*Rectified Linear Unit* adalah fungsi aktivasi yang paling umum digunakan dalam aplikasi jaringan saraf tiruan. Fungsi aktivasi *ReLU* membatasi nilai masukannya dimana nilai yang kurang dari nol akan diubah menjadi nol [10, 16]. Persamaan fungsi aktivasi *Rectified Linear Unit* dapat dilihat seperti pada persamaan dibawah.

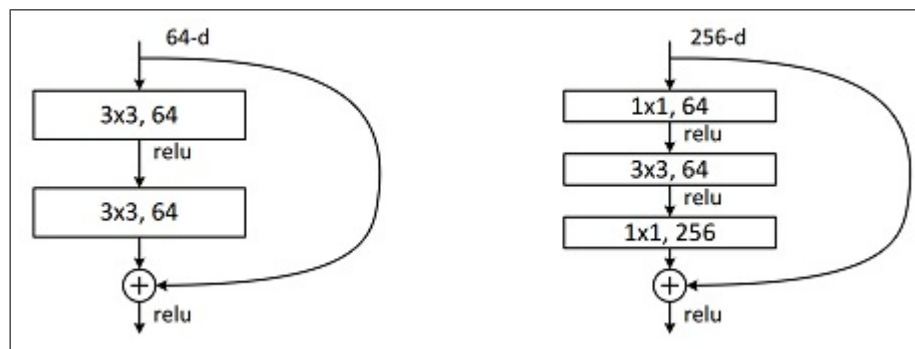
$$f(x) = \max(0, x) = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ 0, & \text{if } x_i < 0 \end{cases}$$



**Gambar 2.5: Rectified Linear Unit**

## 2.5 Residual Network

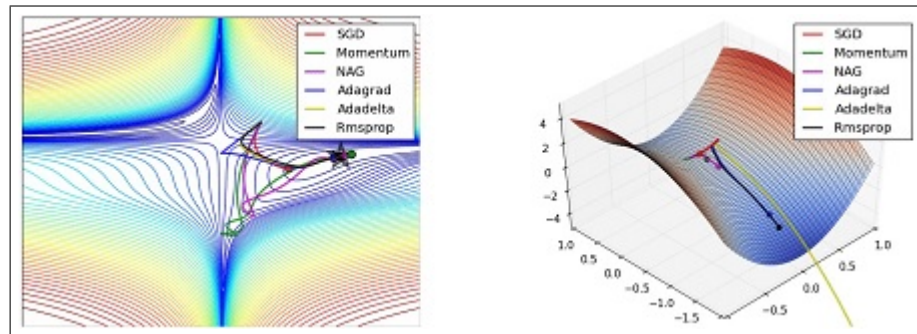
*Residual Network* atau *ResNet* merupakan model jaringan saraf tiruan yang menggunakan *residual block* sebagai dasar dari setiap lapisan. Sebuah *residual block* merupakan sebuah arsitektur jaringan saraf tiruan kecil yang terdiri dari beberapa lapisan. Setiap blok akan menjumlahkan masukan dan keluarannya sehingga layer di dalam suatu blok hanya menambahkan pola-pola yang dipelajari. Hal ini memungkinkan *ResNet* untuk memiliki jumlah blok yang sangat banyak sehingga dapat memetakan suatu fungsi sembarang yang sulit sesuai dengan teorema penaksiran universal [8]. Jenis-jenis *residual block* dapat dilihat pada gambar 2.6.



**Gambar 2.6: Residual Block**

## 2.6 Optimisasi Model

Algoritma optimisasi model yang paling populer dalam melakukan pelatihan model *deep learning* adalah *gradient descent*. *Gradient descent* meminimalisir selisih antara prediksi model dan target sebenarnya dengan merubah bobot-bobot yang terdapat dalam model. Nilai bobot yang ditambahkan berbanding terbalik dengan gradien hasil fungsi kesalahan terhadap masing-masing bobot. Proses optimisasi dilakukan dengan melakukan *backpropagation* yang melibatkan beberapa elemen seperti *learning rate*, dan *loss function*. Tujuannya adalah untuk mencapai titik optimal pada sebuah bidang berdasarkan hasil dari *loss function* [7]. Ilustrasi titik optimal digambarkan pada gambar 2.7.



**Gambar 2.7: Gradient Descent**

### 2.6.1 Backpropagation

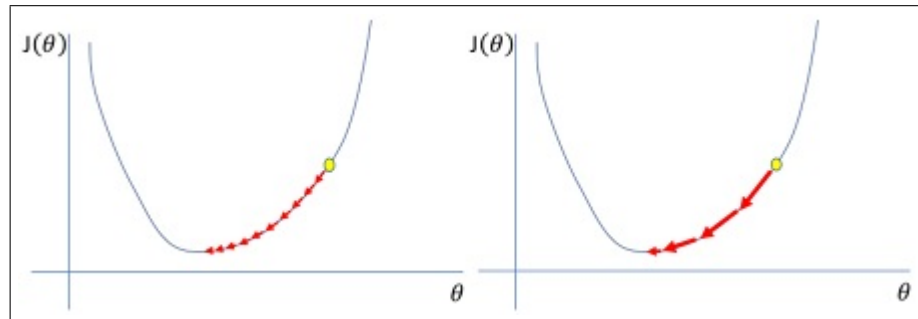
*Backpropagation* adalah sebuah prosedur pelatihan jaringan saraf tiruan yang secara berulang-ulang kali menyesuaikan bobot setiap sel hingga selisih antara keluaran dan target menjadi lebih kecil. Algoritma ini merupakan mengorganisasikan sebuah model jaringan untuk secara mandiri mencari titik optimal secara berkala. Optimisasi dilakukan dengan melakukan *forward-pass* pada satu atau sebagian atau semua data yang ada kedalam model untuk mendapatkan hasil prediksi. Hasil tersebut kemudian diukur selisihnya dengan target yang sebenarnya menggunakan sebuah fungsi kesalahan. Turunan parsial masing-masing bobot terhadap selisih kesalahan ini adalah kuantitas negatif yang harus ditambahkan bobot yang berkaitan [19]. Skema algoritma *backpropagation* dapat dilihat pada gambar 2.9.



**Gambar 2.8: Skema Backpropagation**

### 2.6.2 Learning Rate

*Learning rate* adalah sebuah nilai skalar yang menentukan seberapa besar sebuah bobot pada jaringan saraf akan ditambahkan. Nilai *learning rate* umumnya bernilai kecil karena besar turunan parsial pada setiap iterasi akan berubah-ubah. Nilai *learning rate* yang besar akan merubah dalam skala yang besar sehingga akurasi model cenderung tidak stabil. Nilai *learning rate* yang kecil akan menghasilkan model yang stabil tetapi memerlukan jumlah iterasi yang banyak seperti pada gambar 2.9 [22].



**Gambar 2.9: Ilustrasi Perbedaan *Learning Rate***

### 2.6.3 Mean Squared Error

*Mean squared error* adalah sebuah persamaan estimasi kesalahan dengan merata-ratakan hasil dari pangkat dua selisih antara dua vektor. Fungsi kesalahan ini dipakai untuk mengukur kesalahan model regresi dimana keluaran yang diukur bersifat kontinu. Persamaan ini akan selalu bernilai positif dan nilai yang mendekati nol menandakan bahwa selisih antara dua vektor semakin kecil seperti pada persamaan dibawah [21].

$$MSE = \frac{1}{n} \sum_{i=1}^n \left( \frac{d_i - f_i}{\sigma_i} \right)^2$$

## 2.7 Estimasi Pose Dua Dimensi

Estimasi pose dua dimensi melakukan lokalisasi titik kunci anatomi atau bagian tubuh manusia berdasarkan fungsi dari bagian tersebut pada sebuah gambar dua dimensi. Setiap gambar unik dapat berisi jumlah orang yang berbeda dan dapat



muncul dengan posisi dan ukuran yang berbeda-beda. Interaksi antara manusia dengan benda disekitarnya dapat menyebabkan oklusi dan hilangnya titik kunci anatomi yang dicari. Kompleksitas permasalahan ini juga semakin meningkat berbanding lurus dengan jumlah orang yang ada dalam suatu gambar [5] [6].

Pendekatan yang umum dilakukan adalah dengan mendeteksi setiap orang dalam setiap gambar secara individu. Hal pertama yang dideteksi adalah lokalisasi kemunculan suatu benda yang dianggap sebagai objek manusia dan kemudian mencari titik kunci anatomi dari objek tersebut. Pendekatan *top-down* seperti ini cocok dipakai untuk mendapatkan *single-person* dengan mudah tetapi kompleksitas yang proporsional dengan jumlah orang. Sebaliknya dengan pendekatan *bottom-up* dimana deteksi dilakukan dengan mencari alamat piksel pada gambar yang dianggap adalah titik kunci anatomi dan dilanjutkan dengan membangun pose manusia dari kandidat yang ada. Pendekatan *bottom-up* cocok untuk menyelesaikan masalah *multi-person* dikarenakan pencarian kandidat titik kunci dapat dilakukan terlebih dahulu tanpa harus mengetahui jumlah orang. Perbedaan pendekatan *top-down* dan *bottom-up* dapat dilihat pada gambar 2.10 [3].



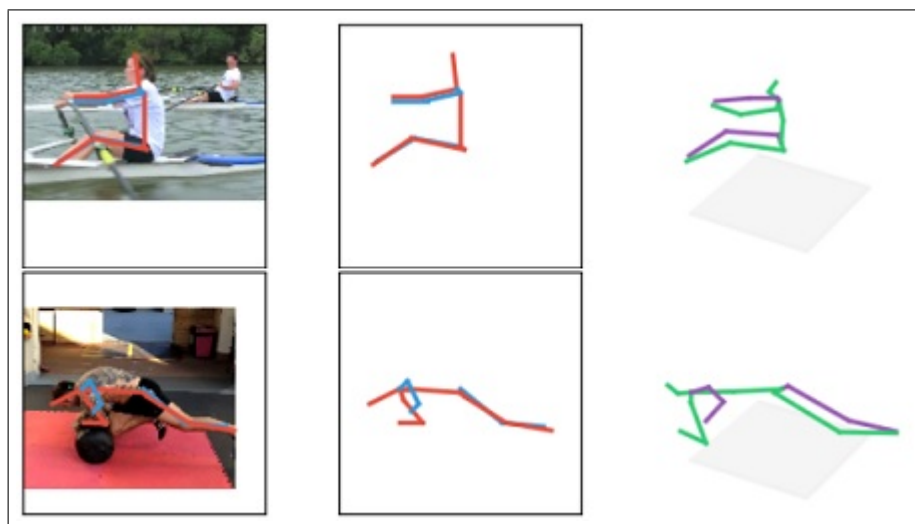
**Gambar 2.10: Pendekatan *Top-Down* dan *Bottom-Up***



## 2.8 Estimasi Pose Tiga Dimensi

Algoritma estimasi pose tiga dimensi merekonstruksi titik kunci tiga dimensi tubuh manusia dari sebuah gambar dua dimensi. Proses pencarian titik kunci dibagi menjadi dua golongan yaitu pencarian secara langsung menggunakan gambar sebagai masukan dan pencarian bertahap dengan mencari titik kunci anatomi dua dimensi terlebih dahulu kemudian dilanjutkan dengan pencarian titik kunci tiga dimensi. Titik kunci yang didapatkan dapat berada dalam koordinat lokal dan global. Pencarian secara langsung tidak menghasilkan akurasi yang tinggi dan cenderung buruk dalam mendeteksi orientasi yang benar [4].

Titik kunci tiga dimensi yang berada dalam koordinat lokal menjadikan posisi pinggang sebagai titik nol yang berarti lokasi pinggang akan selalu bernilai  $\vec{hip} = [0.0, 0.0, 0.0]$ . Posisi titik kunci lain relatif terhadap titik pinggang. Hal ini menyebabkan pose yang ditangkap hanya bersifat lokal yang berarti hanya memiliki satu orientasi lokal. Pencarian pose tiga dimensi lokal dari gambar dua dimensi diilustrasikan pada gambar 2.11 [14].



**Gambar 2.11: Pencarian Pose Tiga Dimensi Lokal**

Titik kunci tiga dimensi dapat juga berada dalam koordinat global yang berarti orientasi kamera saat mengambil gambar diperhitungkan sebagai sebuah observasi virtual. Kompleksitas yang dihadapi berkaitan dengan orientasi kamera yang berubah. Estimasi pose tiga dimensi global umumnya dilakukan dengan

mendeteksi titik kunci anatomi tubuh manusia kedalam suatu dunia virtual tiga dimensi sehingga dapat diamati dari berbagai sudut yang berkorelasi dengan kamera [24, 15]. Pencarian pose tiga dimensi global dari gambar dua dimensi diilustrasikan pada gambar 2.12.



**Gambar 2.12: Pencarian Pose Tiga Dimensi Global**

## 2.9 PyTorch

PyTorch merupakan *framework* untuk melakukan *deep learning* dengan bahasa pemrograman python. Penggiat data sebagai dasar aplikasi umumnya menggunakan bahasa pemrograman ini dalam melakukan riset. PyTorch menyediakan implementasi grafik jaringan saraf tiruan yang bersifat dinamis. Hal ini memungkinkan pengguna untuk mengubah arsitektur model secara cepat dalam setiap iterasi [17].

## **BAB III**

### **PENDEKATAN**

#### **3.1 Langkah-Langkah yang dilakukan**

**Tabel 3.1: Sebuah tabel**

1	2	3
4	5	6
7	8	9

Sumber: Bego Lu

diagram

#### **3.2 Alur Pelatihan Model**

diagram

#### **3.3 Pra Pemrosesan Data**

download -> data loader

#### **3.4 Arsitektur**

2 kali resnet

#### **3.5 Pelatihan Model**

brp epoch dll

## **BAB IV**

### **HASIL DAN ANALISIS**

#### **4.1 Plot Grafik**

#### **4.2 Matplotlib**

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berisi ringkasan dari metodologi dan kesimpulan penting dari hasil evaluasi.

#### **5.2 Saran**

Berisi saran-saran untuk pengembangan riset ini ke langkah ke depan.

## DAFTAR PUSTAKA

- [1] Abiodun, O., Jantan, A., Omolara, O., Dada, K., Umar, A., Linus, O., Arshad, H., Aminu Kazaure, A., Gana, U., dan Kiru, M. (2019). Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, PP:1–1.
- [2] Agostinelli, F., Hoffman, M., Sadowski, P., dan Baldi, P. (2014). Learning Activation Functions to Improve Deep Neural Networks. *arXiv e-prints*, page arXiv:1412.6830.
- [3] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., dan Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [4] Chen, T., Fang, C., Shen, X., Zhu, Y., Chen, Z., dan Luo, J. (2020). Anatomy-aware 3D Human Pose Estimation in Videos. *arXiv e-prints*, page arXiv:2002.10322.
- [5] Felzenszwalb, P. dan Huttenlocher, D. (2005). Pictorial structures for object recognition. *International Journal of Computer Vision*, 61:55–79.
- [6] Gkioxari, G., Hariharan, B., Girshick, R., dan Malik, J. (2014). Using k-poselets for detecting people and localizing their keypoints. pages 3582–3589.
- [7] Guliyev, N. J. dan Ismailov, V. E. (2015). A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function. *arXiv e-prints*, page arXiv:1601.00013.
- [8] He, K., Zhang, X., Ren, S., dan Sun, J. (2015). Deep Residual Learning for Image Recognition. *arXiv e-prints*, page arXiv:1512.03385.
- [9] Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3:31.

- [10] Hinton, G. E. Rectified linear units improve restricted boltzmann machines  
vinod nair.
- [11] Ionescu, C., Papava, D., Olaru, V., dan Sminchisescu, C. (2014). Human3.6m:  
Large scale datasets and predictive methods for 3d human sensing in natural  
environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,  
36(7):1325–1339.
- [12] Kingma, D. P. dan Ba, J. (2014). Adam: A Method for Stochastic  
Optimization. *arXiv e-prints*, page arXiv:1412.6980.
- [13] Kratsios, A. (2019). The Universal Approximation Property:  
Characterizations, Existence, and a Canonical Topology for Deep-Learning.  
*arXiv e-prints*, page arXiv:1910.03344.
- [14] Martinez, J., Hossain, R., Romero, J., dan Little, J. J. (2017). A simple yet  
effective baseline for 3d human pose estimation. In *ICCV*.
- [15] Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P., Seidel,  
H.-P., Rhodin, H., Pons-Moll, G., dan Theobalt, C. (2019). XNect: Real-time  
Multi-Person 3D Motion Capture with a Single RGB Camera. *arXiv e-prints*,  
page arXiv:1907.00837.
- [16] Nwankpa, C., Ijomah, W., Gachagan, A., dan Marshall, S. (2018). Activation  
Functions: Comparison of trends in Practice and Research for Deep Learning.  
*arXiv e-prints*, page arXiv:1811.03378.
- [17] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen,  
T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E.,  
DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai,  
J., dan Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance  
Deep Learning Library. *arXiv e-prints*, page arXiv:1912.01703.
- [18] Ruder, S. (2016). An overview of gradient descent optimization algorithms.  
*arXiv e-prints*, page arXiv:1609.04747.

- [19] Rumelhart, D. E., Hinton, G. E., dan Williams, R. J. (1986). Learning Representations by Back-propagating Errors. *Nature*, 323(6088):533–536.
- [20] Sharma, V., Rai, S., dan Dev, A. (2012). A comprehensive study of artificial neural networks.
- [21] Torabi, M. dan Rao, J. (2013). Estimation of mean squared error of model-based estimators of small area means under a nested error linear regression model. *Journal of Multivariate Analysis*, 117:76 – 87.
- [22] You, K., Long, M., Wang, J., dan Jordan, M. I. (2019). How Does Learning Rate Decay Help Modern Neural Networks? *arXiv e-prints*, page arXiv:1908.01878.
- [23] Zhang, J. (2019). Basic Neural Units of the Brain: Neurons, Synapses and Action Potential. *arXiv e-prints*, page arXiv:1906.01703.
- [24] Zhou, X., Huang, Q., Sun, X., Xue, X., dan Wei, Y. (2017). Towards 3D Human Pose Estimation in the Wild: a Weakly-supervised Approach. *arXiv e-prints*, page arXiv:1704.02447.



## **LAMPIRAN**

Bisa diketik sesuai kebutuhan