



UNIVERSITAS
GADJAH MADA

Kampus
Merdeka
INDONESIA JAYA

Bisa.ai

Multi-task AI on Google Colab: Simultaneous Image Segmentation and Depth Estimation

Oskar Natan, S.ST., M.Tr.T.



UNIVERSITAS GADJAH MADA

Outline

- **Introduction**
 - Semantic segmentation (SS)
 - Depth estimation (DE)
- **Multi-task Learning**
 - Multi-task architecture
 - Proses training dan evaluasi model
- **Cloud Computing**
 - Intro to cloud services
- **Demonstrasi program**
 - Quick setup google colab
 - Implementasi python script
 - Train-val-test
- **Github repo**
 - <https://github.com/oskarnatan/mtl-bisaai>

LOCALLY ROOTED, GLOBALLY RESPECTED



UNIVERSITAS
GADJAH MADA

Introduction

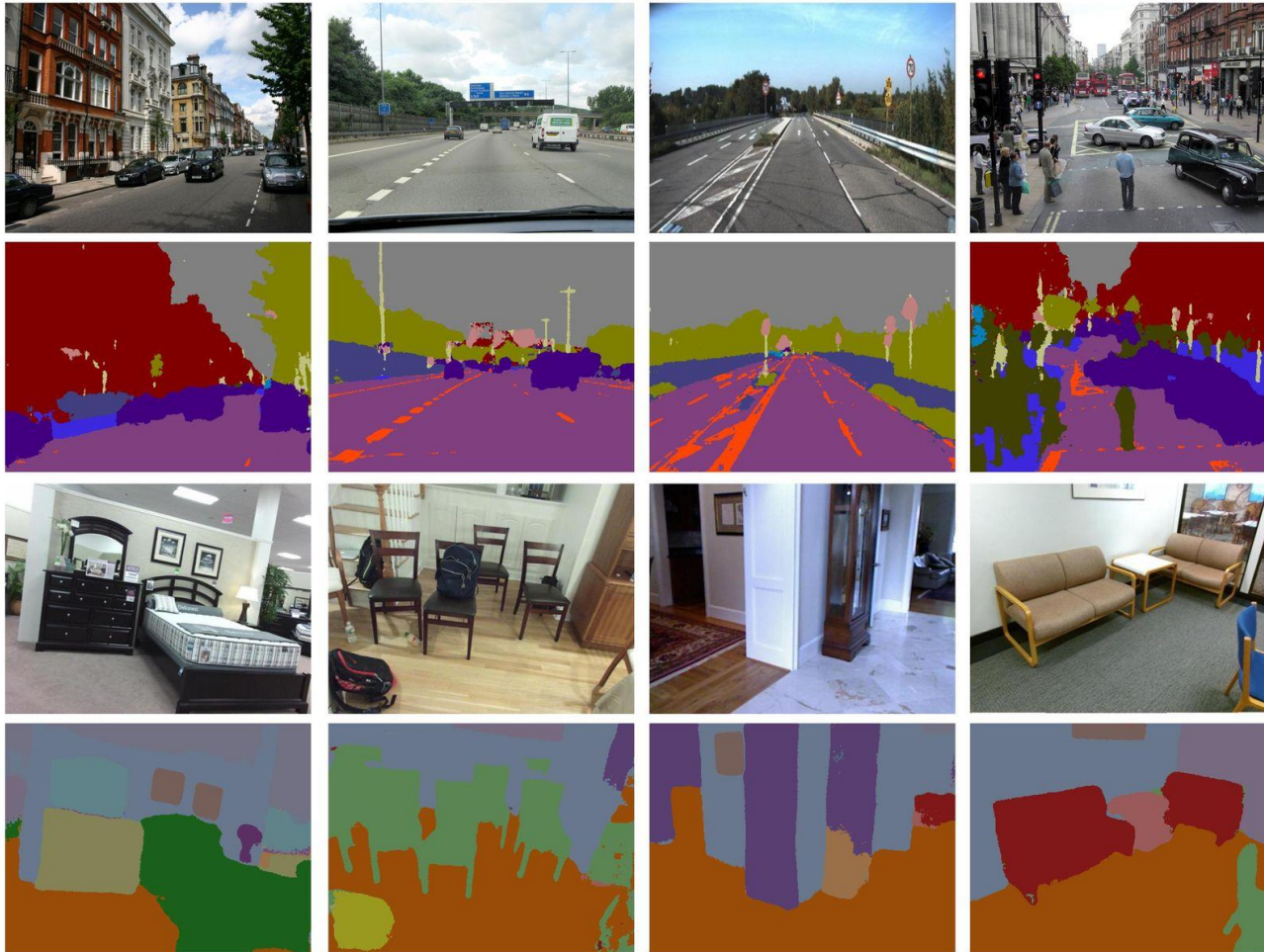
LOCALLY ROOTED, GLOBALLY RESPECTED

ugm.ac.id

Apa itu Image Segmentation?



UNIVERSITAS GADJAH MADA



**Sumber gambar: Lambda Labs*

- Dalam image processing dan computer vision, image segmentation (segmentasi citra) adalah proses mempartisi citra digital menjadi beberapa segmen citra (set piksel).
 - Tujuan dari segmentasi adalah untuk menyederhanakan dan/atau mengubah representasi suatu citra menjadi sesuatu yang lebih mudah untuk dianalisis.
 - Singkatnya, segmentasi citra adalah proses pemberian label (proses klasifikasi) ke setiap piksel dalam gambar.
 - Pixel-wise classification

Pixel-wise Classification



UNIVERSITAS GADJAH MADA



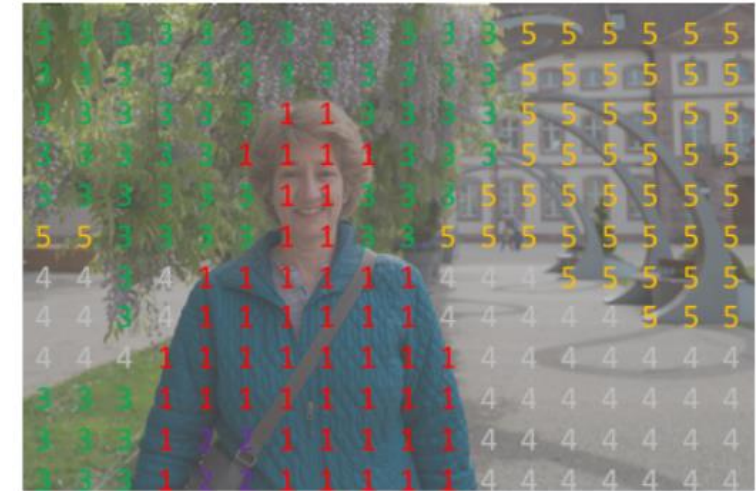
Input



1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5
3	3	3	3	3	3	1	1	3	3	3	3	5	5	5	5	5	5	5
3	3	3	3	3	1	1	1	1	3	3	3	5	5	5	5	5	5	5
3	3	3	3	3	1	1	3	3	3	5	5	5	5	5	5	5	5	5
5	5	3	3	3	3	1	1	3	3	5	5	5	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	5	5	5	5	5	5
4	4	3	4	1	1	1	1	1	1	4	4	4	4	4	5	5	5	5
4	4	4	1	1	1	1	1	1	1	4	4	4	4	4	4	4	4	4
3	3	3	1	1	1	1	1	1	1	4	4	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	4	4	4	4	4	4	4	4	4
3	3	3	1	2	2	1	1	1	1	4	4	4	4	4	4	4	4	4

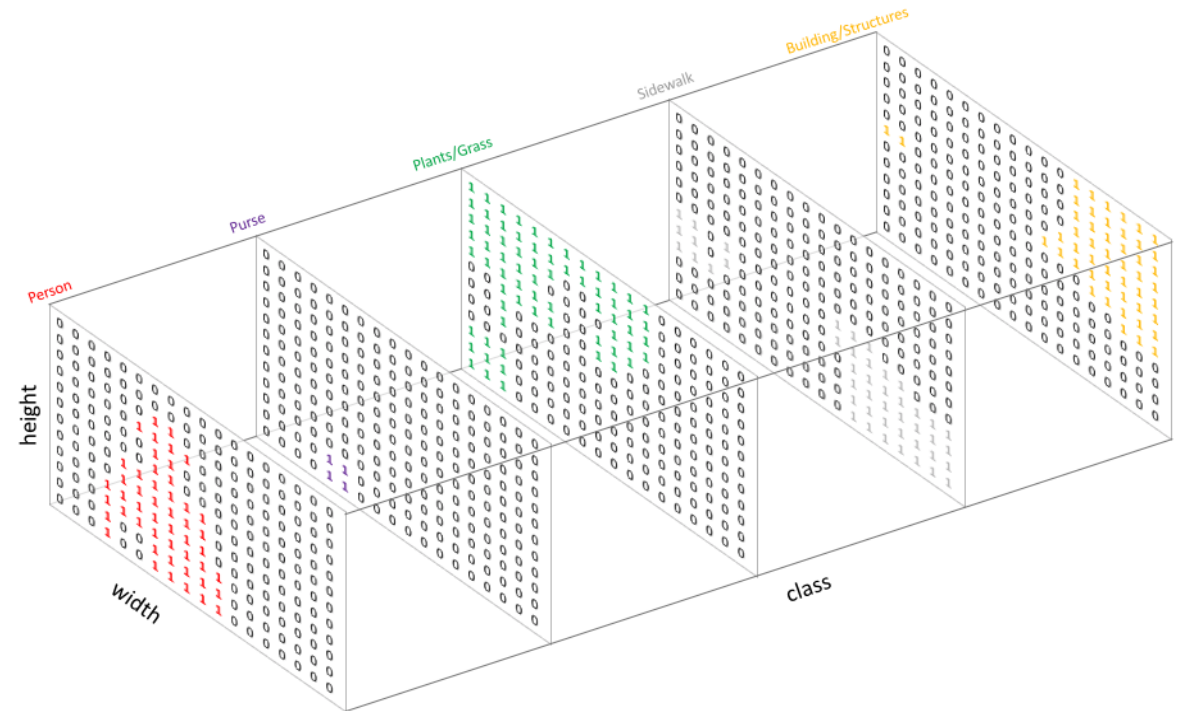
Semantic Labels



- Model DL memproses input gambar RGB ($3 \times H \times W$) dan menghasilkan output matrix ($1 \times H \times W$), dimana setiap elemen matrix berupa class/label tertentu dari sebuah obyek

**Sumber gambar: jeremyjordan.me*

One-hot encoding



Categorical		One hot encoded				
Nama Obyek	Class	Orang	Tas	Tumbuhan	Trotoar	Bangunan
Orang	1	1	0	0	0	0
Tas	2	0	1	0	0	0
Tumbuhan	3	0	0	1	0	0
Trotoar	4	0	0	0	1	0
Bangunan	5	0	0	0	0	1

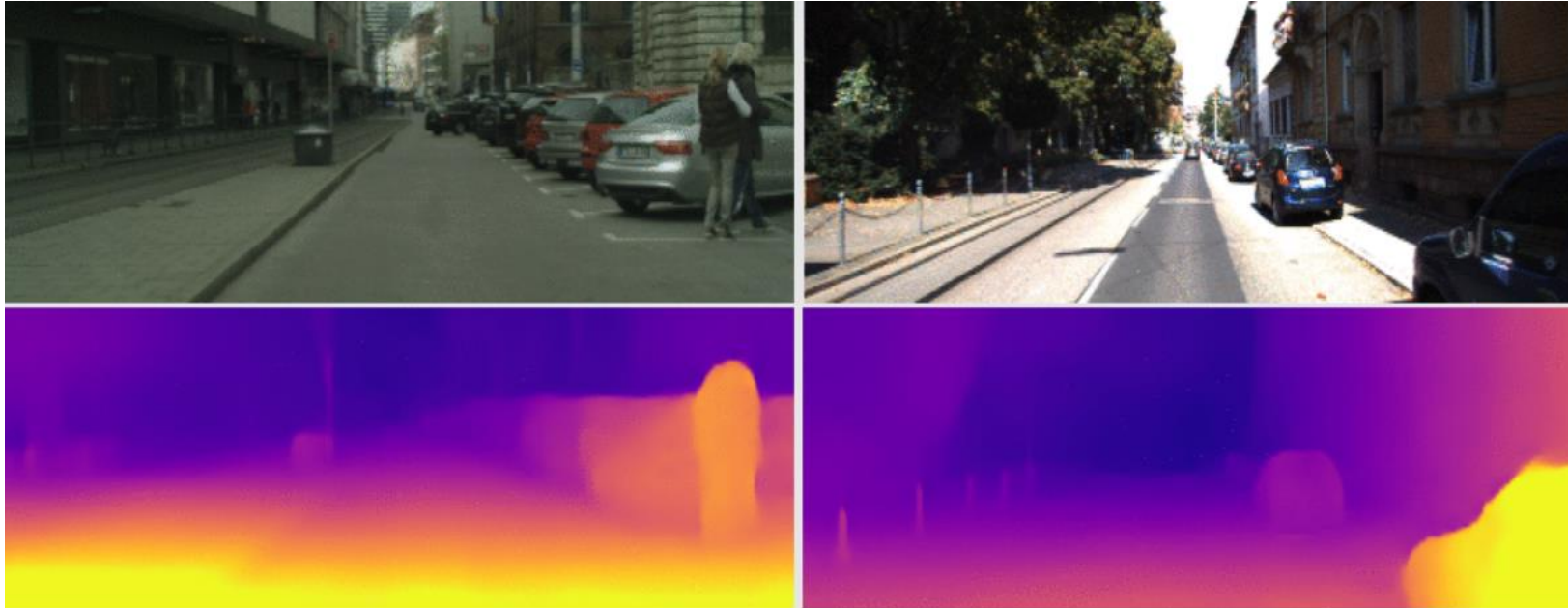
- Model tidak memprediksi matrix dengan dimensi $(1 \times H \times W)$, melainkan memprediksi matrix dengan dimensi $(N_c \times H \times W)$. N_c adalah jumlah label class.

**Sumber gambar: jeremyjordan.me*

Apa itu Depth Estimation?



UNIVERSITAS GADJAH MADA



- Depth estimation adalah kemampuan visual untuk melihat dunia dalam koordinat tiga dimensi (3D).
 - $H \times W \times D$, D adalah depth yang merepresentasikan jarak suatu obyek, atau secara spesifik informasi jarak dari setiap piksel pada gambar.

**Sumber gambar: Google AI blog*

Estimasi dengan Stereo Camera



UNIVERSITAS GADJAH MADA



left image



right image



depth map

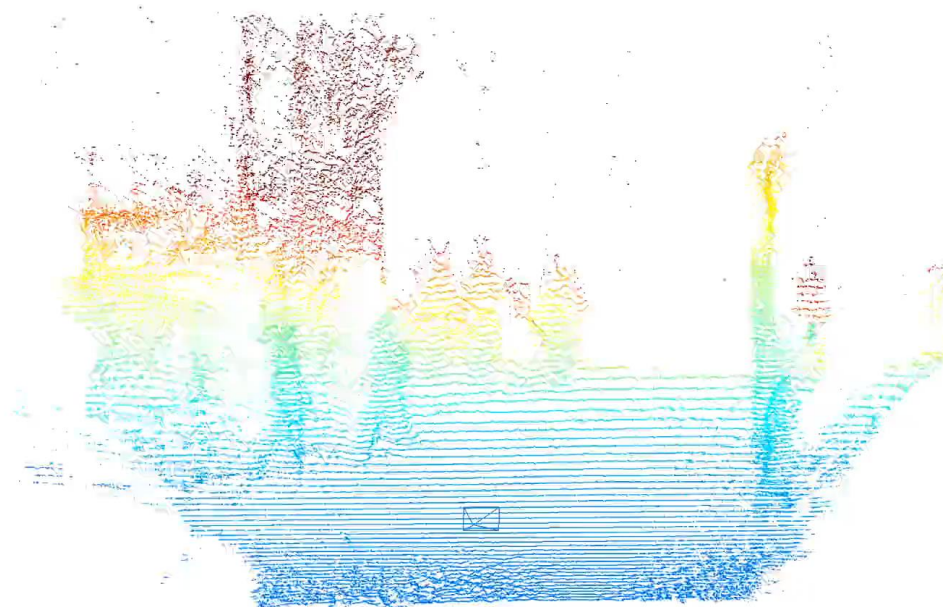
- 2 camera dengan konfigurasi tertentu dapat digunakan untuk mengestimasi depth pada setiap piksel gambar
 - Kelemahan: (1) harus tahu berbagai parameter camera (2) jika peletakan berubah (vibrasi, perubahan posisi, dll) maka depth tidak dapat diestimasi dengan baik

**Sumber gambar: CMU Lecture*

Aplikasi Depth Sensing



UNIVERSITAS GADJAH MADA



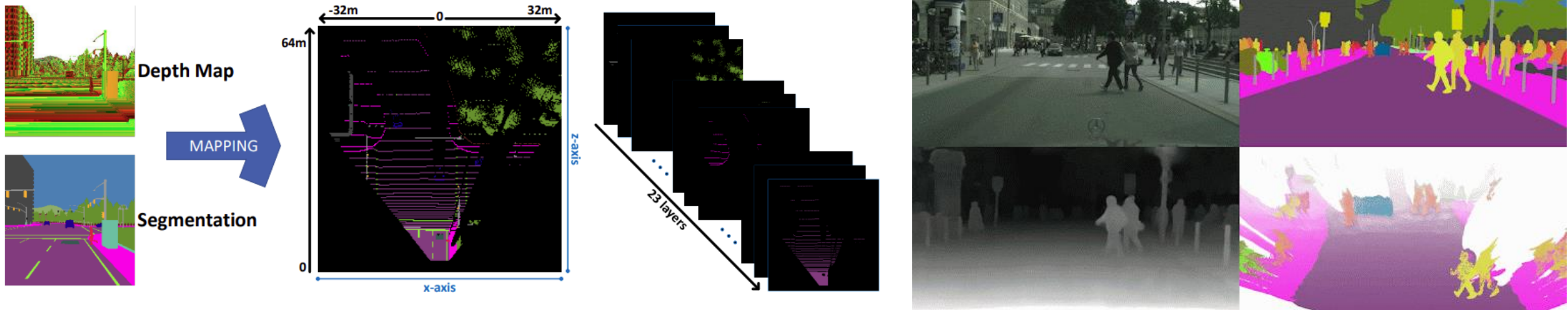
- Depth to cloud
 - Sebagai pengganti LiDAR dengan jumlah point cloud yang lebih banyak dan lebih padat (resolusi lebih tinggi)

**Sumber video: Stereolabs*

Semantic Depth Cloud Mapping



UNIVERSITAS GADJAH MADA



- Dalam autonomous driving, proyeksi antara segmentation map dan depth map memungkinkan model AI untuk melakukan persepsi dalam perspektif BEV (bird's eye view)
 - Mengetahui free dan occupied area lebih baik
 - Mengetahui drivable region lebih baik
 - Representasi ini menyimpan informasi yang lebih baik sebagai acuan untuk controller

*O. Natan and J. Miura, "End-to-end Autonomous Driving with Semantic Depth Cloud Mapping and Multi-agent," IEEE Trans. Intelligent Vehicles, 2022.

*Google Ai blog

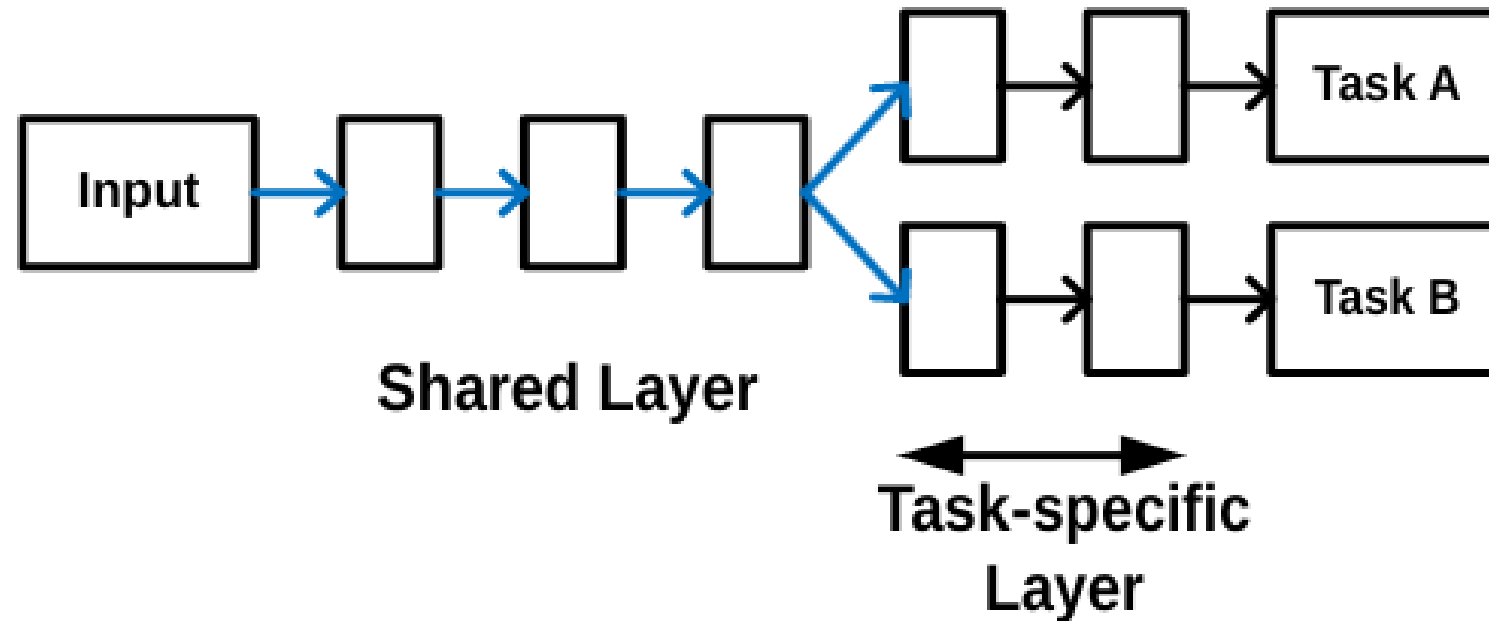


UNIVERSITAS
GADJAH MADA

Multi-task Learning

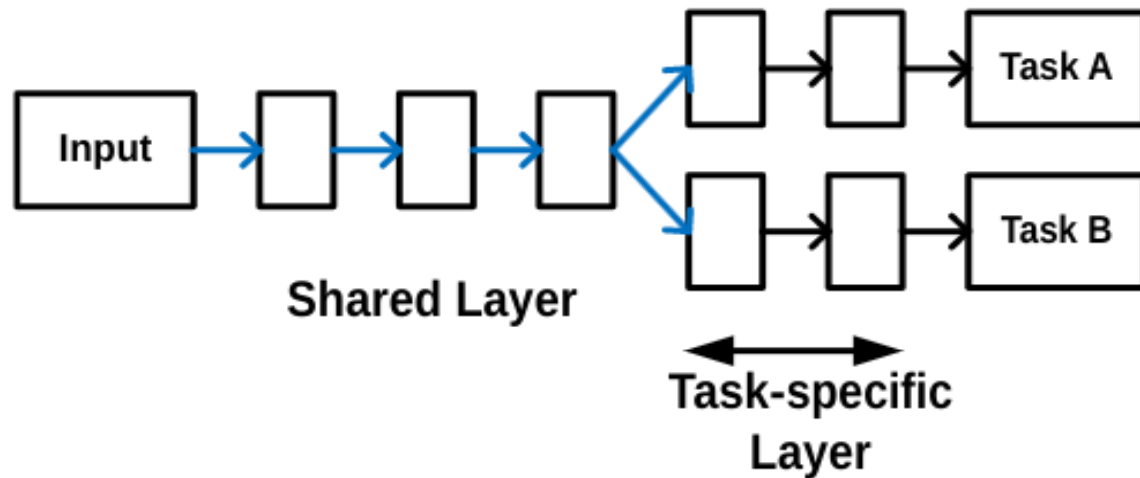
LOCALLY ROOTED, GLOBALLY RESPECTED

Apa itu Multi-task Learning (MTL)?

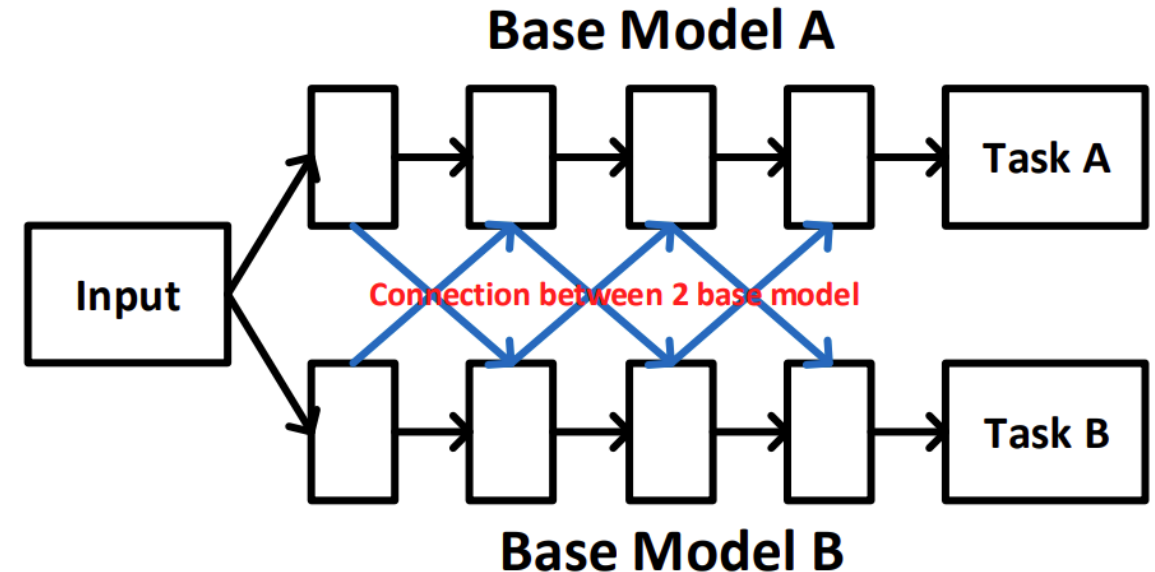


- MTL adalah paradigma learning dimana beberapa task/tugas diselesaikan secara **simultan**.
 - Hal ini dapat menghasilkan peningkatan **efisiensi learning dan performa model** jika dibandingkan dengan melatih model secara terpisah.
- Tujuan MTL adalah untuk **meningkatkan generalisasi** dengan menggunakan mekanisme **feature sharing**. (feature yang diextract oleh backbone yang sama)
 - Apa yang dipelajari untuk menyelesaikan task A, dapat membantu untuk menyelesaikan task B

Jenis-jenis MTL



Hard sharing: Setiap task-specific layer memproses feature yang sama yang diextract oleh backbone yang sama.



Soft sharing: Mekanisme feature sharing terjadi pada seluruh layer di setiap backbone.

Single Task Architecture



UNIVERSITAS GADJAH MADA

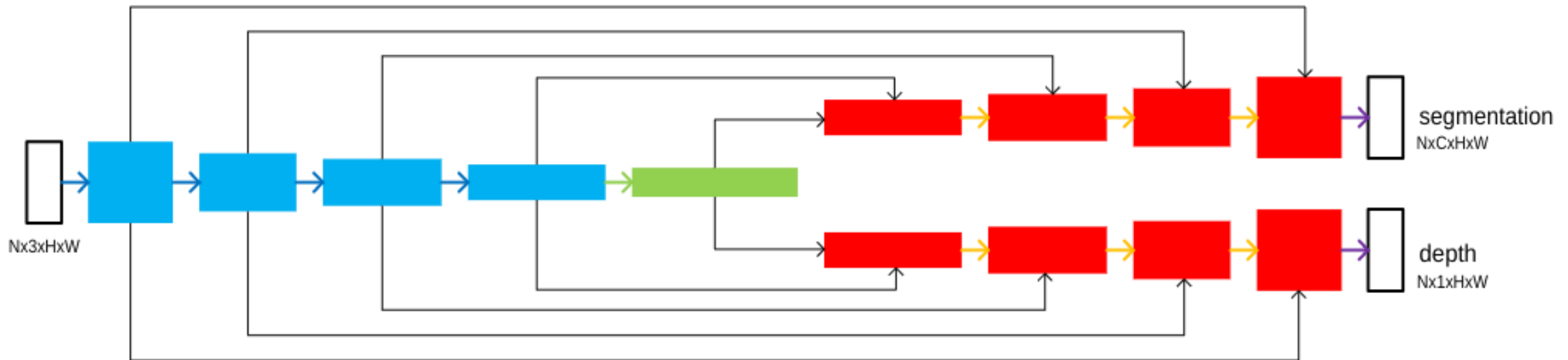


- Standard UNet untuk single task model SemSeg **atau** DepEst.
 - Pada semantic segmentation, output channel C = jumlah class label yang hendak diprediksi. Sementara pada depth estimation, output channel $C = 1$ (nilai depth setiap pixel)

Multi-Task Architecture



UNIVERSITAS GADJAH MADA

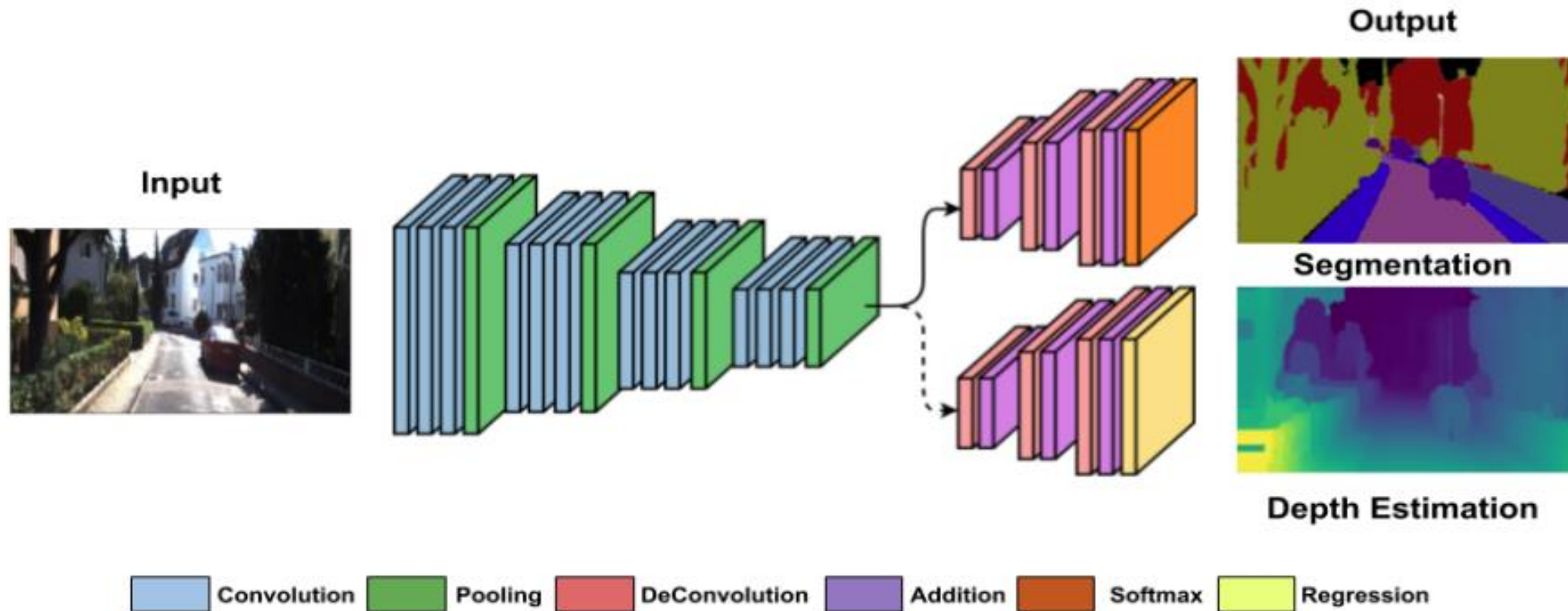


- Multi-task model: SemSeg **dan** DepEst.
 - Pada semantic segmentation, output channel C = jumlah class label yang hendak diprediksi. Sementara pada depth estimation, output channel $C = 1$ (nilai depth setiap pixel)

Contoh MTL Architecture



UNIVERSITAS GADJAH MADA



- Input: RGB, Output: Segmentation dan Depth

**Sumber gambar: arxiv 1901.05808*



UNIVERSITAS
GADJAH MADA

Convolutional Neural Networks

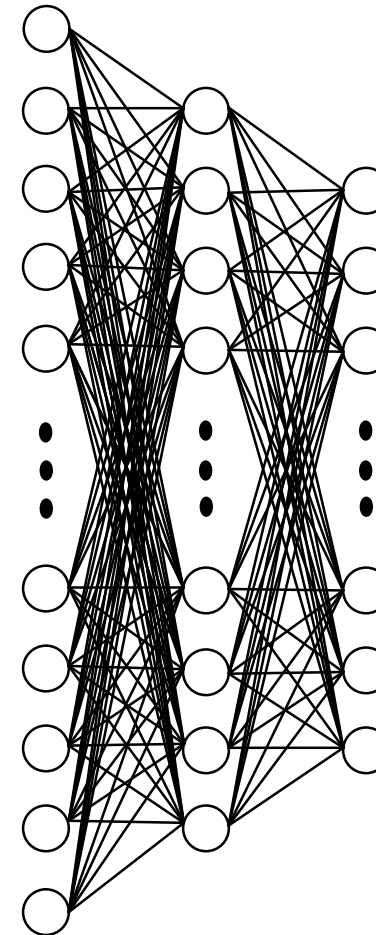
LOCALLY ROOTED, GLOBALLY RESPECTED

CNN (ConvNet)

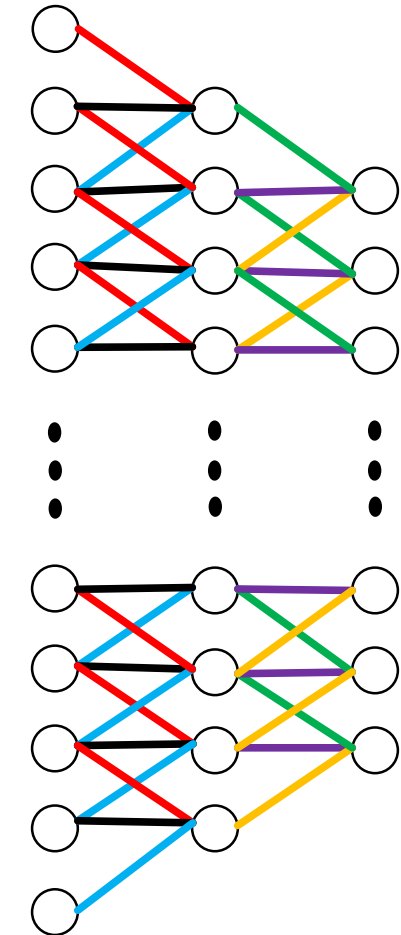


UNIVERSITAS GADJAH MADA

- CNN adalah salah satu jenis deep neural nets yang paling sering digunakan untuk image processing dan computer vision.
- Berbeda dengan MLP yang “fully-connected”, CNN justru hanya “**locally-connected**” dimana setiap neuron dalam layer-nya hanya terkoneksi dengan beberapa neuron pada layer lain.



Independent weights



Shared weights

Convolution



UNIVERSITAS GADJAH MADA

CONVOLUTION

KERNEL

PADDING

STRIDE

POOLING

UPSAMPLING

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved
Feature

Input

0	1	2
3	4	5
6	7	8

Kernel

0	1
2	3

*

=

Output

19	25
37	43

Convolution: proses cross correlation antara input dan kernel filter

$$(0 \times 0) + (1 \times 1) + (3 \times 2) + (4 \times 3) = 19$$

$$(1 \times 0) + (2 \times 1) + (4 \times 2) + (5 \times 3) = 25$$

$$(3 \times 0) + (4 \times 1) + (6 \times 2) + (7 \times 3) = 37$$

$$(4 \times 0) + (5 \times 1) + (7 \times 2) + (8 \times 3) = 43$$

*Sumber gambar: Dive into DL Book & towardsdatascience.com

Kernel



UNIVERSITAS GADJAH MADA

CONVOLUTION

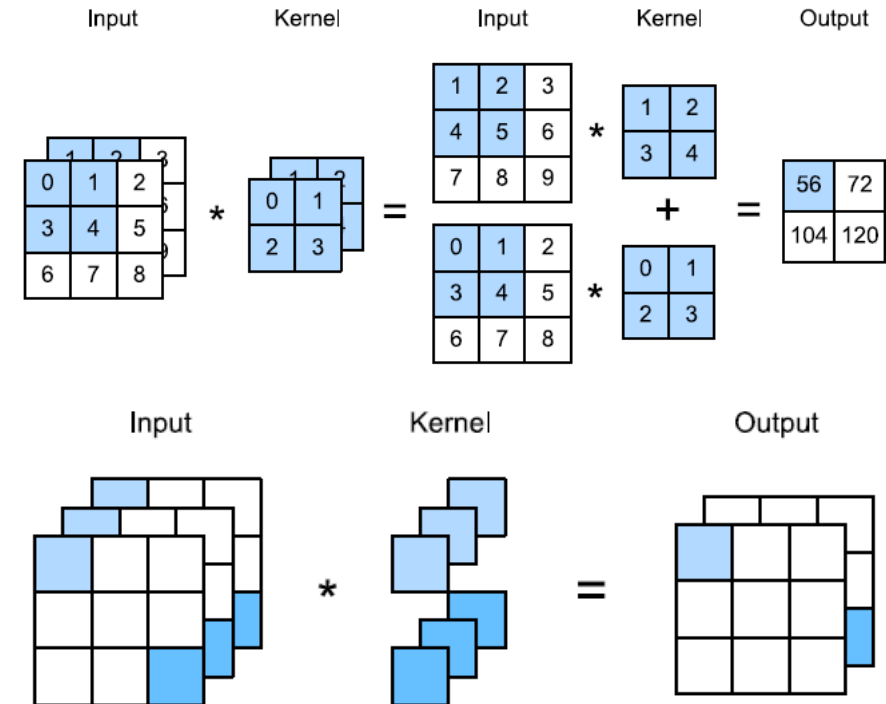
KERNEL

PADDING

STRIDE

POOLING

UPSAMPLING



Kernel: shared weights yang dituning selama proses training

**Sumber gambar: Dive into DL Book*

Padding & Stride



UNIVERSITAS GADJAH MADA

CONVOLUTION

KERNEL

PADDING

STRIDE

POOLING

UPSAMPLING

Input Kernel Output

0	0	0	0	0
0	0	1	2	0
0	3	4	5	0
0	6	7	8	0
0	0	0	0	0

 *

0	1
2	3

 =

0	3	8	4
9	19	25	10
21	37	43	16
6	7	8	0

Padding: menambahkan elemen pada tepian input

Input Kernel Output

0	0	0	0	0
0	0	1	2	0
0	3	4	5	0
0	6	7	8	0
0	0	0	0	0

 *

0	1
2	3

 =

0	8
6	8

Stride: mengatur pergeseran kernel filter

**Sumber gambar: Dive into DL Book*

Pooling & Upsampling



UNIVERSITAS GADJAH MADA

CONVOLUTION

KERNEL

PADDING

STRIDE

POOLING

UPSAMPLING

Input

0	1	2
3	4	5
6	7	8

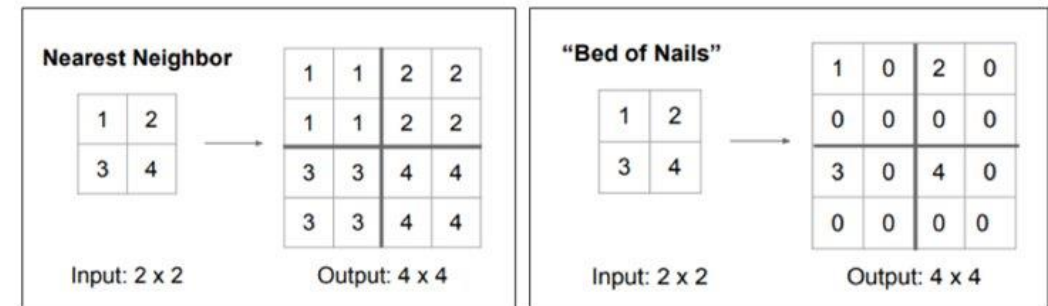
2 x 2 Max
Pooling

Output

4	5
7	8

$\max(0, 1, 3, 4) = 4$
 $\max(1, 2, 4, 5) = 5$
 $\max(3, 4, 6, 7) = 7$
 $\max(4, 5, 7, 8) = 8$

Pooling bekerja dengan mengambil nilai tertinggi (max pooling) atau nilai rata-rata (avgpooling)



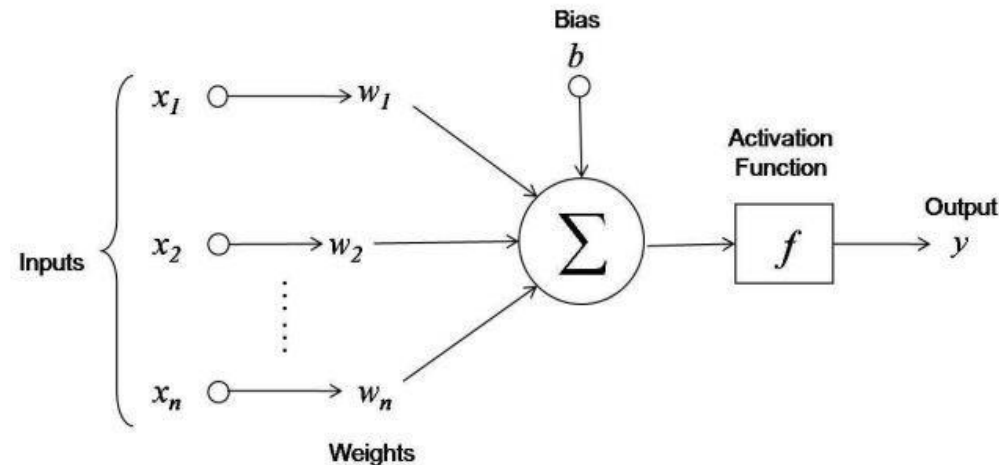
Tujuan upsampling adalah untuk mendistribusikan satu nilai ke resolusi yang lebih tinggi.

**Sumber gambar: Dive into DL Book & jeremyjordan.me*

Activation Function

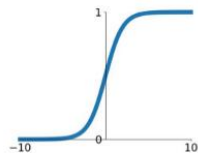


UNIVERSITAS GADJAH MADA



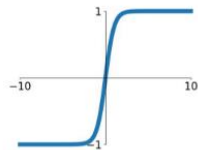
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



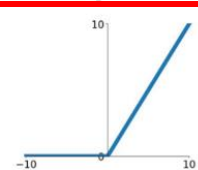
tanh

$$\tanh(x)$$



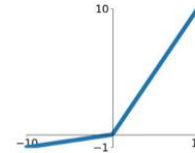
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

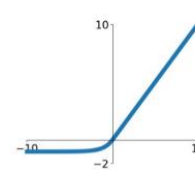


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



- Aktivasi sigmoid digunakan untuk mengaktivasi neuron pada layer output SS
 - Semantic segmentation --> persoalan pixel-wise **binary classification**
- Aktivasi ReLU digunakan untuk mengaktivasi neuron pada layer output DE
 - Depth estimation --> persoalan **regresi** untuk range nilai +F

*Sumber gambar: towardsdatascience.com



UNIVERSITAS
GADJAH MADA

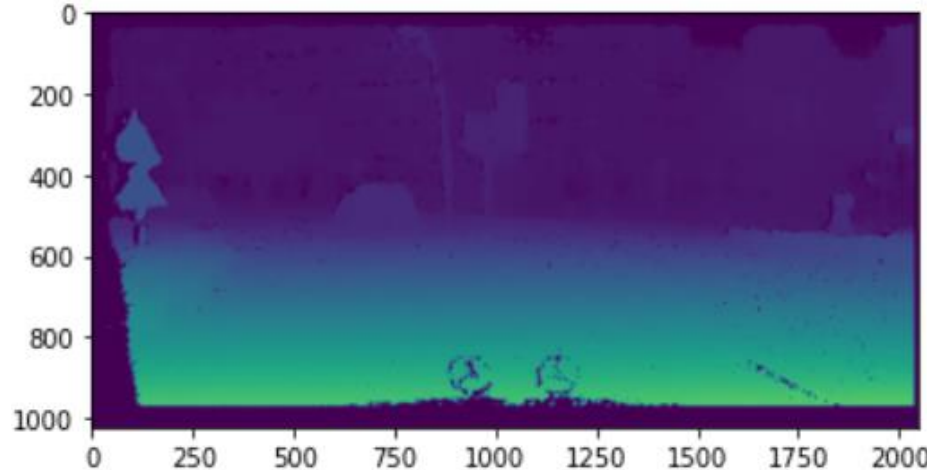
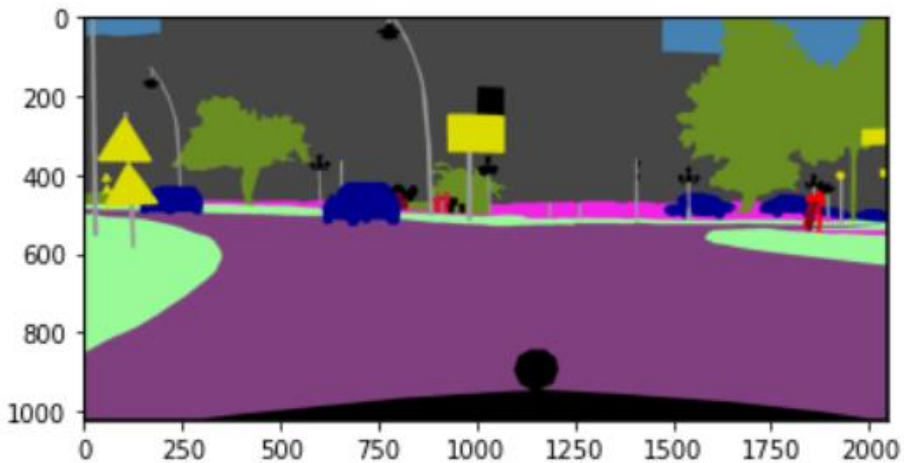
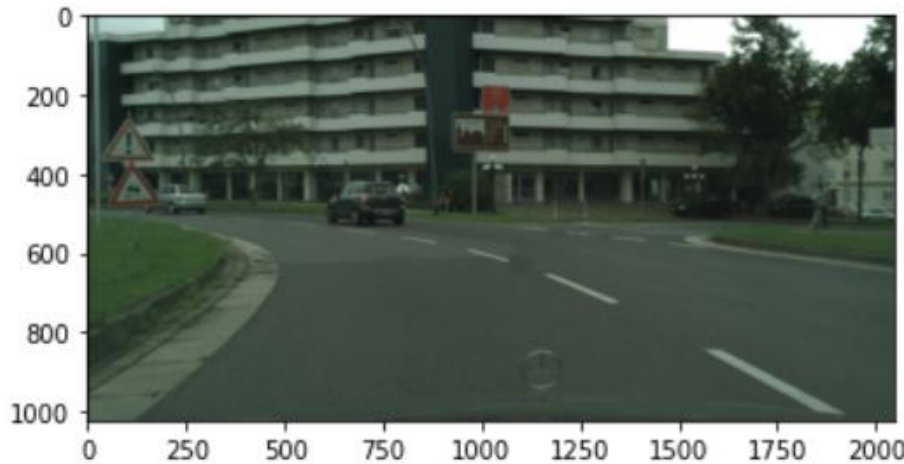
Training

LOCALLY ROOTED, GLOBALLY RESPECTED

Dataset



UNIVERSITAS GADJAH MADA



- Dalam paradigma supervised learning, diperlukan suatu dataset berlabel untuk melatih sebuah model DL
 - SS --> label class
 - DE --> label jarak
- Dataset untuk semantic segmentation dan depth estimation
 - Cityscapes dataset
 - NYUD dataset

**Sumber gambar: cityscapes dataset*

Learning Process



UNIVERSITAS GADJAH MADA

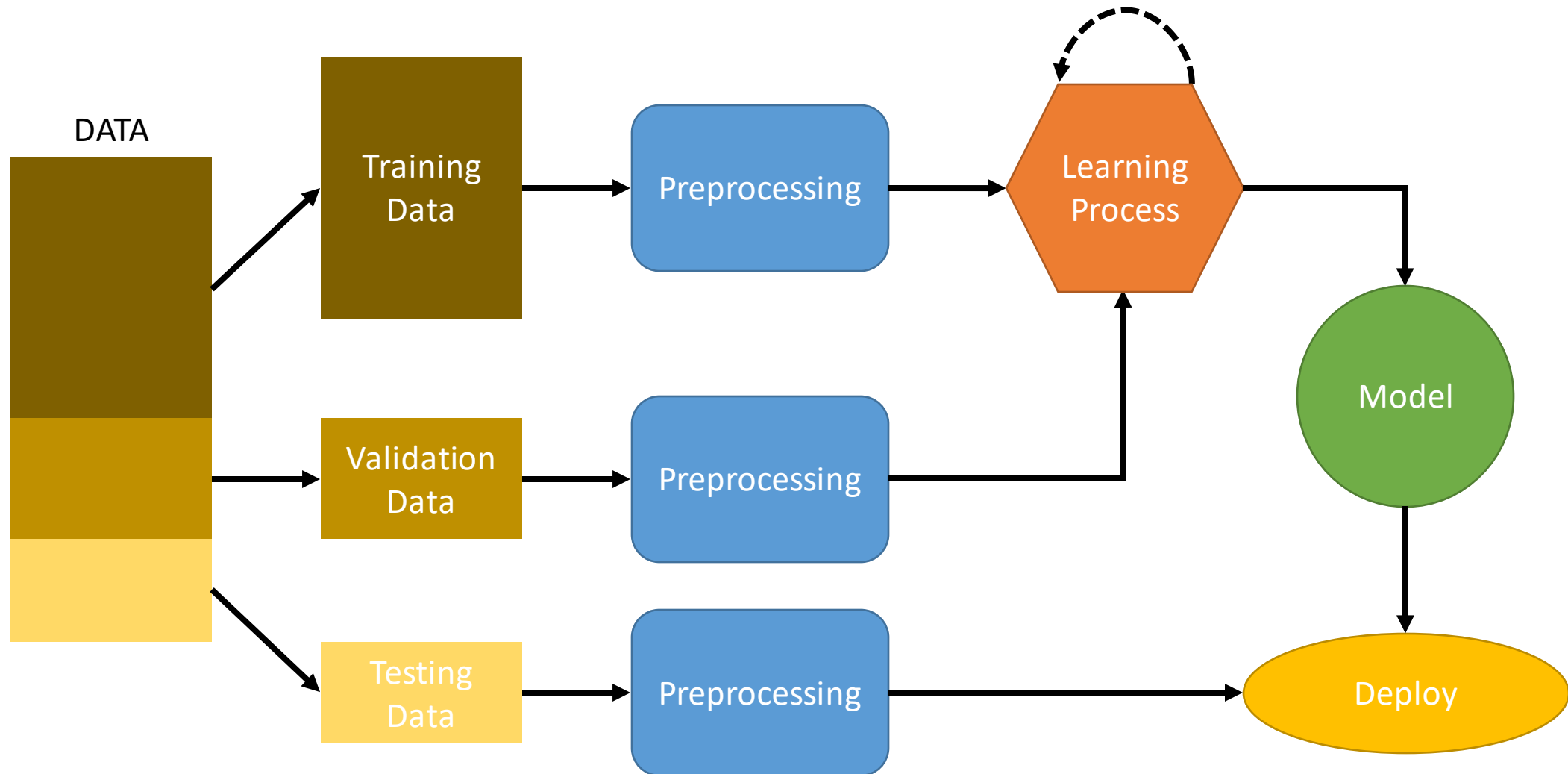


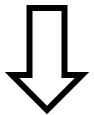
Image - Tensor - Batch



UNIVERSITAS GADJAH MADA

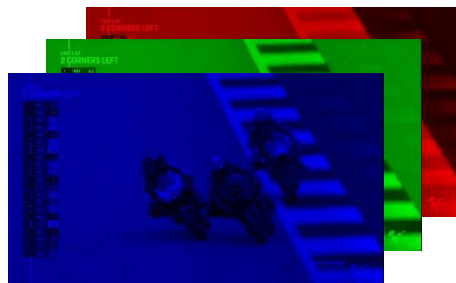
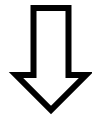
- Komputer membaca sebuah gambar dalam bentuk array berdimensi $c \times h \times w$

Image ke-0



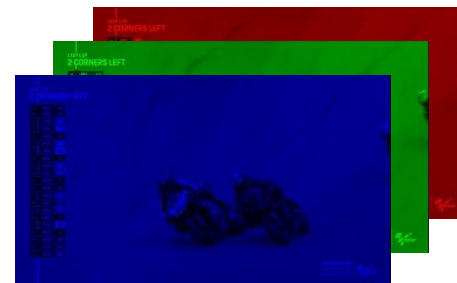
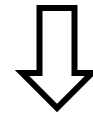
Tensor_0: $3 \times h \times w$

Image ke-1

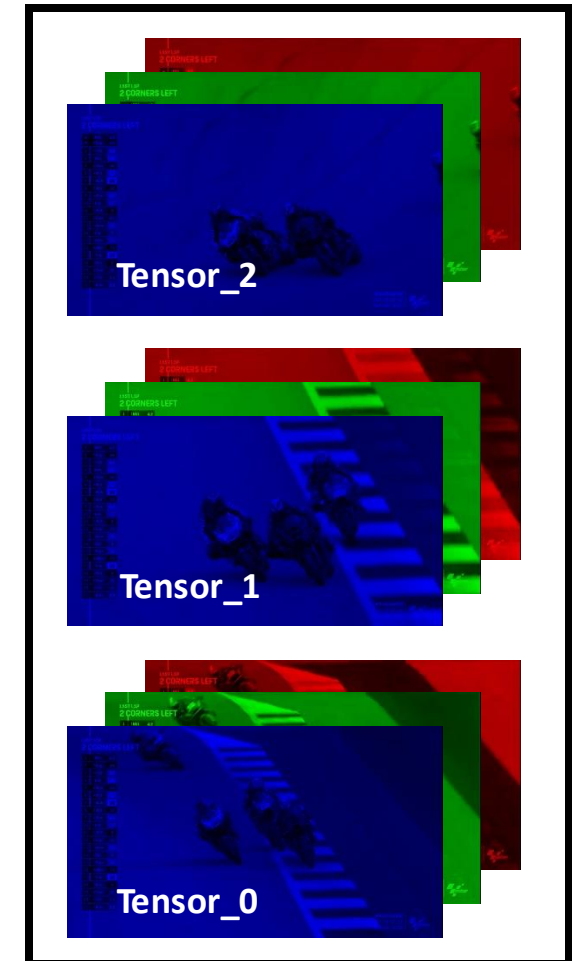


Tensor_1: $3 \times h \times w$

Image ke-2



Tensor_2: $3 \times h \times w$

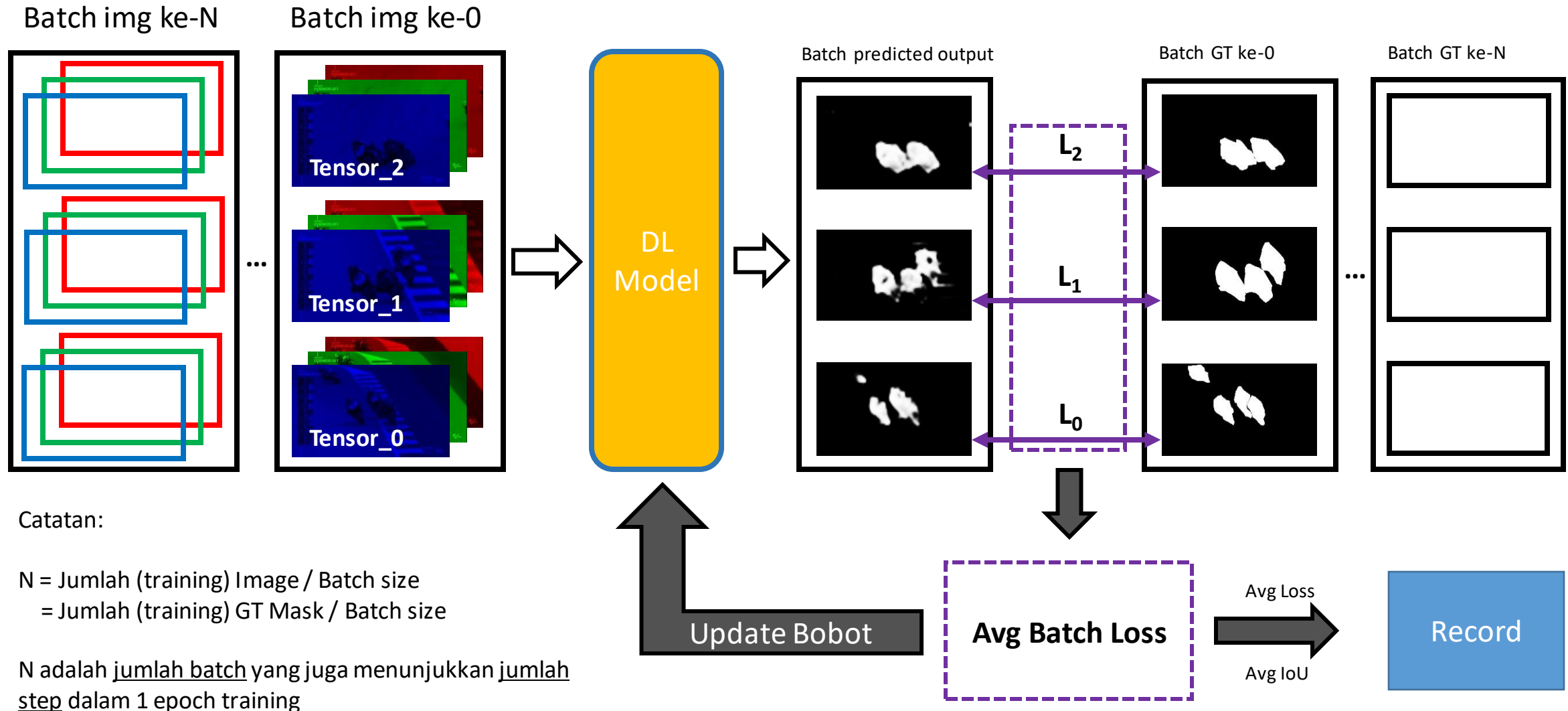


Batch: $3 \times 3 \times h \times w$

Batch Processing



UNIVERSITAS GADJAH MADA

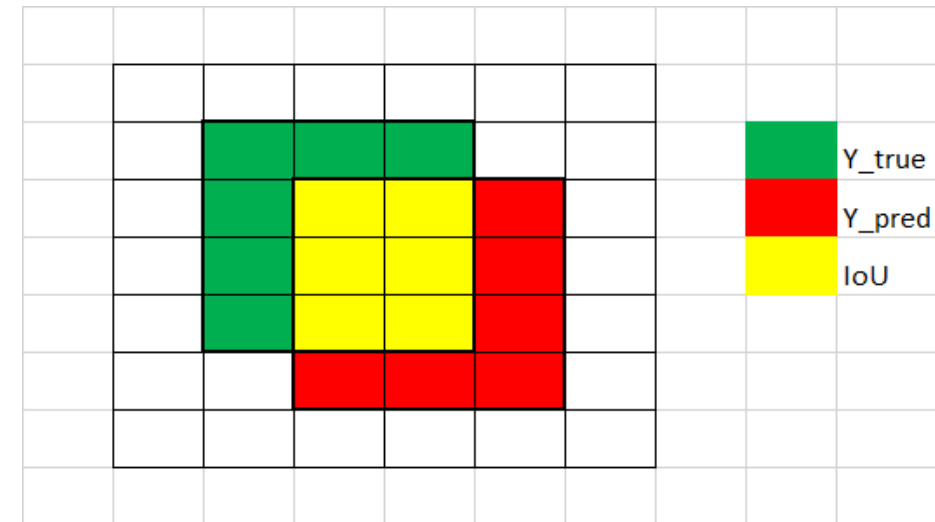


MTL Loss & Metric Functions



- Binary Cross Entropy (BCE) digunakan sebagai loss function pada persoalan pixel-wise classification (0 atau 1).
- MAE digunakan sebagai loss function pada persoalan regresi (0 s/d 1).
- Multi-task loss:
 - Loss MTL = BCE + MAE, atau:
 - Loss MTL = $a \times \text{BCE} + b \times \text{MAE}$
 - Di mana a dan b adalah bobot loss function (menentukan kefokuskan model terhadap salah satu task)
 - Bobot loss function ini juga dapat dituning secara adaptif

- Perhitungan metric dapat dilakukan secara terpisah:
 - IoU untuk semantic segmentation
 - MAE untuk depth estimation

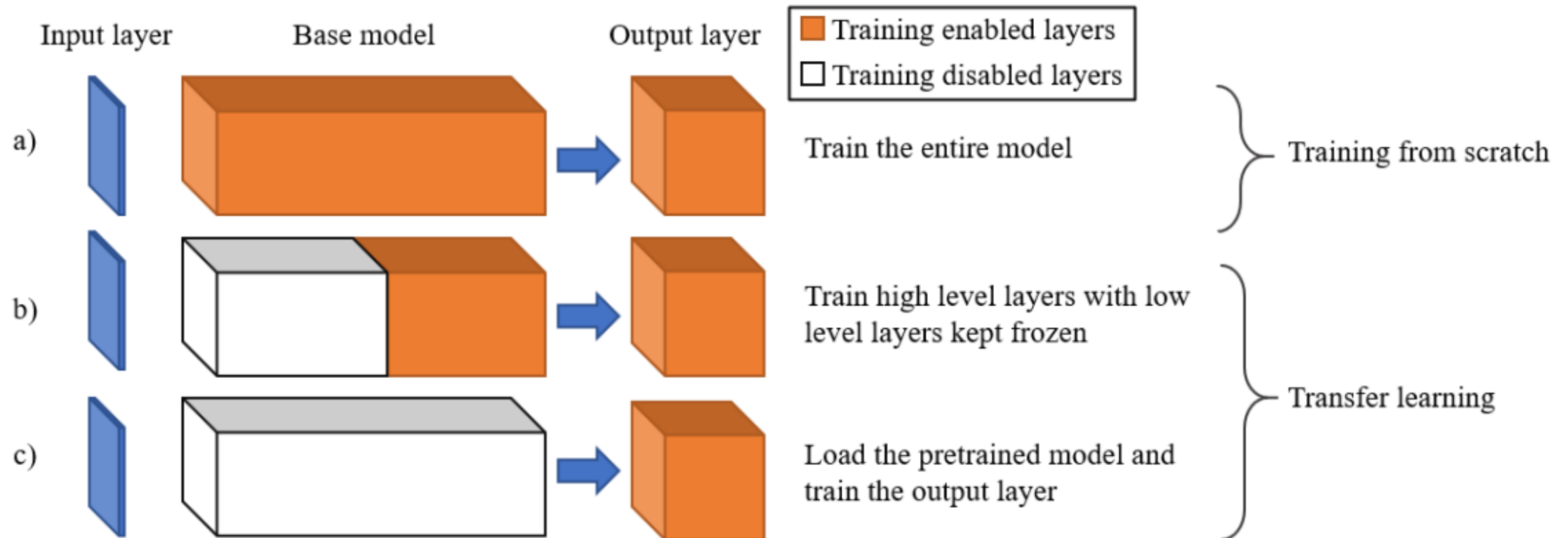


$$J(Y_t, Y_p) = \frac{Y_t \& Y_p}{Y_t \mid Y_p} \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

Transfer Learning



UNIVERSITAS GADJAH MADA



- Dengan transfer learning dari suatu pretrained model, maka model kita lebih cepat convergence dan (umumnya) performa lebih baik
 - Akan tetapi, model (terutama bagian backbone/base) menjadi tidak fleksibel untuk dimodifikasi

**Sumber gambar: arXiv 2107.04537*



UNIVERSITAS
GADJAH MADA

Cloud Computing

LOCALLY ROOTED, GLOBALLY RESPECTED

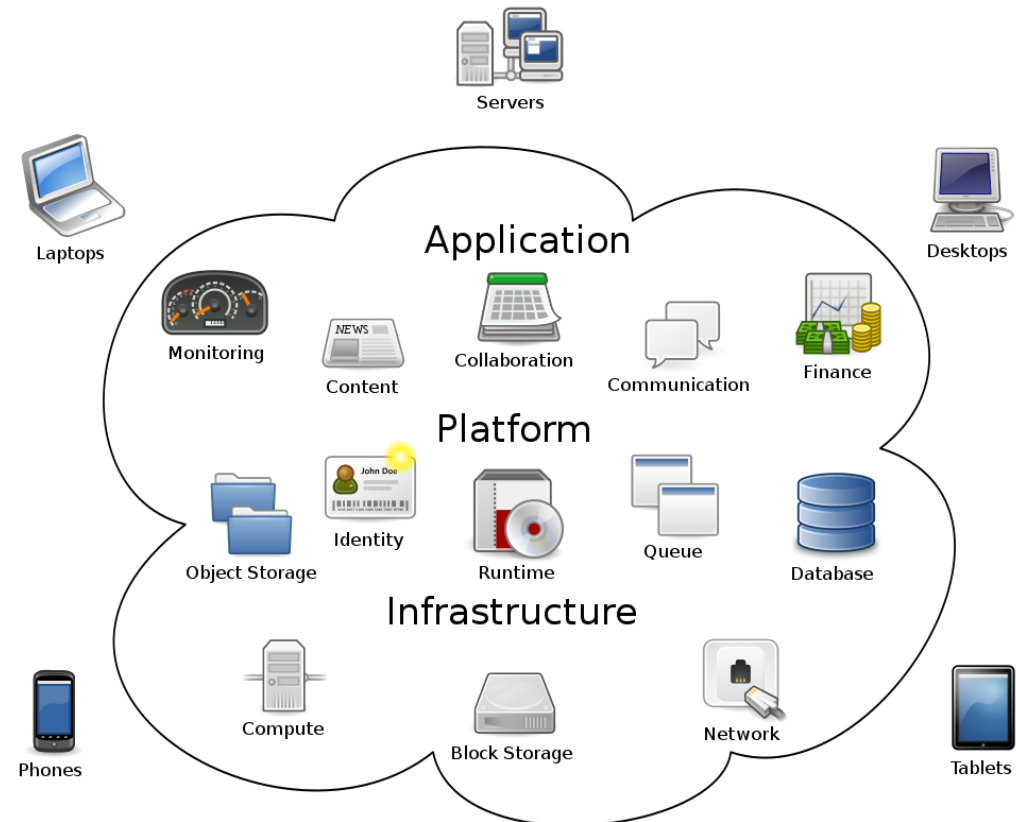
ugm.ac.id

Apa itu Cloud Computing?



UNIVERSITAS GADJAH MADA

- Cloud computing adalah sumber daya sistem komputer berdasarkan permintaan, terutama untuk penyimpanan data (storage) dan kekuatan komputasi (computing power), tanpa pengelolaan aktif secara langsung oleh pengguna.
- Cloud computing environment sering memiliki fungsi yang didistribusikan di beberapa lokasi.
- Cloud computing bergantung pada berbagi sumber daya untuk mencapai koherensi dan biasanya menggunakan model "bayar sesuai penggunaan" (pay-as-you-go) yang dapat membantu mengurangi biaya modal.
 - Tetapi juga dapat menyebabkan biaya operasional yang membengkak bagi pengguna yang tidak teliti.



**Sumber gambar: wikipedia*

Cloud Services



UNIVERSITAS GADJAH MADA



- Amazon Web Services
- Google Cloud Platform
- Microsoft Azure Cloud
- Alibaba Cloud
- Huawei Cloud
- IBM Cloud



Google Cloud Platform



Alibaba Cloud



HUAWEI CLOUD



IBM Cloud

**Sumber gambar: google images*

Google Colab



UNIVERSITAS GADJAH MADA

- Google Colab adalah IPYNB environment gratis yang sepenuhnya berjalan di google cloud.
 - Tidak memerlukan setup (instalasi library, driver GPU, dll).
 - Colab mendukung banyak library machine learning populer (pytorch, TF, dll).
 - Terdapat limitation untuk free account: Jumlah GPU, kapasitas drive, dll



- More Tutorials:
 - <https://machinelearningmastery.com/google-colab-for-machine-learning-projects/>
 - https://www.tutorialspoint.com/google_colab/what_is_google_colab.htm

**Sumber gambar: google colab*



UNIVERSITAS
GADJAH MADA

Code

<https://github.com/oskarnatan/mtl-bisaai>

LOCALLY ROOTED, GLOBALLY RESPECTED