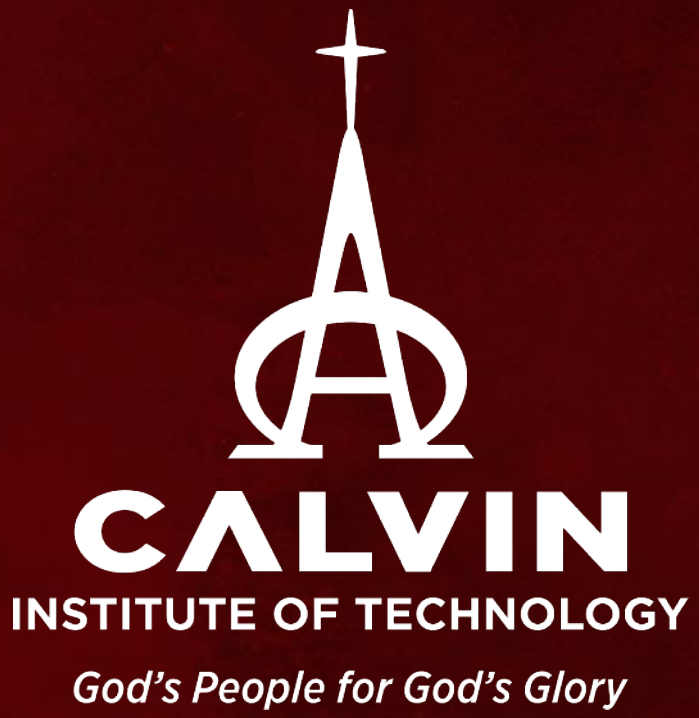


MultiTask Learning

Hendrik Santoso Sugiarto

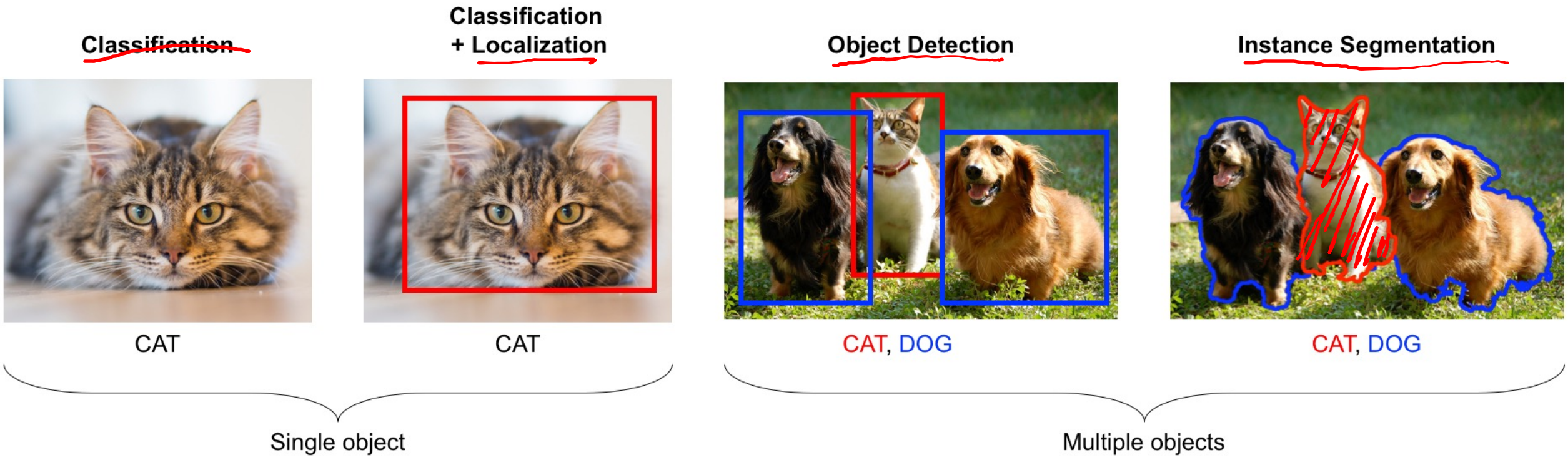
Capaian Pembelajaran

- Multitask
  - Mengerti sistem pendeteksi objek
  - Mengerti sistem segmentasi objek
  - Mengerti sistem T5
- Siamese
  - Mengerti sistem pengenalan wajah
  - Mengerti sistem pengenalan tulisan
- Multiobjective
  - Mengerti sistem transfer style

Multitask

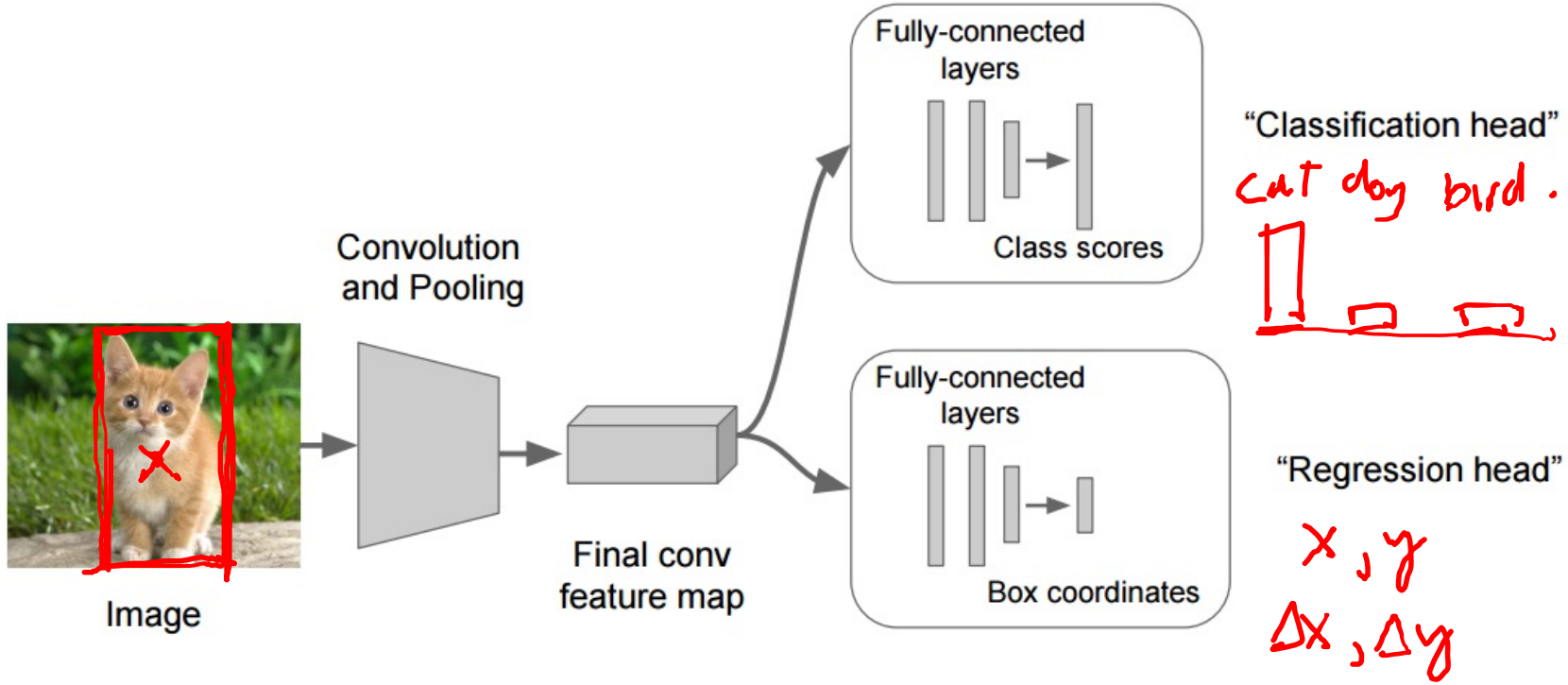
Deteksi Objek

- Lokalisasi vs deteksi vs segmentasi



Lokalisasi

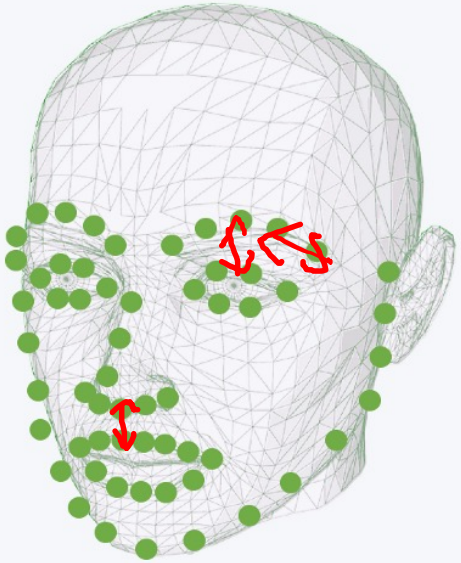
- Multitask Learning
- Prediksi:
  1. Apakah ada objek?
  2. Lokasi pusat kotak (x,y)
  3. Tinggi kotak
  4. Lebar kotak
  5. Kategori objek



Deteksi Landmark

- Multitask learning dapat juga digunakan untuk memprediksi lokasi banyak poin sekaligus

FACIAL LANDMARK POINTS

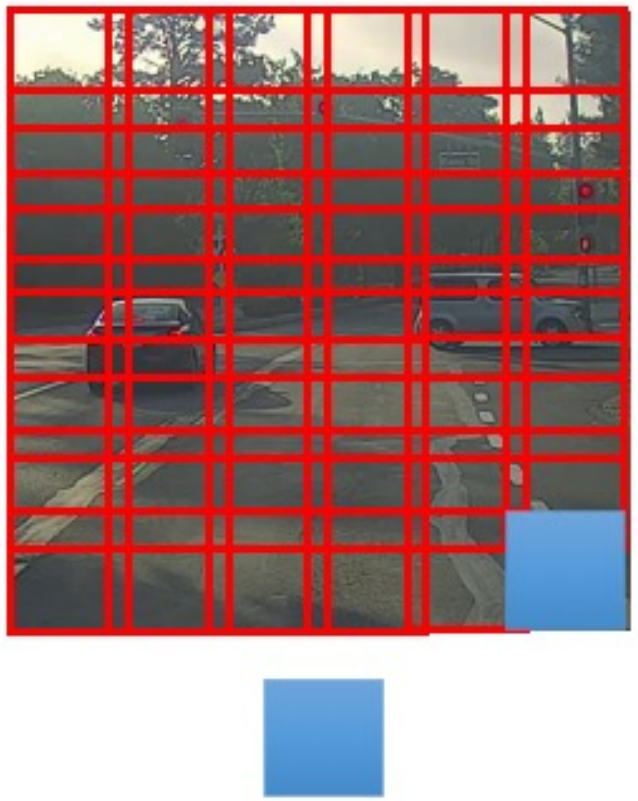


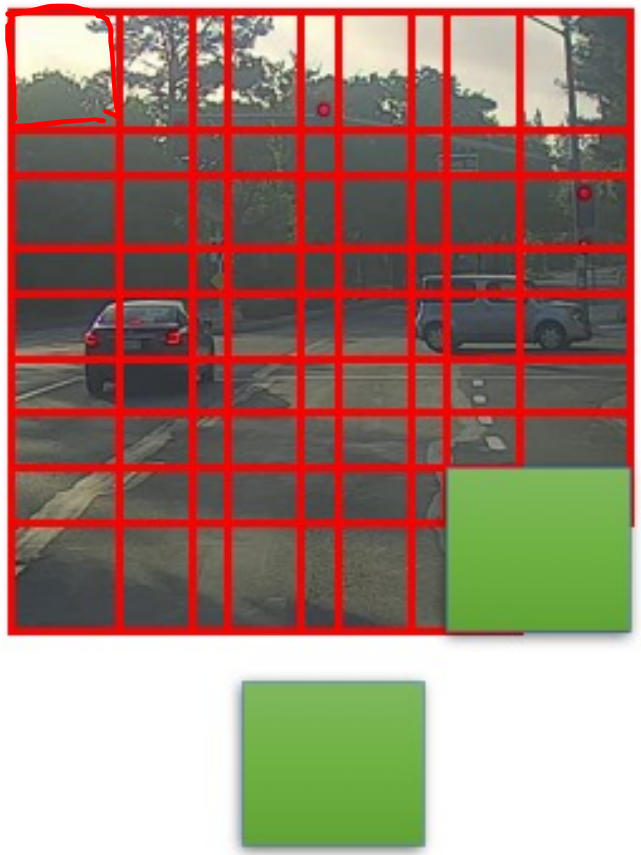
Feature	Point range
Left jaw line	0-7
Chin	8
Right jaw line	9-16
Left eyebrow	17-21
Right eyebrow	22-26
Bridge of nose	27-30
Bottom of nose	31-35
Left eye	36-41
Right eye	42-47
Outer edge of lips	48-59
Inner edge of lips	60-67

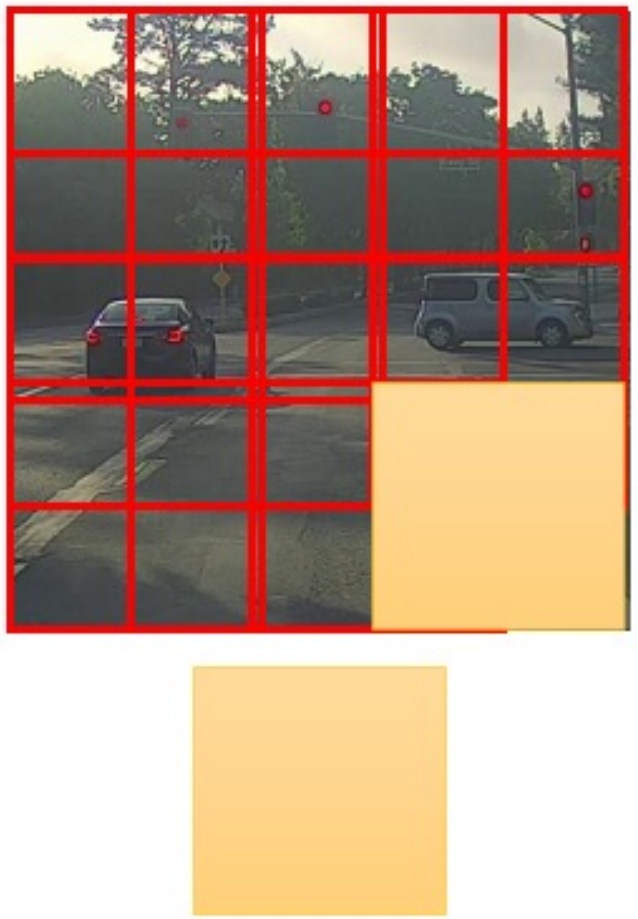
68^2

CALVIN INSTITUTE OF TECHNOLOGY

Sliding Window



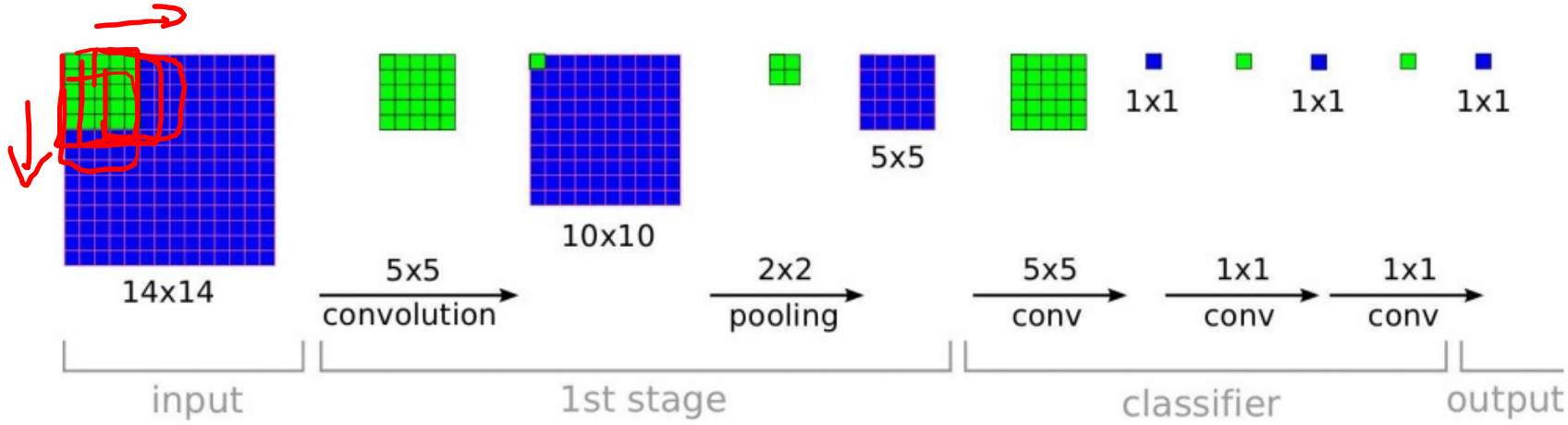


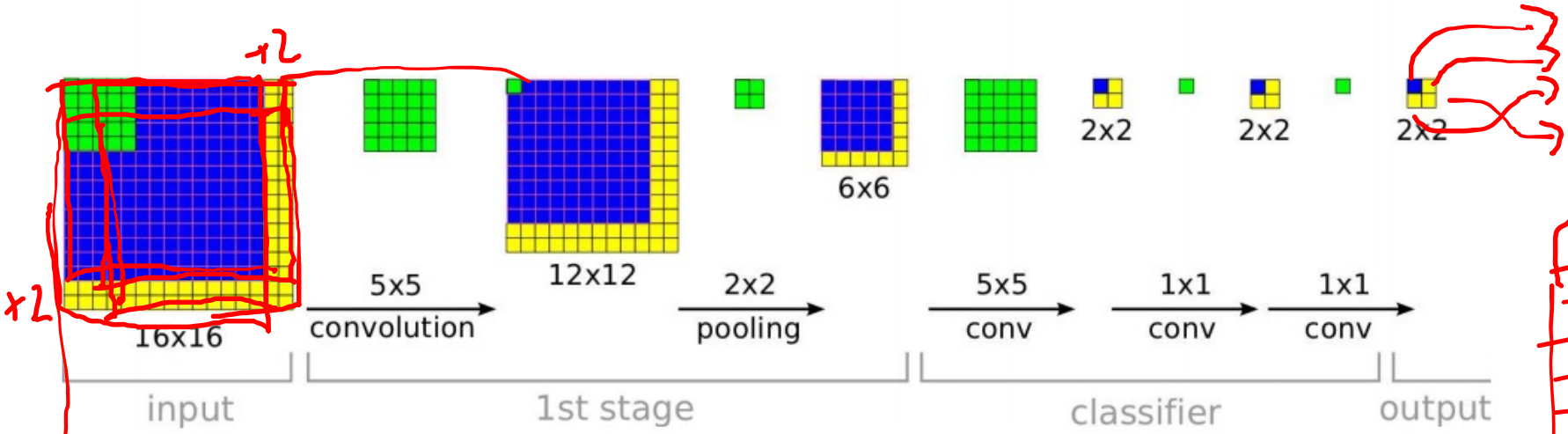


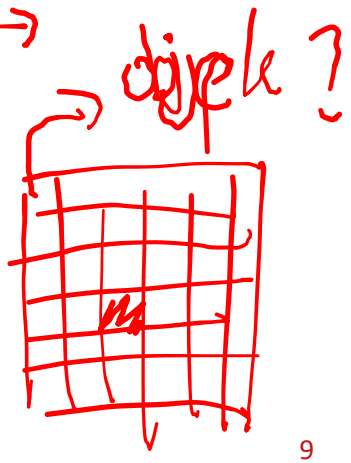
Deteksi Objek dengan Sliding Window

- Latih model untuk mendeteksi objek tertentu
- Coba prediksi pada semua kemungkinan sliding window (dengan berbagai ukuran) untuk melihat apakah terdapat objek pada lokasi tersebut
- Problem: sangat lambat dan mahal

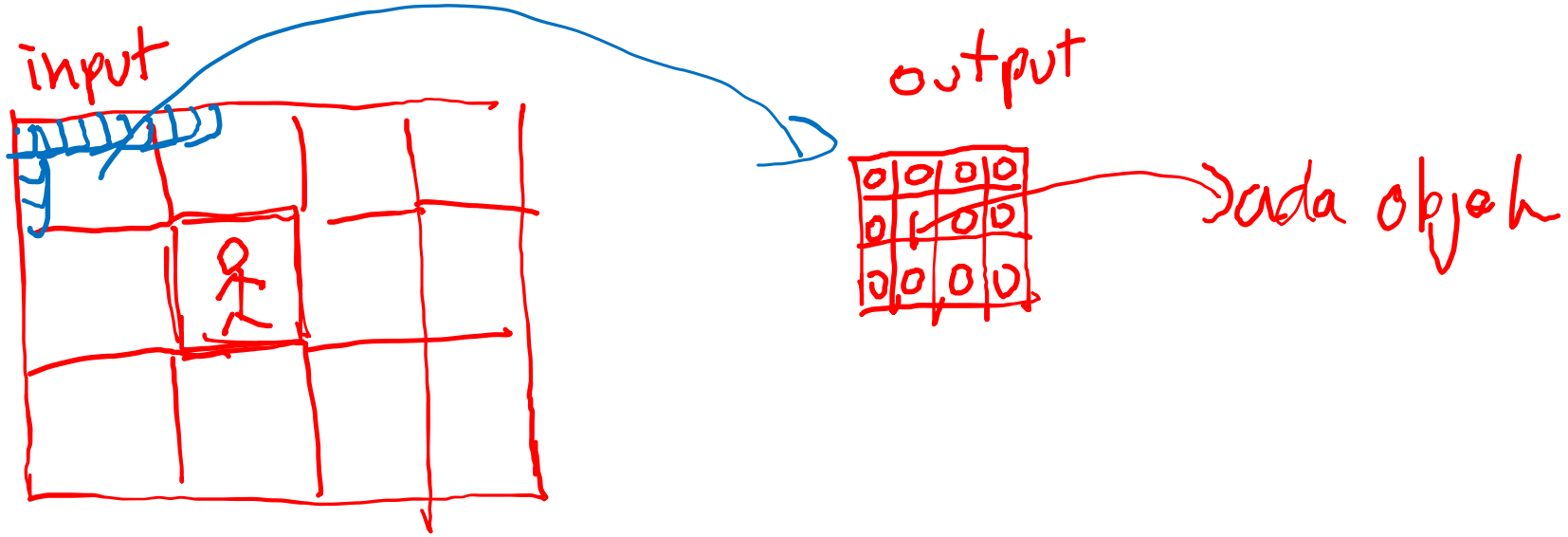
Implementasi konvolusi terhadap sliding window





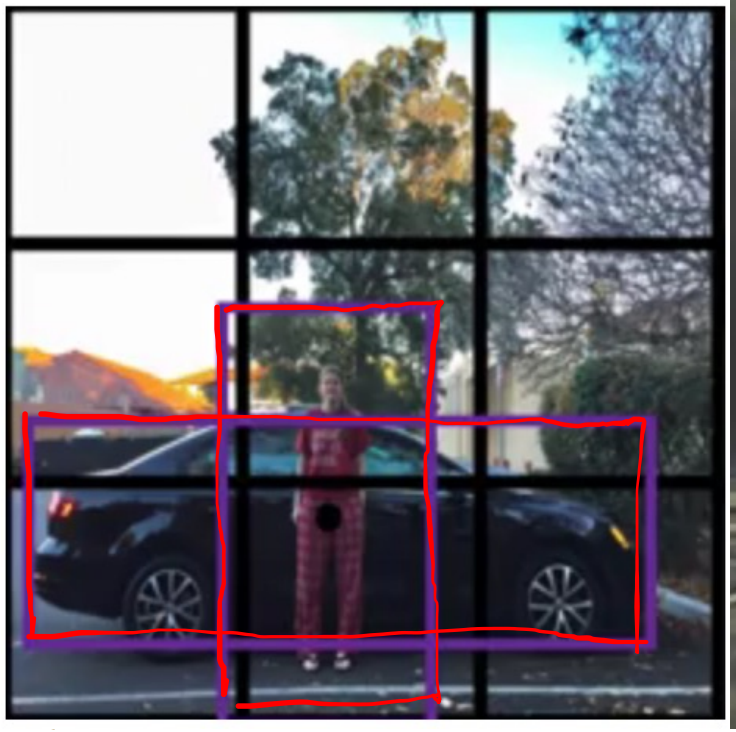


CALVIN INSTITUTE OF TECHNOLOGY

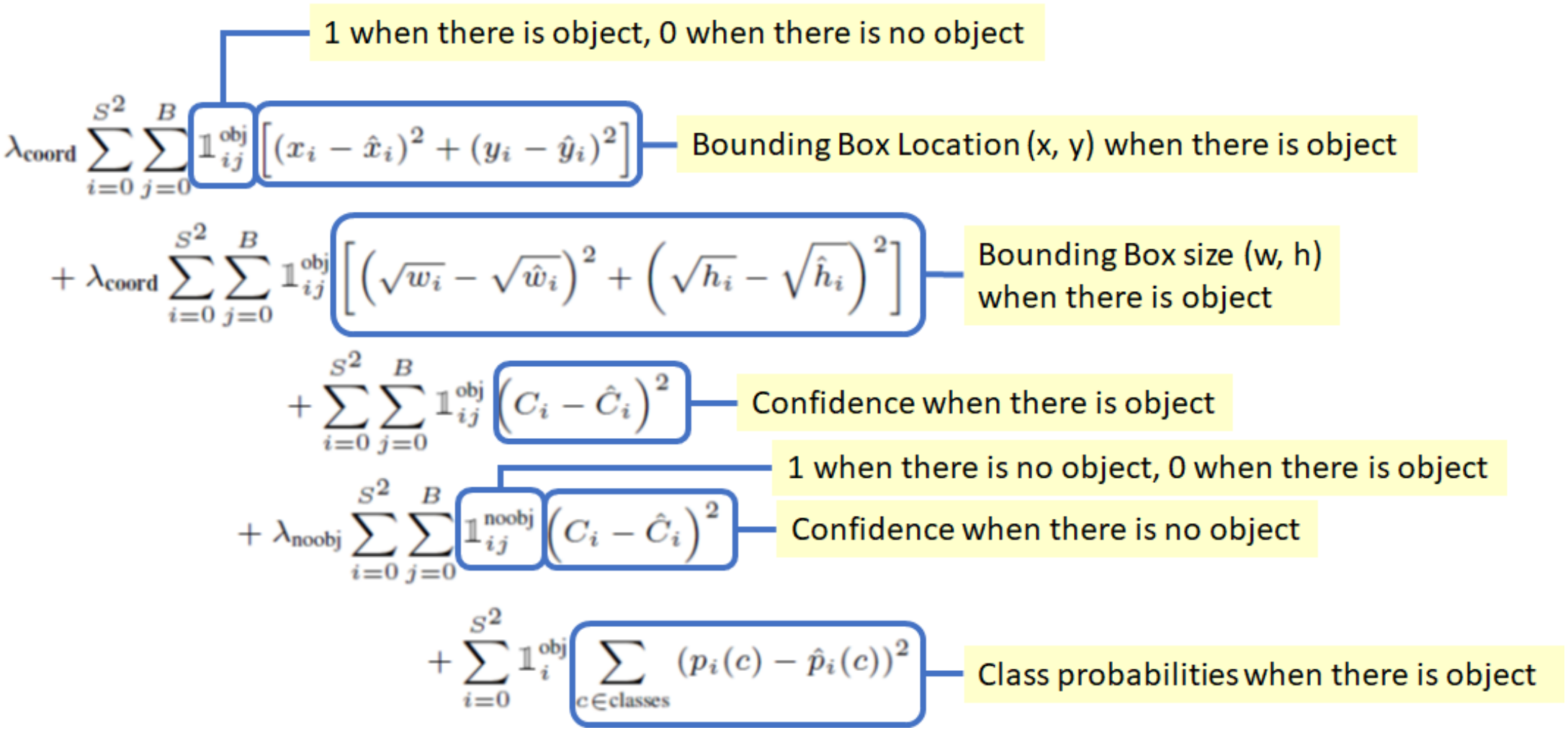


YOLO (you only look once)

- Komponen:
  - Konvolusi sliding windows
  - Intersection over union
  - Non-max suppresion
  - Anchor boxes



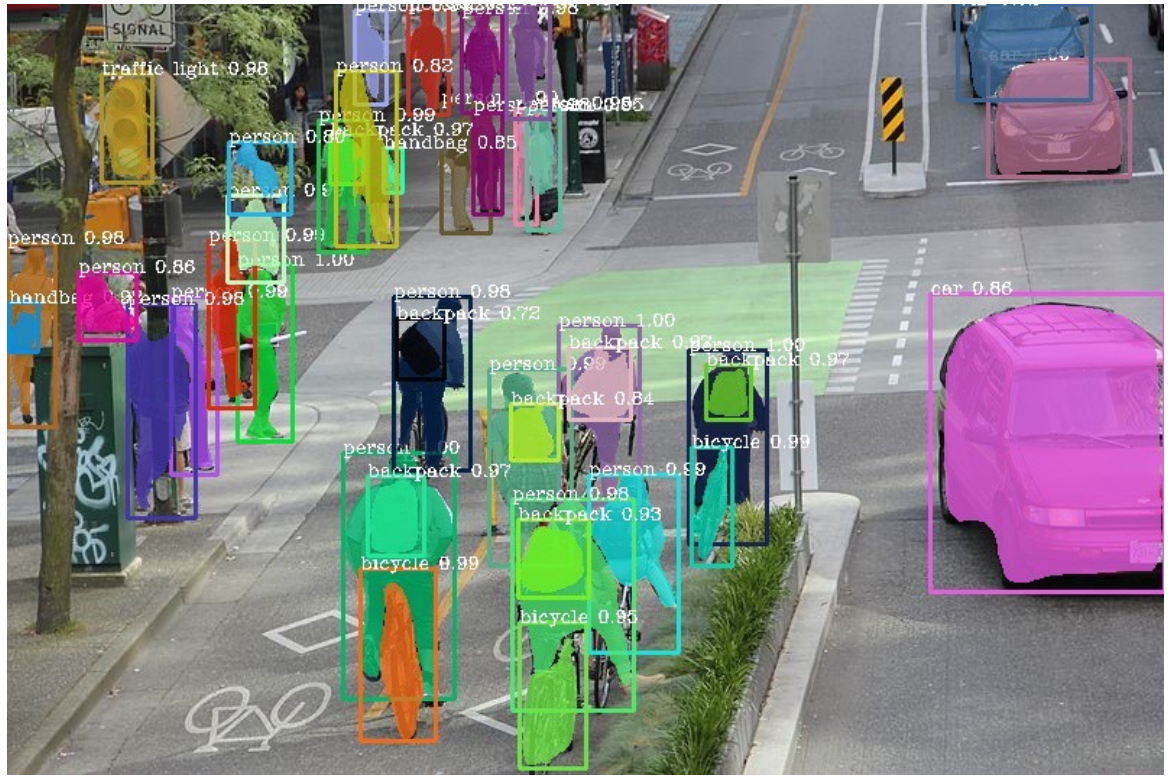
YOLO loss function



CALVIN INSTITUTE OF TECHNOLOGY

Segmentasi

- Problem dari sliding window adalah banyak perhitungan terbuang untuk mendeteksi ruang kosong
- Kita bisa melakukan segmentasi terlebih dahulu sebelum mendeteksi



Region Proposal

- https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html
- R-CNN (Region CNN):
  - Ajukan area.
  - Klasifikasi area yang diajukan satu-persatu
  - Output label + Bounding box
- Fast R-CNN:
  - Ajukan area.
  - Gunakan konvolusi dengan sliding windows untuk klasifikasi semua area yang diajukan
- Faster R-CNN:
  - Gunakan jejaring konvolusi untuk mengajukan area

Seberapa efisien?

Faster!

FASTER!

Better!

	R-CNN	Fast R-CNN
Training Time:	84 hours	**9.5 hours**
(Speedup)	1x	**8.8x**
Test time per image	47 seconds	**0.32 seconds**
(Speedup)	1x	**146x**
mAP (VOC 2007)	66.0	**66.9**

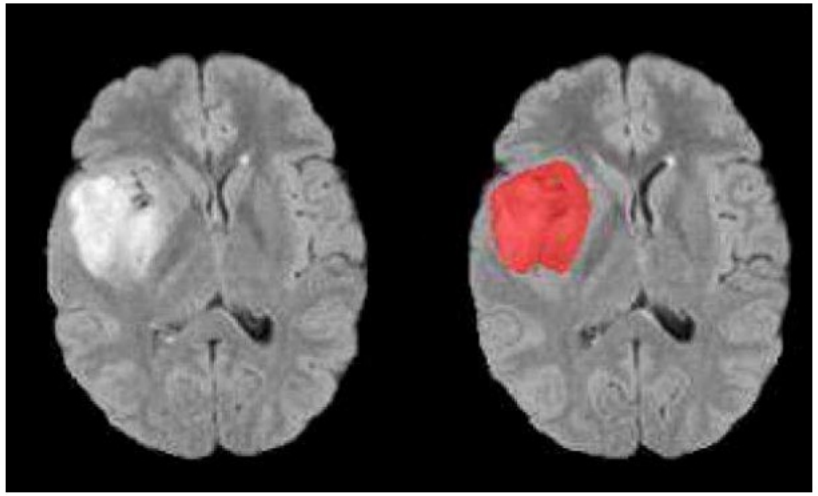
Using VGG-16 CNN on Pascal VOC 2007 dataset

	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image (with proposals)	50 seconds	2 seconds	**0.2 seconds**
(Speedup)	1x	25x	**250x**
mAP (VOC 2007)	66.0	**66.9**	**66.9**

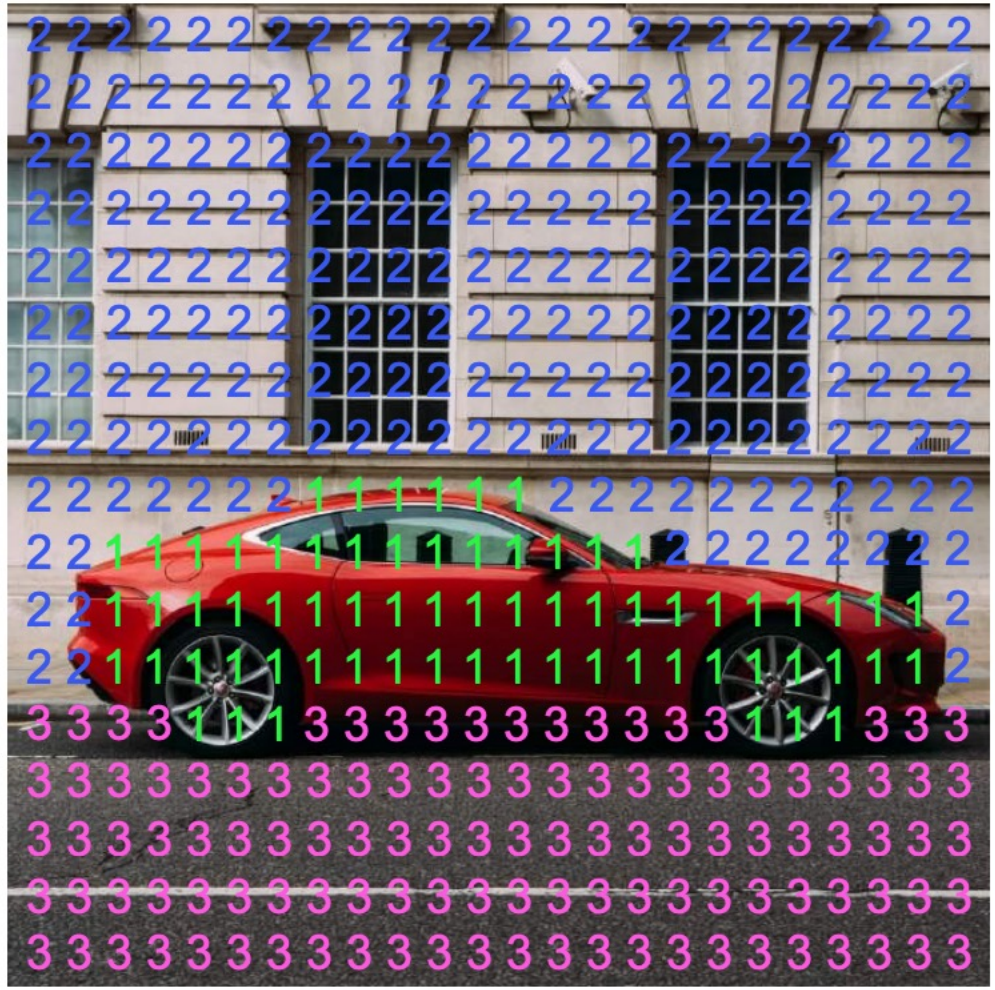
CALVIN INSTITUTE OF TECHNOLOGY

Semantic Segmentation

- Motivasi:
  - Deteksi area kanker



Per pixel class label



1. Car
2. Building
3. Road

U-NET

input

output

Conv, RELU

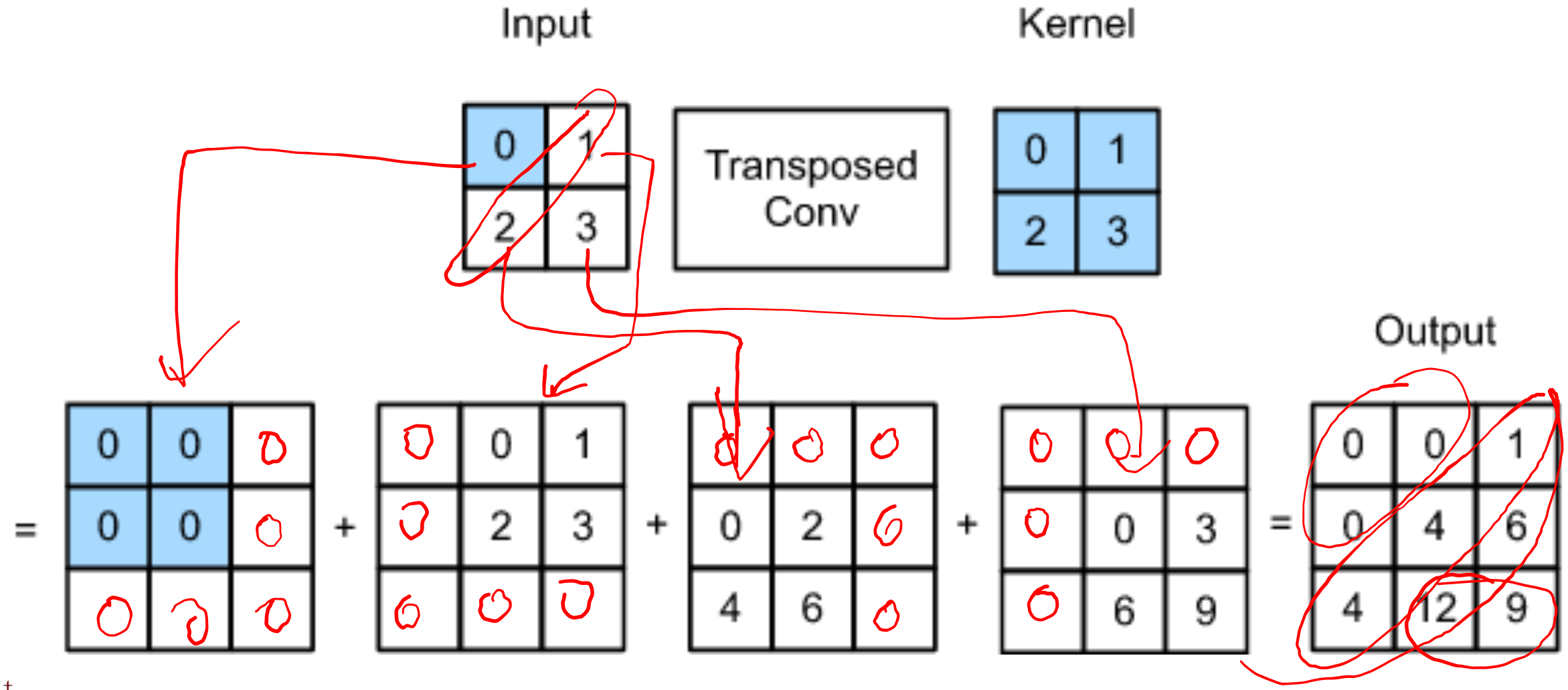
Max Pool

Trans Conv

Skip Connection

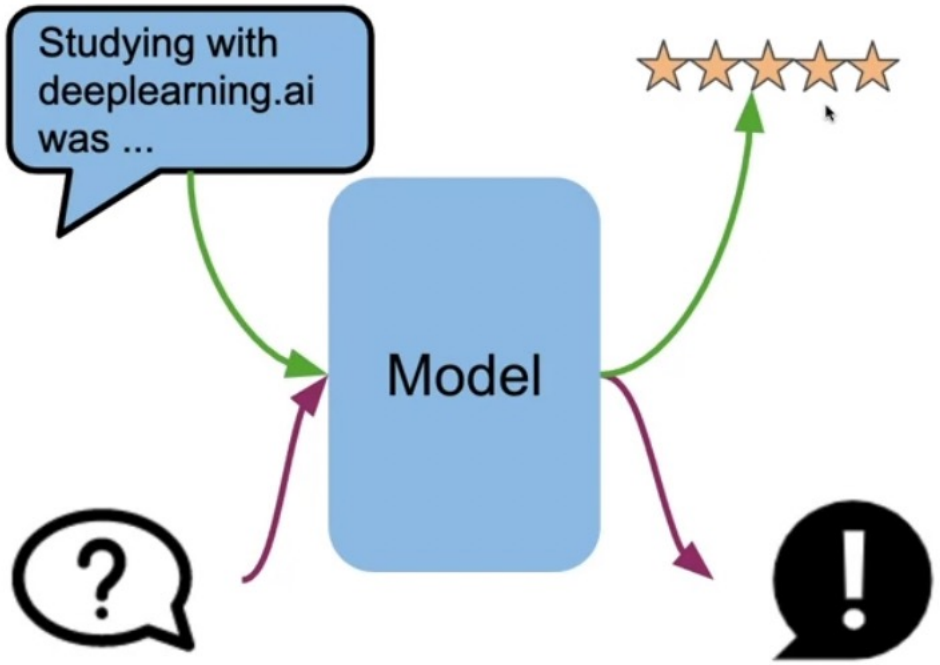
Conv (1x1)

Transpose Convolution

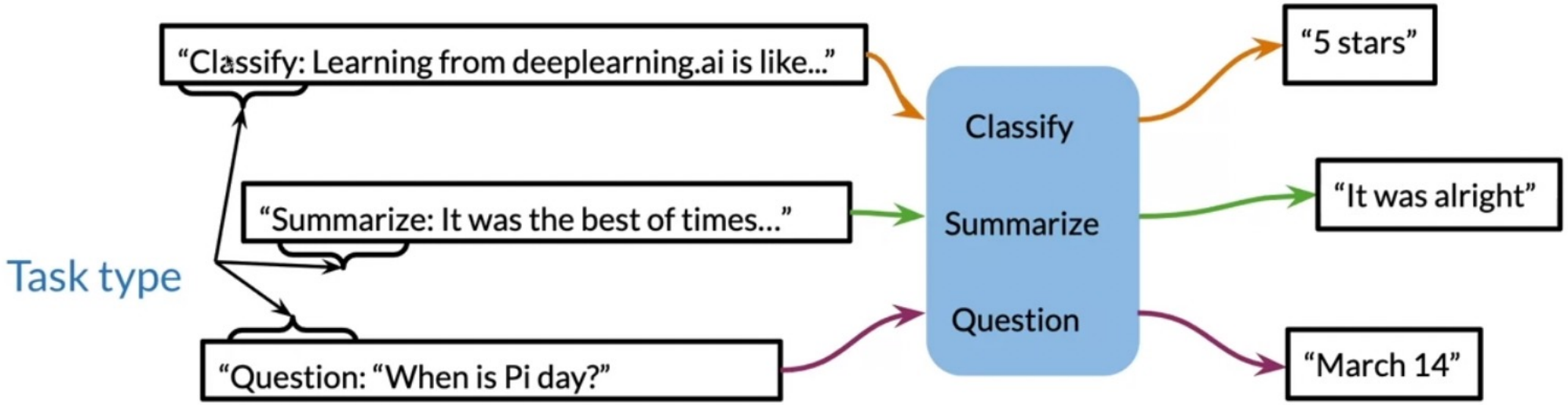


T5: Encoder vs Encoder-Decoder

T5: Multi-task



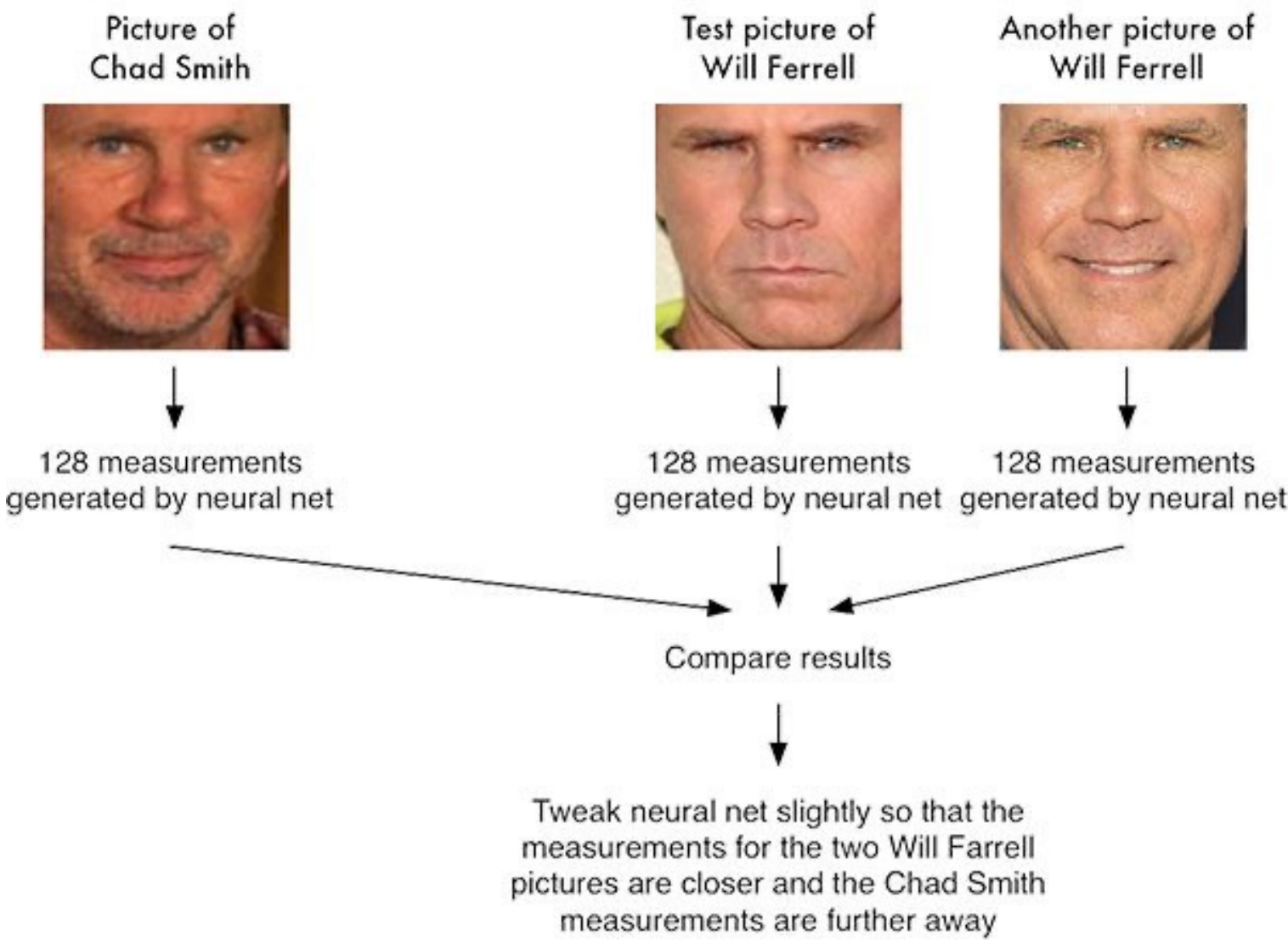
T5: Text-to-text



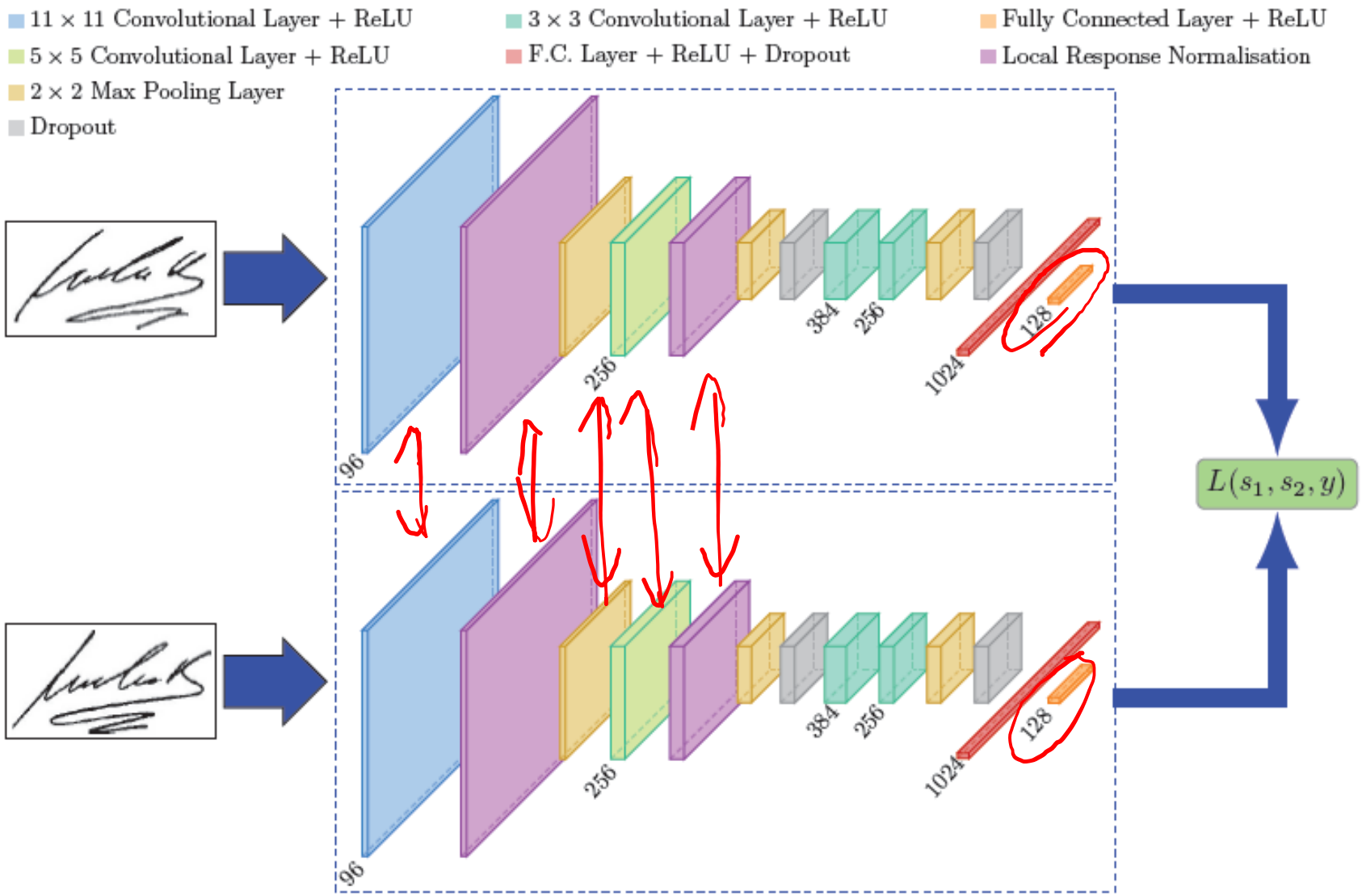
Siamese

Pengenalan wajah

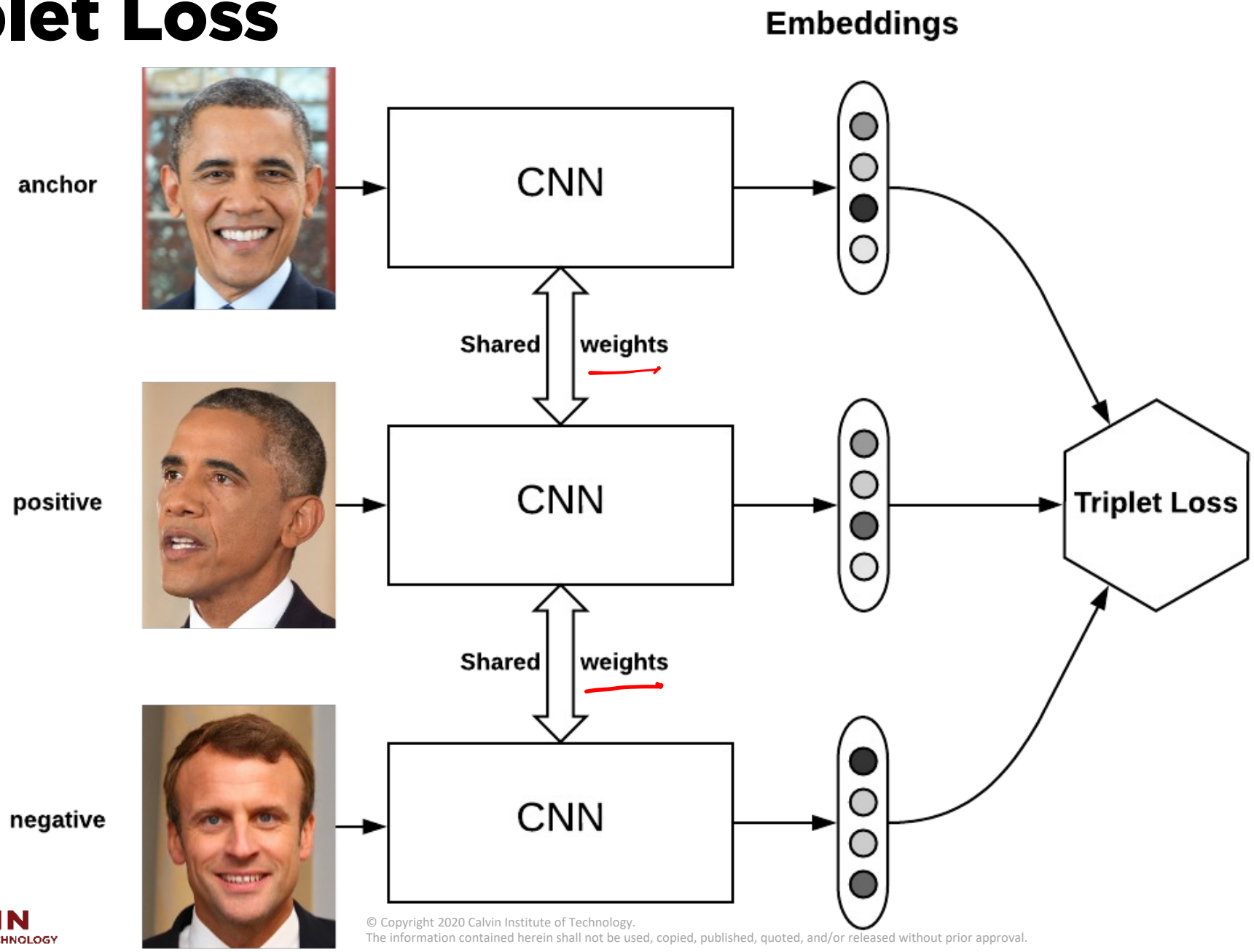
- Problem: 1 shot learning
- Kita biasanya hanya mempunyai 1 foto per nama
- Menggunakan perbandingan pair (kemiripan antara 2 foto)



Siamese Network



Triplet Loss

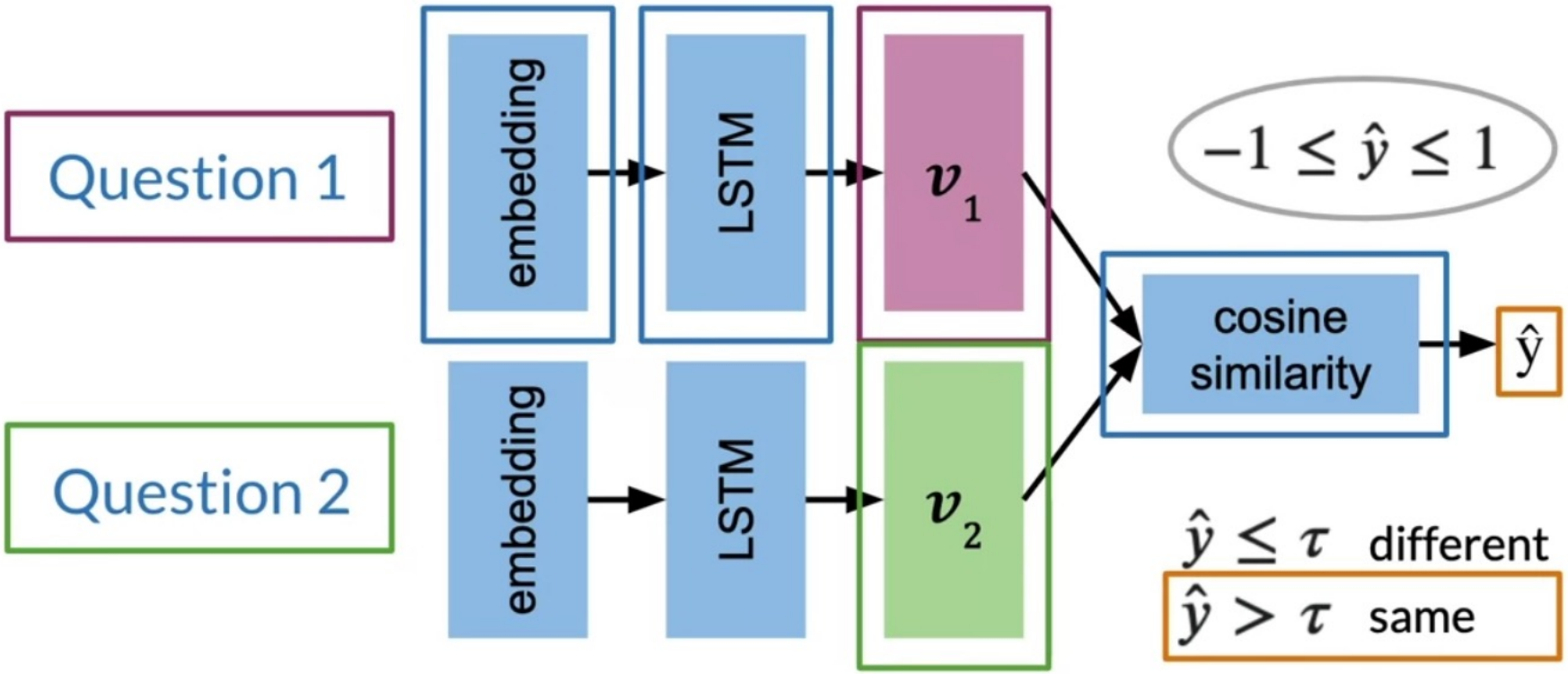


Duplikasi

How old are you? = What is your age?

Where are you | from? ≠ Where are you | going?

Arsitektur



Loss Function

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{||v_1||\,||v_2||}$$

$$s(v_1, v_2)$$

How old are you? — Anchor

What is your age? — Positive

$$s(A, P) \approx 1$$

Where are you from? — Negative

$$s(A, N) \approx -1$$

$$\text{diff} = s(A, N) - s(A, P)$$

Uji Pemahaman

Apakah 2 subnetworks pada arsitektur Siamese share parameter yang sama?

- Benar
- Salah

Triplet Loss

Triplet Loss

Simple loss:

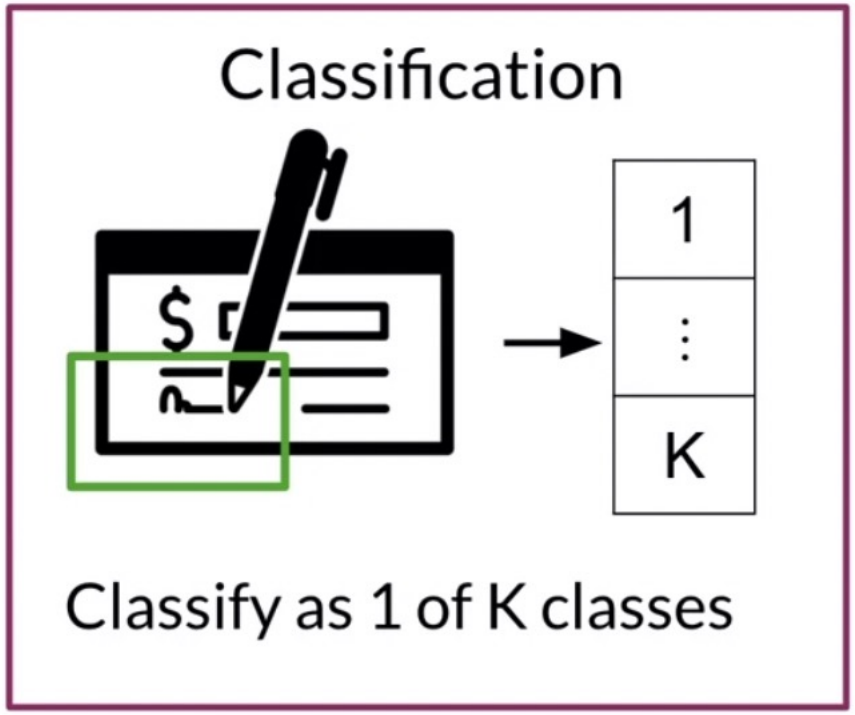
$$\text{diff} = \text{s}(A, N) - \text{s}(A, P)$$

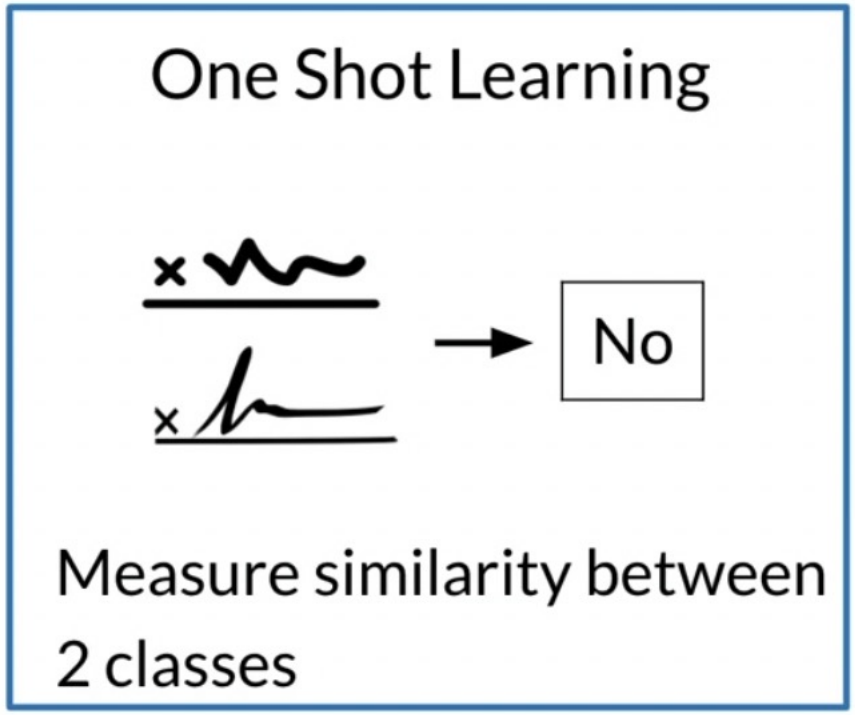
How old are you? Anchor

What is your age? Positive

Where are you from? Negative

Classification vs One Shot Learning





One Shot Learning

No need for retraining !

Learn a similarity score!

0.9

$$s(sig1, sig2) > \tau$$ ✅

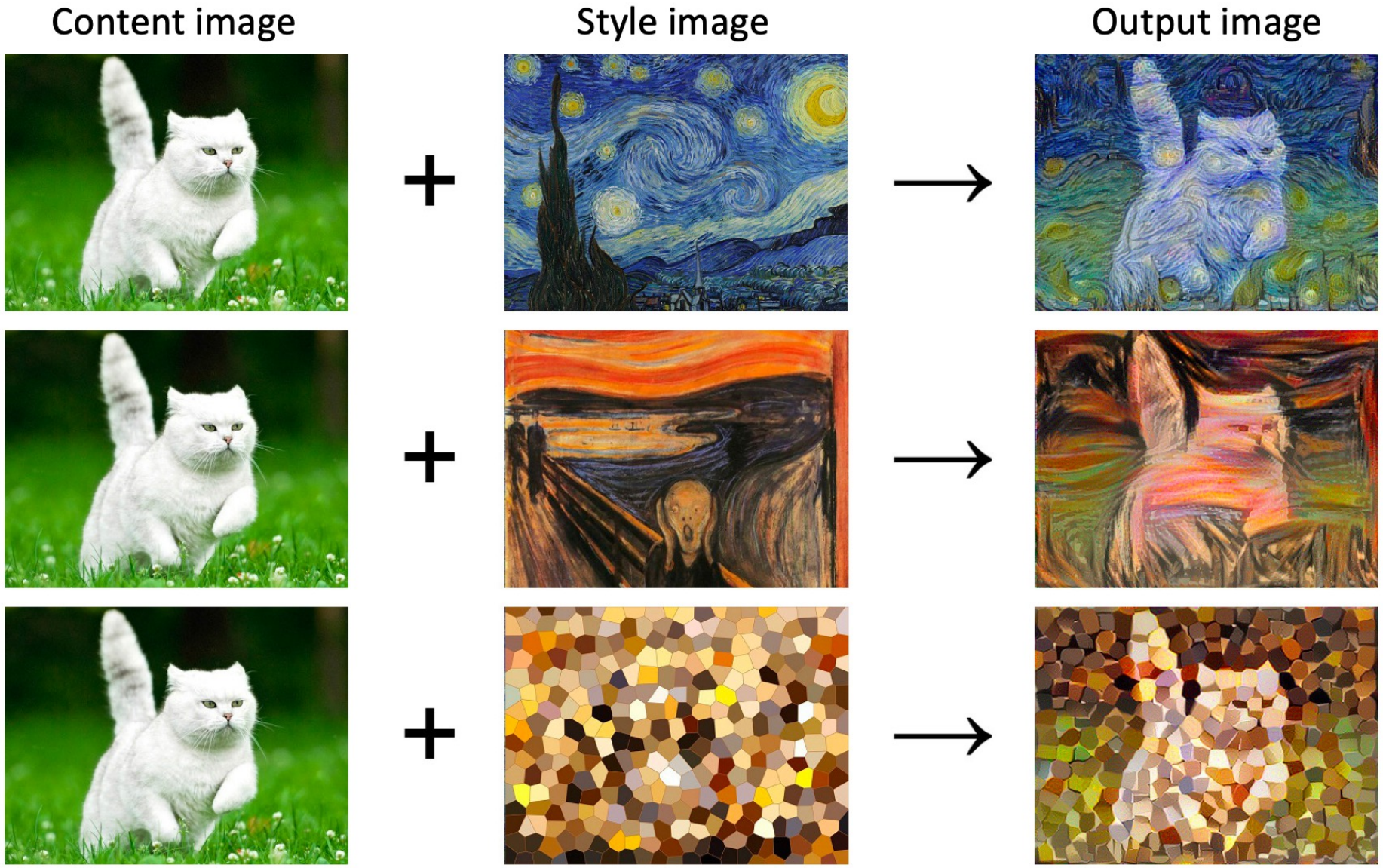
$$s(sig1, sig2) \leq \tau$$ ❌

Uji Pemahaman

- Manakah nilai similarity tertinggi?
  - Antara anchor dan positive
  - Antara anchor dan negative
  - Antara positive dan negative

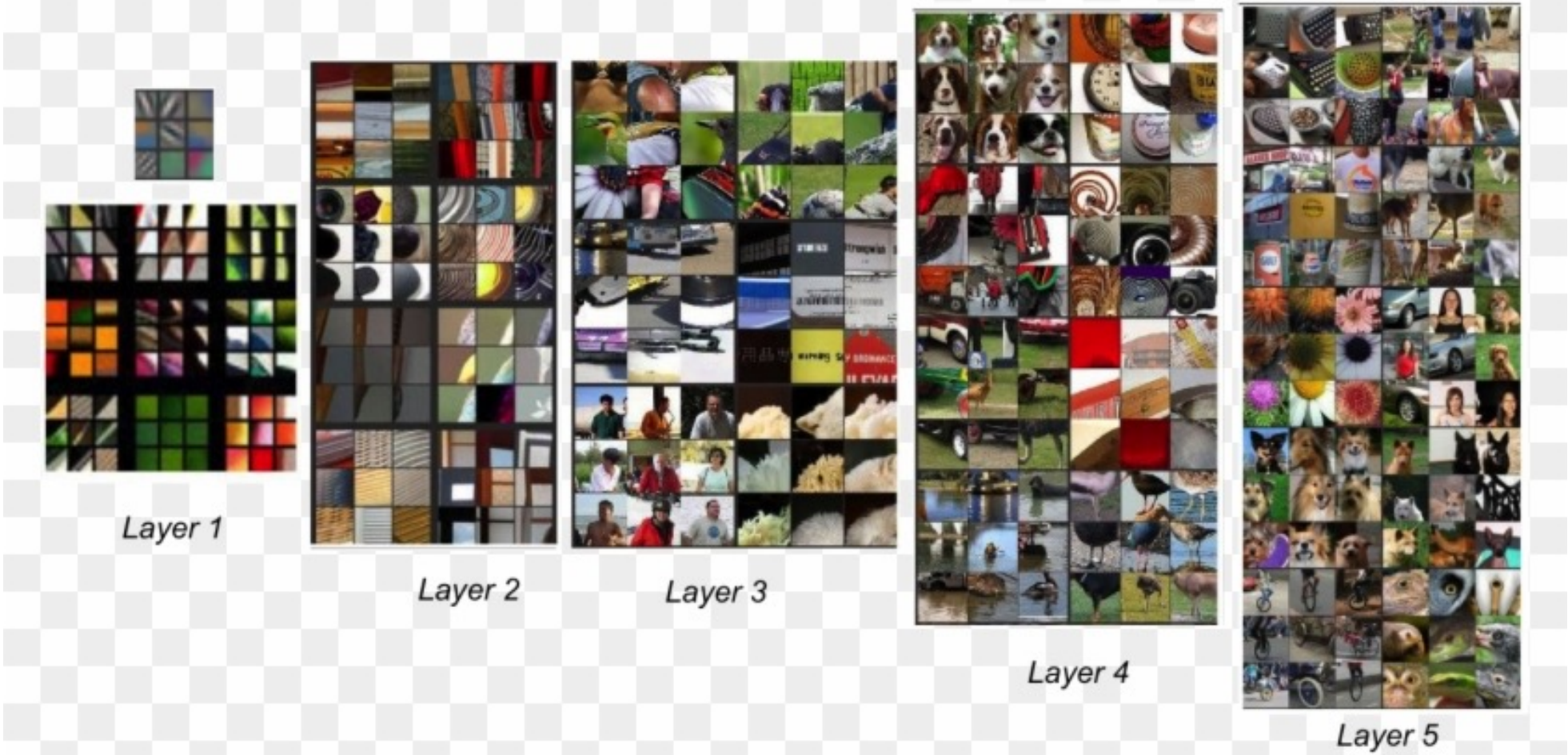
Multiobjective

Content + Style



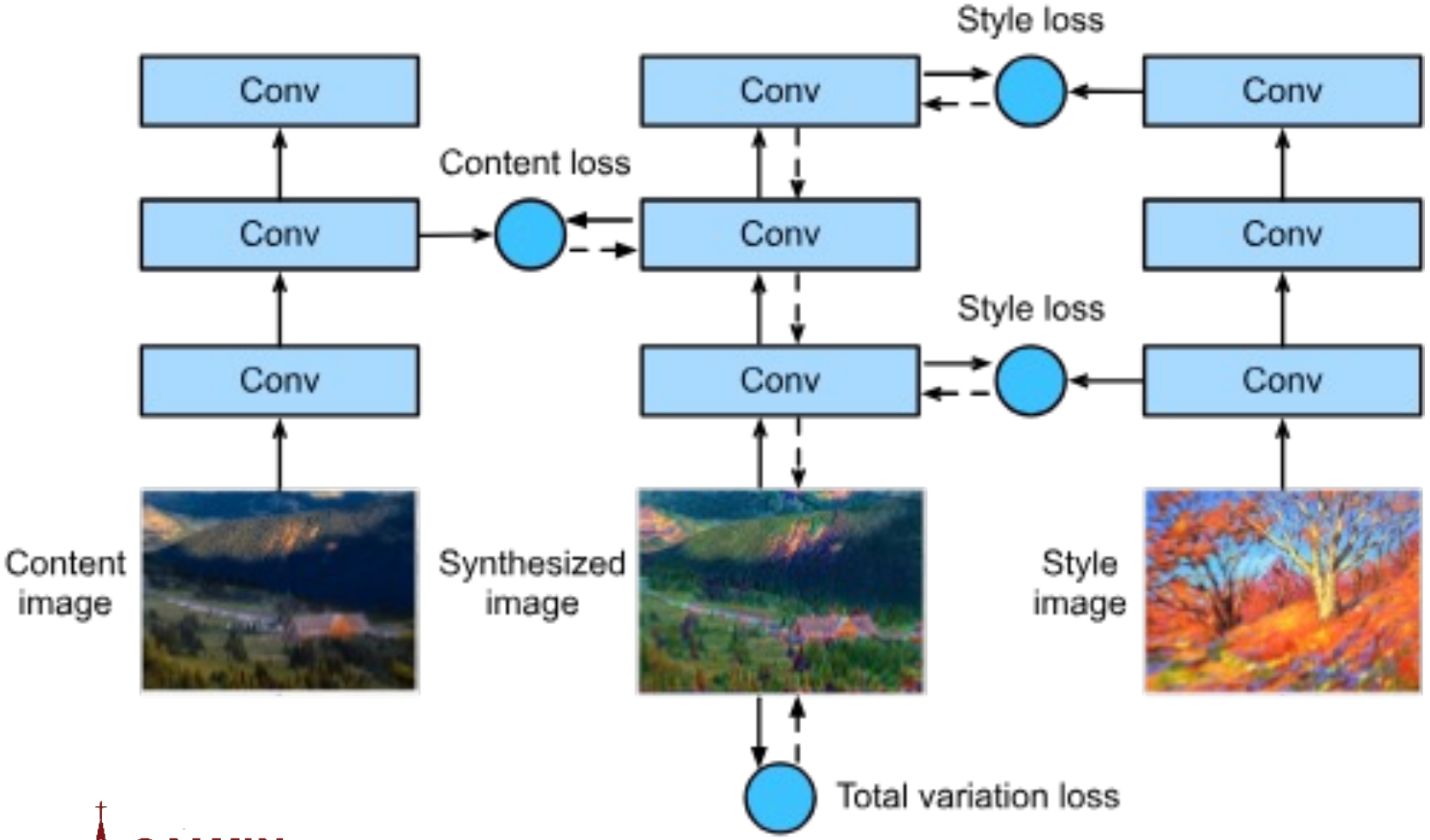
Visualizing hidden layer

- https://arxiv.org/pdf/1311.2901.pdf



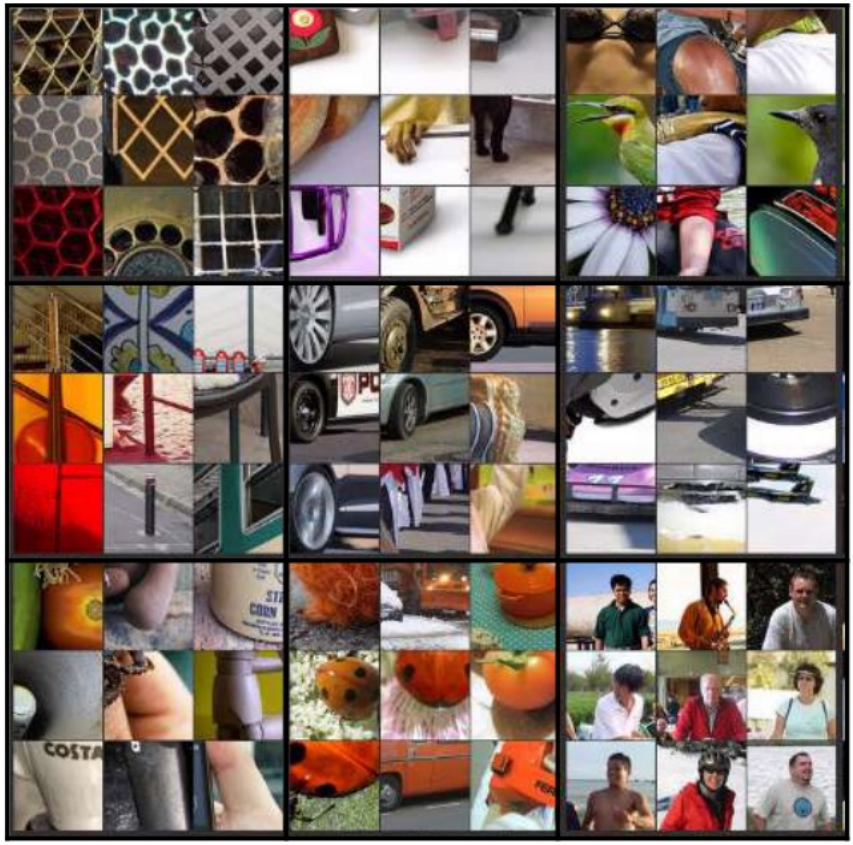
Multi Objective Function

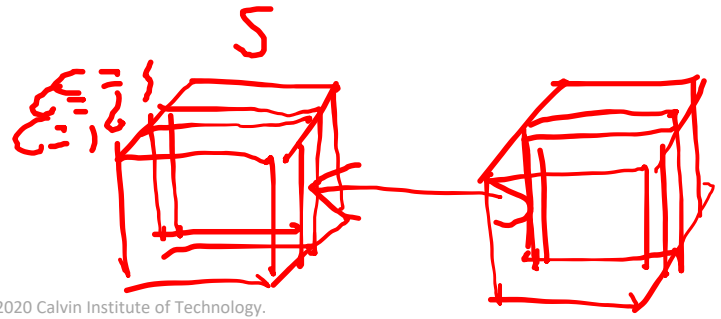
- $J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$



Style Transfer Loss $J(G) = \alpha J_{content}(C, G) + \beta J_{style}(S, G)$

- Content loss function:
  - Misalkan $a^{(l)C}$ dan $a^{(l)G}$ adalah layer aktivasi
  - jika $a^{(l)C}$ dan $a^{(l)G}$ mirip, maka konten kedua gambar mirip
  - $J_{content}(C, G) = \frac{1}{2}\left|\left|a^{(l)C} - a^{(l)G}\right|\right|^2$
- Style loss function:
  - Style didefinisikan sebagai korelasi antar aktivasi across channels
  - Misalkan $a^{(l)}_{i,j,k}$ merupakan aktivasi pada (i,j,k)
  - Maka $G^{(l)}_{kk'} = \sum_{ij} \sum_{ij} a^{(l)}_{i,j,k} a^{(l)}_{i,j,k'}$
  - $J_{style}(S, G) = \frac{1}{\left(2n_H^{(l)} n_W^{(l)} n_C^{(l)}\right)^2} \sum_k \sum_{k'} \left(G^{(l)S}_{kk'} - G^{(l)G}_{kk'}\right)$





GAN

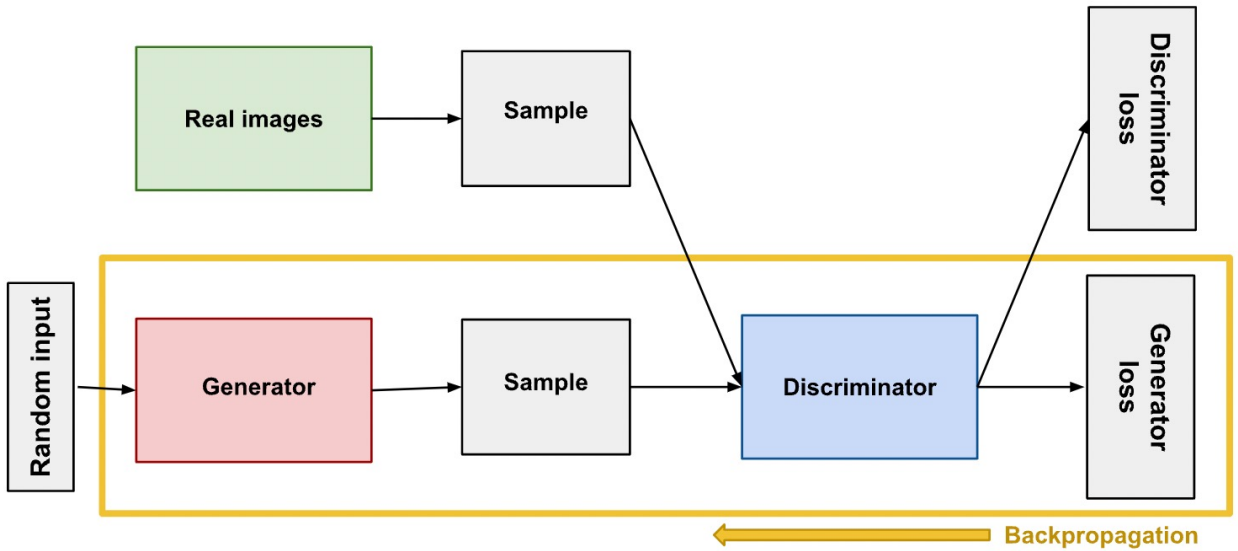


Figure 1: Backpropagation in generator training.

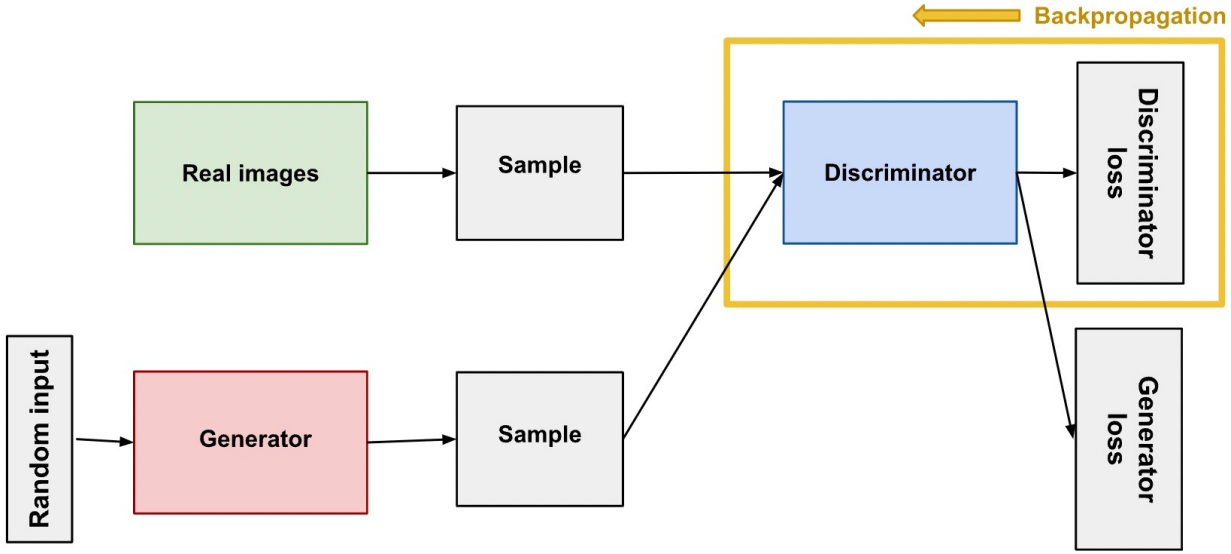


Figure 1: Backpropagation in discriminator training.

Summary

- Masing-masing sebutkan konsep utama yang diingat

Tuhan Memberkati

God's People for God's Glory
CALVIN
INSTITUTE OF TECHNOLOGY