

# Dasar-Dasar Capsule Networks (CapsNet)

Materi diambil dari

Suyanto, "Deep Learning: Modernisasi Machine Learning untuk Big Data", Penerbit Informatika, Bandung, 2019. Bab 4 Capsule Network

Aurélien Géron. Capsule Networks (CapsNets) – Tutorial

# Kerangka Bahasan

1. Pengantar
2. Ide Dasar dan Motivasi CapsNet
3. Arsitektur CapsNet
4. Algoritma Pembelajaran Routing by Agreement

# Pengantar

- CapsNet merupakan pengembangan dari Convolutional Neural Network (disingkat ConvNet atau CNN).
- CapsNet memperbaiki performansi CNN yang memiliki kelemahan dalam mengenali relasi spasial antar komponen spesifik pada objek-objek di dalam citra.

# Kerangka Bahasan

1. Pengantar
2. Ide Dasar dan Motivasi CapsNet
3. Arsitektur CapsNet
4. Algoritma Pembelajaran Routing by Agreement

# Proses Hirarkis

- Pengelihatan manusia memiliki kemampuan untuk memproses citra beresolusi tinggi dengan fitur berdimensi rendah.
- Secara implisit, proses penglihatan manusia bekerja secara hirarkis.
- Terdapat beberapa lapis dalam sistem penglihatan manusia yang memiliki fitur dalam level yang berbeda-beda.
- Setiap lapis merangkum fitur dari lapis sebelumnya sehingga memiliki fitur yang berdimensi lebih rendah.
- Hal ini dalam CNN dimungkinkan dengan adanya pooling layer

# Kelemahan CNN

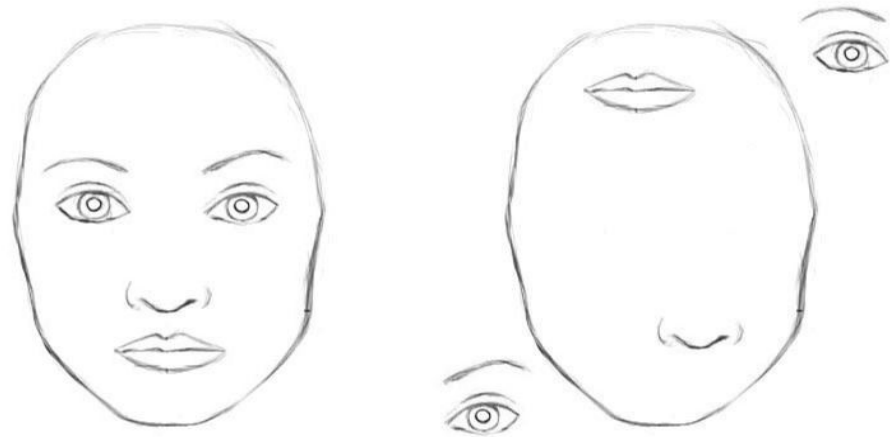
- Hal ini membuat semakin dalam jaringan CNN, semakin kecil fitur yang dihasilkan.
- Misalkan kita memiliki CNN dengan beberapa layer.
- Convolutional layer pertama mengekstrak fitur yang sederhana seperti garis dan tepi dari objek.
- Convolutional layer kedua mengekstrak fitur yang lebih kompleks seperti bentuk-bentuk kecil dari objek.
- Convolutional layer ketiga dan seterusnya mengekstrak fitur yang lebih kompleks lagi sampai pada layer terakhir mengekstrak fitur untuk satu objek utuh.

## Kelemahan CNN (cont)

- Antar setiap convolutional layer diberikan pooling layer yang bertugas untuk merangkum fitur pada area tertentu (misalkan  $2 \times 2$ ) menjadi fitur yang lebih kompleks dengan dimensi yang lebih kecil.
- Hal ini tentu akan mereduksi kompleksitas pemrosesan pada layer-layer berikutnya.
- Namun, hal ini juga memberikan kelemahan pada CNN. Karena adanya proses perangkuman fitur tersebut, pooling layer mereduksi informasi spasial dari fitur tersebut.

## Kelemahan CNN (cont)

- Kedua gambar memiliki fitur yang lengkap dari entitas wajah: dua mata, dua alis, satu hidung, dan satu mulut.



- Namun, fitur kedua gambar memiliki orientasi yang berbeda



## Kelemahan CNN (cont)

- Mata manusia dapat dengan mudah mengenali bahwa citra sebelah kiri merupakan wajah manusia dan citra sebelah kanan bukan wajah manusia.
- Hal ini disebabkan mata manusia dapat mengenali informasi spasial dari komponen-komponen yang ada pada citra.
- Pooling layer mereduksi informasi spasial tersebut.
- Sehingga, pada proses pengenalan citra wajah, CNN mengenali wajah dari fitur-fitur komponen yang ada wajah tersebut, seperti keberadaan mata kiri dan kanan, garis wajah, bibir dan hidung.
- Namun, CNN tidak dapat mengenali informasi spasial yang ada pada komponen-komponen wajah karena pooling layer mereduksi informasi tersebut.

## Kelemahan CNN (cont)

- Sehingga bagi CNN, posisi dari setiap komponen-komponen tersebut tidak penting.
- Akibatnya, citra sebelah kanan juga akan dianggap sebagai citra wajah.

# Kerangka Bahasan

1. Pengantar
2. Ide Dasar dan Motivasi CapsNet
3. **Arsitektur CapsNet**
4. Algoritma Pembelajaran Routing by Agreement

# Capsule

- Berbeda dengan arsitektur dari deep neural network pada umumnya yang terdiri dari banyak layer, pada CapsNet kita menambahkan layer-layer tersebut di dalam sebuah layer.
- Atau dapat dikatakan, bahwa CapsNet memiliki layer yang bersarang. Satu capsule merupakan satu set layer bersarang.
- Setiap capsule mengeluarkan output berupa vektor yang merepresentasikan nilai probabilitas dari keberadaan sebuah entitas di dalam citra dan orientasi dari vektor merepresentasikan properti dari entitas tersebut.

# Routing

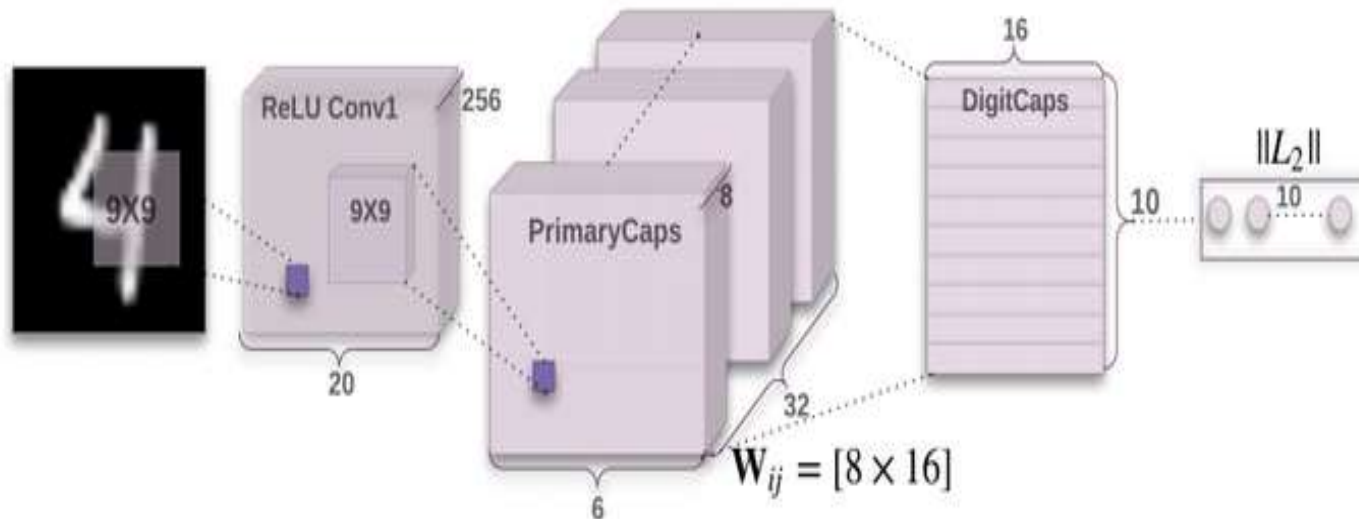
- Vektor dari sebuah capsule dikirimkan ke setiap capsule lain yang berperan sebagai parent dari capsule tersebut.
- Untuk setiap parent dari sebuah capsule dapat dihitung sebuah vektor prediksi dari perkalian antara bobot capsule tersebut dengan sebuah matriks bobot.
- Parent yang memiliki hasil perkalian tertinggi, maka ikatannya dengan capsule tersebut akan diperkuat, sedangkan ikatan capsule dengan parent lain akan diperlemah.
- Skema routing ini lebih baik daripada mekanisme pooling yang ada pada CNN yang hanya meneruskan informasi yang dianggap paling “kuat”.
- Di dalam CapsNet juga terdapat mekanisme squashing yang merupakan fungsi non-linear.

## Routing (cont)

- Fungsi squashing ini dilakukan terhadap vektor output dari setiap capsule.
- Informasi dari posisi setiap entitas pada capsule level rendah dapat dilihat dari capsule mana saja yang aktif.
- Sedangkan informasi posisi pada capsule level lebih tinggi dilihat dari nilai vektor output dari capsule tersebut.

# Arsitektur CapsNet

- Contoh sebuah arsitektur sederhana dari CapsNet 3 layer, yg digunakan untuk mengenali data angka tulisan tangan 0 hingga 9 menggunakan data MNIST.



## Arsitektur CapsNet (cont)

- Contoh CapsNet tsb hanya terdiri dari dua convolutional layer dan satu fully connected layer.
- Convolutional layer pertama adalah Conv1 yang memiliki kernel berukuran  $9 \times 9$  dengan stride 1 dan fungsi aktivasi ReLU.
- Layer ini mengubah nilai intensitas pixel menjadi aktivitas dari fitur lokal yang digunakan sebagai input dari primary capsule.

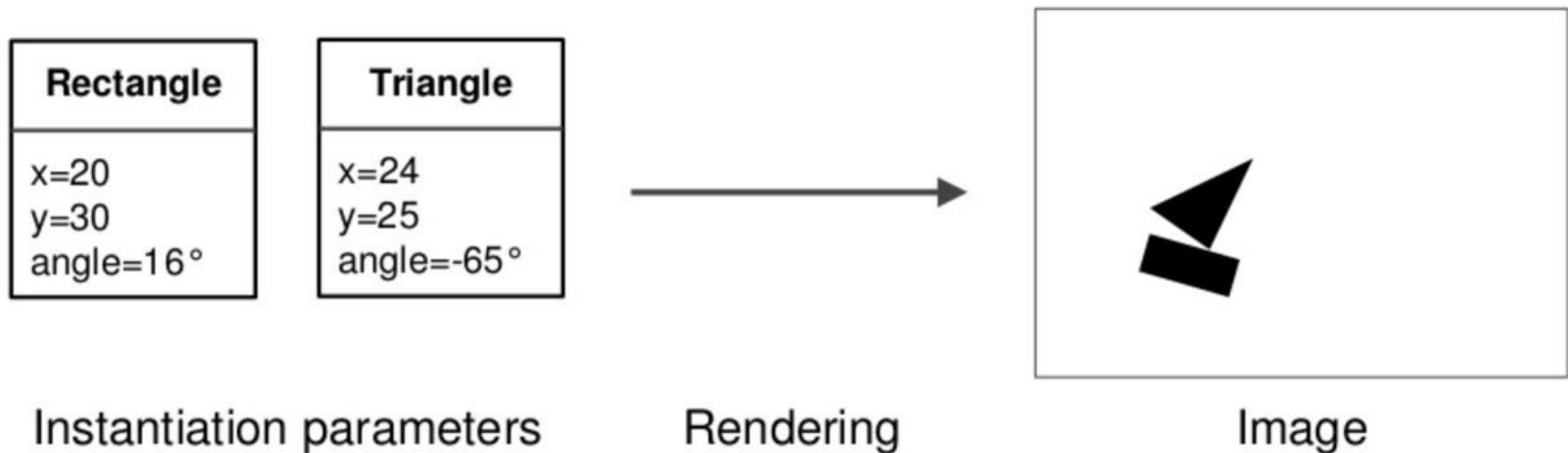


# Graphics

- Primary capsule merupakan level terendah dari entitas multidimensi dan dari sudut pandang inverse graphics, mengaktifkan primary capsule identik dengan membalik proses rendering pada citra.
- Pada proses rendering objek, kita mendefinisikan pose dari objek seperti posisi, ukuran dan orientasi dan membangun objek dari informasi tersebut.
- Inverse rendering adalah kebalikannya.
- Jika diberikan sebuah objek, kita mengekstrak informasi pose tersebut dari objek yang diberikan.

# Graphics (cont)

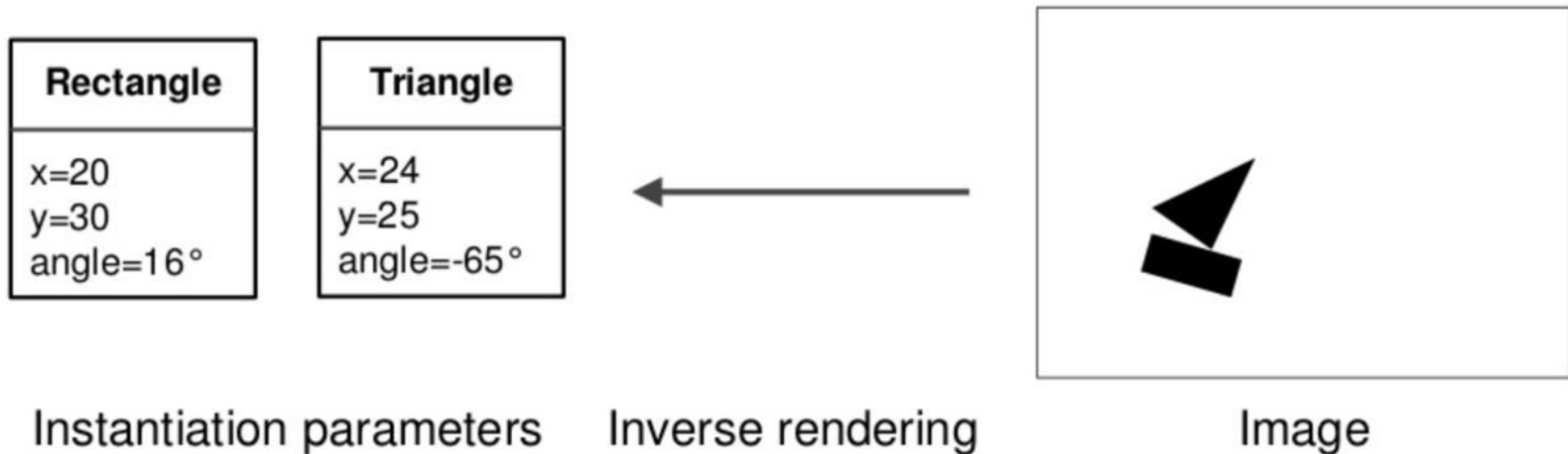
- Contoh sebuah proses rendering pada computer graphics



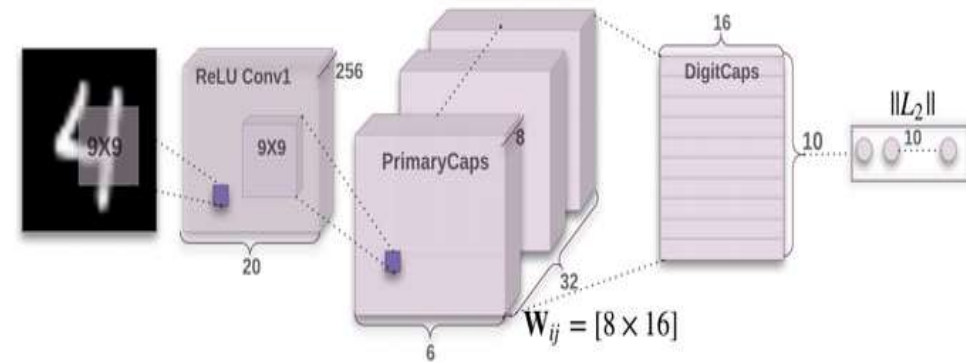
Aurélien Geron. Capsule Networks (CapsNets) – Tutorial

# Graphics (cont)

- Contoh sebuah proses inverse rendering



# Arsitektur CapsNet



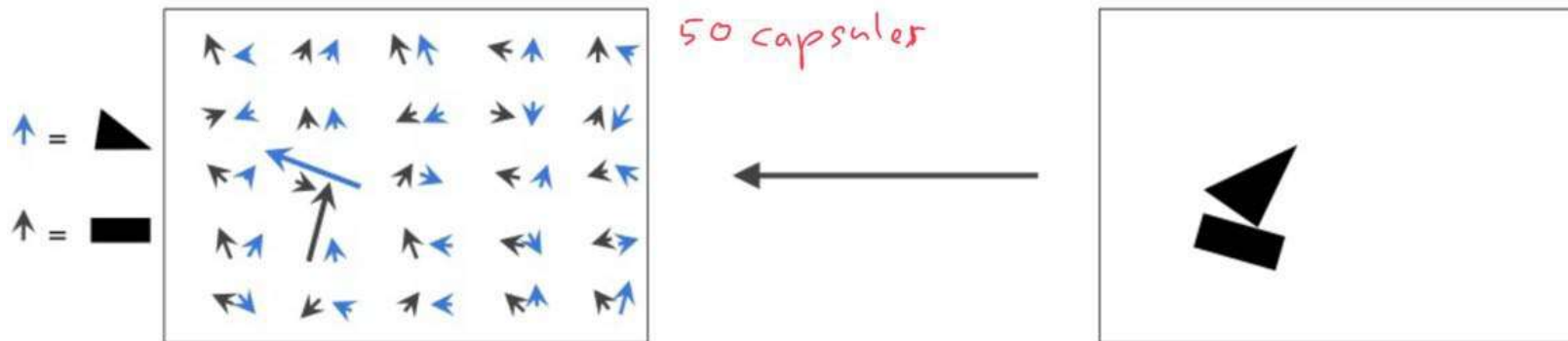
- Primary capsule layer tsb merupakan convolutional capsule layer yang mengandung 32 channel convolutional 8D capsule, dengan setiap primary capsule mengandung 8 unit konvolusi .
- Setiap primary capsule mengambil input dari Conv1 dengan ukuran 256 x 81 yang overlap dengan lokasi dari bagian tengah dari capsule.
- Primary capsule dapat dianggap sebagai sebuah convolutinal layer dengan fungsi squashing sebagai fungsi aktivasinya.

## Arsitektur CapsNet (cont)

- Layer berikutnya adalah DigitCaps memiliki satu capsule per kelas output label, dalam kasus pengenalan angka jumlah label = 10, dengan vektor output berdimensi 16 dan setiap capsule pada DigitCaps menerima input dari setiap capsule pada layer di bawahnya.

# Activation Vector

- Sebuah capsule mencoba memprediksi keberadaan dan instantiation parameters dari sebuah obyek tertentu di lokasi tertentu.
- Orientation, bisa estimasi pose (misal rotation, ketebalan, tingkat skewed dll). Contoh di sini parameter rotasi.



**Activation vector:**

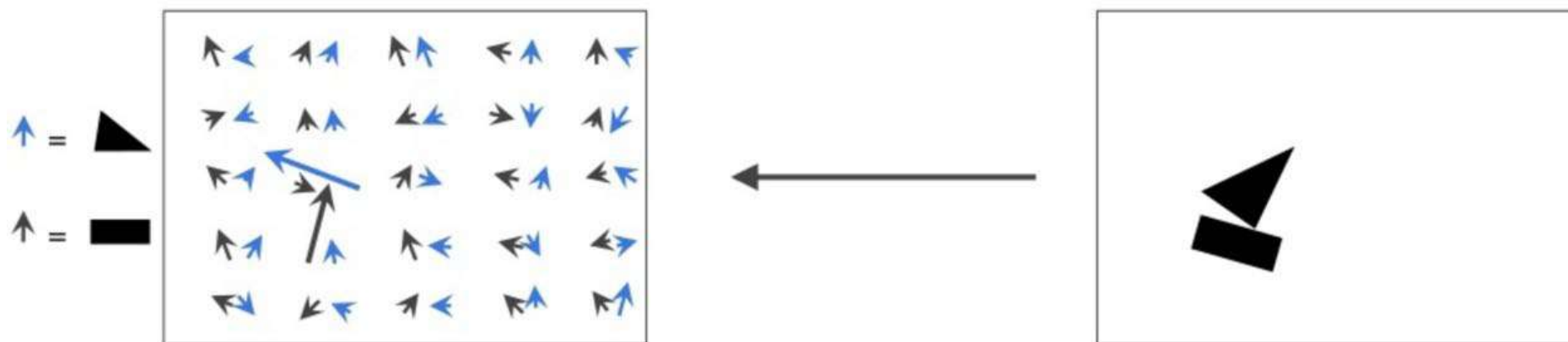
**Length** = estimated probability of presence

**Orientation** = object's estimated pose parameters

## Activation Vector (cont)

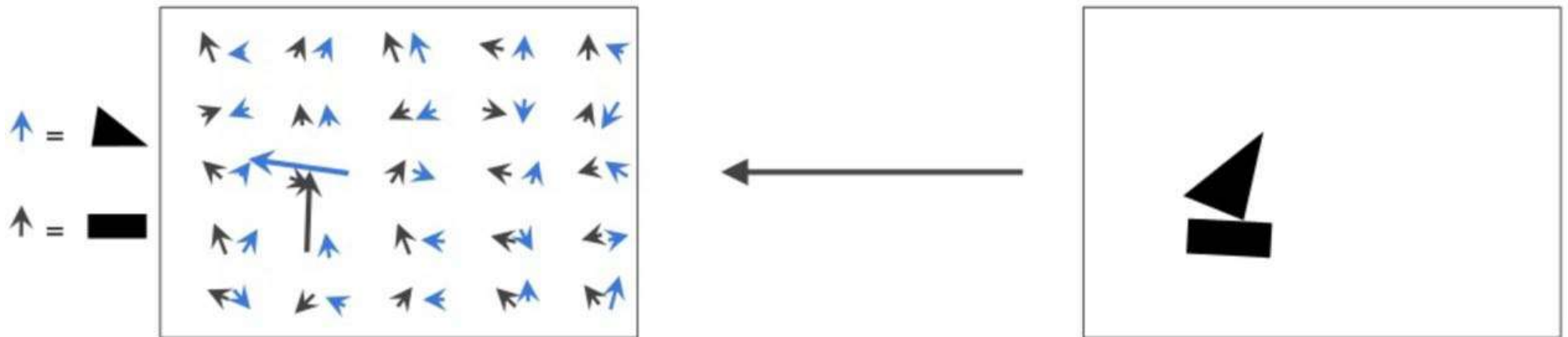
- Capsule bisa berisi lebih banyak dimensi untuk pose, bisa 10 atau bahkan lebih.

# Equivariance





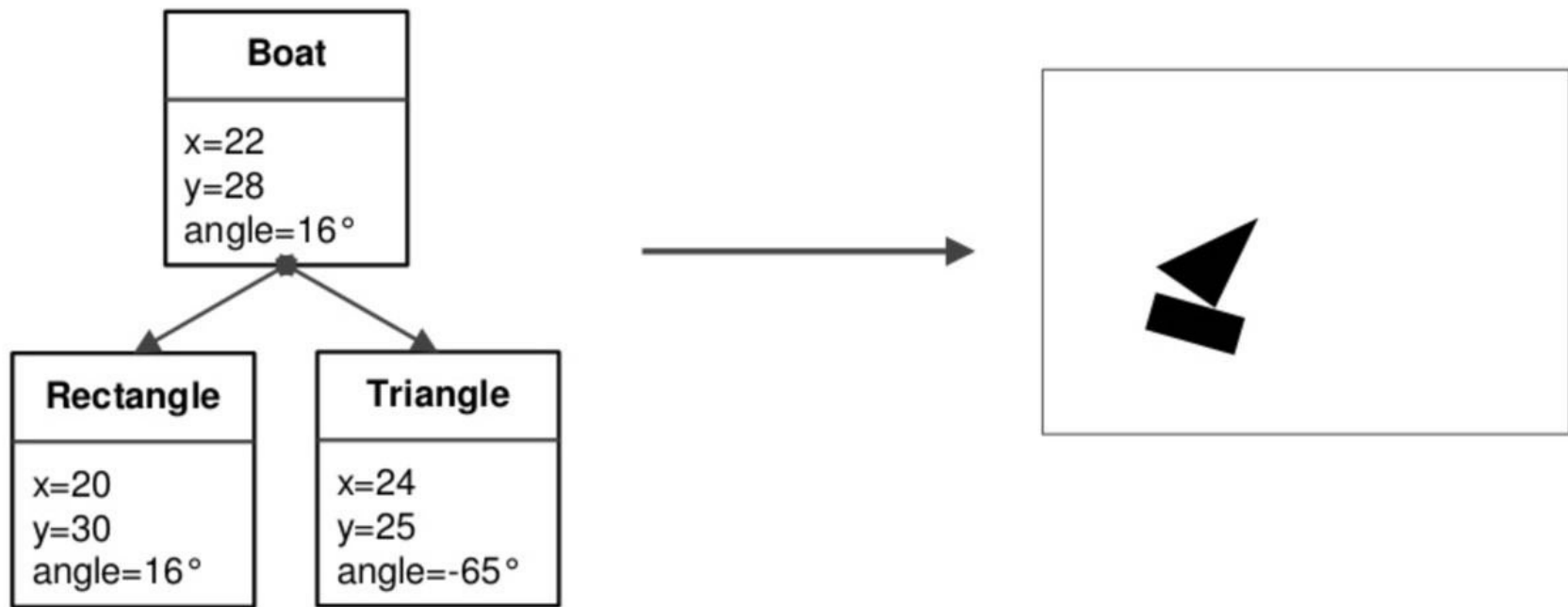
## Equivariance (cont)



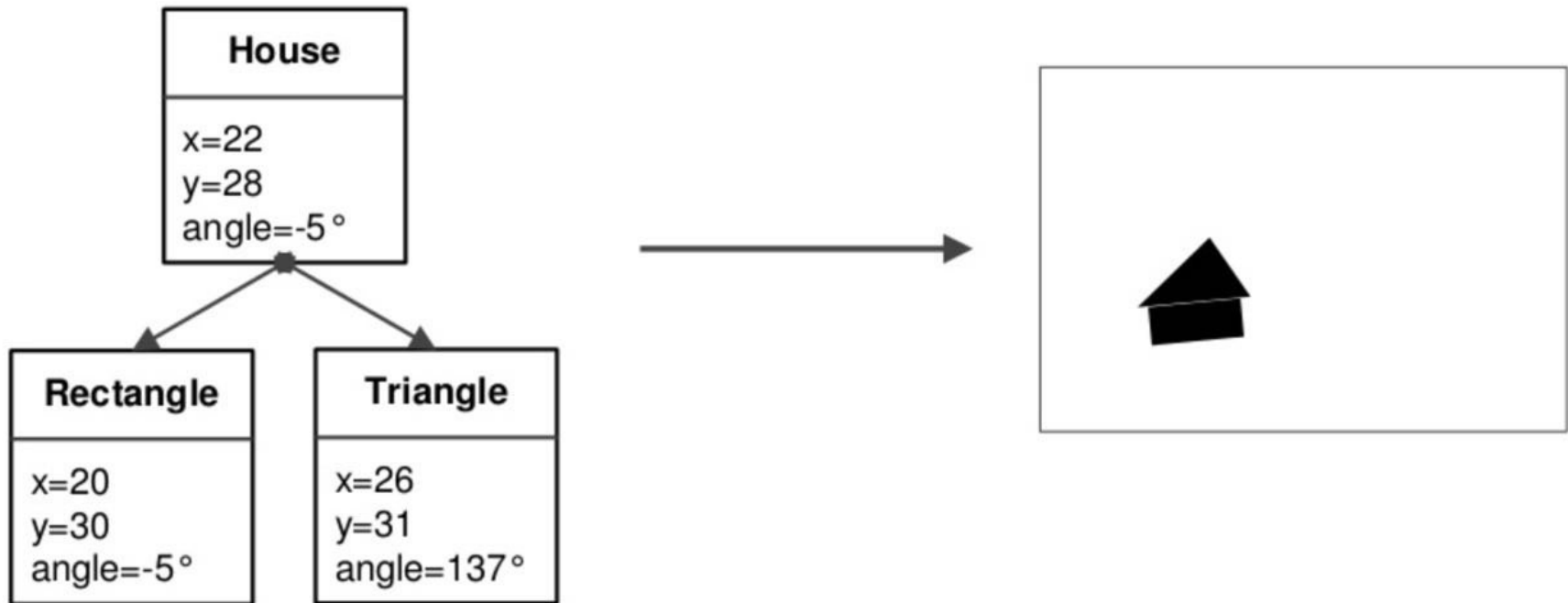
## Equivariance (cont)

- Proses downsample Pooling pada CNN bisa menghilangkan informasi lokasi dan pose dan obyek.
- Informasi lokasi dan pose tsb mungkin tdk menjadi masalah kalau kita hendak melakukan klasifikasi atas gambar secara keseluruhan.
- Namun akan menjadi tantangan bila kita hendak melakukan task yg membutuhkan lokasi dan pose yg akurat, seperti misalnya image segmentasi, atau object detection yg detil.

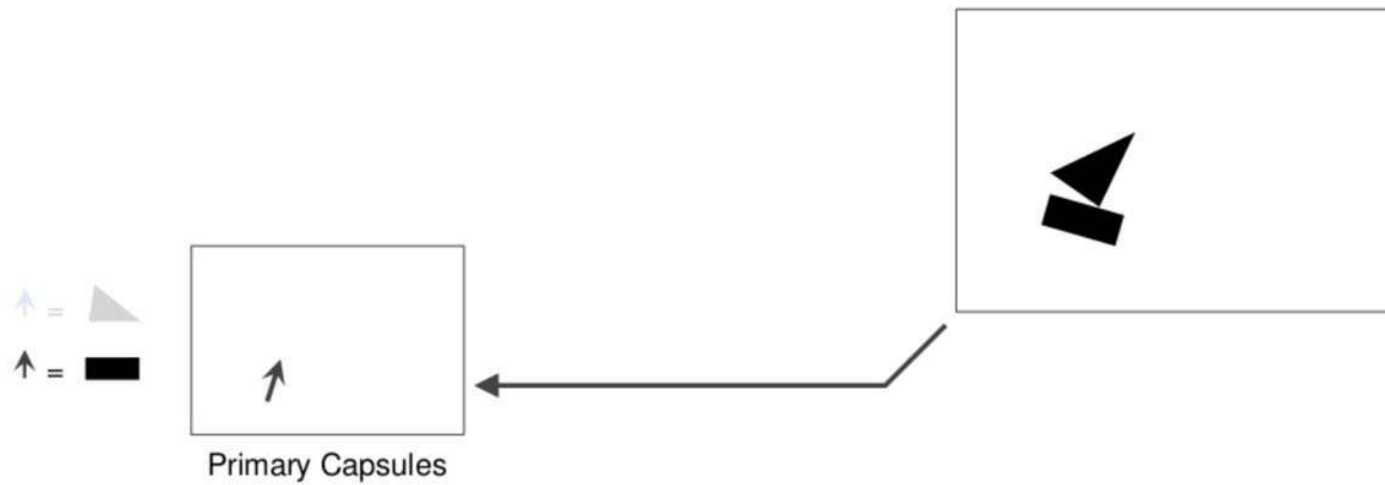
# Struktur Hirarkis atas Bagian2



# Struktur Hirarkis atas Bagian2

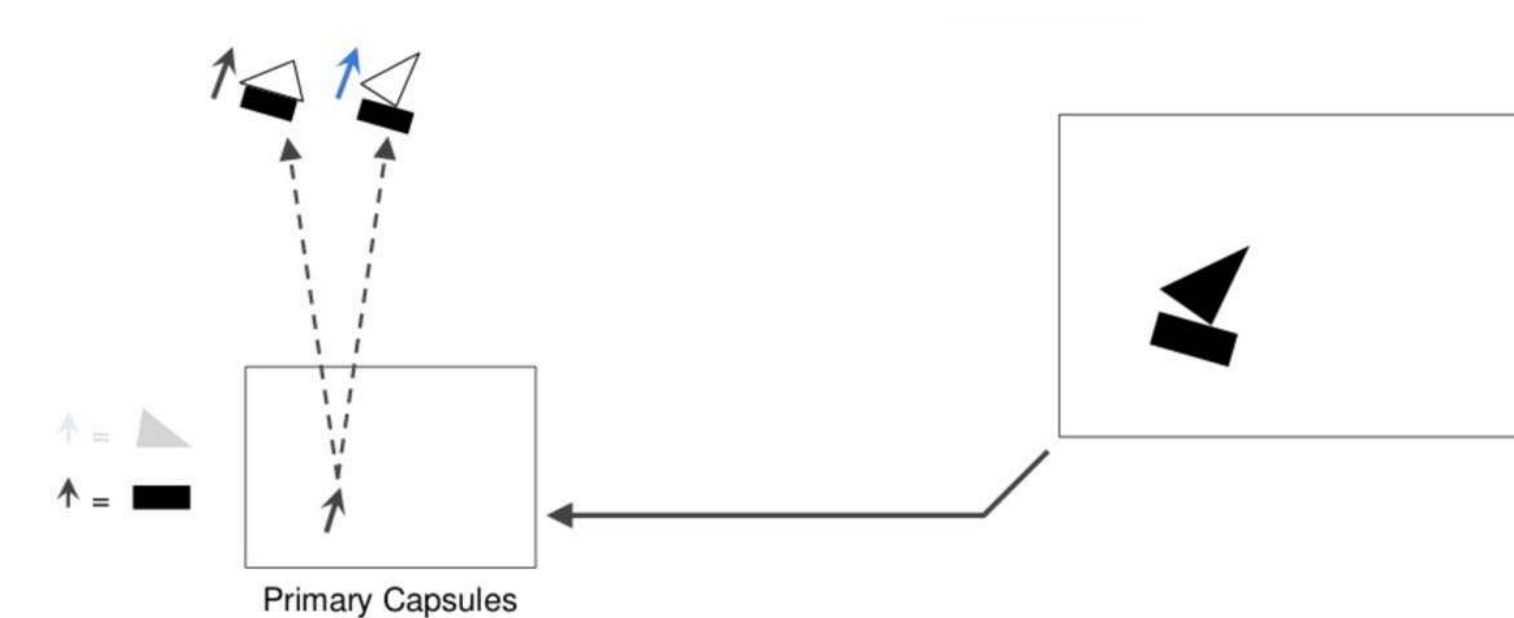


# Primary Capsule

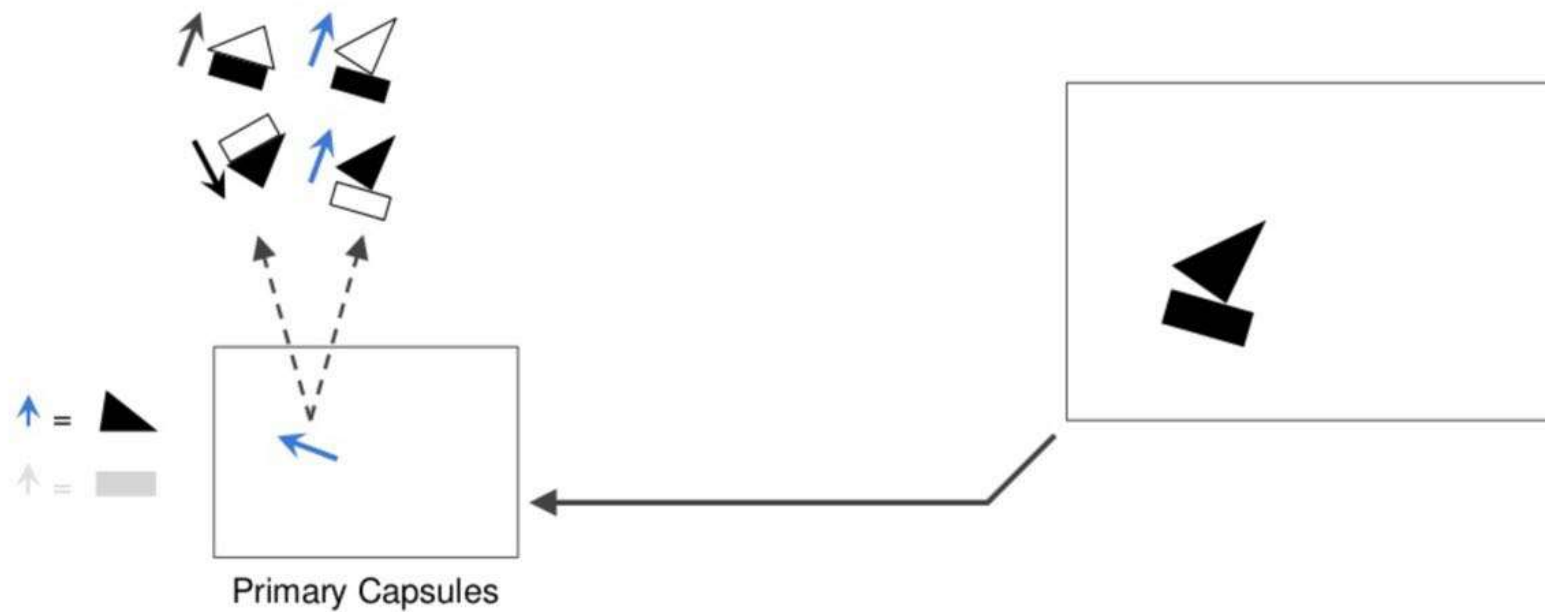


# Memprediksi Output Layer Berikutnya

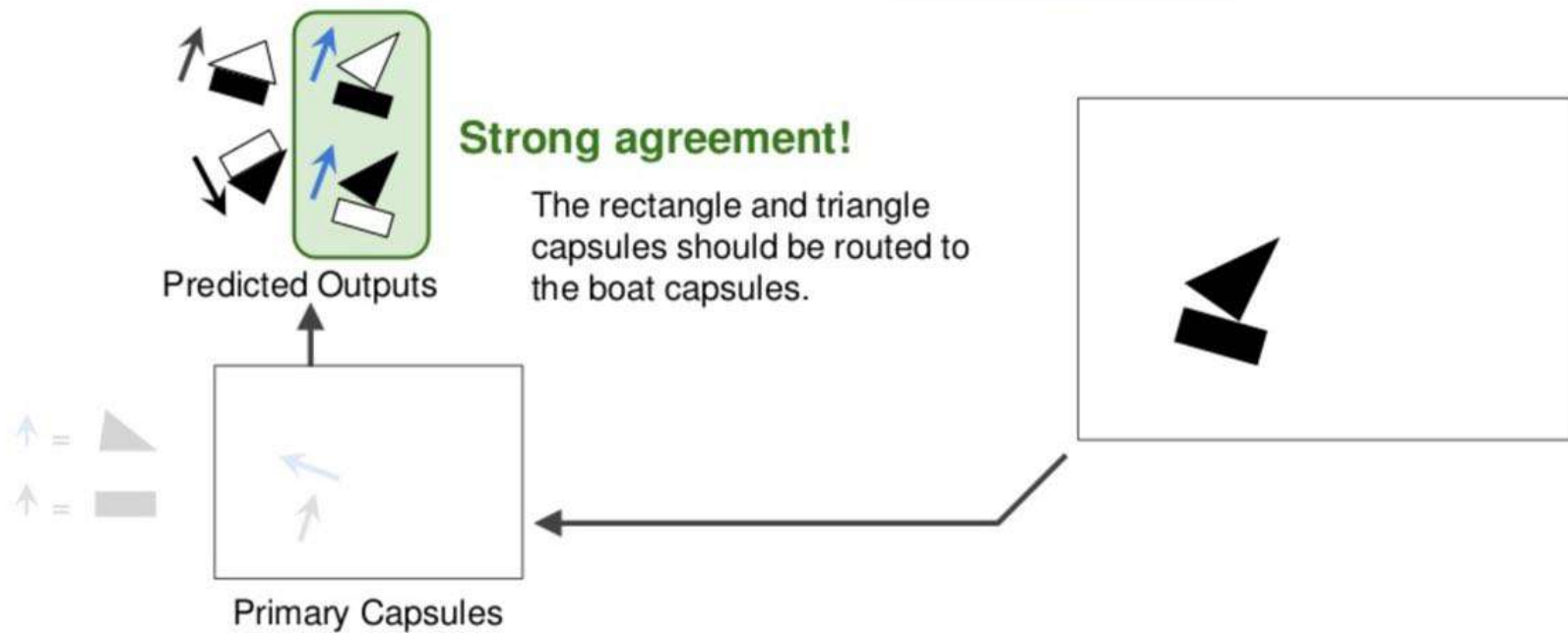
- Misalnya ada dua buah capsul pada layer berikutnya, house capsule dan boat capsule.



## Memprediksi Output Layer Berikutnya (cont)



# Routing by Agreement





Terima kasih

# Daftar Istilah

- Capsule. A capsule is a small group of neurons whose activity vector represents the instantiation parameters of a specific type of entity such as an object or an object part.
- Multi-layer capsule
- Layer bersarang
- Routing