AIL721: Deep Learning

Instructor: James Arambam

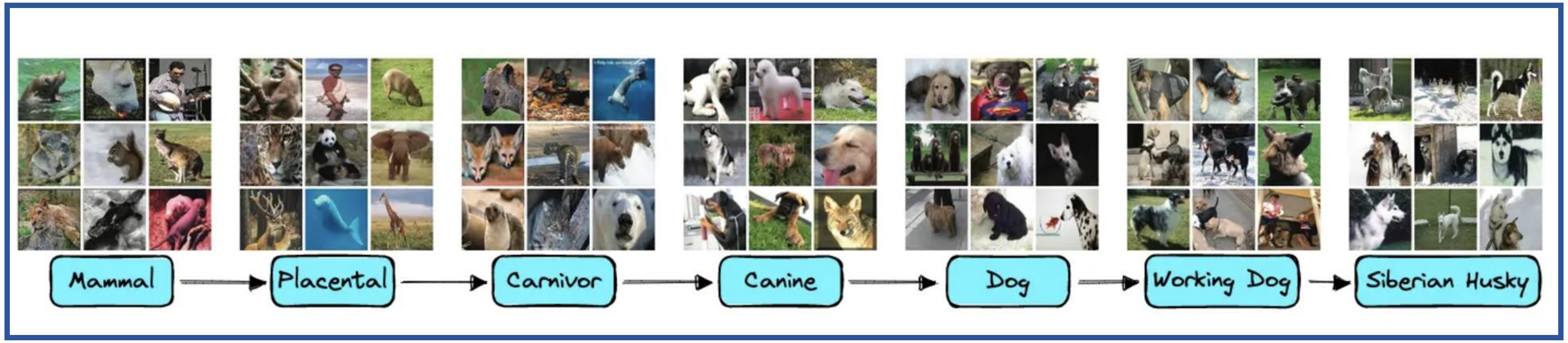
ScAI | Yardi School of Artificial Intelligence
Indian Institute of Technology Delhi

Applications of CNN

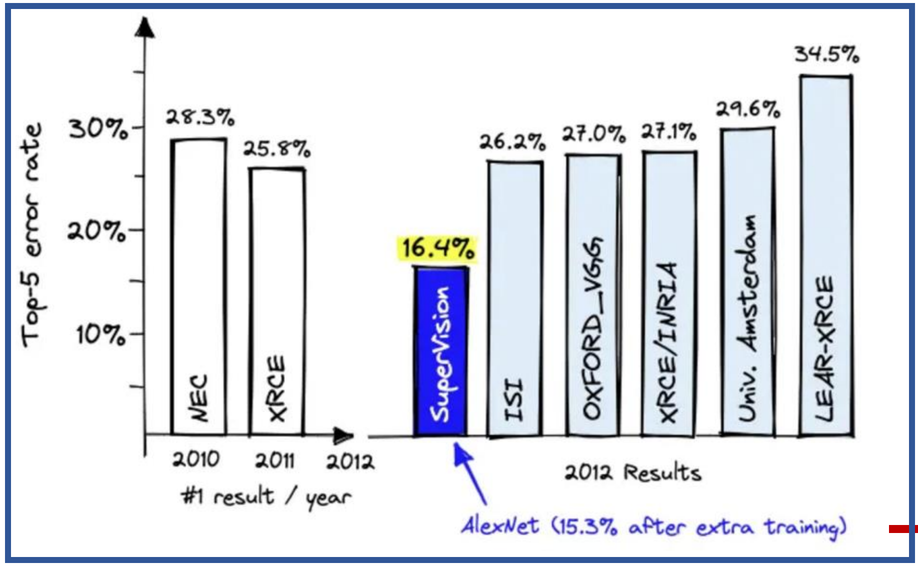
A Brief History of Neural Networks

- **ImageNet Competition**: created a database of more than 14 million hand-annotated images.

Fei-Fei Li's group at Stanford University



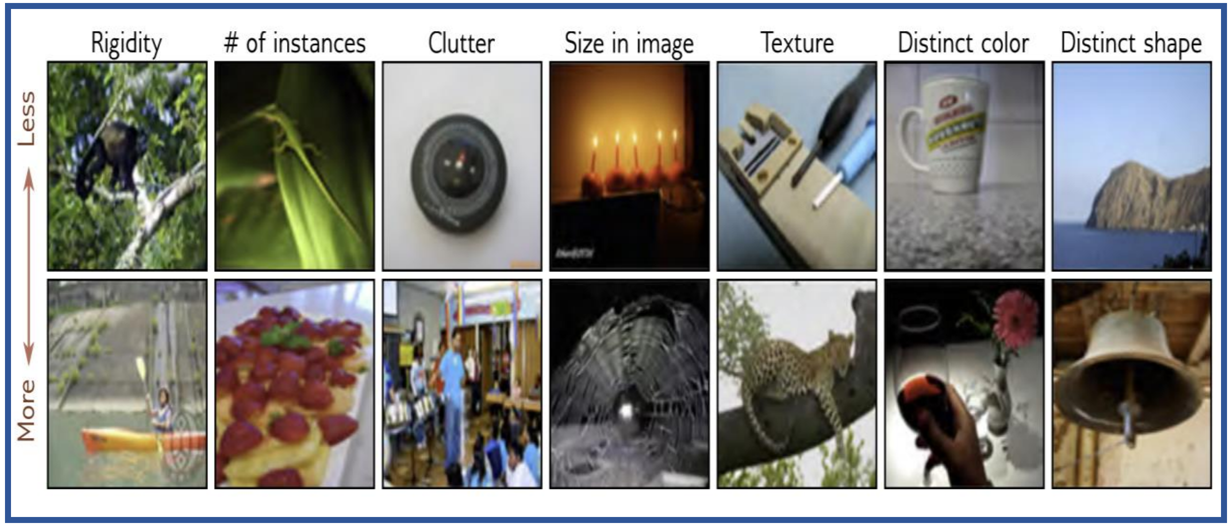
- **AlexNet, 2012**: A convolutional neural network (CNN) based approach by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton.

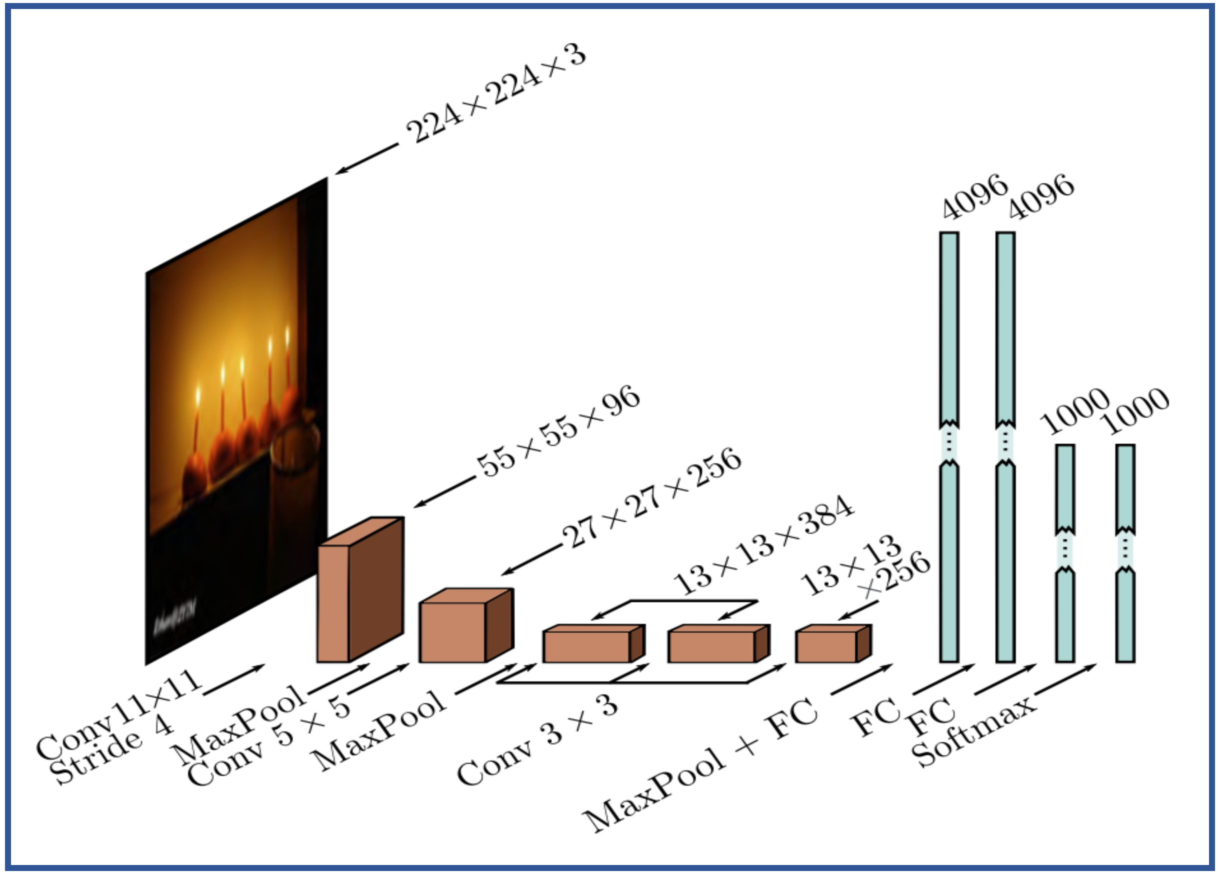


[Images from www.pinecone.io & Wikipedia]

Convolutional Neural Network

❑ ImageNet Classification





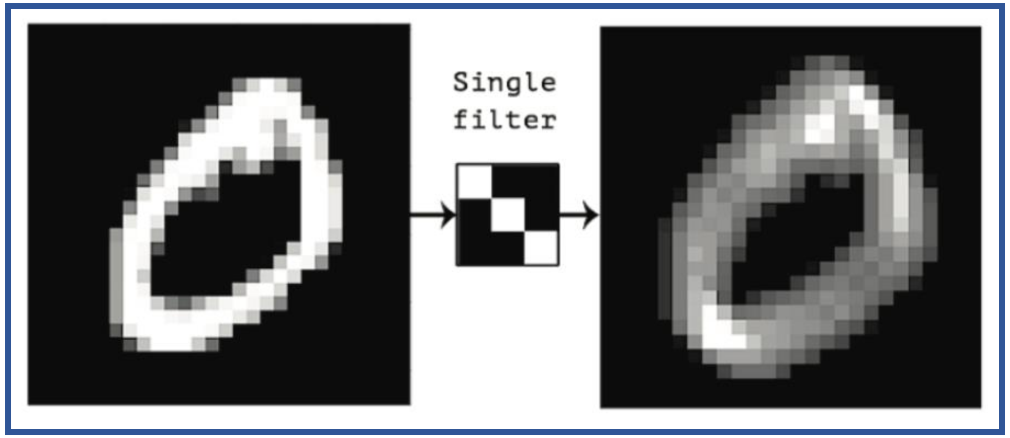
AlexNet

Input: 224x224 RGB image

Output: Probability distribution over 1000 classes.

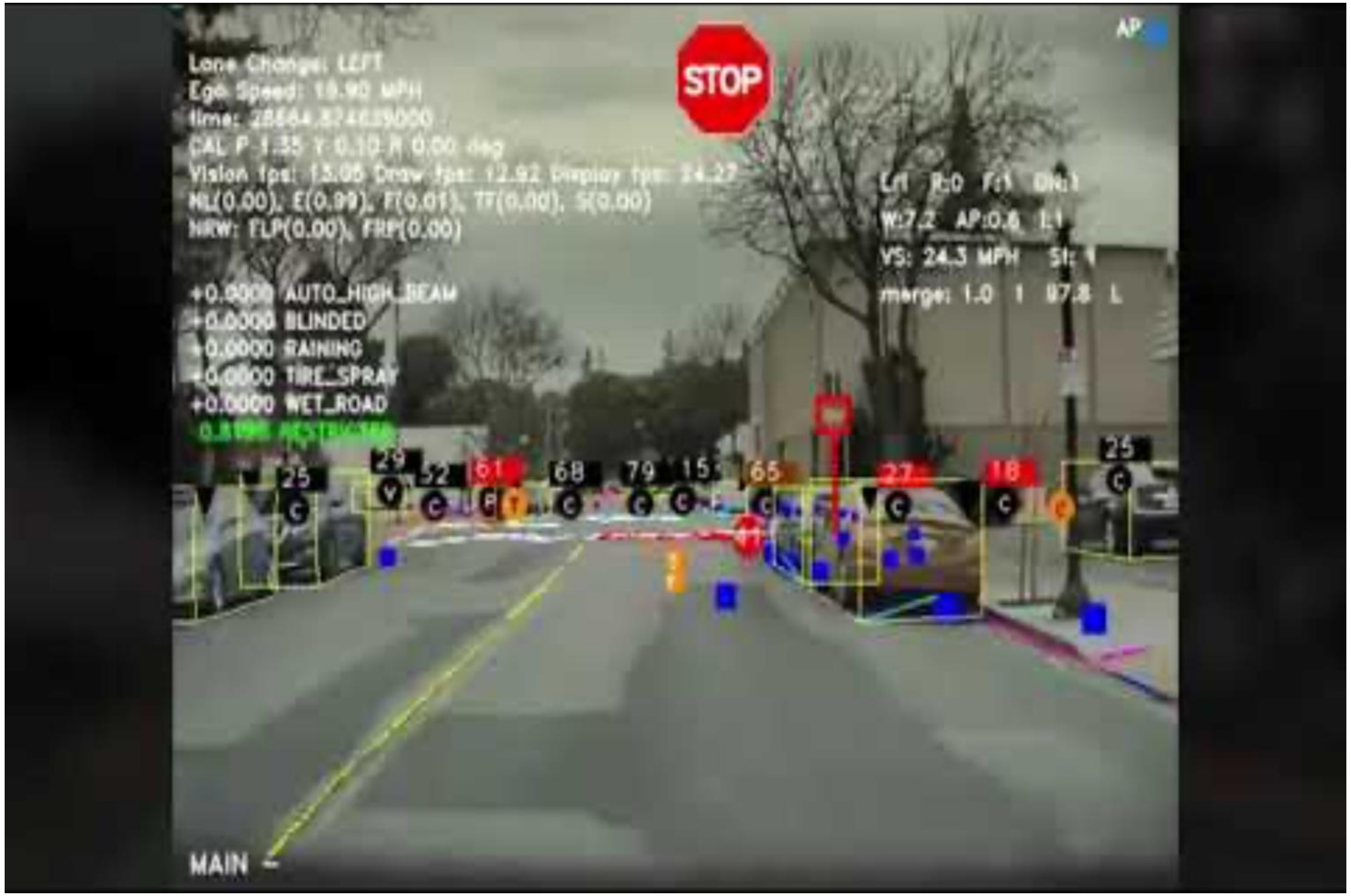
[Image from Prince 2023 UDL Book]

Convolutional Neural Network



[Source: Bishop Book]

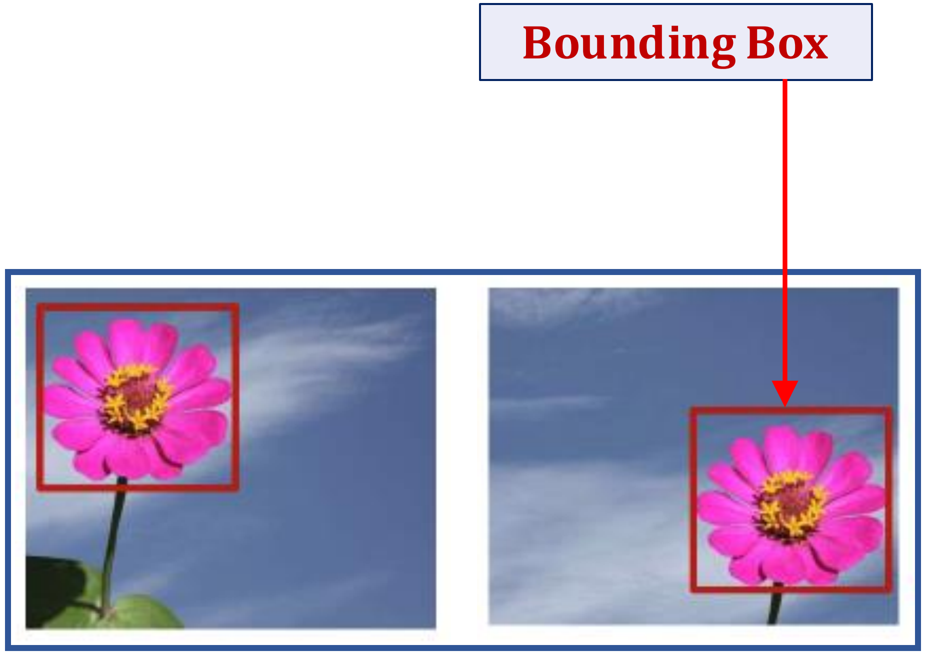
Convolutional Neural Network

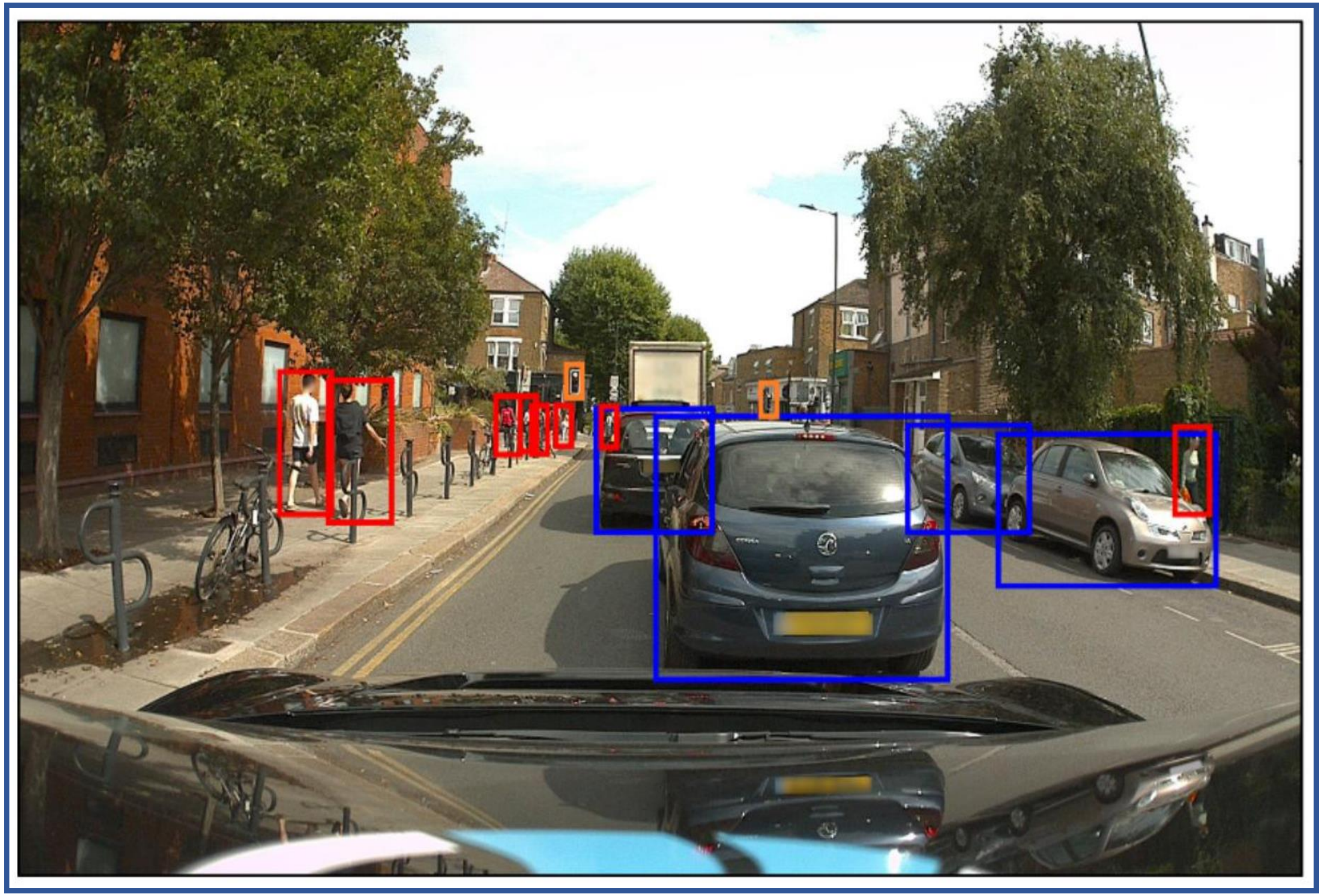


[Video from youtube.com]

Convolutional Neural Network

❑ Object Detection & Localization

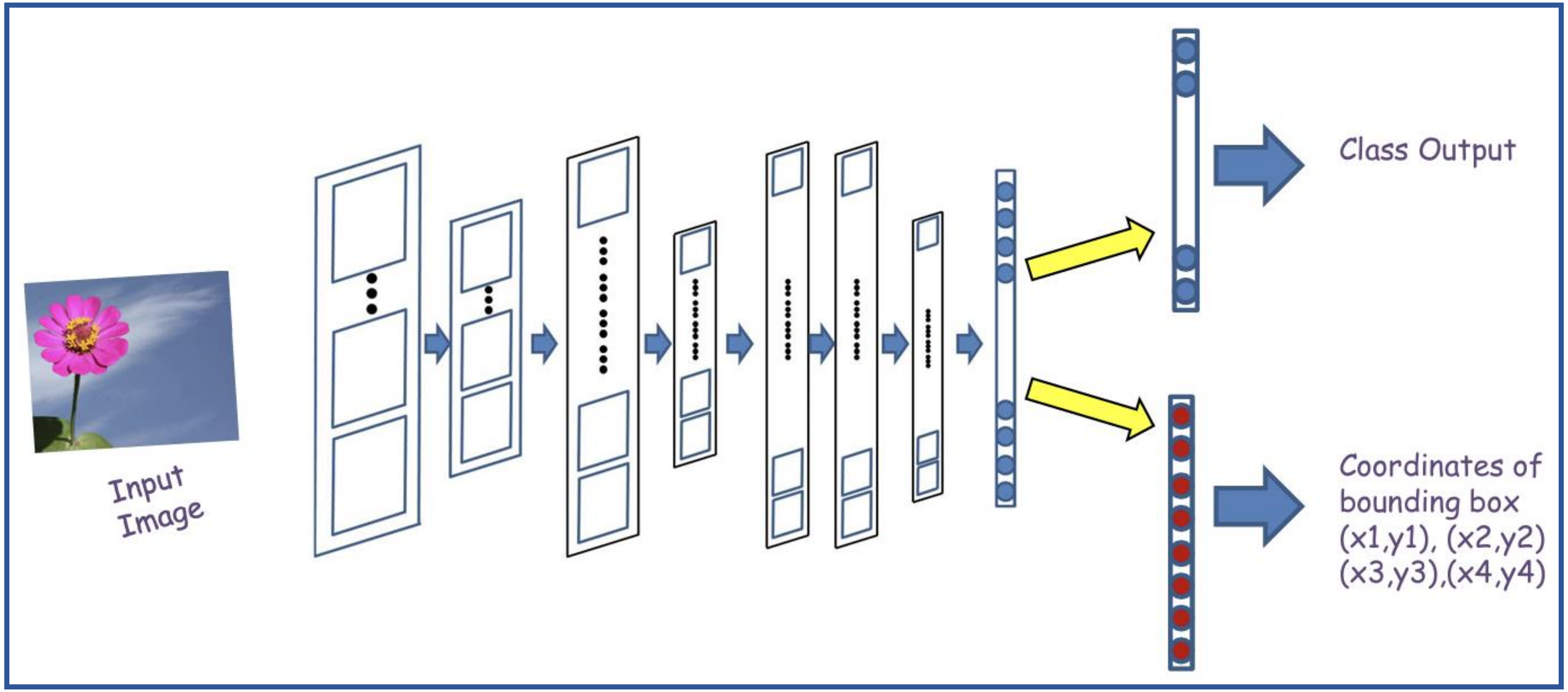




[Image from Bishop 2024 DL Book]

Convolutional Neural Network

❑ Object Detection & Localization



[Examples from Bhiksha et al. CMU DL Course]

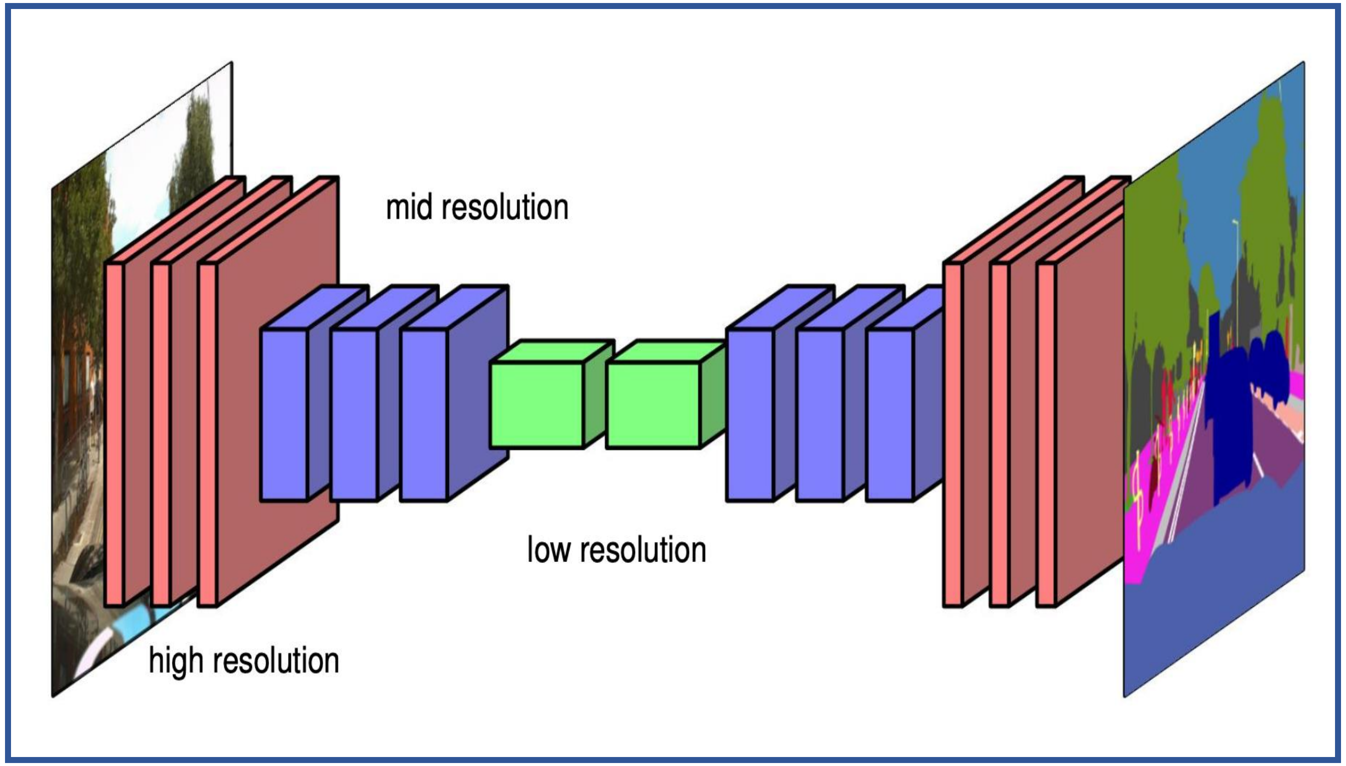
Convolutional Neural Network

❑ Image Segmentation

[Image from Bishop 2024 DL Book]

Convolutional Neural Network

❑ Image Segmentation

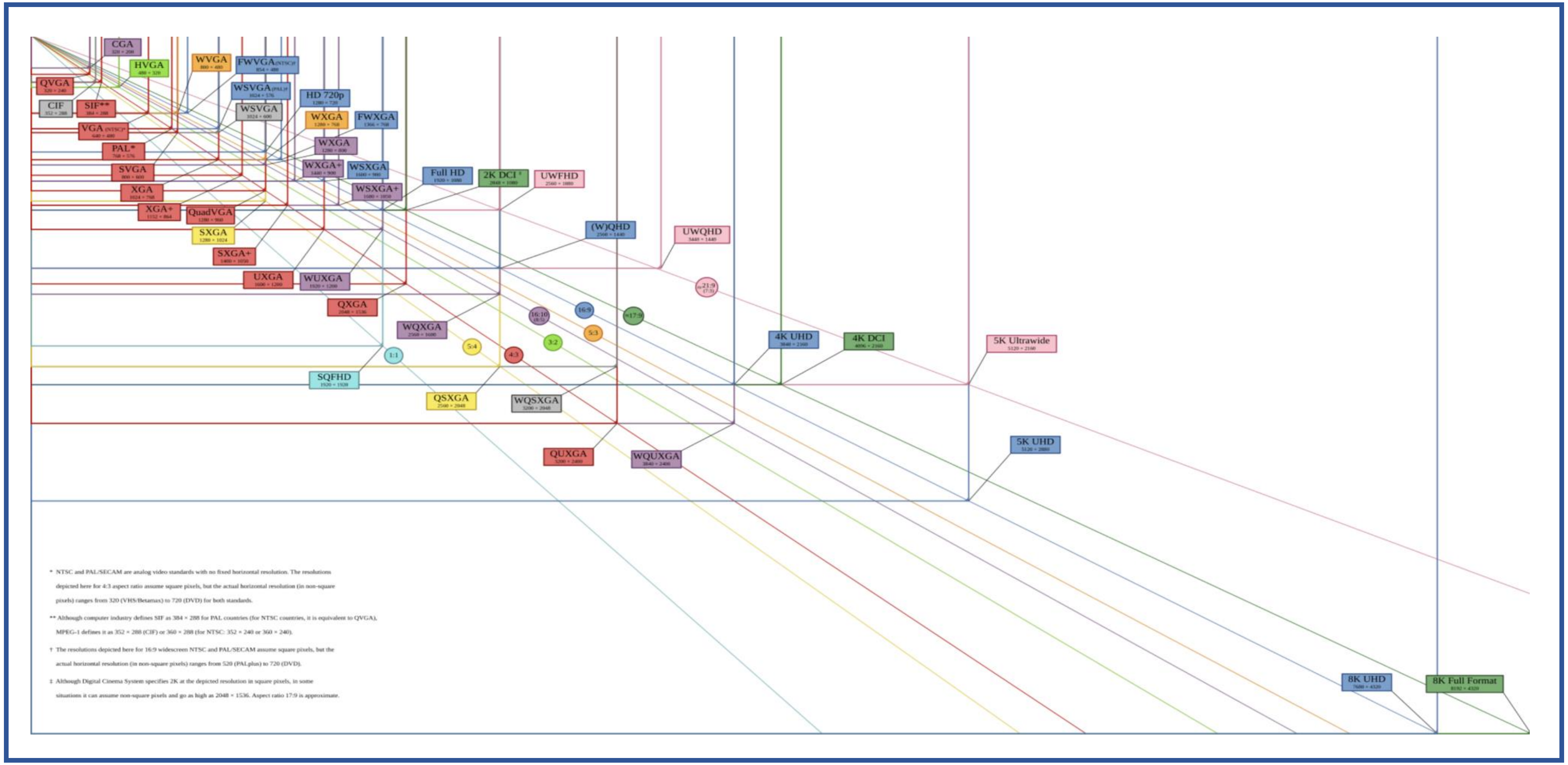


Downsampling **Upsampling**

[Image from Bishop 2024 DL Book]

Convolutional Neural Network

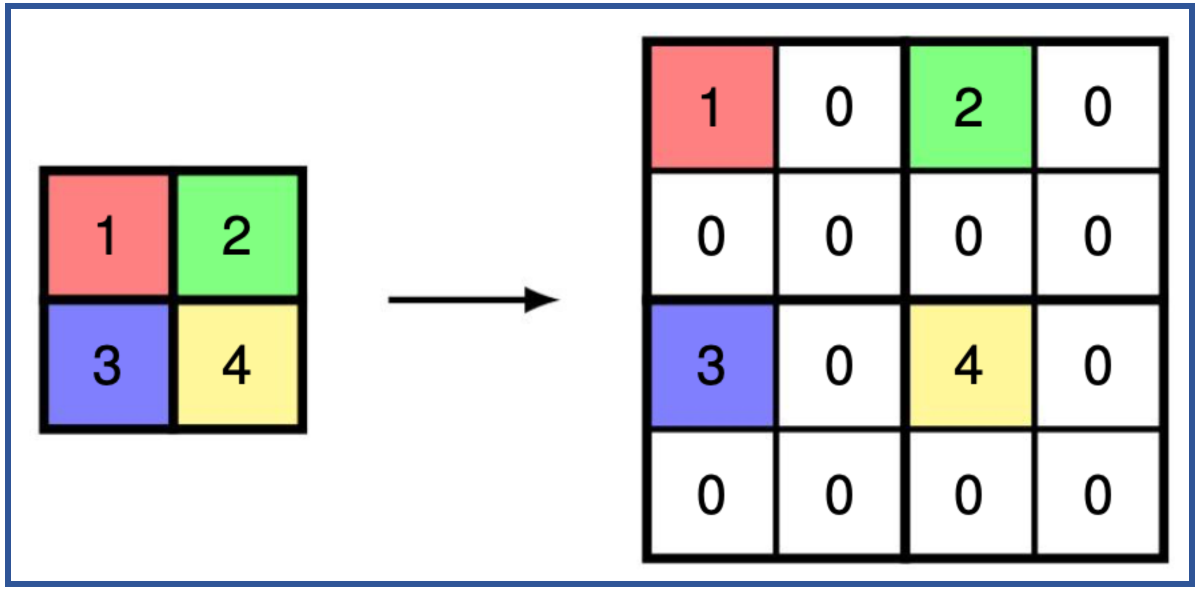
❑ Image Segmentation

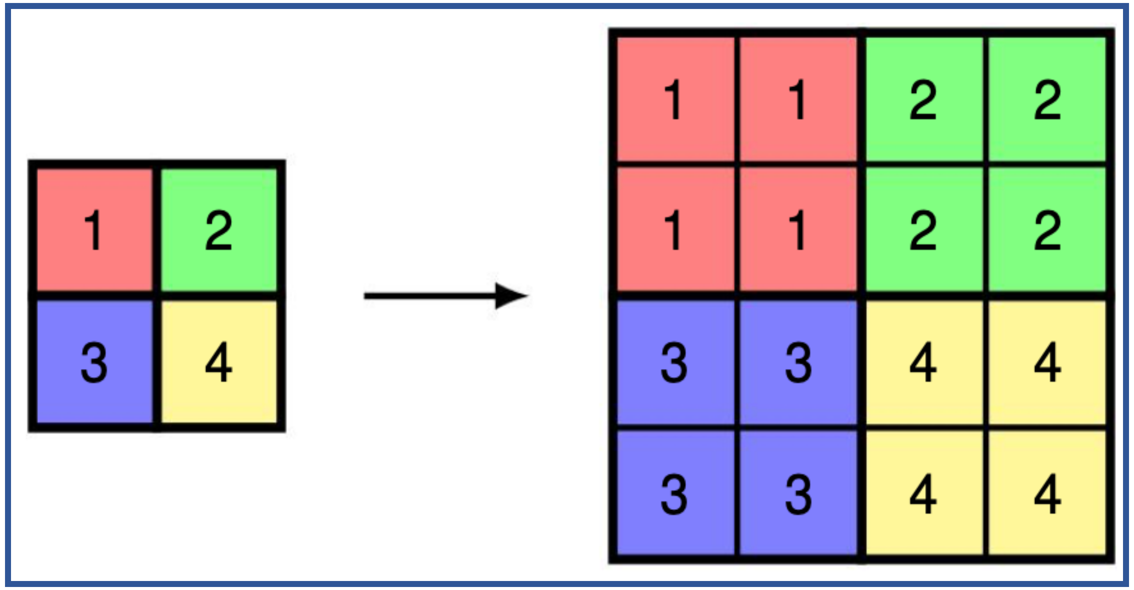


[Image from Bishop 2024 DL Book]

Convolutional Neural Network

❑ Upsampling





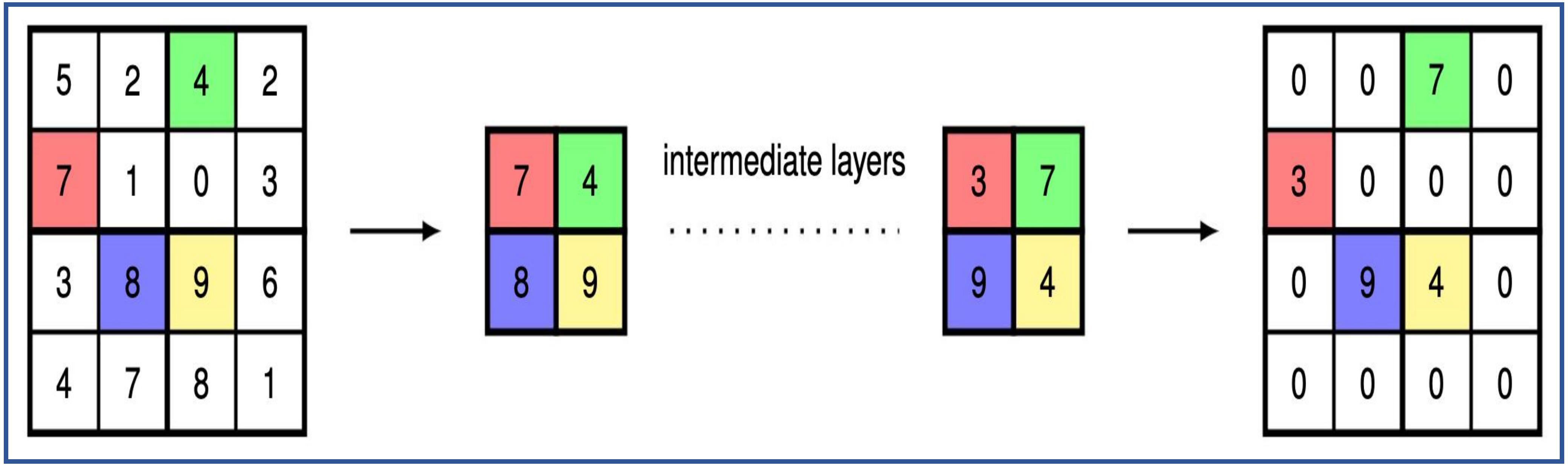
Max Unpooling

Average Unpooling

[Image from Bishop 2024 DL Book]

Convolutional Neural Network

❑ Upsampling

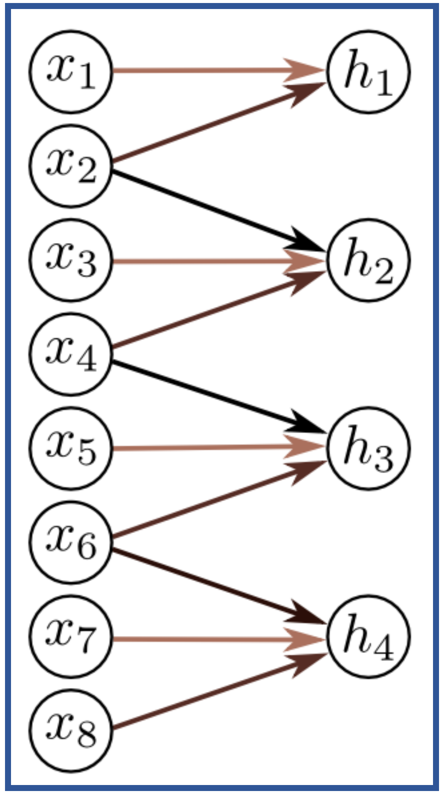


Preserves Spatial Information

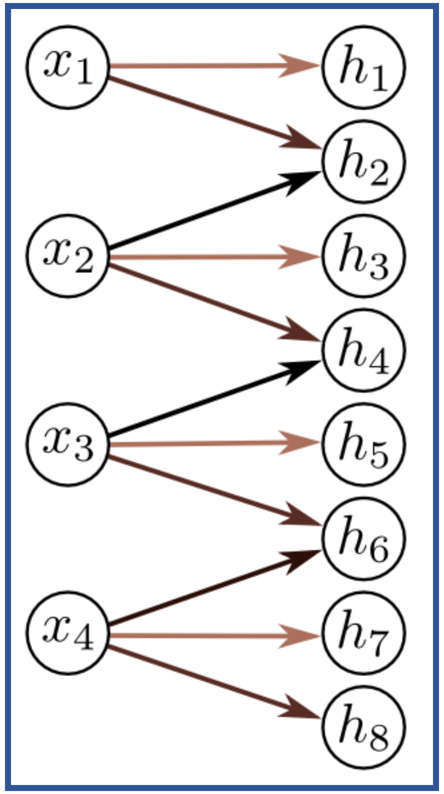
[Image from Bishop 2024 DL Book]

Convolutional Neural Network

❑ Upsampling



Convolution

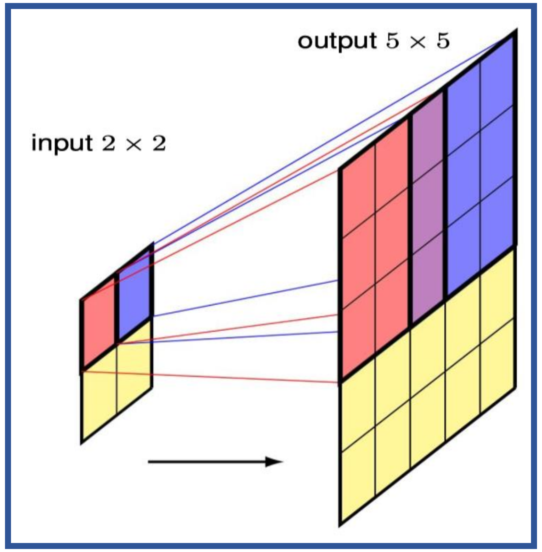


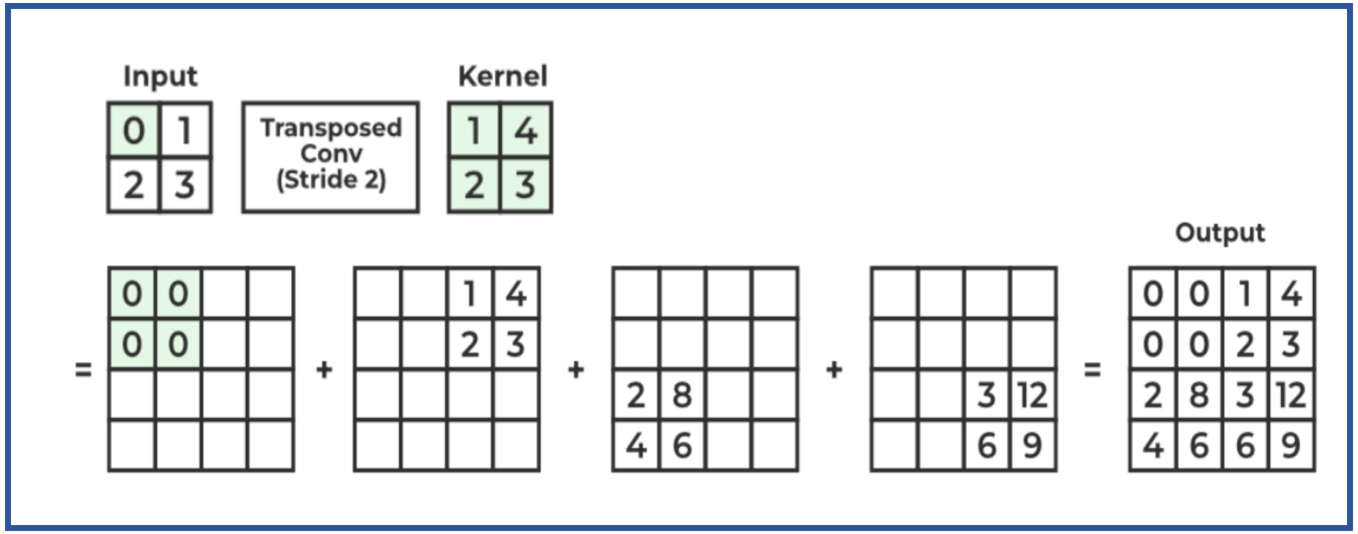
Transposed Convolution

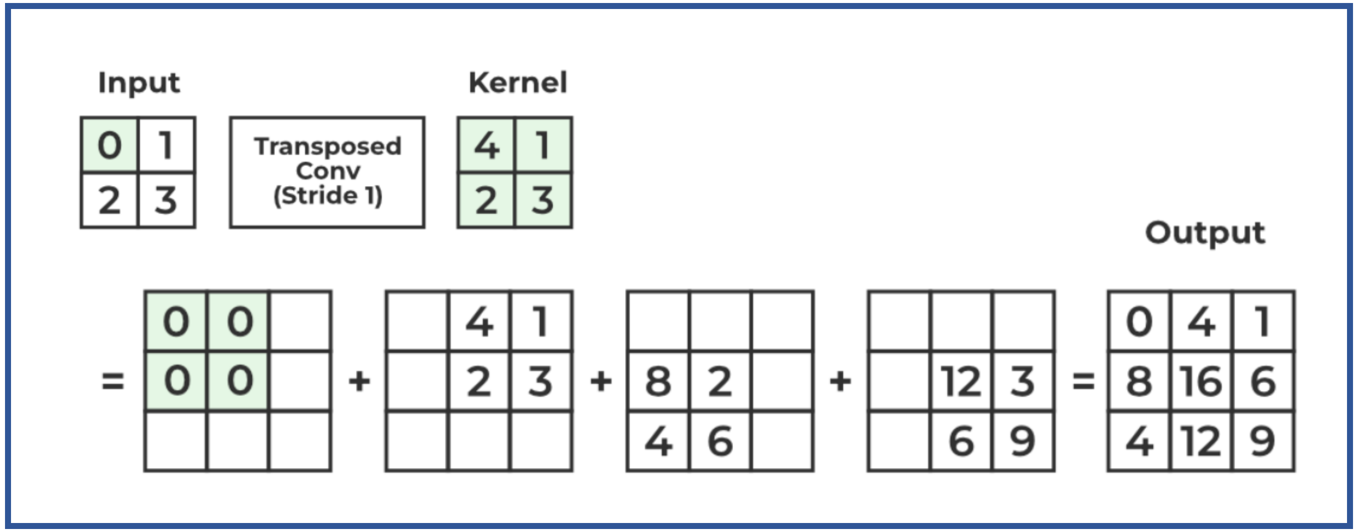
[Image from Prince 2023 UDL]

Convolutional Neural Network

❑ Transposed Convolution



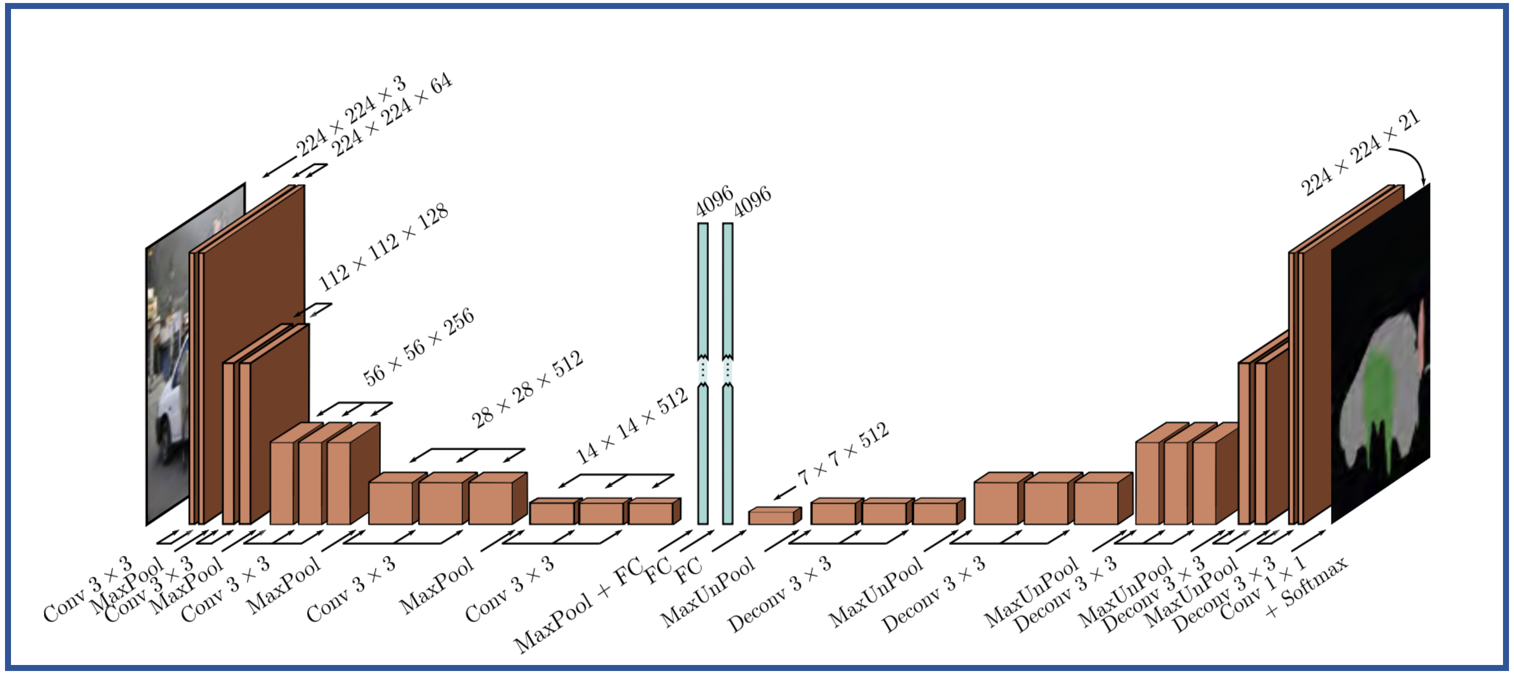


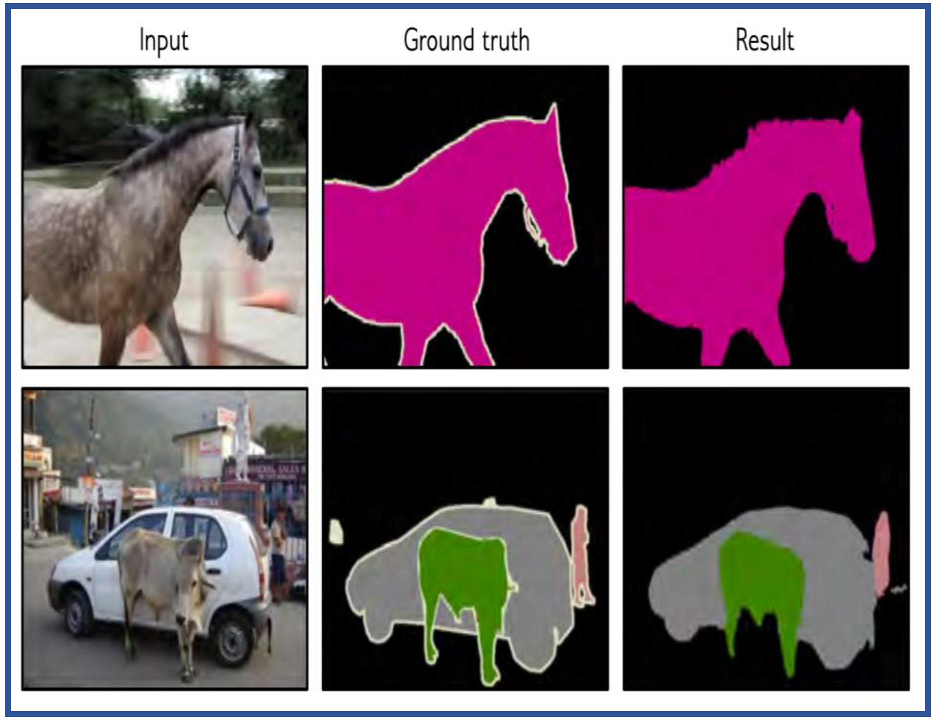


[Image from Prince 2023 UDL]

Convolutional Neural Network

❑ Image Segmentation

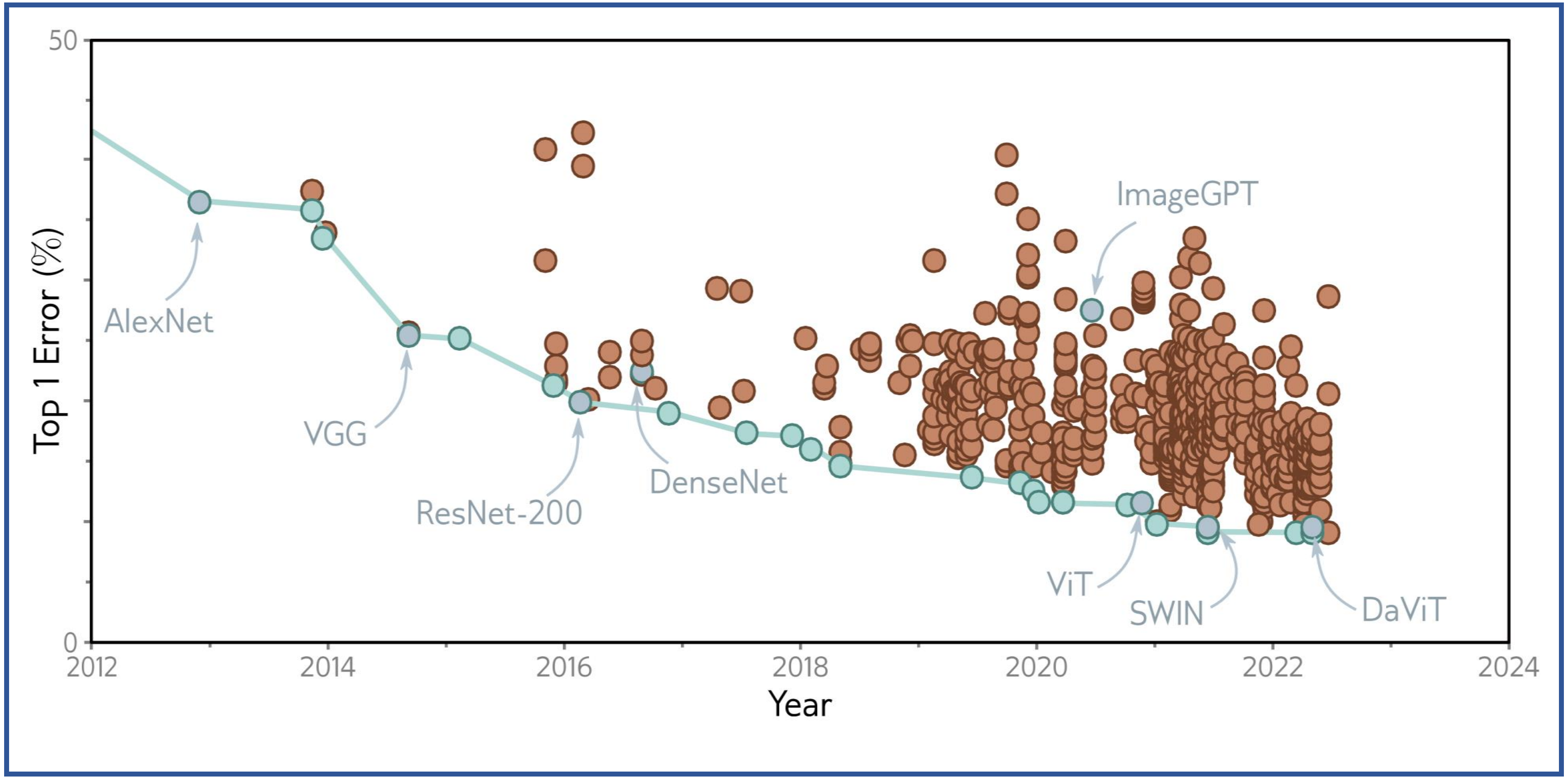




[Image from Prince 2023 UDL Book]

Convolutional Neural Network

❑ ImageNet Competition

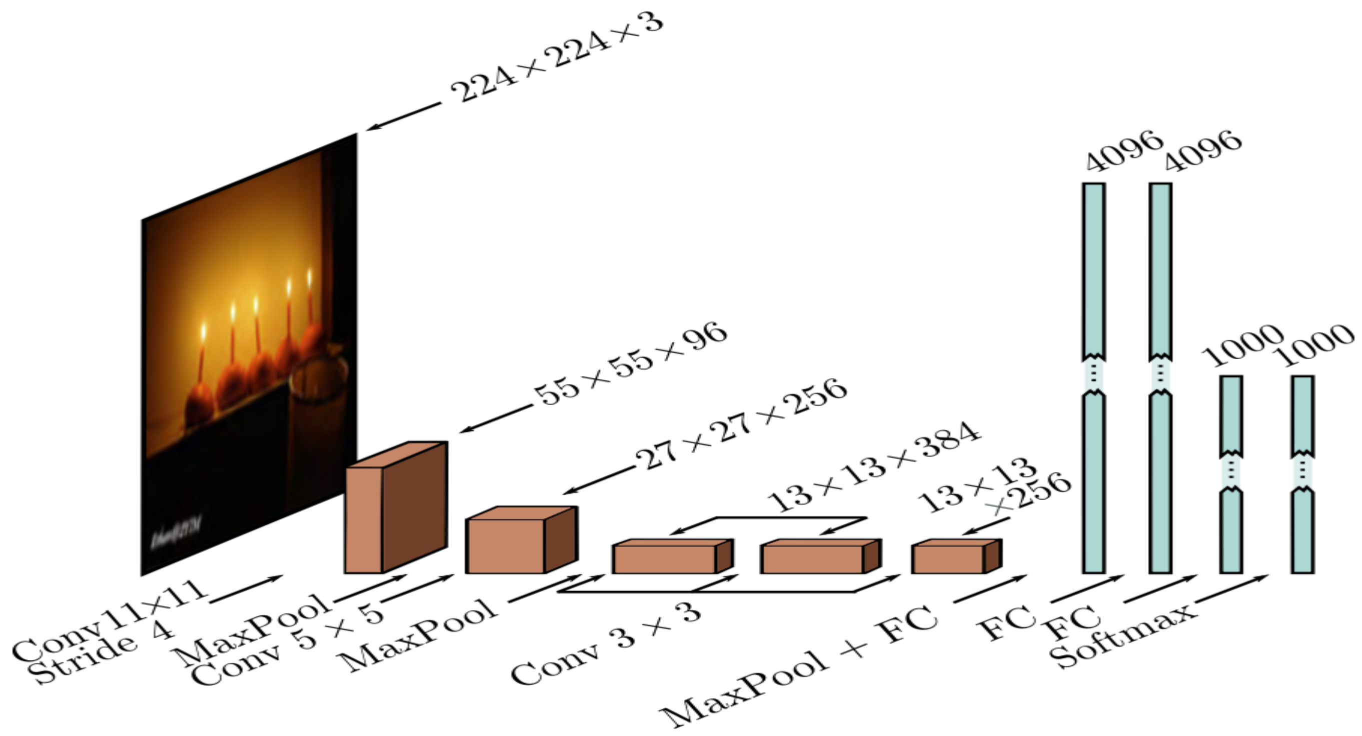


[Image from Prince 2023 UDL Book]

Residual Networks

Residual Network

AlexNet



[Image from Prince 2023 UDL Book]

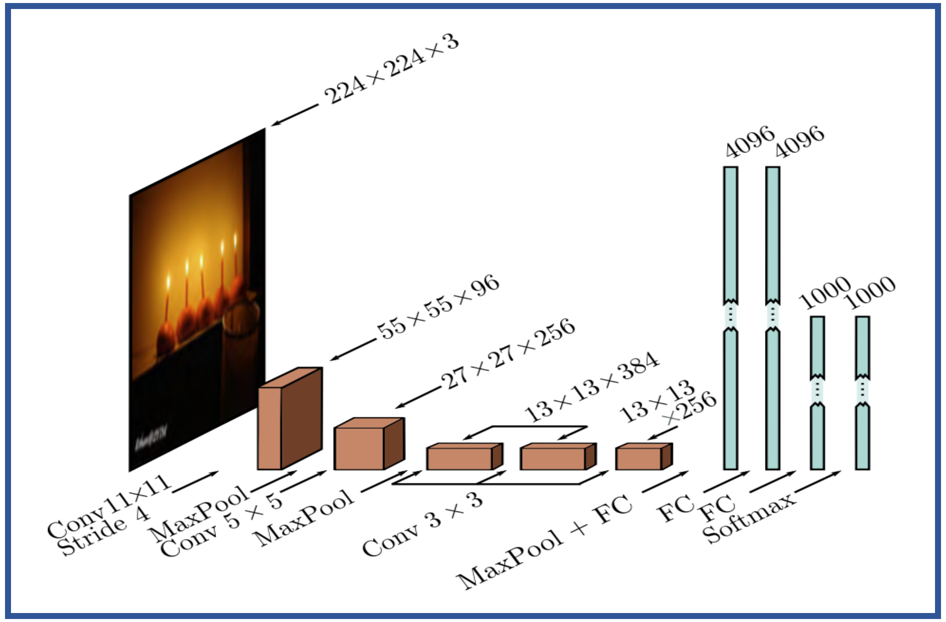
Residual Network

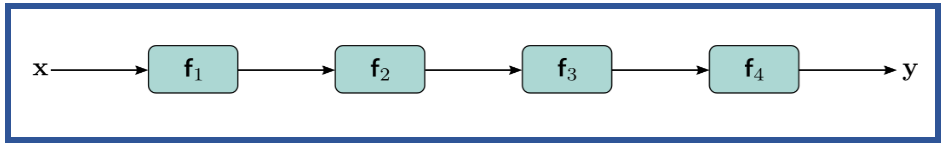
❑ Sequential Processing

$$\begin{aligned} \mathbf{h}_1 &= \mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1] \\ \mathbf{h}_2 &= \mathbf{f}_2[\mathbf{h}_1, \boldsymbol{\phi}_2] \\ \mathbf{h}_3 &= \mathbf{f}_3[\mathbf{h}_2, \boldsymbol{\phi}_3] \\ \mathbf{y} &= \mathbf{f}_4[\mathbf{h}_3, \boldsymbol{\phi}_4], \end{aligned}$$

$$\mathbf{y} = \mathbf{f}_4\Big[\mathbf{f}_3\Big[\mathbf{f}_2\Big[\mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1], \boldsymbol{\phi}_2\Big], \boldsymbol{\phi}_3\Big], \boldsymbol{\phi}_4\Big]$$

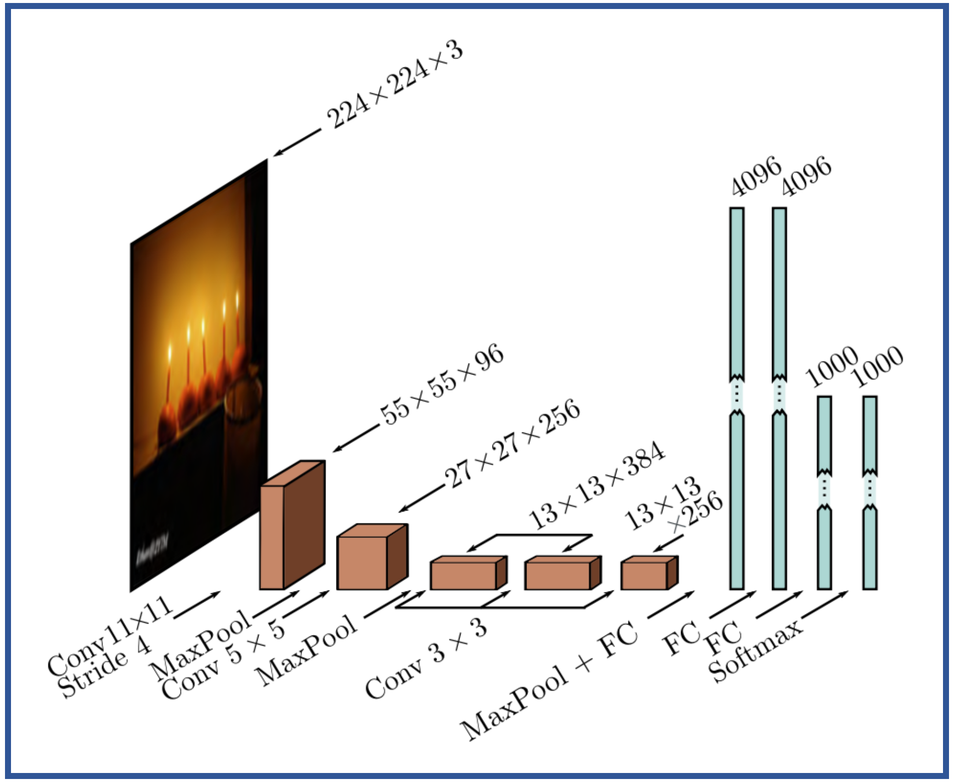
AlexNet





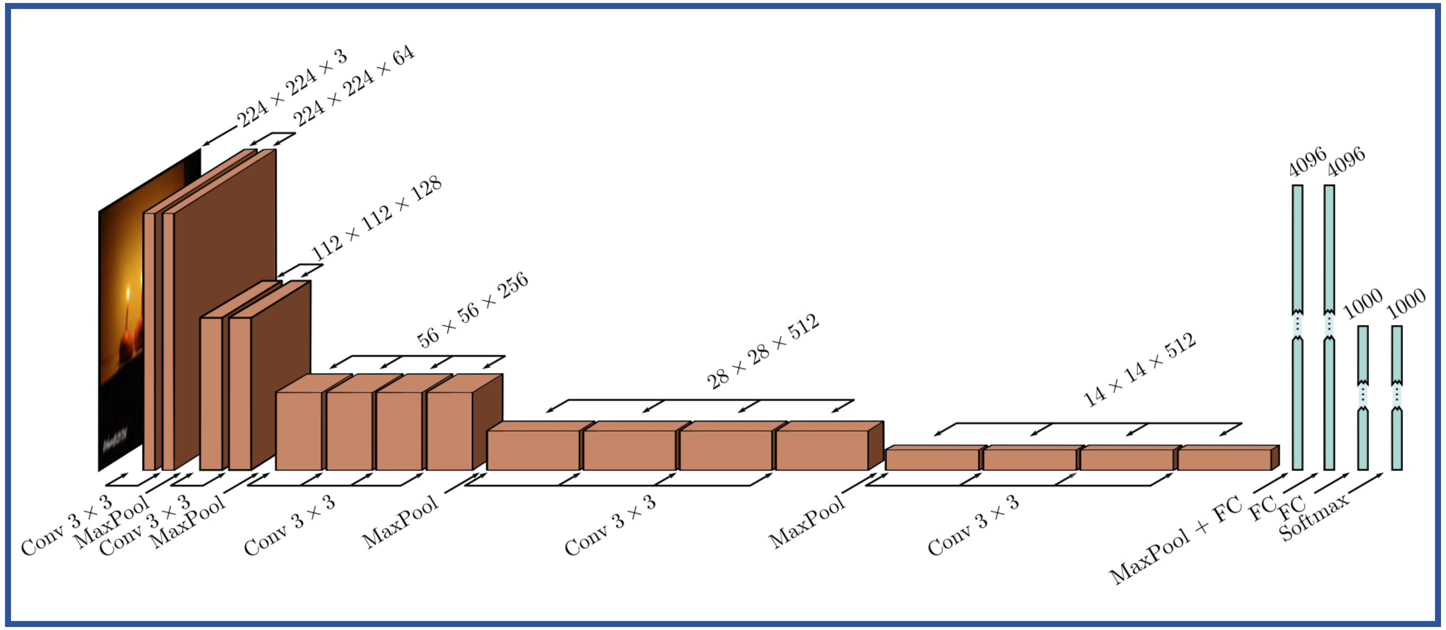
[Image from Prince 2023 UDL Book]

Residual Network



AlexNet

Error rate: 16.4%



VGG Network

Error rate: 6.8%

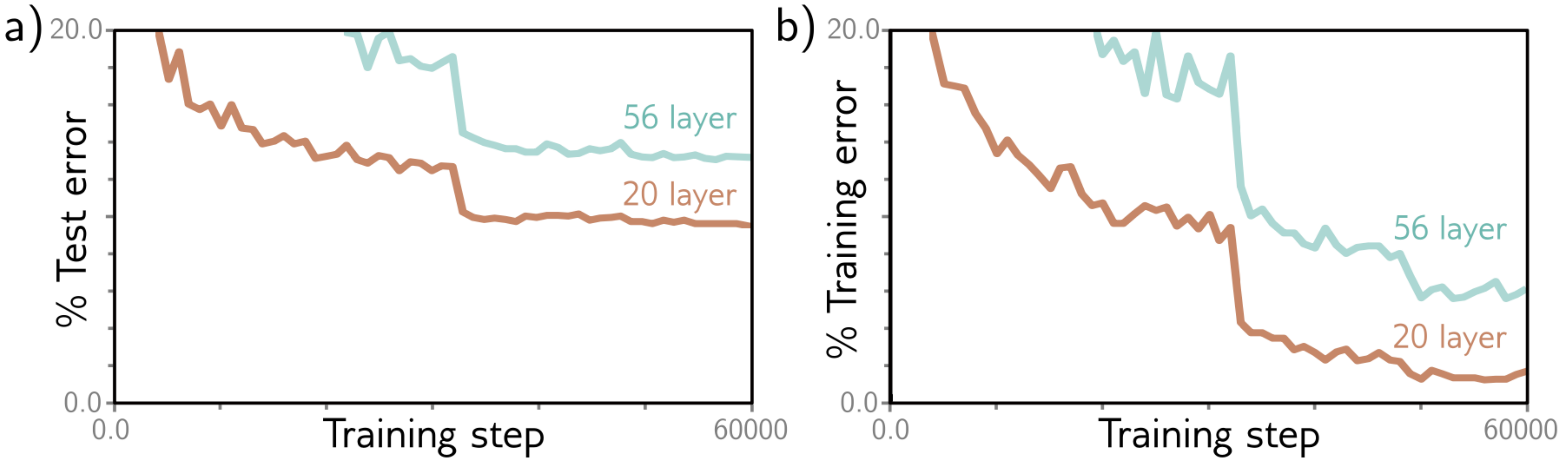
What if we add more layers?

[Image from Prince 2023 UDL Book]

Residual Network

Not Really!

Performance drop is in both training and test data



[Image from Prince 2023 UDL Book]

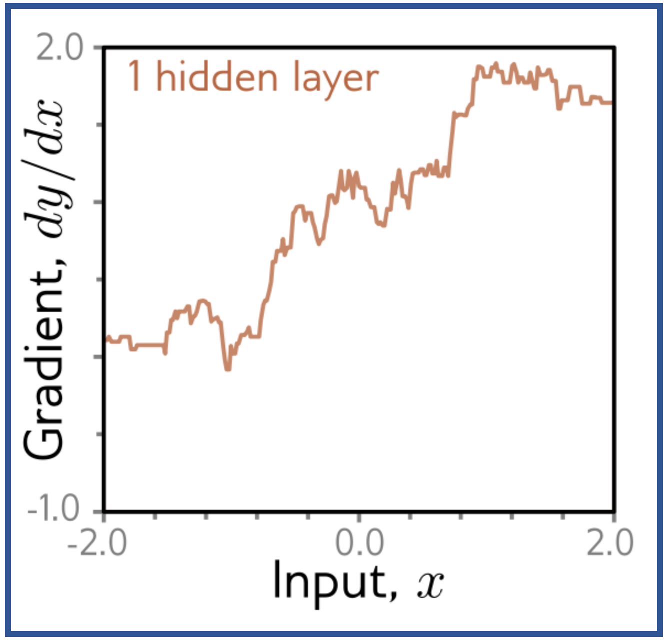
Residual Network

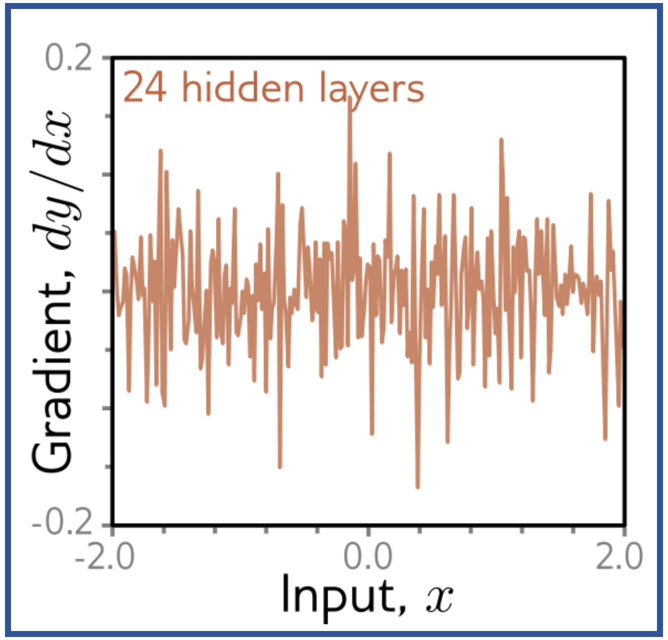
- ❑ Not well understood **why?**
- ❑ One conjecture:

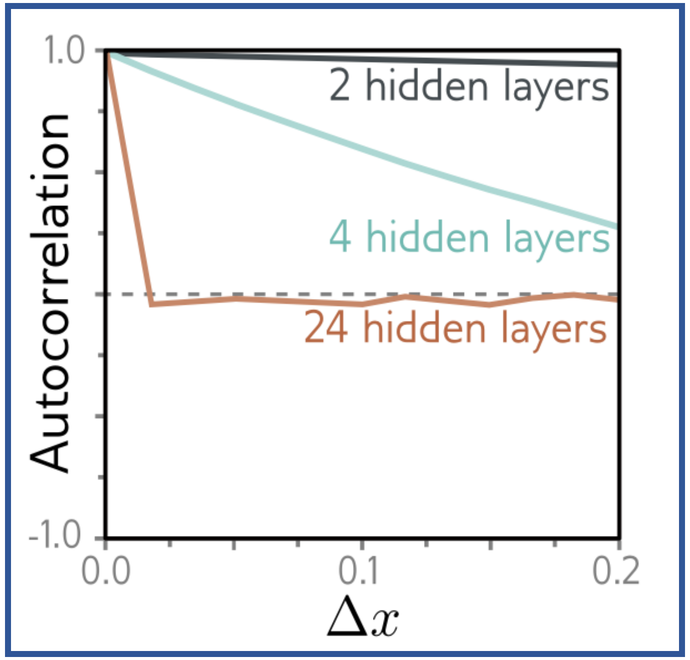
Shattered Gradient

Invariant to Translation

- ▪ **Loss gradients** change unpredictably when we modify parameters in early network layers.





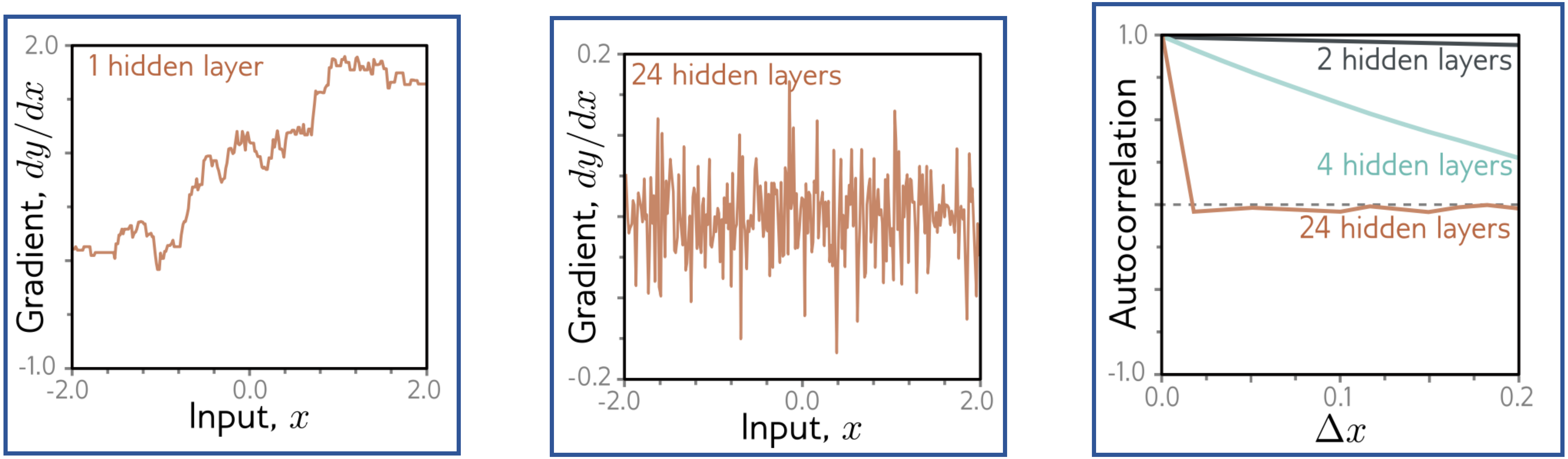


[Image from Prince 2023 UDL Book]

Residual Network

- Nearby gradients are **correlated** for **shallow networks**.
- But **correlation** quickly drops to zero for **deep networks**.

Shattered Gradient



[Image from Prince 2023 UDL Book]

Residual Network

- Changes in early network layers **modify** the output in a complex way.

$$
\begin{aligned}
\mathbf{h}_1 &= \mathbf{f}_1[\mathbf{x}, \phi_1] \\
\mathbf{h}_2 &= \mathbf{f}_2[\mathbf{h}_1, \phi_2] \\
\mathbf{h}_3 &= \mathbf{f}_3[\mathbf{h}_2, \phi_3] \\
\mathbf{y} &= \mathbf{f}_4[\mathbf{h}_3, \phi_4],
\end{aligned}
$$

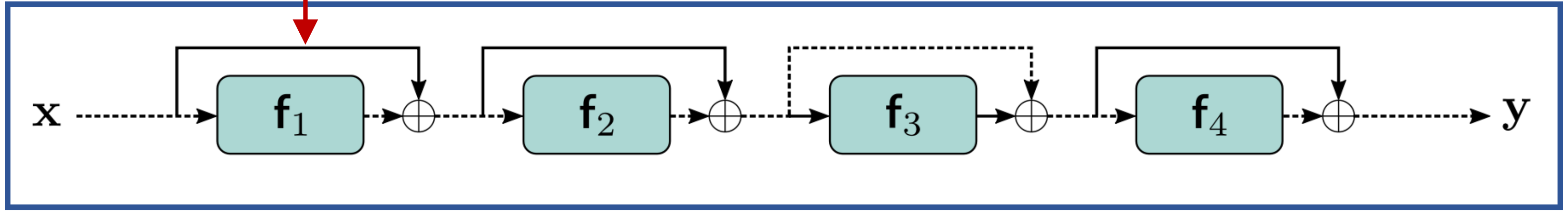
$$
\frac{\partial \mathbf{y}}{\partial \mathbf{f}_1} = \frac{\partial \mathbf{f}_2}{\partial \mathbf{f}_1}\frac{\partial \mathbf{f}_3}{\partial \mathbf{f}_2}\frac{\partial \mathbf{f}_4}{\partial \mathbf{f}_3}
$$

[Image from Prince 2023 UDL Book]

Residual Network

❑ Residual or Skip Connections:

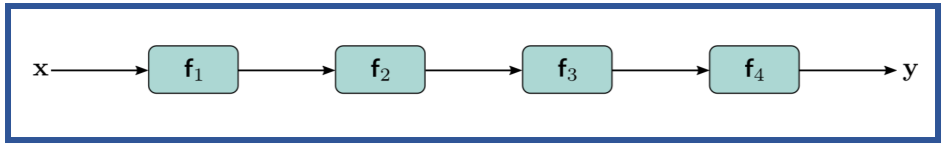
Residual Block or Residual Layer



$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{x} + \mathbf{f}_1[\mathbf{x}, \phi_1] \\
\mathbf{h}_2 &= \mathbf{h}_1 + \mathbf{f}_2[\mathbf{h}_1, \phi_2] \\
\mathbf{h}_3 &= \mathbf{h}_2 + \mathbf{f}_3[\mathbf{h}_2, \phi_3] \\
\mathbf{y} &= \mathbf{h}_3 + \mathbf{f}_4[\mathbf{h}_3, \phi_4],
\end{aligned}$$

Sequential

$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{f}_1[\mathbf{x}, \phi_1] \\
\mathbf{h}_2 &= \mathbf{f}_2[\mathbf{h}_1, \phi_2] \\
\mathbf{h}_3 &= \mathbf{f}_3[\mathbf{h}_2, \phi_3] \\
\mathbf{y} &= \mathbf{f}_4[\mathbf{h}_3, \phi_4],
\end{aligned}$$



[Image from Prince 2023 UDL Book]

Residual Network

❑ Residual or Skip Connections:

$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{x} + \mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1] \\
\mathbf{h}_2 &= \mathbf{h}_1 + \mathbf{f}_2[\mathbf{h}_1, \boldsymbol{\phi}_2] \\
\mathbf{h}_3 &= \mathbf{h}_2 + \mathbf{f}_3[\mathbf{h}_2, \boldsymbol{\phi}_3] \\
\mathbf{y} &= \mathbf{h}_3 + \mathbf{f}_4[\mathbf{h}_3, \boldsymbol{\phi}_4],
\end{aligned}$$

$$\begin{aligned}
\mathbf{y} = \mathbf{x} &+ \mathbf{f}_1[\mathbf{x}] \\
&+ \mathbf{f}_2\big[\mathbf{x} + \mathbf{f}_1[\mathbf{x}]\big] \\
&+ \mathbf{f}_3\Big[\mathbf{x} + \mathbf{f}_1[\mathbf{x}] + \mathbf{f}_2\big[\mathbf{x} + \mathbf{f}_1[\mathbf{x}]\big]\Big] \\
&+ \mathbf{f}_4\Big[\mathbf{x} + \mathbf{f}_1[\mathbf{x}] + \mathbf{f}_2\big[\mathbf{x} + \mathbf{f}_1[\mathbf{x}]\big] + \mathbf{f}_3\big[\mathbf{x} + \mathbf{f}_1[\mathbf{x}] + \mathbf{f}_2[\mathbf{x} + \mathbf{f}_1[\mathbf{x}]]\big]\Big]
\end{aligned}$$

Sequential

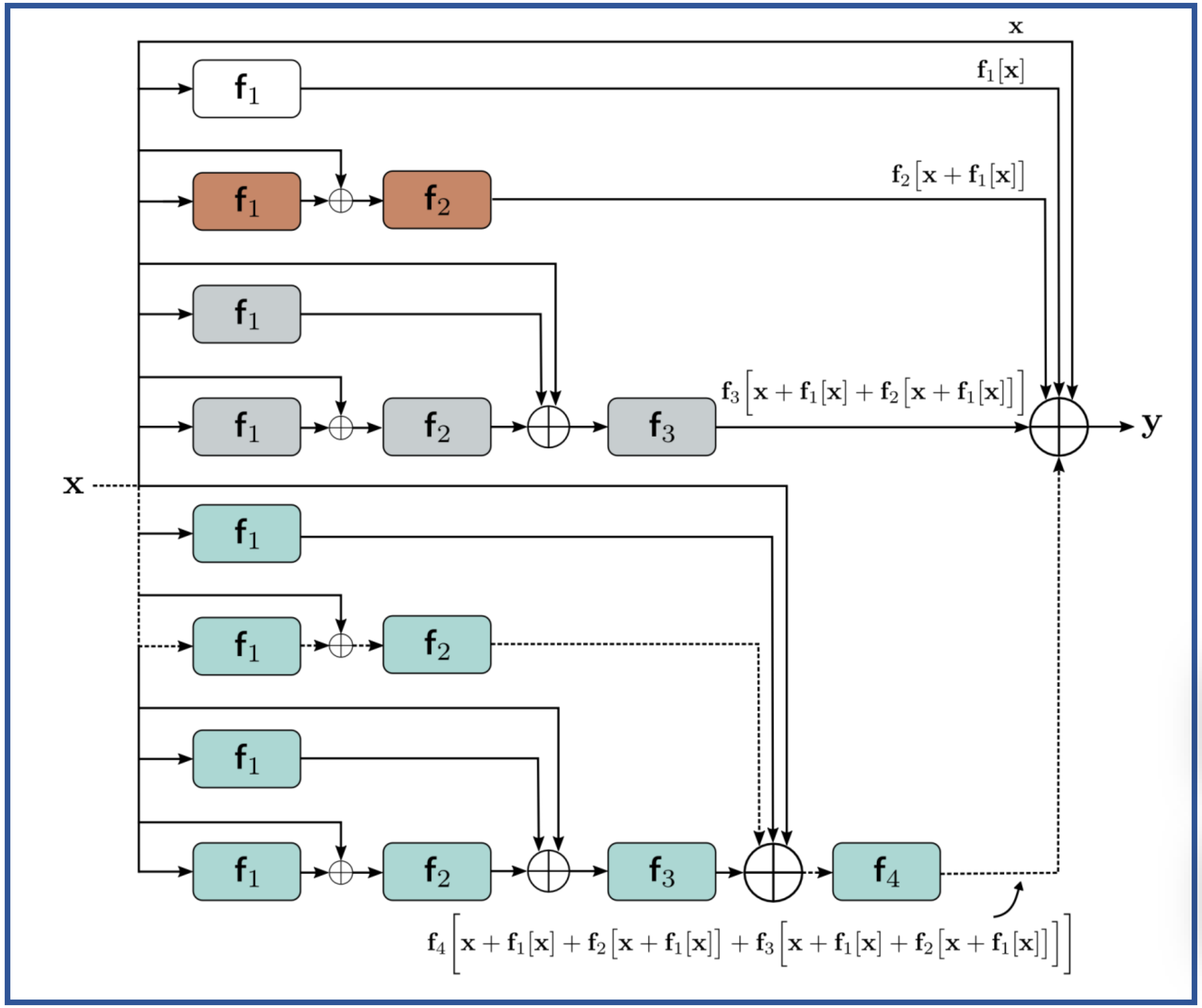
$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1] \\
\mathbf{h}_2 &= \mathbf{f}_2[\mathbf{h}_1, \boldsymbol{\phi}_2] \\
\mathbf{h}_3 &= \mathbf{f}_3[\mathbf{h}_2, \boldsymbol{\phi}_3] \\
\mathbf{y} &= \mathbf{f}_4[\mathbf{h}_3, \boldsymbol{\phi}_4],
\end{aligned}$$

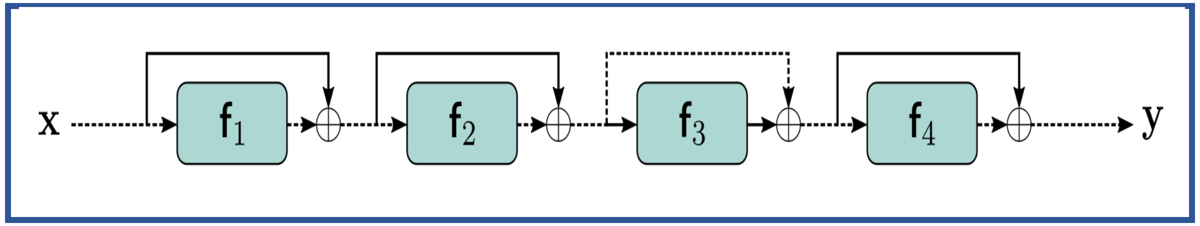
$$\mathbf{y} = \mathbf{f}_4\Big[\mathbf{f}_3\Big[\mathbf{f}_2\big[\mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1], \boldsymbol{\phi}_2\big], \boldsymbol{\phi}_3\Big], \boldsymbol{\phi}_4\Big]$$

[Image from Prince 2023 UDL Book]

Residual Network

❑ Residual or Skip Connections:





$$
\begin{aligned}
y = x &+ f_1[x] \\
&+ f_2\big[x + f_1[x]\big] \\
&+ f_3\Big[x + f_1[x] + f_2\big[x + f_1[x]\big]\Big] \\
&+ f_4\Big[x + f_1[x] + f_2\big[x + f_1[x]\big] + f_3\big[x + f_1[x] + f_2[x + f_1[x]]\big]\Big]
\end{aligned}
$$

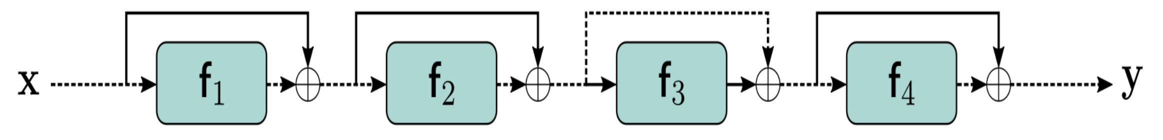
Final Output = Input + Sum of 4 smaller networks

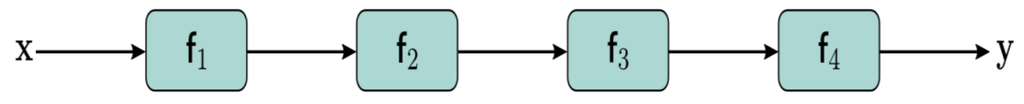
Ensemble of smaller networks

[Image from Prince 2023 UDL Book]

Residual Network

❑ Residual or Skip Connections:





$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{x} + \mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1] \\
\mathbf{h}_2 &= \mathbf{h}_1 + \mathbf{f}_2[\mathbf{h}_1, \boldsymbol{\phi}_2] \\
\mathbf{h}_3 &= \mathbf{h}_2 + \mathbf{f}_3[\mathbf{h}_2, \boldsymbol{\phi}_3] \\
\mathbf{y} &= \mathbf{h}_3 + \mathbf{f}_4[\mathbf{h}_3, \boldsymbol{\phi}_4],
\end{aligned}$$

$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{f}_1[\mathbf{x}, \boldsymbol{\phi}_1] \\
\mathbf{h}_2 &= \mathbf{f}_2[\mathbf{h}_1, \boldsymbol{\phi}_2] \\
\mathbf{h}_3 &= \mathbf{f}_3[\mathbf{h}_2, \boldsymbol{\phi}_3] \\
\mathbf{y} &= \mathbf{f}_4[\mathbf{h}_3, \boldsymbol{\phi}_4],
\end{aligned}$$

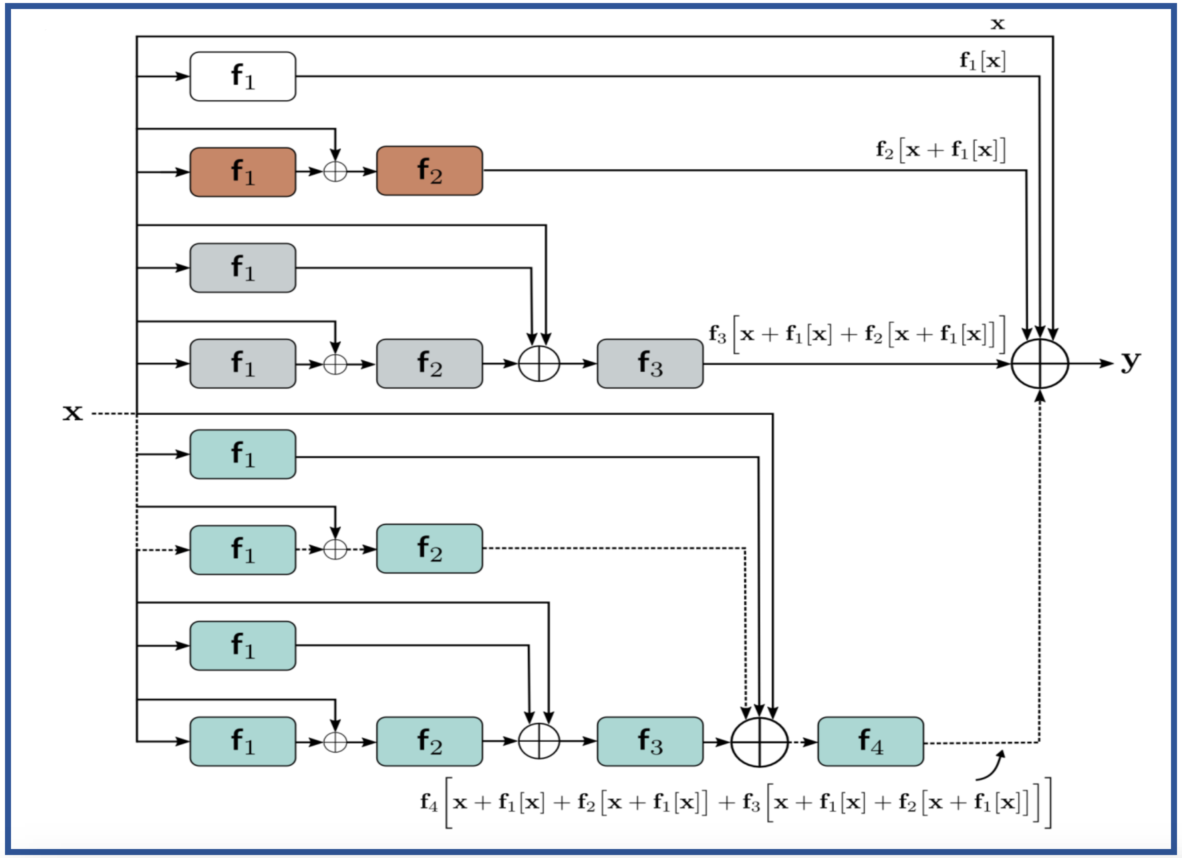
$$\frac{\partial \mathbf{y}}{\partial \mathbf{f}_1} = \mathbf{I} + \frac{\partial \mathbf{f}_2}{\partial \mathbf{f}_1} + \left(\frac{\partial \mathbf{f}_3}{\partial \mathbf{f}_1} + \frac{\partial \mathbf{f}_2}{\partial \mathbf{f}_1} \frac{\partial \mathbf{f}_3}{\partial \mathbf{f}_2} \right) + \left(\frac{\partial \mathbf{f}_4}{\partial \mathbf{f}_1} + \frac{\partial \mathbf{f}_2}{\partial \mathbf{f}_1} \frac{\partial \mathbf{f}_4}{\partial \mathbf{f}_2} + \frac{\partial \mathbf{f}_3}{\partial \mathbf{f}_1} \frac{\partial \mathbf{f}_4}{\partial \mathbf{f}_3} + \frac{\partial \mathbf{f}_2}{\partial \mathbf{f}_1} \frac{\partial \mathbf{f}_3}{\partial \mathbf{f}_2} \frac{\partial \mathbf{f}_4}{\partial \mathbf{f}_3} \right)$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{f}_1} = \frac{\partial \mathbf{f}_2}{\partial \mathbf{f}_1} \frac{\partial \mathbf{f}_3}{\partial \mathbf{f}_2} \frac{\partial \mathbf{f}_4}{\partial \mathbf{f}_3}$$

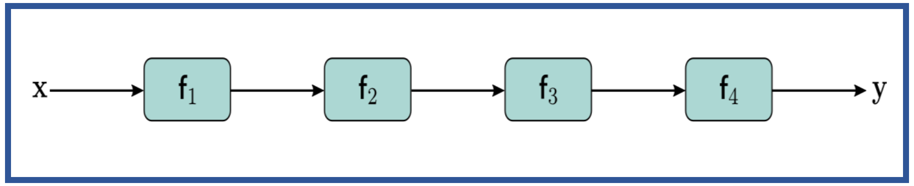
[Image from Prince 2023 UDL Book]

Residual Network

❑ Residual or Skip Connections:



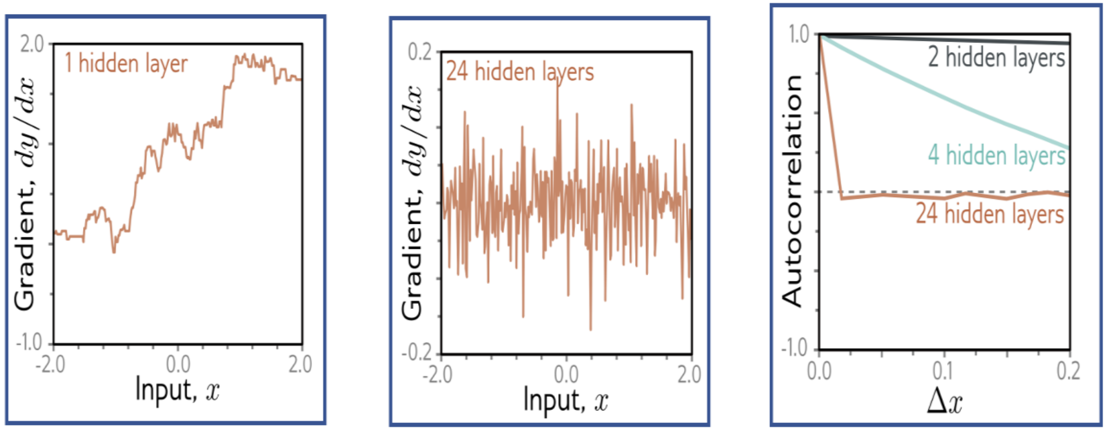
Networks with residual links **suffer less** from **shattered gradients.**



❑ Nearby gradients are **correlated** for **shallow networks**.

Shattered Gradient

❑ But **correlation** quickly drops to zero for **deep networks**.



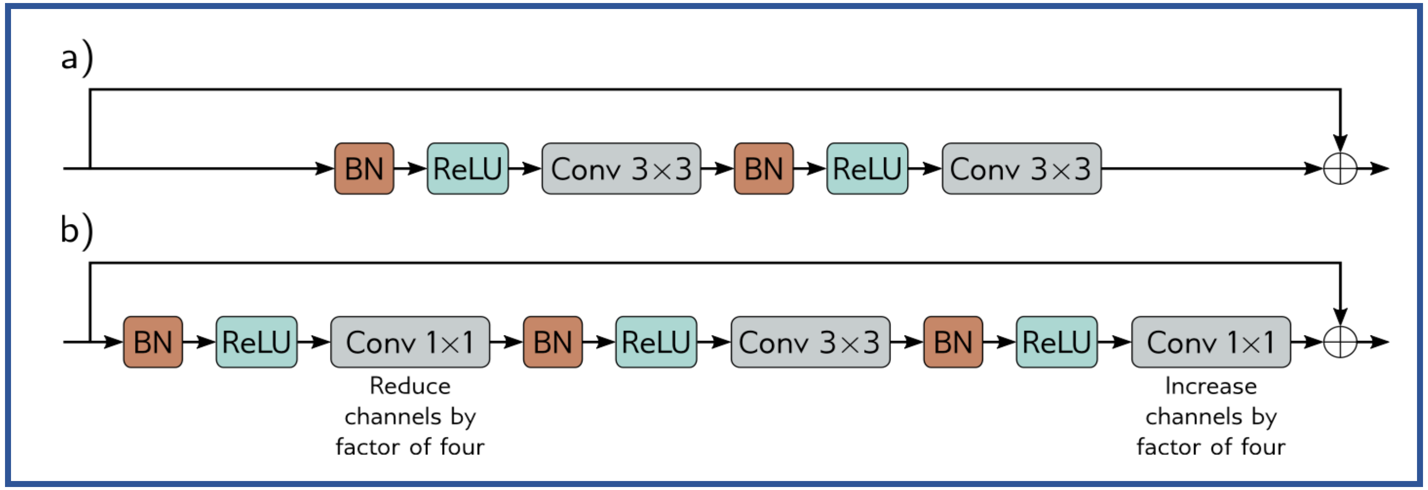
[Image from Prince 2023 UDL Book]

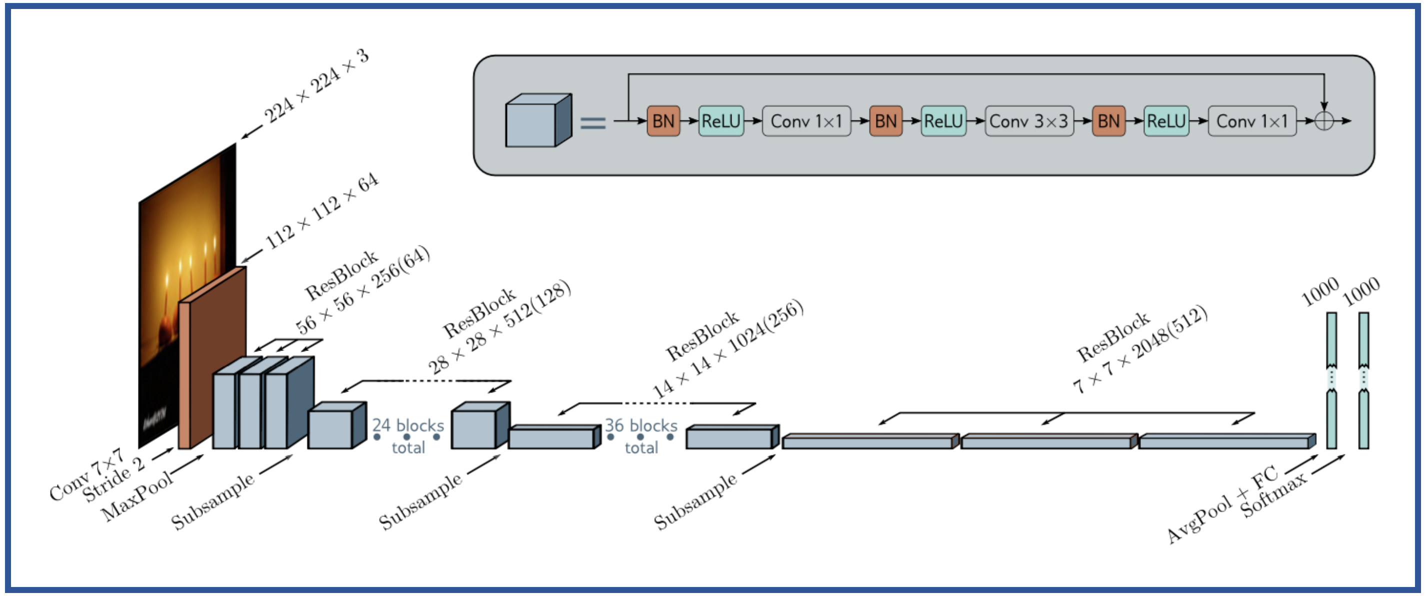
Residual Network

❑ ResNet

❖ Error Rate:

- ResNet-200: 4.8%
- VGG: 6.8%
- AlexNet: 16.4%

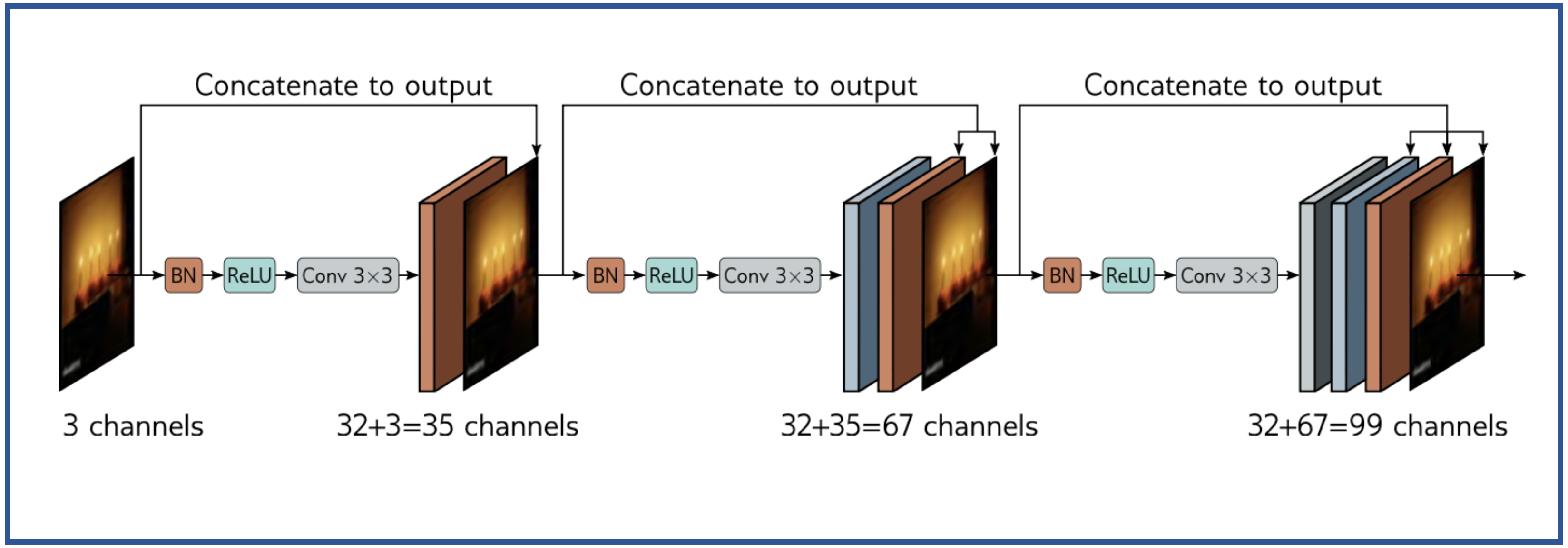




[Image from Prince 2023 UDL Book]

Residual Network

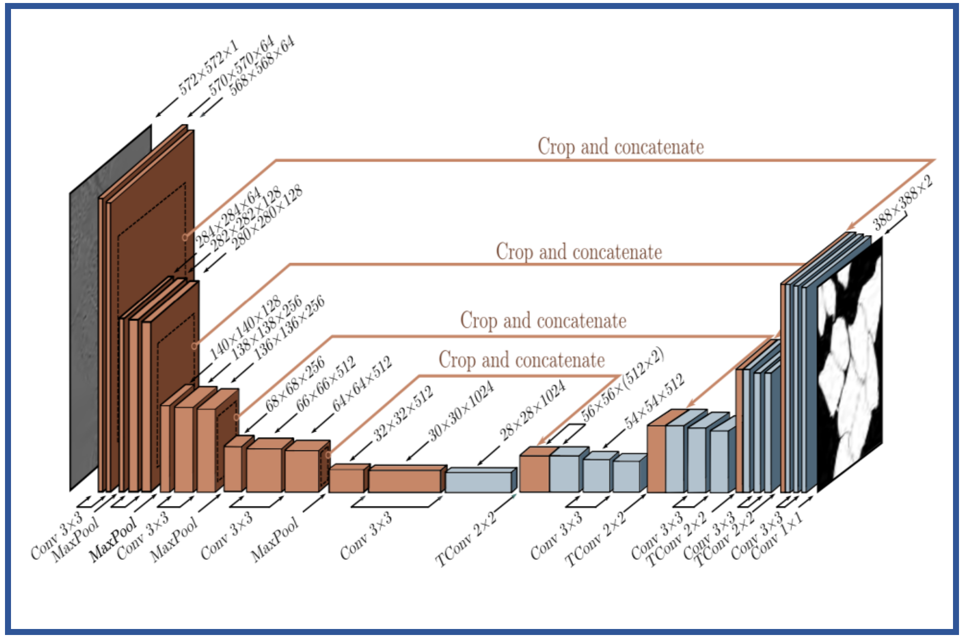
❑ DenseNet

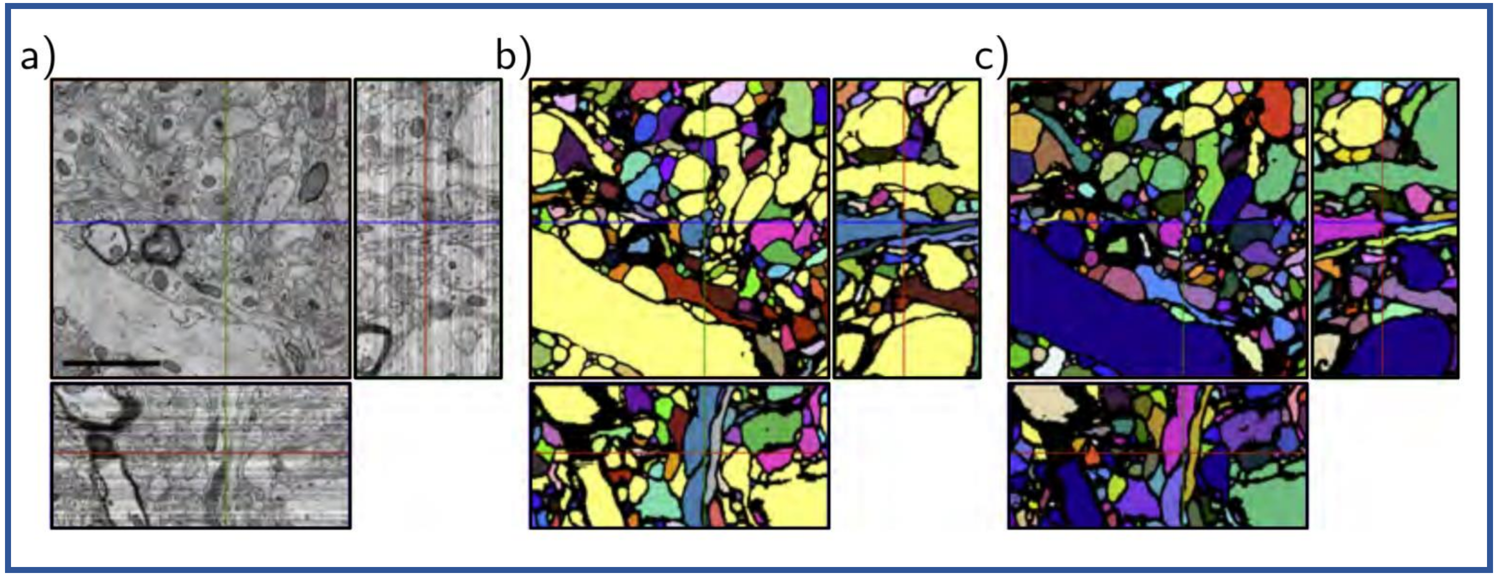


[Image from Prince 2023 UDL Book]

Residual Network

❑ U-Net:

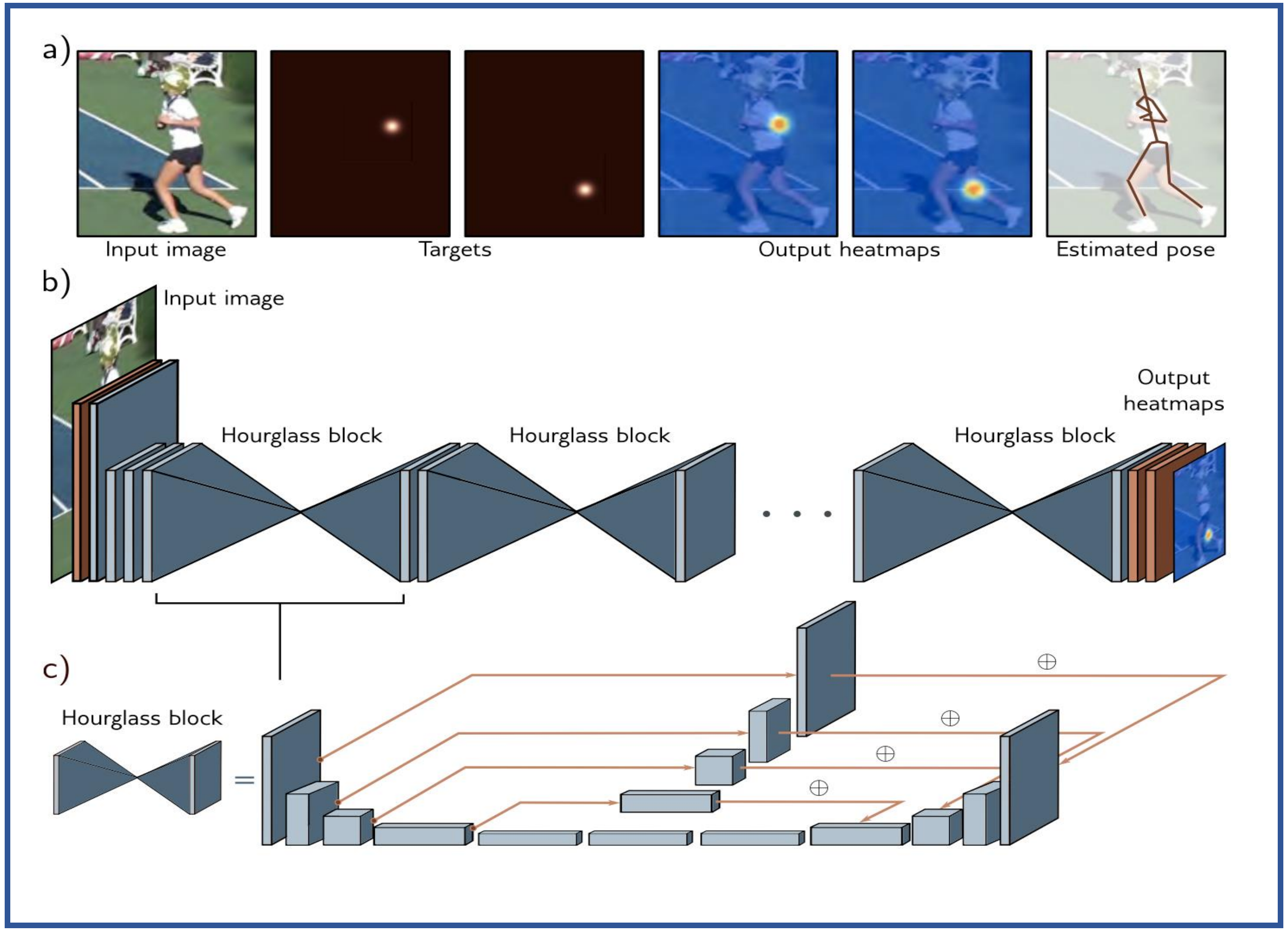




[Image from Prince 2023 UDL Book]

Residual Network

❑ Hourglass Networks:

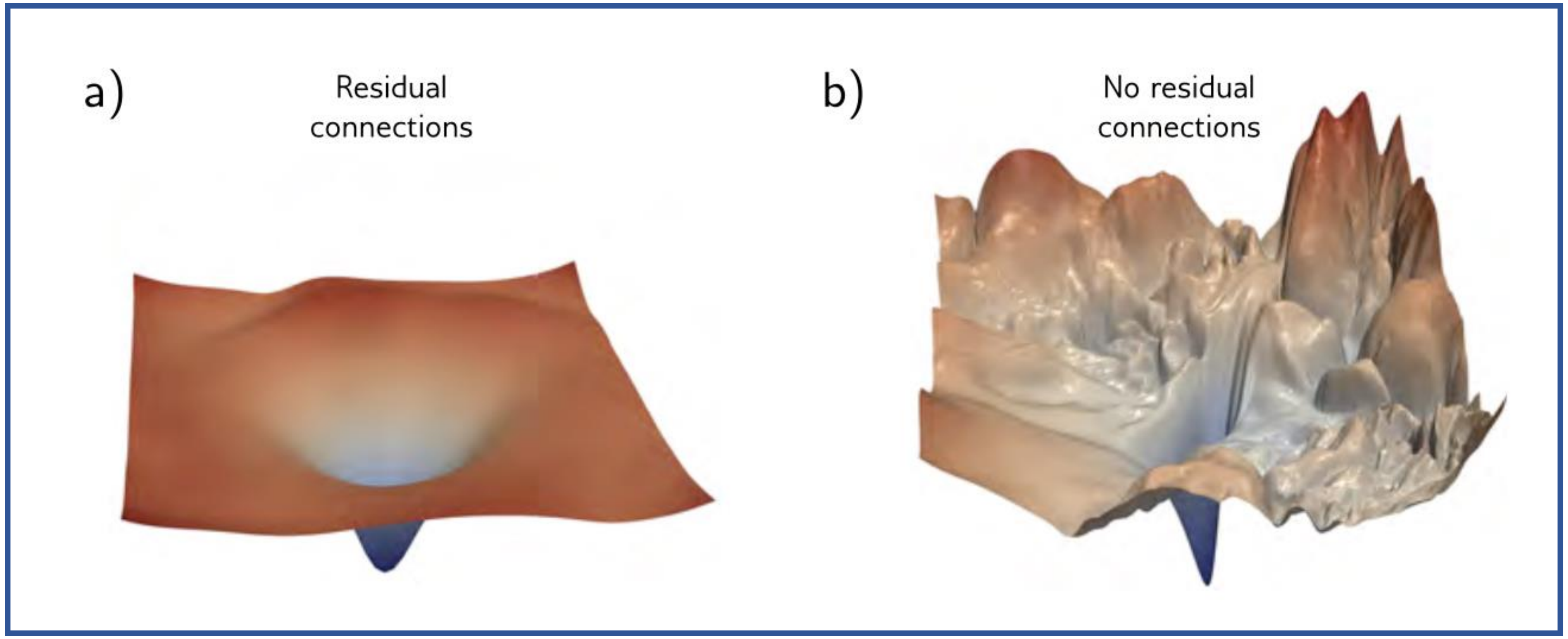


[Image from Prince 2023 UDL Book]

Residual Network

❑ Why does residual network perform well?

- Enables much deeper network to be trained.



[Image from Prince 2023 UDL Book]

ScAI

End