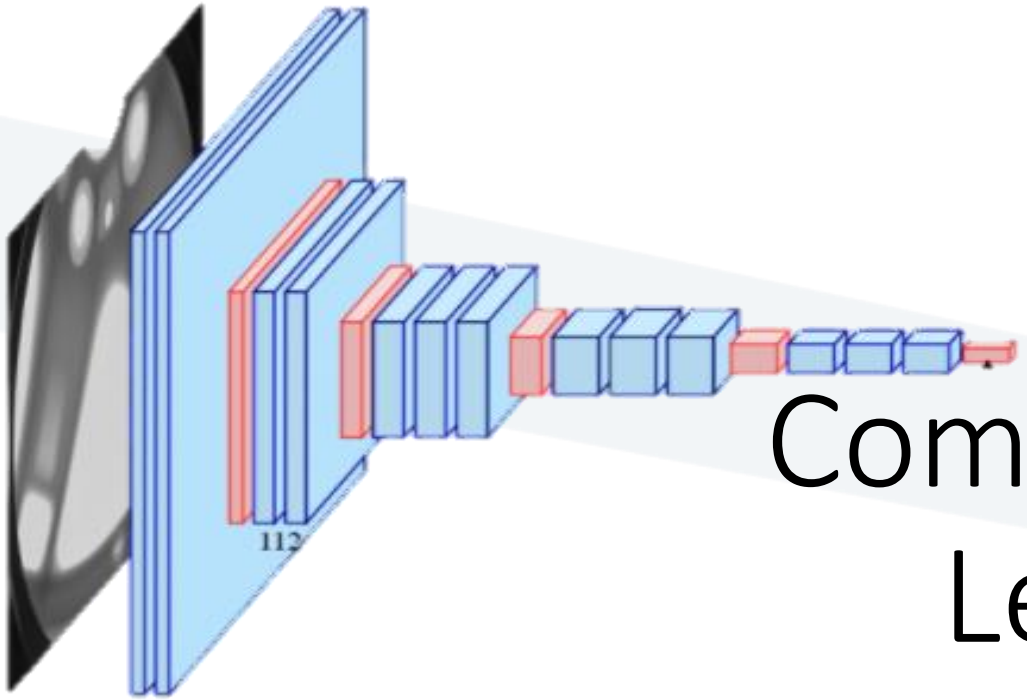
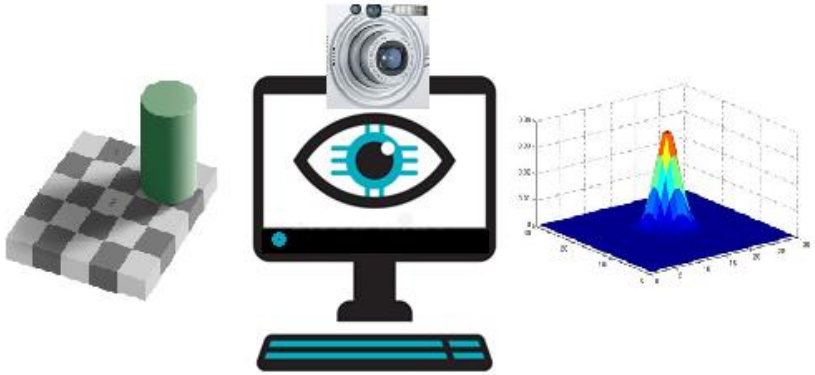


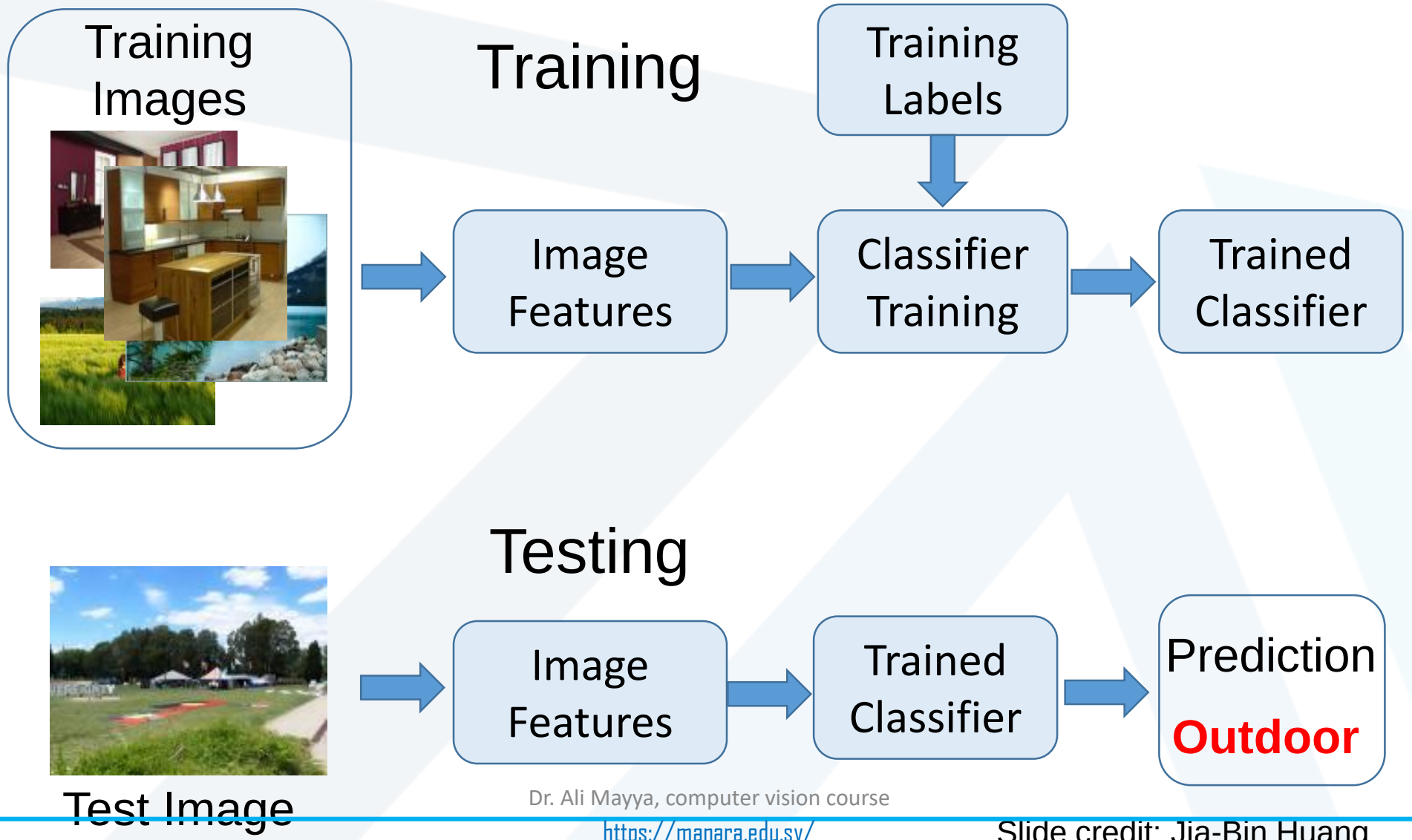


Computer Vision
Lecture 10
Deep Learning for visual recognition

Dr. Ali Mahmoud Mayya
Computer Science Faculty
Tishreen University, Syria

2023

Traditional Image Categorization: Training & Testing phase



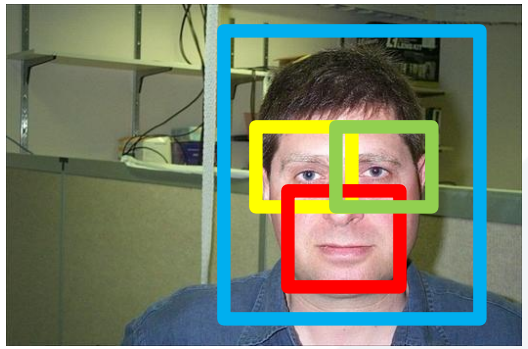
Slide credit: Jia-Bin Huang

Learning Feature Hierarchy
التعلم الهرمي للسمات

Goal: Learn **useful higher-level features** from images

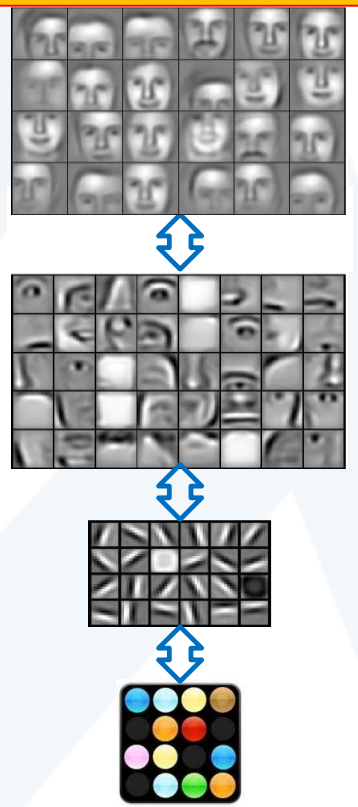
الهدف: التدريب على التعرف على السمات عالية المستوى من الصورة

Input data



Lee et al., ICML 2009; CACM 2011

Feature representation
هرمية تمثيل السمات



3rd layer
“Objects”

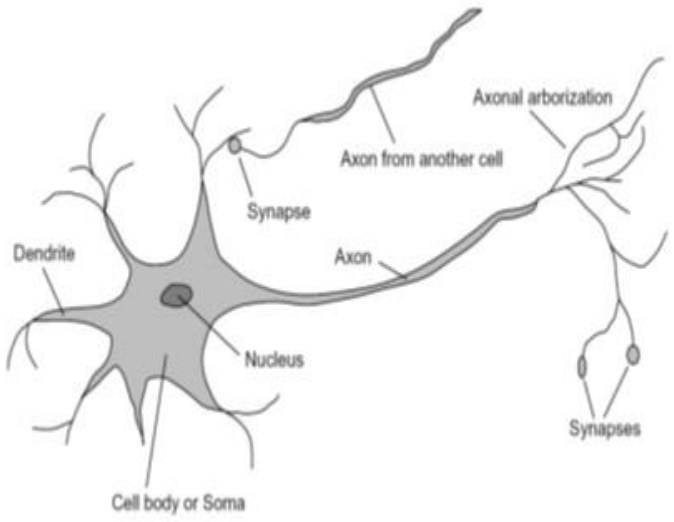
2nd layer
“Object parts”

1st layer
“Edges”

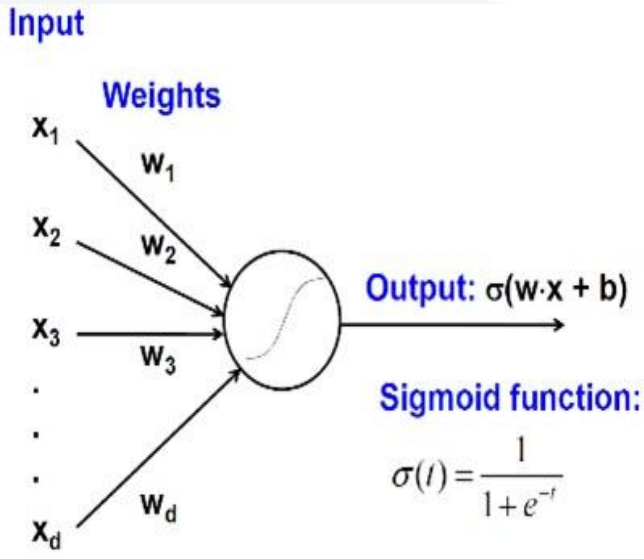
Pixels

Slide: Rob Fergus

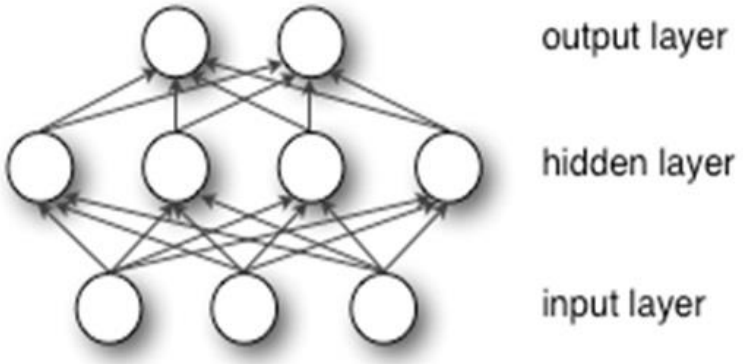
Neural networks basics (remind!)

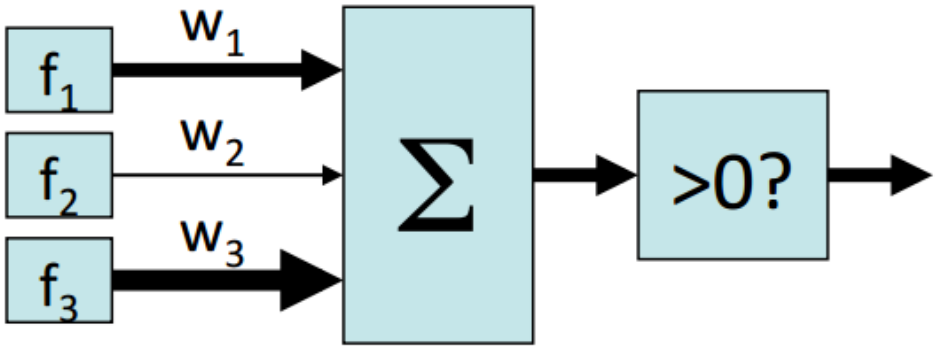


A biological neuron
العصبون البيولوجي في الدماغ



An artificial neuron (Perceptron)
- a linear classifier
العصبون الصناعي (مصنف خطي)

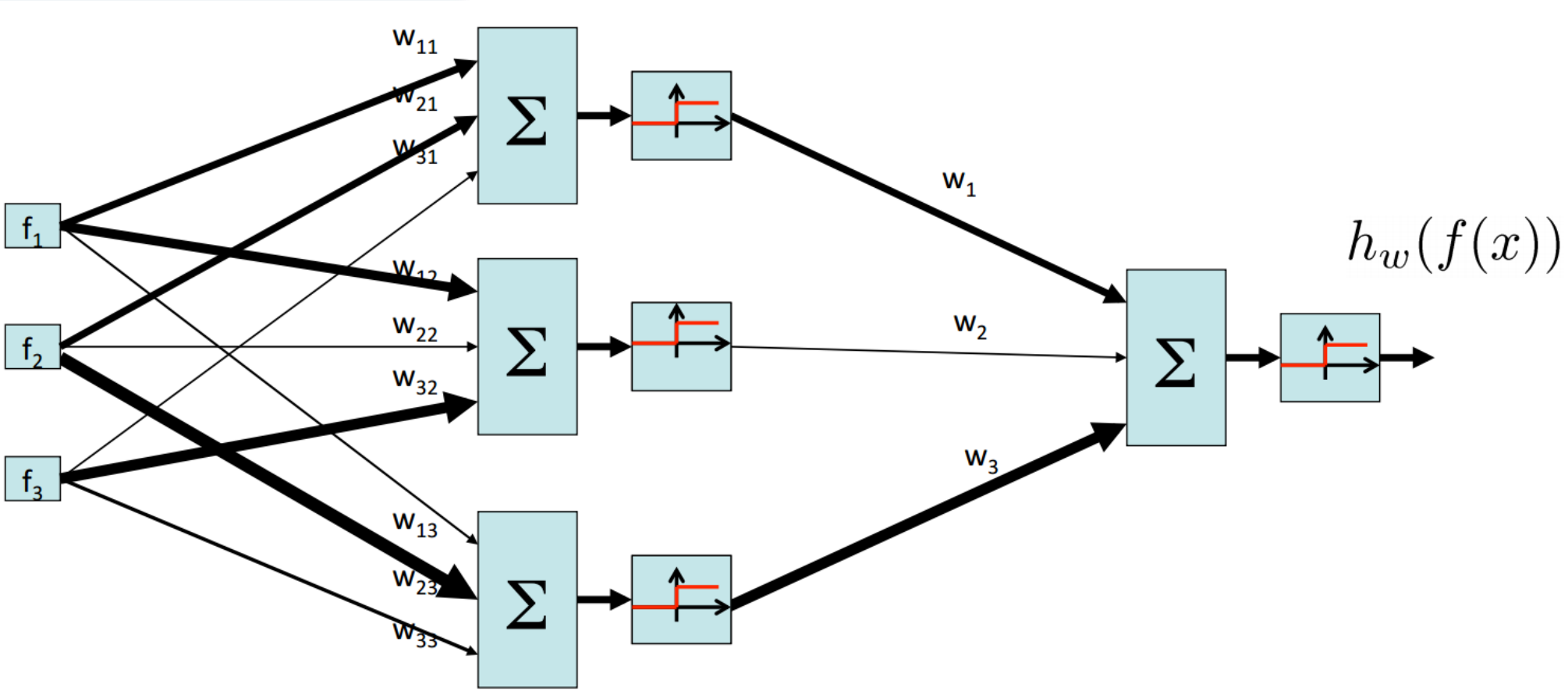




Inputs are feature values
Each feature has a weight
Sum is the activation

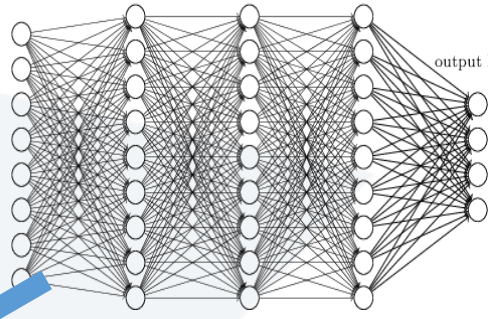
$$\text{activation}_w(x) = \sum_i w_i \cdot f_i(x) = w \cdot f(x)$$

Neural networks basics (remind!): Multilayer Perceptron



Now Deep learning ☺

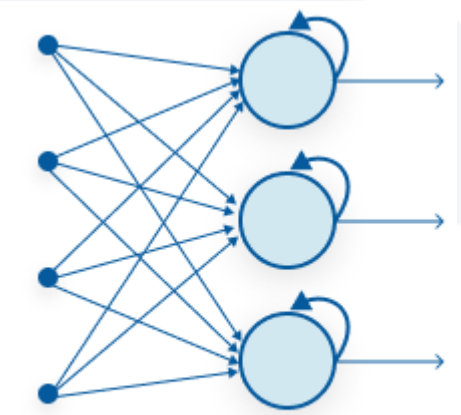
Big labeled datasets



Deep learning

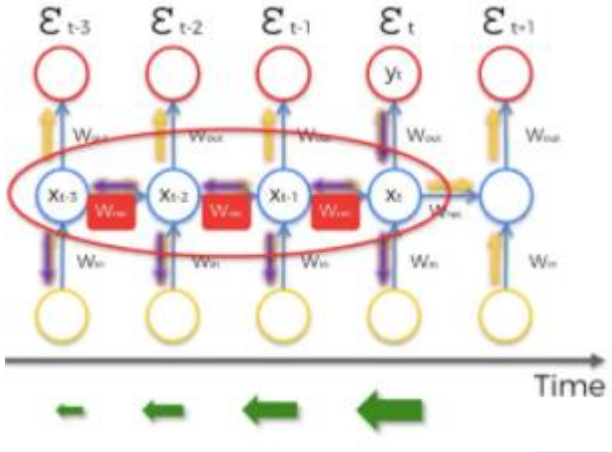
GPU technology

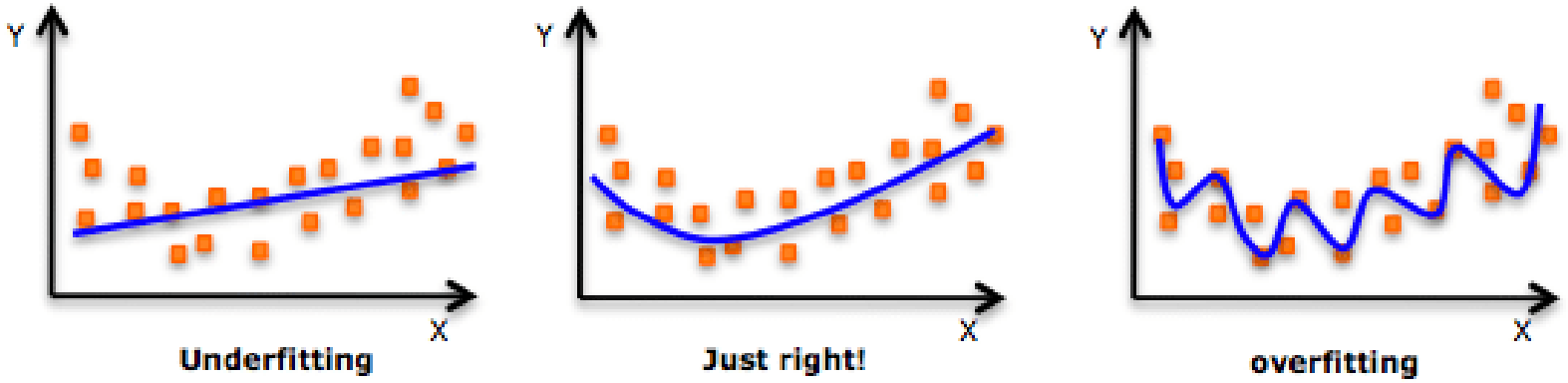
Convolutional Neural Networks (CNN, ConvNet, DCN)



Recurrent Neural Network

- لم يعد بوسع الشبكات التقليدية مواكبة الضخامة في كمية البيانات التي تدخل مرحلة التدريب
- زيادة عدد الطبقات في الشبكات التقليدية خلق مشاكل منها زمن التدريب الطويل وبطء عملية التعلم خصوصاً بحالة **Backpropagation** أو مشكلة **Overfitting** والتي يتأثر فيها التدريب بالضجيج في البيانات فنحصل على عملية تدريب غير منتظمة يكون فيها النموذج منحازاً بشكل كبير لمجموعة التدريب وباختبار النموذج نحصل على نتائج سيئة
- مشكلة **Gradient Vanishing** في الشبكات متعددة الطبقات

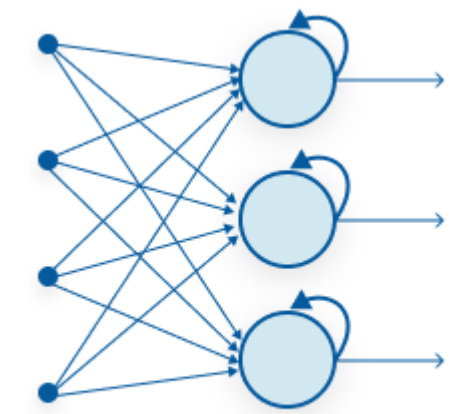




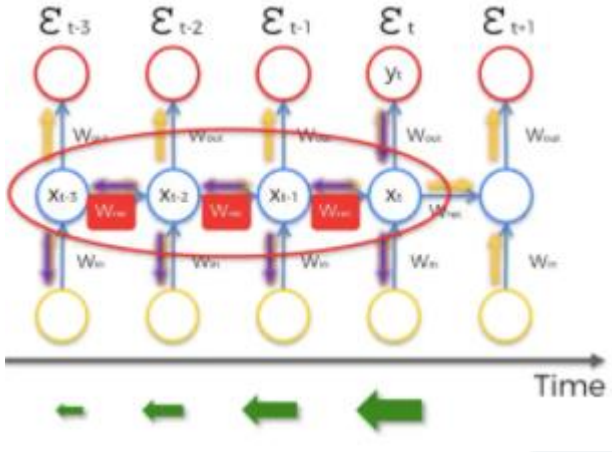
An example of overfitting, underfitting and a model that's "just right!"

https://manara.edu.sy/

Convolutional Neural Networks (CNN, ConvNet, DCN)



Recurrent Neural Network



- تستطيع بعض شبكات التعلم العميق أن تحتفظ ببعض بياناتها وتقدمها للمراحل المستقبلية مما يخلق نوع من الذاكرة في هذه الشبكات ومن أمثلتها **Recurrent NN (RNN)** و **Long-Short Term Memory (LSTM)**
- هناك بعض الشبكات التي تستطيع أن تنجز مراحل المعالجة المسبقة واستخلاص السمات معاً فدخلها صورة وخرجها سمات ومنها **Convolutional Neural Networks (CNN)**
- في شبكات التعلم العميق ليس بالضرورة أن يتصل العصبون في طبقة ما بكل العصبونات في الطبقة التالية

Convolutional Neural Networks (CNN)

أحد أنواع شبكات التعلم العميق ذات التغذية الأمامية **Feed Forwarding Deep Learning NN.**

تستخدم غالباً لتمييز أنماط الصور والتعرف على مكونات الصورة.

تتألف من مجموعة من الطبقات الالتفافية والتي تتبع بطبقات انتخاب.

في كل طبقة يتم التعرف على جزء من معلومات الصورة.

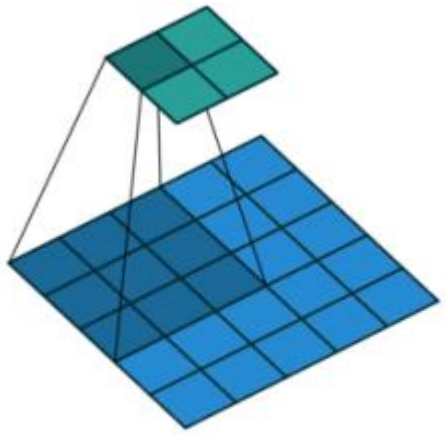
كلما دخلنا لعمق الشبكة أكثر كلما أصبحت المعلومات التي نحصل عليها من الشبكة أفضل وأقل حجماً.

تتألف من أربع مراحل أساسية موضحة بالشكل

يتم زيادة العمق من خلال زيادة عدد الطبقات الالتفافية **Convolutional Layers** والتي تؤدي للحصول على معلومات أكثر أهمية من الصورة مع زيادة العمق

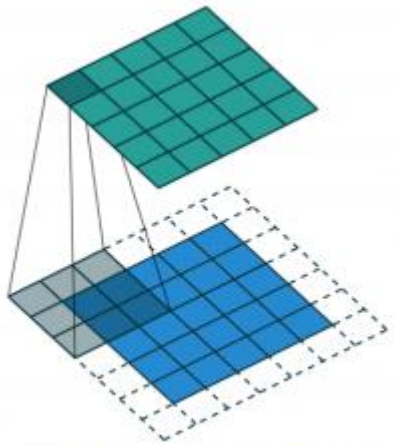
https://www.youtube.com/watch?v=oI2rvjbzVmI

تتضمن تطبيق مرشحات بطريقة الالتفاف على الصورة ونحصل منها على صورة سمات محددة تتم هذه العملية باستخدام مرشحات بأبعاد محددة K*K وباستخدام معاملين: الحشو Padding والخطوة Stride



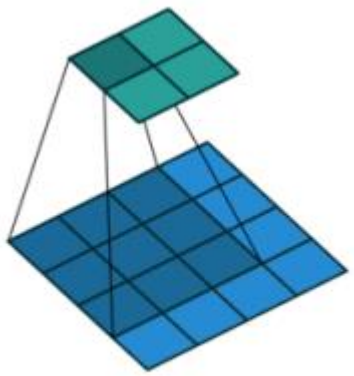
No padding, stride 2

تطبيق الالتفاف بمرشح 3*3 بدون حشو، حجم خطوة المرشح هنا 2 هنا الصورة الأصلية 5*5 ستصبح بحجم 2*2



Padding 1, stride 1

تطبيق الالتفاف بمرشح 3*3 مع حشو، حجم الخطوة 1
حجم الصورة بعد الالتفاف هو ذاته حجمها قبل الالتفاف.
الصورة كانت بحجم 5*5 وبقيت كذلك.



No padding, stride 1

تطبيق الالتفاف بمرشح 3*3 وبدون حشو، حجم الخطوة 1.
حجم الصورة بعد الالتفاف
أقل من حجمها قبله.
الصورة كانت بحجم 4*4 وأصبحت بحجم 2*2.

- مصطلحات هامة
- Stride الخطوة التي يقوم بها المرشح لدى انتقاله من بكسل إلى بكسل مجاور
- حشو Padding إضافة بكسلات حدية على جوانب الصورة لتناسب حجم القناع

باستخدام القانون الآتي نستطيع معرفة حجم خرج طبقة Convolution:

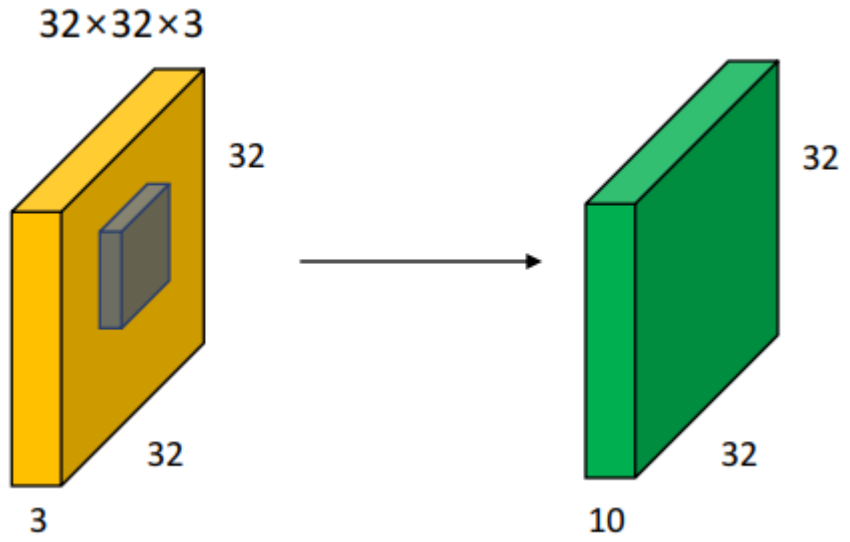
Output volume size= NoF*((N−K+2P)/S+1)

NoF: عدد المرشحات، k: حجم المرشح، N حجم الدخل، S: مقدار الخطوة stride

P: مقدار الحشو

مثال: تتضمن تطبيق مرشح 3*5*5 على الصورة بدون استخدام حشو وباستخدام STRIDE=1 نحصل على صورة بحجم 3*28*28

مثال: إذا كان لدينا صورة دخل بحجم 3*32*32 و10 مرشحات بحجم 5*5 وباستخدام حشو pad=2 وخطوة stride=1 ما هو شكل المصفوفة الناتجة عن الطبقة الالتفافية وما هو حجم بارامترات الطبقة كاملةً.



Output volume size= NoF*((N−K+2P)/S+1)

Output volume size=(32 + 2×2 - 5)/1 + 1 = 32 spatially

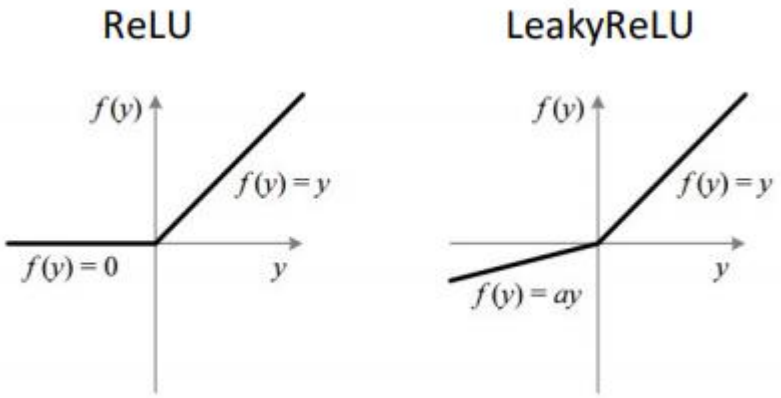
كل مرشح لديه 5*5*3+1 بارامتر = 76 بارامتر (1 للانحياز)

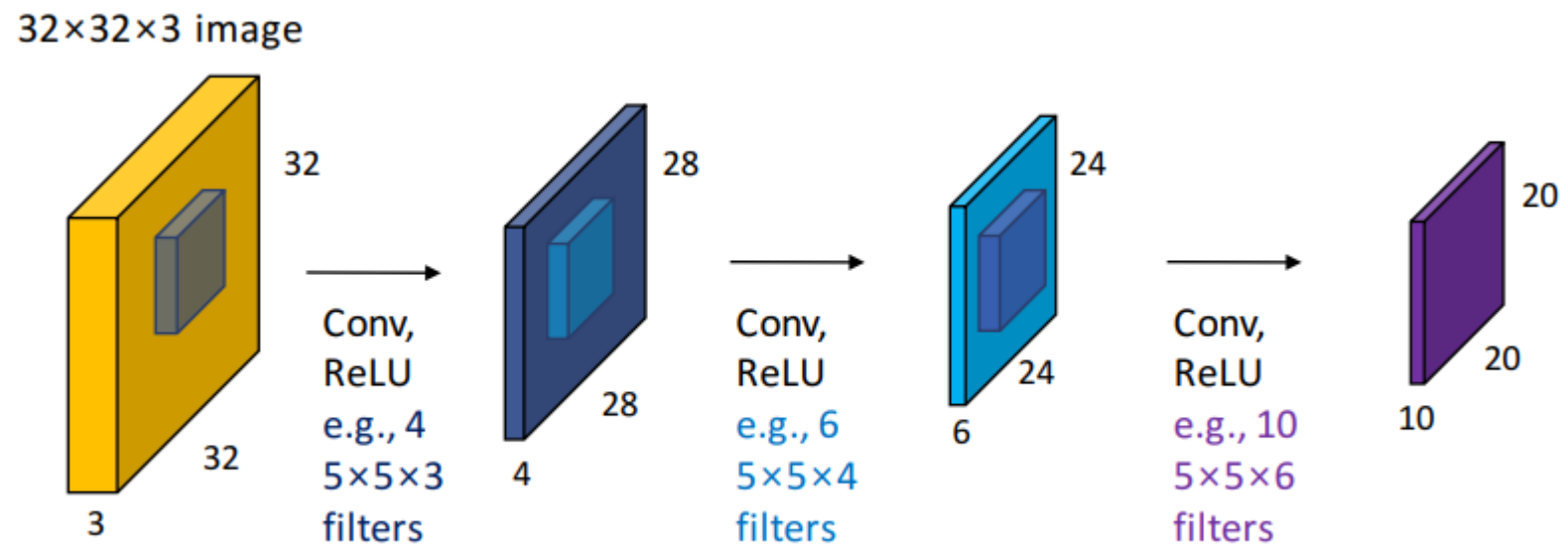
لدينا 10 مرشحات --> حجم البارامترات= 760 بارامتر

الطبقة الالتفافية

Convolutional Neural Networks (CNN)

- **مثال:** وضح شكل شبكة CNN لصورة بدخل 3*32*32 وباستخدام مرشحات 5*5 وعملية انتخاب بخطوة stride =1 بدون حشو، بحيث تتضمن الطبقة الالتفافية الأولى 4 مرشحات والثانية 6 مرشحات والثالثة 10 مرشحات. **(لاحظ لم نأخذ بعين الاعتبار الآن أي طبقة انتخاب)**





التابع غير الخطي المستخدم لإزالة النقاط الضجيجية هو Relu

(32-4,32-4,4)

(28-4,28-4,6)

(24-4,24-4,10)

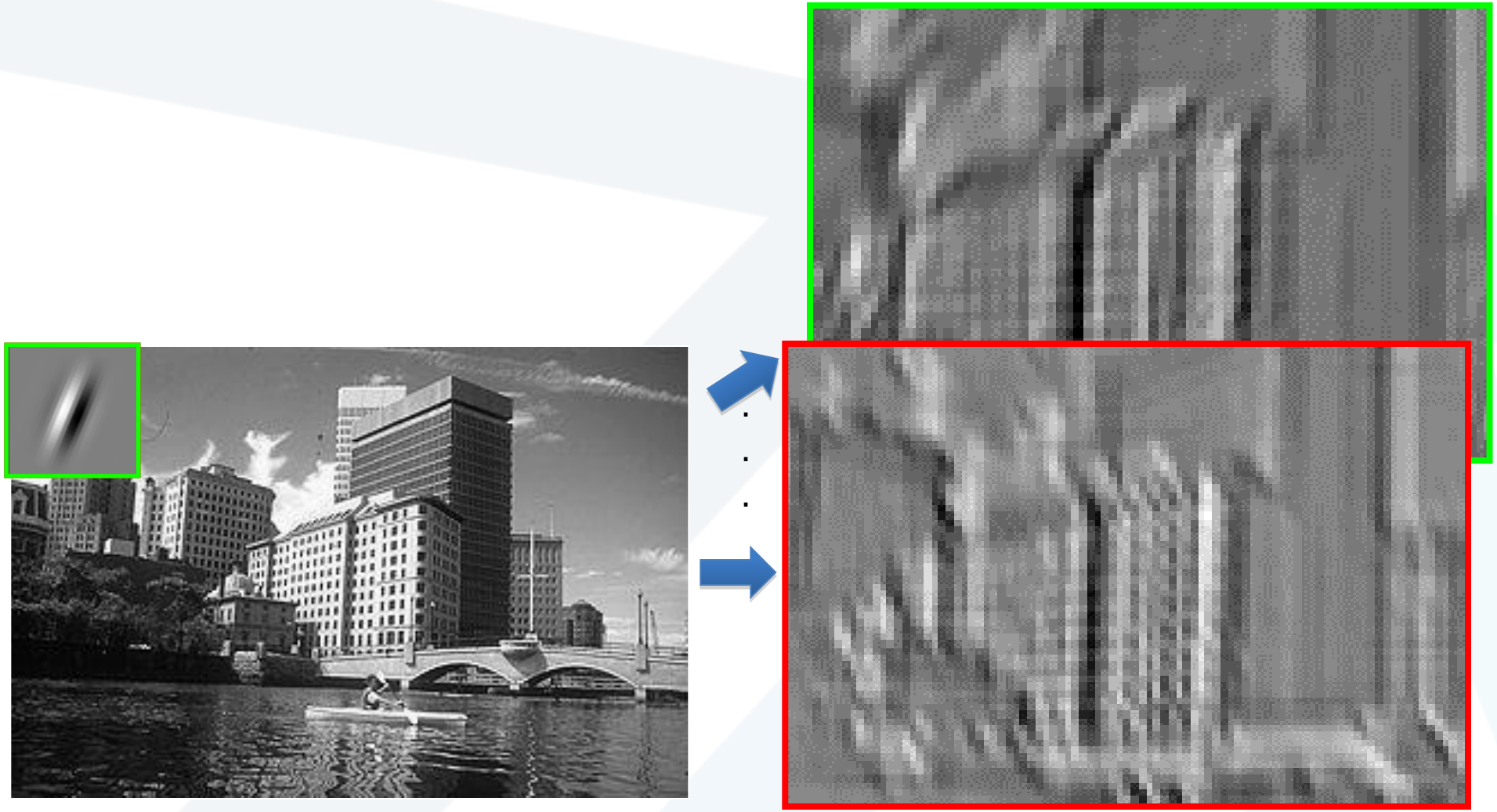
حجم خرج الطبقة= ((N−K+2P)/S+1)

((32-5+0)/1+1)=28

الطبقة الالتفافية

Convolutional Neural Networks (CNN)

- Weighted moving sum



Input

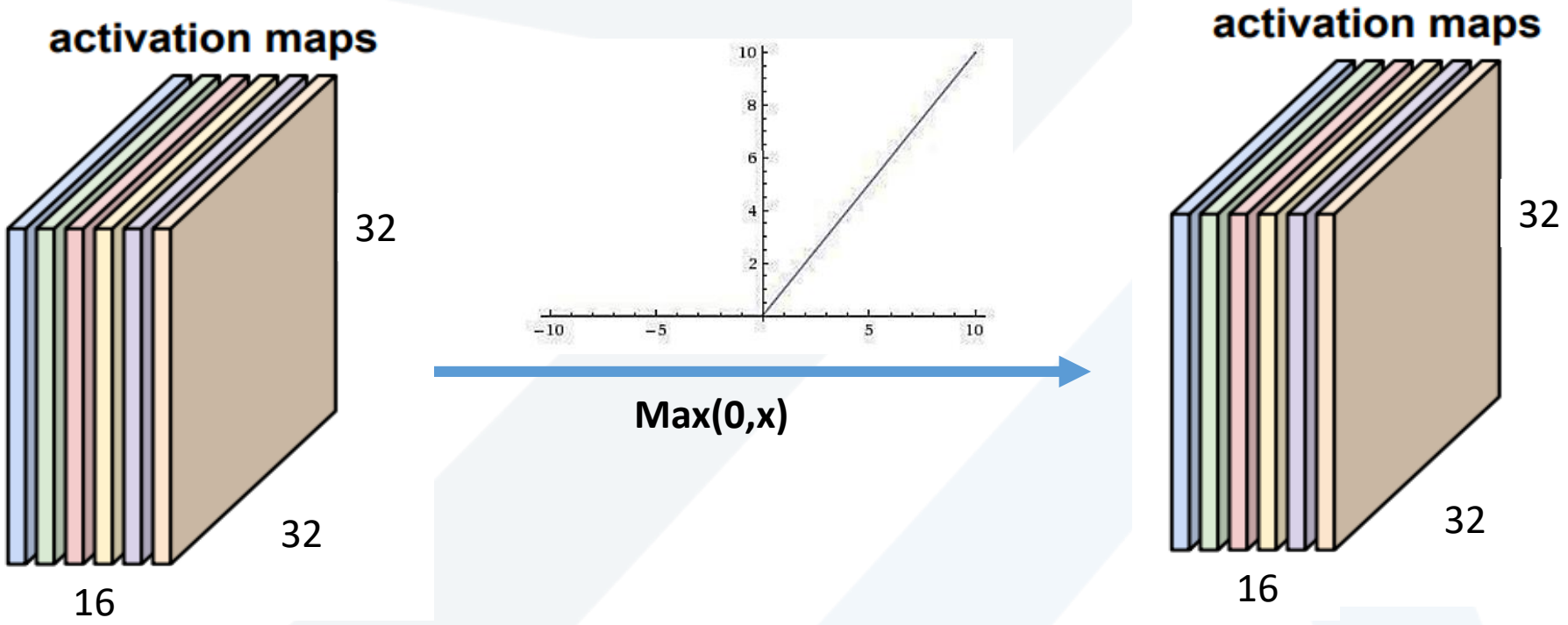
Feature Activation Map

slide credit: S. Lazebnik

الطبقة غير الخطية
Non Linear

Convolutional Neural Networks (CNN)

غالبا في هذه المرحلة يتم استخدام تابع غير خطي وهو RelU وهو أحد توابع التفعيل في الشبكات العصبونية.

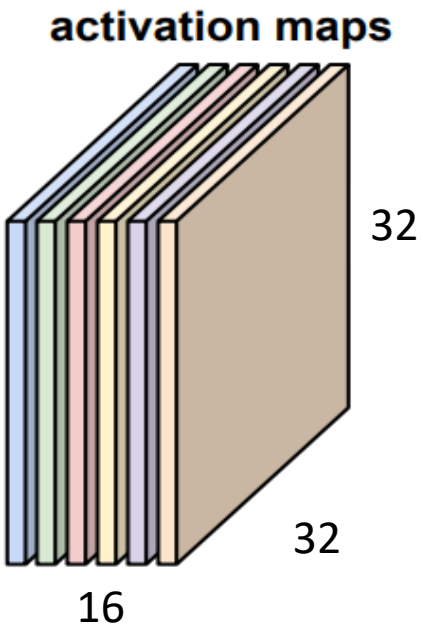


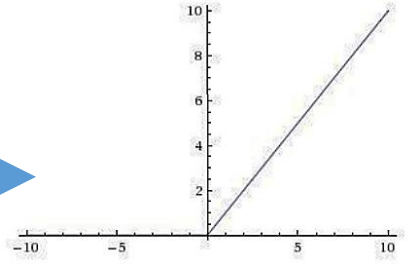
الهدف من هذه العملية: إلغاء القيم السالبة مع تسريع عملية التدريب للمراحل اللاحقة، مشكلته إلغاء بعض القيم التي قد تكون مهمة

طبقة الانتخاب
Max Pooling

Convolutional Neural Networks (CNN)

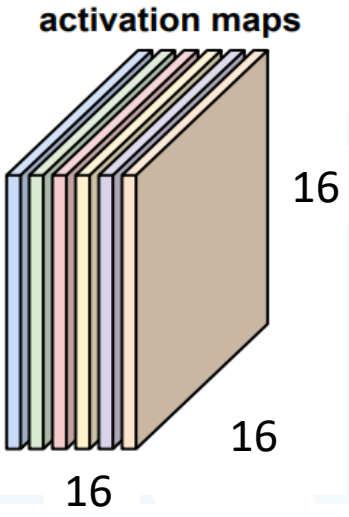
تقوم بعملية Down Sampling أو تخفيض لأبعاد الصورة بطريقة تقلل بيانات الصورة ويمكن استخدام عدة طرق Max Pooling, Average Pooling ....

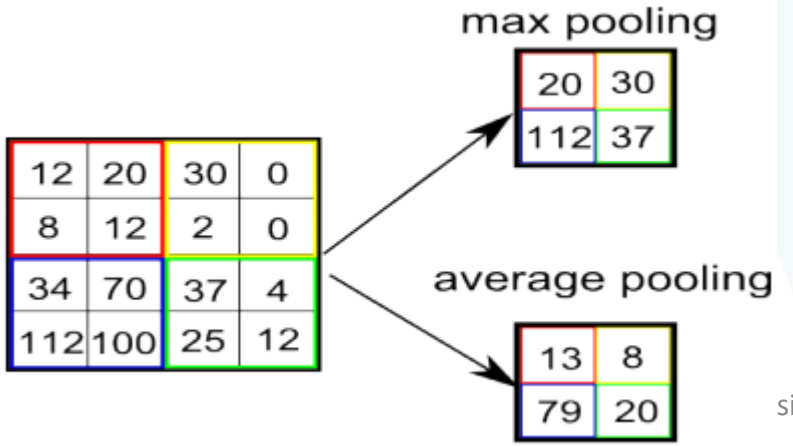




Max(0,x)
RelU

Max- Pooling

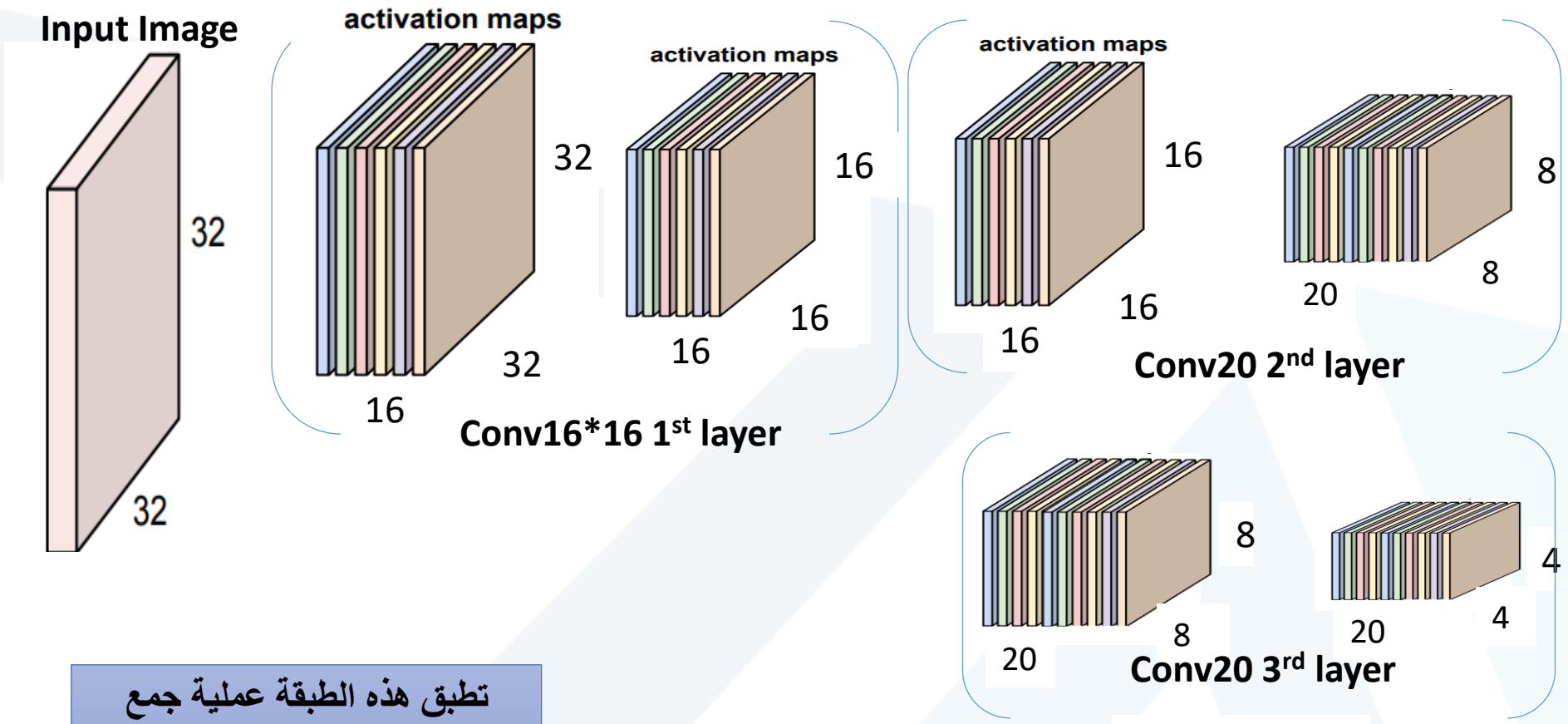




الهدف من هذه العملية: تقليل أبعاد الصورة والحصول منها على أهم المعلومات وجعل السمات منيعة ضد الضجيج والتشويش

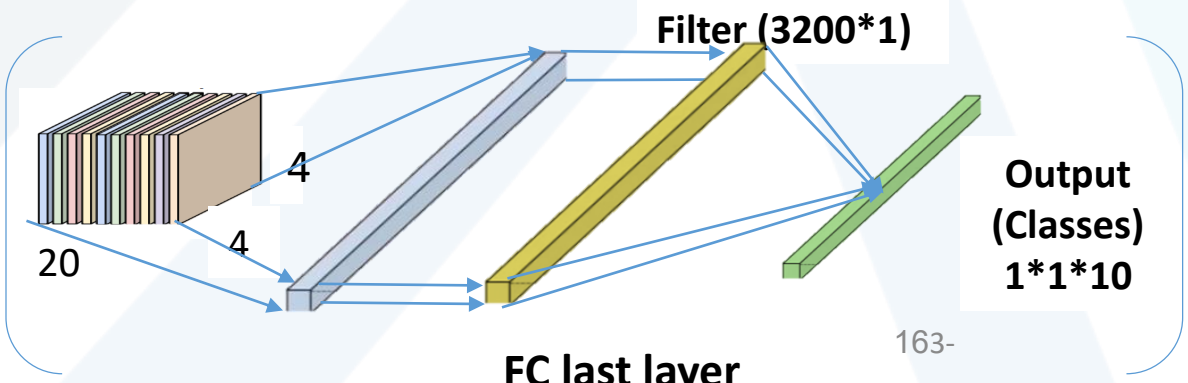
الطبقة كاملة الاتصال
Fully Connected (FC)

Convolutional Neural Networks (CNN)



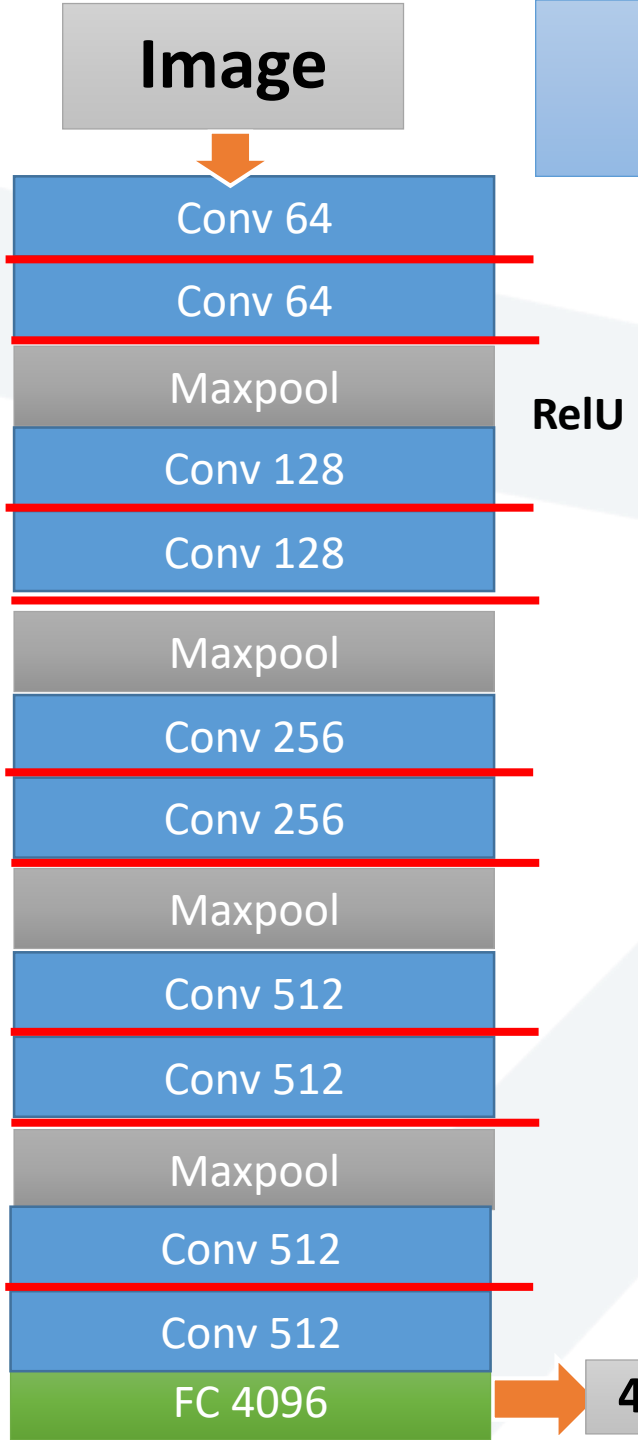
تطبق هذه الطبقة عملية جمع موزون للسمات الناتجة من الطبقة الالتفافية الأخيرة

كل السمات في دخل هذه الطبقة ستشارك بعملية جمع موزون لإعطاء سمة واحدة في الخرج

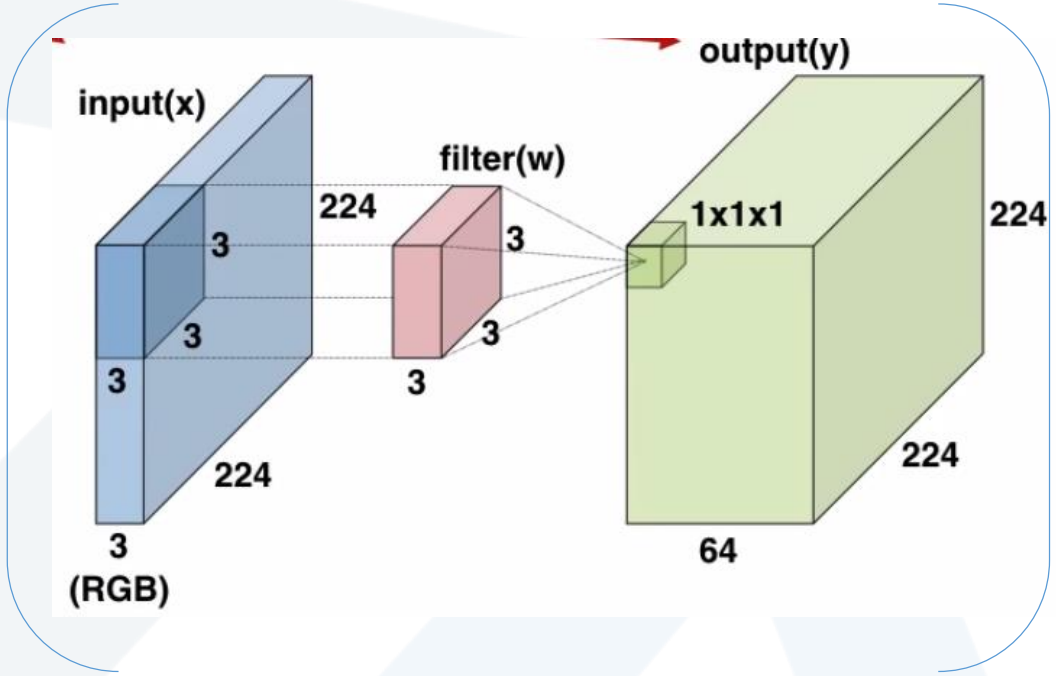


Convolutional Neural Networks (CNN)

نموذج شبكات CNN
2Conv1MaxPool

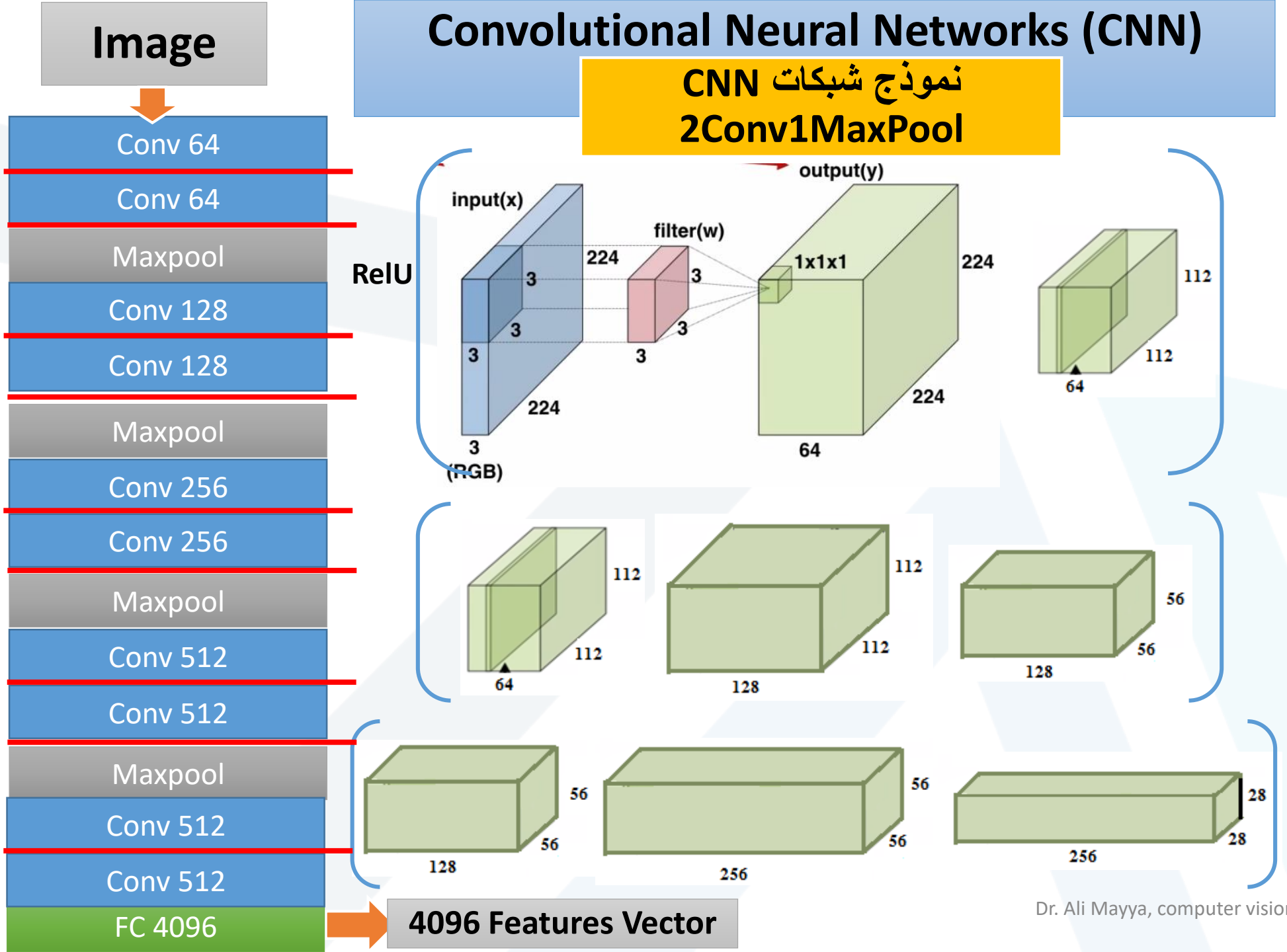


RelU



Convolutional Neural Networks (CNN)

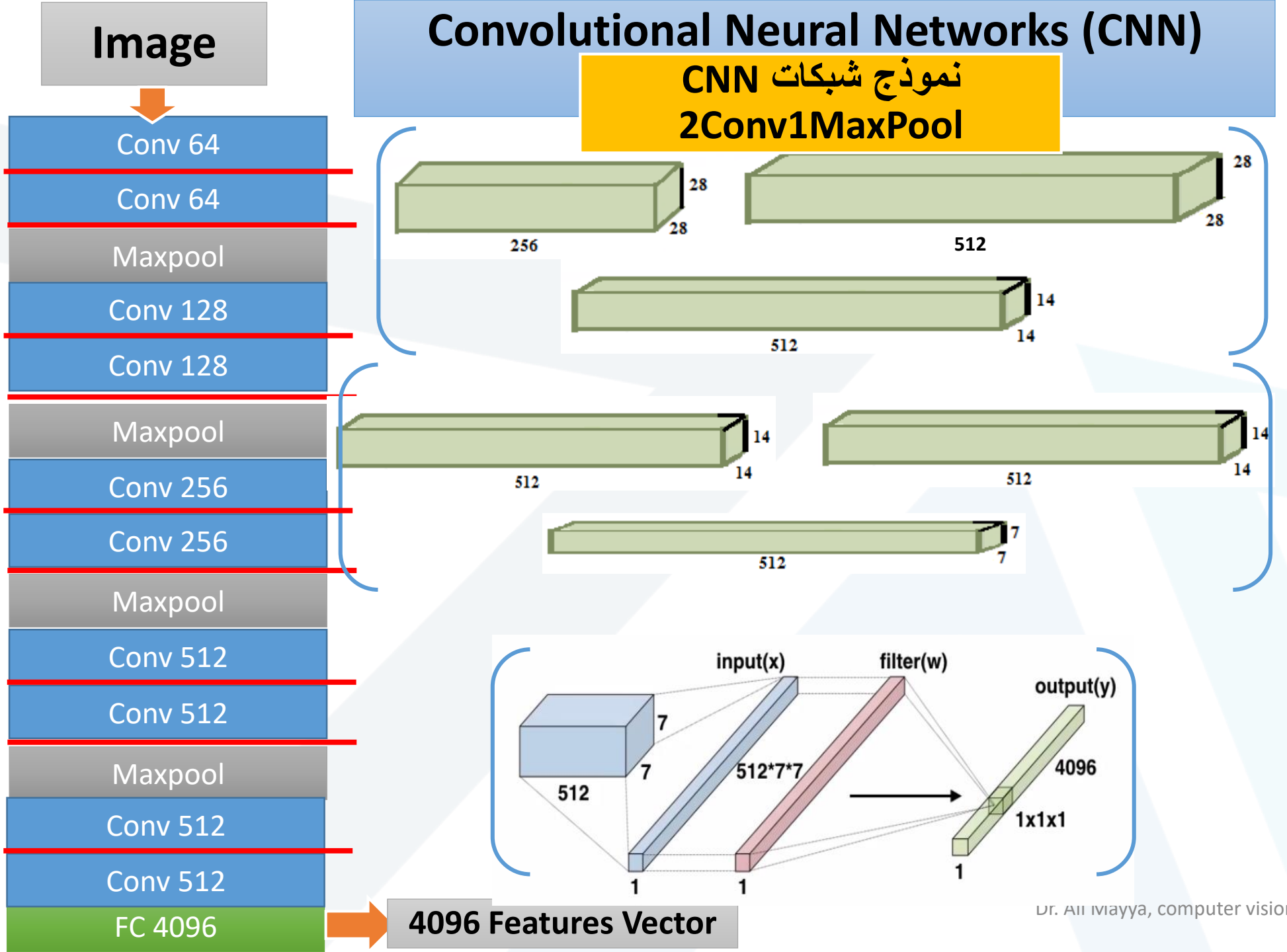
نموذج شبكات CNN
2Conv1MaxPool



Dr. Ali Mayya, computer vision course

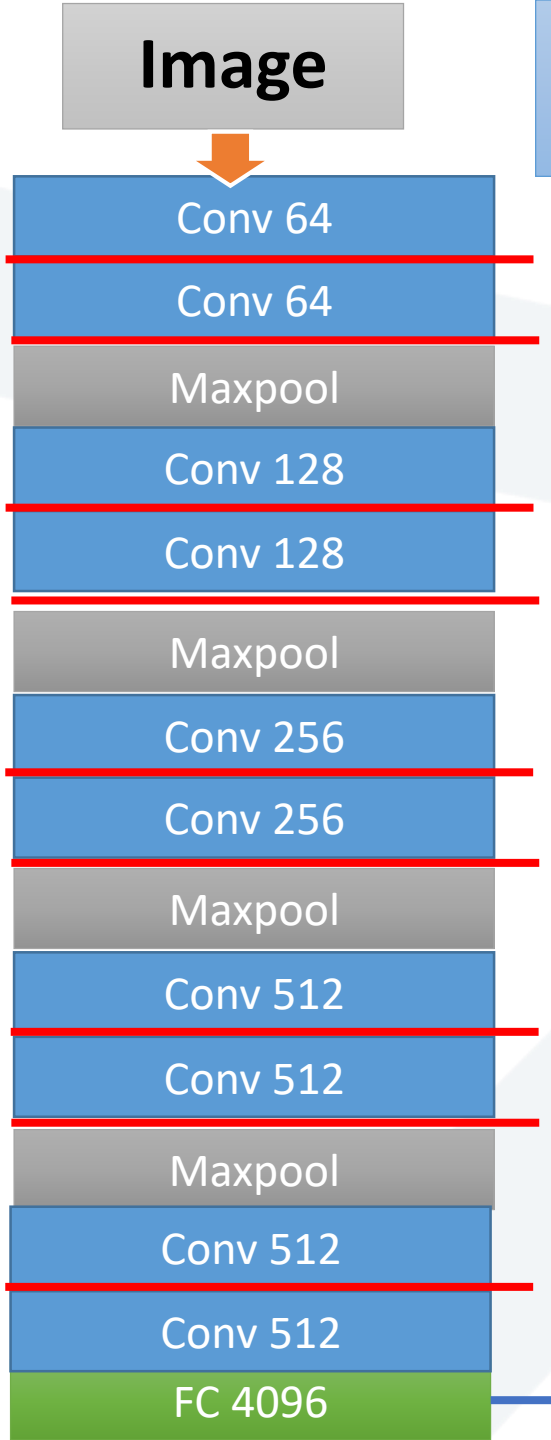
Convolutional Neural Networks (CNN)

نموذج شبكات CNN
2Conv1MaxPool



Convolutional Neural Networks (CNN)

نموذج شبكات CNN
2Conv1MaxPool



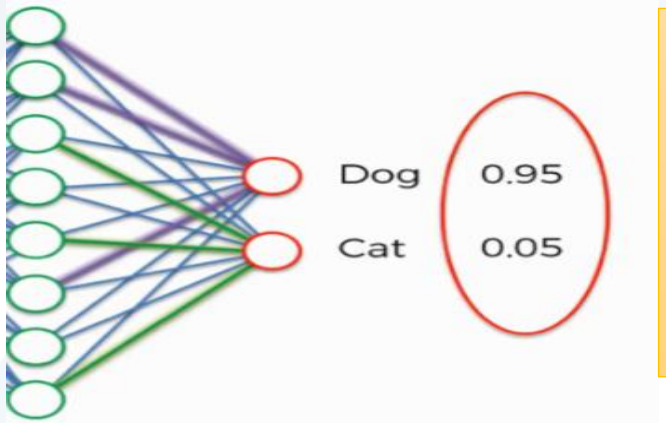
(512*7*7)
25088*1
Vector

Dense Layer
طبقة التخفيض

4096 Features Vector

Softmax Activation Function

Classes
As a probability distribution
الأصناف كتوزع طبيعي (كاحتمالية)

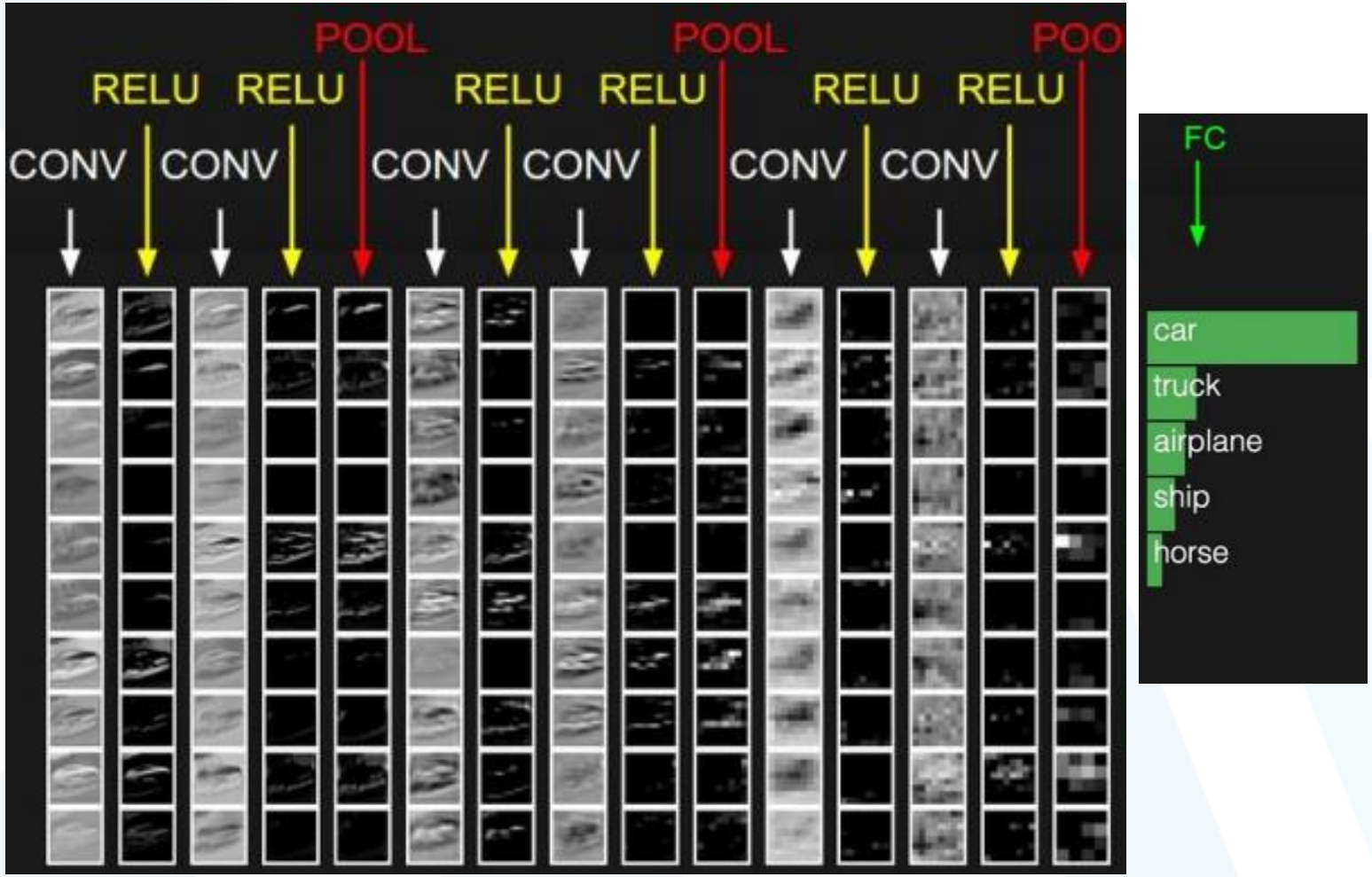


طبقة التخفيض:
تقلل أبعاد الشعاع إلى عدد محدد
يمثل عدد الأصناف المراد التعرف عليها

Convolutional Neural Networks (CNN)

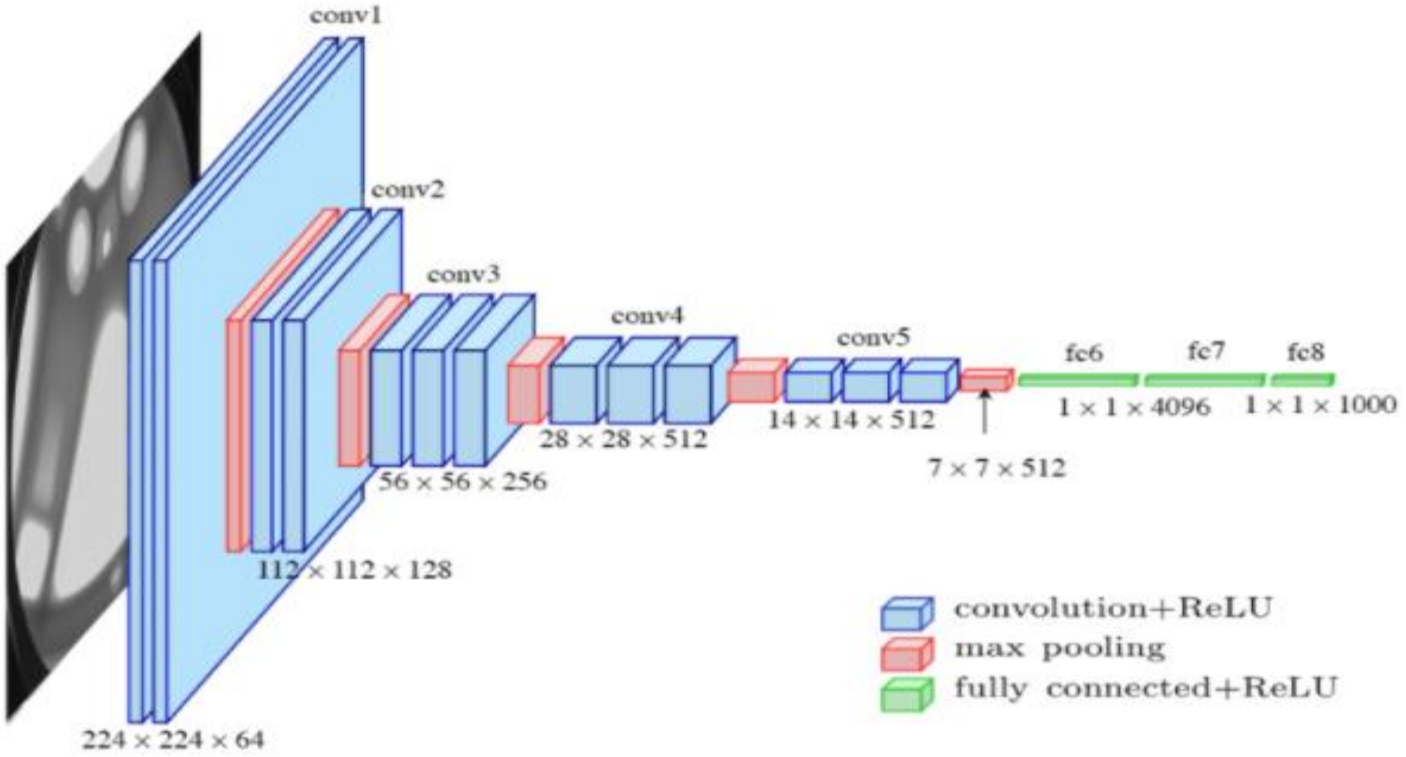
مثال عن نموذج 2Conv1MaxPool

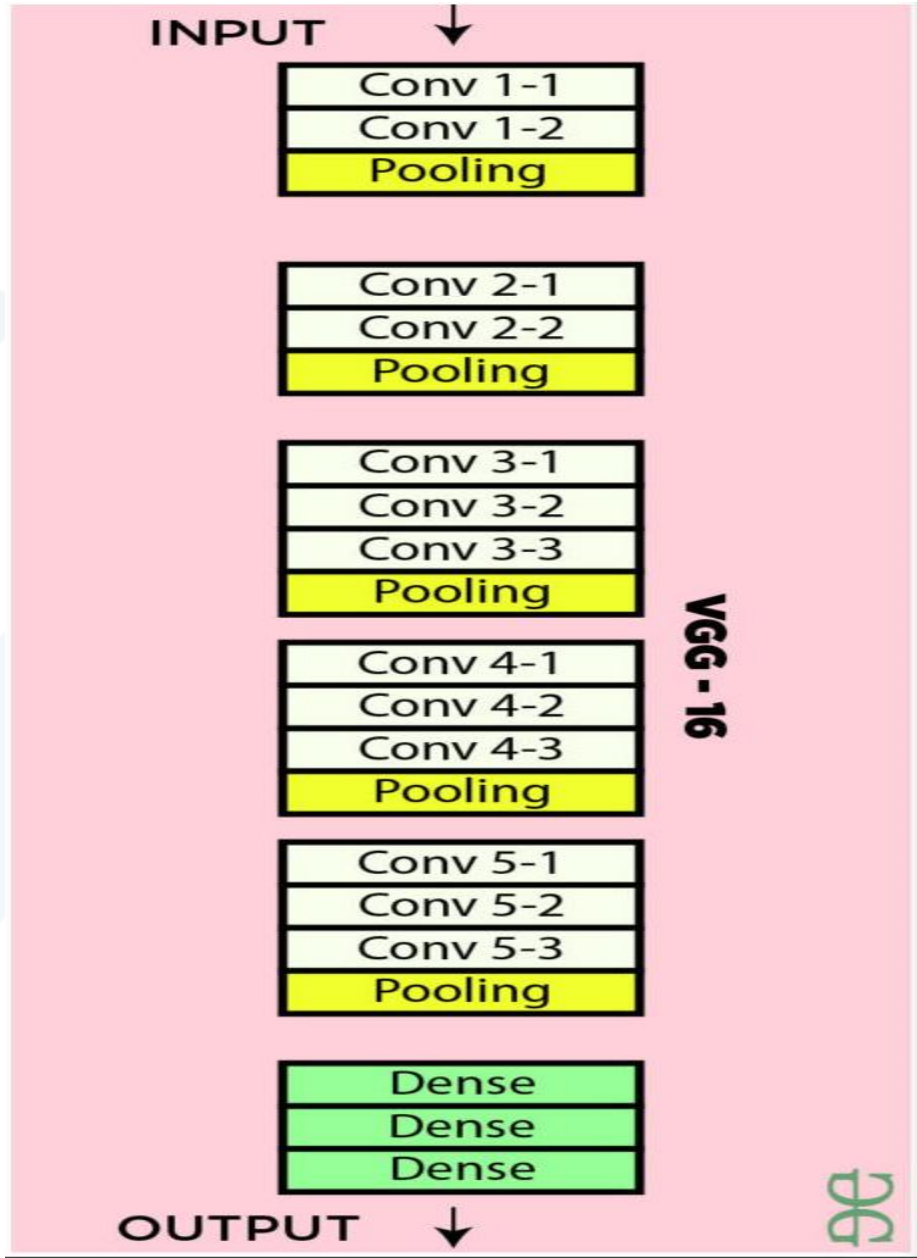
التعرف على مكونات صورة (تمييز صنف سيارة من بين 5 أنماط محتملة)



شبكة VGG

Convolutional Neural Networks (CNN)



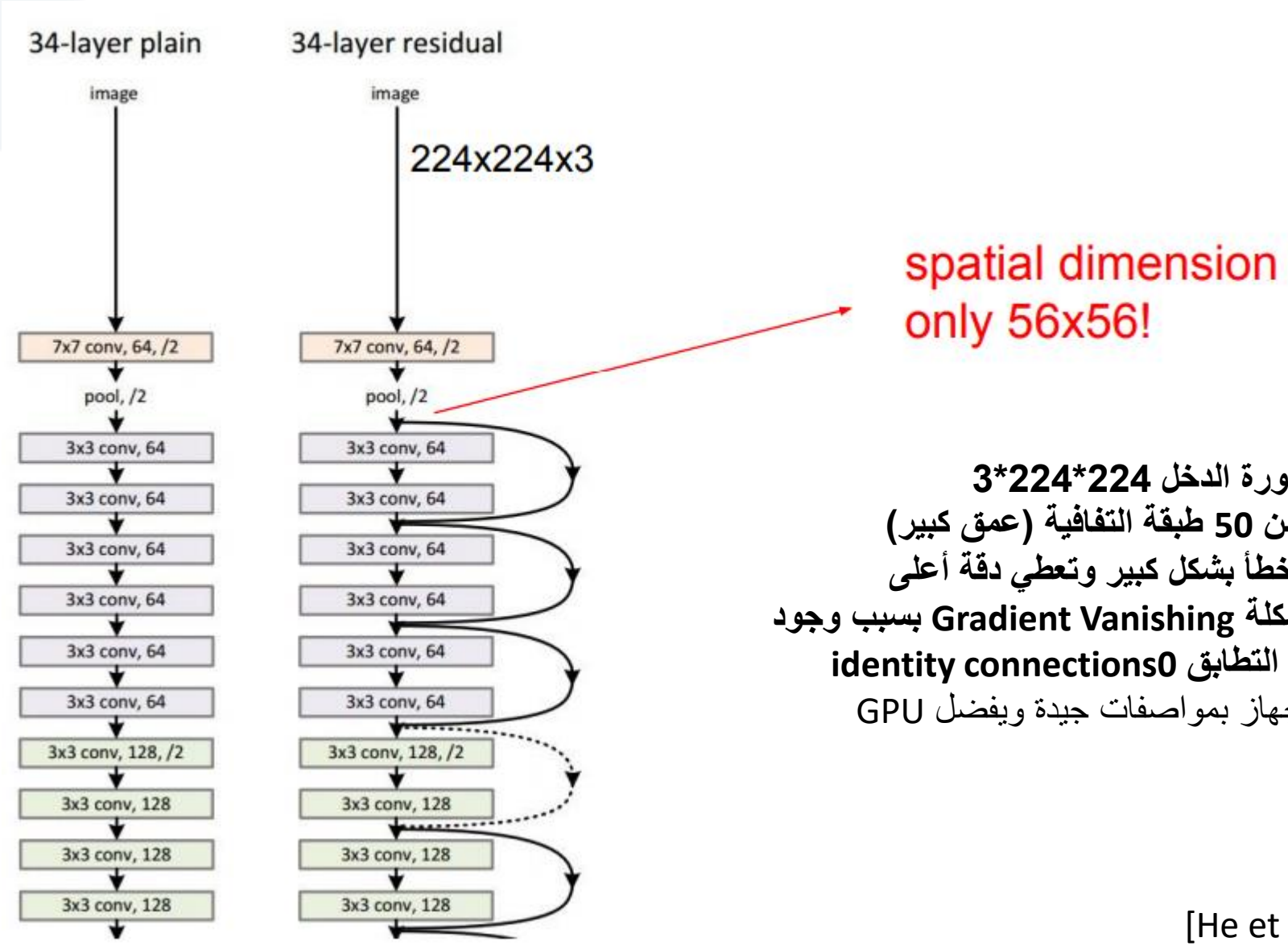


حجم صورة الدخل 3*224*224
حجم شعاع السمات 4096
مدربة على مجموعة بيانات ImageNet التي تتضمن 1000 صنف لذلك
تتضمن طبقة Dense الأخيرة 1000 عصبون
تتميز بأداء عالي لكن زمن تدريب طويل
High Performance, Slow Training

[Simonyan and Zisserman, 2014]

ResNet شبكة

Convolutional Neural Networks (CNN)

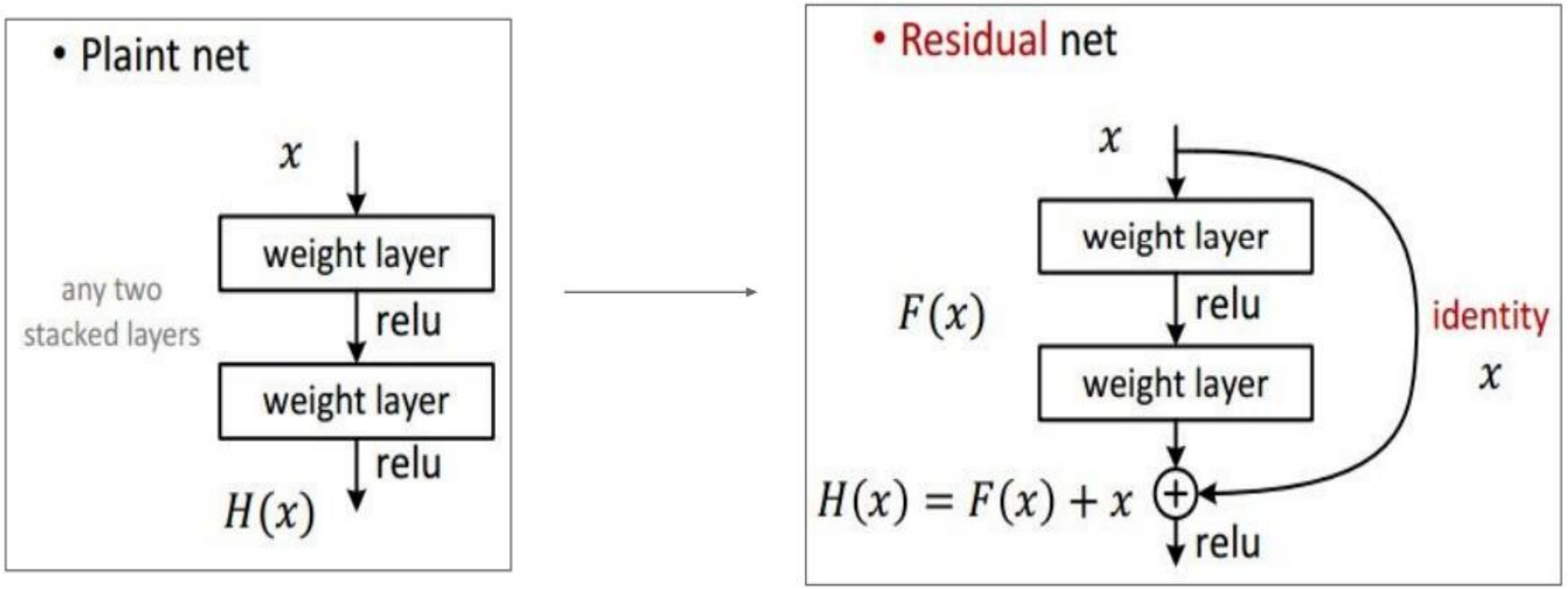


حجم صورة الدخل 224*224*3
مؤلفة من 50 طبقة التفافية (عمق كبير)
تقلل الخطأ بشكل كبير وتعطي دقة أعلى
تحل مشكلة Gradient Vanishing بسبب وجود
وصلات التطابق identity connections0
تحتاج جهاز بمواصفات جيدة ويفضل GPU

[He et al., 2015]

ResNet شبكة

Convolutional Neural Networks (CNN)



يوجد تغذية من الدخل x إلى خرج الطبقة التالية أو التي تليها

يفيد هذا الخط في منع مشكلة Vanishing Gradient والتي تحصل عند زيادة عمق الشبكة حيث تقل قيمة مشتق الخطأ مع زيادة عدد الطبقات حتى تصبح 0 وتتوقف الأوزان عن التعديل، بعد أن كانت عملية زيادة العمق كثيراً تسبب تراجع الأداء قبل شبكة ResNet

يمكن للخط أن يتجاوز طبقة أو أكثر وتضاف قيمته لقيمة خرج الطبقة التي يتجاوزها

مدربة على مجموعة بيانات ImageNet -تتضمن طبقة Average Pooling لها 2048 عصبون ولذا يجب تخفيضها إلى 1000 عصبون في الطبقة كاملة الاتصال (طبقة التصنيف النهائية التي تمتلك تابع تفعيل Softmax)

صممت من قبل Hu بالعام 2015 وربح فيها مسابقة ImgeNet Competition

شبكة EfficientNet

Convolutional Neural Networks (CNN)

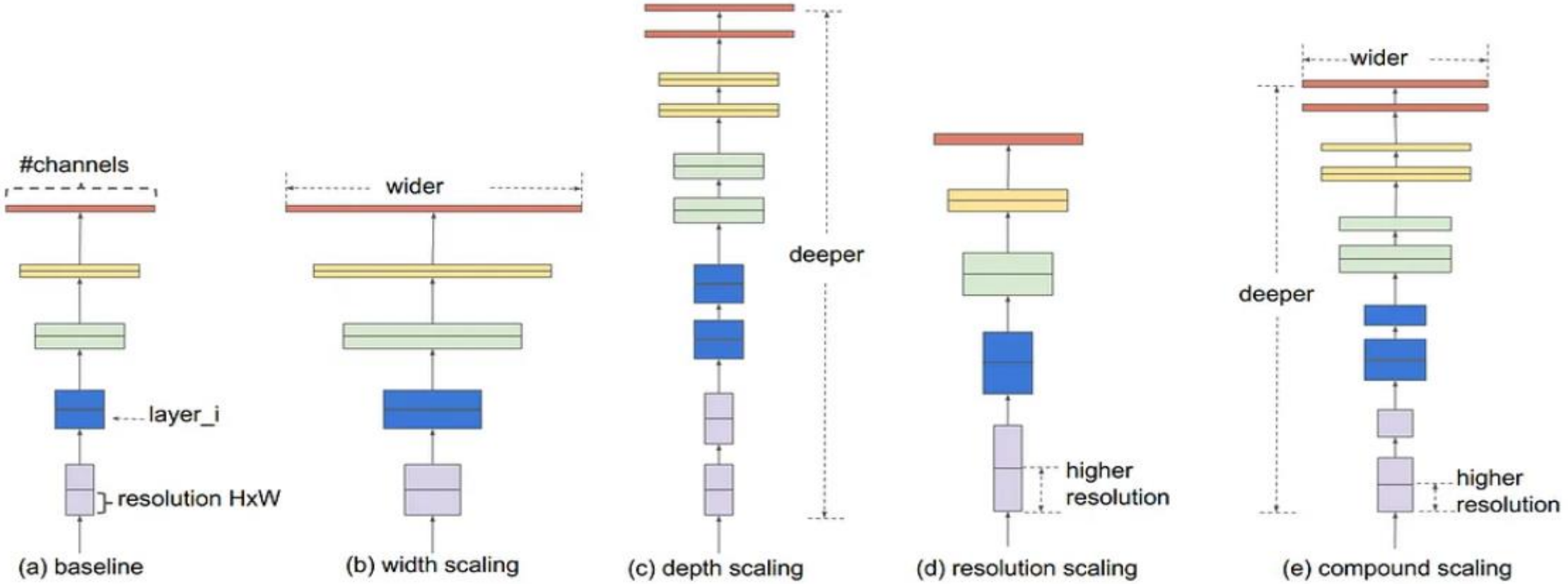
تقوم هذه الشبكة على مبدأ Compound Scaling وهو مؤلف من ثلاث أجزاء:

Width Scaling ويعني تغيير العرض أو عدد المرشحات المستخدمة في طبقة Convolution كلما كان أكثر كلما كان Width أكبر.

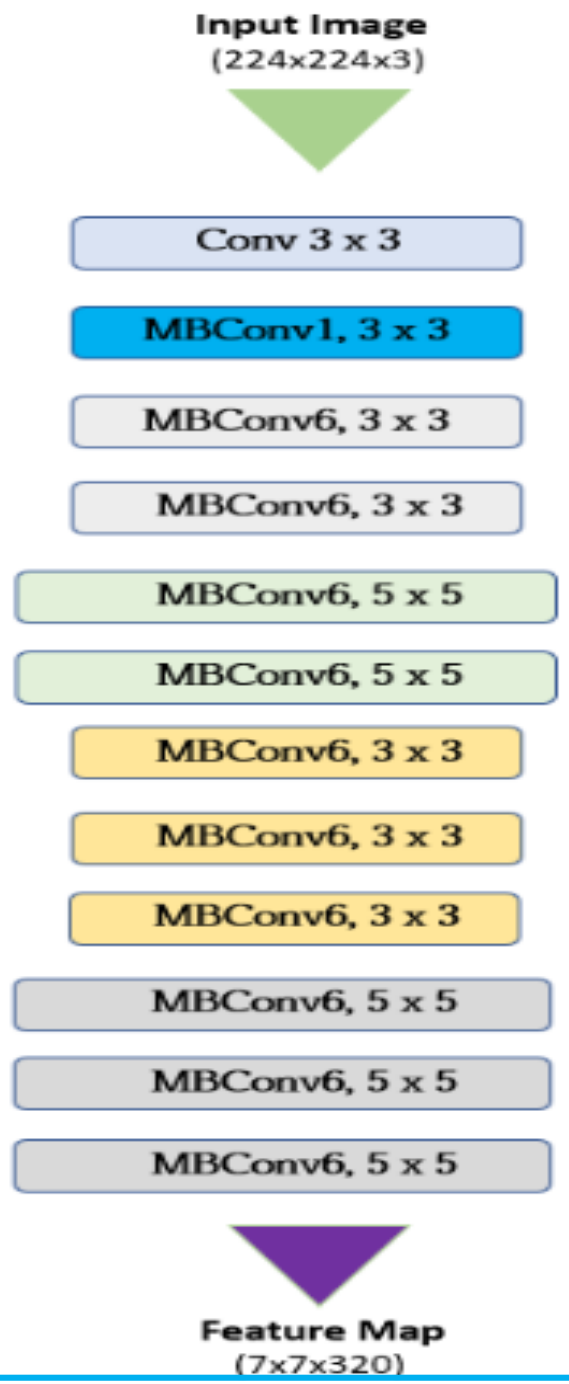
Depth Scaling ويعني تغيير العمق وهو تغيير عدد الطبقات فكلما كان عدد الطبقات أكبر كان العمق أكبر.

Resolution Scaling ويعني تغيير أبعاد الصورة فمع زيادة الأبعاد يزداد معامل Resolution مؤدياً لدقة أعلى وتدريب أبطا.

تقوم هذه الشبكة بتحديد قيم لهذه البارامترات معاً للحصول على أفضل أداء.



شبكة EfficientNet



تتألف بنية شبكة EfficientNetB0 من طبقات MBConv وهي اختصار لعبارة Mobile Inverted Residual Bottleneck Convolution وهي تستخدم مبدأ تدريب يسمى Neural Network Search أو البحث عن البنية الأفضل والأكثر ملاءمة للمشكلة المدروسة.

تتضمن في البداية طبقة Conv3*3

تتضمن بعدها عدة طبقات MBConv

من ثم تتضمن طبقة Conv 1*1 مع طبقة Max Pooling ثم طبقة Average Pooling مؤلفة من 1280 عصبون.

الشبكة مدربة على مجموعة بيانات ImageNet (1000 صنف) لذلك يكون حجم طبقة التصنيف النهائية (الطبقة كاملة الاتصال) 1000عصبون.

تستخدم شبكة EfficientNet :

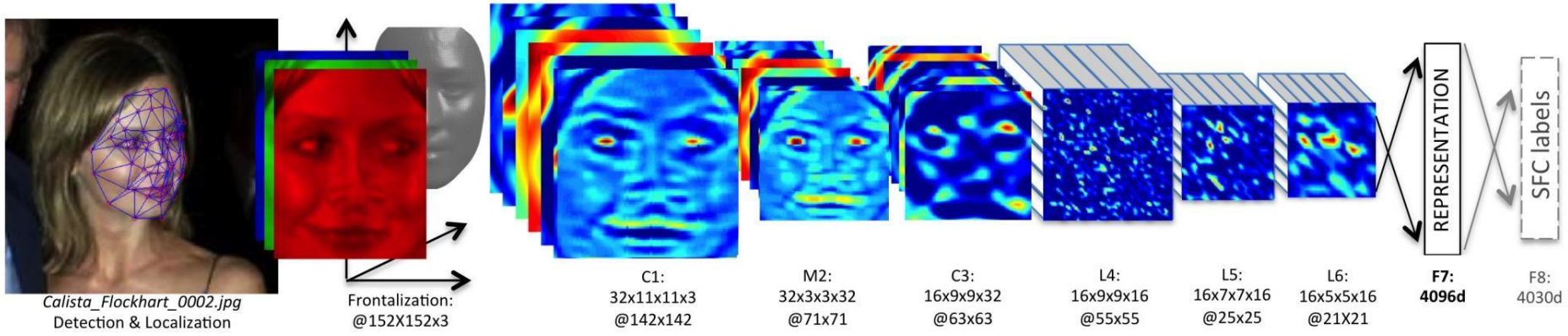
- معامل Width Scaling يسمى phi
- معامل Depth Scaling يسمى Alpha
- معامل Resolution Scaling يسمى rho.
- تقوم من خلالها بضبط بارامترات العرض والعمق والدقة بشكل منتظم باستخدام طبقات MBConv.

تتغلب هذه الشبكة على جميع شبكات CNN السابقة من حيث الأداء.

تطبيقات

Convolutional Neural Networks (CNN)

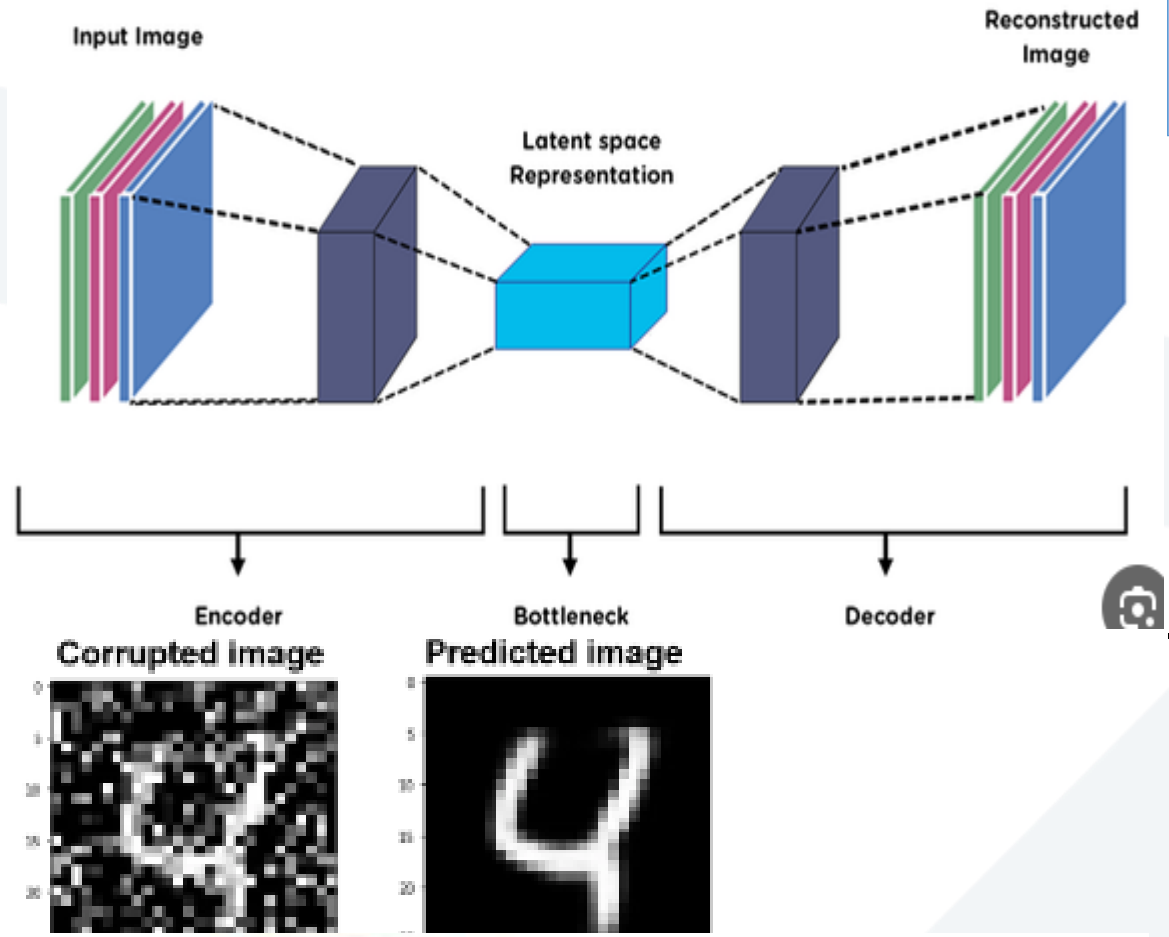
- Used in Facebook, Google, Microsoft
- Image Recognition, Image description and captioning, ….
- Classification application (Face recognition, disease classification, object detection, etc.).
- Detection, tracking, segmentation.

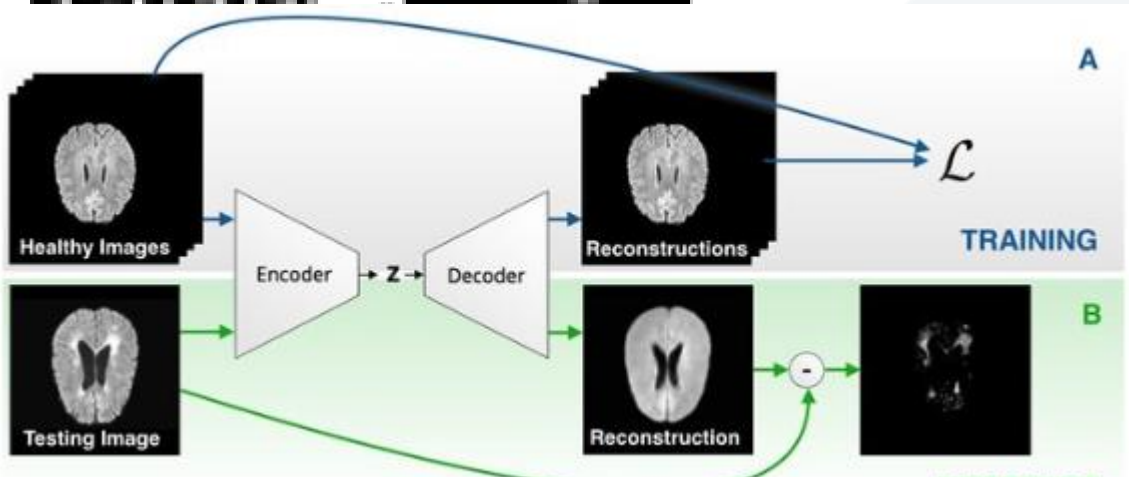


Taigman et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification, CVPR'14

Slide: R. Fergus

Auto-encoder deep models





- شبكات عصبونية عميقة تقوم على مبدأ التعلم بدون مشرف Unsupervised Learning ومؤلفة من جزأين:
- Encoder يستقبل صورة على دخله ويقوم باستخلاص سماتها من خلال عدة طبقات Convolution تليها طبقات تخفيض Pooling
- Decoder يستقبل على دخله خرج Encoder ويقوم بتطبيق عدة عمليات Upsampling لإعادة بناء الصورة بحجمها الأصلي مجدداً.
- تسمى البنية بين الجزأين بجزء Bottleneck
- تستخدم في تطبيقات مثل Anomaly Detection حيث يتم تدريب هذه الشبكات على صور طبيعية لمكونات بدون أي تشوهات أو مشاكل ثم تختبر على صور تتضمن مشاكل أو تشوهات حيث يتم اعتماد عتبة تشابه محددة بين الصورة الناتجة عن الشبكة وبين الصورة الأصلية لتحديد وجود تشوه أو مشكلة في الصورة من عدمها.
- من التطبيقات الأخرى De-noise أو إزالة الضجيج.

References

1. Fei-Fei Li & Andrej Karpathy & Justin Johnson, Convolutional Neural Networks, 2016.
2. EFSTRATIOS GAVVES, Lecture 1: Introduction to Neural Networks and Deep Learning, UVA DEEP LEARNING COURSE – EFSTRATIOS GAVVES.
3. K He, Deep Residual Learning for Image Recognition, https://arxiv.org/pdf/1512.03385, 2015.
4. Simonyan and Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, https://arxiv.org/pdf/1409.1556 , 2014.
5. C. Gonzalez & R. E. Woods,Digital Image Processing 2nd edition, 2001.
6. C. Gonzalez & R. E. Woods,Digital Image Processing 3rd edition, 2008.
7. By Scott E Umbaugh, Computer Imaging: Digital Image Analysis and Processing, 2005.