

8. Visió artificial

Models d'intel·ligència artificial



Introducció

- La visió artificial és una de les àrees més antigues de la intel·ligència artificial.
- Els primers sistemes de visió artificial van ser desenvolupats a la dècada dels 60.
- Els sistemes de visió artificial són capaços d'analitzar imatges i vídeos per tal d'extreure'n informació.
- Veurem quins són els conceptes bàsics de la visió artificial i com s'apliquen en la pràctica.

Visió

- Procés de **percepció**, on el sistema visual és capaç de construir una representació (*imatge*) a partir de la informació captada per la retina.
- Aquest procés pot ser **actiu** (quan l'observador mou els ulls) o **passiu** (quan l'observador no mou els ulls).
- La visió artificial pura és un procés **passiu**, molts conceptes, però, com la **localització** o la **reconstrucció 3D** requereixen un procés **actiu**.

Enfocaments

- Hi ha dos enfocaments principals per a la visió artificial:
 - **Extracció de característiques:**
 - S'apliquen una serie de **transformacions** a la imatge per tal d'extreure característiques rellevants (*vores, textura, fluix òptic, segments, entre d'altres*).
 - **Basat en models:**
 - S'utilitzen models matemàtics (*geomètrics o estadístics*) per tal de representar la imatge.
- En la pràctica, sovint es combinen ambdós enfocaments.

El color (I)

- Propietat de la llum que depèn de la seva longitud d'ona.
- Els humans el percebem el a partir d'unes cèl·lules receptors de la retina: els **cons**.
 - Hi ha tres tipus de cons:
 - **L** (longitud d'ona llarga)
 - **M** (longitud d'ona mitjana)
 - **S** (longitud d'ona curta)
 - Cada tipus de cons és sensible a un rang de longituds d'ona i, per tant, a un rang de colors.

El color (II)

- **Principi de tricromia:** qualsevol color es pot representar com una combinació de tres colors primaris.
- **Colors primaris**
 - Aquells que no es poden descompondre en altres colors.
 - **blau, verd i vermell.**
- Espais de color: RGB, HSV, YUV, ...
- El més utilitzat en visió artificial és el **RGB (Red, Green, Blue).**

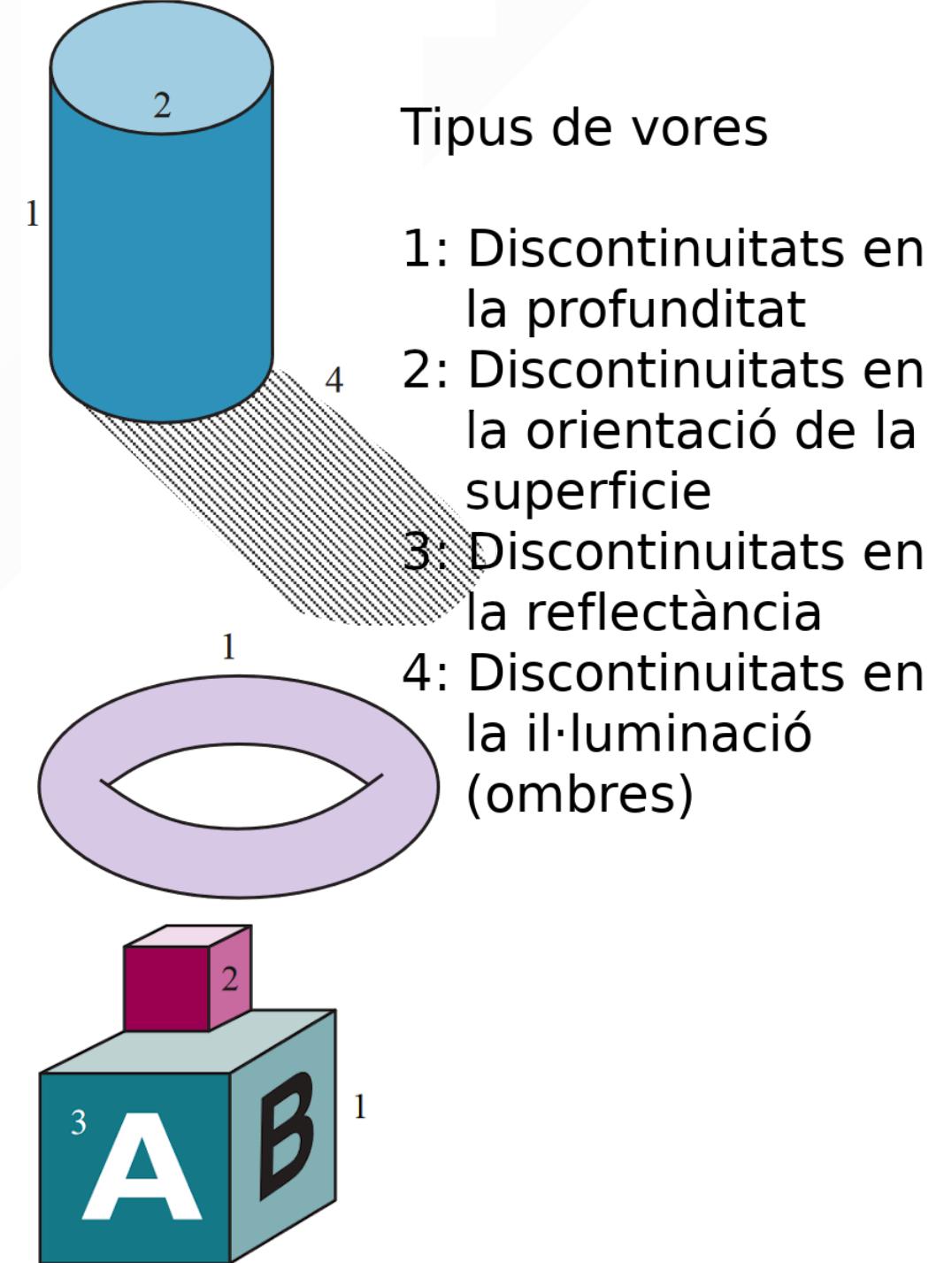
Característiques de les imatges

Definició

- En una imatge hi ha molta informació que no és rellevant.
- Per les tasques de visió artificial es solen utilitzar **característiques** de les imatges.
- Les característiques són aquelles parts de la imatge que són rellevants per a la tasca que es vol realitzar.
- Ens centrem en quatre característiques de les imatges quasi sempre rellevants.
 - *Vores, textura, fluix òptic i segmentació.*

Vores

- Línees que separen regions de diferent intensitat.
- Permeten identificar objectes.
- Simplifiquen la imatge i permeten reduir la quantitat d'informació.
- Passem d'una imatge molt gran a una matriu de vores: **matriu de**



Tipus de vores

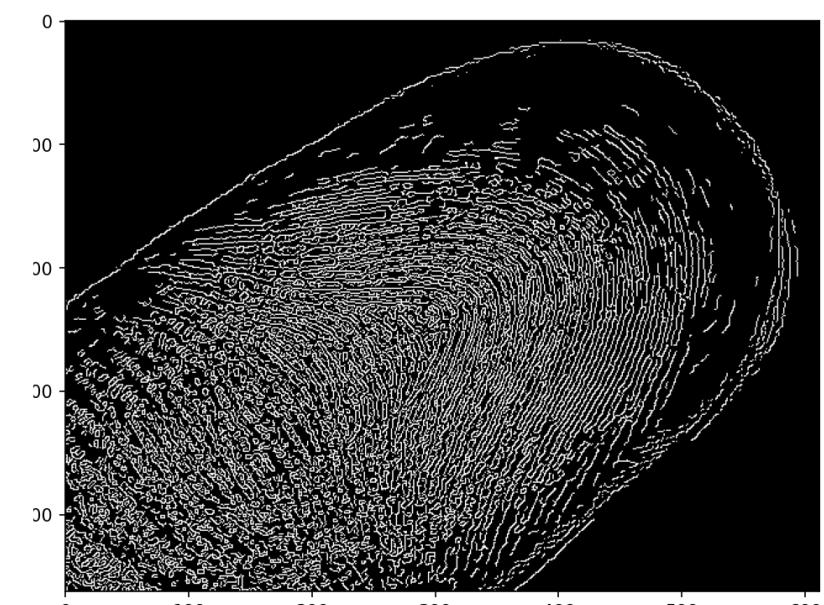
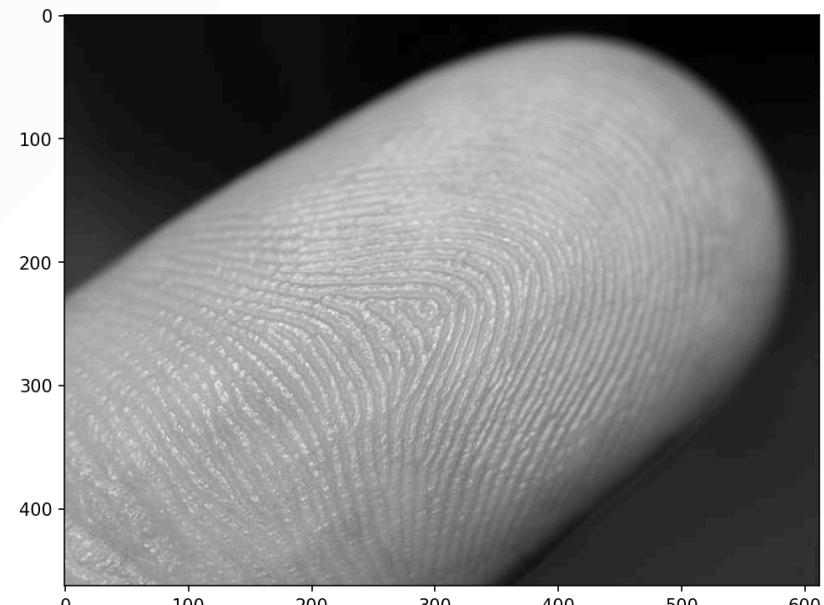
- 1: Discontinuitats en la profunditat
- 2: Discontinuitats en la orientació de la superficie
- 3: Discontinuitats en la reflectància
- 4: Discontinuitats en la il·luminació (ombres)

Detecció de vores

- Tasca de visió artificial que consisteix en detectar les vores d'una imatge.
- Hi ha molts algoritmes per detectar vores, però el més utilitzat és l'algoritme de **Canny**, per John F. Canny, que el va publicar el 1986.
- Objectius:
 - **Bona detecció**: detectar totes les vores.
 - **Bona localització**: les vores han de ser el més pròximes possible a les vores reals.
 - **Minimitzar les respostes falses**: no detectar vores on no n'hi ha

Algoritme de Canny

- Fa servir quatre passos per detectar les vores d'una imatge:
 1. Es redueix el soroll amb el **filtre de Gauss**.
 2. Calcula el gradient de la imatge amb el **filtre de Sobel**.
 3. Es detecten les vores amb el **mètode de supressió de non-màxims**.
 4. Es decideixen quines vores són vàlides amb el **mètode de la**



Textura

- En visió artificial entenem com a textura un **patró de píxels** que es observable en una imatge.
 - Ex: Finestres en un edifici, taques en una vaca, etc.
- Ajuden, al igual que les vores, a **identificar objectes**.

A



B



B'



Característiques de la textura

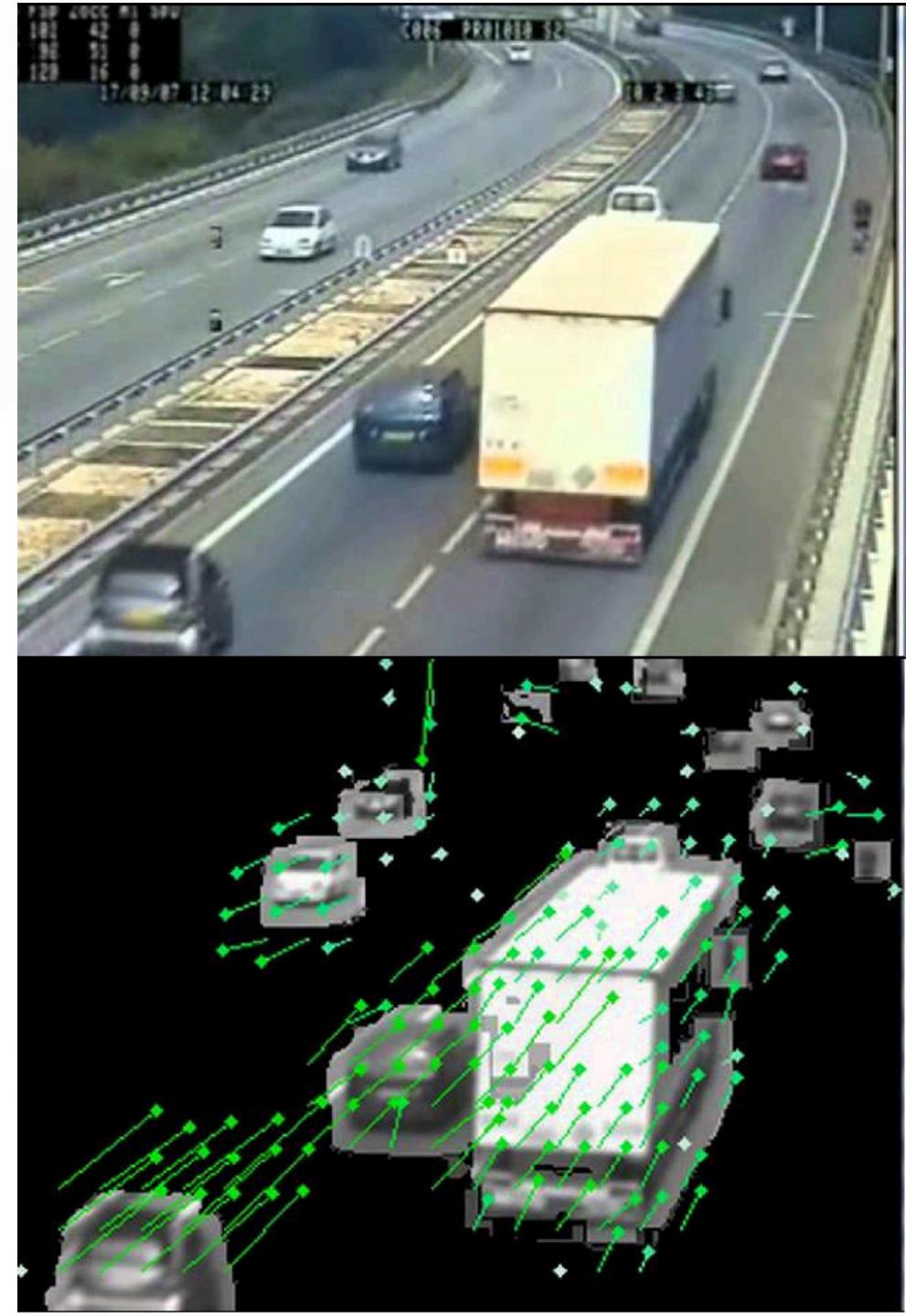
- La textura pot ser més o menys regular, per lo que es freqüent utilitzar un model de **tesel·les** per descriure-la.
Vejam algunes característiques:
 - **Tesel·la:** patró que es repeteix en una imatge.
 - **Tesel·lació:** procés de cobrir una superfície amb tesel·les.
 - **Tipus:**
 - **Regulars:** es repeteixen sempre de la mateixa manera.
 - **Irregulars:** no hi ha un patró clar de repetició.
 - **Escala:** la textura pot ser més o menys gran.

Utilitats de la textura

- **Identificació:** permet identificar objectes. Ex: un cavall té una textura diferent a la d'una zebra.
- **Correspondència:** permet trobar zones corresponents en diferents imatges. Important en la reconstrucció 3D.
- **Segmentació:** permet separar la imatge en diferents regions.
- **Reconstrucció:** permet reconstruir la imatge a partir de les tesel·les.
- **Classificació:** permet classificar objectes.

Fluix òptic

- El **fluix òptic** és la **velocitat aparent** amb la que es mouen els objectes entre dues imatges.
- Els algoritmes de visió artificial són capaços de calcular el fluix òptic a partir de diferents imatges.
- El fluix òptic és important per moltes tasques, com poden ser la **reconstrucció 3D**, la **compensació de moviment** o la **compressió**.



Segments

- Anomenen **segments** a les **regions** de la imatge que tenen alguna propietat comuna (color, textura, forma, etc.).
- Per definit els segments hi ha dós enfocaments principals:
 - **Basat en límits:** es busquen els límits de les regions. Es pot entendre com un problema de *classificació* on cada pizel pertany o no a un segment i es soluciona amb tècniques de machine learning i models preentrenats.
 - **Basat en regions:** s'agrupen els pixels en regions segons alguna propietat comuna. Es pot entendre com un problema de *b* i es soluciona amb tècniques com k-means, etc.

Tipus de segmentació

- Tipus de segmentació:
 - **Segmentació binària:** es segmenta la imatge en dues regions: objecte i fons.
 - **Segmentació semàntica:** es segmenta en categories predefinides.
 - **Segmentació d'instàncies:** es segmenta en instàncies d'objectes.
 - **Segmentació panòptica:** es segmenta en categories predefinides, però també es segmenten les instàncies d'objectes.



(a) Image



(b) Semantic Segmentation



(c) Instance Segmentation



(d) Panoptic Segmentation



Tasques de visió artificial

en triunfo
zepo. De formule ervan
chemt en daardoor zorgt

0630(appel non surtaxé)
regal.com

Glycerin Hexanedioic Behenyl
Acid SabaKemel Extract
Citrae Cetearyl Glucoside
Lium
Benzf

Tasques

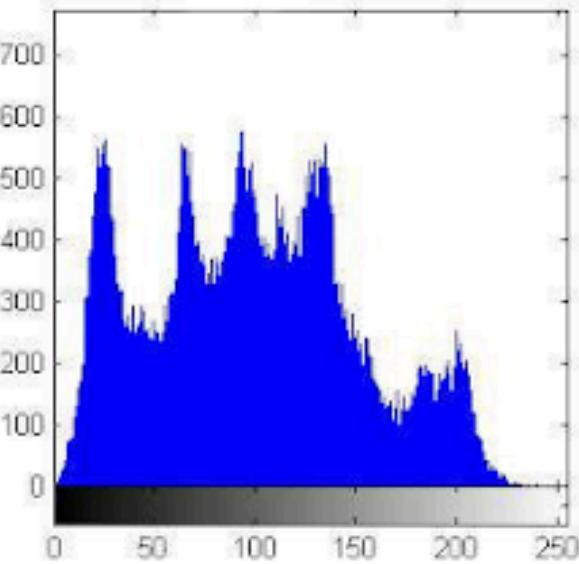
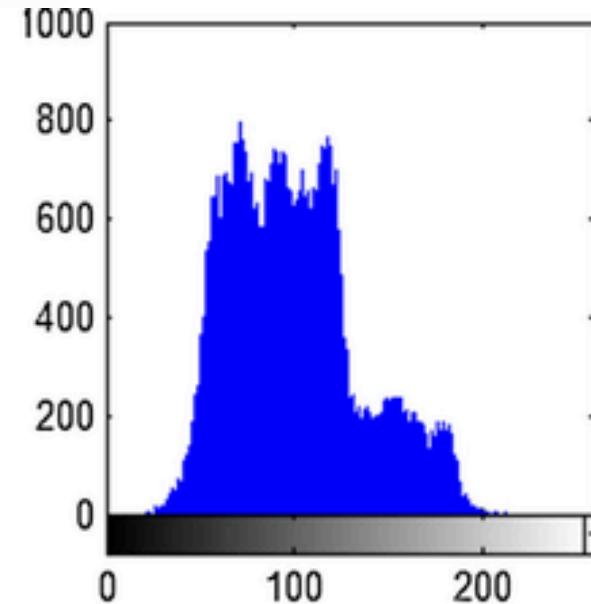
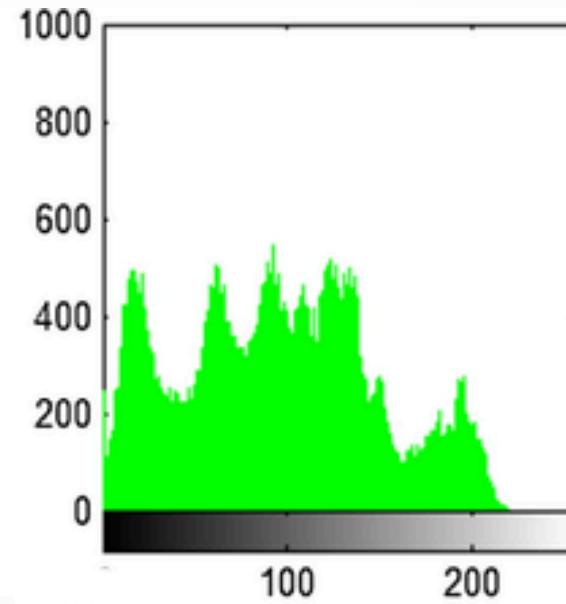
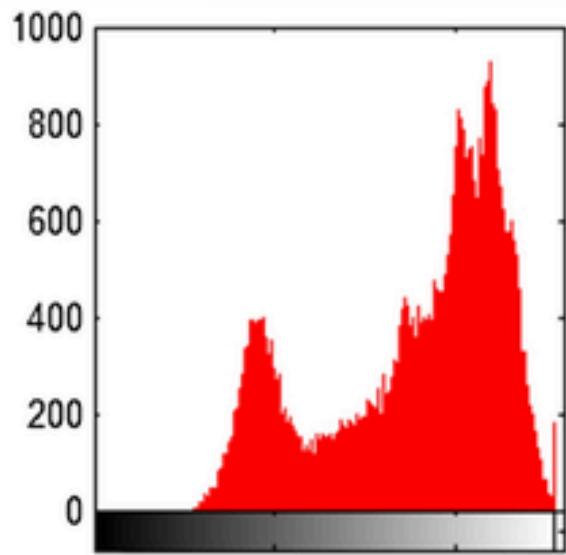
- Les tasques de visió artificial són aquelles que es poden realitzar a partir d'imatges.
- Veurem algunes de les més importants:
 - **Preprocessament d'imatges**
 - **Classificació d'imatges i Reconeixement d'objectes**
 - **Reconstrucció 3D**
 - **Localització**
 - **Segmentació**
 - **Reconstrucció**

Processament d'imatges

- El **processament** d'imatges és el conjunt de tècniques que s'apliquen a les imatges per tal de millorar-ne la qualitat o per tal d'extreure'n informació.
- Històricament, el processament d'imatges era la única forma de obtindre resultats en visió artifical, amb l'aparició de les xarxes neuronals, però, aquesta tasca ha perdut importància.
- Tot i això, segueix sent una tasca important en visió artificial, especialment en tasques de visió artificial més tradicionals o **quan no hi ha GPUs disponibles**.
- Veurem algunes de les tècniques més comunes.

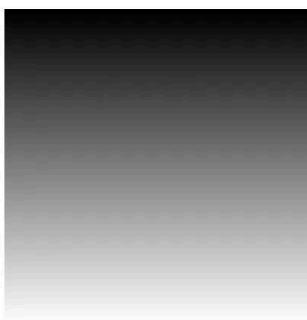
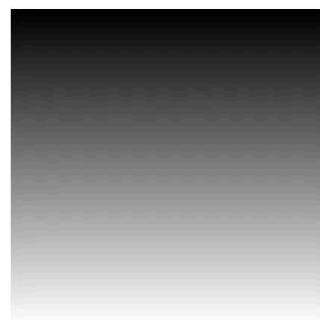
Histogrames

- El **histograma** d'una imatge és la representació gràfica de la distribució dels píxels en funció de la seva intensitat.
- Els histogrames són molt útils per entendre la distribució dels píxels en una imatge.
- Son molt utilitzats en el preprocessament d'imatges per tal de normalitzar-les.
- Els histogrames es poden calcular per cada canal de color (R, G, B) o per la imatge en escala de grisos.
- S'utilitzen molt en la **normalització** d'imatges.



Equalització de l'histograma

- L'**equalització de l'histograma** és una tècnica que es fa servir per tal de millorar el contrast d'una imatge.
- L'objectiu és que la distribució dels píxels sigui més uniforme.
- Es divideix l'histograma en *bins* i es redistribueixen els píxels de manera que la distribució sigui més uniforme.
- Els efectes moltes vegades no són realistes, però si solen ser útils per a tasques de visió artificial.



Filtratge

- El **filtratge** és una tècnica que es fa servir per tal de millorar la qualitat de la imatge.
- Hi ha molts tipus de filtres, però els més comuns són els filtres de **suavitzat** i els filtres de **realçament**.
- Els filtres de suavitzat són útils per tal de reduir el soroll de la imatge.
- Els filtres de realçament són útils per tal de millorar el contrast de la imatge.
- Els filtres es poden aplicar a tota la imatge o a una regió concreta.

Filtres de suavitzat

- El soroll és un problema comú en les imatges.
- Podem reduir el soroll de la imatge aplicant filters de suavitzat.
- Els filters més comuns són el **filtre de mitjana** i el **filtre de Gauss**.
- Filtre de mitjana: substitueix cada píxel per la mitjana dels píxels del seu entorn.
- Filtre de Gauss: substitueix cada píxel per la mitjana ponderada dels píxels del seu entorn.
 - Els píxels tenen un pes més gran com més propers estan al píxel central.

Filtres de realçament

- Els filters de realçament són útils per tal de millorar el contrast de la imatge. Molt utilitzats en la detecció de vores.
- Els filters més comuns són:
 - **Filtre de Sobel:**
 - Calcula el gradient de la imatge, és a dir, la intensitat de canvi de la imatge.
 - **Filtre de Laplace:**
 - calcula el laplacià de la imatge, és a dir, la segona derivada de la imatge.



Thresholding

- El **thresholding** és una tècnica que es fa servir per tal de binaritzar una imatge.
- El thresholding es fa aplicant un **llindar** a la imatge.
- Els píxels que tenen una intensitat superior al llindar es converteixen en blancs i els que tenen una intensitat inferior es converteixen en negres.
- És una forma simple de **segmentació**: es vol separar la imatge en dues regions: *objecte* i *fons*.

Global Thresholding

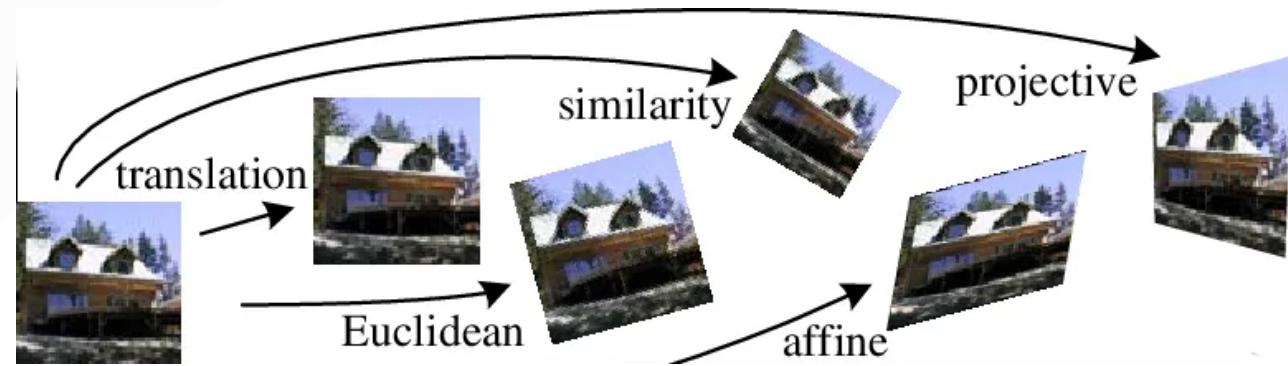


Adaptive Thresholding



Transformacions

- Les **transformacions** són tècniques que es fan servir per tal de canviar la forma de la imatge.
- Les transformacions més comunes són: **rotació, escala, desplaçament i canvis de perspectiva.**
- Es divideixen en **lineals** i **no lineals**: segons si canvien la forma de la imatge.



Extracció del fluix òptic (*optical flow*)

- L'extracció del fluix òptic és pot fer amb diferents tècniques, però es poden dividir en dos grans grups:
 - **Discrets:** es calcula el fluix òptic per punts concrets de la imatge.
 - L'algorisme més comú és el de **Horn-Schunck**.
 - Més ràpid que el dens, però menys precís.
 - **Denses:** es calcula el fluix òptic per cada píxel de la imatge.
 - Els algorismes més comú son el de **Lucas-Kanade** i el de **Farnebäck**.

Extracció del fluix óptic (*optical flow*)



(a) Sparse Optical Flow – Lukas Kanade



(b) Dense Optical Flow - Gunnar Farneback

Llibreries

- Hi ha moltes llibreries que es poden fer servir per tal de fer el preprocessament d'imatges.
- Les més comunes són:
 - **OpenCV**: llibreria de visió artificial i machine learning.
 - Per visió artificial, és la més utilitzada.
 - **Pillow**: llibreria de processament d'imatges.
 - **Scikit-image**: llibreria de processament d'imatges.
 - **Mahotas**: llibreria de processament d'imatges.
 - **SimpleCV**: llibreria de visió artificial.

Classificació d'imatges i reconeixement d'objectes

- Aquestes tasques consisteixen en **identificar** els objectes que hi ha a la imatge.
- La **classificació d'imatges** consisteix en **identificar** l'objecte que hi ha a la imatge.
- El **reconeixement d'objectes** consisteix en **identificar** els objectes que hi ha a la imatge i **localitzar-los**.
- Ambdues tasques són molt importants en visió artificial i són la base de moltes aplicacions.

Classificació d'imatges

- La majoria de sistemes actuals de classificació d'imatges es basen en l'**aparença** (textura, color, forma, etc.) de l'objecte; però, hi ha sistemes que també fan servir l'**estructura** de l'objecte.
- Dues dificultats principals:
 - **Variabilitat de l'objecte**: els objectes poden tenir moltes aparences diferents (dos gossos poden ser molt diferents).
 - **Variabilitat de la imatge**: la mateixa imatge pot tenir moltes aparences diferents (llum, ombra, etc.).
- Les xarxes neuronals convolucionals són les més utilitzades per aquesta tasca.

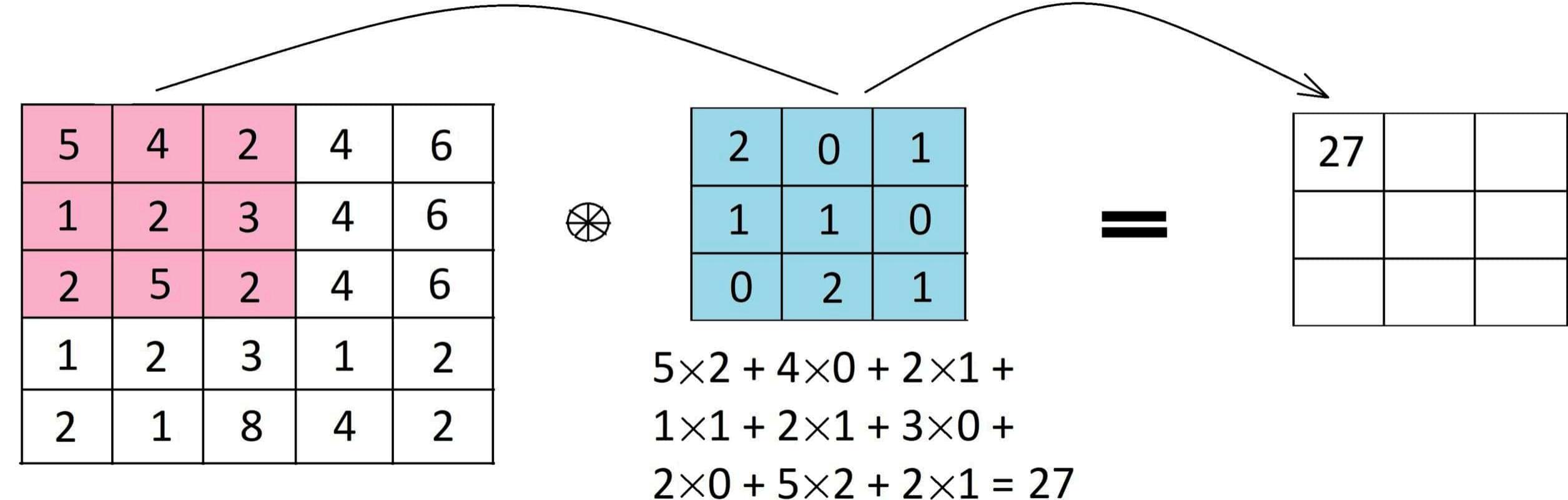
Xarxes neuronals convolucionals

- Les **xarxes neuronals convolucionals** (XNN) són un tipus de xarxes neuronals, especialment dissenyades per processar dades en forma de matrius; com poden ser les imatges.
- Les xarxes neuronals convolucionals són molt bones per a tasques de classificació d'imatges.
- Com la resta de xarxes neuronals, les xarxes neuronals convolucionals necessiten ser entrenades amb moltes dades numèriques.
- Veurem a continuació com es passarán les imatges per la xarxa.

Convolució

- La **convolució** permet reduir la quantitat d'informació de la imatge i ens permetrà enviar a la xarxa solament les **característiques més rellevants**.
- Aquest procés millora la precisió de la xarxa i la fa més ràpida.
- La convolució es fa amb **filtres** que es van aplicant a la imatge.
- Els filters soLEN ser matrius de mida petita (3×3 , 5×5 , etc.).
- El resultat de la convolució es una **imatge més petita** que l'original, anomenada **mapa de característiques**.
- Si no volem reduir la mida de la imatge, podem fer servir

Convolució



Funcions d'activació

- Després de la convolució, s'aplica una **funció d'activació**.
- Les funcions d'activació són funcions que apliquen una **no linealitat** a la imatge.
- La més utilitzada en xarxes neuronals convolucionals és la **ReLU**. Els valors negatius es converteixen en zero i els positius es mantenen igual.
- La funció d'activació és molt important per tal de que la xarxa mantingui la **capacitat de generalització**.
- Després de la funció d'activació, es pot aplicar un **pooling**.

Pooling

- El **pooling** és una tècnica que es fa servir per tal de reduïr la mida de la imatge encara més.
- Hi ha diferents tipus de pooling, però el més comú és el **max pooling**.
- Es sol utilitzar una finestra de mida petita (2×2 , 3×3 , etc.) i es pren el valor màxim de la finestra.
- El resultat és un **mapa de característiques poolat**. Aquest mapa de característiques es passarà a la següent capa.
- El pooling obliga a la xarxa a ser **invariant a petites transformacions**.

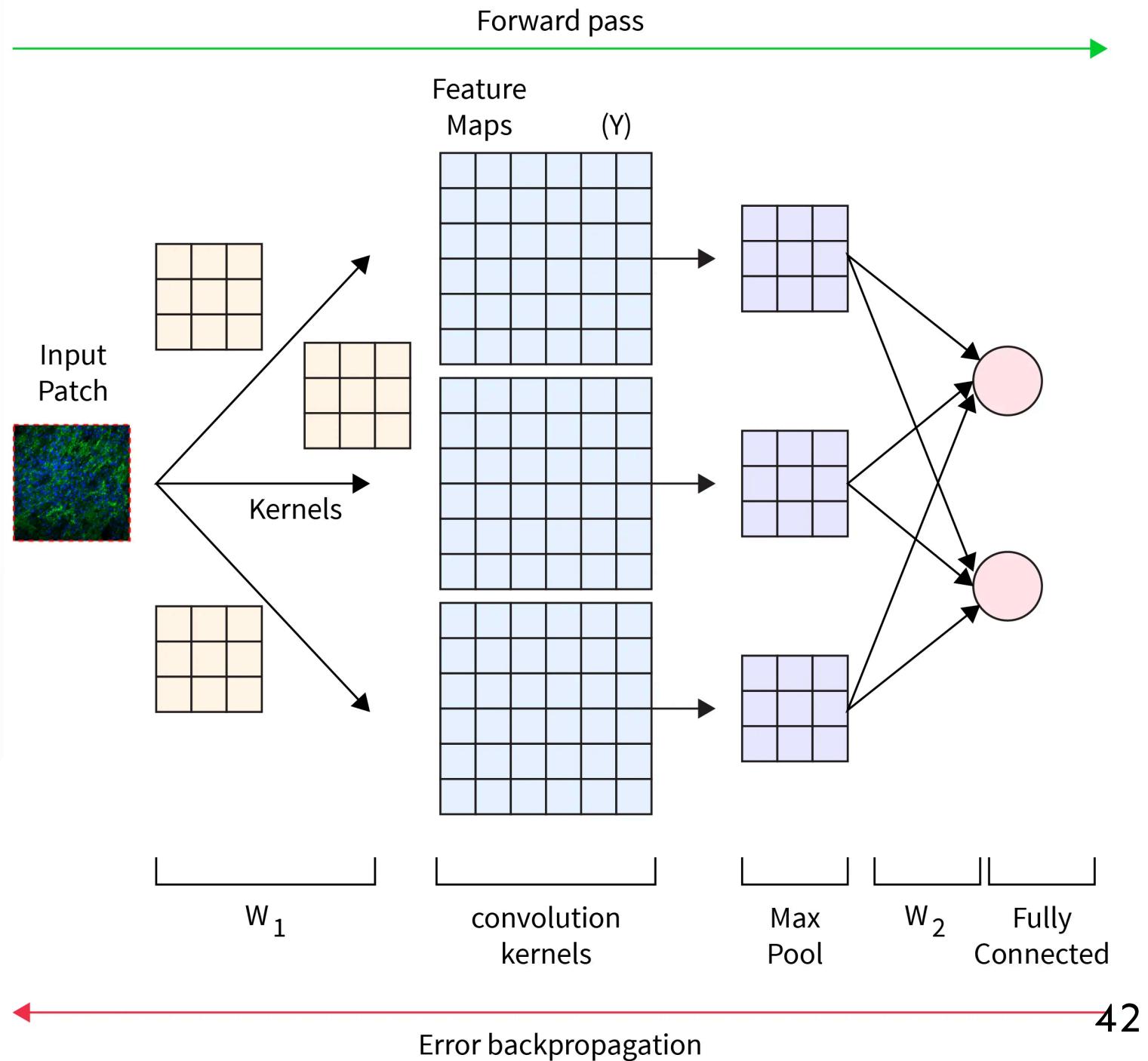
Regularització

- Després de les capes de convolució i pooling, es solen afegir capes de **regularització**.
- Les capes de regularització són capes que ajuden a la xarxa a **generalitzar**.
- Les capes de regularització més comunes són les capes de **dropout**.
- Aquestes capes eliminan un percentatge de les neurones de la xarxa, fent que no s'actualitzin en cada iteració.
- Això fa que la xarxa no es **sobreajusti**.

Aplanament i capes totalment connectades

- Entre les capes de regularització i les capes totalment connectades, es sol fer un **aplanament**.
- L'aplanament és el procés de convertir el **mapa de característiques** en un **vector**.
- Aquest vector es passarà a les capes totalment connectades.
- Les capes totalment connectades són les capes que es fan servir per tal de **classificar** la imatge.
- Aquestes capes són les que es fan servir per tal de **reduir la dimensió** del vector de característiques.

Estructura d'una xarxa neuronal convolucional



Funcionament d'una CNN (I)

- En les imatges els pixels individuals no tenen gaire sentit
 - Sabem que un 8 tindrà pixels negres en la part central però no sabem exactament on.
- Els patrons locals si que poden ser importants
 - Sabem que el 0 i el 8 tenen cercles, el 1 i el 7 tenen línies verticals, etc.
- Les relacions entre patrons també son interessants
 - El 1 té dues línies, el 6 una línia i un cercle, etc.
- Estratègia general: **extreure patrons locals i després combinar-los per extreure patrons més globals**

Funcionament d'una CNN (II)

- Les xarxes neuronals convolucionals (CNN) són una forma de fer això
 - Una capa està formada per una convolució + ReLU
 - La convolució mesura la similitud entre un filtre i la finestra. Cada filtre detecta un patró diferent.
 - La ReLU posa a zero els valors negatius i poténcia els positius, identificant patrons.
 - Si posem una capa darrere, que reba les dades d'altres capes i les combini, l'efecte serà el de tindre una finestra més gran.

Funcionament d'una CNN (III)

- Si continuem afegint capes, les finestres es faran més grans i més complexes
- Això permetrà identificar patrons més globals
- Finalment, les capes totalment connectades combinaran tots els patrons per tal de classificar la imatge
- Aquesta és la idea bàsica d'una CNN
 - Extreure patrons locals
 - Combinar-los per extreure patrons globals
 - Classificar la imatge

Arquitectures de xarxes neuronals convolucionals

- Hi ha moltes arquitectures de xarxes neuronals convolucionals aprofitables, però les més conegudes són:
 - **VGG-16**: xarxa de 16 capes. Va aconseguir un 92.7% d'exactitud en el dataset ImageNet en 2014.
 - **ResNet**: xarxa de 152 capes, basada en la idea de **residual learning**. Va aconseguir un 96.4% d'exactitud en el dataset ImageNet en 2015.
 - **Inception**: xarxa de 22 capes, basada en la idea de **factorització de convolucions**. Va aconseguir un 97.3% d'exactitud en el dataset ImageNet en 2015.

Reconeixement d'objectes

- El **reconeixement d'objectes** és una tasca més complexa que la classificació d'imatges.
- Mentre que la classificació d'imatges consisteix en **identificar** l'objecte que hi ha a la imatge, el reconeixement d'objectes consisteix en **identificar** els objectes que hi ha a la imatge i **localitzar-los** (dibuixar un rectangle al voltant de l'objecte - *bounding box*).
- Les classes d'objectes a identificar estaran **predefinides**. D'aquesta manera, el sistema podrà identificar si hi ha un gos, un cotxe, una persona, etc.

Procediment bàsic

- El procediment bàsic per fer el reconeixement d'objectes és el següent:
 1. Definim una *finestra* que es mourà per tota la imatge.
 2. Passem la finestra per tota la imatge i en cada posició passem la imatge per una XNC.
 3. Ens quedem en les puntuacions més altes i ignorem la resta.
 4. Resolem conflictes i reduïm la quantitat de *bounding boxes*.

Problemes en el procediment bàsic