Computer Vision for Medical Imaging
Hands-on Worksheets

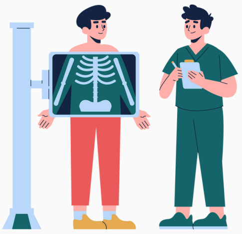
Barbara Klaudel

Table of Contents

Diagnostic Imaging 1

X-ray, CT, MRI and Ultrasound

1. Match sentences with the imaging type that they describe. One sentence can match several types of imaging.



X-ray

A. Images are acquired with ionizing radiation.

B. Uses strong magnetic fields and radio waves.

C. Images are acquired with sound.

D. Non-invasive and does not use ionizing radiation.

E. Images are 2D.

F. Images are a 3D representation.

G. Images are monochromatic.

H. Provides excellent soft tissue contrast.

I. Can be used to visualize blood flow.

J. Often used for bone imaging.

CT

K. Can differentiate between fluid-filled and solid structures.

L. Requires the patient to lie still in a confined space.

M. Can be performed at the bedside.

N. Provides real-time imaging.

MRI

O. Requires the use of contrast agents for some examinations.

P. Uses computer-processed combinations of many X-ray images.

R. Can cause heating of metal implants.

S. Images can be affected by gas in the body.

T. Typically has lower spatial resolution compared to other modalities.

Ultrasound

TheLion.AI

Barbara Klaudel
2024

DICOM Standard 1
DICOM applications, WW & WL

1. Which of the following operations are defined by the DICOM standard?

A. Query
B. Retrieve
C. Display
D. Store
E. Transmit
F. Print

2. In DICOM, **window width (WW)** and **window level (WL)** are used to adjust the contrast and brightness of medical images. The window level is the center of the window, while the window width is the range of CT numbers (Hounsfield Units) that will be displayed as shades of gray.

2.1 Select appropriate term.
A. Changing the **widow width / window level** affects image contrast.
B. Adjusting the **window width / window level** impacts image brightness.

2.2 You are given a CT image of the chest with the following properties:

- Image size: 512 x 512 pixels
- Bit depth: 12 bits
- Hounsfield Unit range: -1000 to 3095

For the following structures fill in WL and WW and describe what can we visualized with these settings.

A. **Lung tissue** WL: ☐, WW: ☐

- Best visualizes:
- Allows detection of:

B. **Soft tissue** WL: ☐, WW: ☐

- Best visualizes:
- Allows differentiation between:

C. Bone WL: ☐, WW: ☐

- Best visualizes:
- Allows detection of:

DICOM Standard 2
Tags and transfer syntaxes

1. Using a DICOM viewer with tag inspection capabilities (e.g., **DicomBrowser, dcm4che tools**):

   A. Open a sample DICOM file and list the values for the following tags:
   - Patient's Name (0010,0010)
   - Study Date (0008,0020)
   - Modality (0008,0060)
   - Manufacturer (0008,0070)
   - Pixel Spacing (0028,0030)

   B. Identify a private tag (hint: look for odd group numbers). What might be its purpose?

2. DICOM uses a hierarchical information model. Explain the relationship between: **Patient, Study, Series and Image**. How is this hierarchy reflected in DICOM tags?

3. Answer the following questions about the transfer syntax.

A. What is a transfer syntax in DICOM? Why is it important?

B. Using a DICOM toolkit (e.g., DCMTK, pydicom), determine the transfer syntax of your sample DICOM file. What does this transfer syntax indicate about:
a) Byte ordering (little endian vs. big endian)
b) Compression (if any)

C. Research and briefly explain the differences between the following transfer syntaxes:

- 1.2.840.10008.1.2 (Implicit VR Little Endian)
- 1.2.840.10008.1.2.1 (Explicit VR Little Endian)
- 1.2.840.10008.1.2.4.50 (JPEG Baseline)

D. How might different transfer syntaxes affect:
A) File size

B) Processing speed

C) Image quality

DICOM Standard 3

Data anonimization

1. List at least 10 DICOM tags that should typically be anonymized to protect patient privacy.

2. Using a DICOM anonymization tool (e.g., DICOM Confidential, gdcmanon), attempt to anonymize your sample DICOM file.
a) Which tags were modified?

b) Were there any unexpected challenges in the anonymization process?

3. Burnt-in annotations:
a) Explain what burnt-in annotations are in the context of medical imaging.

b) Why do they pose a challenge for anonymization?

c) Suggest a strategy for dealing with burnt-in annotations during the anonymization process.

4. Hidden data:
a) Besides obvious identifiers, where else in a DICOM file might patient information be hidden?

b) Research and explain the purpose of the "Patient Comments" (0010,4000) tag. Why might this be problematic for anonymization?

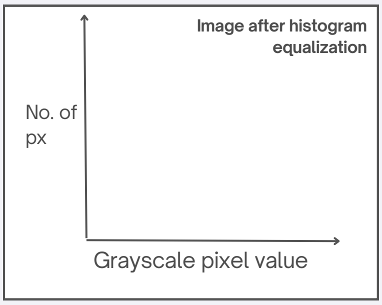
Image Preprocessing 1

Histogram equalization

1. Perform contrast enhancement with histogram equalization.

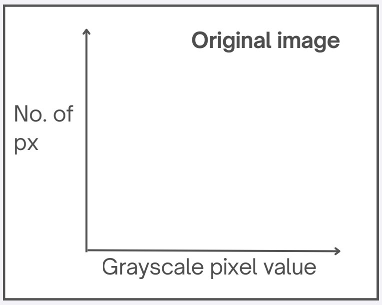
2	2	2	2	2
2	3	5	2	3
2	4	4	3	5
2	4	4	5	3
2	5	1	3	4

Histogram equalization →



A. Count the number of occurrences of each pixel value, fill in the table and plot the result.

Gray scale	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
No. of px								



B. Calculate the PDF, CDF, xx and xx. Use the given formulas.

PDF								
CDF								
Norm CDF								
Hist eq								

max intensity = 7
sum of all px =

PDF = no. of px / sum of all px

CDF -> cumulative sum

Norm CDF -> divide each cell by the max. intensity value

Hist eq-> Round to the nearest unit

C. Fill in the values after histogram equalization. Hist eq. row tells you how to encode the original number. Plot the values. Compare the result with the original image.

D. How do more advanced methods defer from the regular histogram equalization?

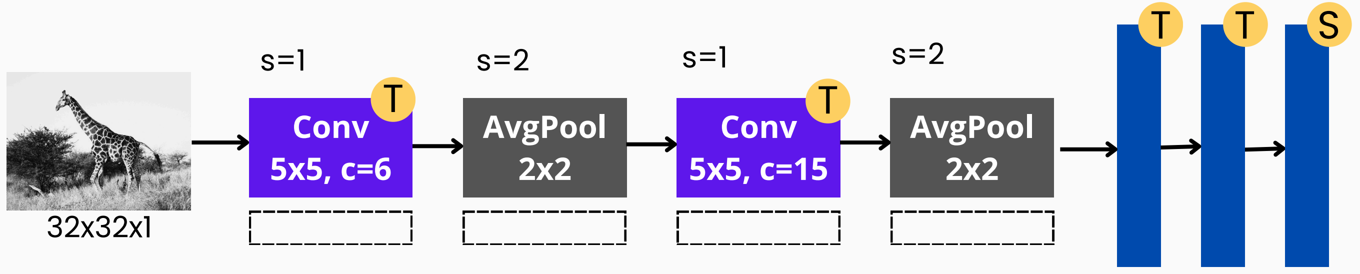
Adaptive histogram - equalization

Contrastive Limited - Adaptive histogram equalization

CNN 1

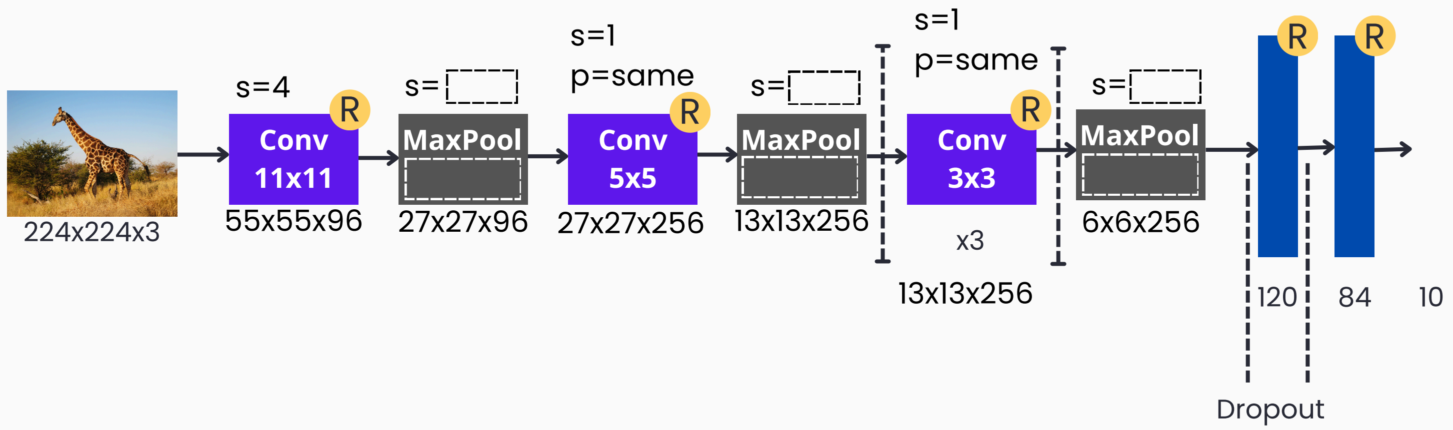
LeNet

1. Fill in output size after each layer.



AlexNet

2. 1. Fill in kernel size and stride in MaxPool layers.



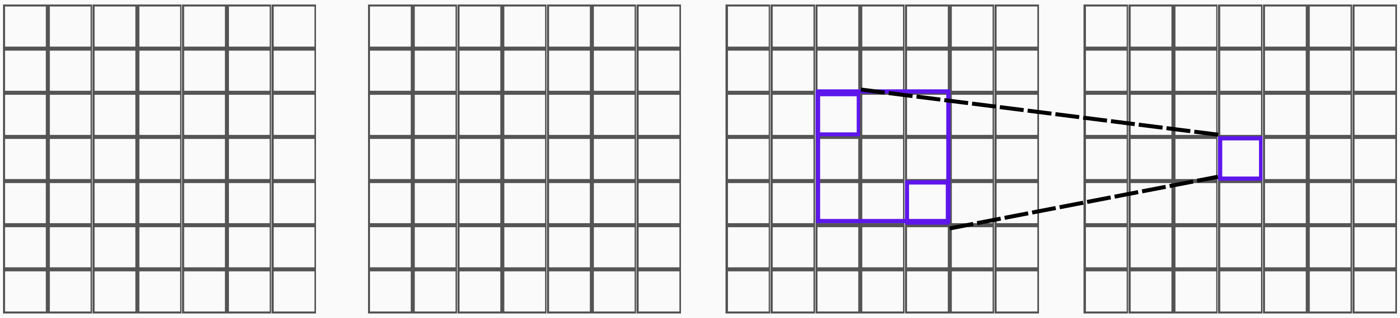
2. 2. Write down 3 significant differences between AlexNet and LeNet.

2. 3. For each difference, explain why AlexNet authors decided for a different option.

CNN 2

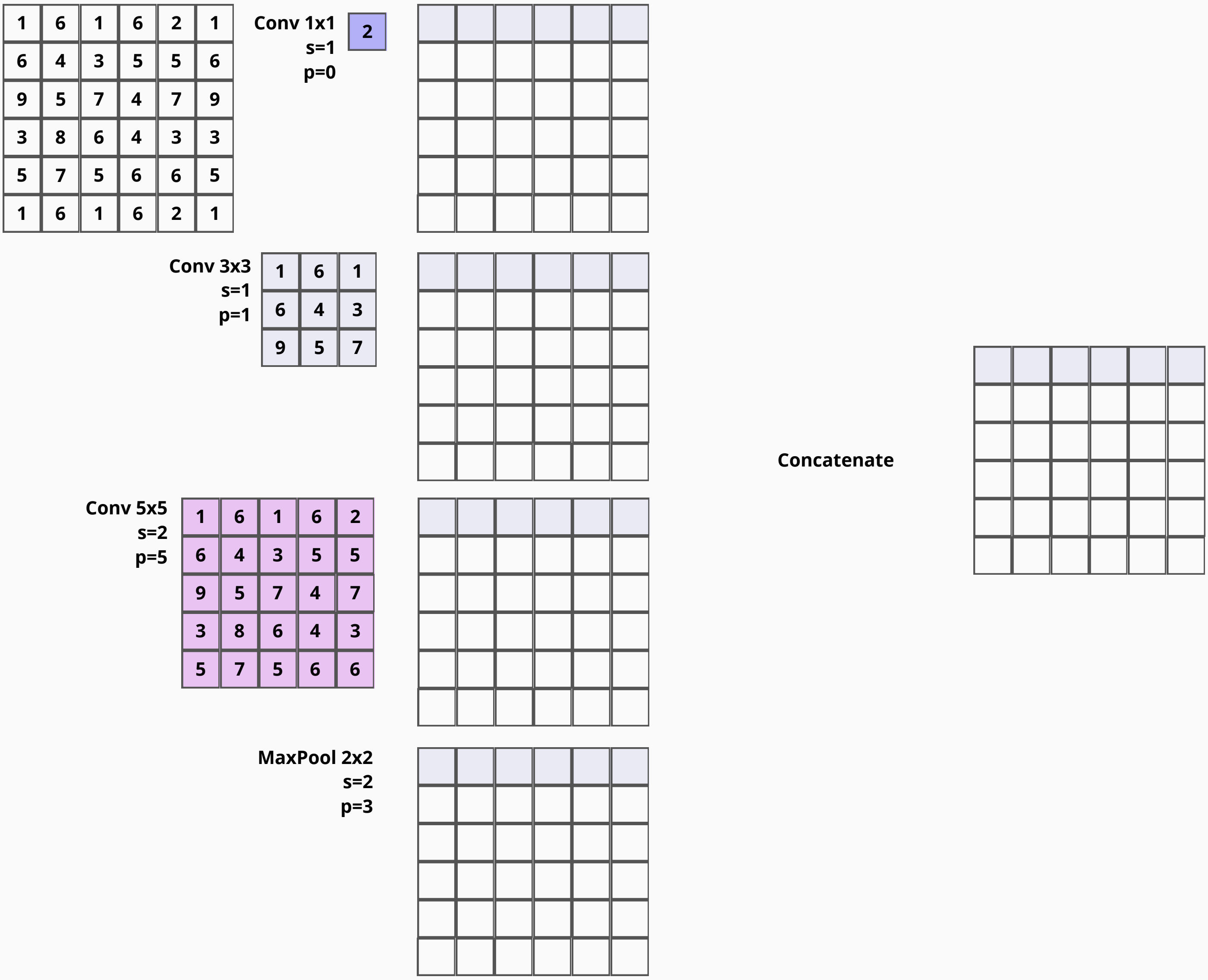
VGG

1. Compare perception fields of three 3x3 filters. How many fields from the first images modifies a single field in the last image? Illustrate your answer.



GoogleLeNet (Inception V1)

2. Calculate the result of the simplified Inception block. For each matrix, fill in only th first row.



1	6	1	6	2	1
6	4	3	5	5	6
9	5	7	4	7	9
3	8	6	4	3	3
5	7	5	6	6	5
1	6	1	6	2	1

Conv 1x1
s=1
p=0

2

Conv 3x3
s=1
p=1

1	6	1
6	4	3
9	5	7

Conv 5x5
s=2
p=5

1	6	1	6	2
6	4	3	5	5
9	5	7	4	7
3	8	6	4	3
5	7	5	6	6

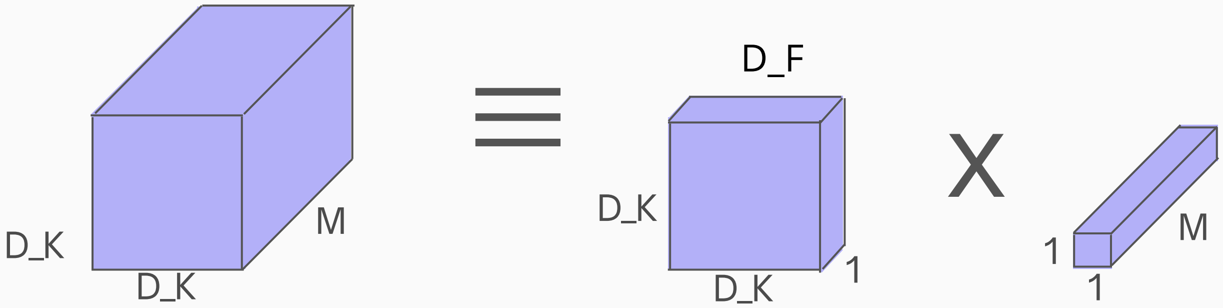
MaxPool 2x2
s=2
p=3

Concatenate

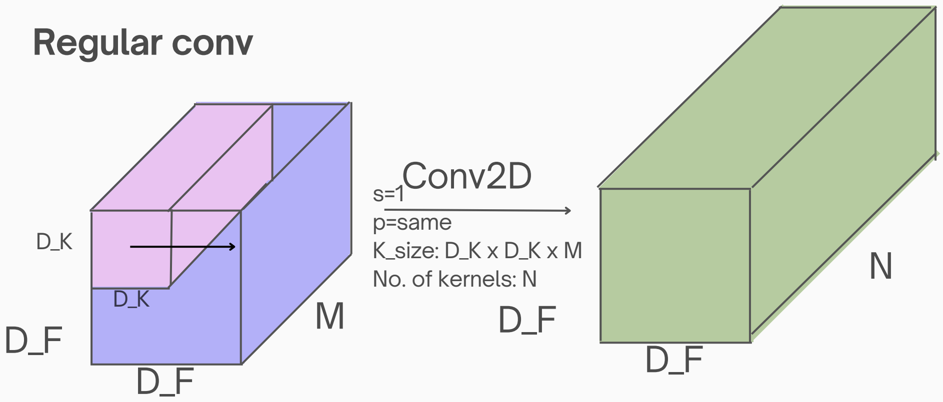
CNN 3

MobileNet

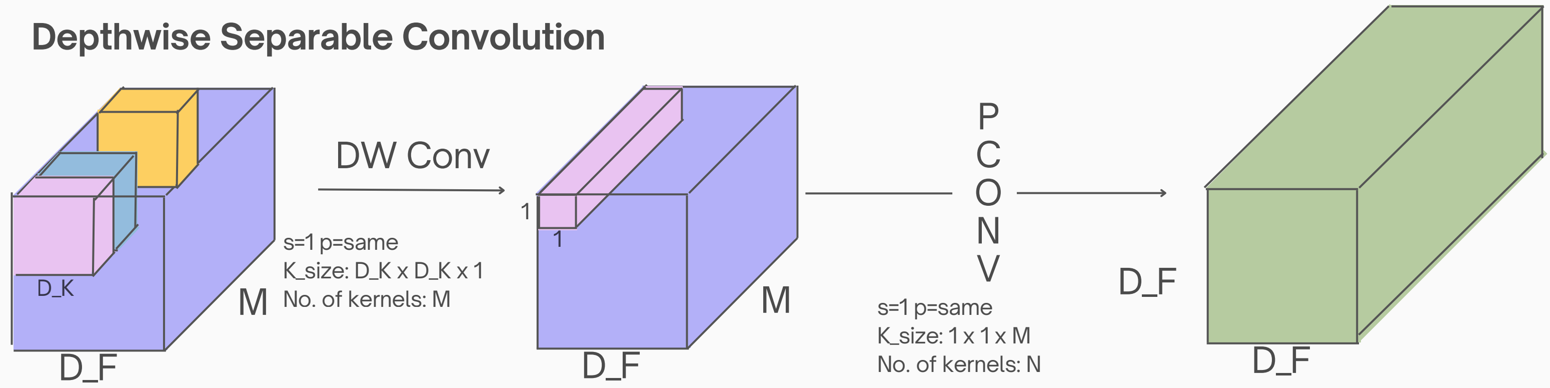
1. How many times can we reduce the number of calculations with the Depthwise Separable Convolutions compared with a regular convolution?
To calculate the final result assume N=32 and D_K = 3.



Regular conv



Depthwise Separable Convolution



$$X_1 = \underbrace{D_F^2}_{\text{No. of positions each kernel takes}} \cdot \underbrace{D_K^2 \cdot M}_{\text{Number of elements in a kernel}} \cdot \underbrace{N}_{\text{Number of kernels}}$$

$$X_2 =$$

$$R = \frac{X_2}{X_1} =$$

CNN 4

SOTA CNNs Summary

1. What new contributions did each of the following architectures make? Add examples of new developments to each architecture.

2012 - AlexNet

2014 - VGG

2014 - GoogleLeNet

2015 - ResNet

2016 - SqueezeNet

2017 - MobileNet

2019 - EfficientNet

2022 - ConvNeXt

Image Segmentation 1

Transposed Convolution

1. Calculate the output of the transpose convolution.
Assume stride=2, padding=1.

8.1 Add padding=1 around the output
8.2 Mark the receptive fields of each input cell on the output cell.
8.3 Calculate thesubscores of each cell in the output matrix.
8.4 Calculate the values of the final output matrix.

Input

2	1
3	2

Kernel

1	0	3
0	3	1
2	1	0

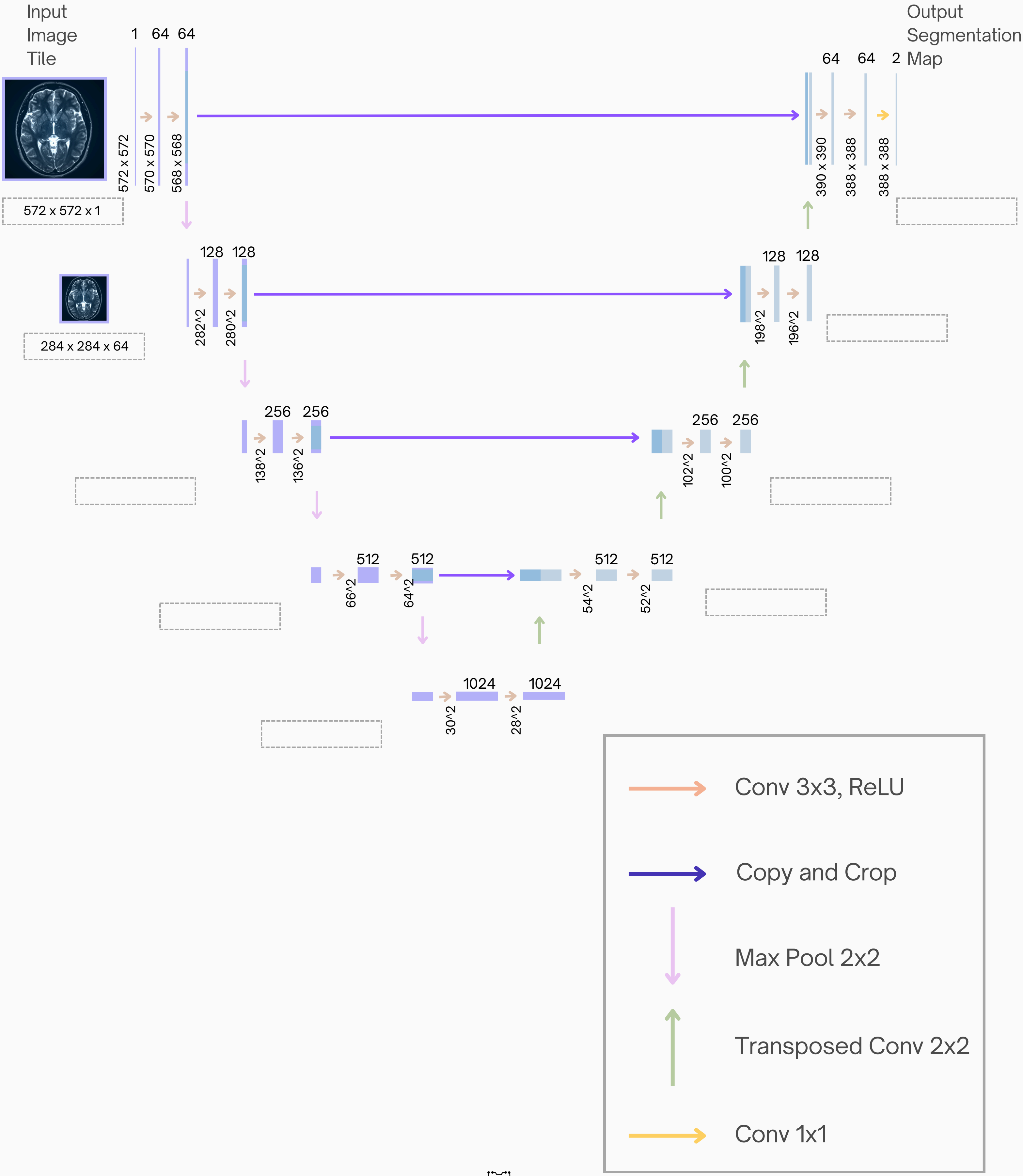
Output

=

Image Segmentation 2

UNet

1. Fill in the input size for each block.



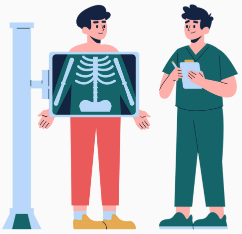
TheLion.AI

Barbara Klaudel
2024

Diagnostic Imaging 1

Diagnostic Imaging - with answers

1. Match sentences with the imaging type that they describe. One sentence can match several types of imaging.



X-ray

A. Images are acquired with ionizing radiation.

X-ray, CT

B. Uses strong magnetic fields and radio waves.

MRI

C. Images are acquired with sound.

Ultrasound

D. Non-invasive and does not use ionizing radiation.

MRI, Ultrasound

E. Images are 2D.

X-ray, Ultrasound (some applications)

F. Images are a 3D representation.

CT, MRI, Ultrasound (3D/4D) applications

G. Images are monochromatic.

X-ray, CT (in standard form)

H. Provides excellent soft tissue contrast.

MRI, Ultrasound

I. Can be used to visualize blood flow.

MRI (with specific sequences), Ultrasound (Doppler)

CT

J. Often used for bone imaging.

X-ray, CT

K. Can differentiate between fluid-filled and solid structures.

Ultrasound, MRI

MRI

L. Requires the patient to lie still in a confined space.

CT, MRI

M. Can be performed at the bedside.

X-ray (portable), Ultrasound

N. Provides real-time imaging.

Ultrasound

O. Requires the use of contrast agents for some examinations.

X-ray, CT, MRI

P. Uses computer-processed combinations of many X-ray images.

CT

R. Can cause heating of metal implants.

MRI

S. Images can be affected by gas in the body.

Ultrasound

T. Typically has lower spatial resolution compared to other modalities.

Ultrasound (compared to CT and MRI)

Ultrasound

TheLion.AI
Barbara Klaudel
2024

DICOM Standard 1
DICOM applications, WW & WL

1. Which of the following operations are defined by the DICOM standard?

[✓ A. Query]
[✓ B. Retrieve]
[✓ C. Display]
[✓ D. Store]
[✓ E. Transmit]
[✓ F. Print]

All answers are correct!

2. In DICOM, **window width (WW)** and **window level (WL)** are used to adjust the contrast and brightness of medical images. The window level is the center of the window, while the window width is the range of CT numbers (Hounsfield Units) that will be displayed as shades of gray.

2.1 Select appropriate term.
A. Changing the [✓ **widow width**] / **window level** affects image contrast.
B. Adjusting the **window width** / [✓ **window level**] impacts image brightness.

2.2 You are given a CT image of the chest with the following properties:

- Image size: 512 x 512 pixels
- Bit depth: 12 bits
- Hounsfield Unit range: -1000 to 3095

For the following structures fill in WL and WW and describe what can we visualized with these settings.

A. **Lung tissue** WL: 1500 HU, WW: -600 HU
- Best visualizes: Lung parenchyma, airways, pulmonary vessels
- Allows detection of: subtle lung nodules or interstitial lung diseases

B. **Soft tissue** WL: 350-400 HU, WW: 40-50 HU
- Best visualizes: Muscles, organs, fat, fluid collections
- Allows differentiation between: various soft tissue structures and detection of pathologies like masses or inflammation

C. Bone WL: 1500-2000 HU, WW: 300-350 HU
- Best visualizes: Bone cortex and trabeculae, calcifications
- Allows detection of: fractures, bone lesions, or subtle calcifications

DICOM Standard 2
Tags and transfer syntaxes

1. Using a DICOM viewer with tag inspection capabilities (e.g., **DicomBrowser, dcm4che tools**):

A. Open a sample DICOM file and list the values for the following tags:

- Patient's Name (0010,0010) e.g. SMITH^JOHN
- Study Date (0008,0020) e.g. 20240303
- Modality (0008,0060) e.g. CT
- Manufacturer (0008,0070) e.g. SIEMENS
- Pixel Spacing (0028,0030) e.g. 0.7\0.7

B. Identify a private tag (hint: look for odd group numbers). What might be its purpose?

2. DICOM uses a hierarchical information model. Explain the relationship between: **Patient, Study, Series and Image**. How is this hierarchy reflected in DICOM tags?

- Patient: Top level, can have multiple studies. Patient ID (0010,0020)
- Study Instance UID (0020,000D) Study: Collection of series related to a specific procedure, can have multiple series.
- Series Instance UID (0020,000E) Series: Set of images from a single acquisition, can have multiple images.
- SOP Instance UID (0008,0018) Image: Individual DICOM file containing pixel data and metadata.

3. Answer the following questions about the transfer syntax.

A. What is a transfer syntax in DICOM? Why is it important?

Specifies the encoding of the data elements and pixel data in a DICOM dataset. It determines how the data is formatted, compressed, and should be interpreted.

B. Using a DICOM toolkit (e.g., DCMTK, pydicom), determine the transfer syntax of your sample DICOM file. What does this transfer syntax indicate about:

a) Byte ordering (little endian vs. big endian)

b) Compression (if any)

C. Research and briefly explain the differences between the following transfer syntaxes:

- 1.2.840.10008.1.2 (Implicit VR Little Endian)

Little-endian byte ordering, no compression, tags are encoded without explicitely statying what type of information they store.

- 1.2.840.10008.1.2.1 (Explicit VR Little Endian)

Little-endian byte ordering, no compression, tags are encoded with the information about what type of information they store.

- 1.2.840.10008.1.2.4.50 (JPEG Baseline)

Pixel data is stored as lossy JPEGs.

D. How might different transfer syntaxes affect:

A) File size Compressed syntaxes (e.g., JPEG) result in smaller file sizes.

B) Processing speed Uncompressed data (e.g., Explicit VR Little Endian) is faster to read but slower to transmit.

C) Image quality Lossy compression (e.g., JPEG Baseline) can reduce image quality, while lossless or uncompressed maintain original quality.

DICOM Standard 3
Data anonimization

1. List at least 10 DICOM tags that should typically be anonymized to protect patient privacy. e.g. : Patient's Name (0010,0010), Patient ID (0010,0020), Date of Birth (0010,0030), Patient's Address (0010,1040), Patient's Telephone Numbers (0010,2154), Referring Physician's Name (0008,0090), Institution Name (0008,0080), Study Date (0008,0020), Accession Number (0008,0050), Study ID (0020,0010)

2. Using a DICOM anonymization tool (e.g., DICOM Confidential, gdcmanon), attempt to anonymize your sample DICOM file.

a) Which tags were modified?

b) Were there any unexpected challenges in the anonymization process?

E.g. Personal information imprinted in the pixel data.

3. Burnt-in annotations:

a) Explain what burnt-in annotations are in the context of medical imaging.

Burnt-in annotations are text or graphics permanently embedded in the pixel data of medical images.

b) Why do they pose a challenge for anonymization?

because they can't be removed by simply modifying DICOM tags; they require image processing.

c) Suggest a strategy for dealing with burnt-in annotations during the anonymization process.

Optical character recognition (OCR) to detect text in images, then apply image processing techniques to remove or obscure the detected regions.

4. Hidden data:

a) Besides obvious identifiers, where else in a DICOM file might patient information be hidden?

Private tags, sequence items, comments or notes fields, image overlays.

b) Research and explain the purpose of the "Patient Comments" (0010,4000) tag. Why might this be problematic for anonymization?

It can contain free-text information entered by healthcare providers, potentially including identifiable details not captured in other structured fields.

Image Preprocessing 1

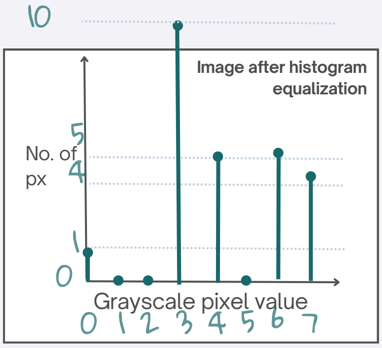
Histogram equalization

1. Perform contrast enhancement with histogram equalization.

2	2	2	2	2
2	3	5	2	3
2	4	4	3	5
2	4	4	5	3
2	5	1	3	4

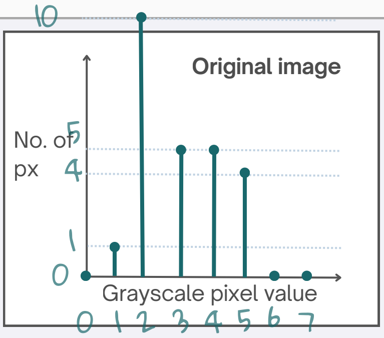
Histogram equalization

3	3	3	3	3
3	4	7	3	4
3	6	6	4	7
3	6	6	7	4
3	7	0	4	6



A. Count the number of occurrences of each pixel value, fill in the table and plot the result.

Gray scale	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
No. of px	0	1	10	5	5	4	0	0



B. Calculate the PDF, CDF, xx and xx. Use the given formulas.

PDF	0	0.04	0.4	0.2	0.2	0.16	0	0
CDF	0	0.04	0.44	0.64	0.84	1	1	1
Norm CDF	0	0.28	3.08	4.48	5.88	7	7	7
Hist eq	0	0	3	4	6	7	7	7

max intensity = 7
sum of all px = 24

PDF = no. of px / sum of all px

CDF -> cumulative sum

Norm CDF -> divide each cell by the max. intensity value

Hist eq-> Round to the nearest unit

C. Fill in the values after histogram equalization. Hist eq. row tells you how to encode the original number. Plot the values. Compare the result with the original image.

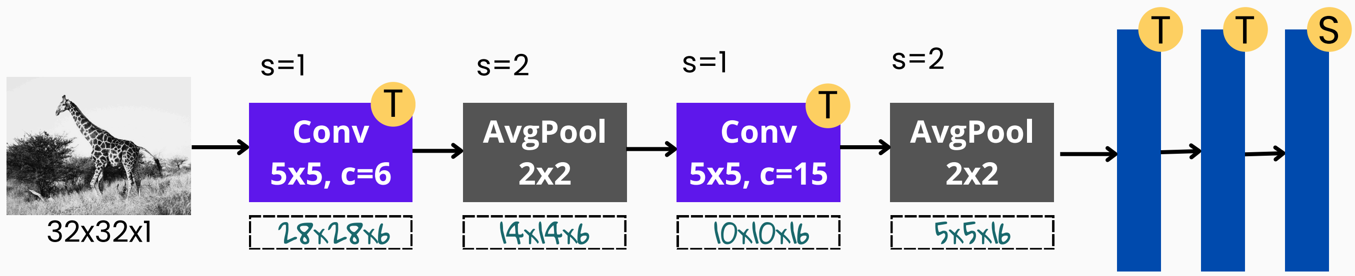
D. How do more advanced methods defer from the regular histogram equalization?

Adaptive histogram equalization -	the adaptive method computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image.
Contrastive Limited Adaptive histogram equalization -	the contrast limiting procedure is applied to each neighborhood from which a transformation function is derived

CNN 1

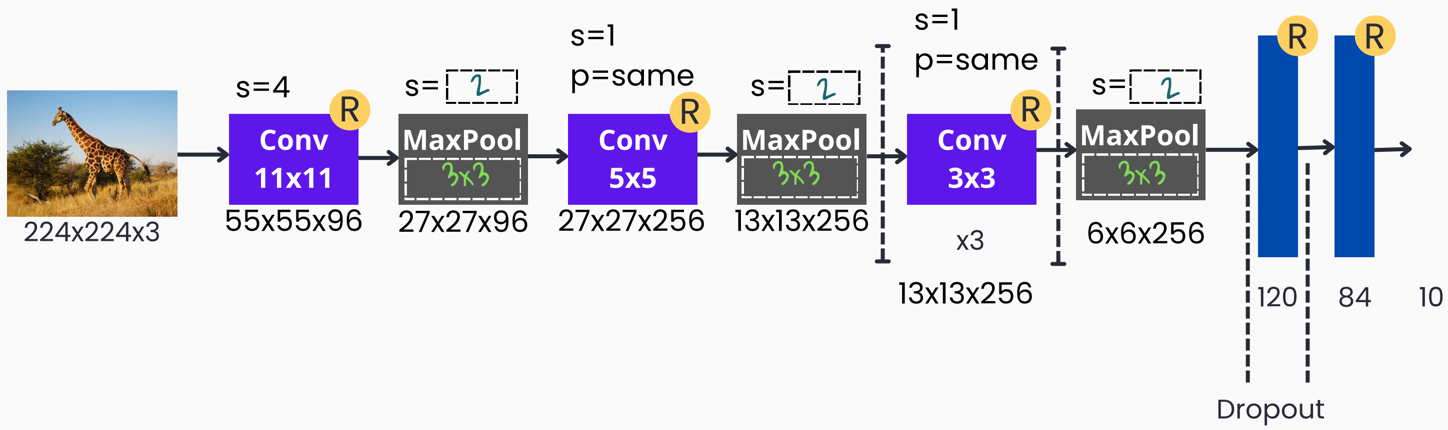
LeNet

1. Fill in output size after each layer.



AlexNet

2. 1. Fill in kernel size and stride in MaxPool layers.



2. 2. Write down 3 significant differences between AlexNet and LeNet.

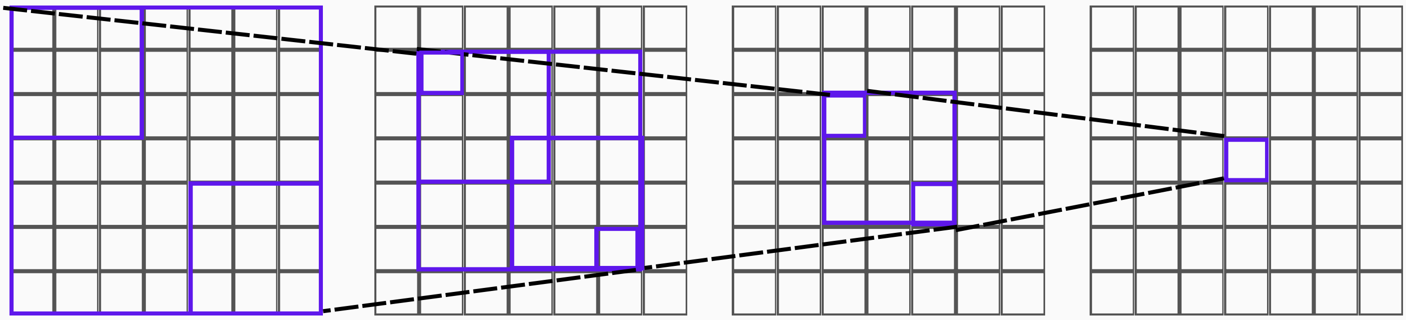
AlexNet uses MaxPool while LeNet uses AvgPool.	AlexNet has more layers than LeNet.
AlexNet uses ReLU instead of Tanh.	ReLU is cheaper to compute which gets more important the deeper the network is. ReLU has more diverse outputs.
AlexNet uses larger kernels than LeNet.	AlexNet works with larger images. AlexNet's input: 224x224x3 LeNet's input: 32x32x1

2. 3. For each difference, explain why AlexNet authors decided for a different option.

CNN 2

VGG

1. Compare perception fields of three 3x3 filters. How many fields from the first images modifies a single field in the last image? Illustrate your answer.



GoogleLeNet (Inception V1)

2. Calculate the result of the simplified Inception block. For each matrix, fill in only th first row.

1	6	1	6	2	1
6	4	3	5	5	6
9	5	7	4	7	9
3	8	6	4	3	3
5	7	5	6	6	5
1	6	1	6	2	1

Conv 1x1
s=1
p=0

2

2	12	2	12	4	2

Conv 3x3
s=1
p=1

1	6	1
6	4	3
9	5	7

80	128	144	123	159	76

Conv 5x5
s=2
p=5

1	6	1	6	2
6	4	3	5	5
9	5	7	4	7
3	8	6	4	3
5	7	5	6	6

0	0	0	0	0	0

MaxPool 2x2
s=2
p=3

0	0	0	0	0	0

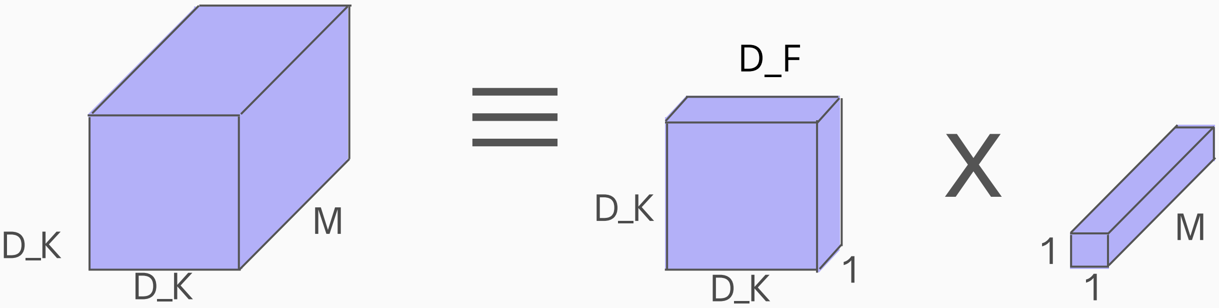
Concatenate

82	140	146	135	163	78

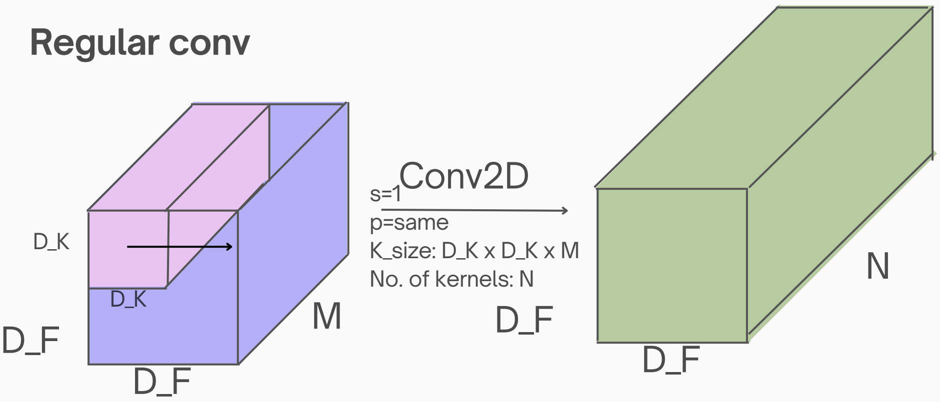
CNN 3

MobileNet

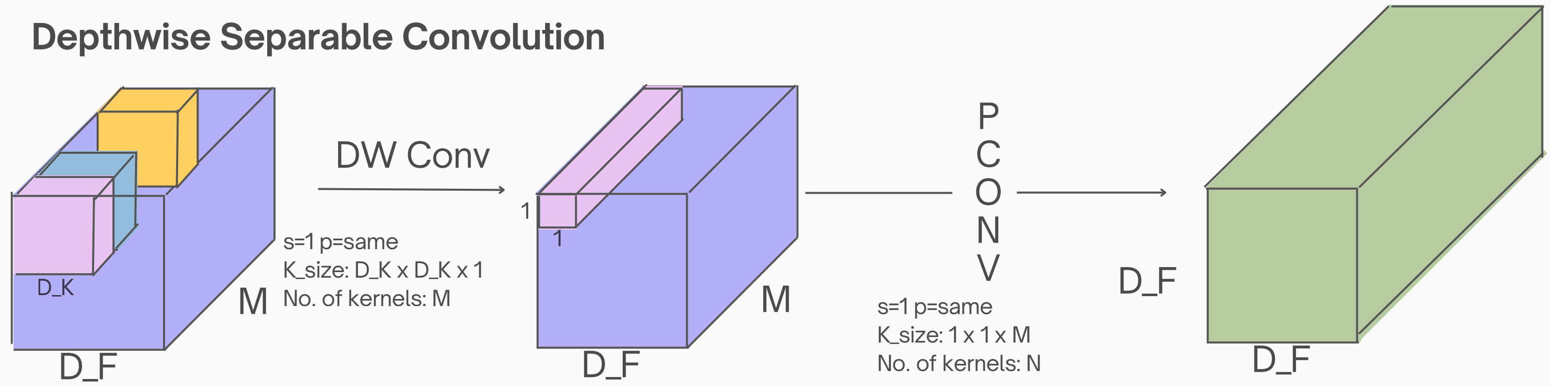
1. How many times can we reduce the number of calculations with the Depthwise Separable Convolutions compared with a regular convolution? To calculate the final result assume N=32 and D_K = 3.



Regular conv



Depthwise Separable Convolution



$$X_1 = \underbrace{D_F^2}_{\text{No. of positions each kernel takes}} \cdot \underbrace{D_K^2 \cdot M}_{\text{Number of elements in a kernel}} \cdot \underbrace{N}_{\text{Number of kernels}}$$

$X_2 =$ DK^2 * M * DF^2 + M * N * DF^2 = M * DF^2 (DK^2 + N)

N = 32
DK = 3

$R = \frac{X_2}{X_1} =$ 1 / N + 1 / DK^2 ~= 14.2% -> roughly 7 times faster

CNN 4

SOTA CNNs Summary

1. What new contributions did each of the following architectures make? Add examples of new developments to each architecture.

2012 - AlexNet
- ReLU instead of Tanh.
- MaxPool instead of AvgPool.
- More layers + larger kernel -> larger inputs
- uses GPU

2014 - VGG
- Use the same 3x3 kernels for ALL conv operations.
- Double the number of channels after each block.
- More layers.

2014 - GoogleLeNet
- Inception blocks.
- Auxiliary classifiers.
- Global average pooling.
- 1x1 convolution.

2015 - ResNet
- Residual connections -> way more layers.

2016 - SqueezeNet
- Small number of parameters.
- No fully connected layers.
- Squeeze and expand blocks.

2017 - MobileNet
- Small number of parameters
- Depthwise Separable COnvolution

2019 - EfficientNet
- Small number of parameters
- Depthwise Separable COnvolution

2022 - ConvNeXt
- Pure CNN taking inspiration from transformers.
- Larger kernel size.
- GERU instead of ReLU.
- Layer normalization instead of bartch normalization.
- Inverted bottleneck.
- Depthwise convolution.

Image Segmentation 1

Transposed Convolution

1. Calculate the output of the transpose convolution.
Assume stride=2, padding=1.

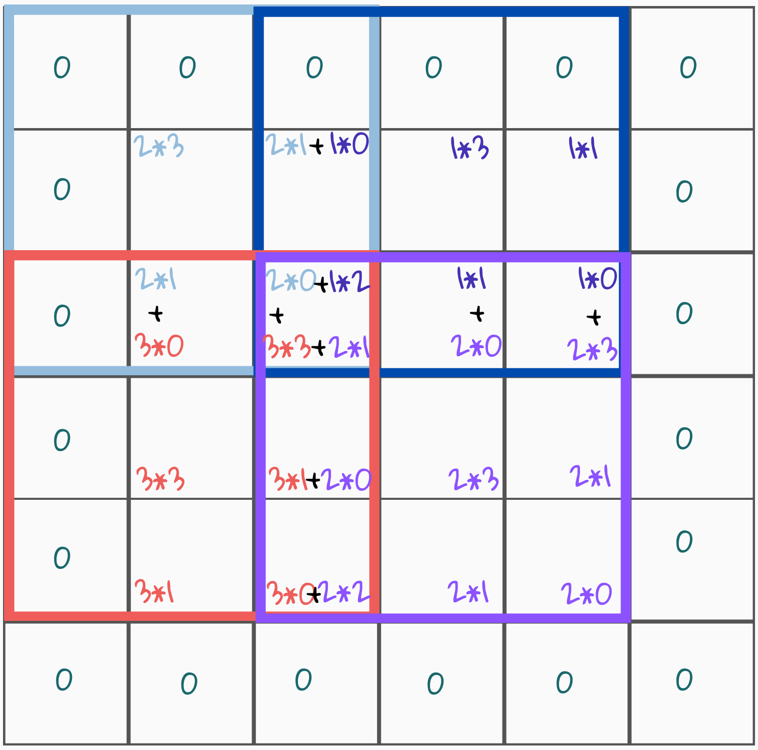
8.1 Add padding=1 around the output
8.2 Mark the receptive fields of each input cell on the output cell.
8.3 Calculate thesubscores of each cell in the output matrix.
8.4 Calculate the values of the final output matrix.

Input

2	1
3	2

Kernel

1	0	3
0	3	1
2	1	0



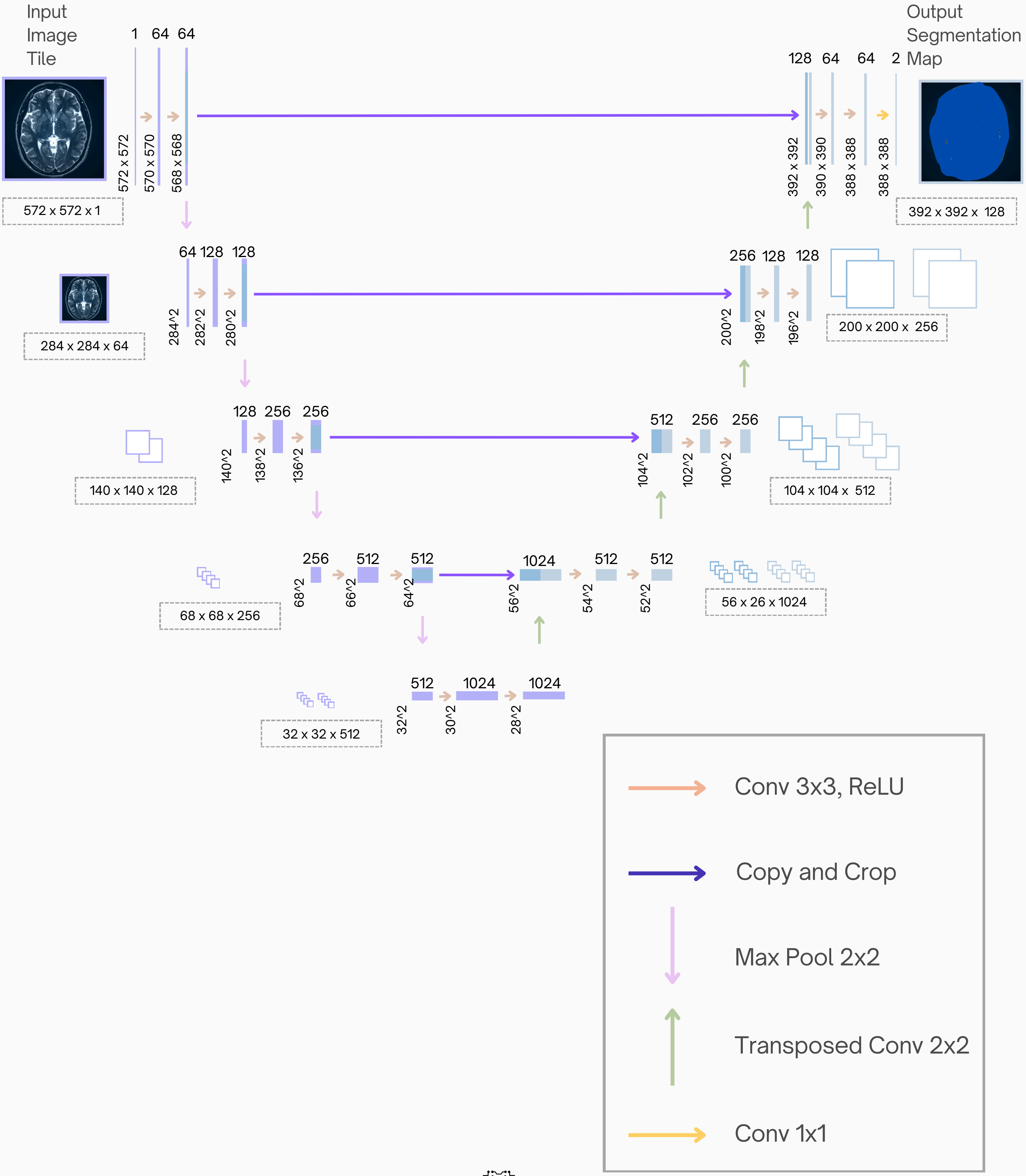
0	0	0	0	0	0
0	2*3	2*1+1*0	1*3	1*1	0
0	2*1 + 3*0	2*0+1*2 + 3*3+2*1	1*1 + 2*0	1*0 + 2*3	0
0	3*3	3*1+2*0	2*3	2*1	0
0	3*1	3*0+2*2	2*1	2*0	0
0	0	0	0	0	0

=

6	2	3	1
2	10	1	6
9	3	6	2
3	4	2	0

Image Segmentation 2
UNet

1. Fill in the input size for each block.



TheLion.AI

Barbara Klaudel
2024