Modern neural network architectures for image analysis

Deep learning course for industry

Mitko Veta

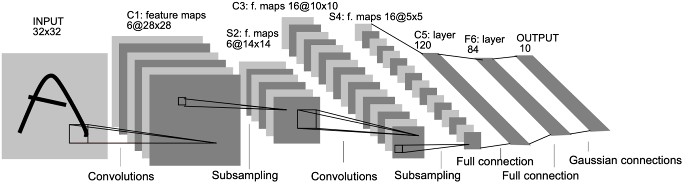
Eindhoven University of Technology
Department of Biomedical Engineering

2020

Learning goals

- Have an overview of relatively recent neural network architectures
  - Image classification
  - Object detection and segmentation
  - Generative models

LeNET



Y. LeCun et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324

What changed since the 1998?

What changed since the 1998?

Explosion of datasets (public and proprietary)

What changed since the 1998?

Explosion of datasets (public and proprietary)

Caffe2 | Microsoft CNTK | TensorFlow | theano

Chainer | PYTORCH | dmlc mxnet | K

ImageNet challenge

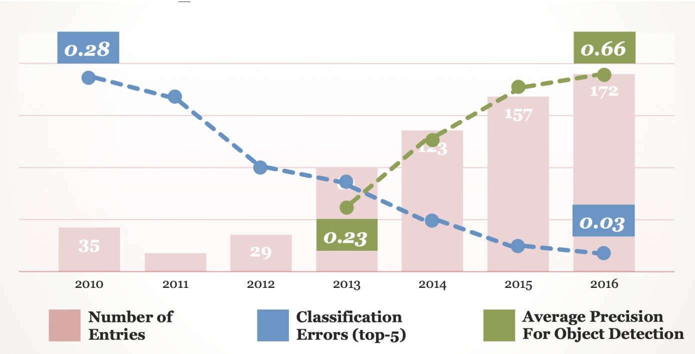


Figure source: image-net.org

Medical image analysis challenges

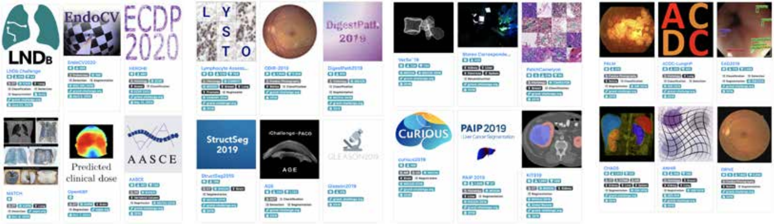
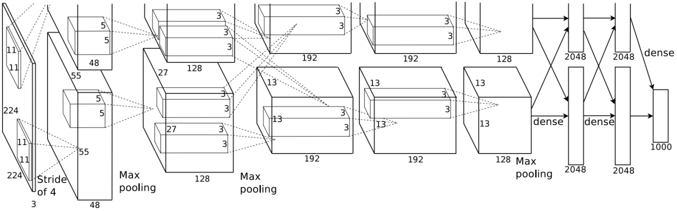


Figure source: grand-challenge.org

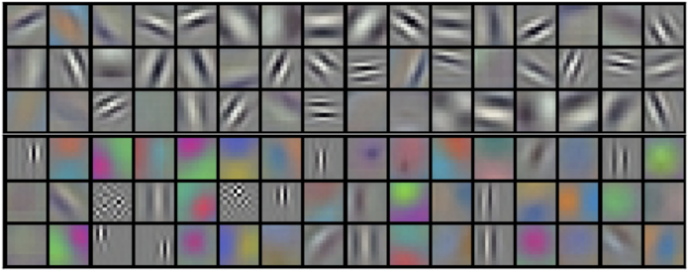
Architectures for image classification

AlexNet



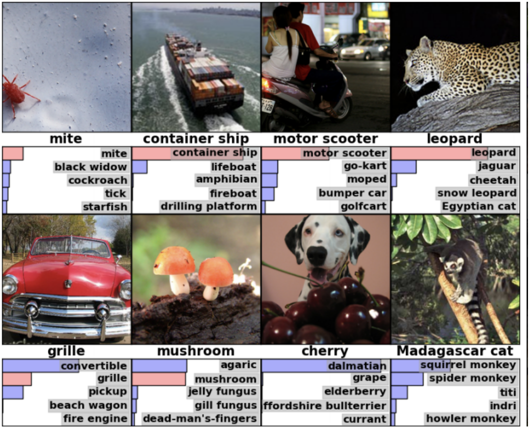
A. Krizhevsky et al. (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105

AlexNet



A. Krizhevsky et al. (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105

AlexNet



A. Krizhevsky et al. (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105

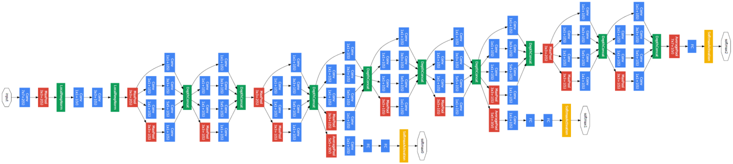
VGG-net

"The image is passed through a stack of convolutional (conv.) layers, where we use filters with a very small receptive field: 3 × 3 (which is the smallest size to capture the notion of left/right, up/down, center)."

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

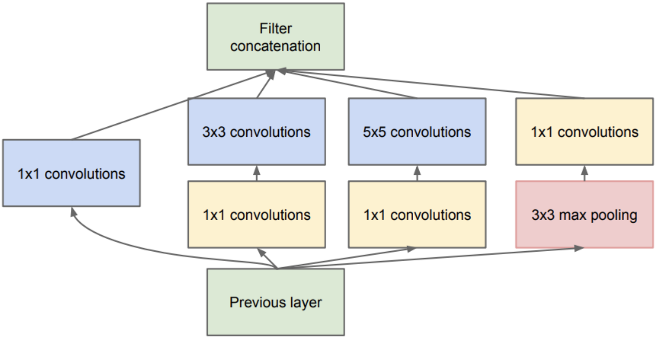
K. Simonyan et al. (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*

Inception-v1



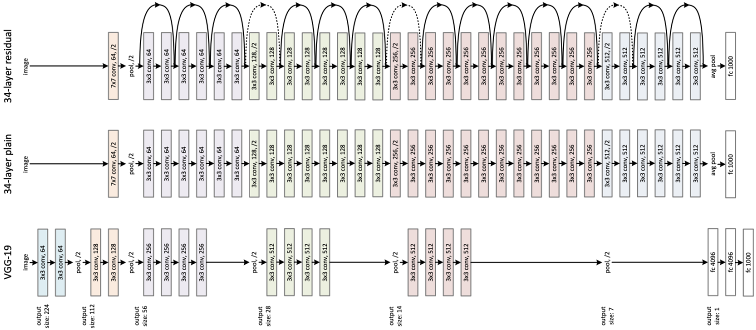
C. Szegedy, W. Liu, et al. (2015). “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9

Inception module



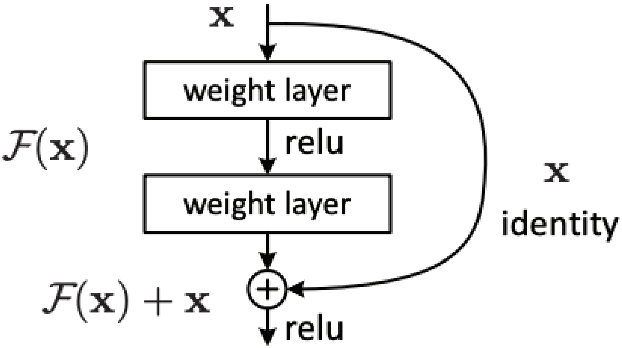
C. Szegedy, W. Liu, et al. (2015). "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9

ResNet



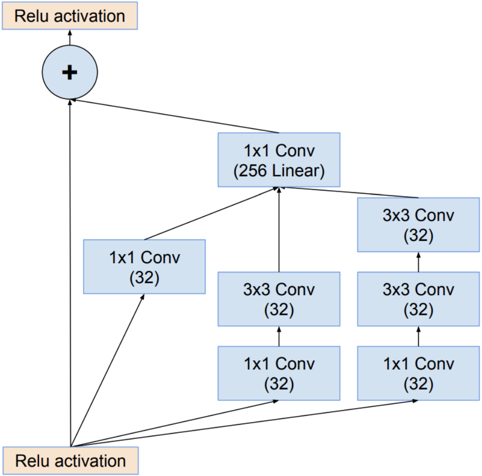
K. He, X. Zhang, et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778

Residual block



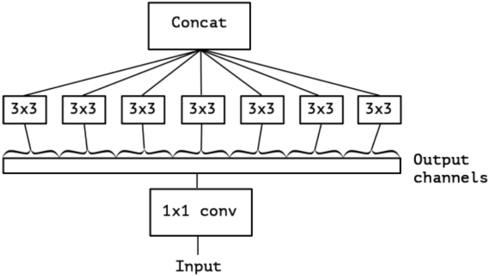
K. He, X. Zhang, et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778

Inception-ResNet



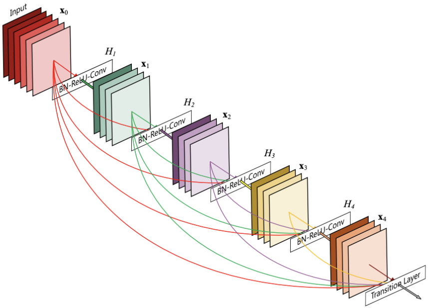
C. Szegedy, S. Ioffe, et al. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning". In: *Thirty-first AAAI conference on artificial intelligence*

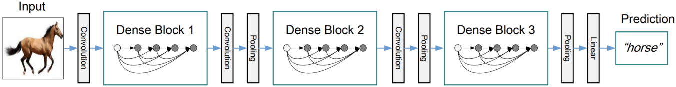
Xception



F. Chollet (2017). "Xception: Deep learning with depthwise separable convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258

Densenet



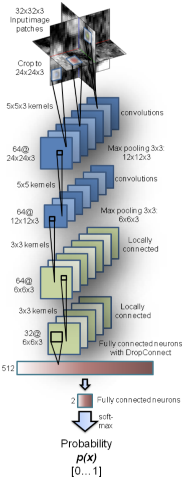


G. Huang et al. (2017). "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708

3D convolutional neural networks

All architectural components and features of 2D networks can be also used with 3D networks (e.g. residual connections).

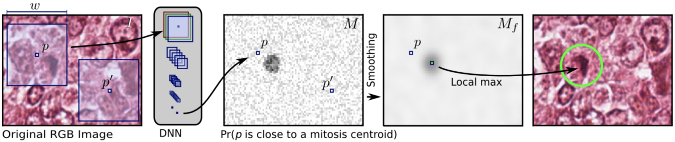
2D architectures can be used for 3D data, either in a slice-by-slice manner or with pseudo-3D inputs.



H. R. Roth et al. (2015). "Improving computer-aided detection using convolutional neural networks and random view aggregation". In: *IEEE transactions on medical imaging* 35.5, pp. 1170–1181

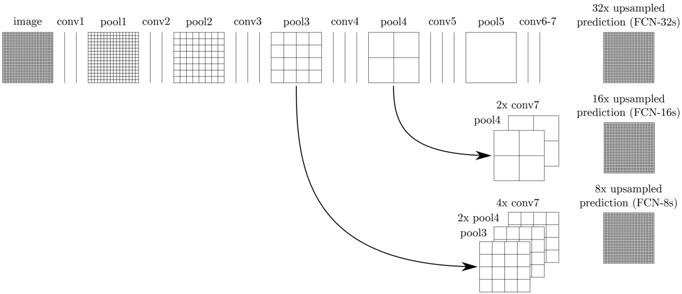
Architectures for object detection and image segmentation

Sliding window object detection



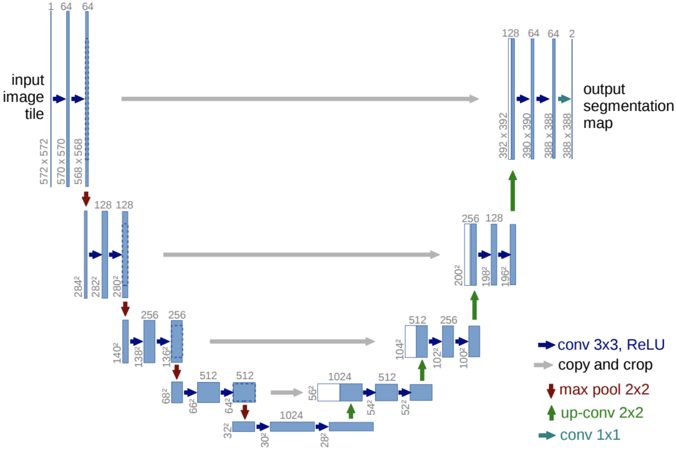
D. C. Cireşan et al. (2013). "Mitosis detection in breast cancer histology images with deep neural networks". In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 411–418

Fully convolutional neural network architectures



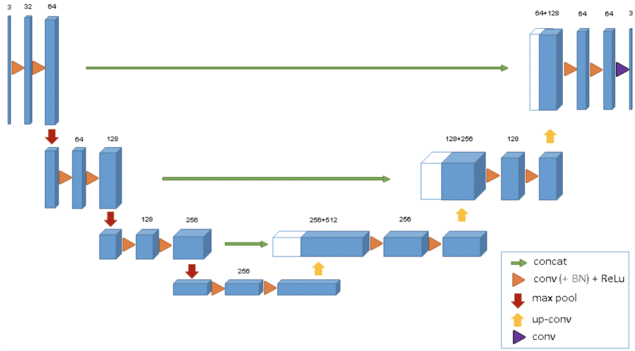
J. Long et al. (2015). "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440

U-Net



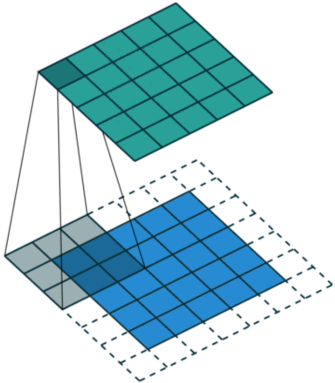
O. Ronneberger et al. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241

3D U-Net



Ö. Çiçek et al. (2016). "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 424–432

Dilated convolutions



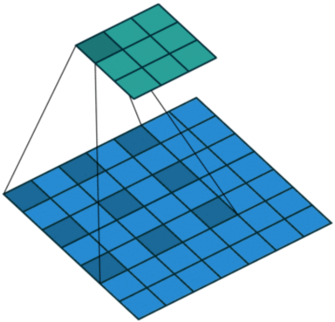
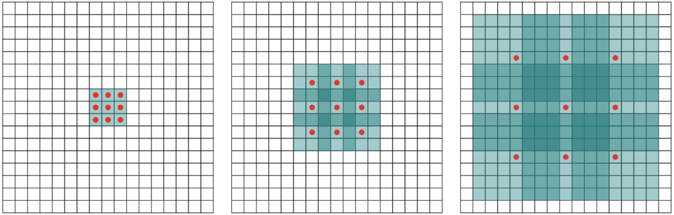


Figure source: github.com/vdumoulin

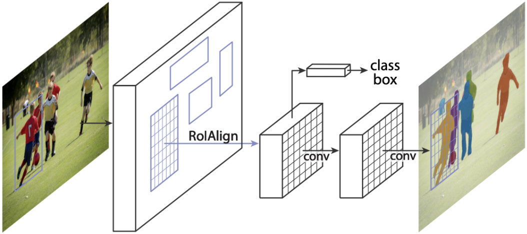
Dilated convolutions



F. Yu et al. (2015). "Multi-scale context aggregation by dilated convolutions". In: *arXiv preprint arXiv:1511.07122*

Region proposal CNNs

Main idea: one end-to-end model that both detects regions and classifies/segments objects in them.



K. He, G. Gkioxari, et al. (2017). "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969

Region proposal CNNs

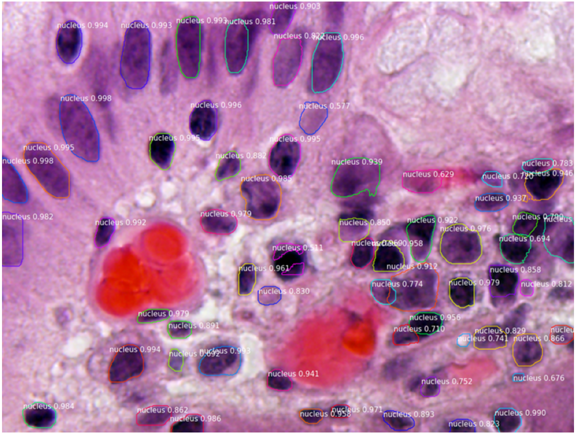


Figure from: github.com/matterport/Mask_RCNN

Generative models

Generative adversarial networks

Two competing models:

- **Generator:** learns to generate new data.
- **Discriminator:** learns to distinguish real from fake (generated) data.

Generative adversarial networks

Two competing models:

- **Generator:** learns to generate new data.
- **Discriminator:** learns to distinguish real from fake (generated) data.

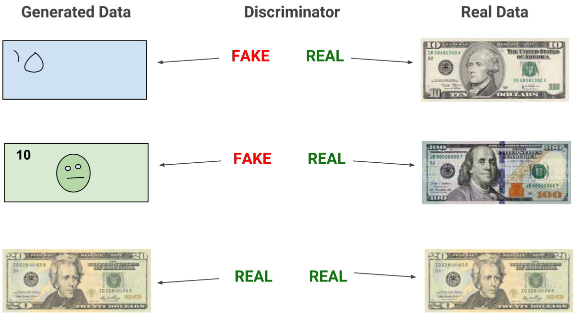
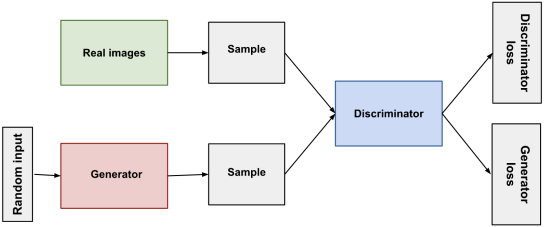


Figure from: developers.google.com/machine-learning/gan

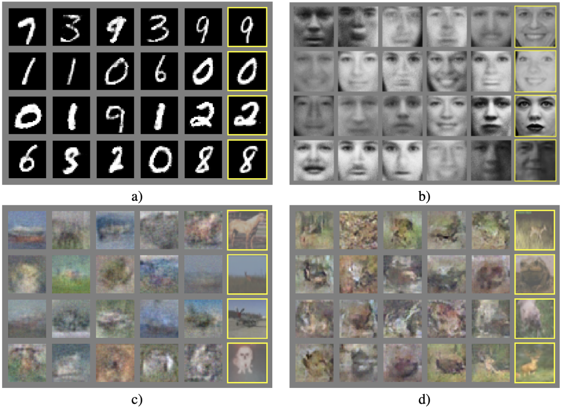
Generative adversarial networks



The input is a random number but it can also be a condition (e.g. a class, object mask etc.).

Figure from: developers.google.com/machine-learning/gan

Generative adversarial networks



I. Goodfellow et al. (2014). "Generative adversarial nets". In: *Advances in neural information processing systems*, pp. 2672–2680

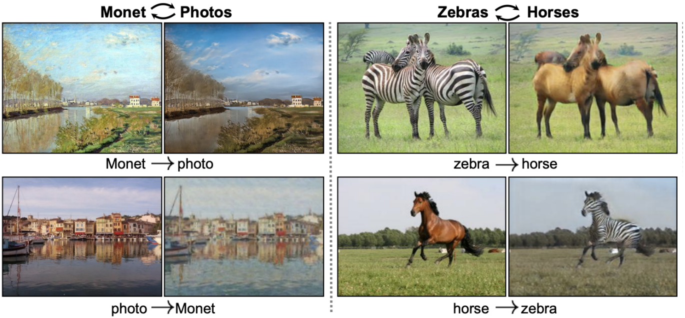
BigGANs

A. Brock et al. (2018). "Large scale gan training for high fidelity natural image synthesis". In: *arXiv preprint arXiv:1809.11096*

Fake faces

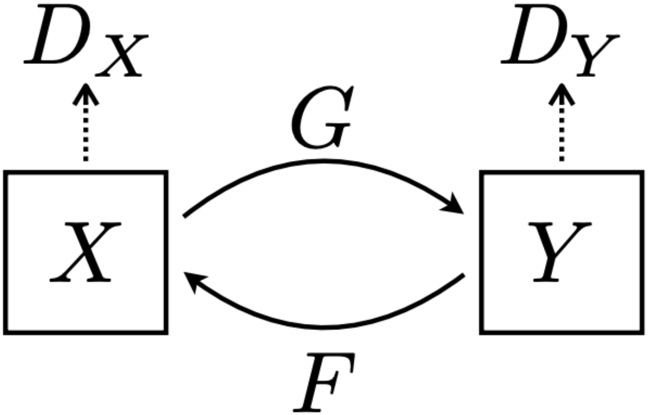
thispersondoesnotexist.com

Cycle-GANs



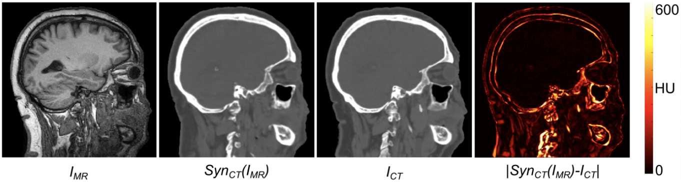
J.-Y. Zhu et al. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232

Cycle-GANs



J.-Y. Zhu et al. (2017). “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232

MR to CT synthesis

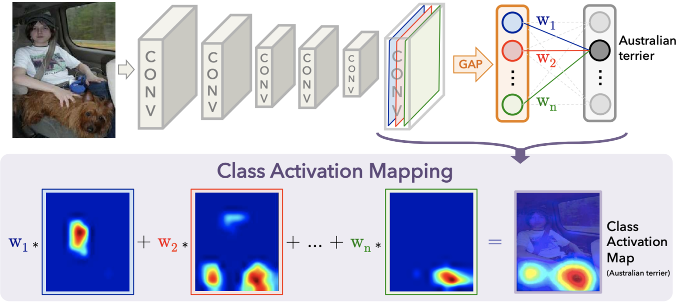


J. M. Wolterink et al. (2017). "Deep MR to CT synthesis using unpaired data". In: *International workshop on simulation and synthesis in medical imaging*. Springer, pp. 14–23

Other application

Interpretable models

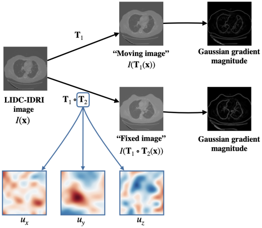
What does the model learn (in human-explainable terms)? Which regions from the image are important for the model output?

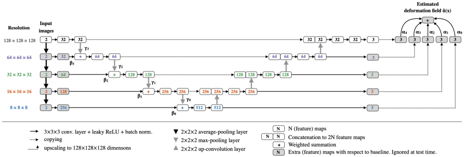


B. Zhou et al. (2016). "Learning deep features for discriminative localization". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929

Image registration

Predict image transformation that registers the moving to the fixed image.

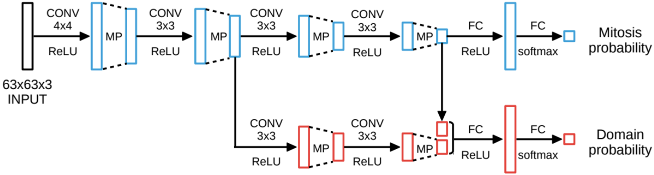




K. A. Eppenhof et al. (2019). "Progressively trained convolutional neural networks for deformable image registration". In: *IEEE transactions on medical imaging*

Adversarial training

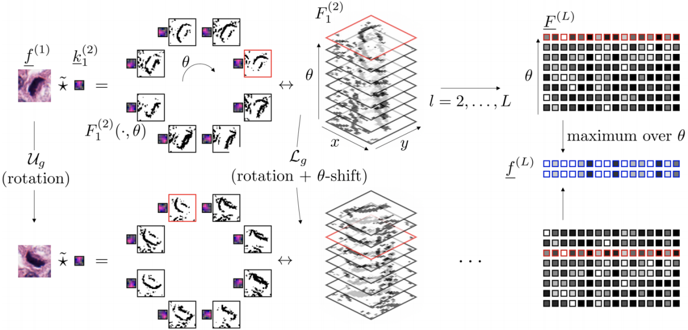
Learn features independent of a confounding factor (such as the domain of origin).



M. W. Lafarge et al. (2017). "Domain-adversarial neural networks to address the appearance variability of histopathology images". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, pp. 83–91

Group convolutions

Roto-translational equivariance.



M. W. Lafarge et al. (2017). "Domain-adversarial neural networks to address the appearance variability of histopathology images". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, pp. 83–91

Outlook for DL and MIA

- Robust models (e.g. to noise and adversarial attacks)
- Deep learning for image acquisition
- Interpretable models
- Prospective clinical validation
- Workflow integration