Application of Deep Learning on Cancer Images

Neo Christopher Chung

Lecture 13, 1000-719bMSB

AI on Medical Images?

- Neural networks demonstrated state-of-the-art performances in natural images.
- Early detection and diagnosis of cancer primarily rely on imaging.
- Yet, cancer diagnosis using neural networks has been lagging behind.

Current state of medical imaging analysis

- Medical images have relied on feature engineering.
- Radiomics (radiology + omics): high-throughput extraction of “engineered” (or “handcrafted”) features from medical images

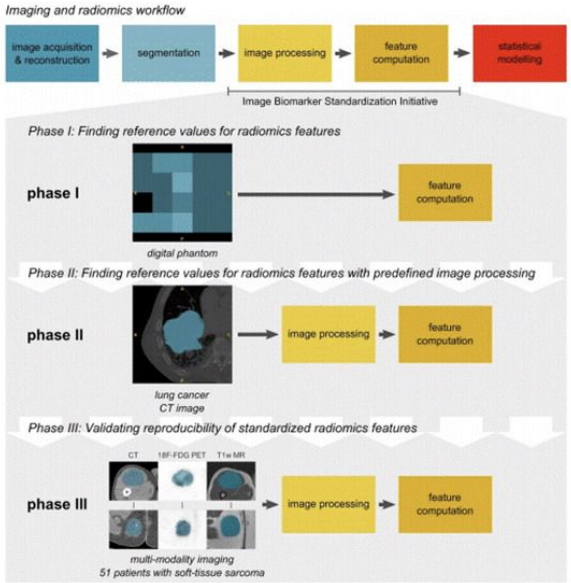
- On one hand, there’s a continued development to utilize radiomic features in downstream ML classifiers, including using neural networks
- On the other hand, there’s many approaches to utilize end-to-end neural networks trained directly on medical images.

Radiomics Papadimitroulas et al 2021

1. Shape features: provide quantitative description of geometric properties of the ROIs/VOIs, such as surface area, total volume, diameter, sphericity or surface-to-volume ratio.
2. First order statistics (histogram-based features): describe the fractional volume for the selected region of voxels and the distribution of the voxels' intensity, for example minimum, maximum, mean, variance, skewness, or kurtosis.
3. Second order statistics (textural features): These features are extracted based on matrices derived from intensity relationships of neighboring voxels in a 3D image
4. Higher order statistics features: These features are obtained by statistical methods after applying filters or mathematical transformations to the image

Standardized Quantitative Radiomics

The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping

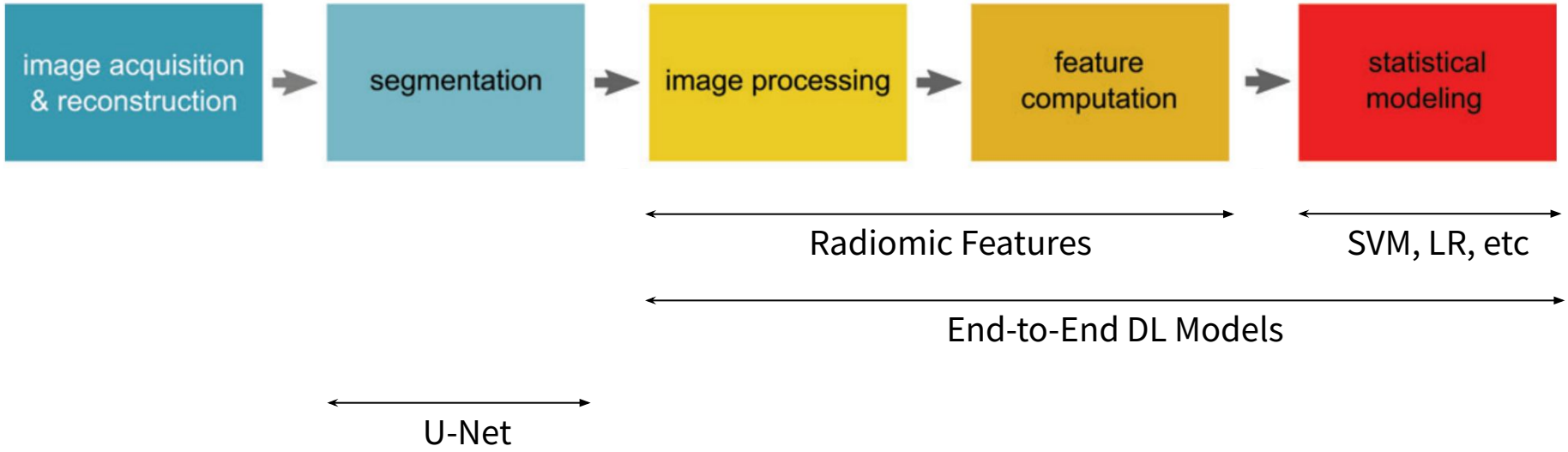


- Twenty-five research teams found agreement for calculation of 169 radiomics features derived from a digital phantom and a human lung cancer on CT scans
- Among these 169 candidate radiomics features, good to excellent reproducibility was achieved for 167 radiomics features by using MRI, fluorine 18 fluorodeoxyglucose PET, and CT images obtained in 51 patients with soft-tissue sarcoma

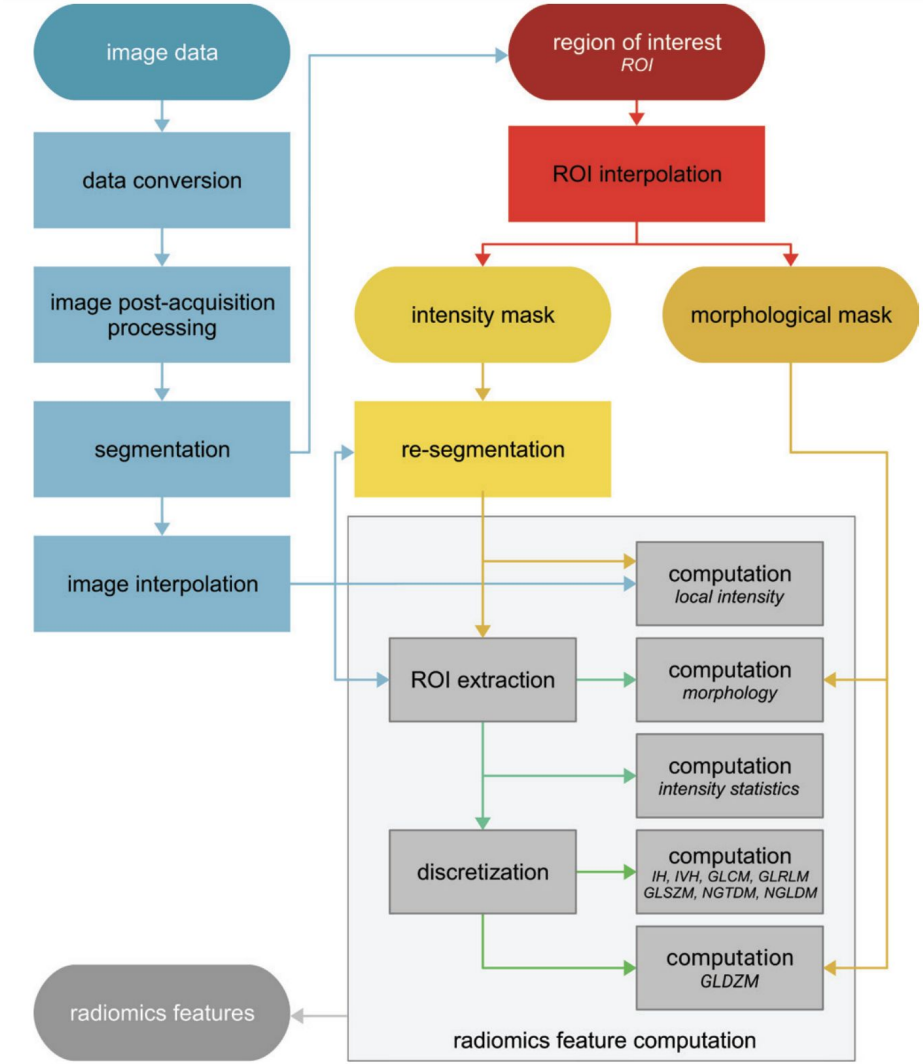
Zwanenburg A et al. Published Online: March 10, 2020
https://doi.org/10.1148/radiol.2020191145

Radiology

Imaging and radiomics workflow



Zwanenburg et al. 2020



Data conversion (eg, conversion to standardized uptake values)
Image post-acquisition processing (eg, image denoising)

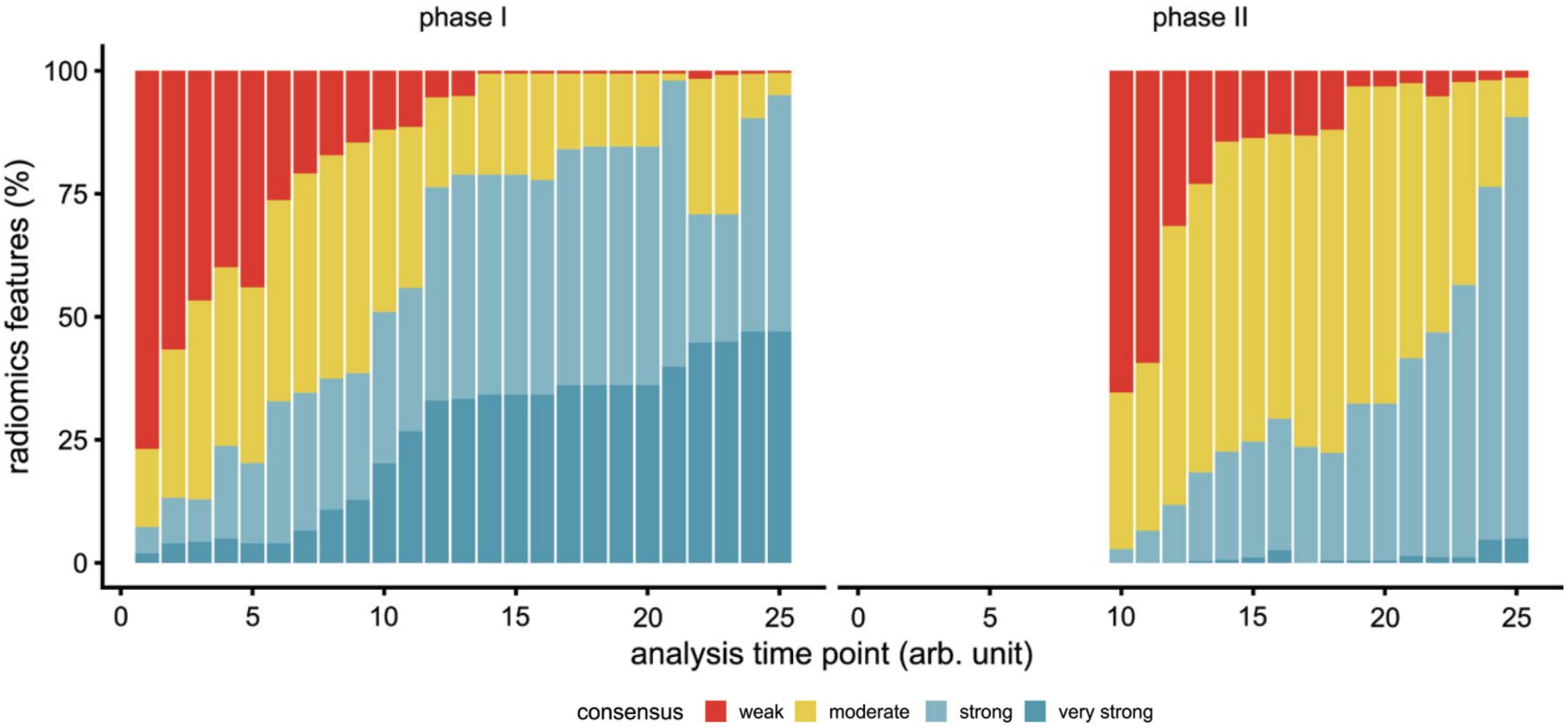
Region of interest (ROI) is created automatically during the **segmentation step (human annotation or predicted segmentation)**, or an **existing ROI** is retrieved

Radiomics features are then computed from the image masked by the ROI and its immediate neighborhood (local intensity features) or the ROI itself (all others).

Separate discretization or processing prior to specific radiomic feature computation

Zwanenburg et al. 2020

Figure 4(a). The overall development of consensus on the validity of (tentative) reference values in phases I and II



Zwanenburg et al. 2020

Deep Learning Radiomics?

- We can use a neural network to extract certain ‘higher order’ features.
- For example, the shallow layer may extract basic features and the deep layer may extract high level conceptual features.
- Can these DL radiomic features be standardized?
- Would the end-to-end DL model perform better?
- How about interpretability?

Segmentation

The purpose of medical image segmentation is to find structures of interest, such as tumors and lesions, and marking the constituting pixels with the same label. Deep learning techniques have proven to be very effective in this task and segmentation is in fact the problem which is most commonly tackled using CNNs

The most well-known CNN architecture used for segmentation for medical images is U-net (and its newer variant nnU-net), which uses upsampling convolutional layers to obtain segmentation maps with the same resolution as the input. This architecture allows training the model using entire images end-to-end, which allows the model to utilize the whole context of the image.

Ronneberger et al. 2015 Convolutional Networks for Biomedical Image Segmentation https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28

Isensee et al. 2021 nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation https://www.nature.com/articles/s41592-020-01008-z

U-Net

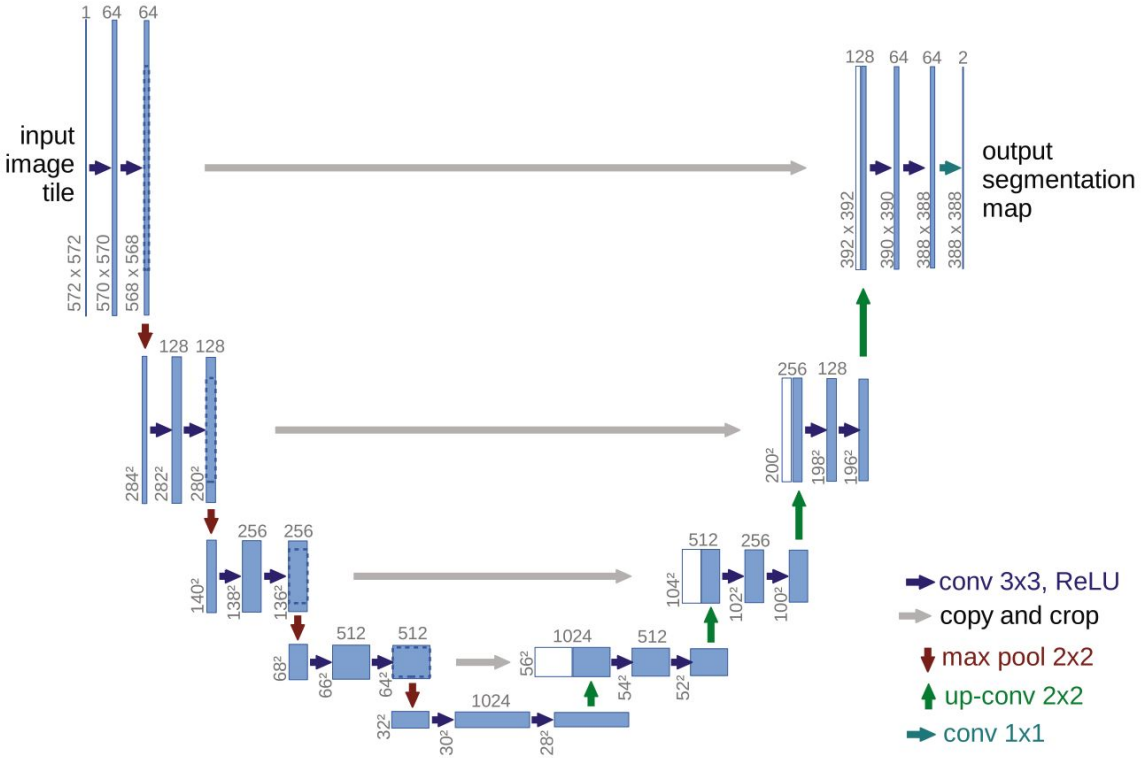


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

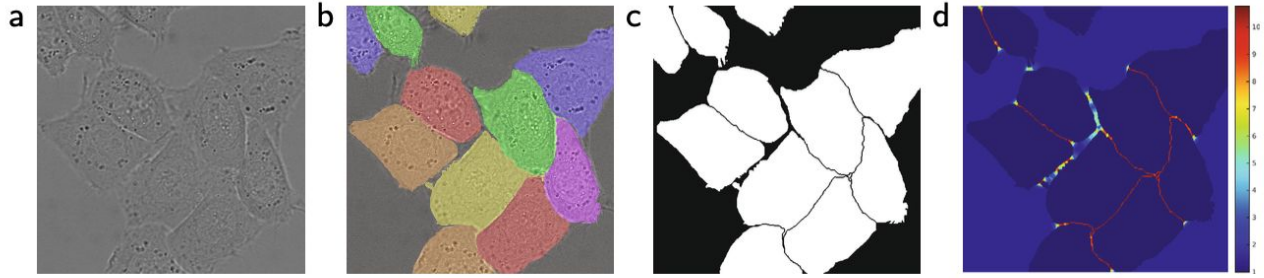


Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (**a**) raw image. (**b**) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (**c**) generated segmentation mask (white: foreground, black: background). (**d**) map with a pixel-wise loss weight to force the network to learn the border pixels.

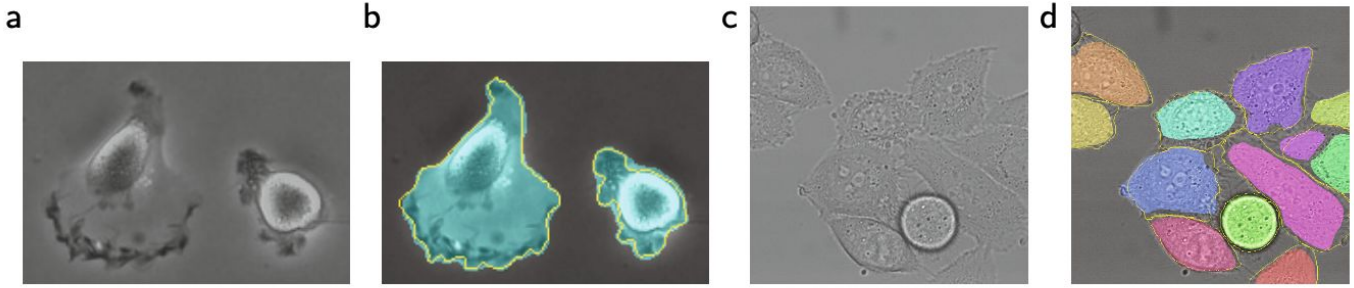


Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).

Batch effects, everywhere!

Compared to gene expression, medical images are much more variable.

- Age, sex, and many other demographic variables
- Different hospitals and imaging systems
- Acquisition protocols and reconstruction algorithms
- Countries and their different approaches (both technical and human factors)

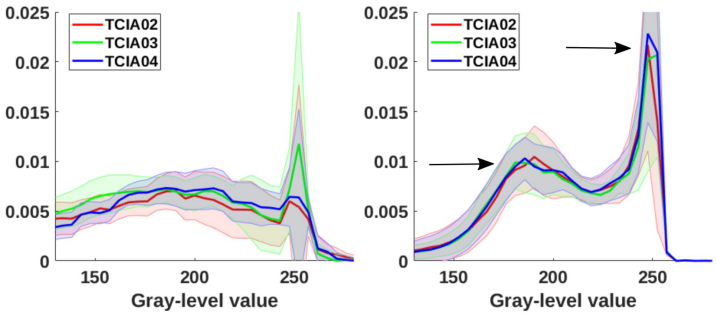


Fig. 2: Kernel density estimations of graylevel values within tumor masks for 3 centers. Curves show median±standard deviation. (a) before standardization. (b) after standardization.

Batch effects (also called Harmonization)

Most studies have been carried out using small, retrospective and monocenter cohorts of patients.

The developed models are rarely tested on external datasets, even less often on several ones.

Collecting data from several centers/countries is however complex for legal, ethical, administrative and technical reasons.

How can we normalize various statistics (e.g., radiomic features) that are confounded by batch effects? → covariates in regression, location & scale, ComBat?

But perhaps, the better solution would be looking at the source of batch effects
→ How can we normalize (harmonize) imaging data across multiple medical centers?

Using DL to remove batch effects?

Let A be a source multicentric domain (a clinical dataset containing heterogeneous images from multiple centers) and B a target reference domain (healthy subjects from homogeneously-calibrated acquisitions)

Our objective is to learn an implicit mapping $\phi : A \rightarrow B$ conditioned on an input image I such that translated images $J = I \circ \phi$ shows reduced center effect while preserving the clinical characteristics of A.

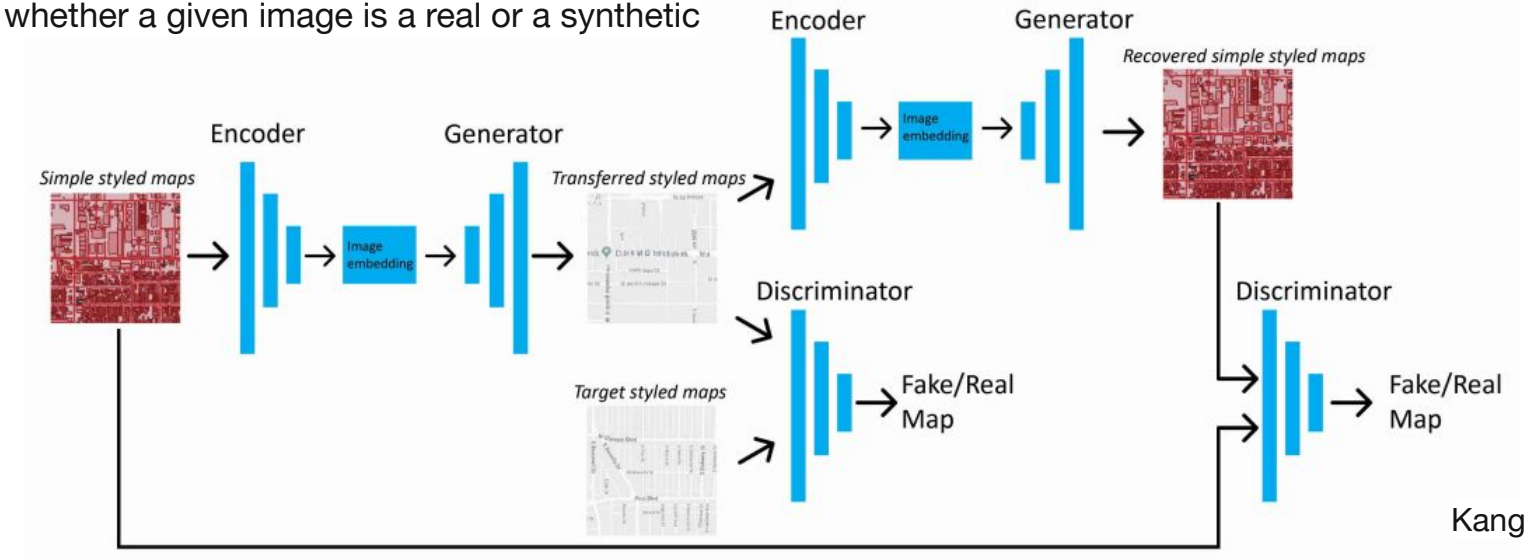
We first learn a mapping $\phi_U : A \rightarrow B$ using an unpaired **image-to-image translation network (CycleGAN architecture)** between A and B. CycleGAN can learn domain transfer without paired training data due to a cyclic loss where translated images $\phi_U(A)$ are back-translated to their original domain, thereby enforcing similarity to the original data.

Hognon et al. Standardization of Multicentric Image Datasets with Generative Adversarial Networks IEEE NSS/MIC 2019

CycleGAN Zhu et al. ICCV 2017 https://junyanz.github.io/CycleGAN/

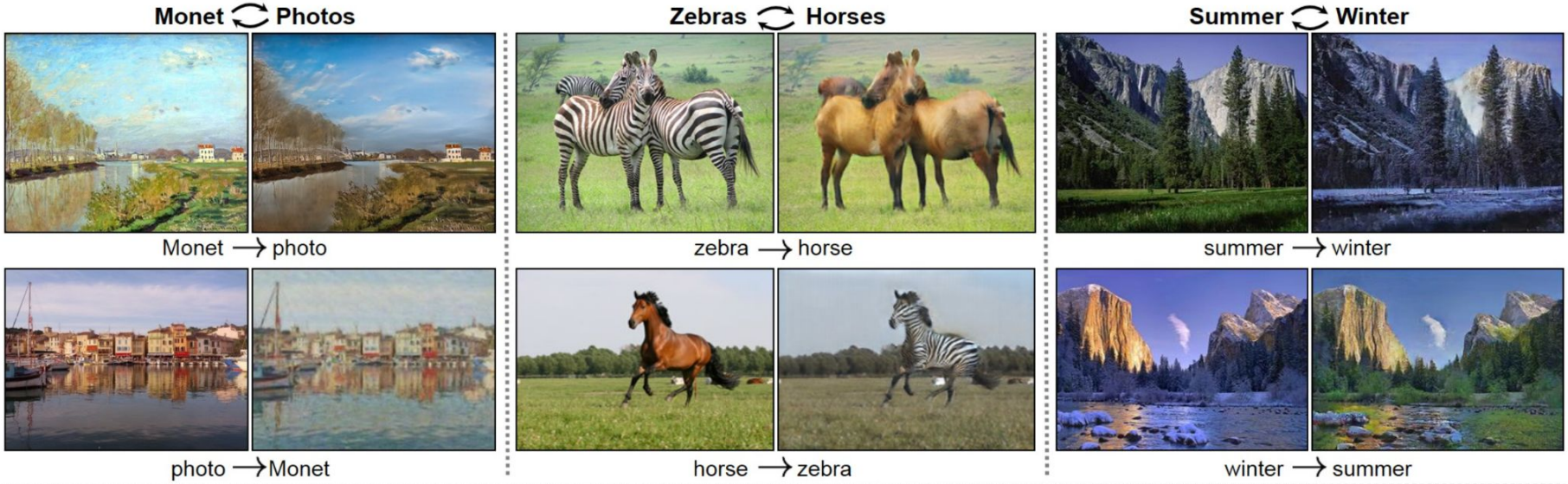
CycleGAN: image-to-image translation model, specifically designed to translate an image from a source domain X to a target domain Y **in the absence of paired examples**. The key innovation is a cycle consistency loss to push F(G(X)) ≈ X (and vice versa).

CycleGAN is an extension of generative adversarial network (GAN), simultaneous training of a generator models and a discriminator models. In GAN, a generator model creates a synthetic image and a discriminator learns (and predicts) whether a given image is a real or a synthetic



Kang et al 2019

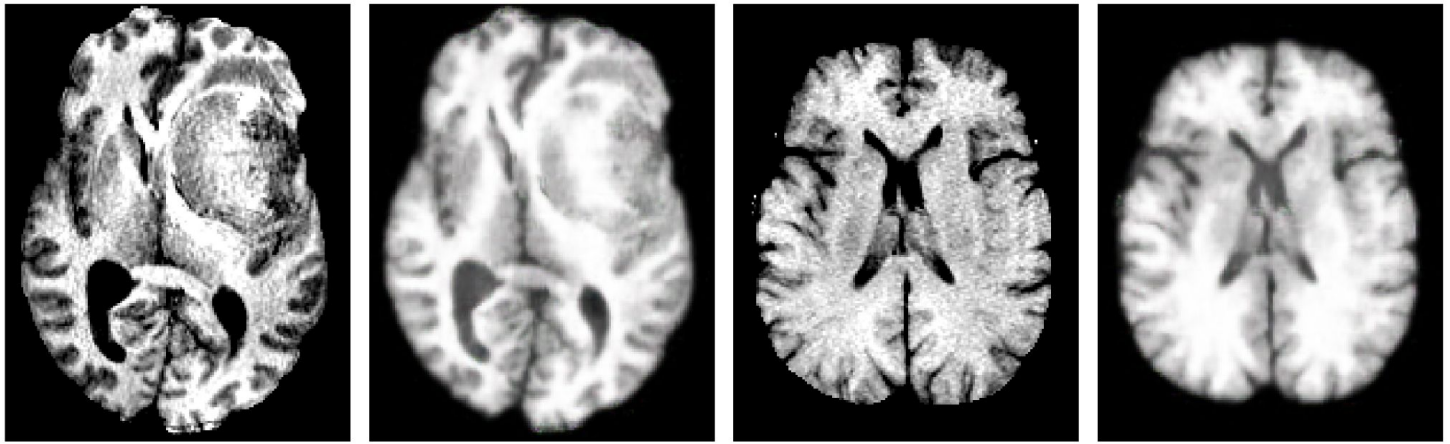
CycleGAN Zhu et al. ICCV 2017 https://junyanz.github.io/CycleGAN/



The loss function in CycleGAN: adversarial loss (as in GAN), cycle consistency loss (below), and identity loss (for color composition).

The overall model translate from an original image **X** (in the source domain) to another image**G(X)** (in the target domain), which is translated back into the source domain **F(G(X))**. Cycle consistency loss = L_1 distance between **X** and **F(G(X))**

Using DL to remove batch effects?

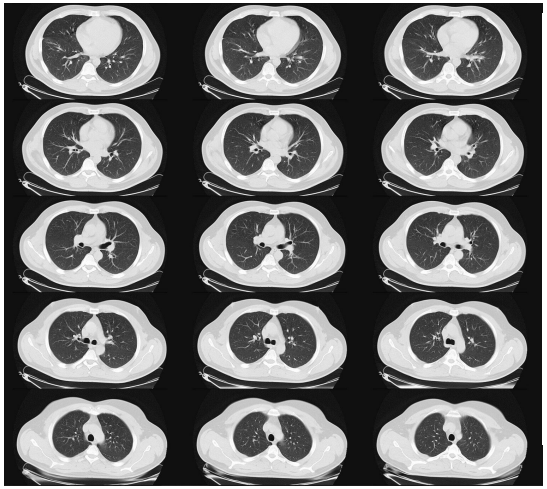


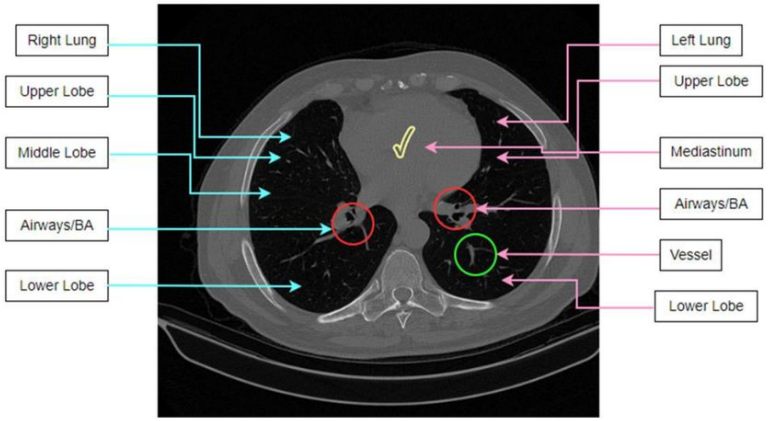
(a) (b) (c) (d)

Fig. 1: Example standardization results of the BRATS dataset (a,c) original images (b,d) standardized images.

Hognon et al. Standardization of Multicentric Image Datasets with Generative Adversarial Networks IEEE NSS/MIC 2019

Lung CT (computed tomography) scans





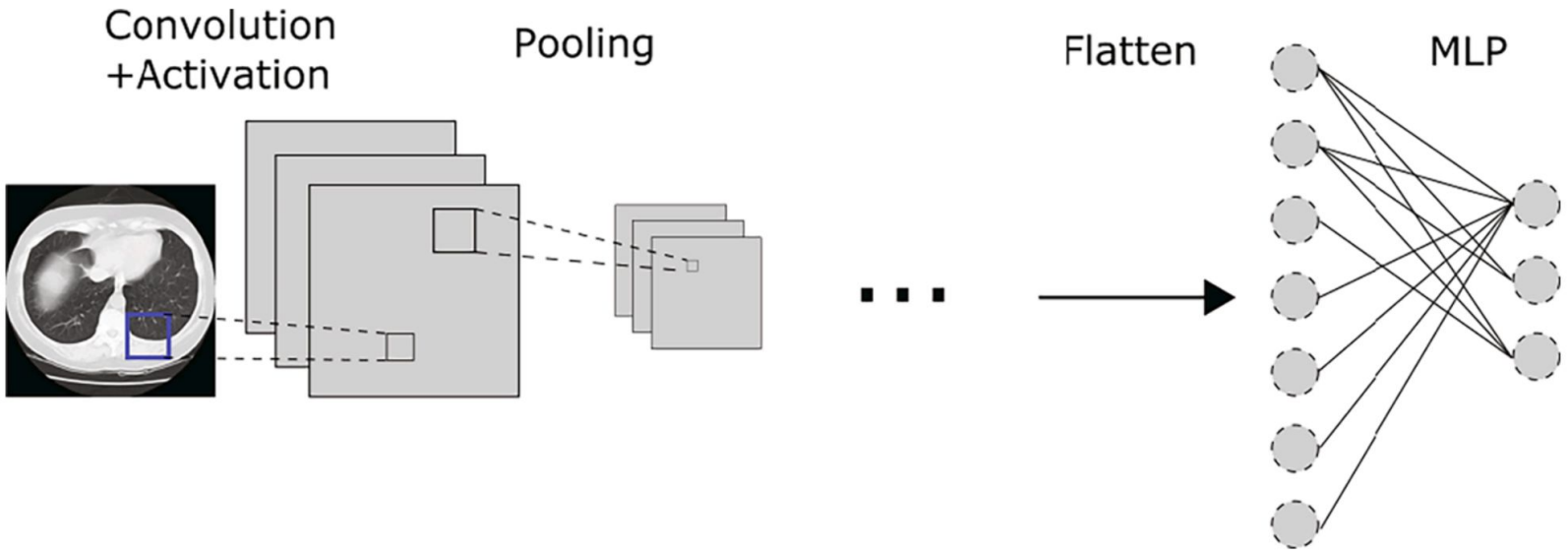
Lung CT (computed tomography) scans

Find a Nodule slice

Extract a Nodule

Classification

We look at classification from having found ROI and cropped the medical image (CT scan) appropriately.



Making a DL Classifier

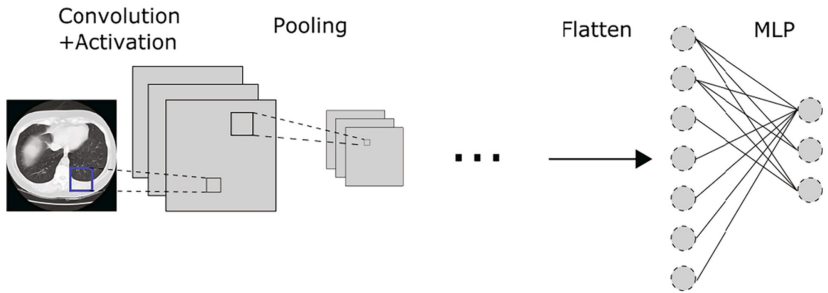
Architecture, # layers, activation functions, etc

Training hyperparameters (learning rates, etc)

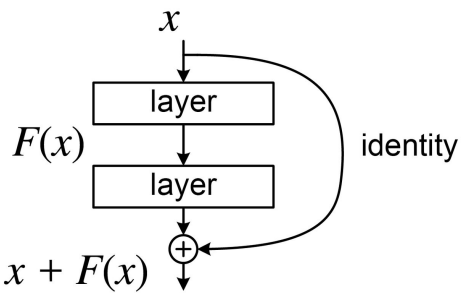
Evaluation metrics

Incorporating clinical variables (i.e., biomarkers)

Image scale, batch normalization, regularization



Finetuning a ResNet-50 model



ResNet (Residual Networks) designed to improve training **deep** neural networks
He et al 2015 https://arxiv.org/abs/1512.03385

"Residual connections": used in LSTM, transformer (e.g., ChatGPT), AlphaFold

ResNet-50: 50 bottleneck residual blocks, architecture
& pre-trained model In pytorch (torchvision.models.resnet50)

Training strategy: input samples upscaled to 224×224, z-normalization, 50 epochs using the Adam optimizer, learning rate of 10−3 (reduced by x0.1, at 20 and 40 epochs)

The final accuracy of **0.891**