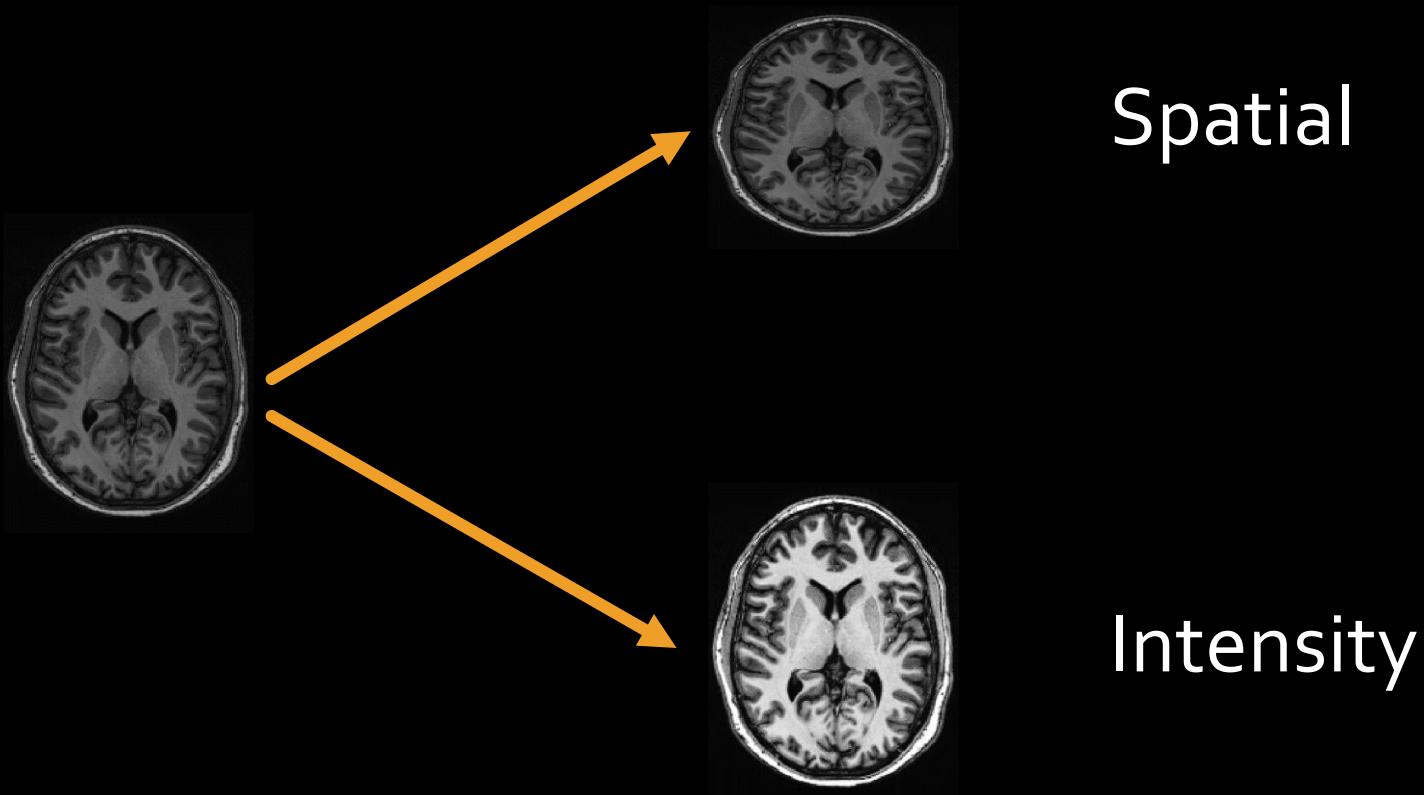


Data Augmentation and Preprocessing

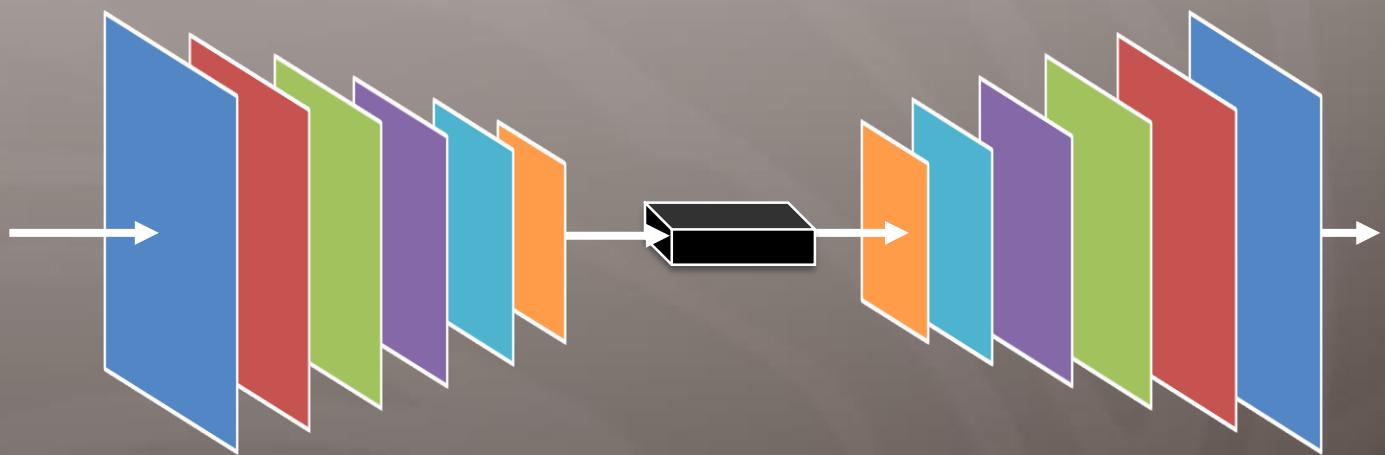
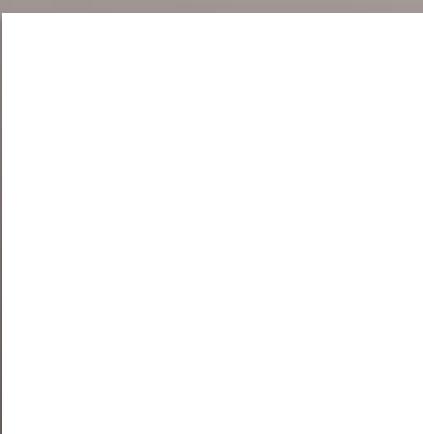
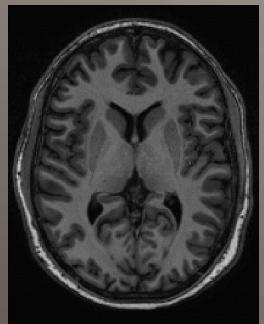
[Spring 2020 CS-8395 Deep Learning in Medical Image Computing]

Instructor: Yuankai Huo, Ph.D.
Department of Electrical Engineering and Computer Science
Vanderbilt University

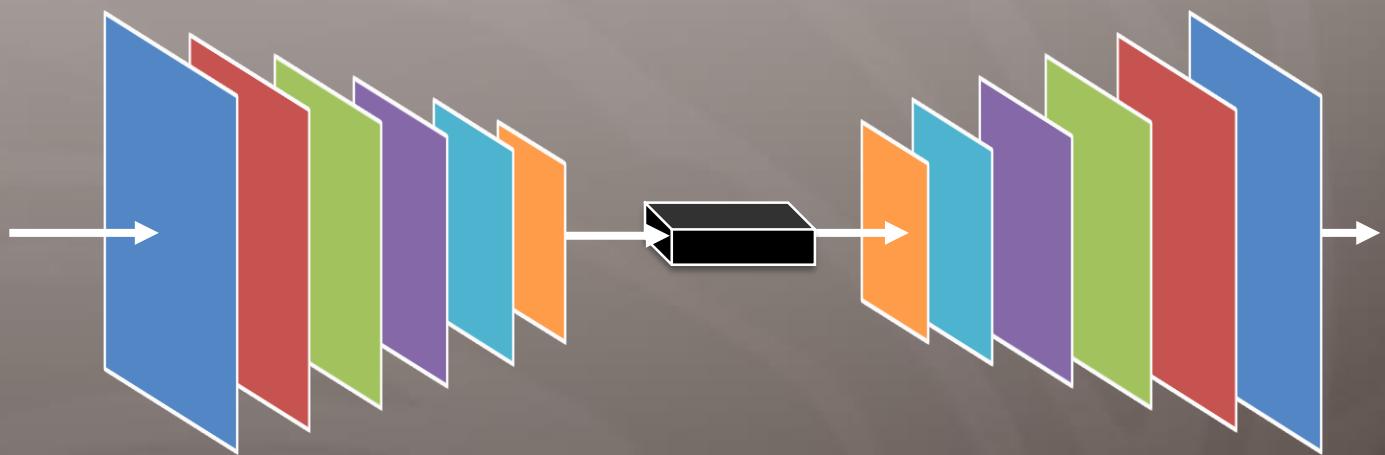
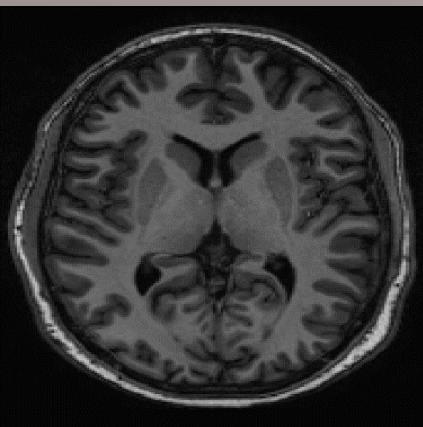
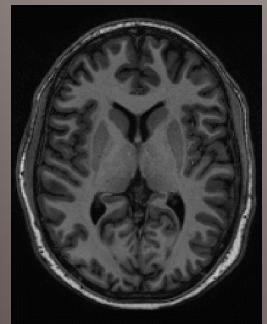
Preprocessing



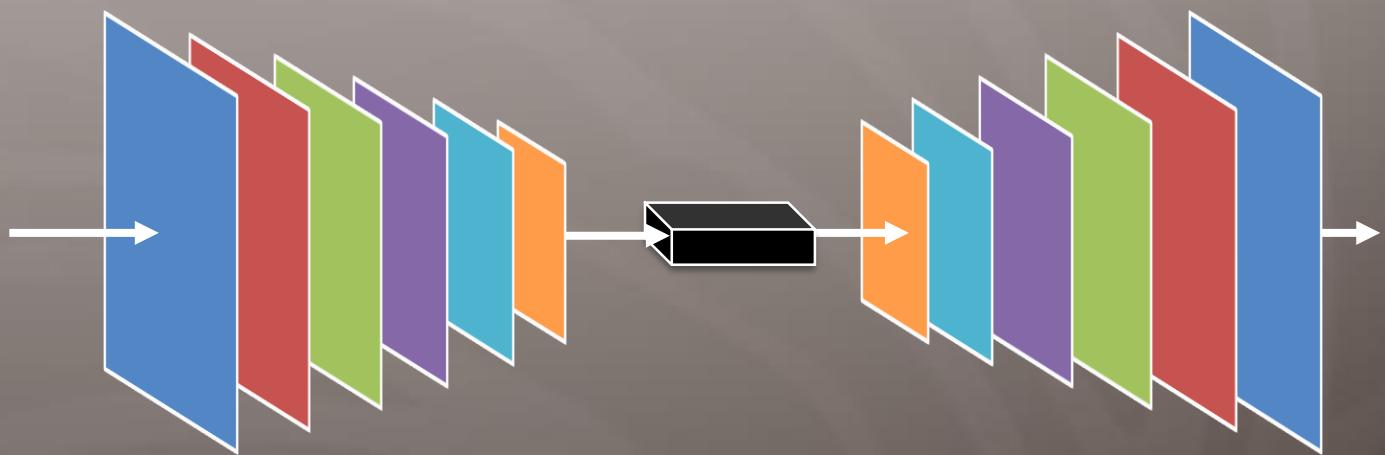
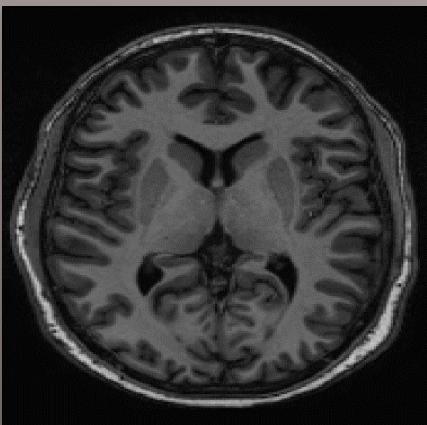
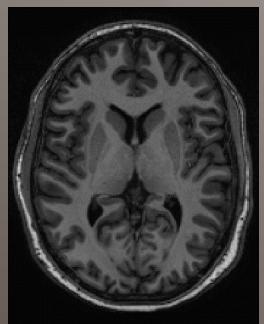
Preprocessing



Preprocessing

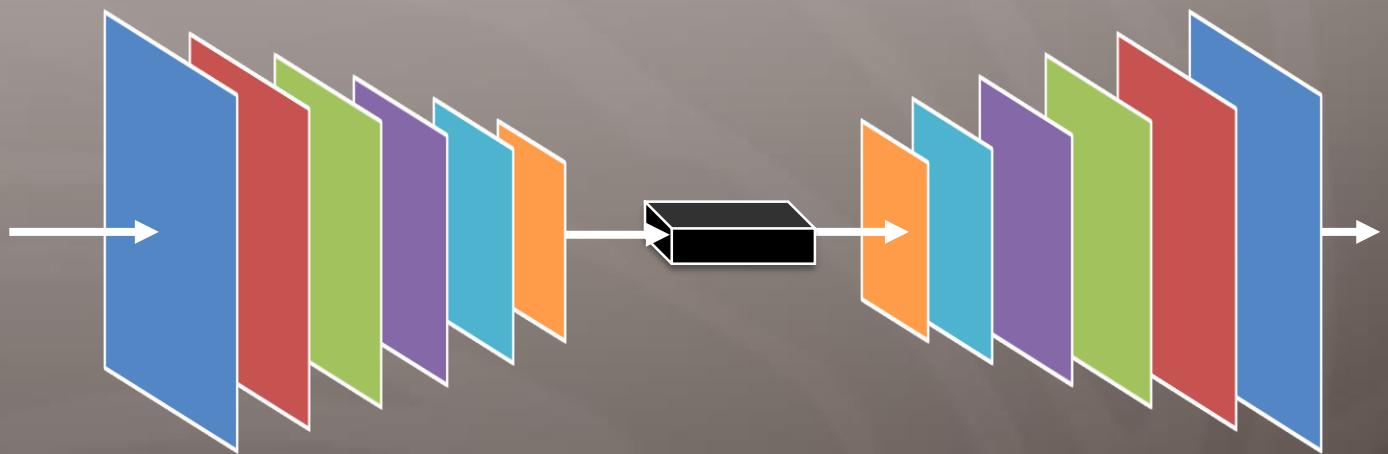
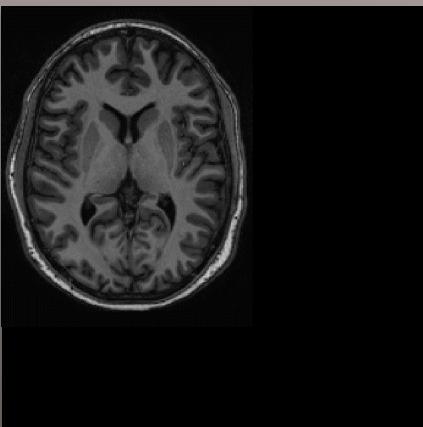
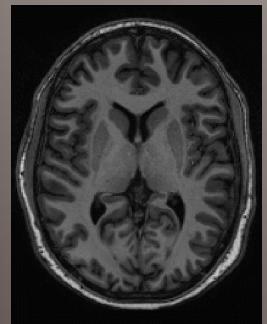


Preprocessing

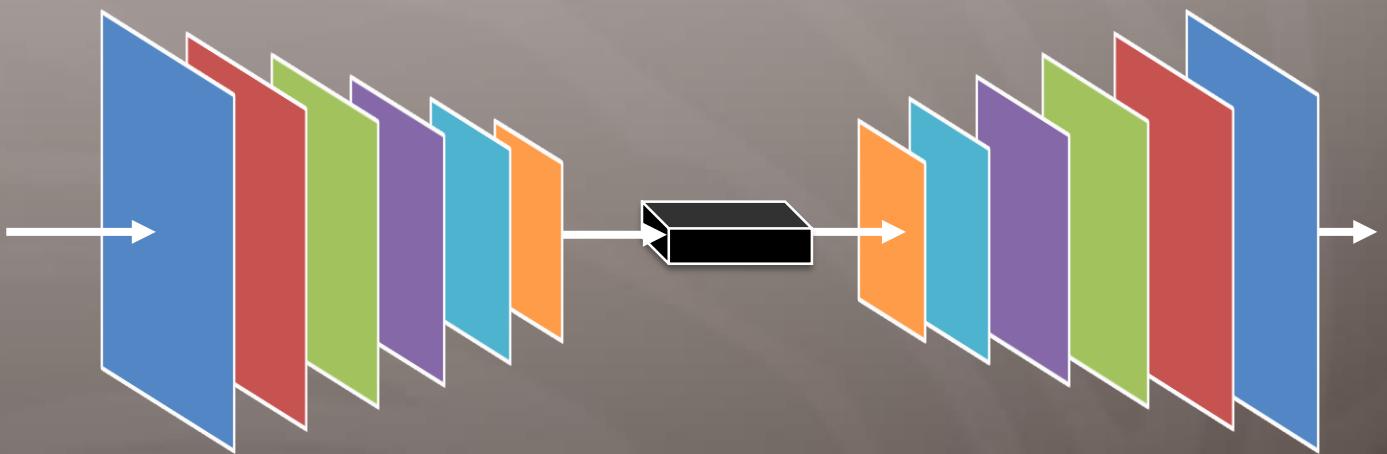
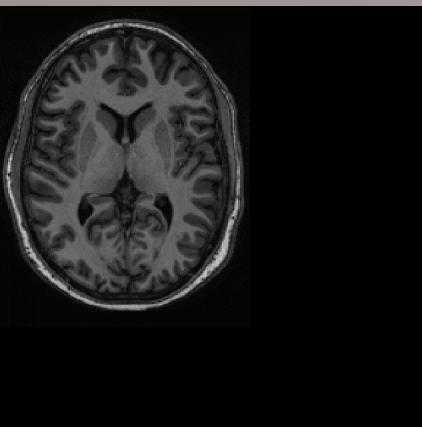
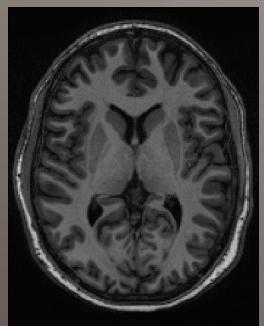


```
cur_cnt = np.reshape(img, self.shape)
```

Preprocessing

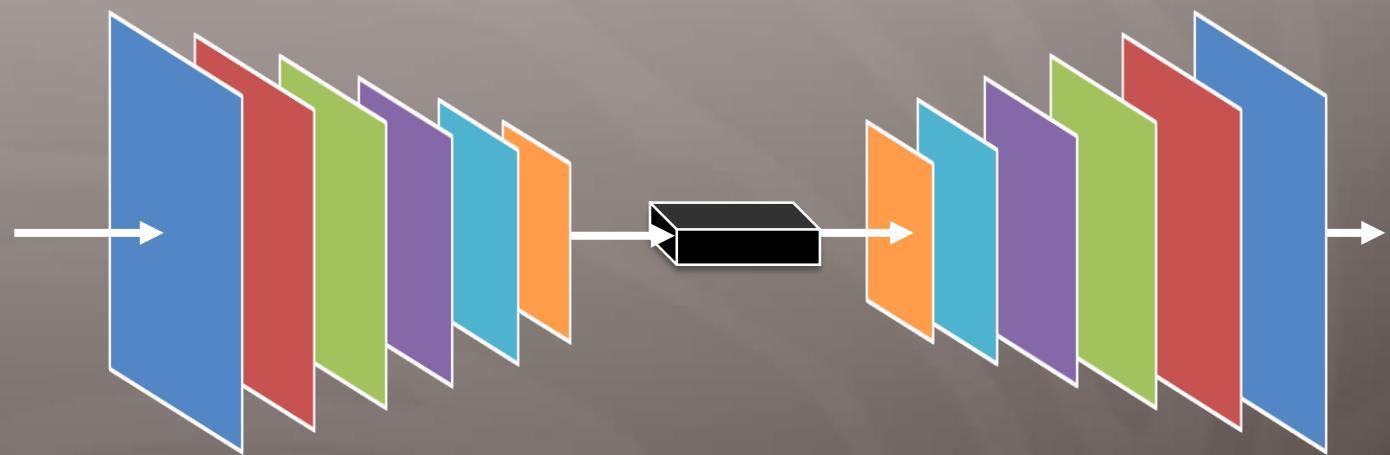
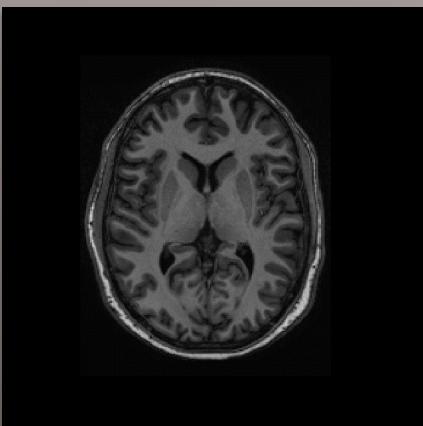
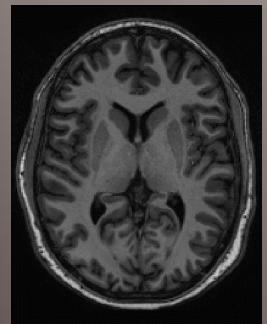


Preprocessing

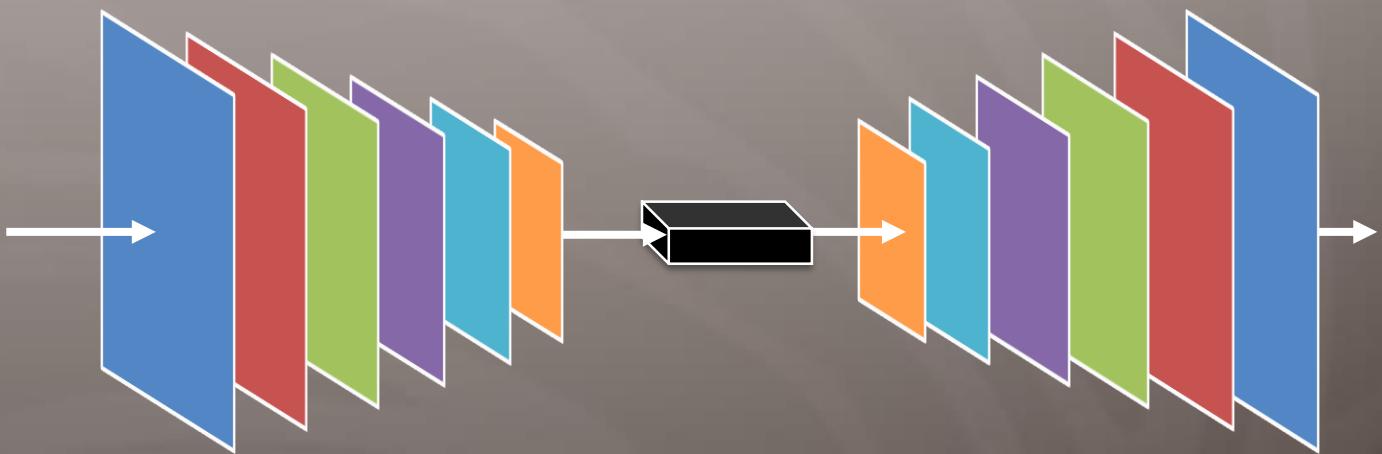


```
cur_cnt[:min(hh, nRows), :min(ww, nCols)] = img[:min(hh, nRows), :min(ww, nCols)]
```

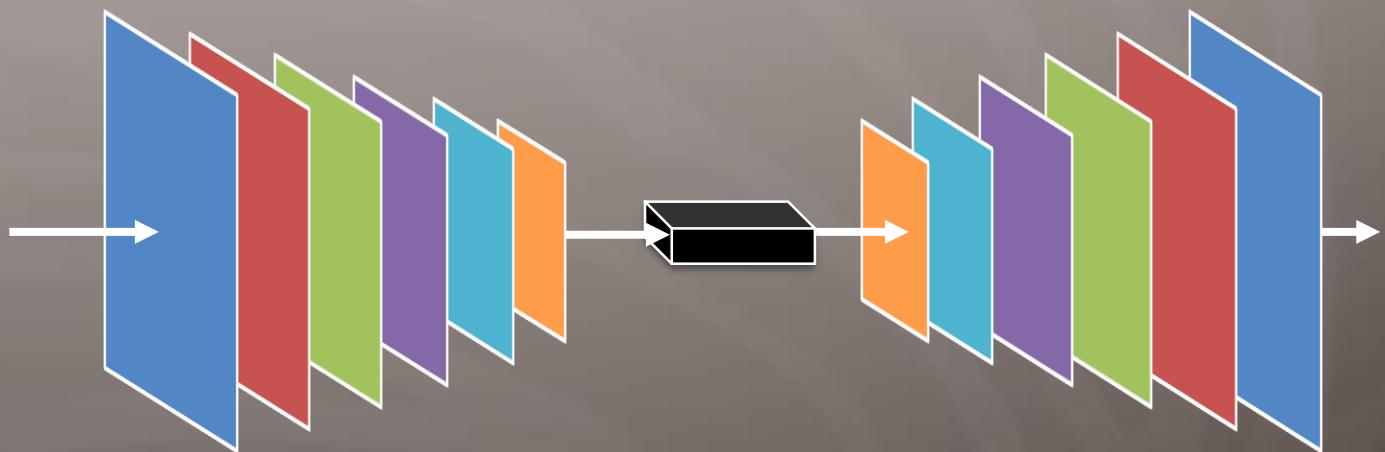
Preprocessing



Preprocessing

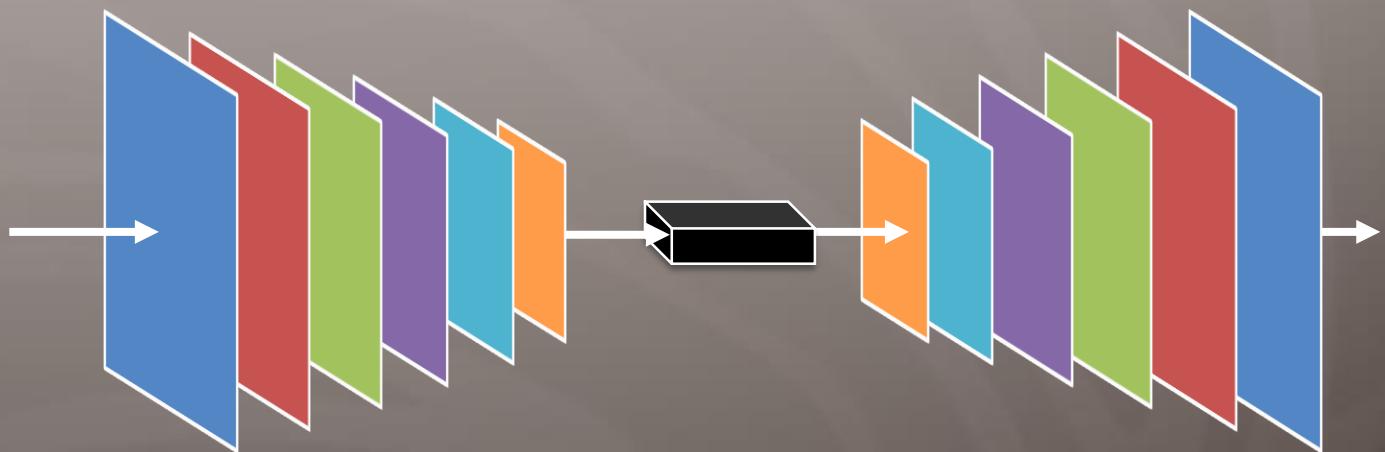


Preprocessing

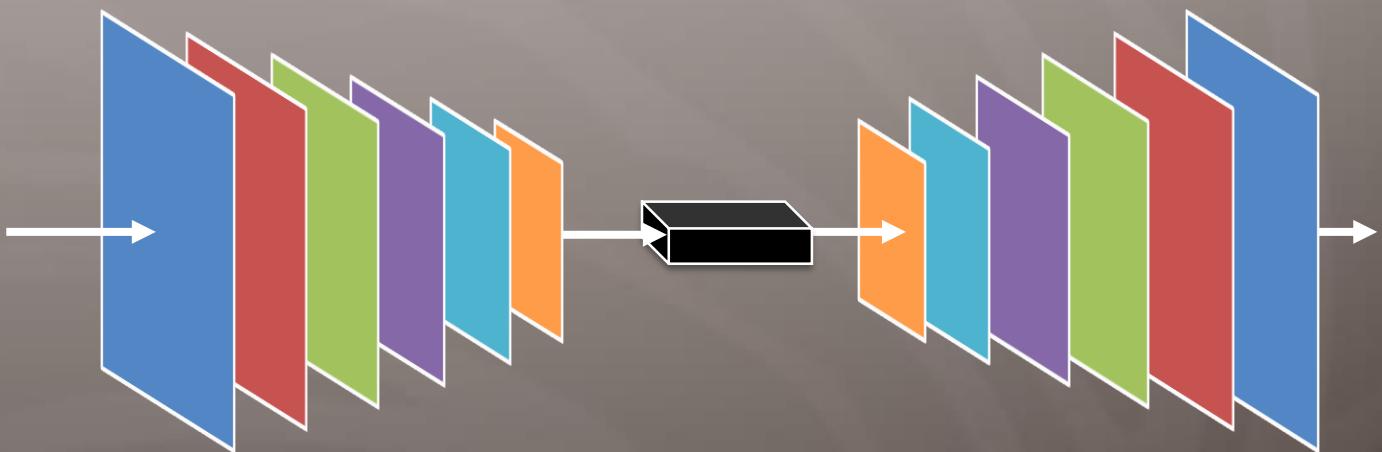


```
cur_cnt = np.reshape(img, self.shape)
```

Preprocessing



Preprocessing



```
cur_cnt[:max(hh, nRows), :max(ww, nCols)] = img[:max(hh, nRows), :max(ww, nCols)]
```

Ultimate Tool: Registration

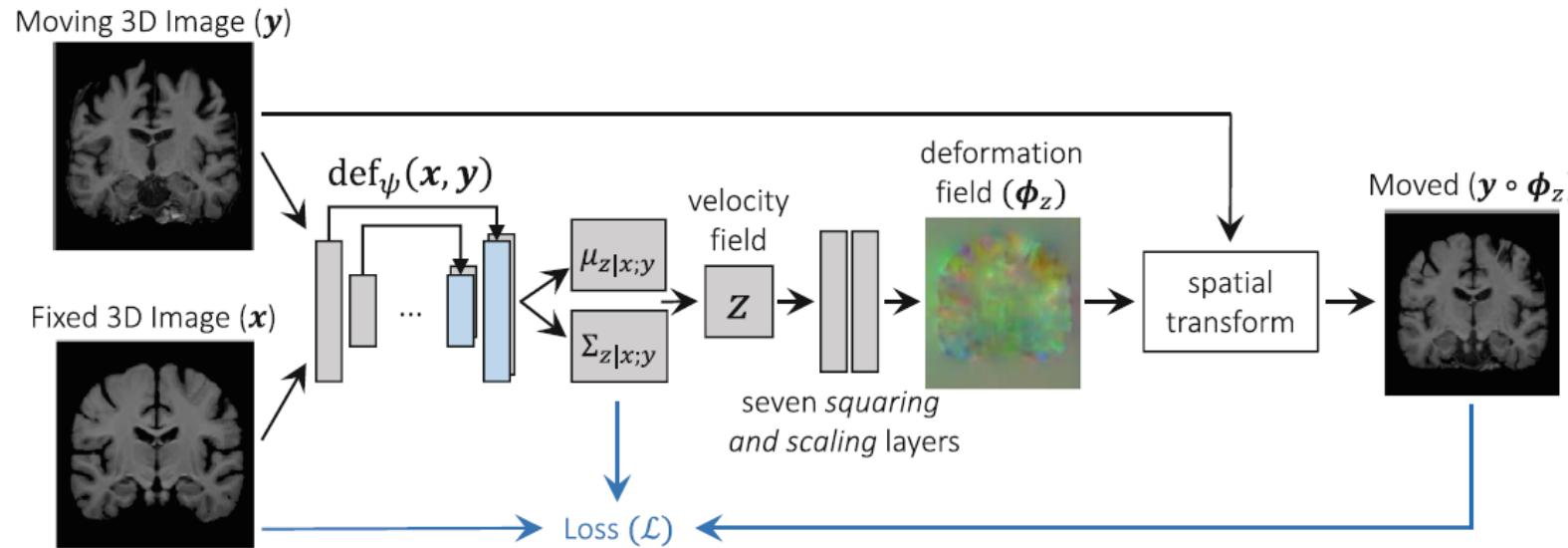
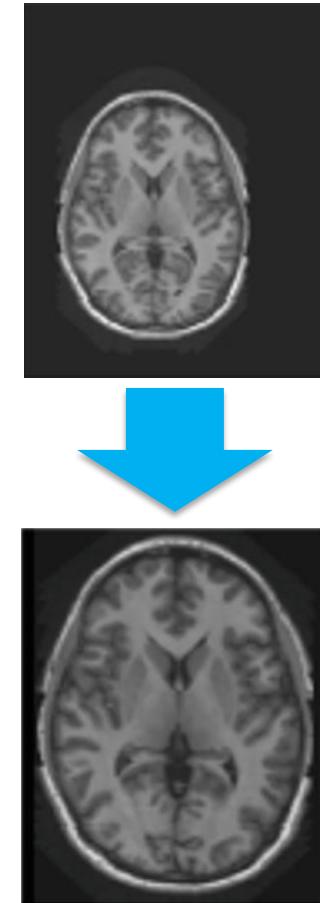
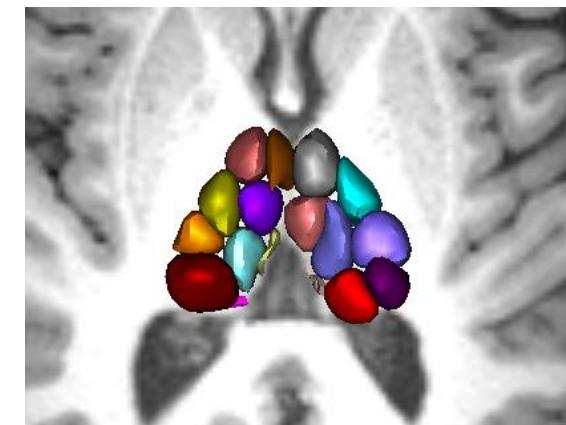
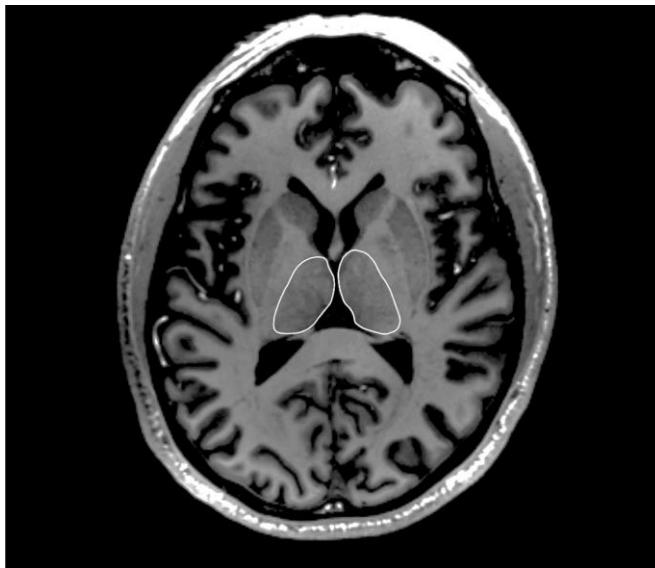


Fig. 1. Overview of end-to-end unsupervised architecture. The first part of the network, $\text{def}_\psi(\mathbf{x}, \mathbf{y})$ takes the input images and outputs the approximate posterior probability parameters representing the velocity field mean, $\mu_{z|x;y}$, and variance, $\Sigma_{z|x;y}$. A velocity field \mathbf{z} is sampled and transformed to a diffeomorphic deformation field ϕ_z using novel differentiable *squaring and scaling* layers. Finally, a spatial transform warps \mathbf{y} to obtain $\mathbf{y} \circ \phi_z$.

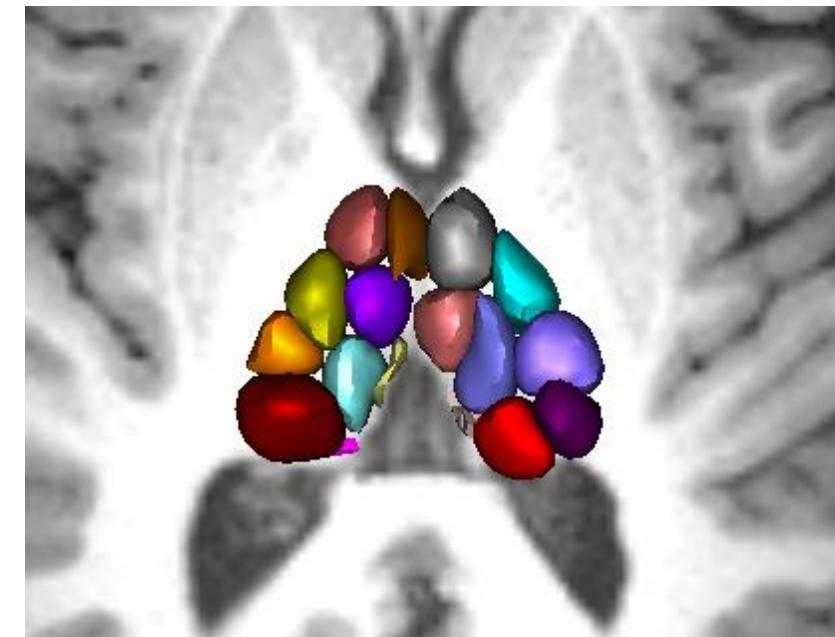
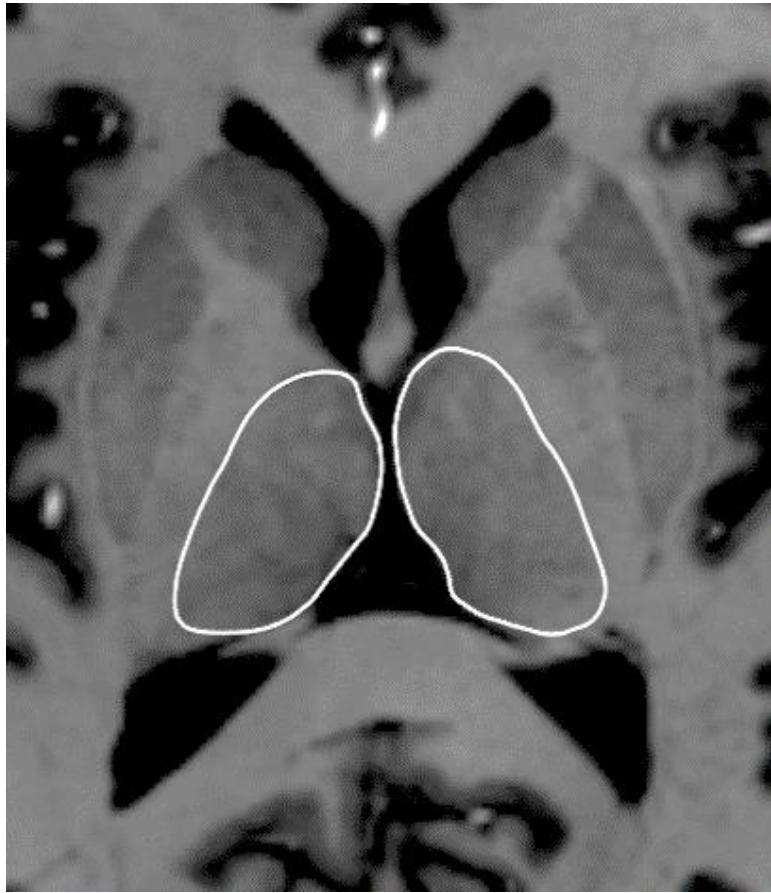


For Example

- Thalamus is a deep brain structure responsible for modulating the flow of information between the cerebral cortex and the basal ganglia
- Functionally and structurally diverse – different nuclei are part of different cortical relay networks
- However, they are not visible in clinical T1 or T2 scans. The thalamus has uniform intensity.



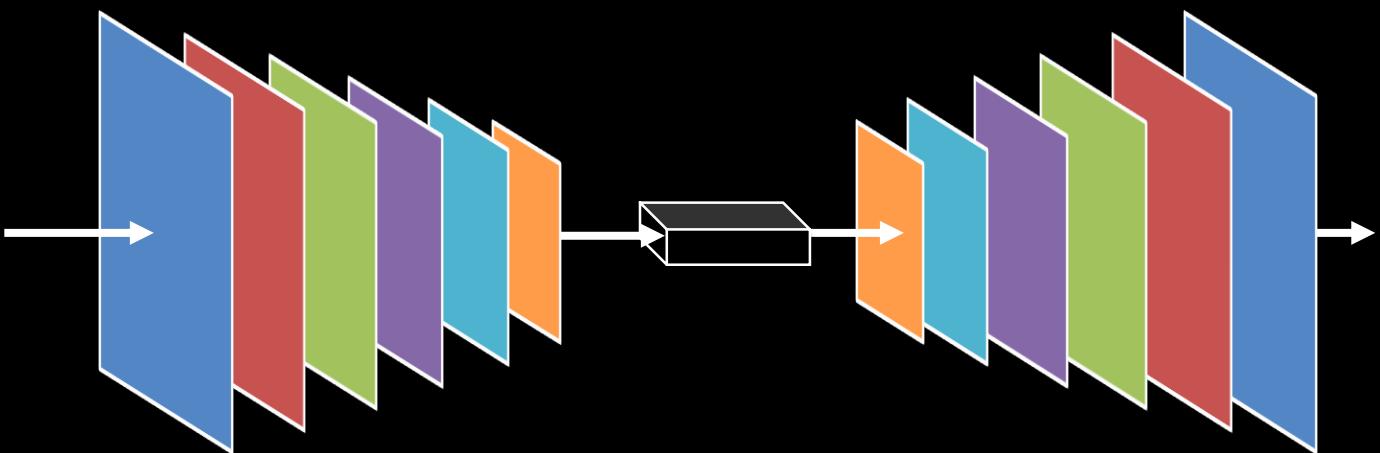
ROI Extraction



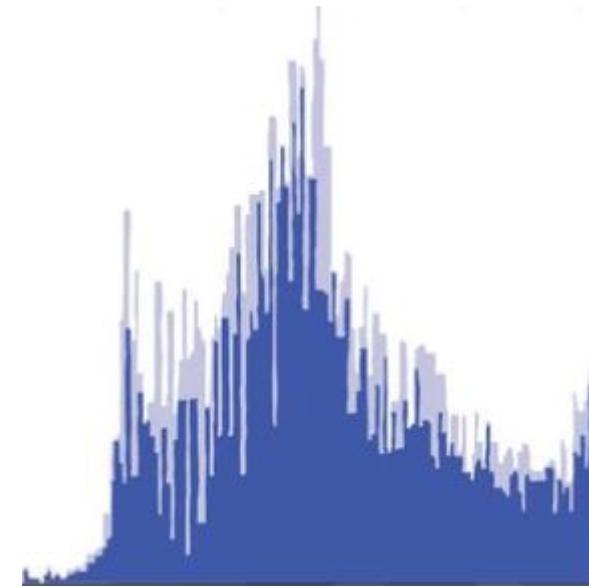
Ultrasound



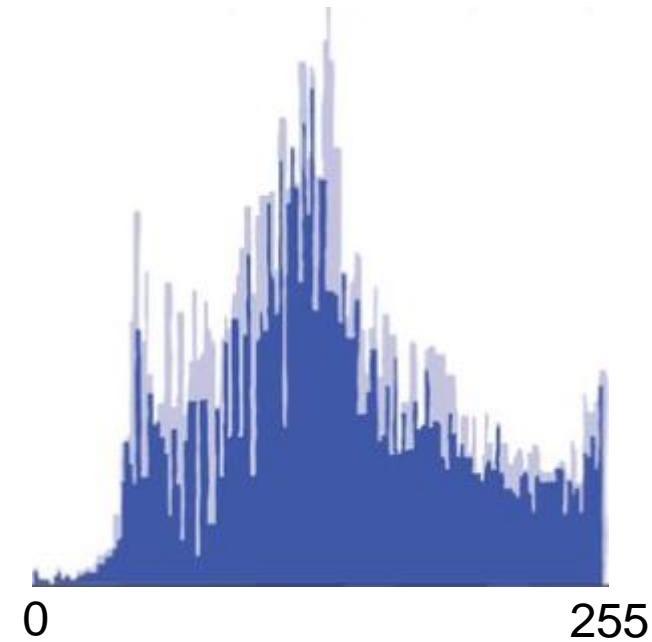
0~255



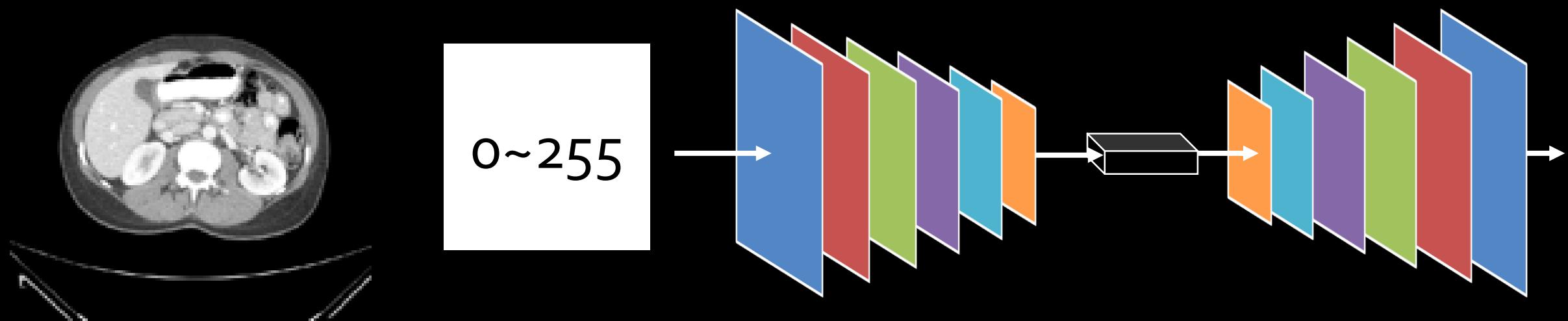
Ultrasound



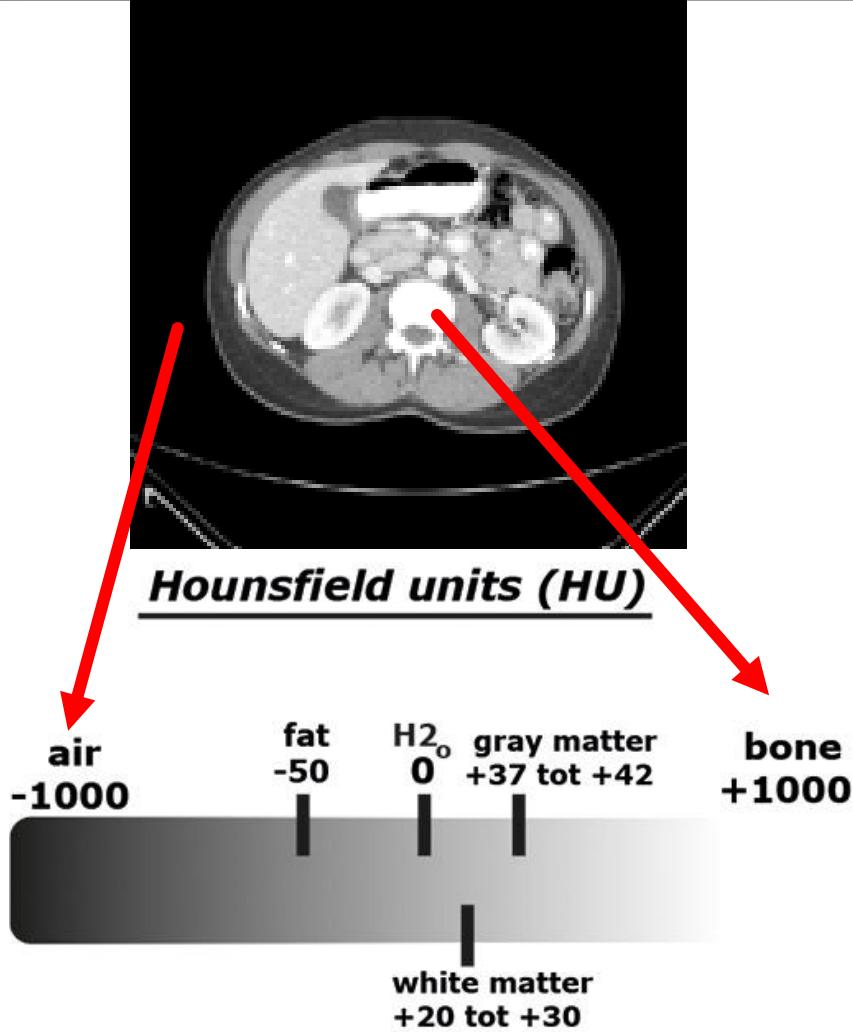
Ultrasound



CT



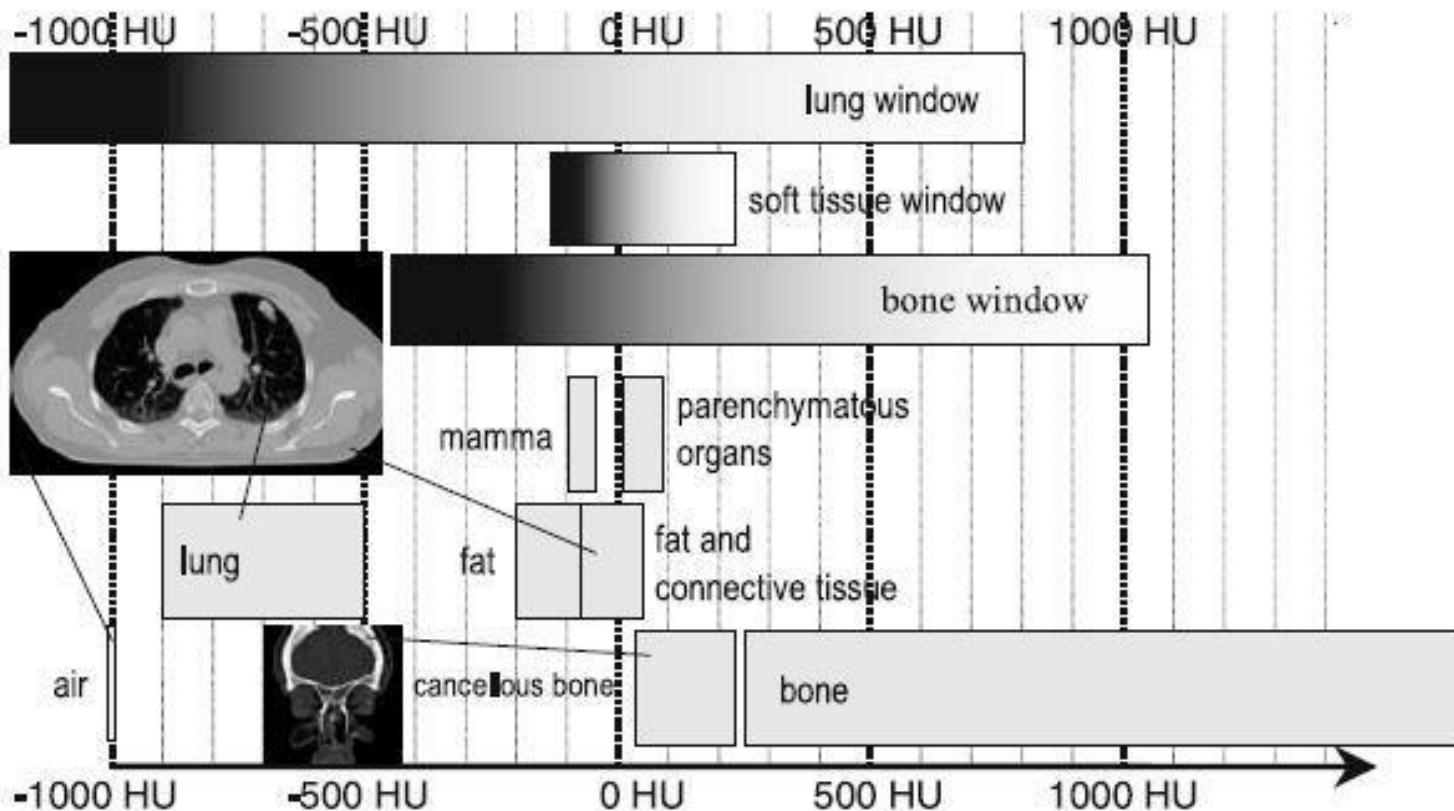
Hounsfield Unit



$$HU = \left(\frac{\mu_X - \mu_{Water}}{\mu_{Water} - \mu_{Air}} \right) \times 1000$$

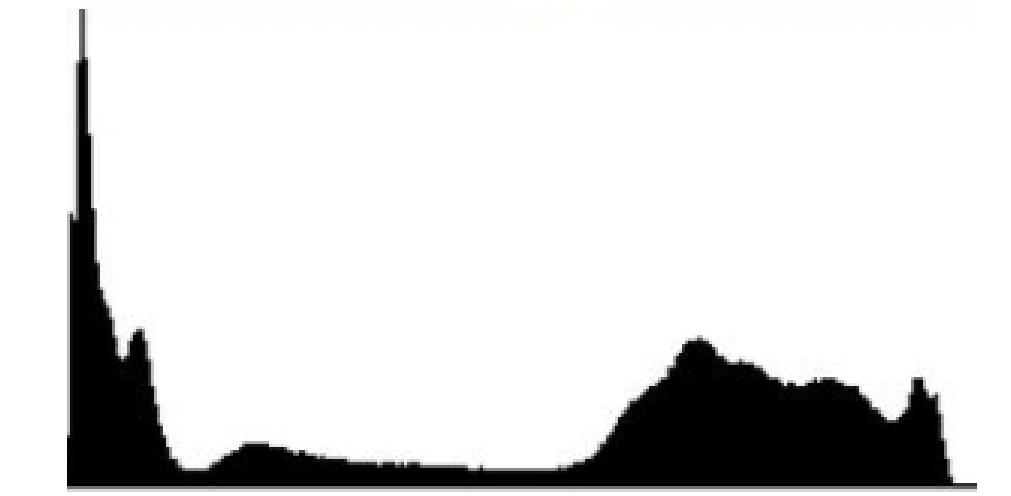
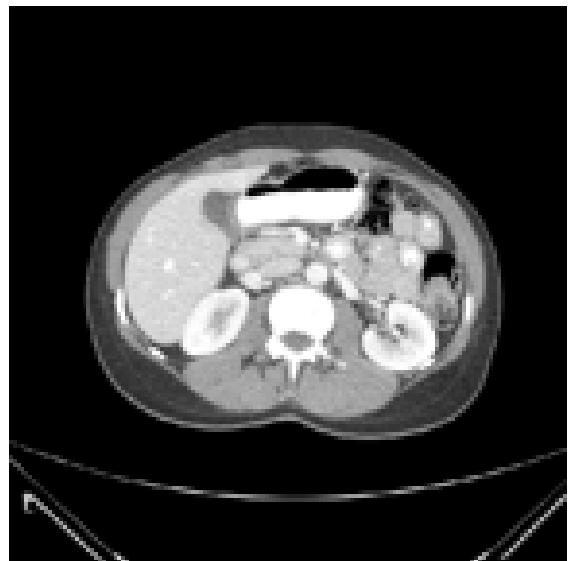
CT

Hounsfield Unit



https://www.researchgate.net/figure/Hounsfield-Units-scale-for-various-tissues-and-their-windows-modified-figure-17_fig3_313270846

CT



CT



CT windows



A. Lung window Level -550, width 1600



B. Soft tissue (mediastinal) window
Level 70, width 450; contrast in arterial phase



C. Bone window Level 570, width 3077



D. Bone window Level 455, width 958



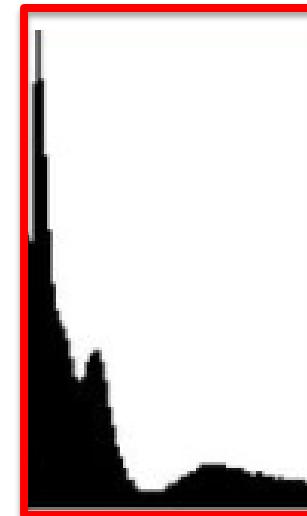
© ELSEVIER, INC. – NETTERIMAGES.COM

https://www.researchgate.net/figure/Hounsfield-Units-scale-for-various-tissues-and-their-windows-modified-figure-17_fig3_313270846

CT windows

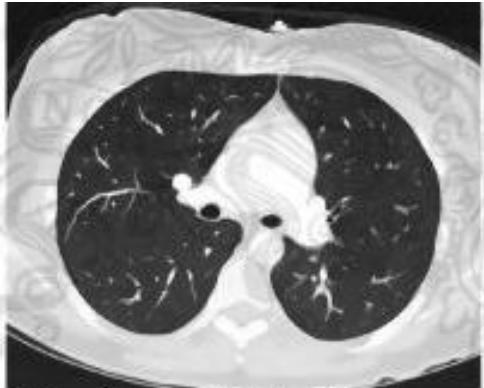


Lung Window



0 255

CT windows



A. Lung window Level -550, width 1600



B. Soft tissue (mediastinal) window
Level 70, width 450, contrast in arterial phase



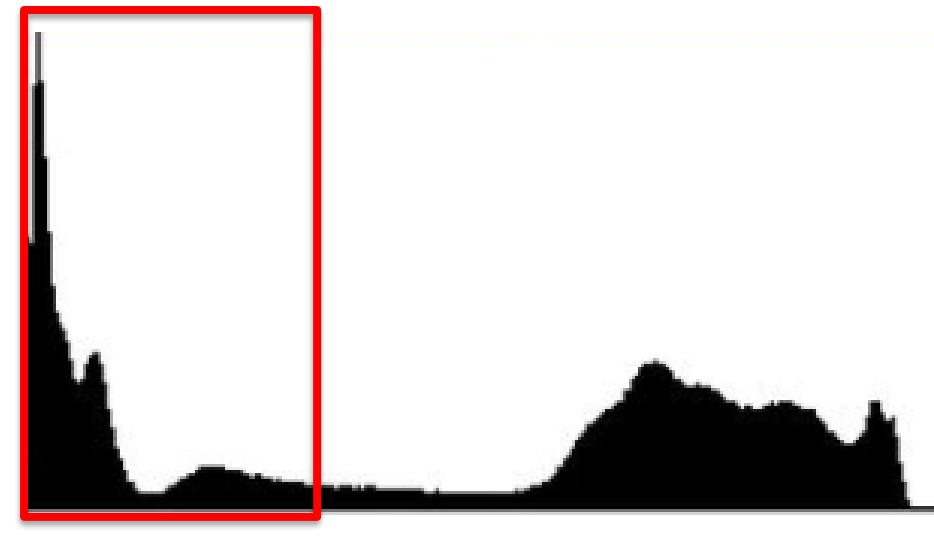
C. Bone window Level 570, width 3077



D. Bone window Level 455, width 958

© ELSEVIER, INC. – NETTERIMAGES.COM

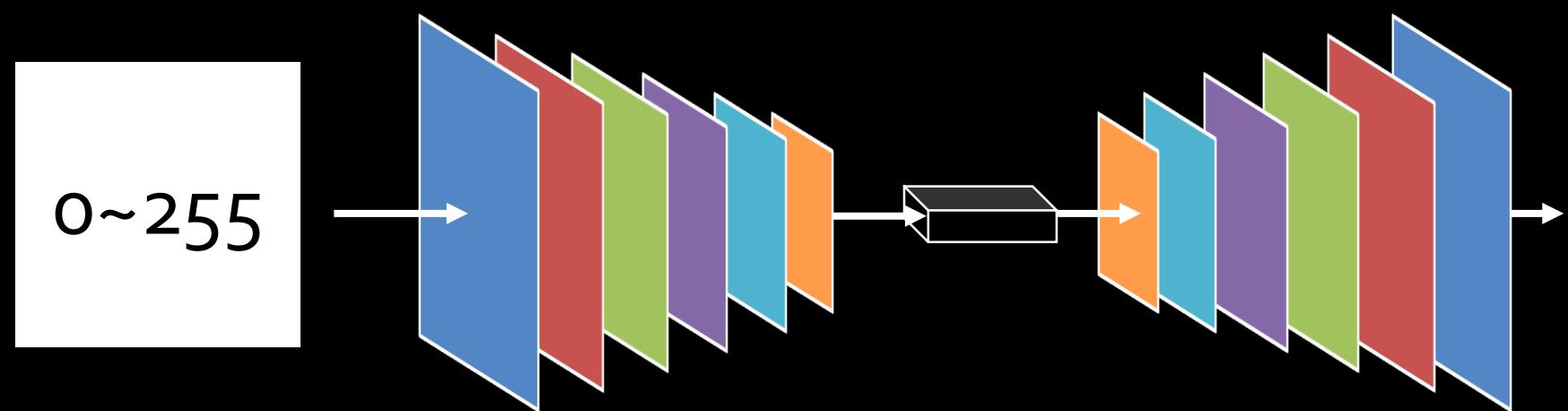
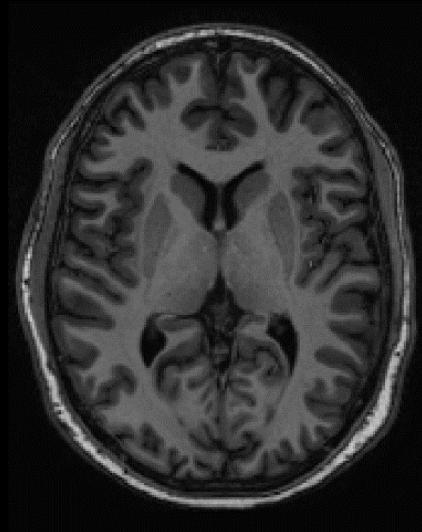
Lung Window



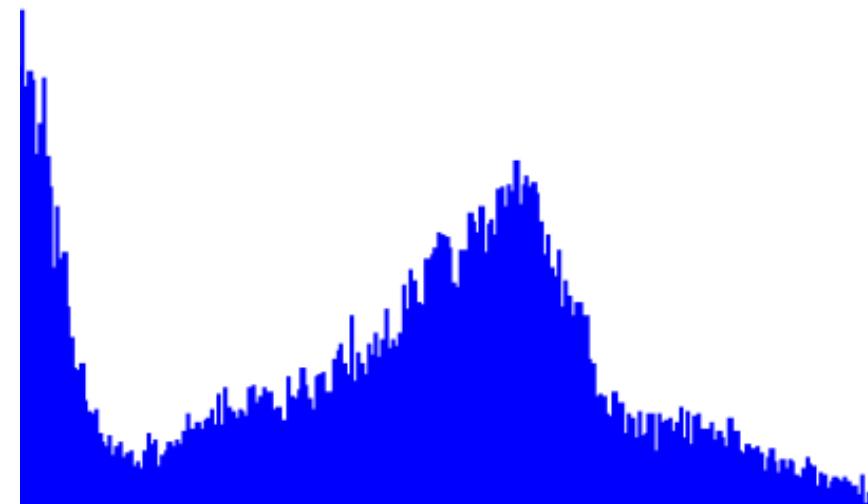
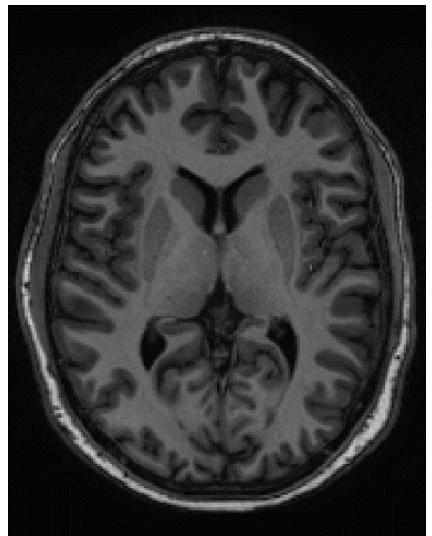
-1000

1000

MRI



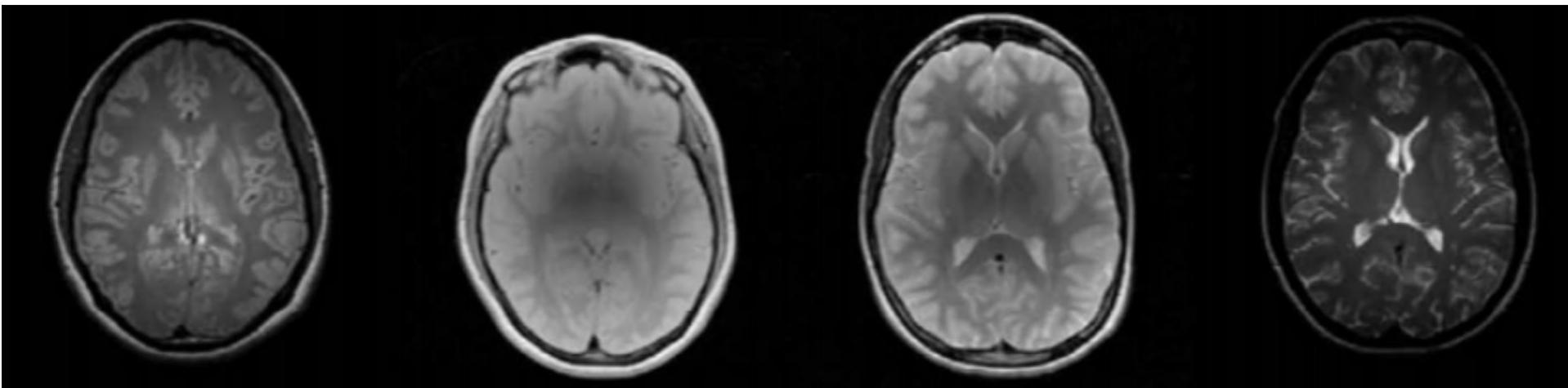
MRI



29398

54033

MRI

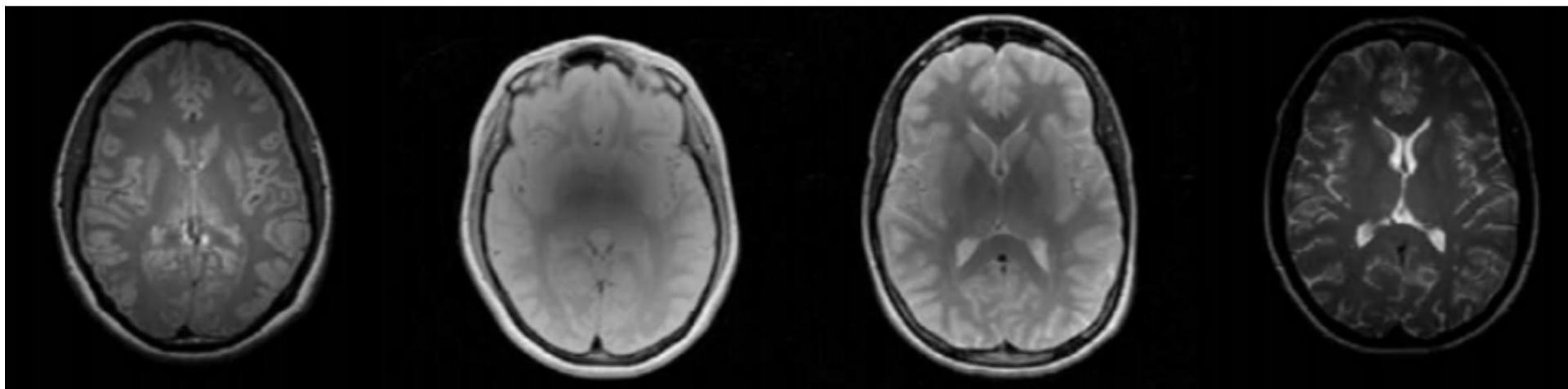


<http://www.cs.ucf.edu/~bagci/teaching/mic17/lec5.pdf>

MRI

MR Intensity Non-Standardness

- Acquisition-to-acquisition signal intensity variations (non-standardness) are inherent in MR images.



PD

PD

T2

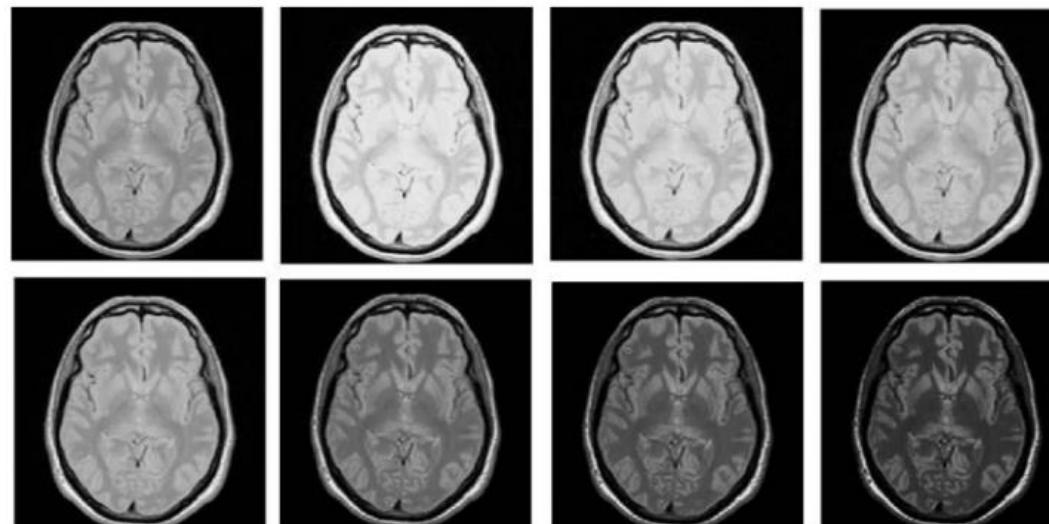
T2

Intensity Non-Standardness

MR image intensities do not possess a tissue-specific numeric meaning even in images acquired for

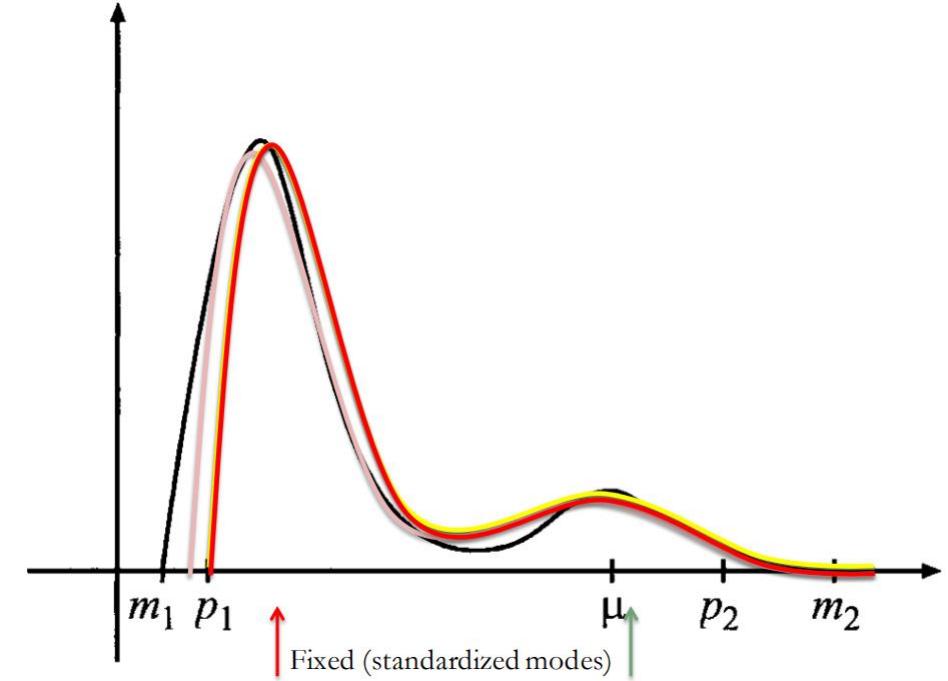
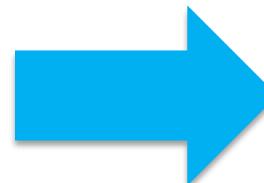
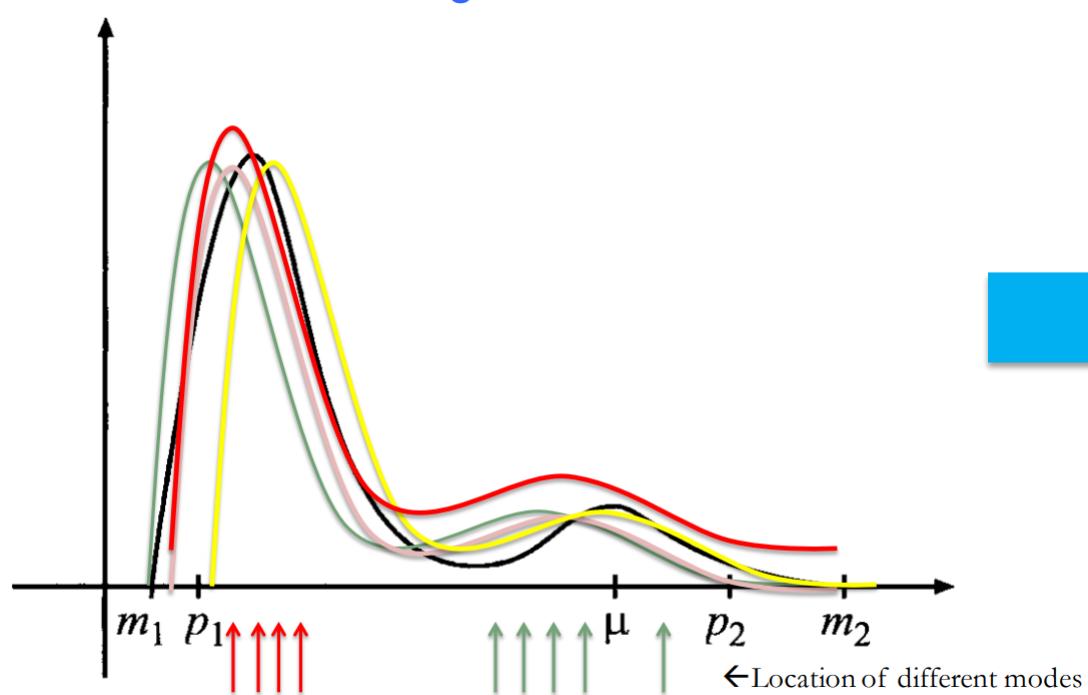
- the same subject,
- on the same scanner,
- for the same body region,

by using the same pulse sequence



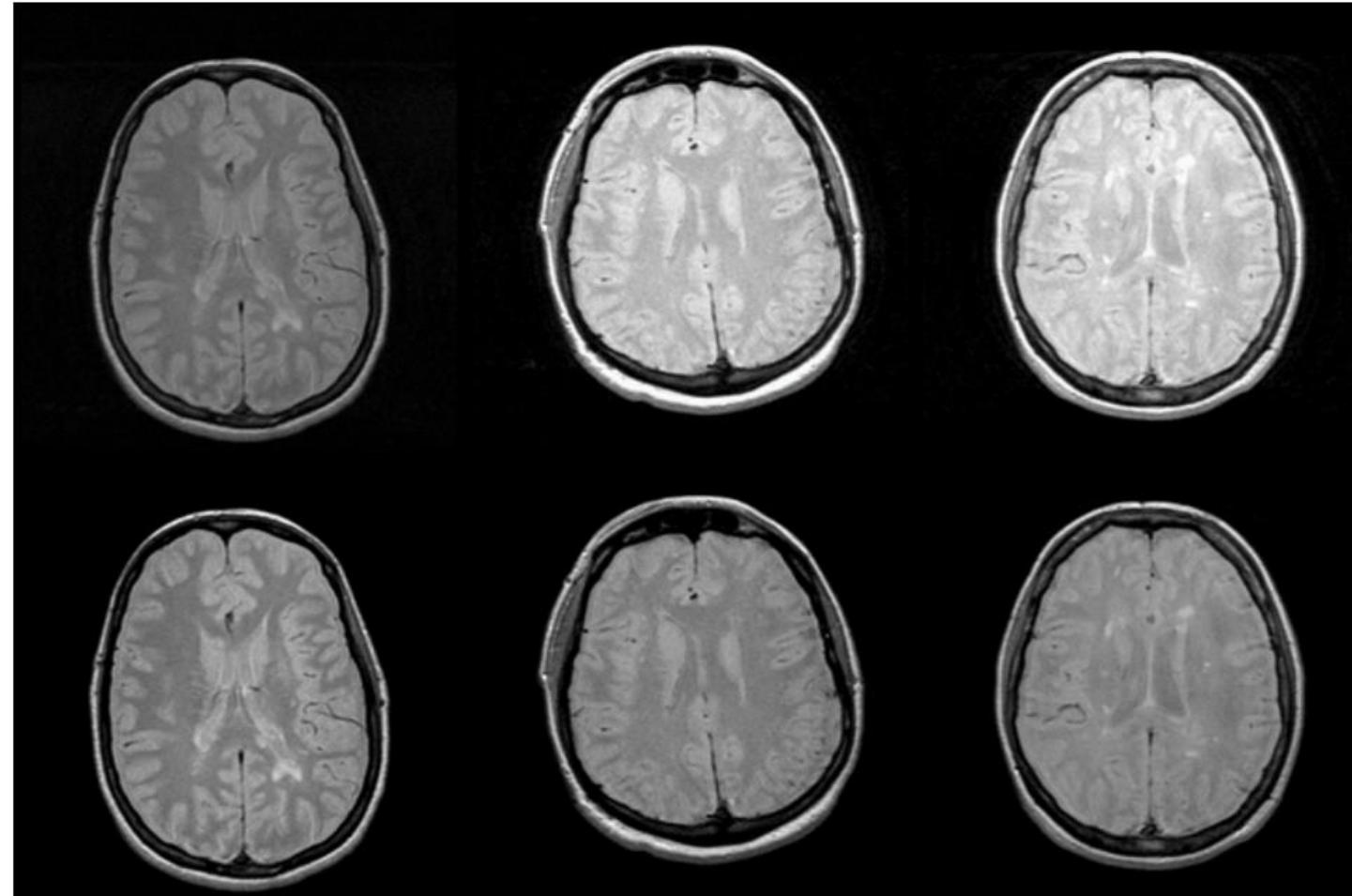
same brain slice,
same person,
same scanners,
different imaging times,
intensities are significantly
different for the same
Tissue type!

Histogram Matching

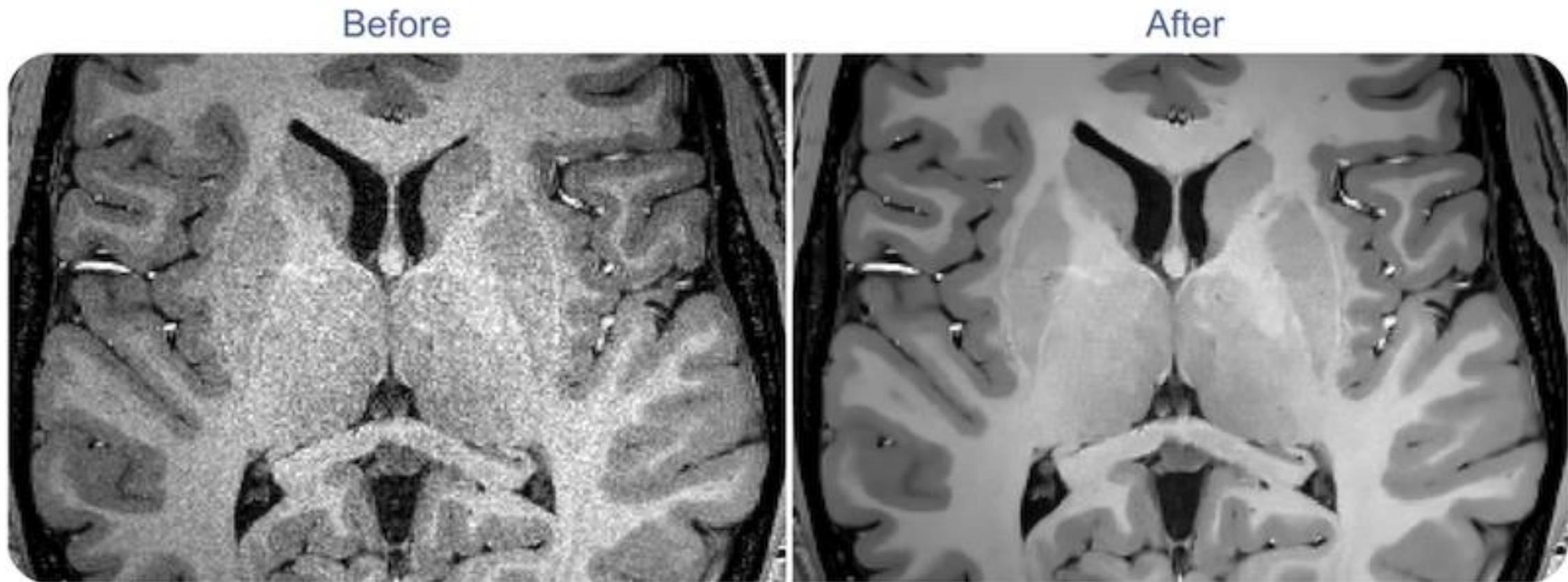


Intensity Standardization

Original
Gray Scale

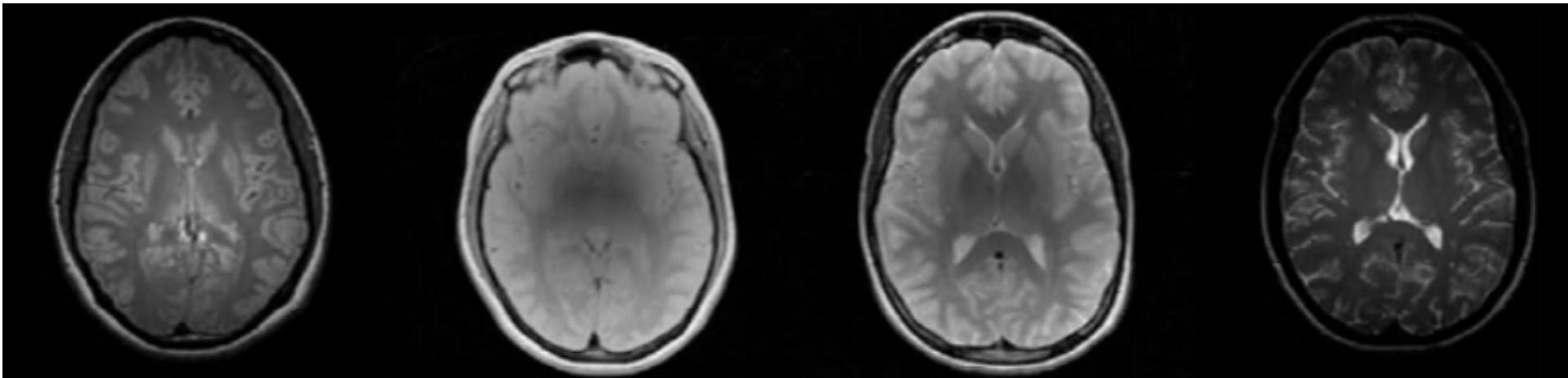


Denoise

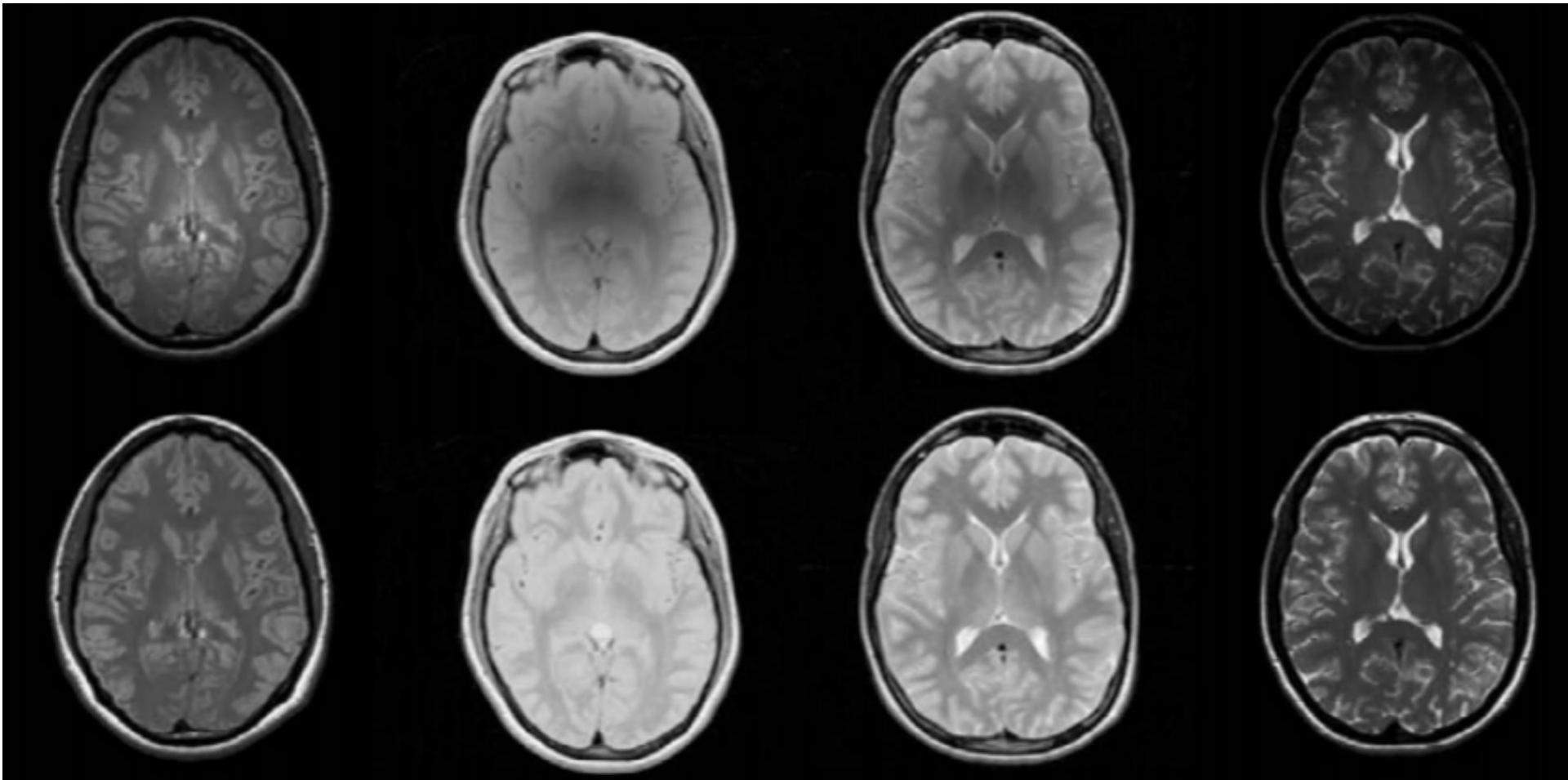


https://www.nitrc.org/project/list_screenshots.php?group_id=806&screenshot_id=738

Intensity Inhomogeneity

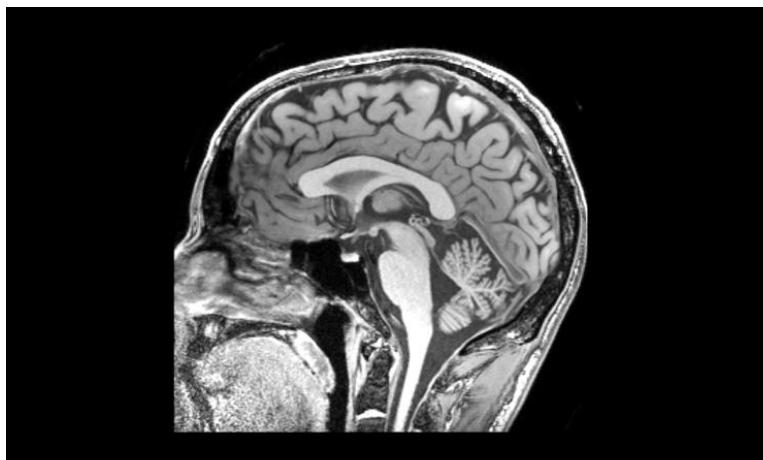
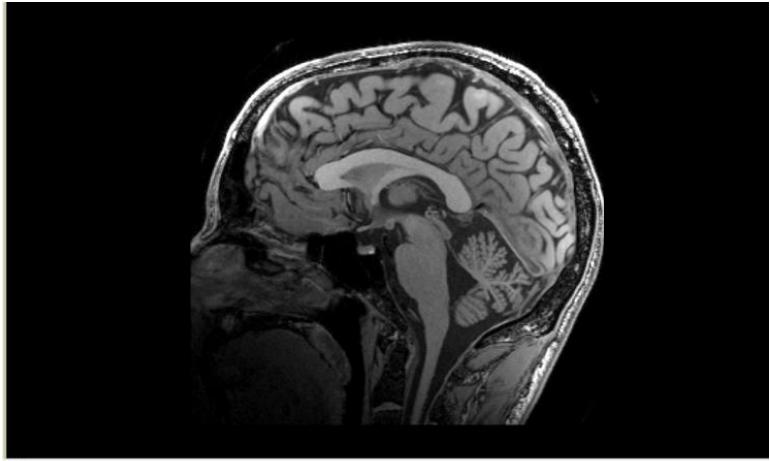


Inhomogeneity Correction



<http://www.cs.ucf.edu/~bagci/teaching/mic17/lec5.pdf>

N4 Correction



N4 Inhomogeneity Correction

We will use N4: Improved N3 Bias Correction (Tustison et al. 2010).

The model assumed in the N4 is: $v(x) = u(x)f(x) + n(x)$

- v is the given image
- u is the uncorrupted image
- f is the bias field
- n is the noise (assumed to be independent and Gaussian)
- x is a location in the image

The data is log-transformed and assuming a noise-free scenario, we have:

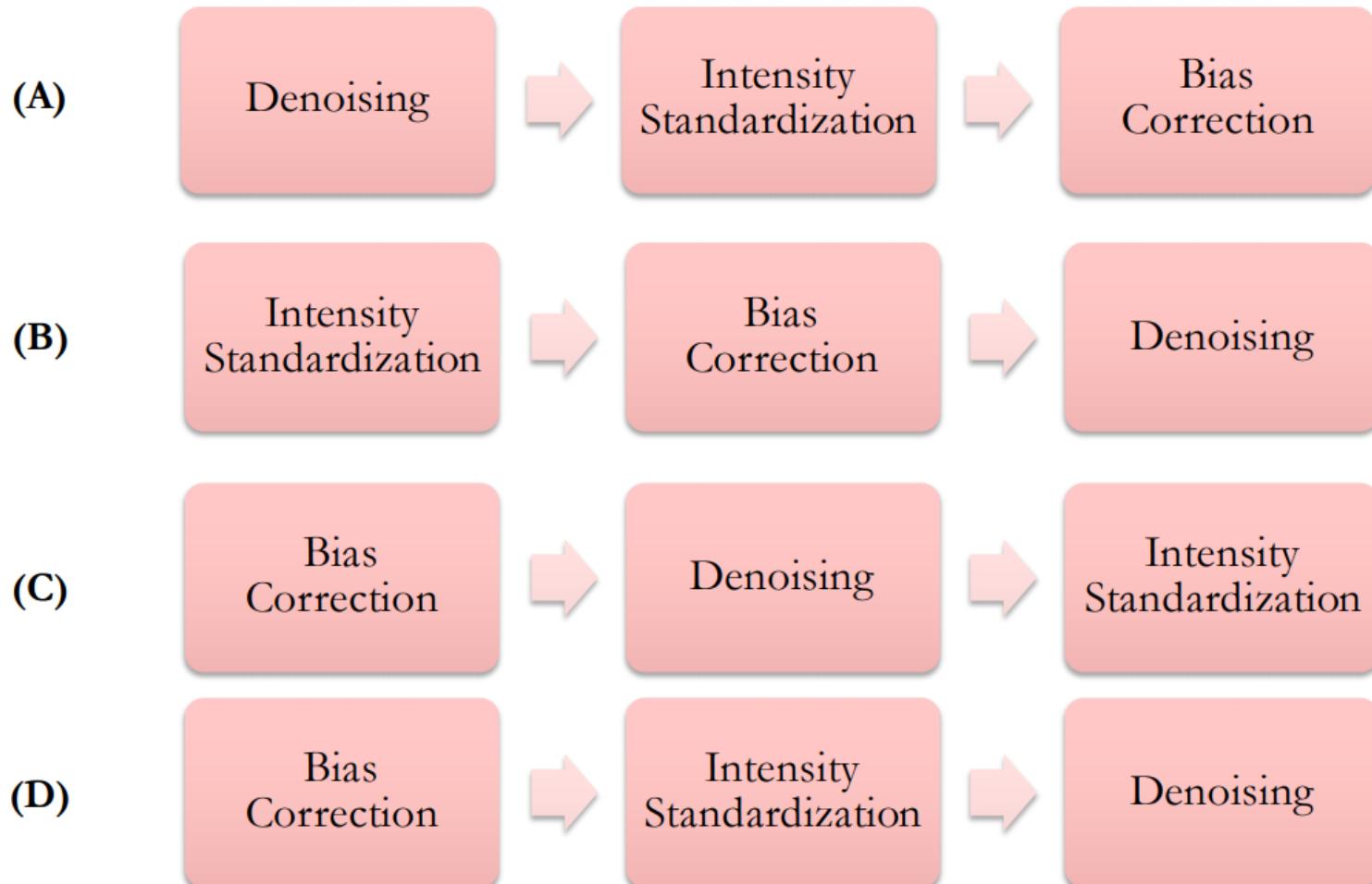
$$\log(v(x)) = \log(u(x)) + \log(f(x))$$

- N4 uses a B-spline approximation of the bias field
- It iterates until a convergence criteria is met
 - when the updated bias field is the same as the last iteration
- It outputs the data back in the original units (not log-transformed)

Deep Learning

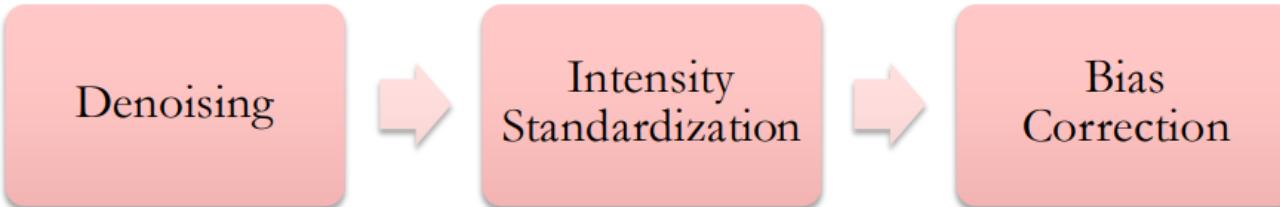
?

Bias Correction, Denoising, Intensity Standardization,

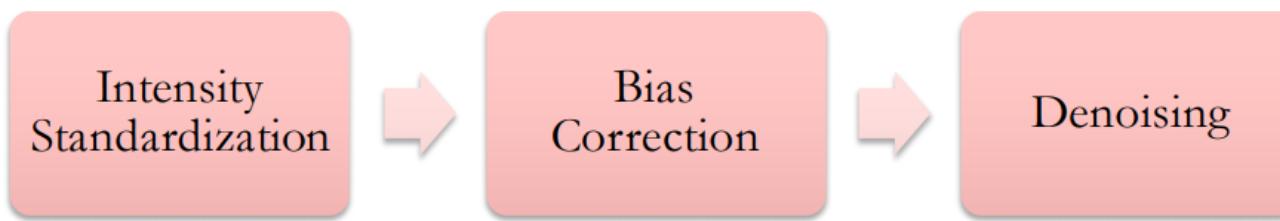


Bias Correction, Denoising, Intensity Standardization,

(A)



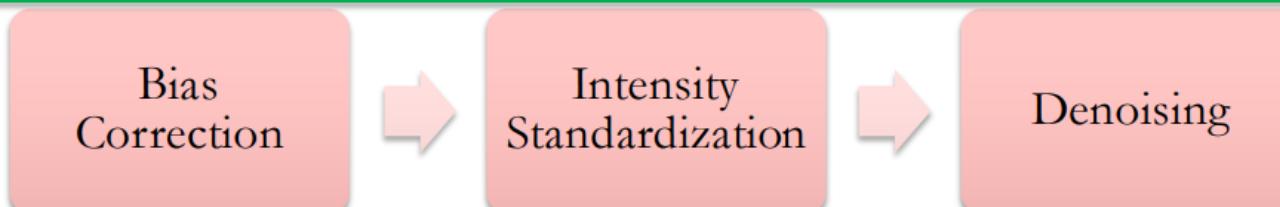
(B)



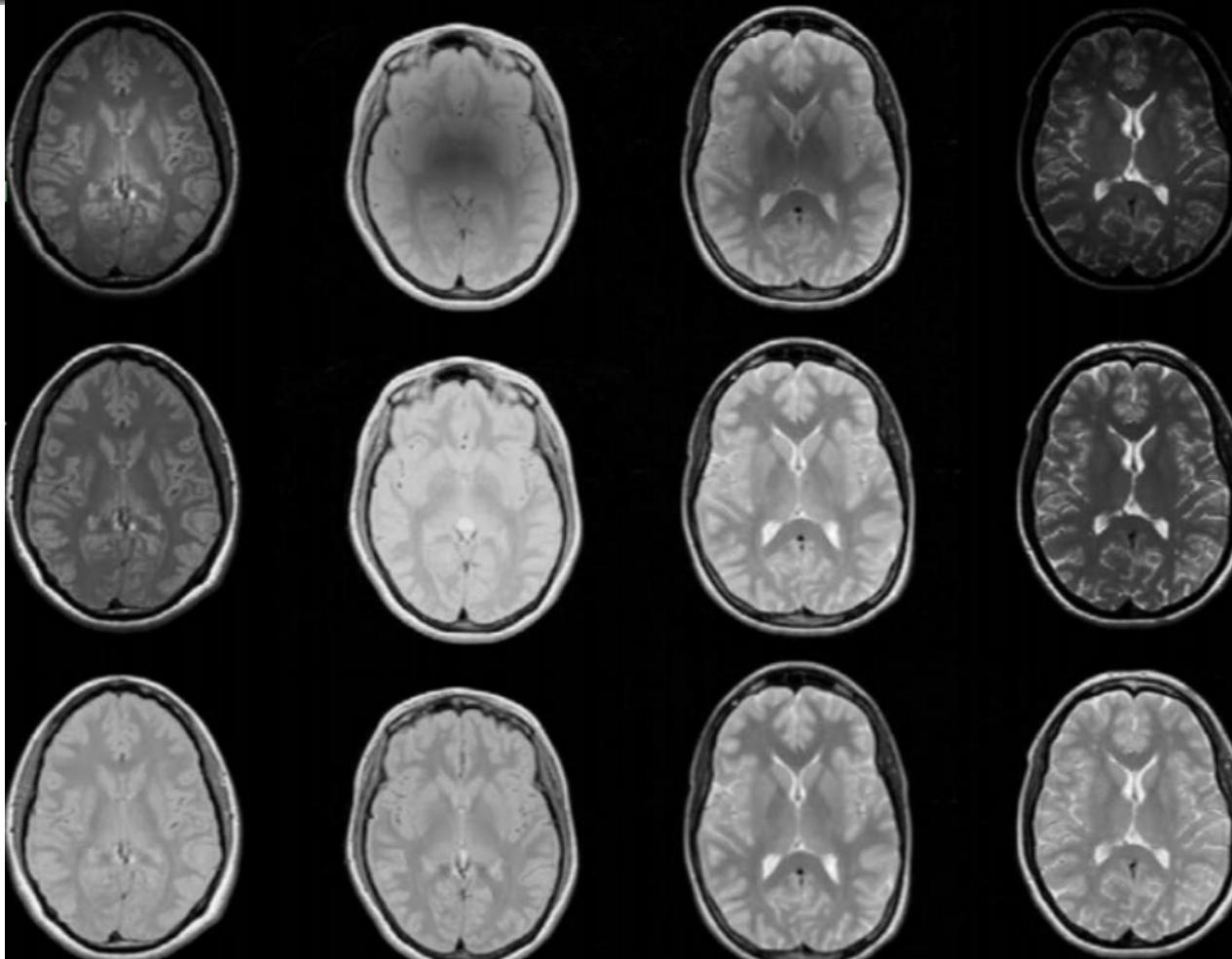
(C)



(D)



After Preprocessing



<http://www.cs.ucf.edu/~bagci/teaching/mic17/lec5.pdf>

Super-resolution



<http://vllab.ucmerced.edu/wlai24/LapSRN/>

Deep Learning

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham,
Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi
Twitter

{cledig, ltheis, fhuszar, jcaballero, aacostadiaz, aaitken, atejani, jtotsz, zehanw, wshi}@twitter.com

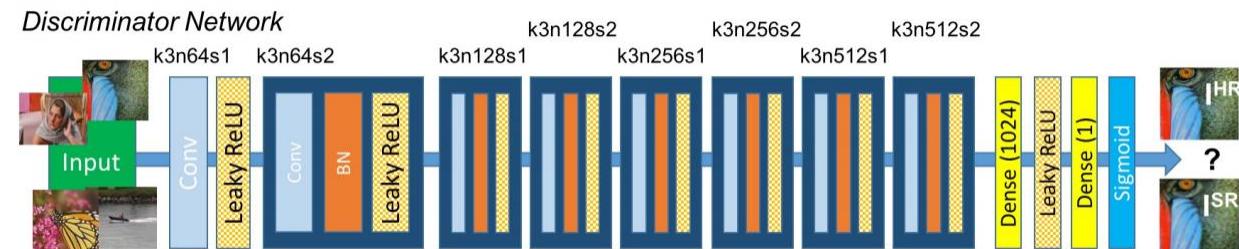
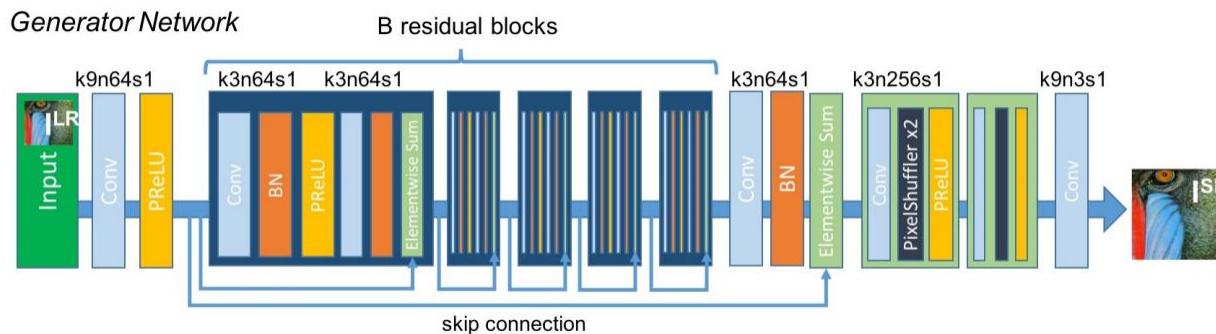
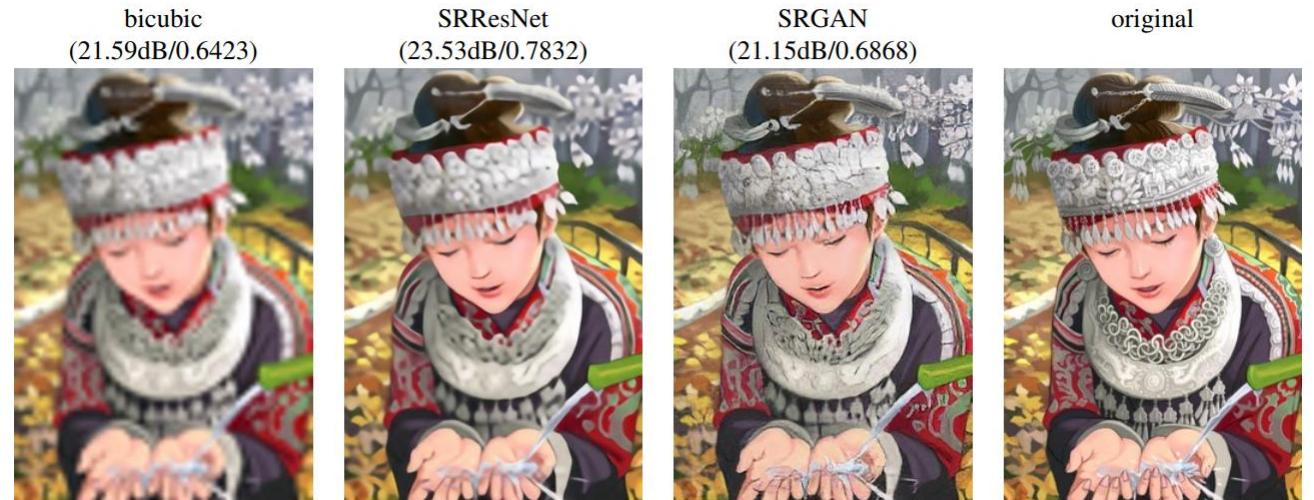
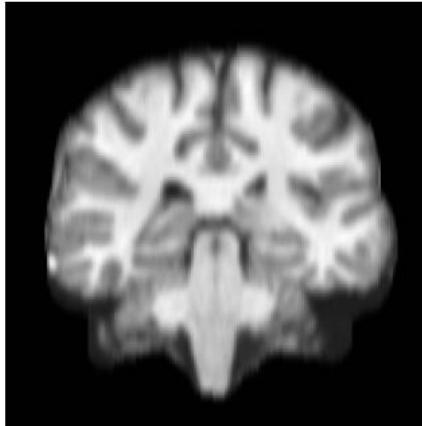
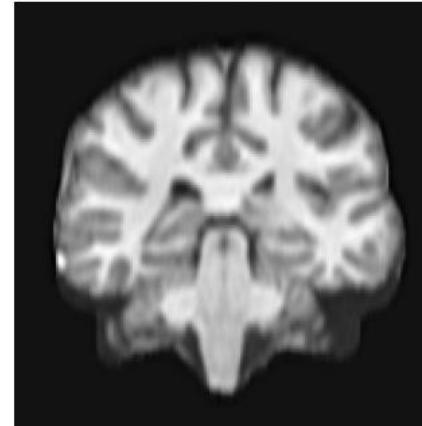


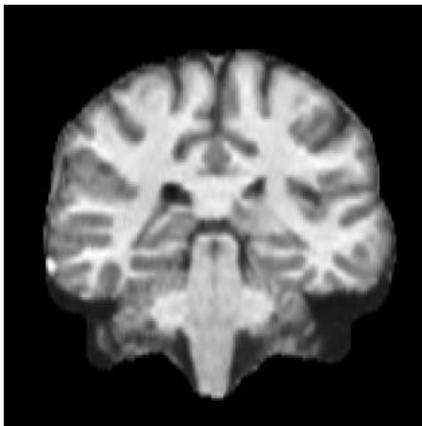
Image Enhancement



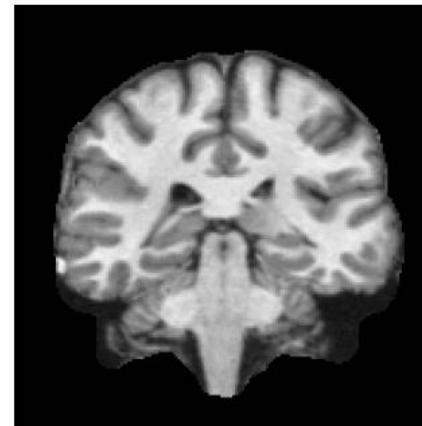
(a) BSP ($k = 3$)



(b) SSR ($k = 3$) [11]



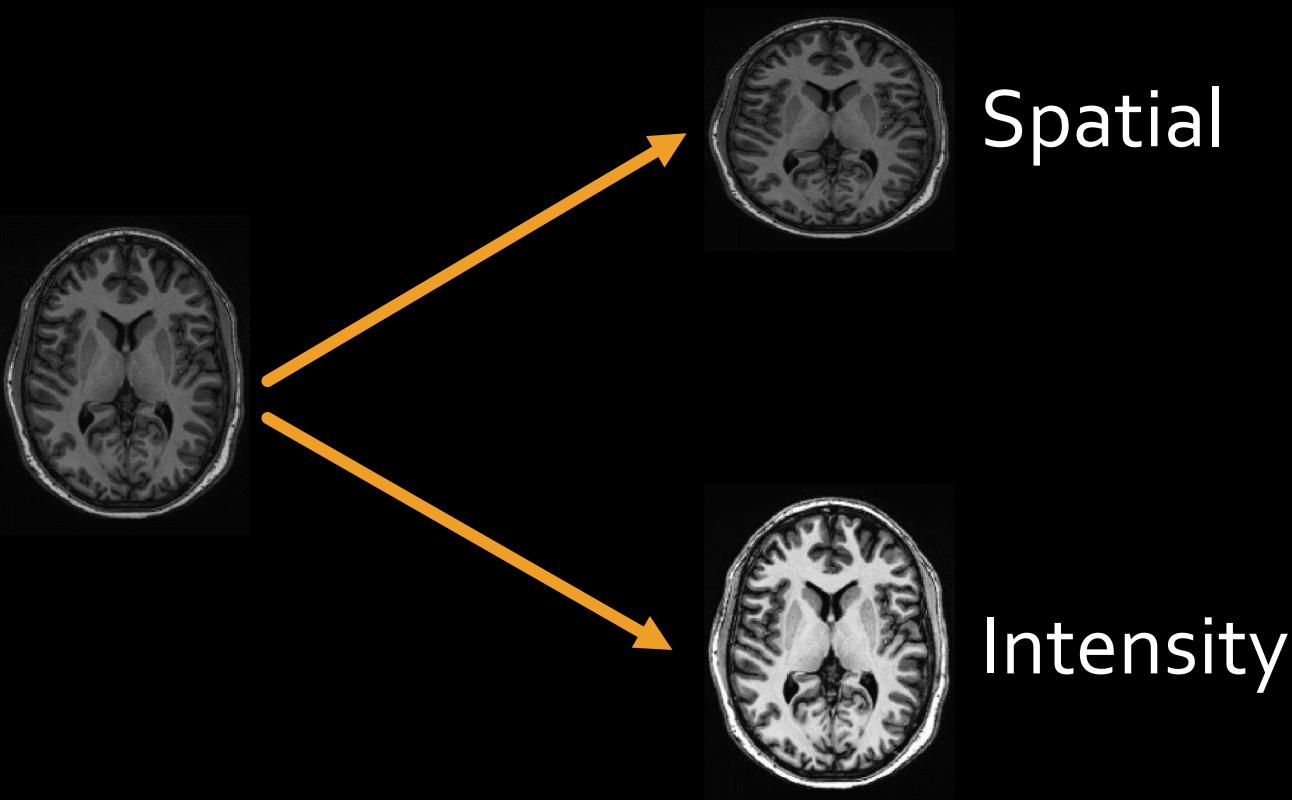
(c) EDSSR ($k = 3$)



(d) HR (1mm)

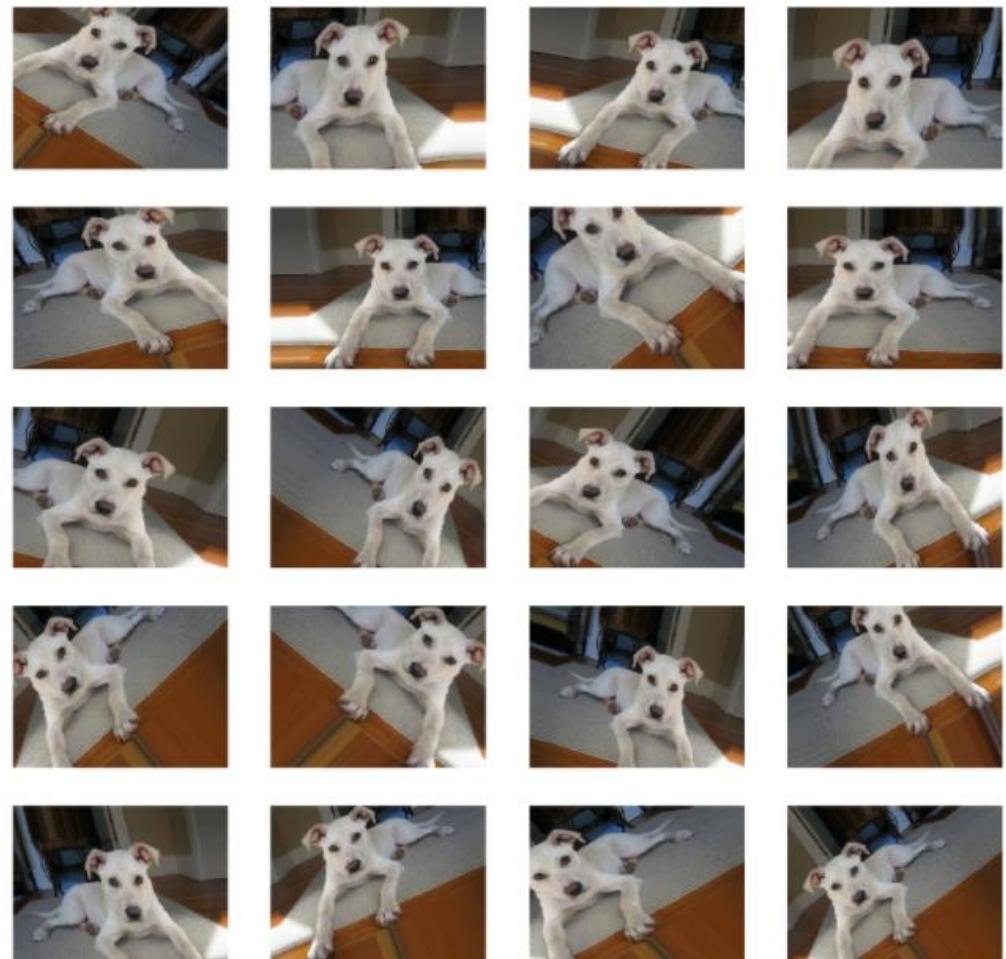
<https://arxiv.org/pdf/1802.09431.pdf>

Preprocessing



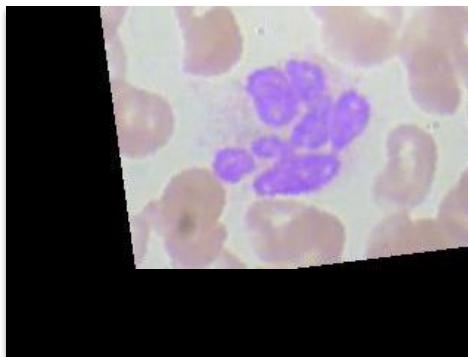
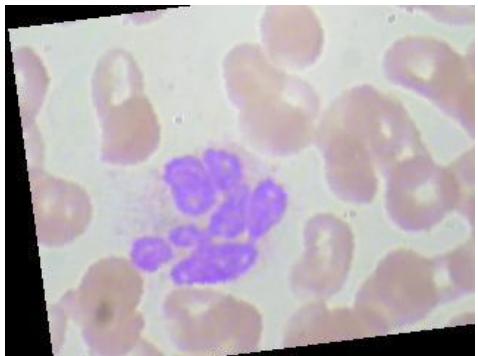
- Reshape
 - Cropping
 - Registration
-
- Ultrasound
 - CT (Hounsfield Unit window)
 - MRI (Histogram, Denoise, Harmonization, Standardization, Enhancement)

Data Augmentation

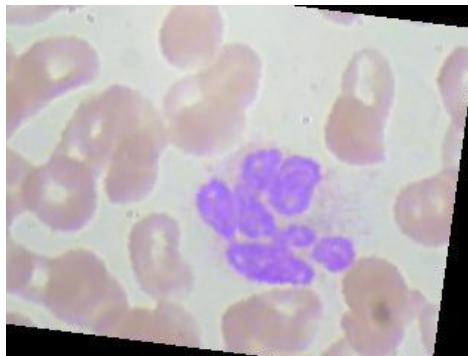


<https://chtseng.wordpress.com/2017/11/11/data-augmentation-%E8%B3%87%E6%96%99%E5%A2%9E%E5%BC%B7/>

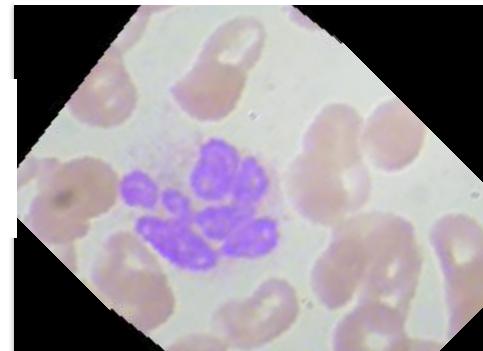
Data Augmentation



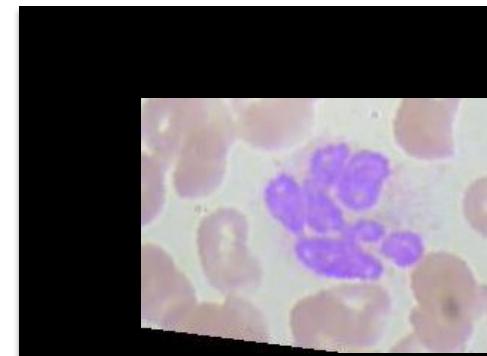
Translation



Flip



Rotation



Combination

Augmentation

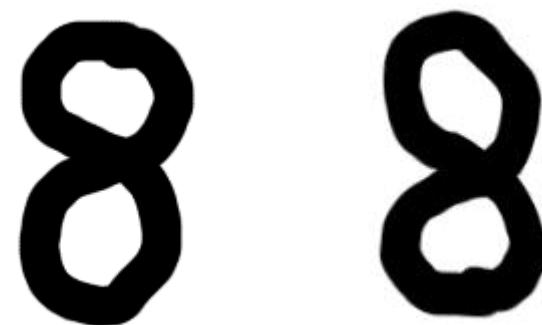
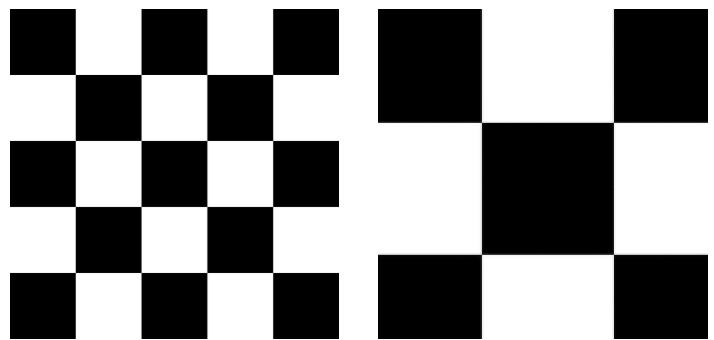
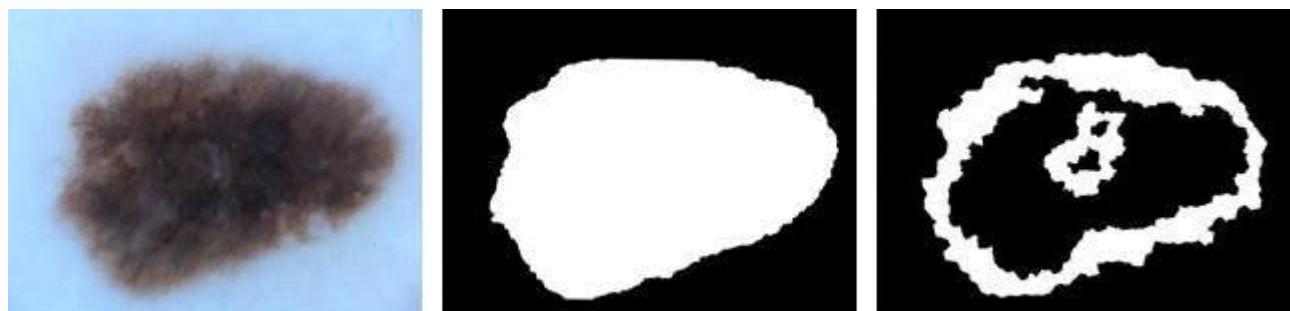
<https://github.com/mdbloice/Augmentor>



Augmentor is an image augmentation library in Python for machine learning. It aims to be a standalone library that is platform and framework independent, which is more convenient, allows for finer grained control over augmentation, and implements the most real-world relevant augmentation techniques. It employs a stochastic approach using building blocks that allow for operations to be pieced together in a pipeline.



Augmentation



<https://github.com/mdbloice/Augmentor>

State of the art Cut-out

Improved Regularization of Convolutional Neural Networks with Cutout

Terrance DeVries¹ and Graham W. Taylor^{1,2}

¹University of Guelph

²Canadian Institute for Advanced Research and Vector Institute



State of the art mixup

mixup: BEYOND EMPIRICAL RISK MINIMIZATION

Hongyi Zhang
MIT

Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz*
FAIR

ABSTRACT

Large deep neural networks are powerful, but exhibit undesirable behaviors such as memorization and sensitivity to adversarial examples. In this work, we propose *mixup*, a simple learning principle to alleviate these issues. In essence, *mixup* trains a neural network on convex combinations of pairs of examples and their labels. By doing so, *mixup* regularizes the neural network to favor simple linear behavior in-between training examples. Our experiments on the ImageNet-2012, CIFAR-10, CIFAR-100, Google commands and UCI datasets show that *mixup* improves the generalization of state-of-the-art neural network architectures. We also find that *mixup* reduces the memorization of corrupt labels, increases the robustness to adversarial examples, and stabilizes the training of generative adversarial networks.

$$\begin{aligned}\hat{x} &= \lambda x_i + (1 - \lambda)x_j, \\ \hat{y} &= \lambda y_i + (1 - \lambda)y_j,\end{aligned}$$

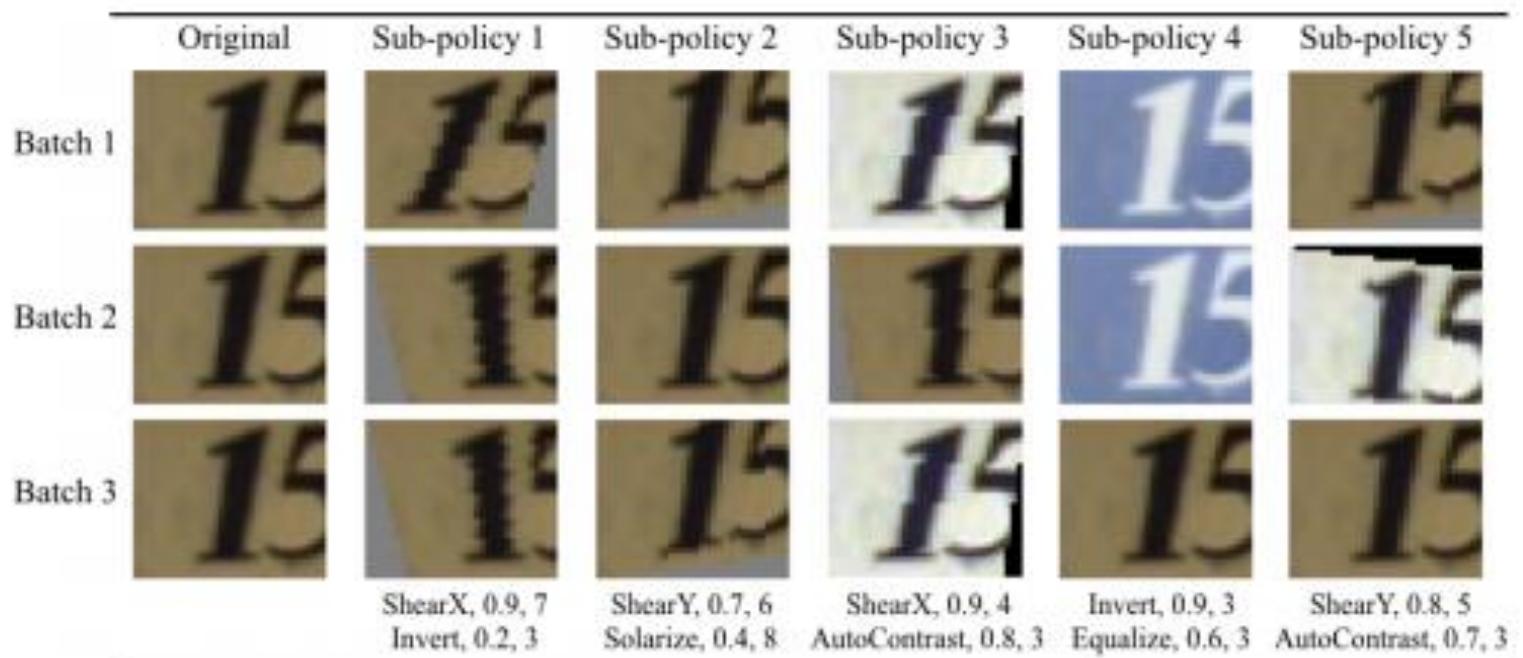
where $\lambda \in [0, 1]$ is a random number



State of the art AutoAugment

AutoAugment: Learning Augmentation Strategies from Data

Ekin D. Cubuk*,†, Barret Zoph†, Dandelion Mané, Vijay Vasudevan, Quoc V. Le
Google Brain



<https://hoya012.github.io/blog/Bag-of-Tricks-for-Image-Classification-with-Convolutional-Neural-Networks-Review/>

Webly Supervised Learning for Skin Lesion Classification

Fernando Navarro^{1(✉)}, Sailesh Conjeti^{1,2}, Federico Tombari¹,
and Nassir Navab^{1,3}

¹ Computer Aided Medical Procedures,
Technische Universität München, Munich, Germany
fernando.navarro@tum.de

² German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
Computer Aided Medical Procedures, Johns Hopkins University, Baltimore

Abstract. Within medical imaging, manual curation of sufficient well-labeled samples is cost, time and scale-prohibitive. To improve the representativeness of the training dataset, for the first time, we present an approach to utilize large amounts of freely available web data through web-crawling. To handle noise and weak nature of web annotations, we propose a two-step transfer learning based training process with a robust loss function, termed as Webly Supervised Learning (WSL) to train deep models for the task. We also leverage *search by image* to improve the search specificity of our web-crawling and reduce cross-domain noise. Within WSL, we explicitly model the noise structure between classes and incorporate it to selectively distill knowledge from the web data during model training. To demonstrate improved performance due to WSL, we benchmarked on a publicly available 10-class fine-grained skin lesion classification dataset and report a significant improvement of top-1 classification accuracy from 71.25% to 80.53% due to the incorporation of web-supervision.

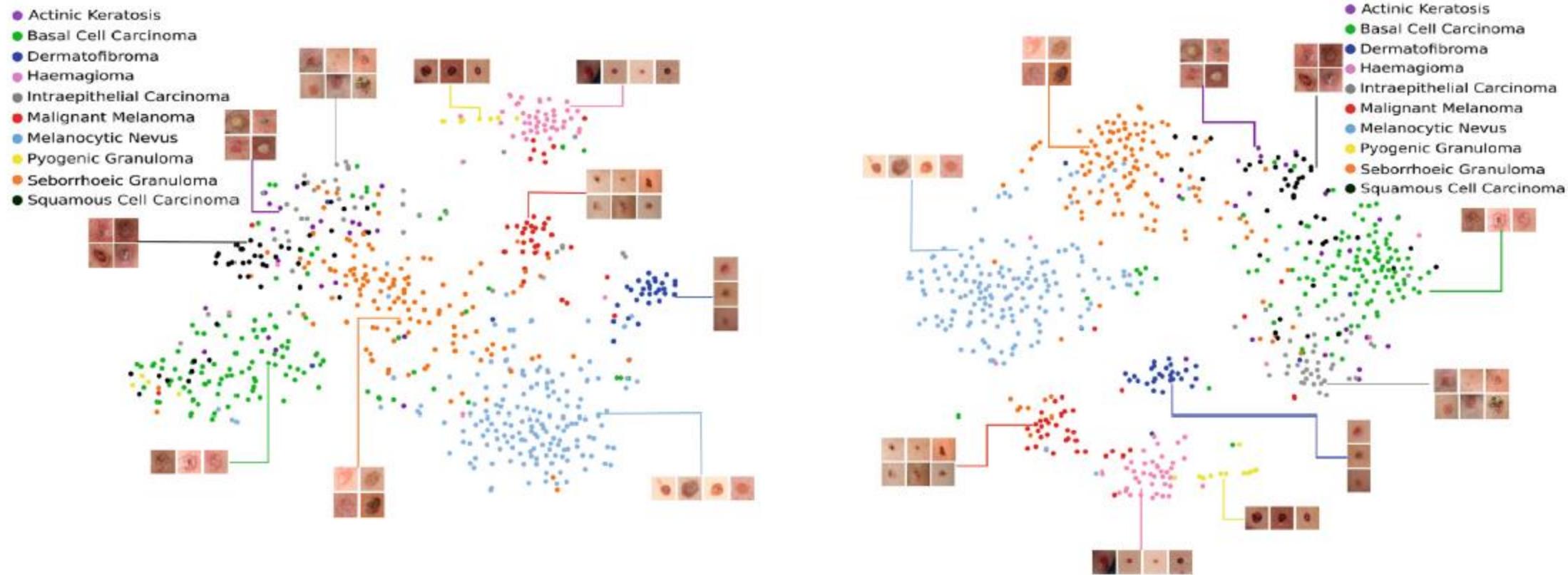


Fig. 1. Comparison of t-SNE embedding space generated from networks trained on limited clean data (Left) against network trained with Webly Supervised Learning (Right) generating compact class clusters with improved separability especially for under-represented classes.



Fig. 2. Type of noise in WSL for Melanoma class as keyword. The images in the first row represent examples of cross-domain noise. The second row represents the cross-category noise.

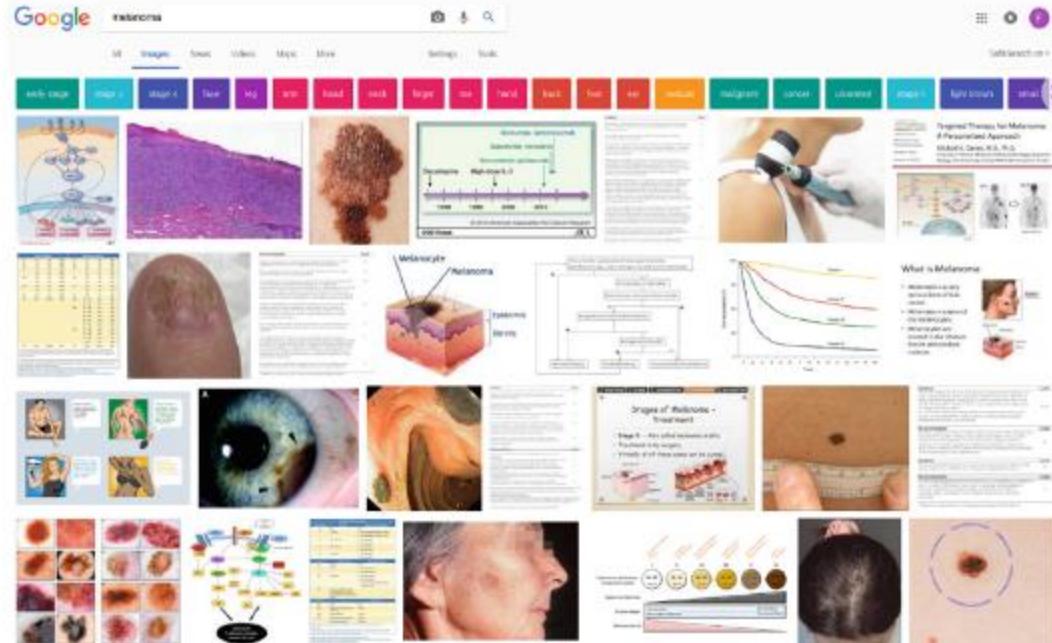


Fig. 3. Comparison of crawled results. The left image shows an example of a search by keyword “melanoma”: the resulting images contain high cross-domain noise. The right image shows the results of a search by image, where the cross-domain noise is significantly reduced sharing strong visual similarity to the query image.

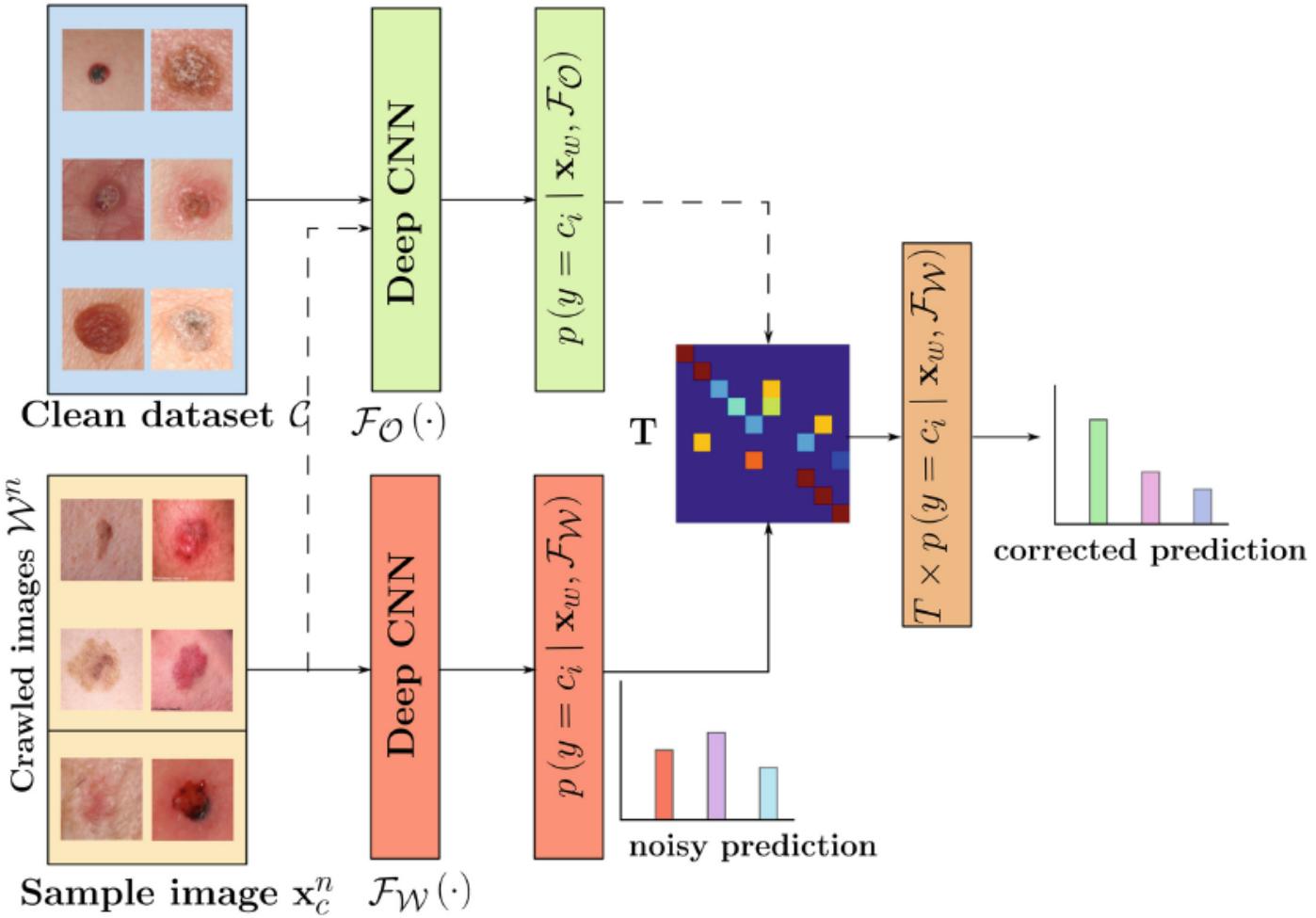


Fig. 4. Overview of the proposed WSL approach consisting of a two-step training. First, training a network on web data, follow by fine-tuning a second network utilizing the latter as strong prior initialization. Noise correction is performed when training on web data.

Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network

Yuhua Chen^{1,2}(✉) , Feng Shi² , Anthony G. Christodoulou² , Yibin Xie² , Zhengwei Zhou² , and Debiao Li^{1,2}

¹ University of California, Los Angeles, CA 90095, USA
chyuhua@ucla.edu

² Biomedical Imaging Research Institute, Cedars-Sinai Medical Center,
Los Angeles, CA 90048, USA

Abstract. High-resolution (HR) magnetic resonance images (MRI) provide detailed anatomical information important for clinical application and quantitative image analysis. However, HR MRI conventionally comes at the cost of longer scan time, smaller spatial coverage, and lower signal-to-noise ratio (SNR). Recent studies have shown that single image super-resolution (SISR), a technique to recover HR details from one single low-resolution (LR) input image, could provide high quality image details with the help of advanced deep convolutional neural networks (CNN). However, deep neural networks consume memory heavily and run slowly, especially in 3D settings. In this paper, we propose a novel 3D neural network design, namely a multi-level densely connected super-resolution network (mDCSRN) with generative adversarial network (GAN)-guided training. The mDCSRN trains and inferences quickly, and the GAN promotes realistic output hardly distinguishable from original HR images. Our results from experiments on a dataset with 1,113 subjects shows that our new architecture outperforms other popular deep learning methods in recovering 4x resolution-downgraded images and runs 6x faster.

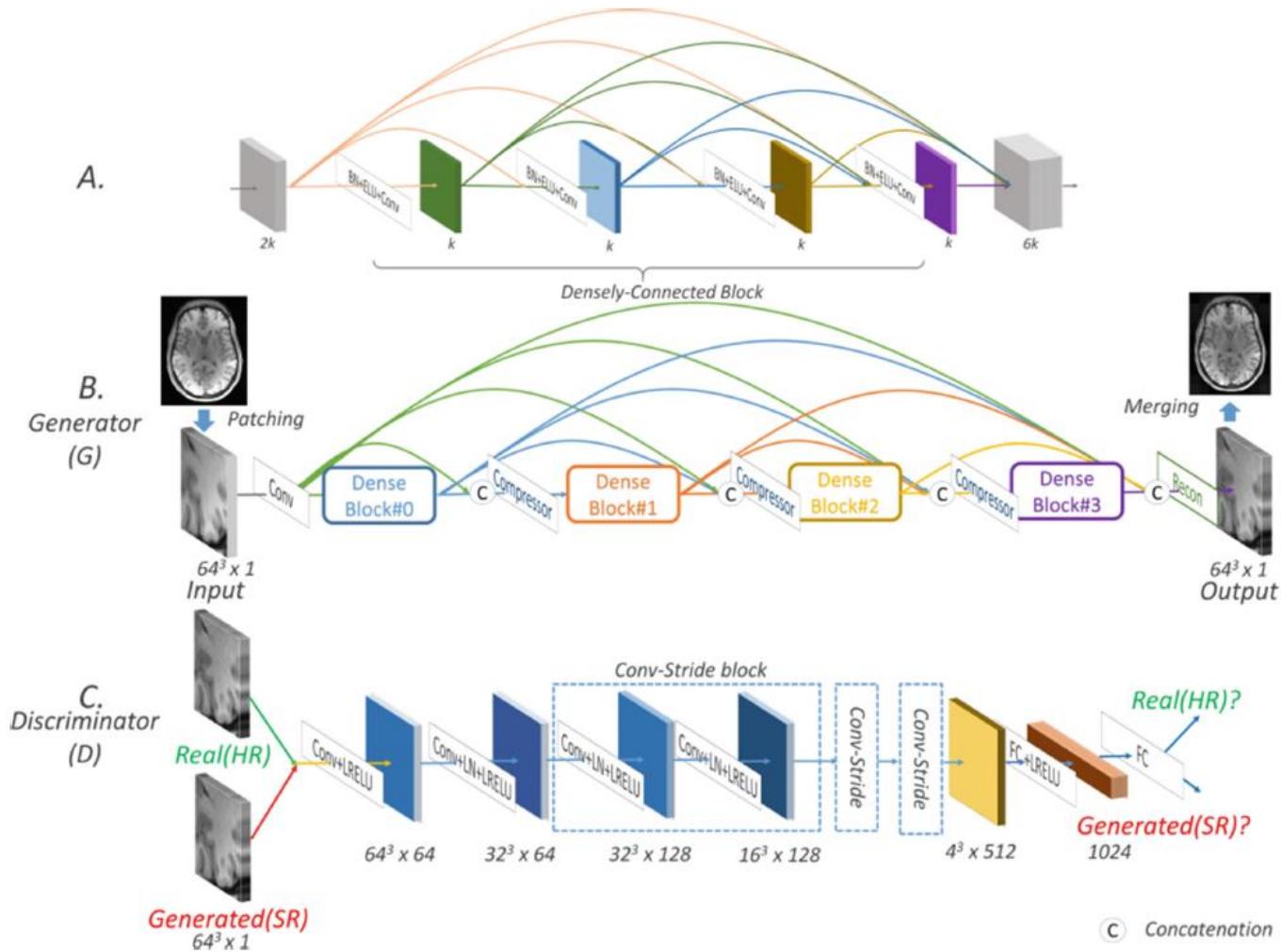


Fig. 1. Architecture of (A) DenseBlock with $3 \times 3 \times 3$ convolutions and (B, C) mDCSRN-GAN Network. The G is $b4u4$ (4 blocks, each has 4 unites) mDCSRN. The first convolutional layer outputs $2k$ ($k = 16$) feature maps, and each compressor shrinks down the feature maps to $2k$ via a $1 \times 1 \times 1$ convolution. The final reconstruction layer is another $1 \times 1 \times 1$ convolution. The D is identical to SRGAN except BatchNorm is replaced by LayerNorm, suggested by WGAN-GP.

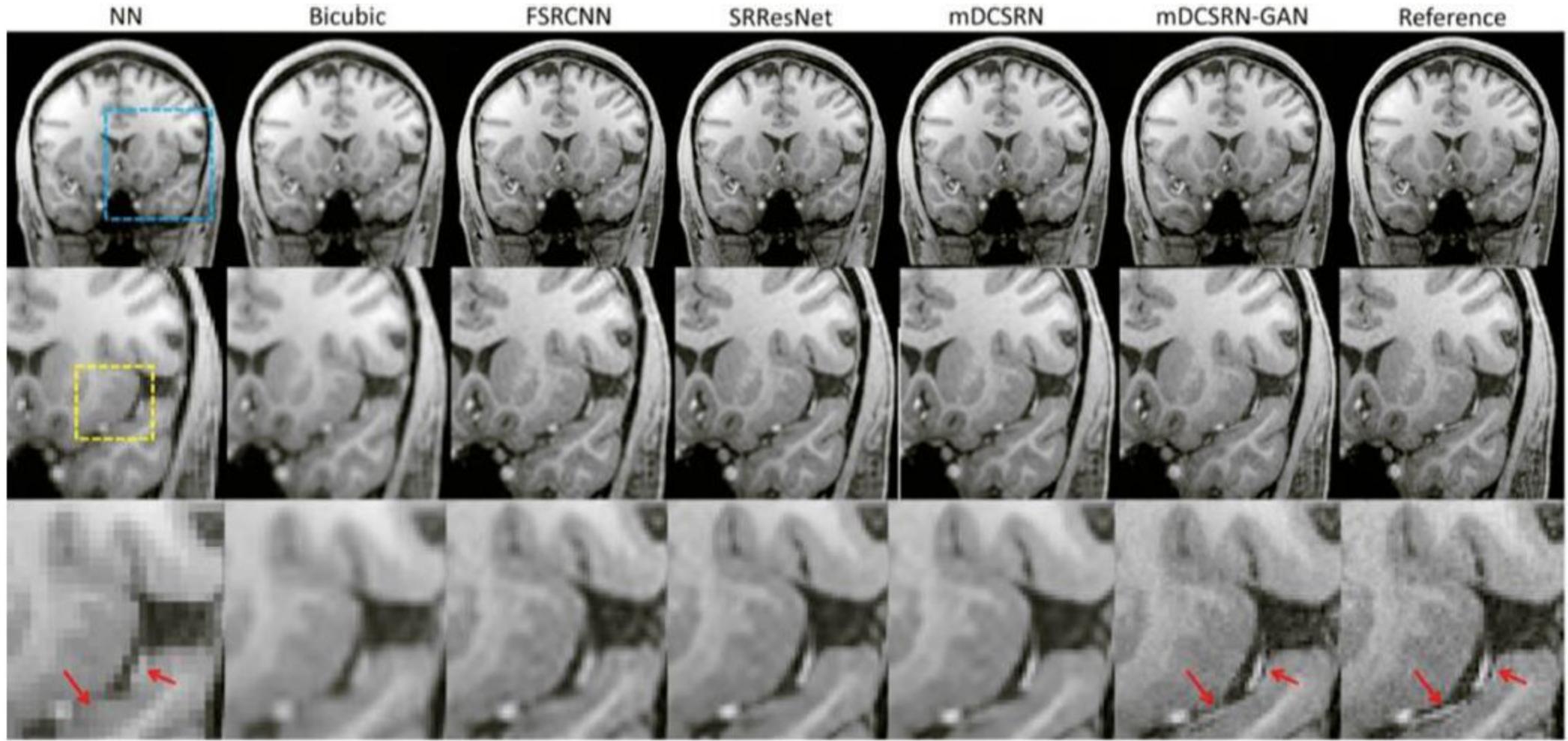


Fig. 2. Illustration of the nearest neighbor(NN) and bicubic interpolation, 3D FSRCNN, 3D SRResNet, mDCSRN, mDCSRN-GAN reconstruction results, and corresponding HR images.

Deep Learning Using K-Space Based Data Augmentation for Automated Cardiac MR Motion Artefact Detection

Ilkay Oksuz^{1(✉)}, Bram Ruijsink^{1,2}, Esther Puyol-Antón¹, Aurelien Bustin¹,
Gastao Cruz¹, Claudia Prieto¹, Daniel Rueckert³, Julia A. Schnabel¹,
and Andrew P. King¹

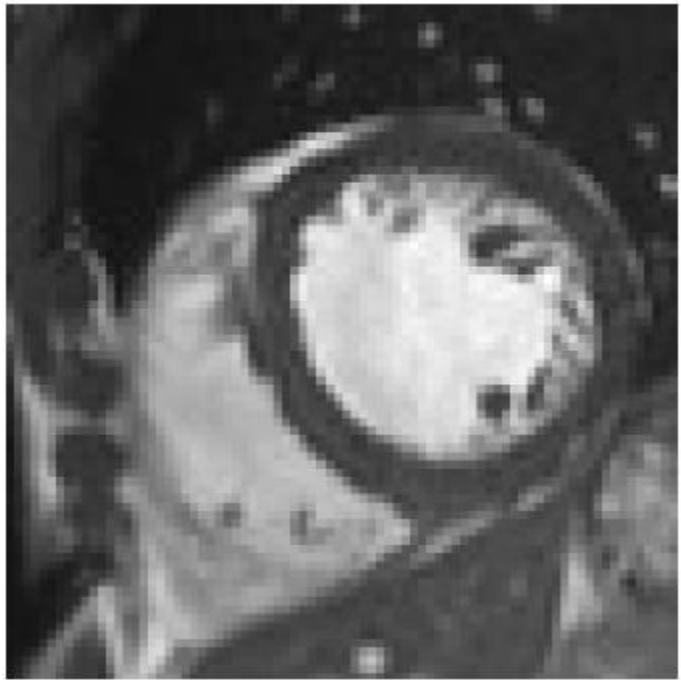
¹ School of Biomedical Engineering and Imaging Sciences,
King's College London, London, UK

ilkay.oksuz@kcl.ac.uk

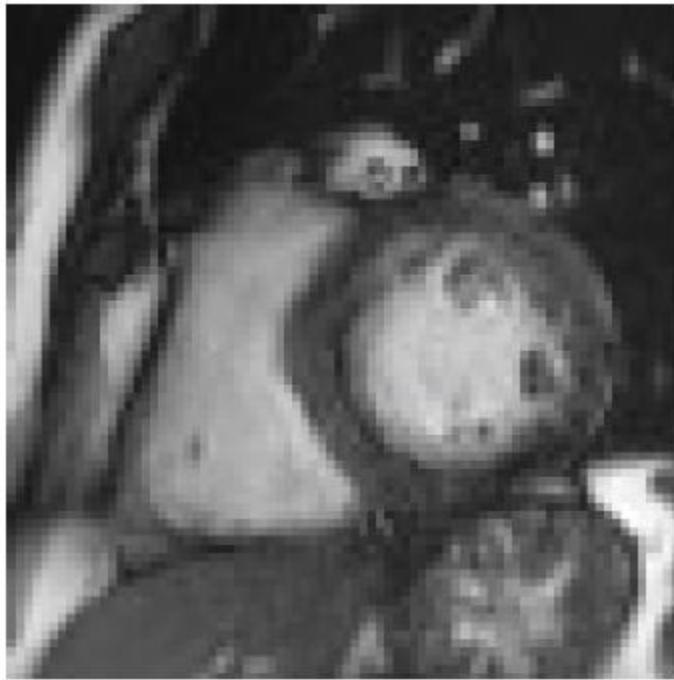
² Guy's and St Thomas' Hospital NHS Foundation Trust, London, UK

³ Biomedical Image Analysis Group, Imperial College London, London, UK

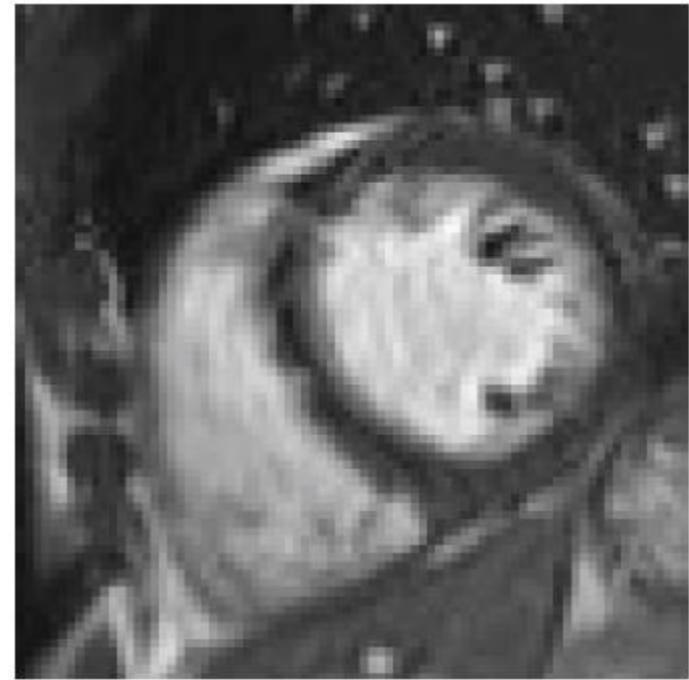
Abstract. Quality assessment of medical images is essential for complete automation of image processing pipelines. For large population studies such as the UK Biobank, artefacts such as those caused by heart motion are problematic and manual identification is tedious and time-consuming. Therefore, there is an urgent need for automatic image quality assessment techniques. In this paper, we propose a method to automatically detect the presence of motion-related artefacts in cardiac magnetic resonance (CMR) images. As this is a highly imbalanced classification problem (due to the high number of good quality images compared to the low number of images with motion artefacts), we propose a novel k-space based training data augmentation approach in order to address this problem. Our method is based on 3D spatio-temporal Convolutional Neural Networks, and is able to detect 2D+time short axis images with motion artefacts in less than 1 ms. We test our algorithm on a subset of the UK Biobank dataset consisting of 3465 CMR images and achieve not only high accuracy in detection of motion artefacts, but also high precision and recall. We compare our approach to a range of state-of-the-art quality assessment methods.



(a) Good quality image



(b) Motion artefact image



(c) K-space corrupted image

Fig. 1. Examples of a good quality CINE CMR image (a), an image with blurring motion artefacts (b), and a k-space corrupted image (c). The k-space corruption process is able to simulate realistic motion-related artefacts. (Please see videos in supplementary material.)

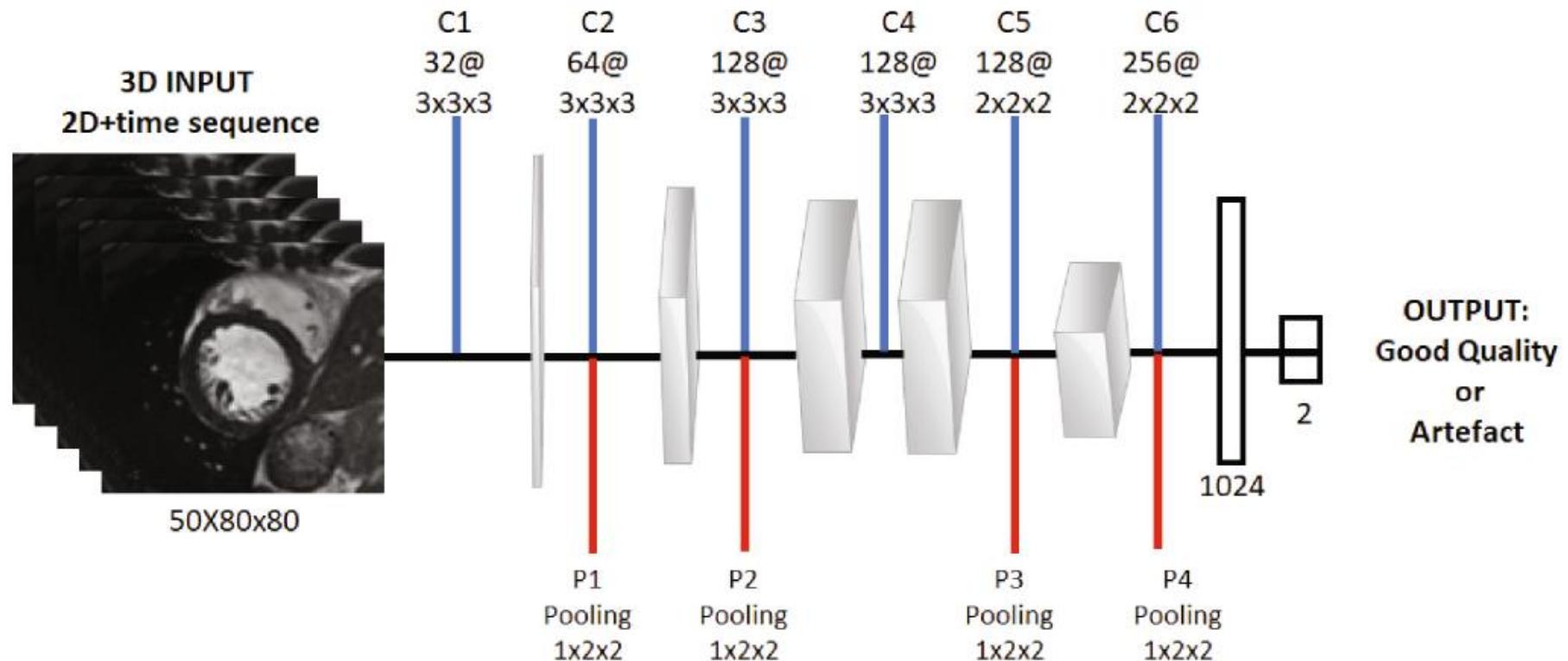


Fig. 2. The CNN architecture for motion artefact detection.

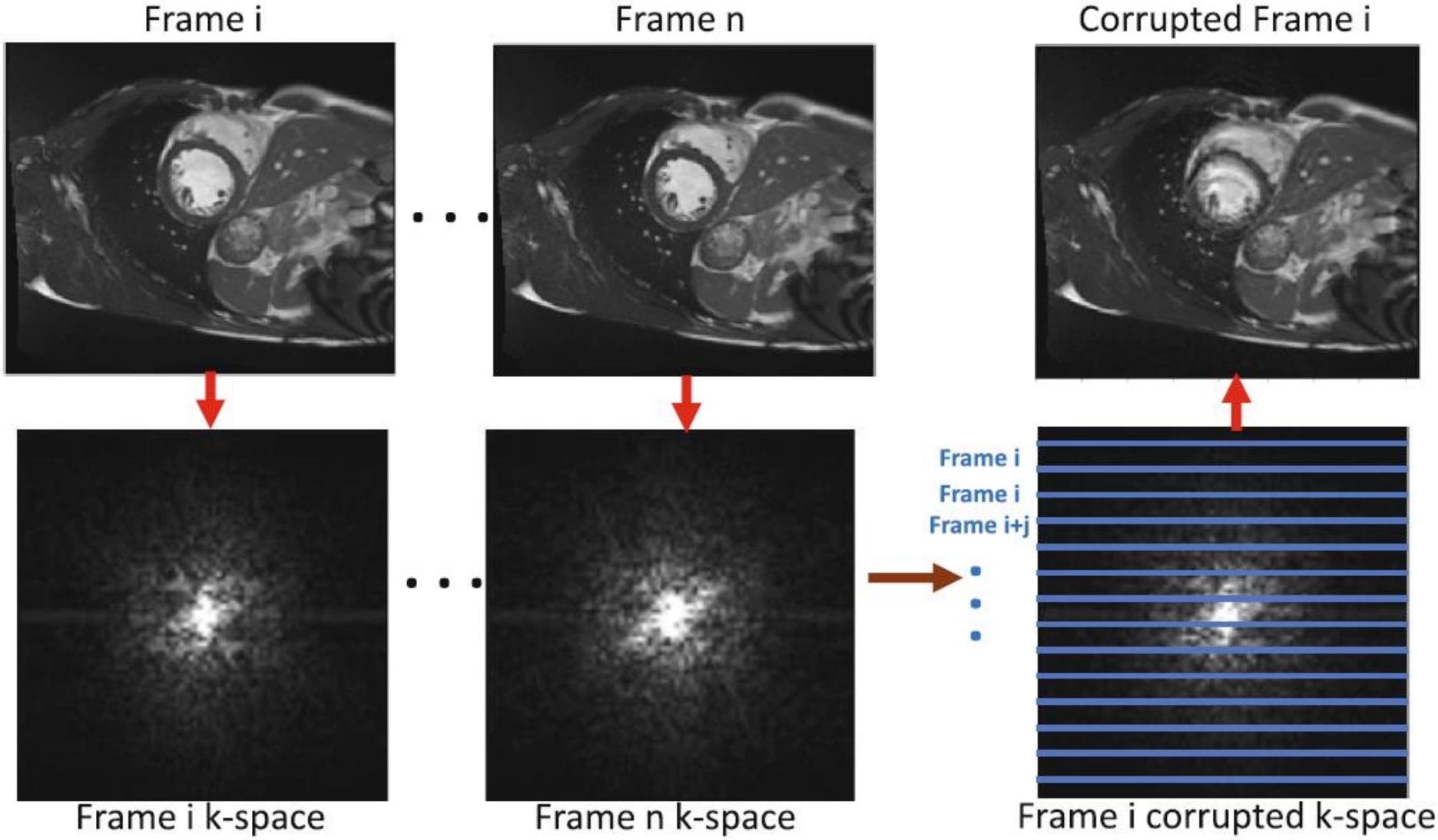


Fig. 3. K-space corruption for motion artefact generation in k-space. We use the Fourier transformation of each frame to generate the k-space of each image and replace k-space lines with lines from different temporal frames to generate corruptions.

Table 1. Mean accuracy, precision, recall and F1 score results of image classification for motion artefacts. 10-fold cross validation is used and each image is labelled once over all folds.

Methods	Accuracy	Precision	Recall	F1-score
K-nearest neighbours	0.952	0.074	0.268	0.116
Linear SVM	0.968	0.721	0.385	0.502
Decision Tree	0.951	0.250	0.385	0.303
Random Forests	0.958	0.320	0.315	0.317
Adaboost	0.960	0.230	0.567	0.327
Naive Bayesian	0.801	0.527	0.183	0.111
Variance of Laplacian	0.958	0.113	0.161	0.133
NIQE [9]	0.958	0.210	0.248	0.227
CNN-no augmentation [14]	0.968	0.700	0.466	0.560
CNN-translational augmentation	0.974	0.750	0.600	0.667
CNN-k-space augmentation	0.977	0.779	0.642	0.704
CNN-k-space+translational augmentation	0.982	0.809	0.652	0.722

Rotation Equivariant CNNs for Digital Pathology

Bastiaan S. Veeling^(✉), Jasper Linmans, Jim Winkens, Taco Cohen,
and Max Welling

University of Amsterdam, Amsterdam, The Netherlands
basveeling@gmail.com

Abstract. We propose a new model for digital pathology segmentation, based on the observation that histopathology images are inherently symmetric under rotation and reflection. Utilizing recent findings on rotation equivariant CNNs, the proposed model leverages these symmetries in a principled manner. We present a visual analysis showing improved stability on predictions, and demonstrate that exploiting rotation equivariance significantly improves tumor detection performance on a challenging lymph node metastases dataset. We further present a novel derived dataset to enable principled comparison of machine learning models, in combination with an initial benchmark. Through this dataset, the task of histopathology diagnosis becomes accessible as a challenging benchmark for fundamental machine learning research.

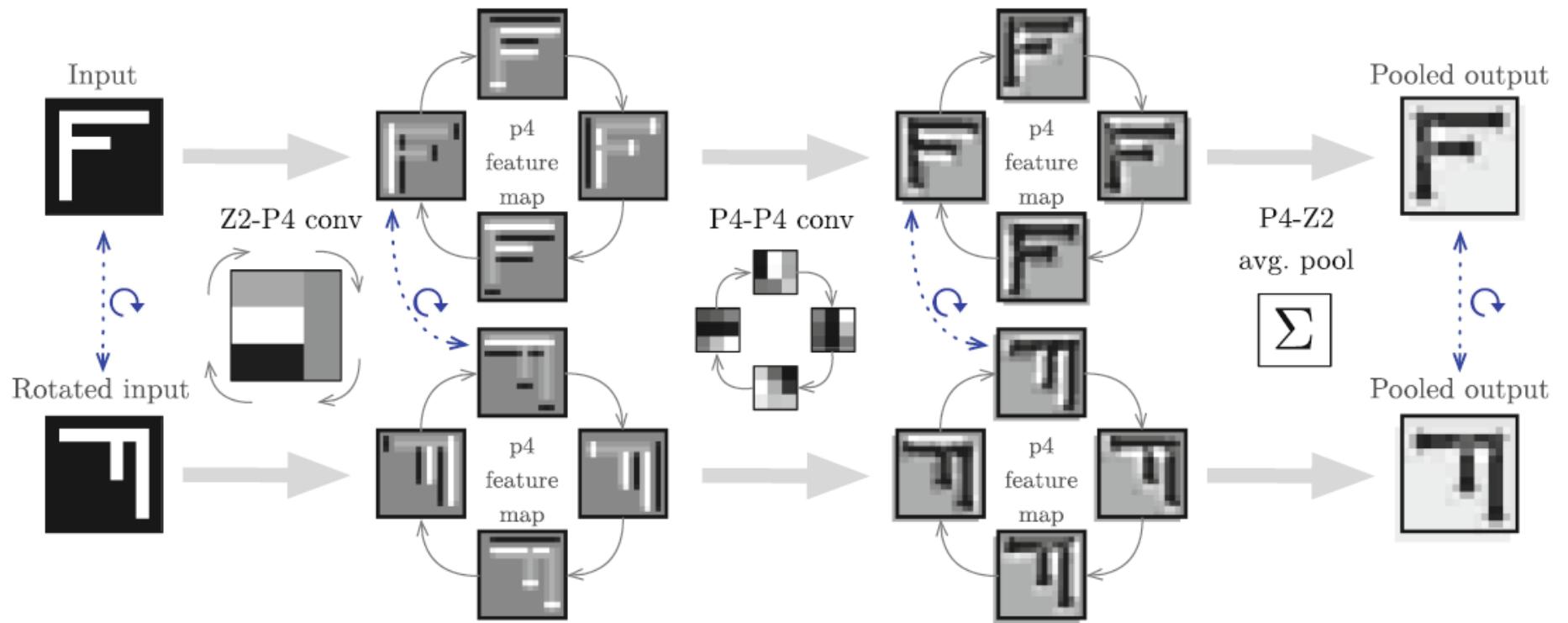


Fig. 1. Given a canonical input and a rotated duplicate, we demonstrate how a 2-layer G-CNN is equivariant in $p4$. Feature maps of one kernel per layer are shown, and the dashed blue arrows indicate how (intermediate) representations of the two inputs correspond. The $\mathbb{Z}^2 \rightarrow p4$ convolution correlates the input with 4 rotated versions of the same kernel. The $p4 \rightarrow p4$ convolution correlates the resulting feature map with the $p4$ -kernel, cyclically-shifting and rotating the kernel for each orientation. The final layer demonstrates how average-pooling over the orientations produces a representation that is locally invariant and globally equivariant to rotation. *Global* average pooling over $p4$ would result in a representation globally invariant to both translation and rotation.

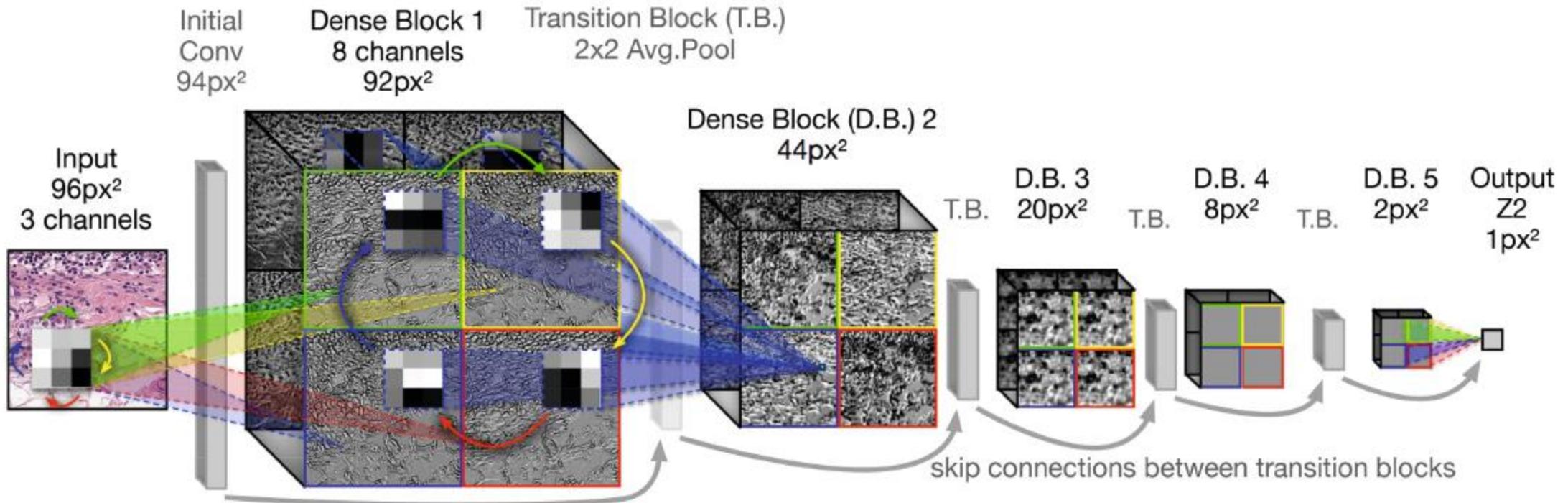


Fig. 2. The proposed equivariant DenseNet architecture for the $p4$ group, consisting of 5 Dense Blocks (D.B.) alternated with Transition Blocks (T.B.). The final layer of the model is a $p4 \rightarrow \mathbb{Z}^2$ group pooling layer followed by a sigmoid activation. The four orientations in $p4$ are illustrated through primary colors. A $\mathbb{Z}^2 \rightarrow p4$ kernel (*left*), $p4 \rightarrow p4$ kernel (*middle*) and $p4 \rightarrow \mathbb{Z}^2$ kernel (*right*) illustrate how equivariance arises in the model.

RIIS-DenseNet: Rotation-Invariant and Image Similarity Constrained Densely Connected Convolutional Network for Polyp Detection

Yixuan Yuan^{1,2}(✉) , Wenjian Qin³, Bulat Ibragimov², Bin Han², and Lei Xing²

¹ Department of Electronic Engineering, City University of Hong Kong,
Hong Kong, China
yxyuan.ee@cityu.edu.hk

² Department of Radiation Oncology, Stanford University, Stanford, USA
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences,
Shenzhen, China

Abstract. Colorectal cancer is the leading cause of cancer-related deaths. Most colorectal cancers are believed to arise from benign adenomatous polyps. Automatic methods for polyp detection with Wireless Capsule Endoscopy (WCE) images are desirable, but the results of current approaches are limited due to the problems of object rotation and high intra-class variability. To address these problems, we propose a rotation invariant and image similarity constrained Densely Connected Convolutional Network (RIIS-DenseNet) model. We first introduce Densely Connected Convolutional Network (DenseNet), which enables the maximum information flow among layers by a densely connected mechanism, to provide end-to-end polyp detection workflow. The rotation-invariant regularization constraint is then introduced to explicitly enforce learned features of the training samples and the corresponding rotation versions to be mapped close to each other. The image similarity constraint is further proposed by imposing the image category information on the features to maintain small intra-class scatter. Our method achieves an inspiring accuracy 95.62% for polyp detection. Extensive experiments on the WCE dataset show that our method has superior performance compared with state-of-the-art methods.

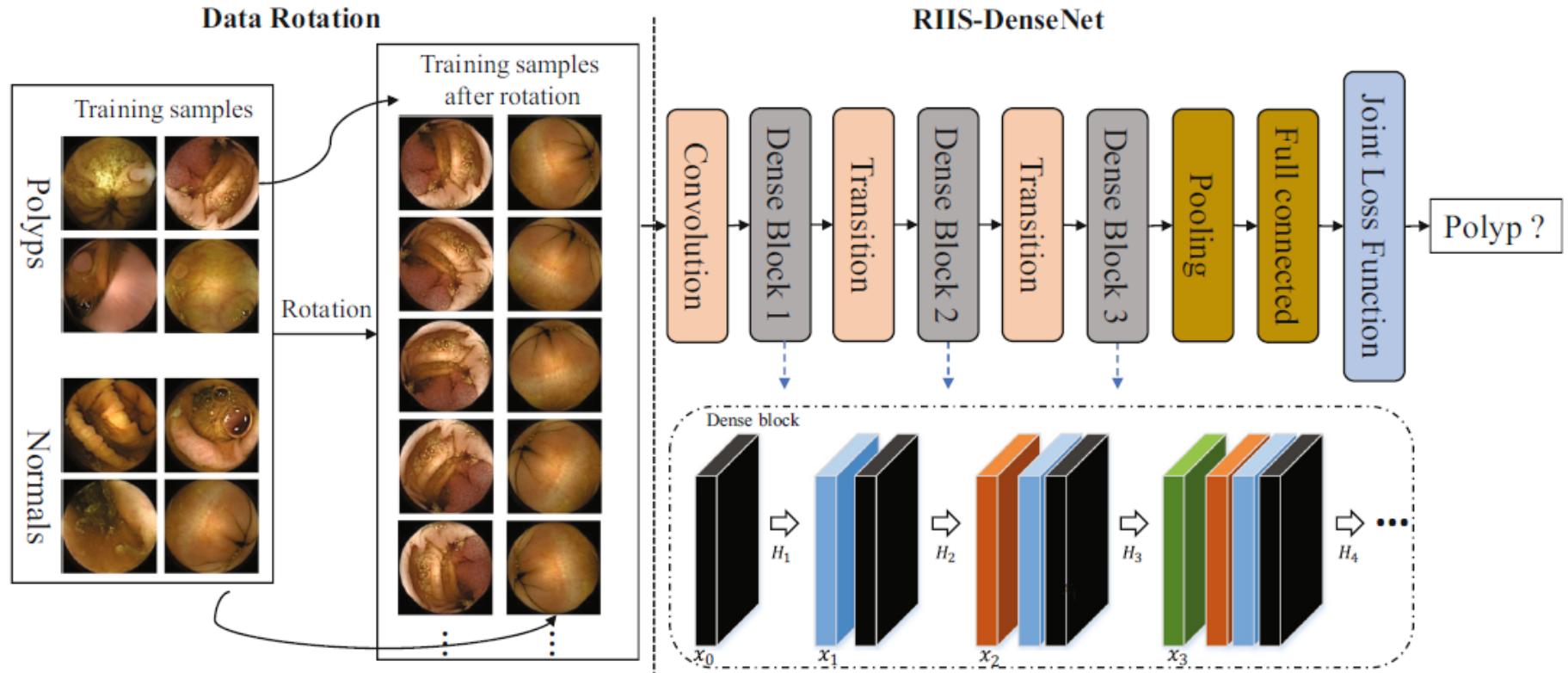
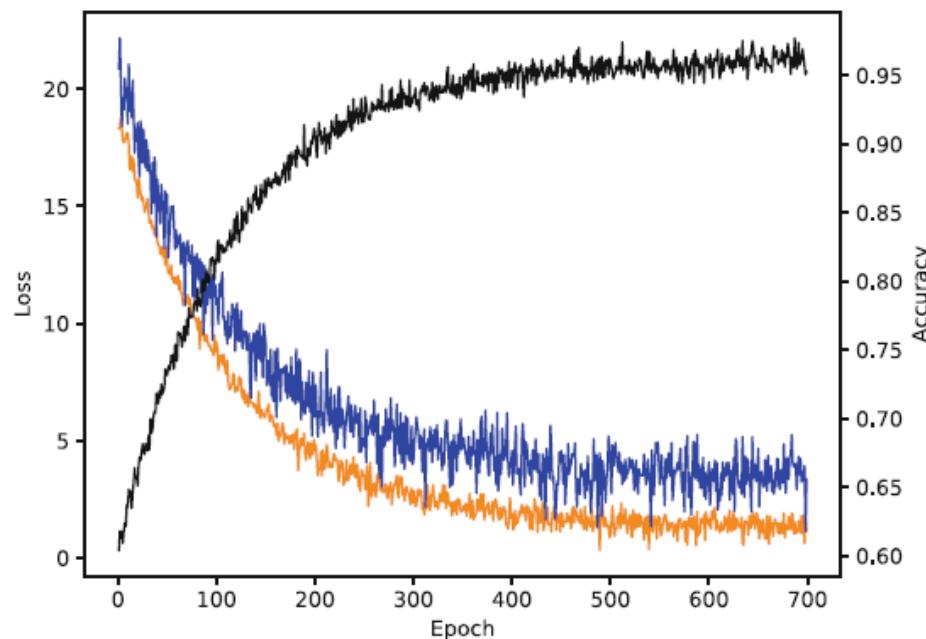
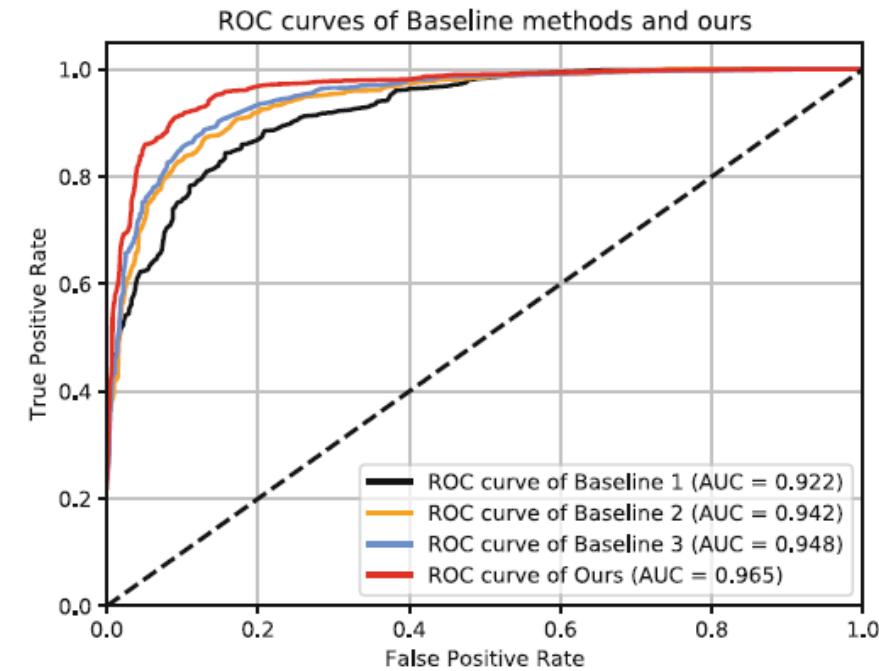


Fig. 1. Workflow of our proposed RIIS-DenseNet. It consists of two parts: data rotation augmentation and RIIS-DenseNet. The RIIS-DenseNet includes three denseblocks, two transitions, one convolution layer, one pooling layer and a novel joint loss function layer.



(a)



(b)

Fig. 2. (a) Loss and accuracy values for different iterations. The blue color represents the test loss while the orange color shows the training loss. The black line represents the test accuracy. (b) ROC curves for different baseline methods and ours.