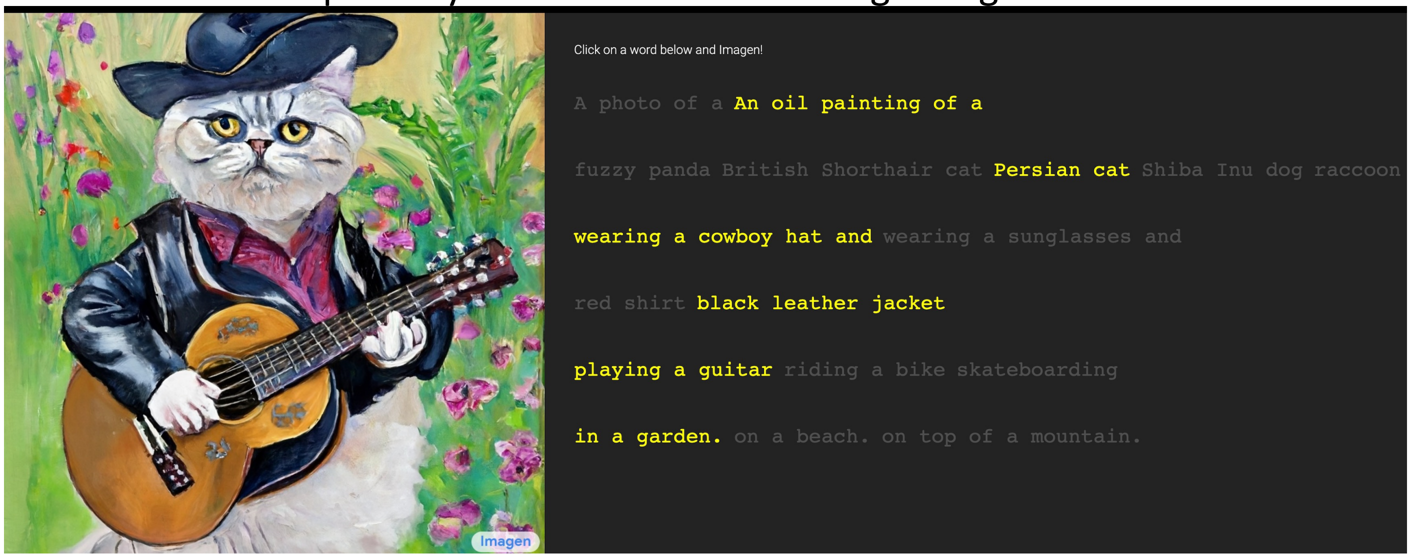
Advanced concepts in deep learning 2

Hassan Rivaz
Associate Professor
Concordia University Research Chair in Medical Image Analysis
Director, IMPACT laboratory
sonography.ai

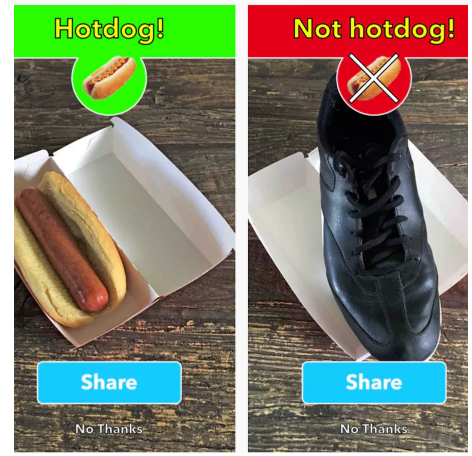
IMPACT Laboratory
Est. 2014
Transforming Healthcare

XAI (Explainable AI)

- Classify different dog breeds (e.g. the Stanford database with 120 breeds) or hotdogs. Generate cute cat pictures!
- Healthcare, self-driving cars
- Black-box nature of DL hinders its adoption in critical applications
- XAI provides interpretable and understandable explanations
- XAI can also partially address some issues regarding ethics and bias



From https://imagen.research.google



From:
www.theverge.com/tldr/2017/5/14/15639784/hbo-silicon-valley-not-hotdog-app-download

IMPACT Laboratory
Est. 2014
Transforming Healthcare

XAI Approaches

- **Perturbation**: Changes to features in the input data (e.g. masking, conditional sampling).
- **BackProb**: Back-propagated gradients from output prediction layer back to input layer, e.g. saliency maps, saliency relevance maps, and class activation maps

Perturbation XAI, RISE

- Randomized Input Sampling for Explanation (RISE), 2018:

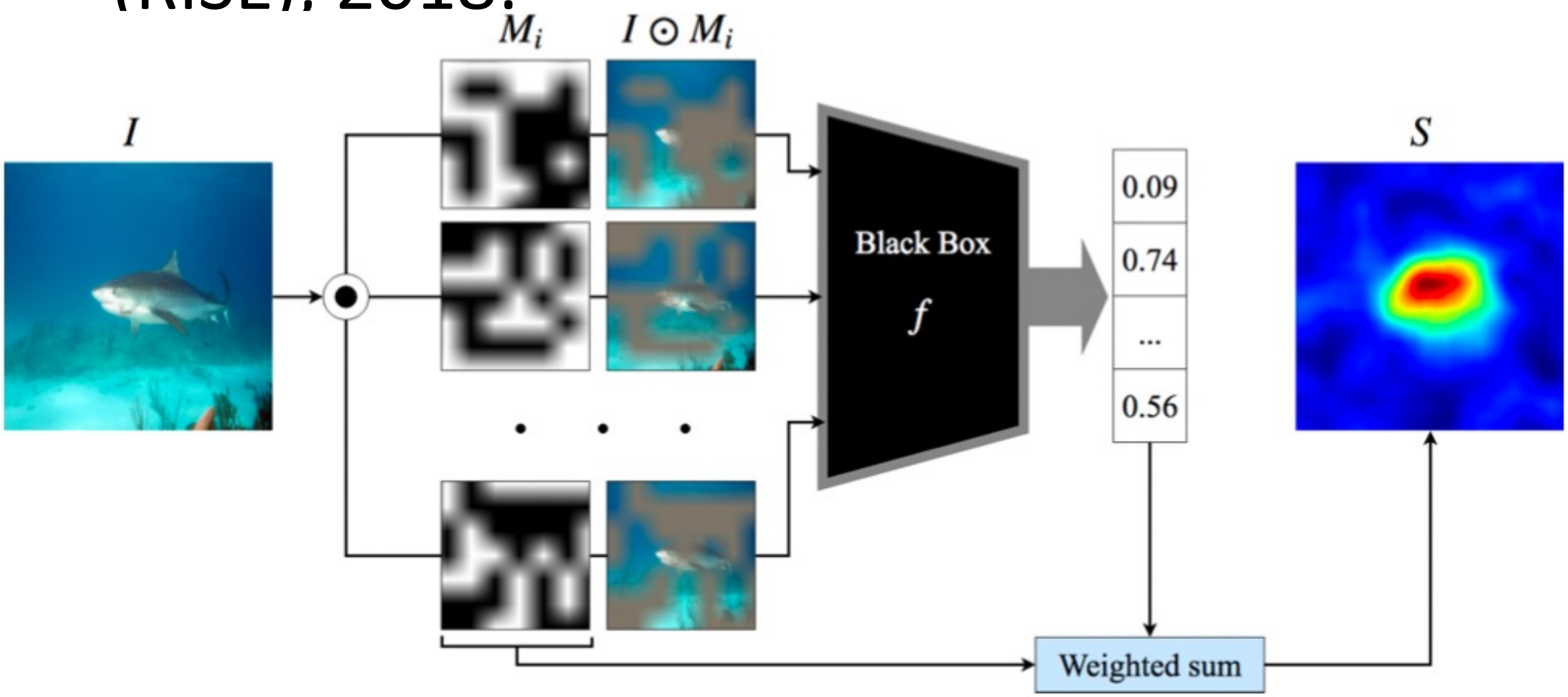
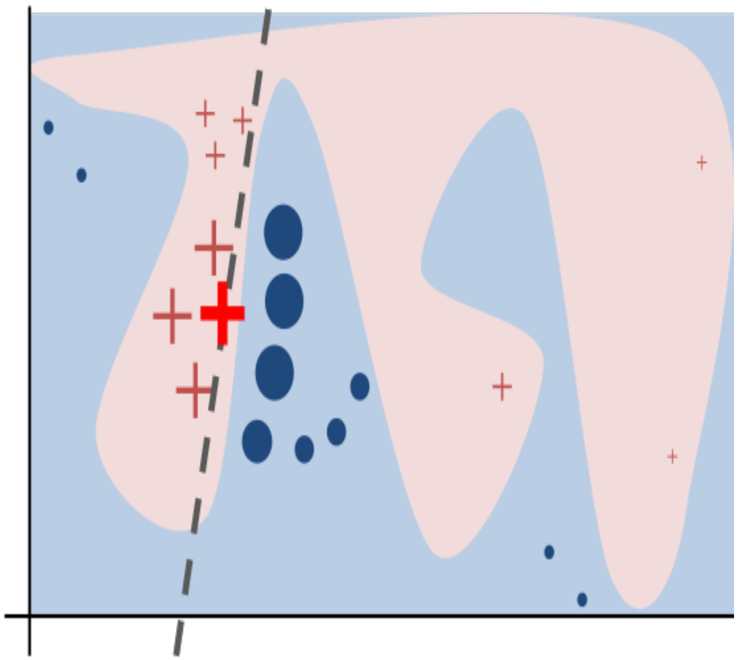


Figure 3: Overview of RISE: Input image I is element-wise multiplied with random masks M_i and the masked images are fed to the base model. The saliency map is a linear combination of the masks where the weights come from the score of the target class corresponding to the respective masked inputs.

IMPACT Laboratory Est. 2014 Transforming Healthcare

Perturbation XAI, LIME

- Black-box model's complex decision function (unknown to LIME) is the blue/pink background
  - It cannot be approximated.
- Bold red cross is the instance being explained.
- LIME samples instances, gets predictions using f, and weighs them by the proximity to the instance being explained (represented by size).
- The dashed line is the learned explanation that is locally faithful.



From M. T. Ribeiro et al. 2016

Perturbation XAI, LIME

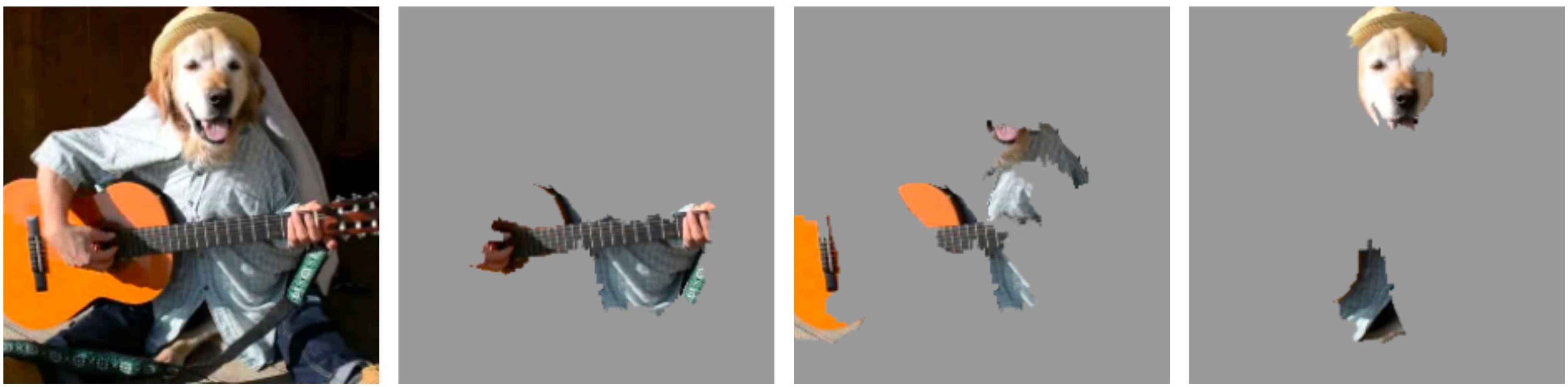
- Iteratively probing a trained model with different variations of the inputs
- Local Interpretable Model-agnostic Explanations (LIME), 2016:

(a) Original Image

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

Perturbation XAI, LIME

- Iteratively probing a trained model with different variations of the inputs
- Local Interpretable Model-agnostic Explanations (LIME), 2016:

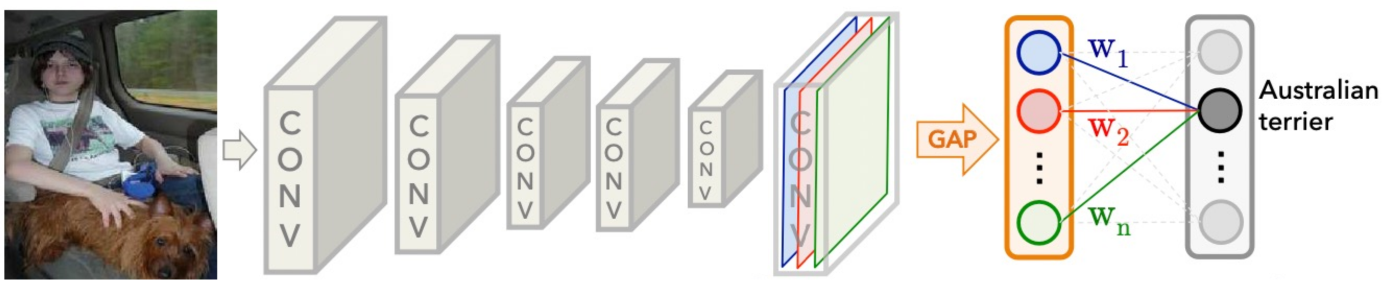


(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

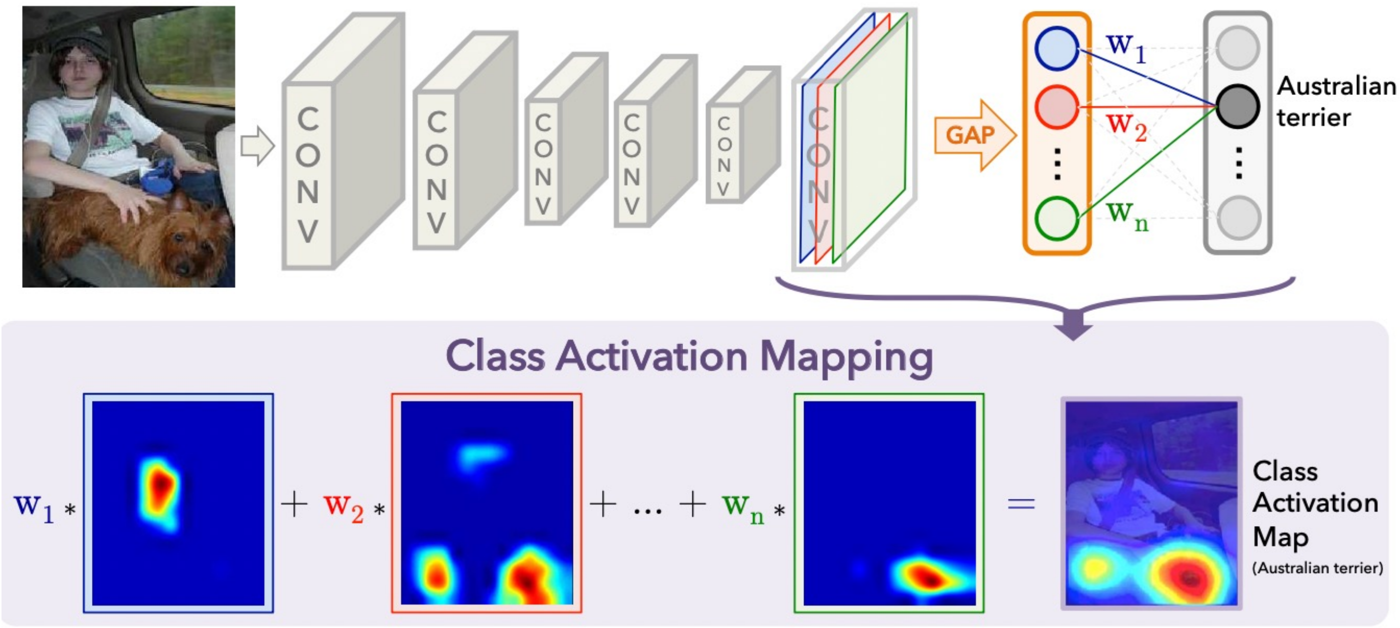
BackProp XAI, CAM

- Class Activation Maps (CAM), Zhou et al. CVPR 2016
- GAP outputs the spatial average of the feature map of each unit at the last convolutional layer



BackProp XAI, CAM

- Class Activation Maps (CAM), Zhou et al. CVPR 2016
- GAP outputs the spatial average of the feature map of each unit at the last convolutional layer
- GAP has remarkable localization ability despite being trained on image-level labels



BackProp XAI, GradCAM

- CAM uses a restricted class of image classification CNNs which do **not contain any fully-connected layers**.
- In contrast, GradCAM make existing state-of-the-art deep models interpretable without altering their architecture

Grad-CAM:
Visual Explanations from Deep Networks via Gradient-based Localization

ICCV 2017

Ramprasaath R. Selvaraju[1*] Michael Cogswell[1] Abhishek Das[1] Ramakrishna Vedantam[1*]
Devi Parikh[1,2] Dhruv Batra[1,2]
[1]Georgia Institute of Technology [2]Facebook AI Research
{ramprs, cogswell, abhshkdz, vrama, parikh, dbatra}@gatech.edu

IMPACT Laboratory
Est. 2014
Transforming Healthcare

The false hope of current approaches to explainable artificial intelligence in health care

The Lancet Digital Health, 2021

Marzyeh Ghassemi, Luke Oakden-Rayner, Andrew L Beam

- For limited complexity input, the relationships between inputs and the outputs is termed **inherent explainability**.
  - Linear regression model, where a coefficient measures the strength of the relationship between the weight of a car and the fuel efficiency.
- In contrast to inherently explainable models, the data and models are too complex in many modern AI use cases
- Another challenge: explanations have no performance guarantees. The performance of explanations is rarely tested, and rely on heuristic measures.
- Many drugs and medical devices function as black boxes, eg. acetaminophen, which, despite having been used for more than a century, has a mechanism of action that is partially understood

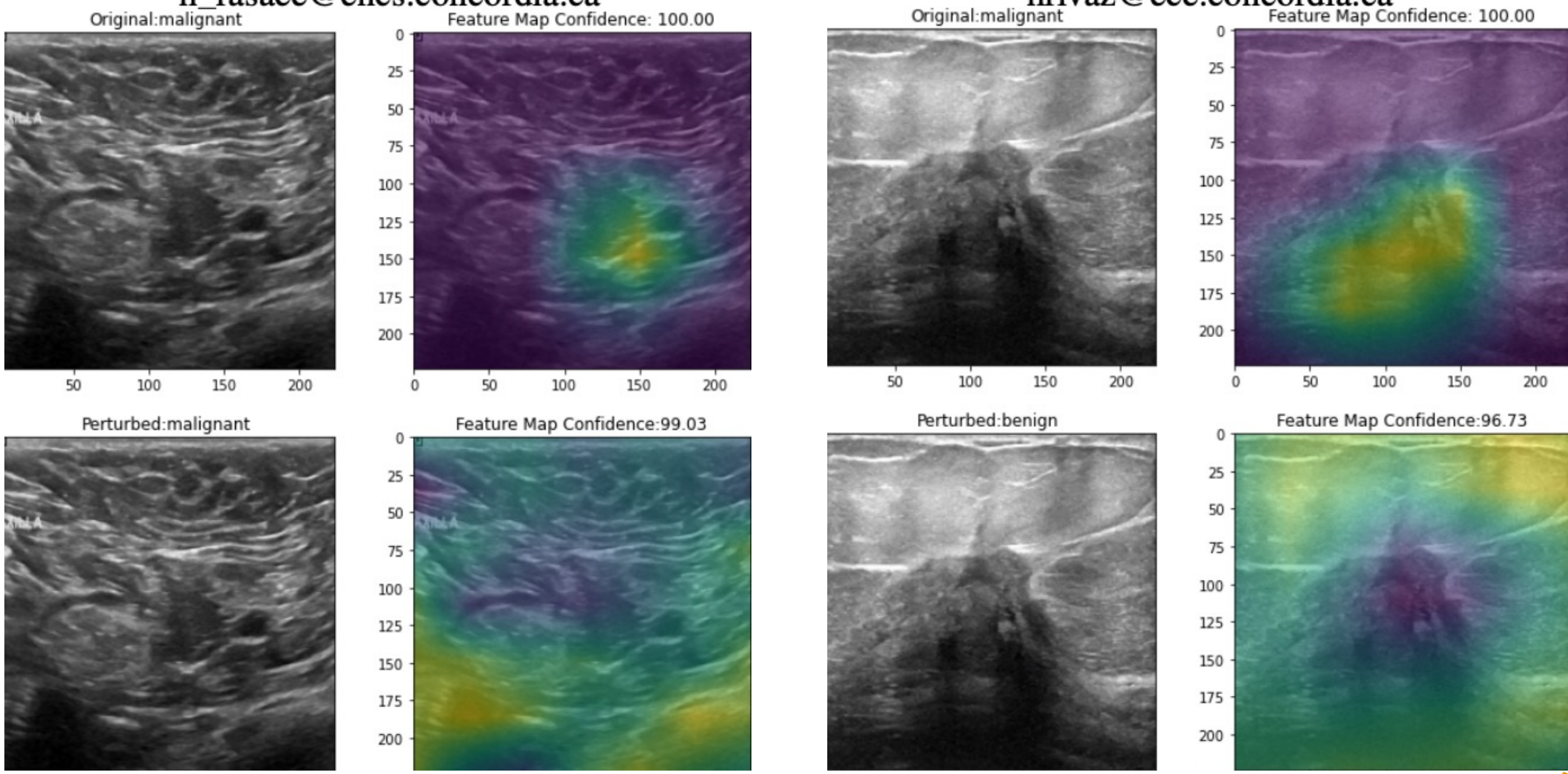
IMPACT Laboratory
Est. 2014
Transforming Healthcare

Explainable AI and susceptibility to adversarial attacks: a case study in classification of breast ultrasound images

IEEE IUS 2021

Hamza Rasaee
Electrical and Computer Engineering
Concordia University
Montreal, Canada
h_rasaee@encs.concordia.ca

Hassan Rivaz
Electrical and Computer Engineering
Concordia University
Montreal, Canada
hrivaz@ece.concordia.ca



IMPACT Laboratory
Est. 2014
Transforming Healthcare

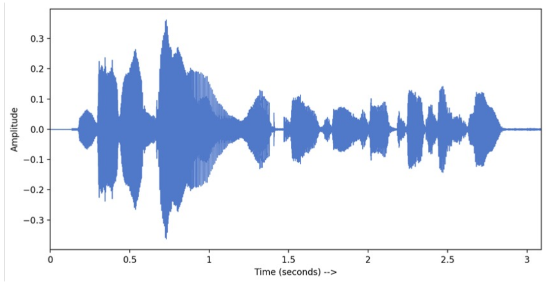
Generative models for XAI?

Made by Imagen. Prompt: the most beautiful punk rock cat singing into a microphone. From: https://twitter.com/DynamicWebPaige/status/1538315418355412997?s=20&t=qqxXBHhjg6MCJxU_mTRloA

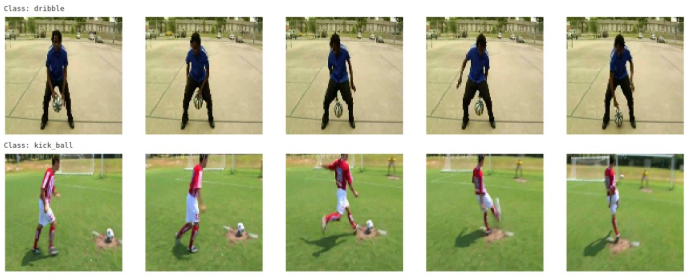
Recurrent Neural Networks

- Many real-life data are actually **sequences** (e.g., time series):

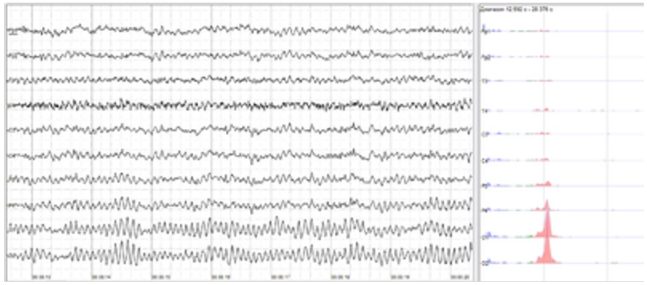
Audio



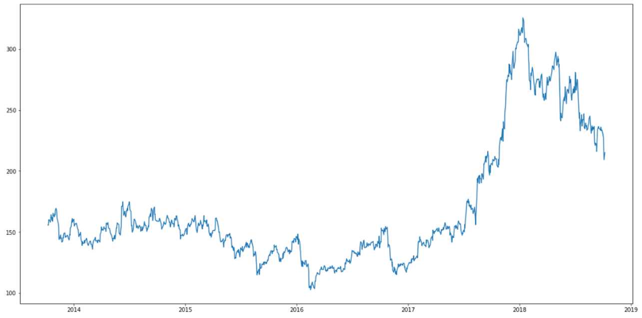
Video



EEG Brain signals



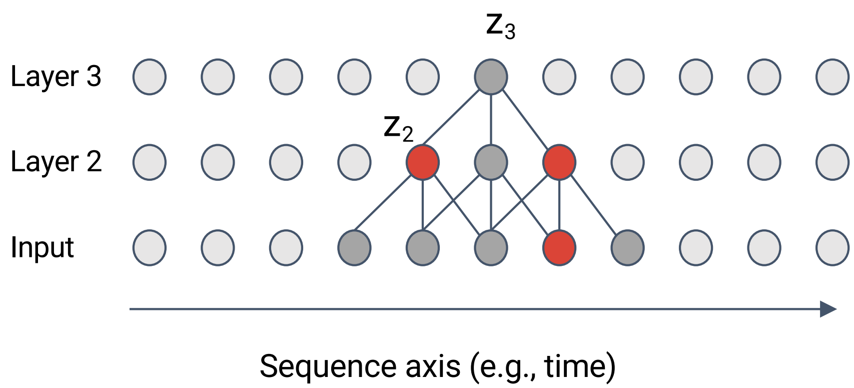
Stock Price Prediction



…………

Courtesy of Prof. Mirco Ravanelli

Recurrent Neural Networks



- CNNs are designed to **capture local dependencies** (i.e, the ones available in the receptive field).

- For some types of signals, we want to learn **long-term dependencies**.

- For instance, we might want to have a model whose neurons depend on the **full history** (i.e., **all** the **past elements**).

Courtesy of Prof. Mirco Ravanelli

Recurrent Neural Networks

- In **Recurrent Neural Networks** (RNNs), the **current output** depends on **all the previous inputs**.
- The general form for an RNN is the following:

$$\boxed{\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}, \mathbf{W})}$$

The **current state $\mathbf{h}_t$** depends on:

- The current input $\mathbf{x}_t$
- The previous state $\mathbf{h}_t$
- A set of learnable parameters **W**

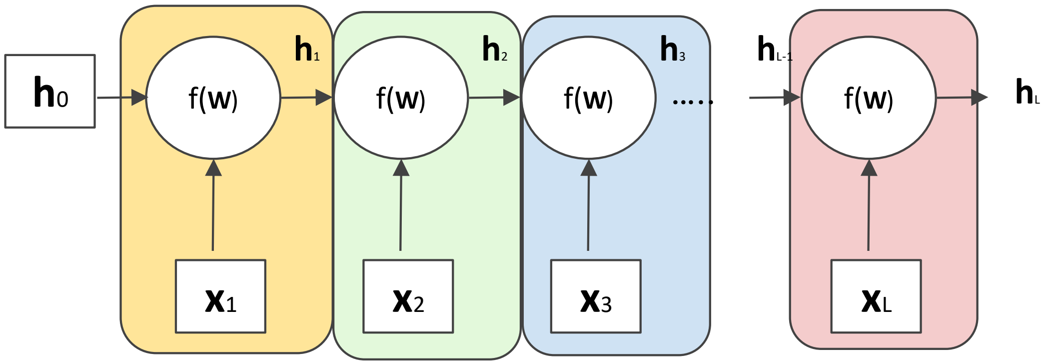
Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Recurrent Neural Networks

- We can **unfold** the recurrent neural networks in this way:

$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}, \mathbf{W})$$



$$\mathbf{h}_0 = 0$$

$$\mathbf{h}_1 = f(\mathbf{x}_1, \mathbf{h}_0, \mathbf{W})$$

$$\mathbf{h}_2 = f(\mathbf{x}_2, \mathbf{h}_1, \mathbf{W})$$

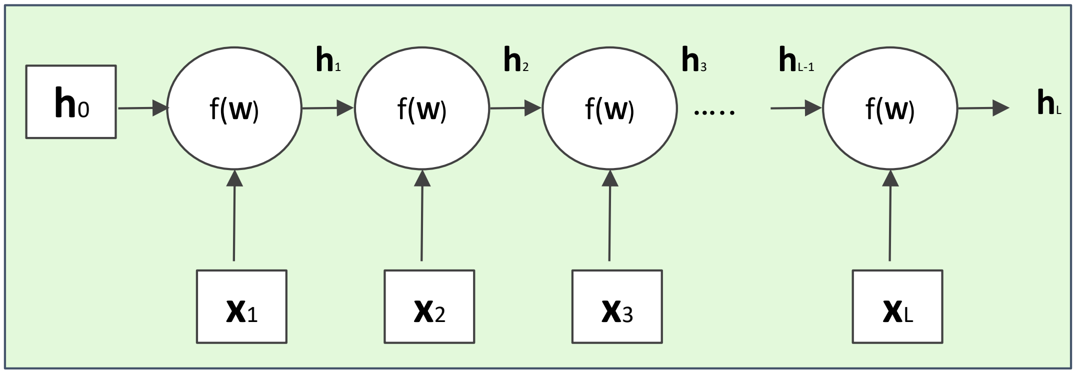
$$\mathbf{h}_3 = f(\mathbf{x}_3, \mathbf{h}_2, \mathbf{W})$$

$$\mathbf{h}_L = f(\mathbf{x}_L, \mathbf{h}_{L-1}, \mathbf{W})$$

- The current state $\mathbf{h}_t$ depends on the current input $\mathbf{x}_t$ and all the previous ones.
- The final state $\mathbf{h}_L$ depends on **all the inputs.**

Courtesy of Prof. Mirco Ravanelli

Recurrent Neural Networks

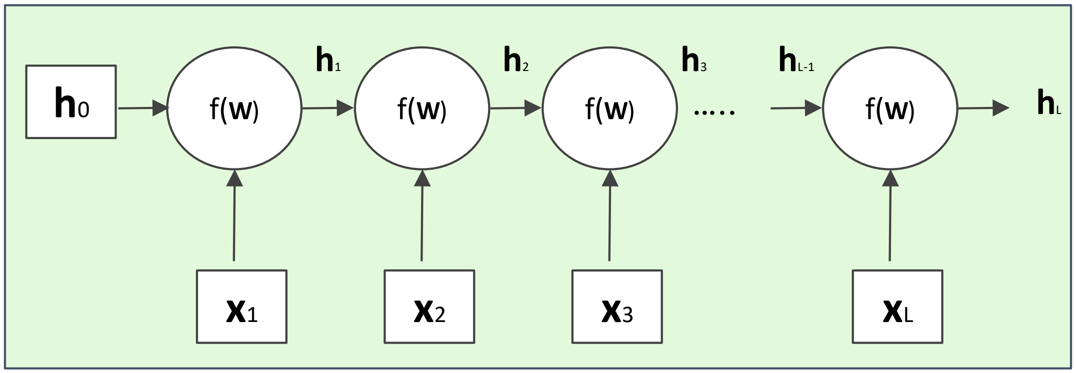


$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}, \mathbf{W})$$

- We apply the same neural network *f* for all the sequence steps (**weight sharing**). Where else do we do this?
- As seen for CNNs, applying the **same weights** over the input helps **find patterns** in the data.
- In CNNs, we share the same filter over the inputs, for RNNs we share a **full neural network**.
- Moreover, RNNs we can find **arbitrary long patterns** because the state vector acts as a memory of the previous inputs.

Courtesy of Prof. Mirco Ravanelli

Recurrent Neural Networks



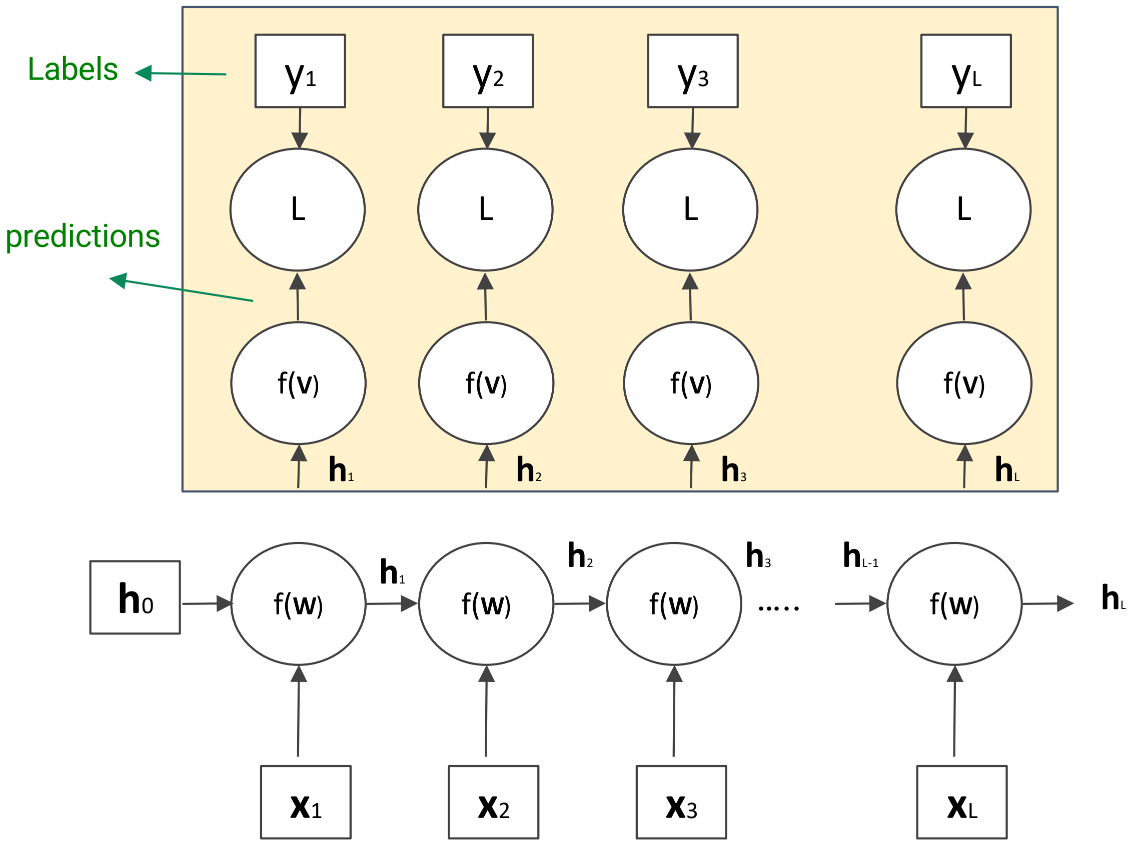
$$\mathbf{h}_t = f(\mathbf{x}_t, \mathbf{h}_{t-1}, \mathbf{W})$$

- Once unfolded, RNN can be represented as a single **big computational graph**.
- We can see a recurrent neural network as a **deep neural network**, which is deep in the **time axis**.
- RNNs can be used in **different types** of **problems,** that include:
    - Problems where we need a **prediction at each time step** (Many-to-many).
    - Problems where we need a single **prediction at the end** (Many-to-one).
    - Problems where we want to turn the **input sequence** of length T into an **output sequence of** length S **(Seq-to-seq**).

Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory Est. 2014 Transforming Healthcare

Many-to-Many Problems



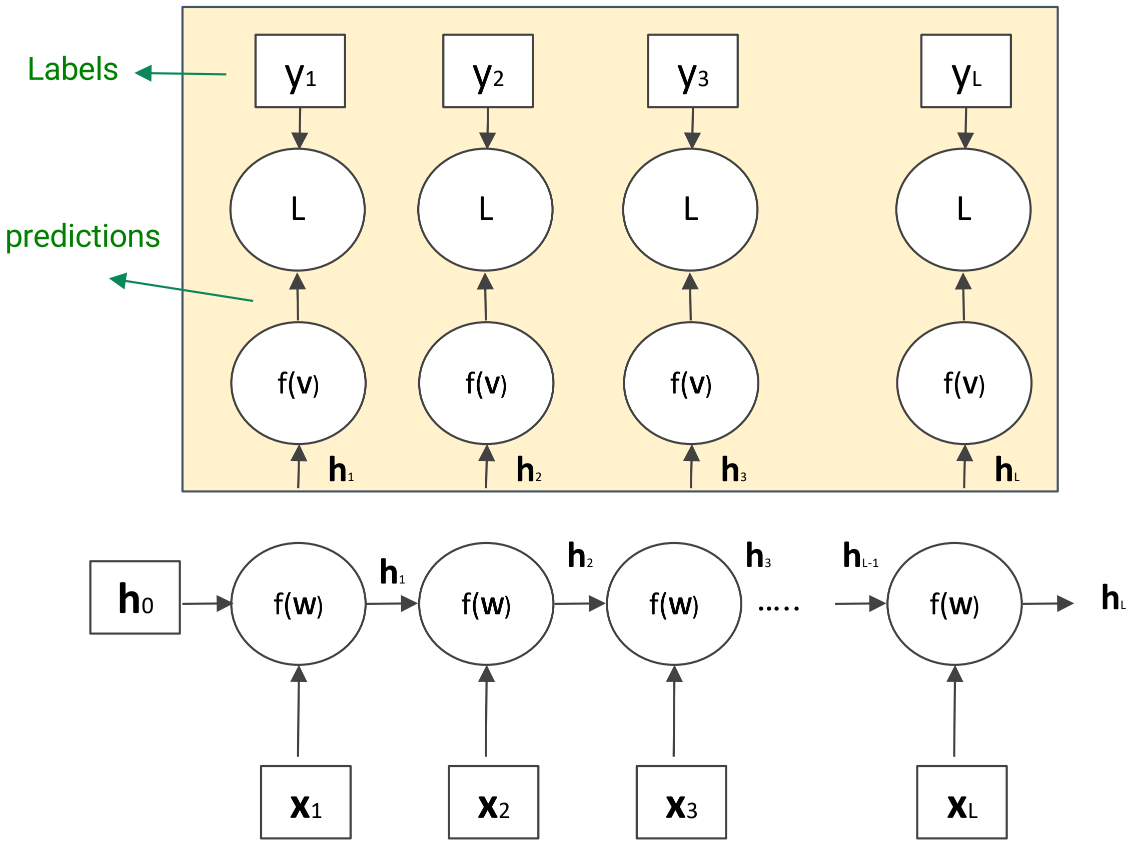
When we need a **single prediction at each time step** (many-to-many problem), we have to:

- Apply a **final transformation** (e.g., linear model) on top of each hidden state.
- Apply a **loss function at each time step** (e.g., Categorical Cross-Entropy for multi-class classification, MSE for regression).
- The **total loss** will be the **sum** (or **average**) of all the losses at each time step.

Courtesy of Prof. Mirco Ravanelli

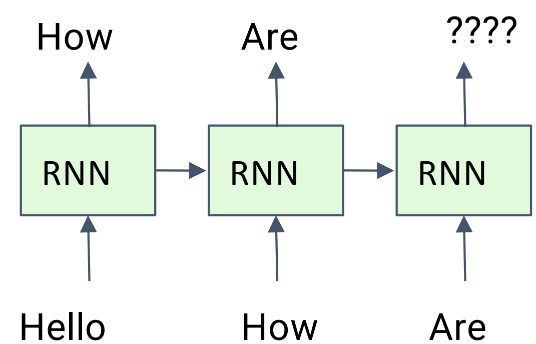
IMPACT Laboratory
Est. 2014
Transforming Healthcare

Many-to-Many Problems



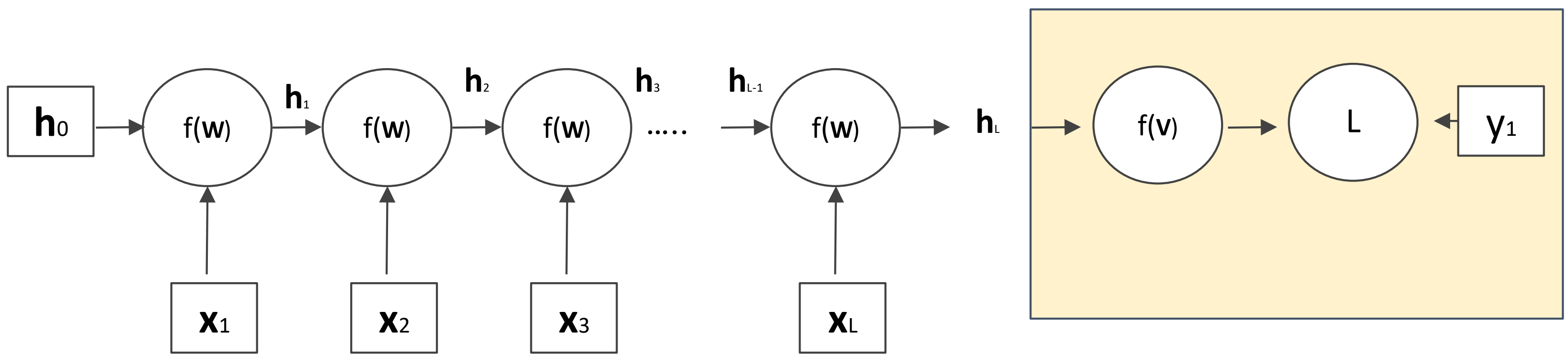
- Examples of problems like this are **stock predictions**, where we need to predict a value every day or hour.

- Another example is **language modeling**, where we want to predict the next word given the previous ones.



Courtesy of Prof. Mirco Ravanelli

Many-to-one Problems



When we need a **single prediction** in output (many-to-one problem), we can do the following:

- Apply a **final transformation** (e.g., linear model) on top of the **last hidden state** (which depends on all the inputs).

- Apply a **loss function** on top of it (e.g., Categorical Cross-Entropy for multi-class classification, MSE for regression).

Examples of problems like this are **speaker identification,** emotion recognition from text, etc.

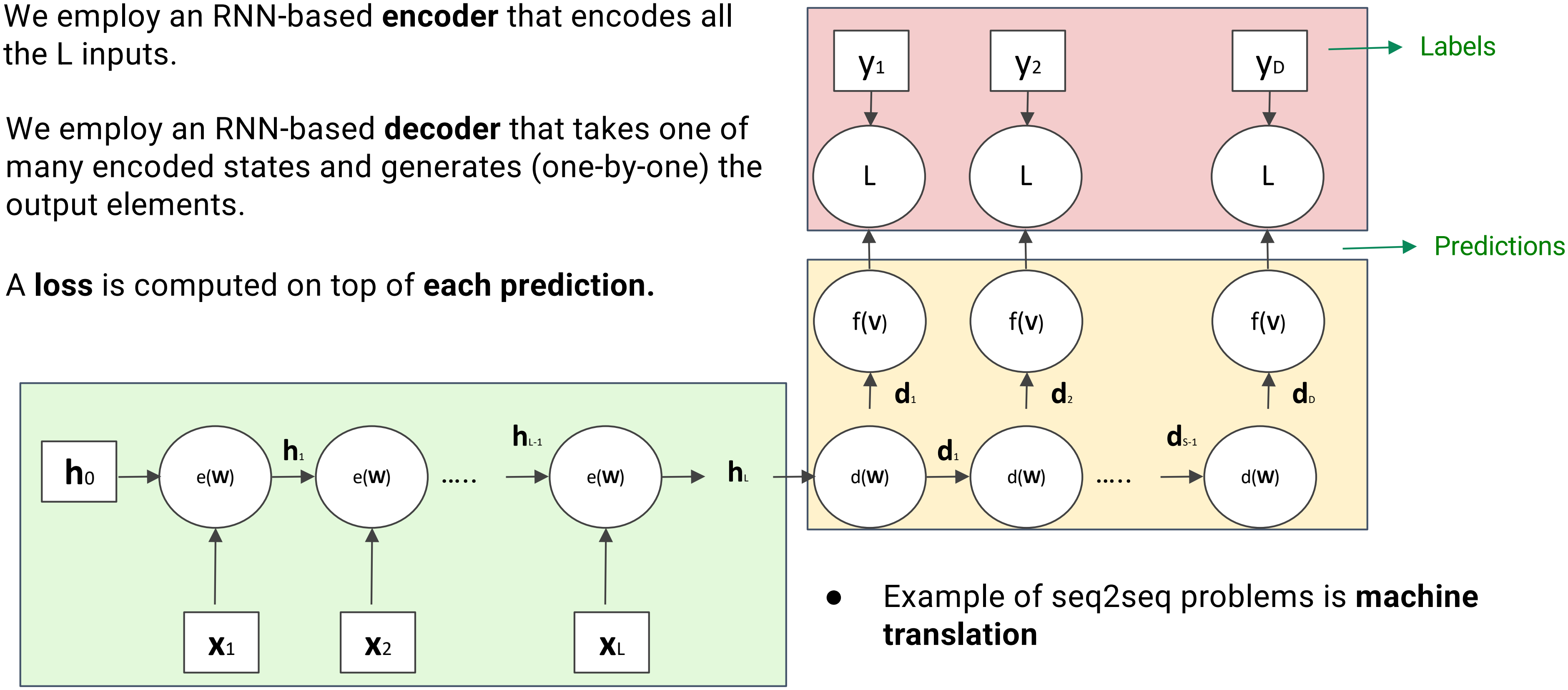
Courtesy of Prof. Mirco Ravanelli

Sequence-to-Sequence Problems

- RNNs can also be used for **sequence-to-sequence problems**:

Attention mechanisms are often use to connect encoder and decoder states.

- We employ an RNN-based **encoder** that encodes all the L inputs.
- We employ an RNN-based **decoder** that takes one of many encoded states and generates (one-by-one) the output elements.
- A **loss** is computed on top of **each prediction.**

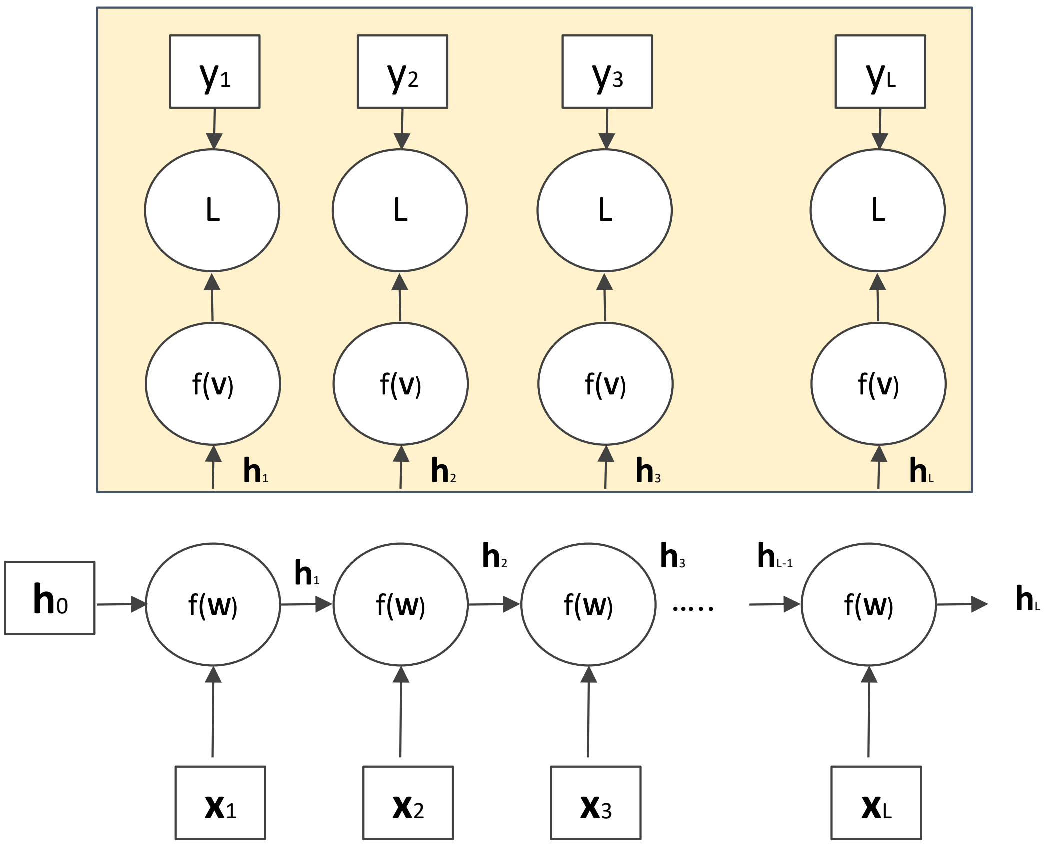


- Example of seq2seq problems is **machine translation**

Courtesy of Prof. Mirco Ravanelli

Training

- *How can we train an RNN?*



- After unfolding it, the RNN becomes a standard **computational graph** that employs long chains of computations.

- We can thus compute the gradient with the **backpropagation algorithm**.

- To highlight that the gradient is mainly propagated over time, it is sometimes called **backpropagation through time**.

- Once we have the gradient, we can update the parameters with **gradient descend** (and all its variants seen so far).

Courtesy of Prof. Mirco Ravanelli

Vanilla RNN

- The simplest RNN is called **Vanilla RNN** (or Elman RNN) is based on the following equation:

$$\boxed{\mathbf{h}_t = tanh(\mathbf{W}^{(in)T}\mathbf{x}_t + \mathbf{W}^{(hh)T}\mathbf{h}_{t-1})}$$

We perform a **linear transformation** of both the current input and the previous state.

We sum them up and apply a **non-linearity** (Tanh or ReLU).

$$\mathbf{W}^{(in)} = \begin{bmatrix} w_{0,1}^{(in)} & w_{0,2}^{(in)} & \dots & w_{0,M}^{(in)} \\ w_{1,1}^{(in)} & w_{1,2}^{(in)} & \dots & w_{1,M}^{(in)} \\ \dots & \dots & \dots & \dots \\ w_{D,1}^{(in)} & w_{D,2}^{(in)} & \dots & w_{D,M}^{(in)} \end{bmatrix} \quad \mathbf{x}_t = [1, x_{t,1}, x_{t,2}, \dots, x_{t,D}]^T$$

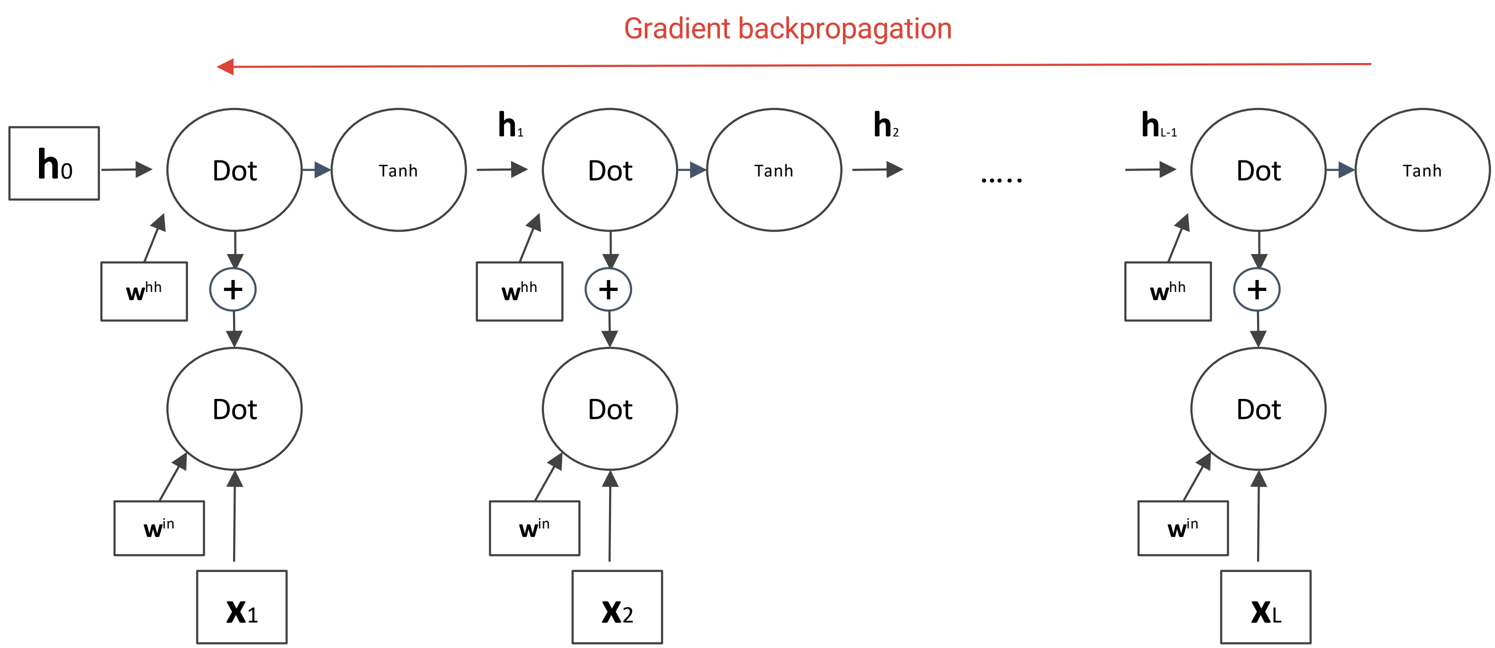
$$\mathbf{W}^{(hh)} = \begin{bmatrix} w_{1,1}^{(hh)} & w_{1,2}^{(hh)} & \dots & w_{1,M}^{(hh)} \\ w_{2,1}^{(hh)} & w_{2,2}^{(hh)} & \dots & w_{2,M}^{(hh)} \\ \dots & \dots & \dots & \dots \\ w_{M,1}^{(hh)} & w_{M,2}^{(hh)} & \dots & w_{M,M}^{(hh)} \end{bmatrix} \quad \mathbf{h}_t = [h_{t,1}, h_{t,2}, \dots, h_{t,M}]^T$$

Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Vanishing Gradient

- When using RNNs, we have to backpropagate the gradient through a **long temporal** chain.
- As we have seen, this might cause **exploding gradients** or **vanishing gradients**.
- The vanishing gradient problem prevents the model to learn **long-term dependencies**.

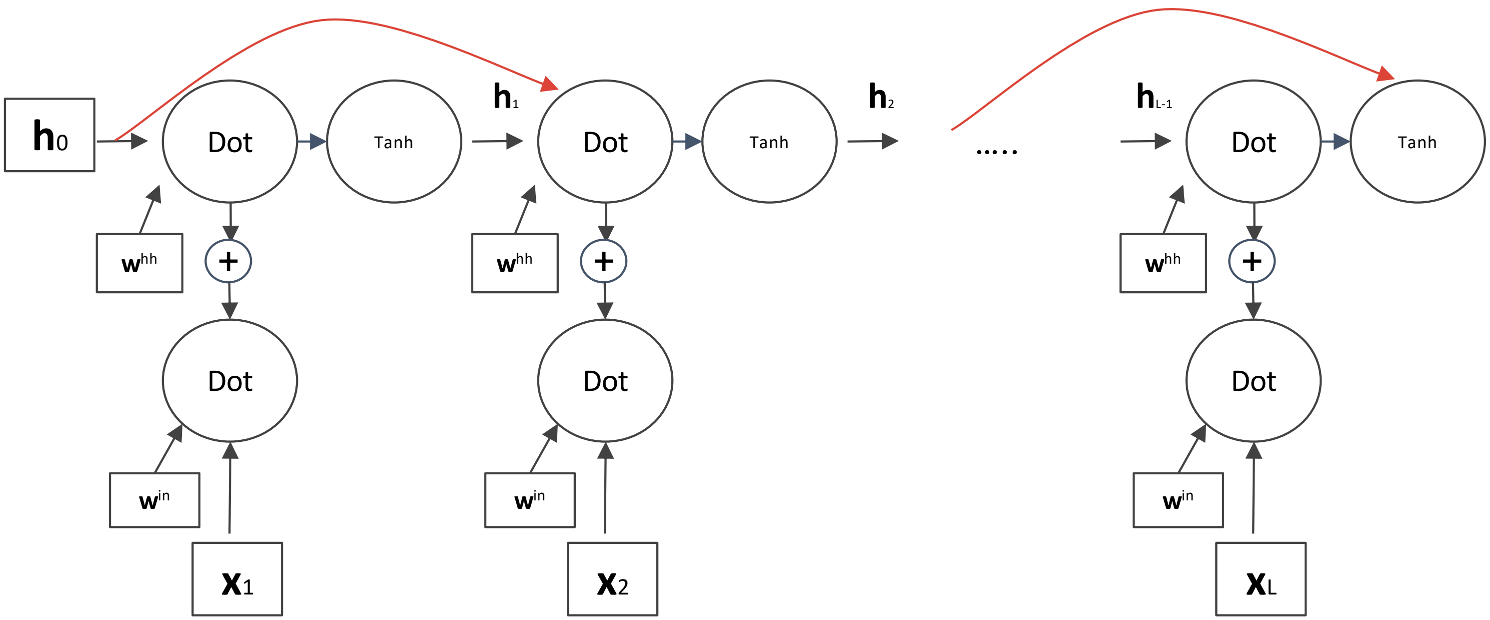


Courtesy of Prof. Mirco Ravanelli

Vanishing Gradient

- As we have seen in the previous lecture, we can add **shortcuts** in the architecture to fight vanishing gradients.

- In this case, we can consider shortcuts across **time steps**.



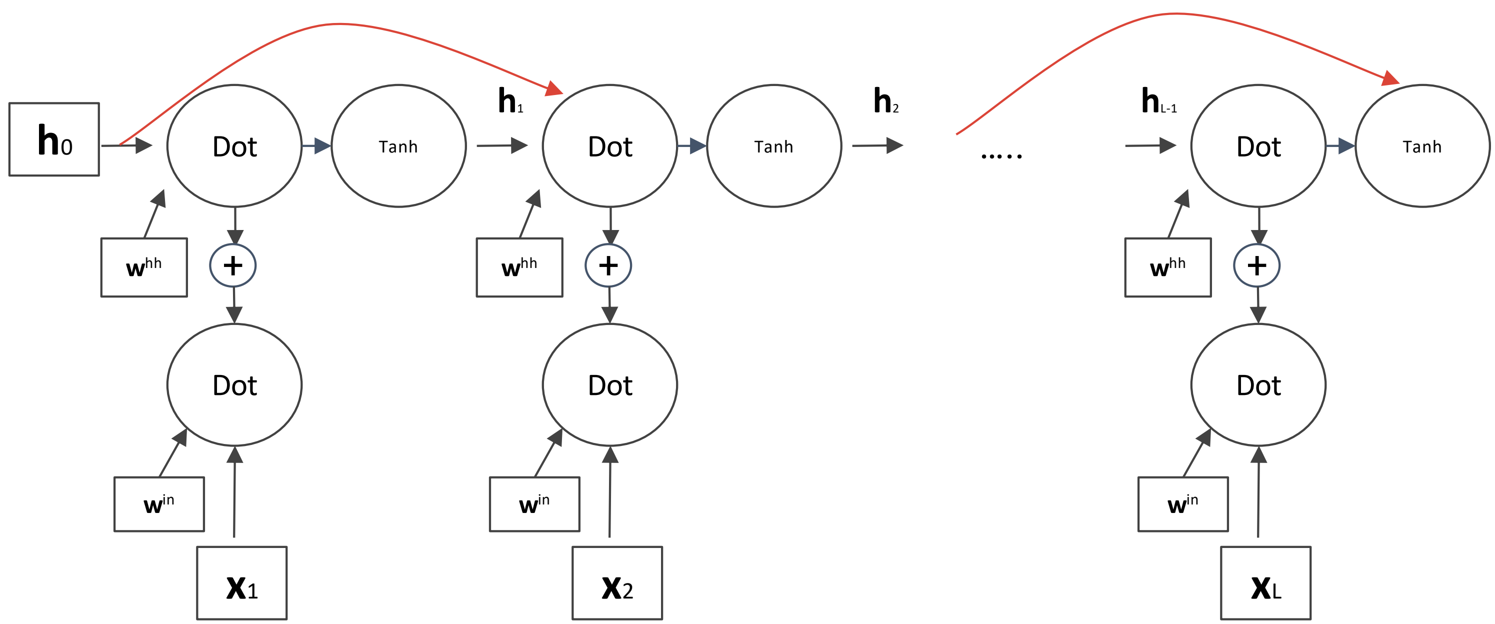
Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Vanishing Gradient

Instead of **hard-coding pre-defined shortcuts**, why don't we try to **learn** them?

Ideally, we want to learn **"dynamic" shortcuts** that connect relevant time steps.



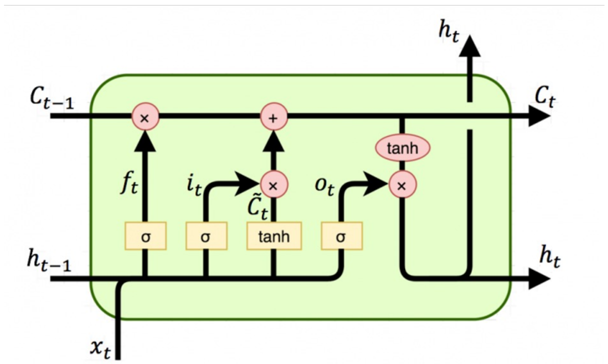
Courtesy of Prof. Mirco Ravanelli

LSTM, Gated RNNs

- This is exactly what we are trying to do with **multiplicative gates**:

Long Short-Term Memory (LSTM)

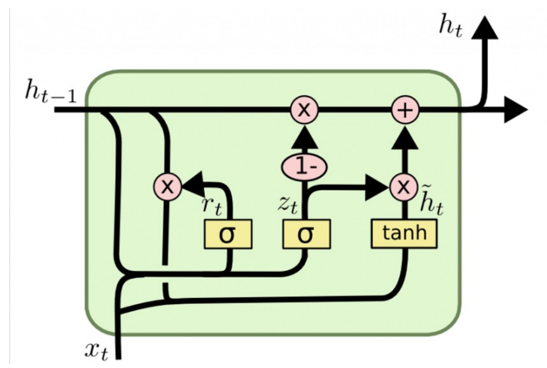
$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \\
h_t &= o_t \circ \sigma_h(c_t)
\end{aligned}
$$



4 multiplicative gates

Gated Recurrent Units (GRU)

$$
\begin{aligned}
z_t &= \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \\
r_t &= \sigma_g(W_r x_t + U_r h_{t-1} + b_r) \\
\hat{h}_t &= \phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t
\end{aligned}
$$



2 multiplicative gates

Light Gated Recurrent Units (Li-GRU)

$$
\begin{aligned}
z_t &= \sigma(BN(W_z x_t) + U_z h_{t-1}) \\
\widetilde{h_t} &= \mathrm{ReLU}(BN(W_h x_t) + U_h h_{t-1}) \\
h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \widetilde{h_t}
\end{aligned}
$$

If the update gate z is 1, we can remember forever the state h (thus learning arbitrary **long-term dependencies**)

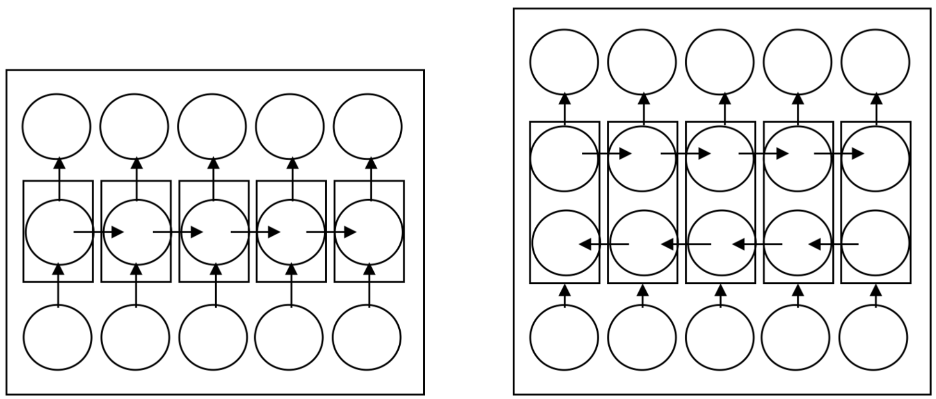
1 multiplicative gate

Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Bidirectional RNN

- In some cases, we want to make a prediction at each time step based on the **whole input elements** and not only the **previous ones**.



(a) (b)

Structure overview

(a) unidirectional RNN
(b) bidirectional RNN

We can employ two RNNs running in **opposite directions**.

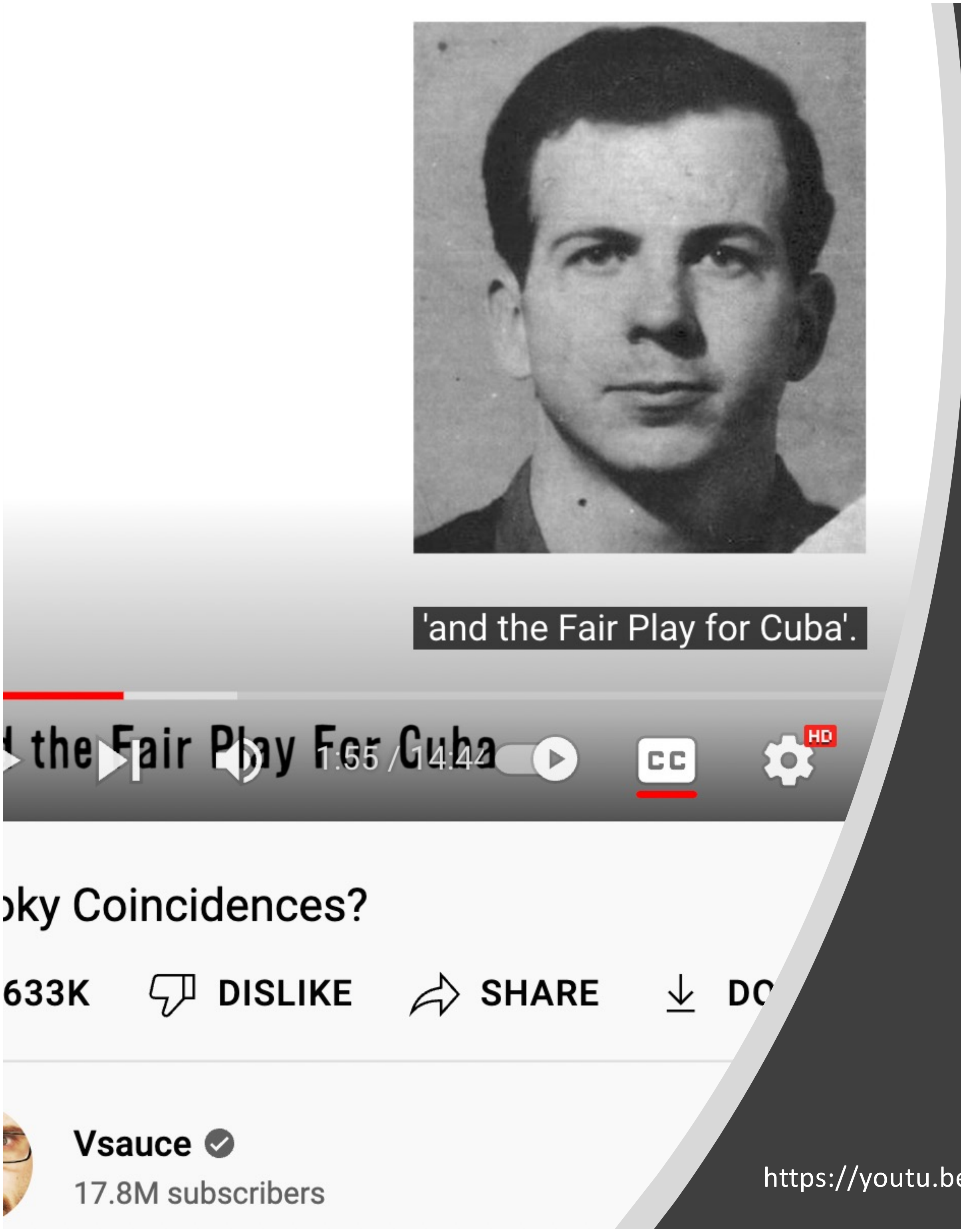
These RNNs use different parameters or share them.

At each layer, we can **combine the forward and the backward state** by concatenating them or summing them up.

Examples of applications are *speech recognition*, *machine translation*, *handwritten recognition*.

Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory
Est. 2014
Transforming Healthcare



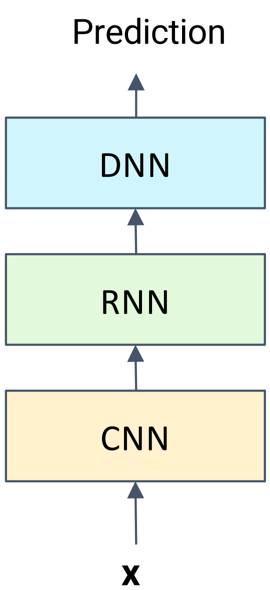
Spooky Coincidences?

https://youtu.be/sHCHEykUxP4?t=115

IMPACT Laboratory
Est. 2014
Transforming Healthcare

CNNs + RNNs + MLPs

- We can make RNNs even deeper by **stacking multiple RNN layers**.
- We can also combine CNNs, RNNs, and DNNs (MLPs):



This model is called CRDNN and is very powerful:

- It first learns **local contexts** with a **CNN**.
- It then captures **long-term dependencies** with an **RNN**.
- Finally, it performs a **final classification** with an MLP.

All the blocks are **jointly trained**.

- This is one of the models used in SpeechBrain (https://speechbrain.github.io/) to process speech signals.

Courtesy of Prof. Mirco Ravanelli

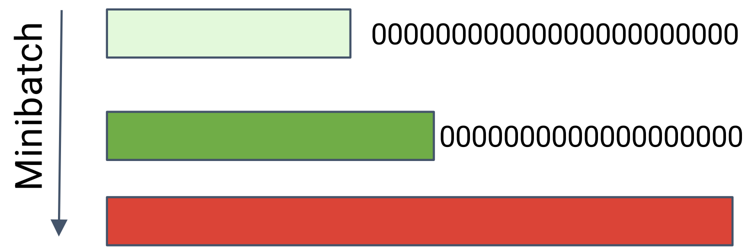
Variable-Length Sequences

- Often, the sequences in input to the RNN have **different lengths**.

- For instance, think about a speech signal: in some cases, we might have long recordings, in some others short ones.

- *How can we manage variable-length sequences?*

We can handle it by **zero-padding:** within each minibatch, we pad with zeros the inputs to match the length of the longest one **.**

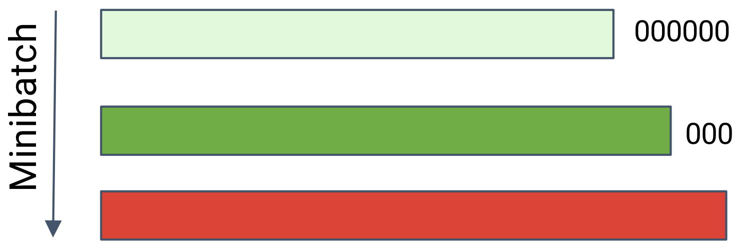


If the length of the inputs is very different, this is **computationally inefficient** because we waste time processing zeros.

Courtesy of Prof. Mirco Ravanelli

Variable-Length Sequences

- One way to mitigate this issue is to **sort** the inputs by length (in ascending or descending order) before creating the minibatches.



This minimizes the need for zero-padding, but **sacrifices randomness** in the minibatch creation:

since we sorted the data, minibatches composed of the same data are shown across epochs.

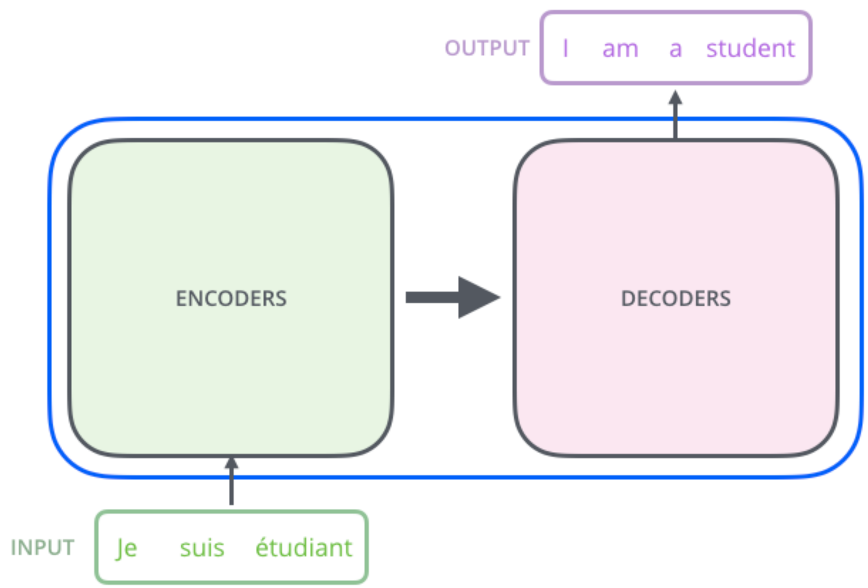
- A compromise solution can be implemented through **bucketing**:
- We split data in buckets such that each bucket contains **inputs** of **similar lengths**.
- When creating minibatches, we sample random inputs from the **same bucket**.

Courtesy of Prof. Mirco Ravanelli

IMPACT Laboratory Est. 2014 Transforming Healthcare

Transformers

A sentence input in one language, and outputs its translation in another.

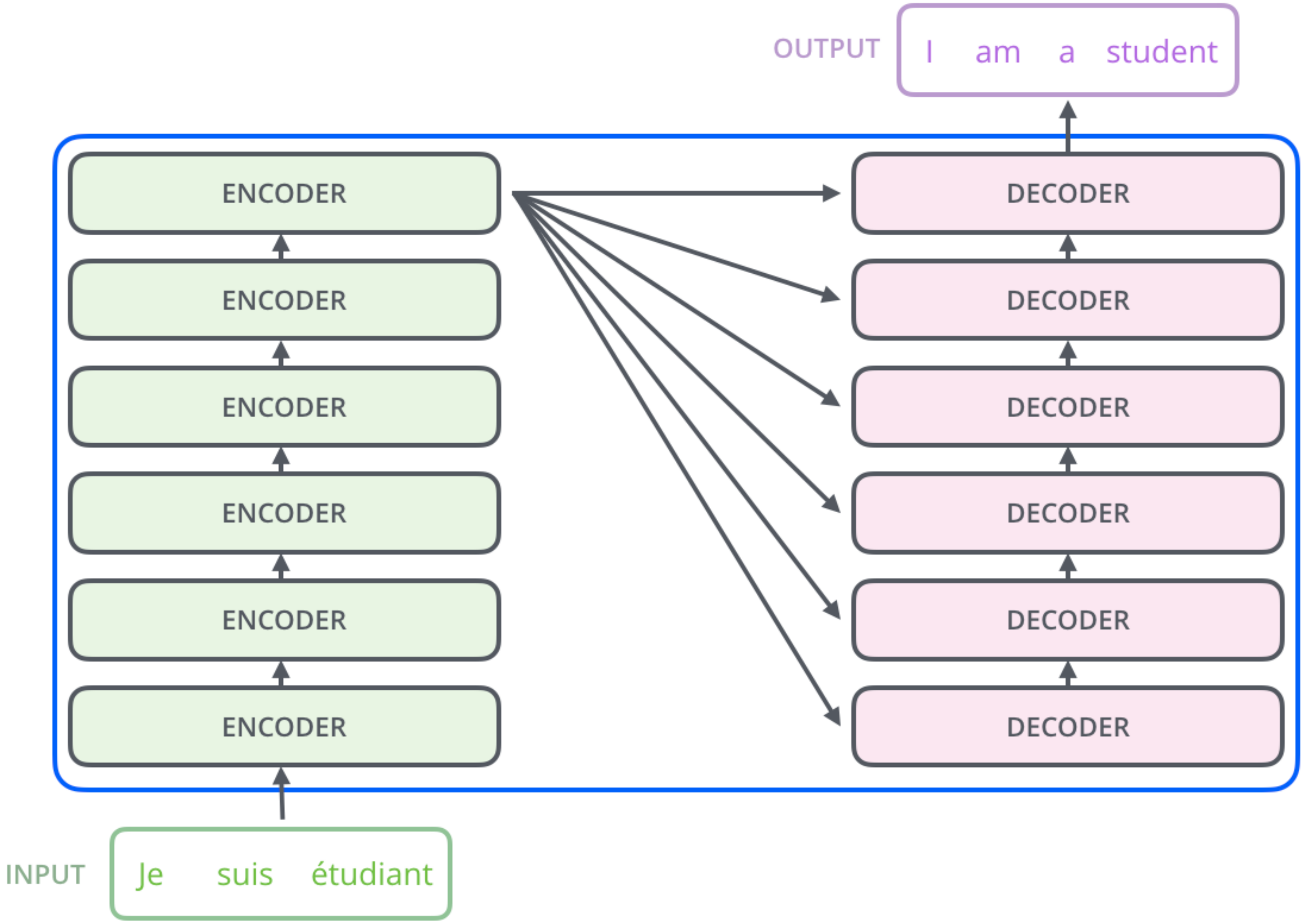


From: https://jalammar.github.io/illustrated-transformer/

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Transformers

Here we are using 6 encoders and 6 decoders



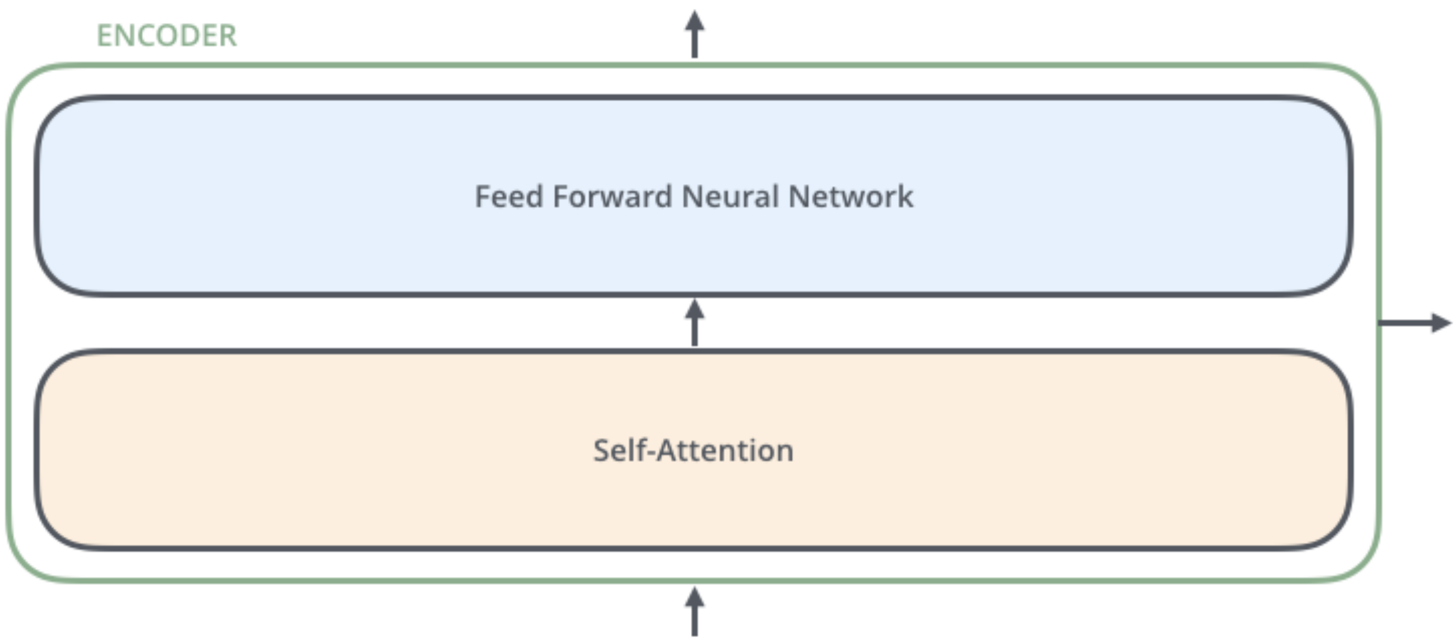
From: https://jalammar.github.io/illustrated-transformer/

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Encoders

The encoders are all identical in structure (yet do not share weights).

Each encoder is broken down into two sub-layers:

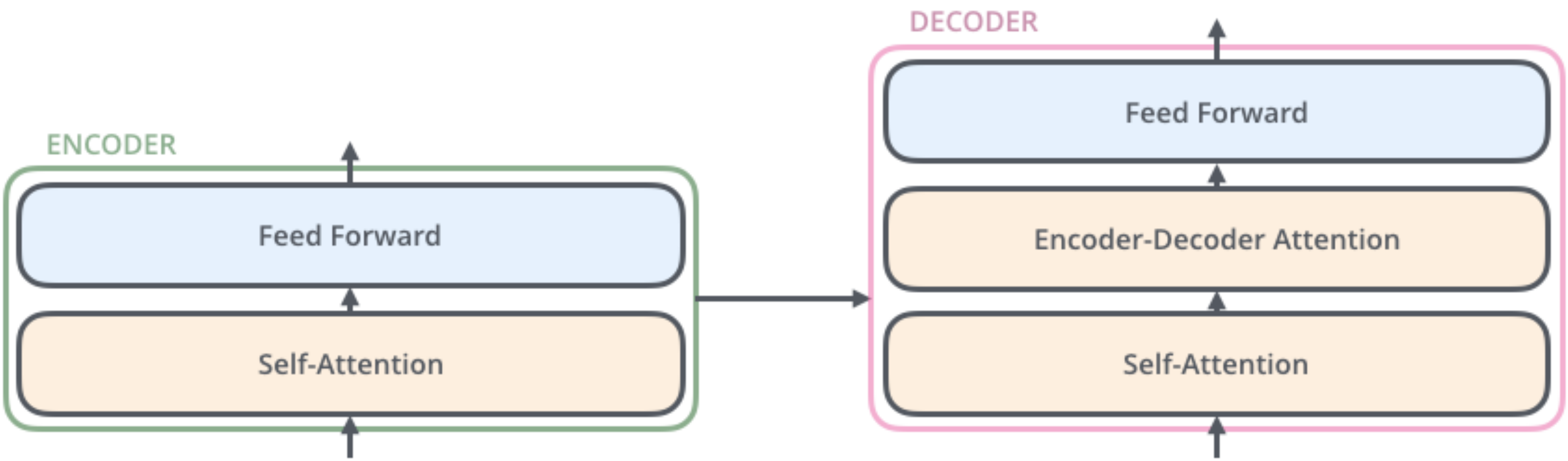


From: https://jalammar.github.io/illustrated-transformer/

ACT Laboratory
Est. 2014
Transforming Healthcare

Decoders

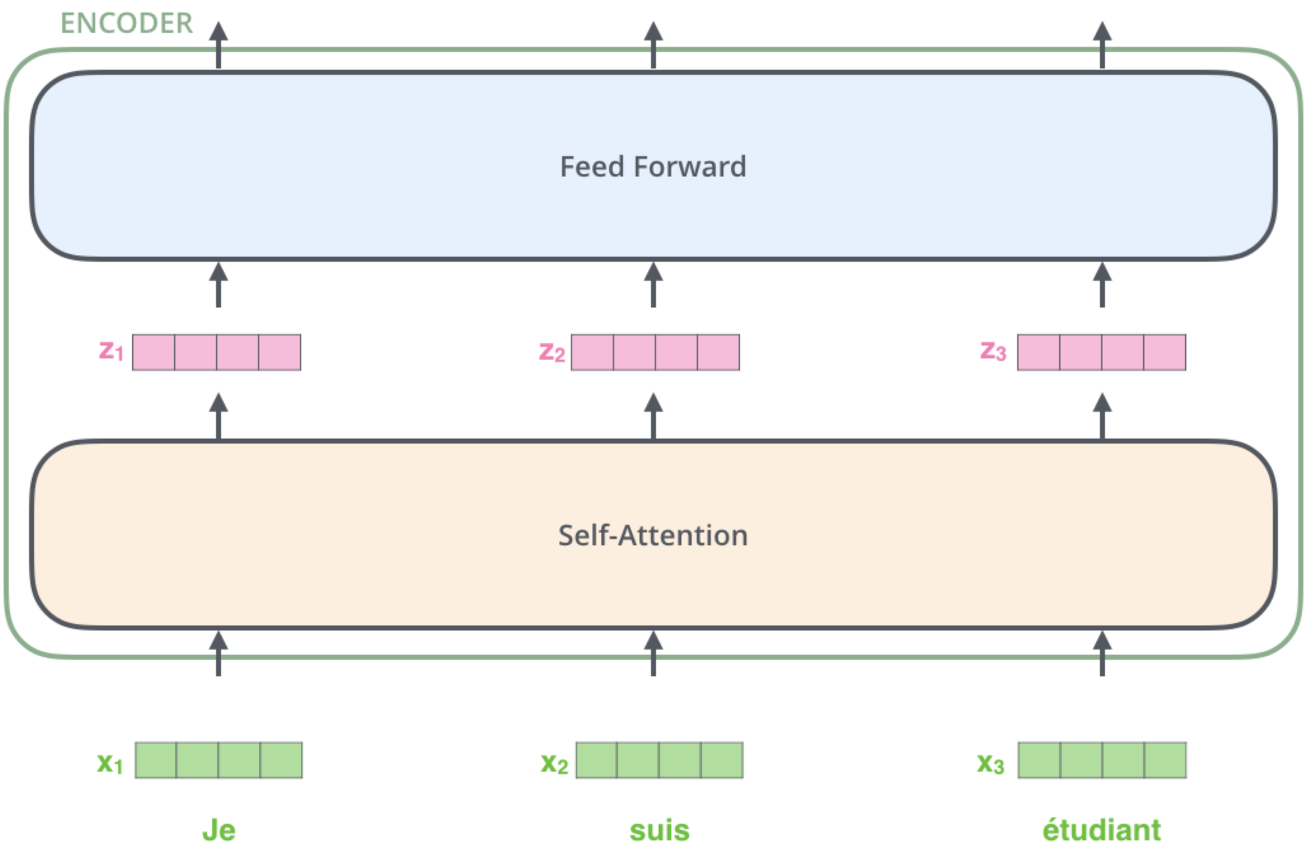
The decoder has both those layers, but between them is an attention layer that helps the decoder focus on relevant parts of the input sentence:



From: https://jalammar.github.io/illustrated-transformer/

The encoder

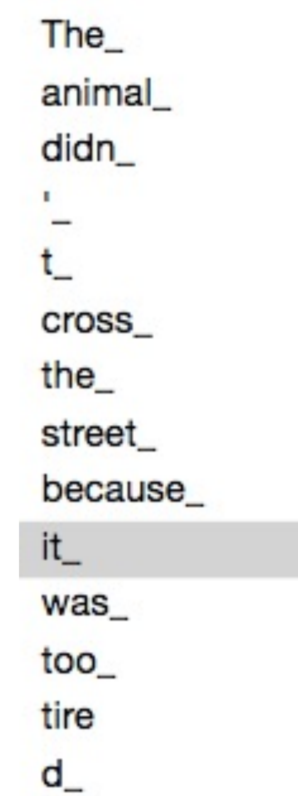
- A key property of the Transformer is that the word in each position flows through its own path in the encoder.
- There are dependencies between these paths in the self-attention layer.
- The feed-forward layer does not have those dependencies, and thus various paths can be executed in parallel while flowing through the feed-forward layer



From: https://jalammar.github.io/illustrated-transformer/

Self-Attention

- What does the word it refer to?
- As the model processes each word (each position in the input sequence), self attention allows it to look at other positions in the input sequence for clues that can help lead to a better encoding for this word.



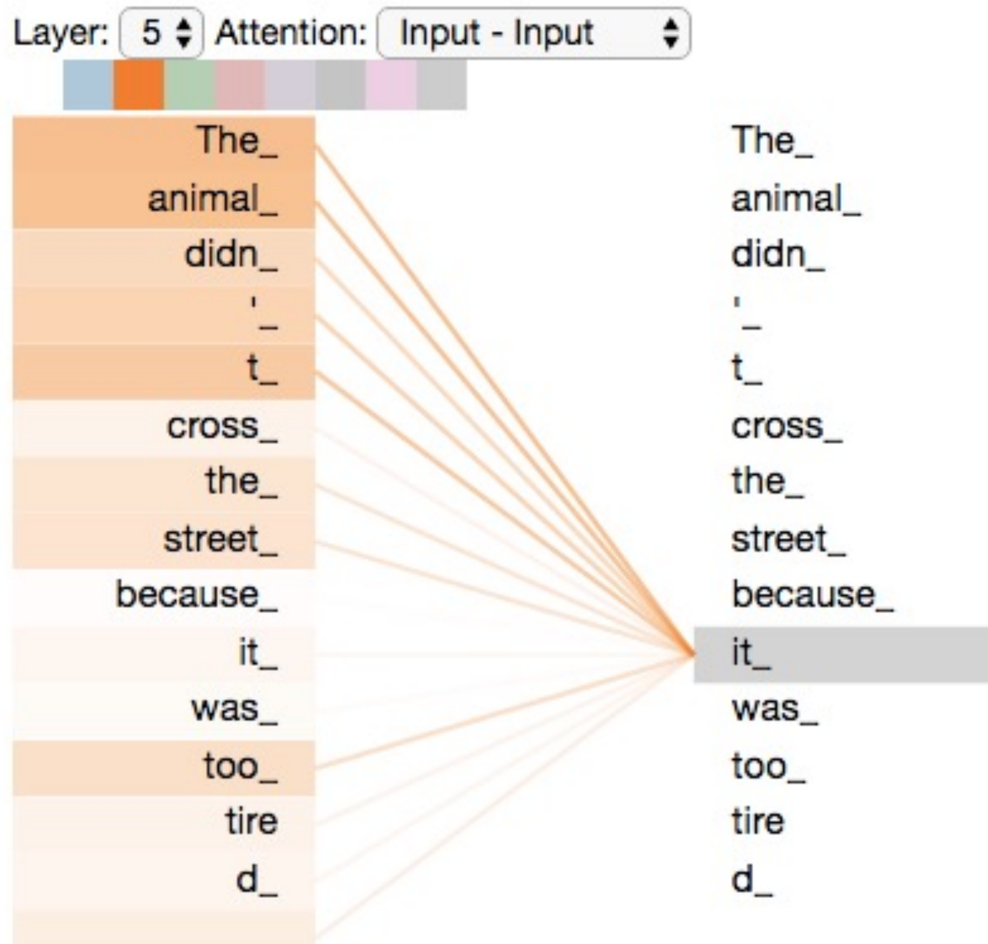
Notebook to visually inspect a Transformer model:
https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello_t2t.ipynb

From: https://jalammar.github.io/illustrated-transformer/

IMPACT Laboratory Est. 2014 Transforming Healthcare

Self-Attention

- What does the word it refer to?
- As the model processes each word (each position in the input sequence), self attention allows it to look at other positions in the input sequence for clues that can help lead to a better encoding for this word.



Notebook to visually inspect a Transformer model:
https://colab.research.google.com/github/tensorflow/tensor2tensor/blob/master/tensor2tensor/notebooks/hello_t2t.ipynb

From: https://jalammar.github.io/illustrated-transformer/

ViT

An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale

ICLR 2021

Alexey Dosovitskiy[*,†], **Lucas Beyer**[*], **Alexander Kolesnikov**[*], **Dirk Weissenborn**[*], **Xiaohua Zhai**[*], **Thomas Unterthiner**, **Mostafa Dehghani**, **Matthias Minderer**, **Georg Heigold**, **Sylvain Gelly**, **Jakob Uszkoreit**, **Neil Houlsby**[*,†]

[*]equal technical contribution, [†]equal advising

Google Research, Brain Team

- In computer vision, convolutional architectures remain dominant
- A pure transformer applied directly to sequences of image patches can perform very well on image classification tasks.
- When pre-trained on large amounts of data and transferred to multiple mid-sized or small image recognition benchmarks (ImageNet, CIFAR, etc.), Vision Transformer (ViT) attains excellent results compared to SOTA CNNs while requiring substantially fewer computational resources to train

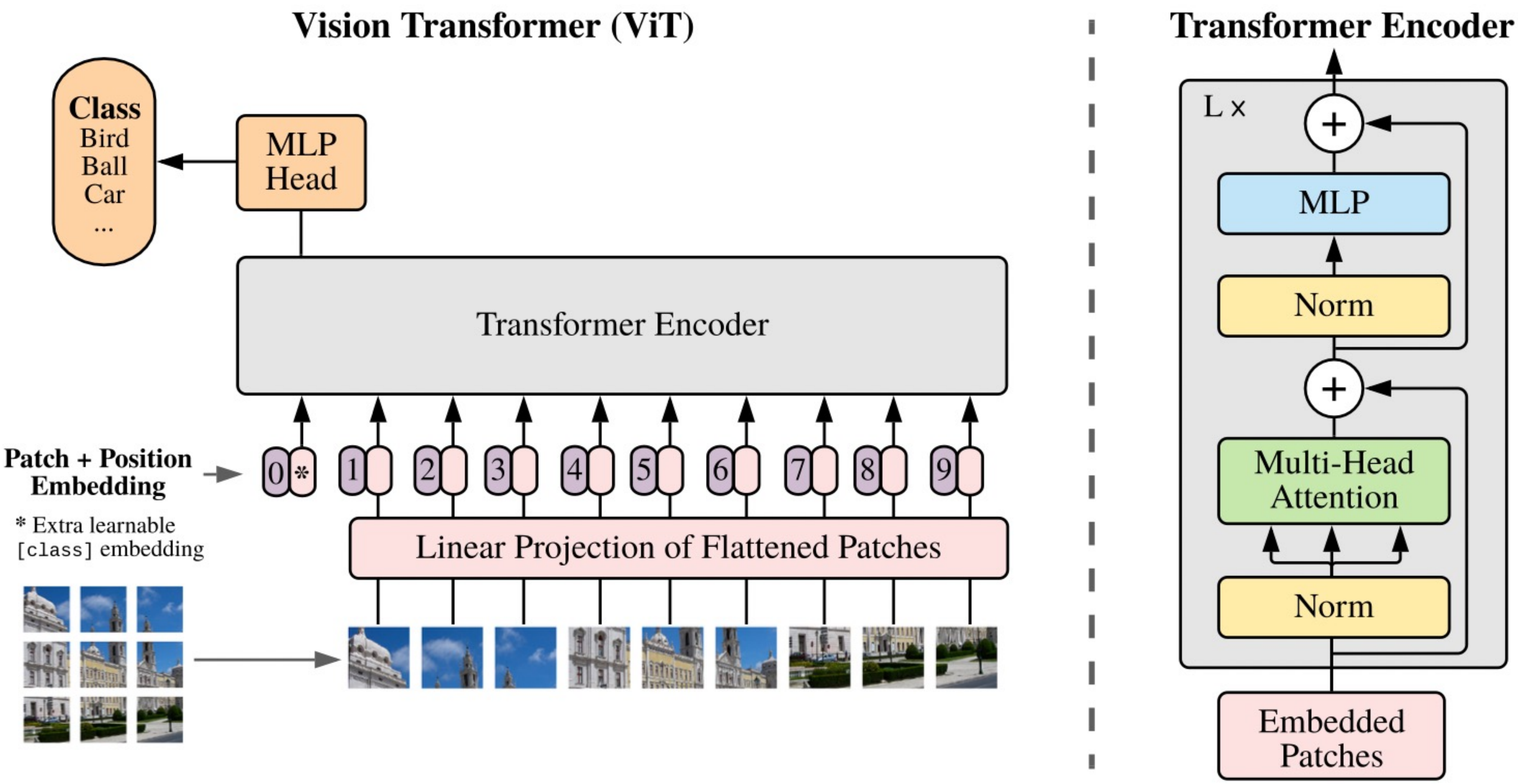
IMPACT Laboratory
Est. 2014
Transforming Healthcare

ViT

- Inspired by the Transformer scaling successes in NLP, apply a standard Transformer to images
- Split an image into patches and provide the sequence of linear embeddings of these patches as an input to a Transformer
- Image patches are treated same as tokens (words) in NLP
- Supervised training of model on image classification

ViT

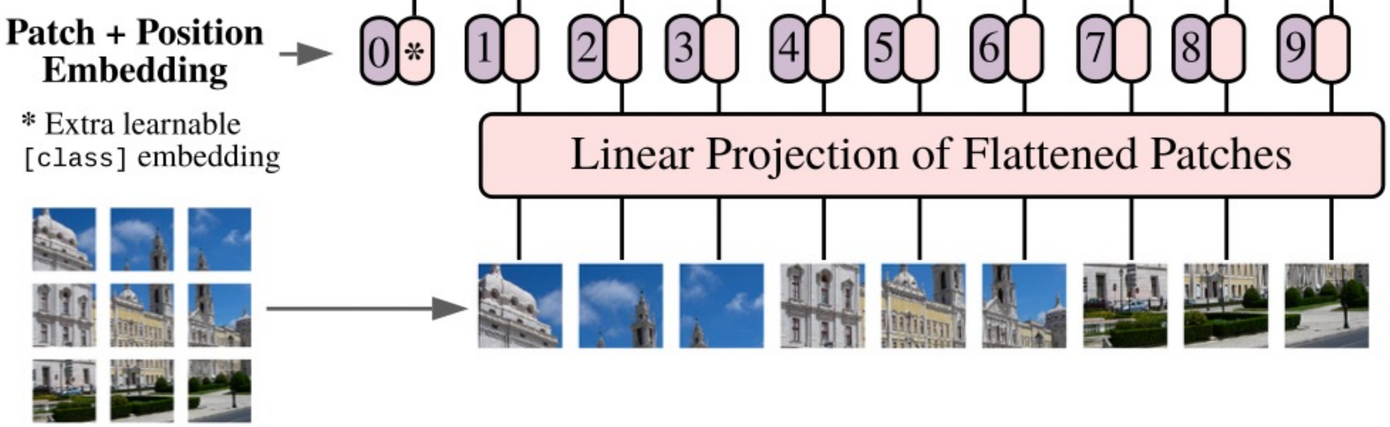
- When trained on mid-sized datasets (eg ImageNet) without strong regularization, these models yield modest accuracies (few points below ResNets)
  - Transformers lack some of the inductive biases of CNNs (translation equivariance and locality)
- The picture changes if the models are trained on larger datasets (14M-300M images).
- **Large scale training trumps inductive bias.**



- Flatten the patches and map to D dimensions with a trainable linear projection
- Position embeddings are added to the patch embeddings to retain positional information (the numbers 1 to 9 in the above example). We use standard 1D position embeddings, since we have not observed significant performance gains from using more advanced 2D-aware position embeddings

Linear projection in ViT

- Naive application of self-attention to images would require that each pixel attends to every other pixel.
  - Computationally expensive
- Instead, apply self-attention to N patches (N=9 here):



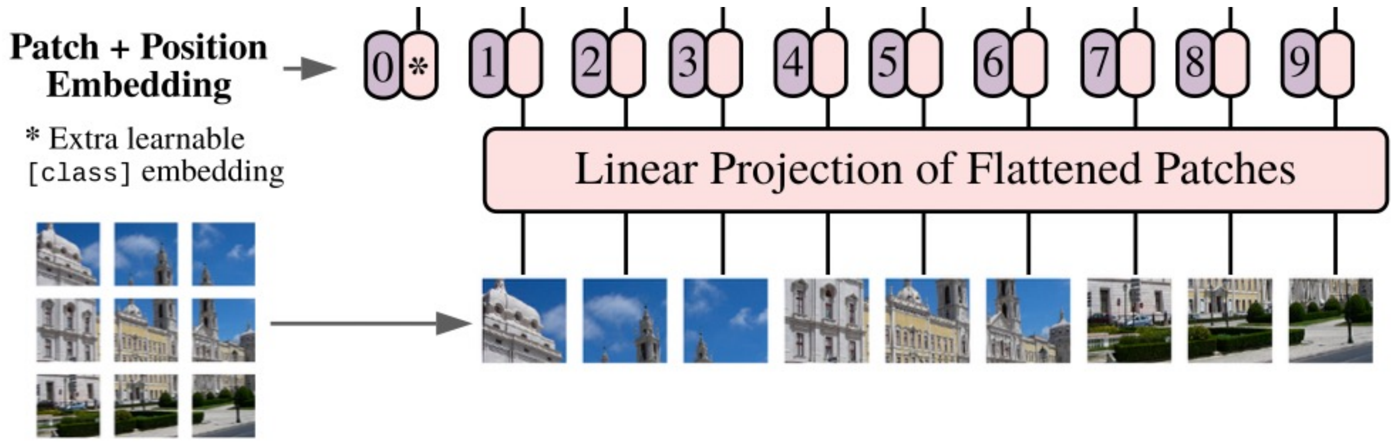
- Reshape image x to flattened 2D patches $\mathbf{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$

C: number of channels, P^2: size of each patch, $N=HW/P^2$: number of patches

$$\mathbf{z}_0 = [\mathbf{x}_{\text{class}};\, \mathbf{x}_p^1\mathbf{E};\, \mathbf{x}_p^2\mathbf{E};\cdots;\, \mathbf{x}_p^N\mathbf{E}] + \mathbf{E}_{pos}, \qquad \mathbf{E} \in \mathbb{R}^{(P^2 \cdot C)\times D},\ \mathbf{E}_{pos} \in \mathbb{R}^{(N+1)\times D}$$

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Hybrid ViT

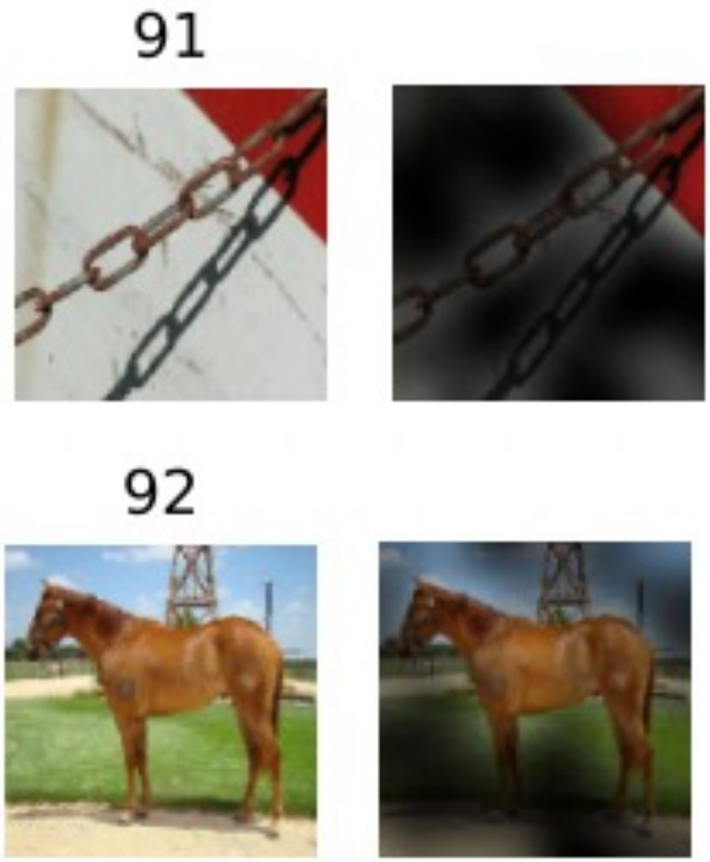


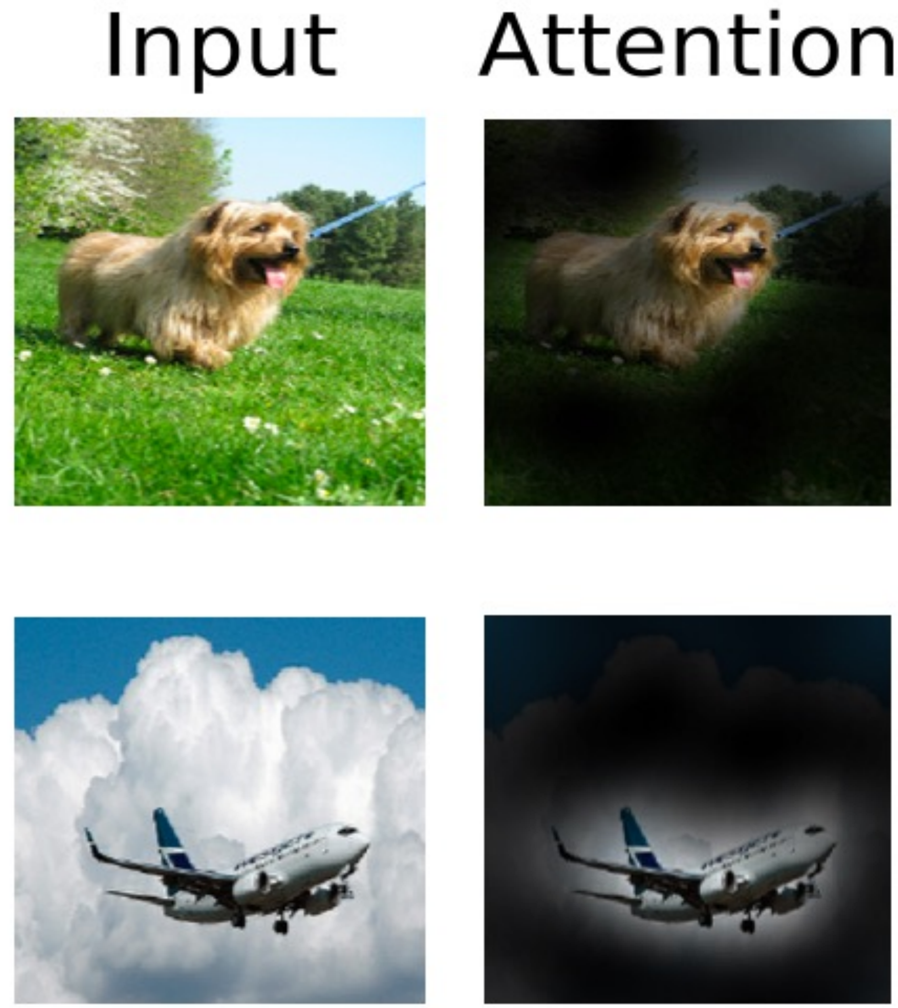
- In the hybrid model, the patch embedding projection E is applied to patches extracted from a CNN feature map.

$$\mathbf{z}_0 = [\mathbf{x}_{\text{class}};\, \mathbf{x}_p^1\mathbf{E};\, \mathbf{x}_p^2\mathbf{E};\cdots;\, \mathbf{x}_p^N\mathbf{E}] + \mathbf{E}_{pos}, \qquad \mathbf{E} \in \mathbb{R}^{(P^2\cdot C)\times D},\, \mathbf{E}_{pos} \in \mathbb{R}^{(N+1)\times D}$$

ViT Attention Maps

- Attention from the output token to the input space:





IMPACT Laboratory
Est. 2014
Transforming Healthcare

Masked Autoencoders Are Scalable Vision Learners

CVPR 2022

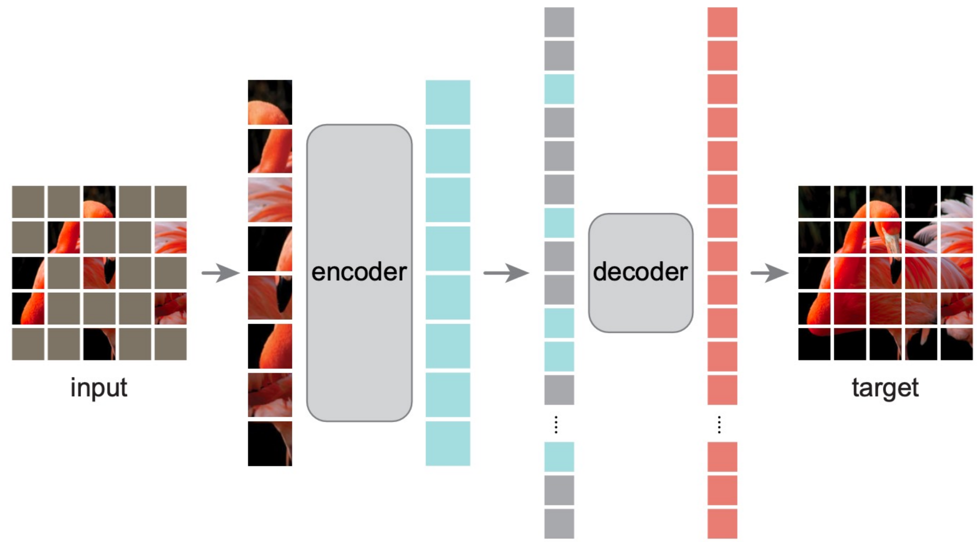
Kaiming He*,† Xinlei Chen* Saining Xie Yanghao Li Piotr Dollár Ross Girshick

*equal technical contribution †project lead

Facebook AI Research (FAIR)

Masked autoencoders (MAE) are scalable self-supervised learners.

1. asymmetric encoder-decoder:
    1. encoder that operates only on the visible subset of patches
    2. lightweight decoder reconstructs the image from the latent representation and mask tokens



IMPACT Laboratory
Est. 2014
Transforming Healthcare

Masked Autoencoders Are Scalable Vision Learners

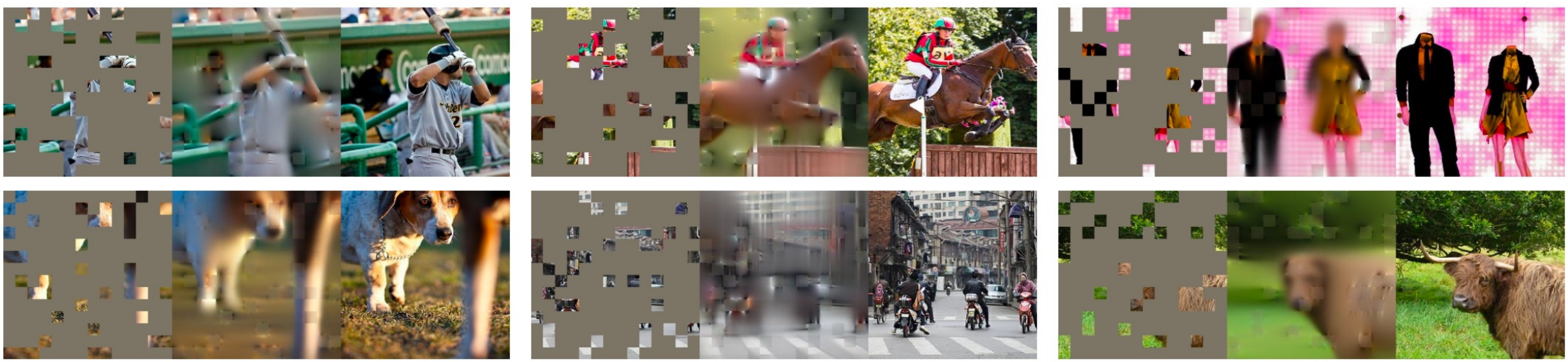
CVPR 2022

Kaiming He*,† Xinlei Chen* Saining Xie Yanghao Li Piotr Dollár Ross Girshick

*equal technical contribution †project lead

Facebook AI Research (FAIR)

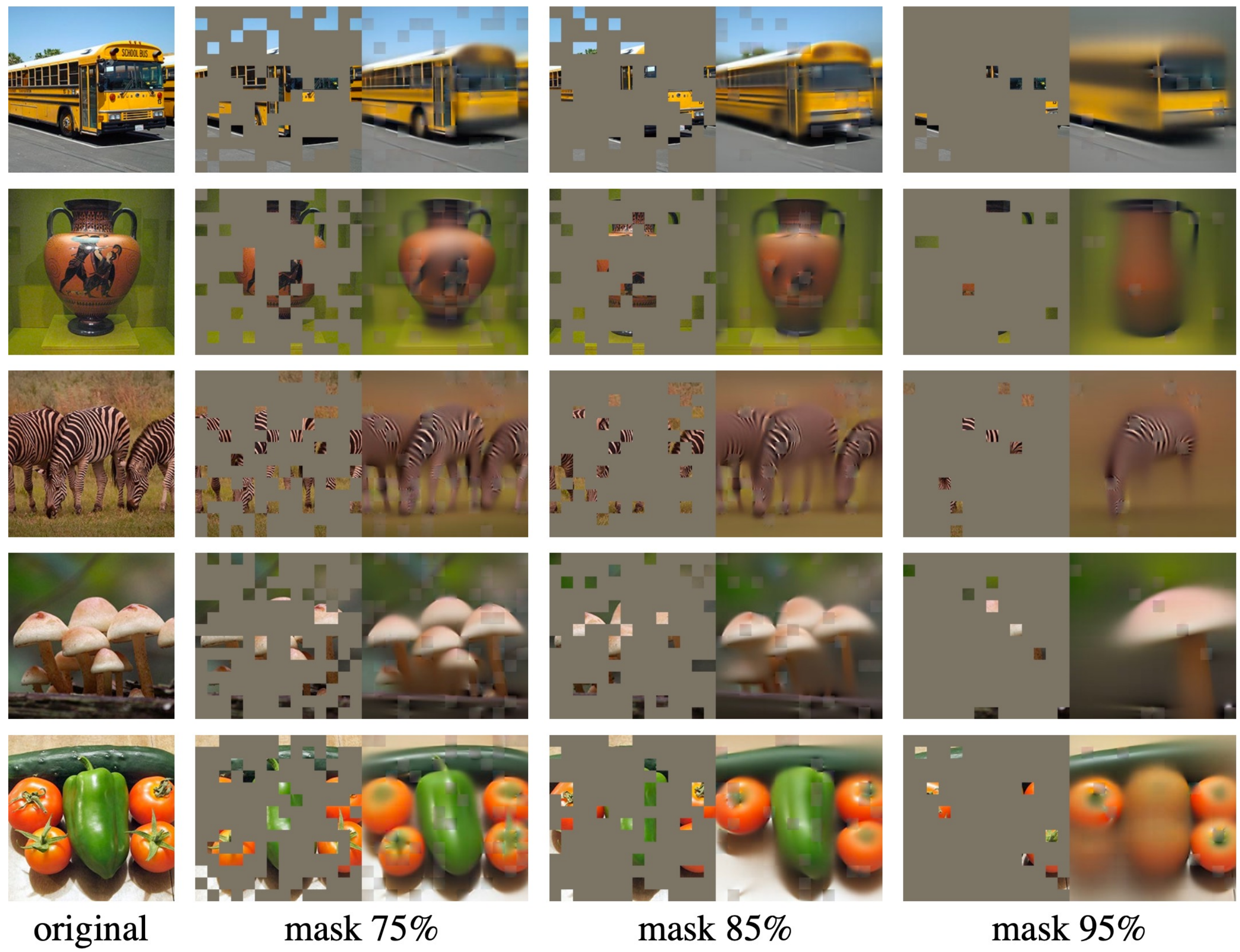
- Images have heavy spatial redundancy
  - missing patch can be recovered from neighboring patches with little high-level understanding
- Mask very high portion of random patches



IMPACT Laboratory
Est. 2014
Transforming Healthcare

Masked Autoencoder

- MAE pre-trained with a masking ratio of 75% but applied on inputs with higher masking ratios.
- The predictions differ plausibly from the original images, showing that the method can generalize.



IMPACT Laboratory
Est. 2014
Transforming Healthcare

Recovering masked regions

AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

ICLR 2021

Alexey Dosovitskiy[*,†], **Lucas Beyer**[*], **Alexander Kolesnikov**[*], **Dirk Weissenborn**[*],
Xiaohua Zhai[*], **Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,**
Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby[*,†]
[*]equal technical contribution, [†]equal advising
Google Research, Brain Team
{adosovitskiy, neilhoulsby}@google.com

Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion

JML 2010

Pascal Vincent PASCAL.VINCENT@UMONTREAL.CA
Département d'informatique et de recherche opérationnelle
Université de Montréal
2920, chemin de la Tour
Montréal, Québec, H3T 1J8, Canada

Hugo Larochelle LAROCHEH@CS.TORONTO.EDU
Department of Computer Science
University of Toronto
10 King's College Road
Toronto, Ontario, M5S 3G4, Canada

Isabelle Lajoie ISABELLE.LAJOIE.1@UMONTREAL.CA
Yoshua Bengio YOSHUA.BENGIO@UMONTREAL.CA
Pierre-Antoine Manzagol PIERRE-ANTOINE.MANZAGOL@UMONTREAL.CA

IMPACT Laboratory
Est. 2014
Transforming Healthcare

44th Annual International Conference of
the IEEE Engineering in Medicine and Biology Society
11-15 July 2022 Glasgow, Scotland, UK
IEEE EMB

Vision Transformers for Classification of Breast Ultrasound Images

B. Gheflati, H. Rivaz

Department of Electrical and Computer Engineering,
Concordia University, Montreal, Canada.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

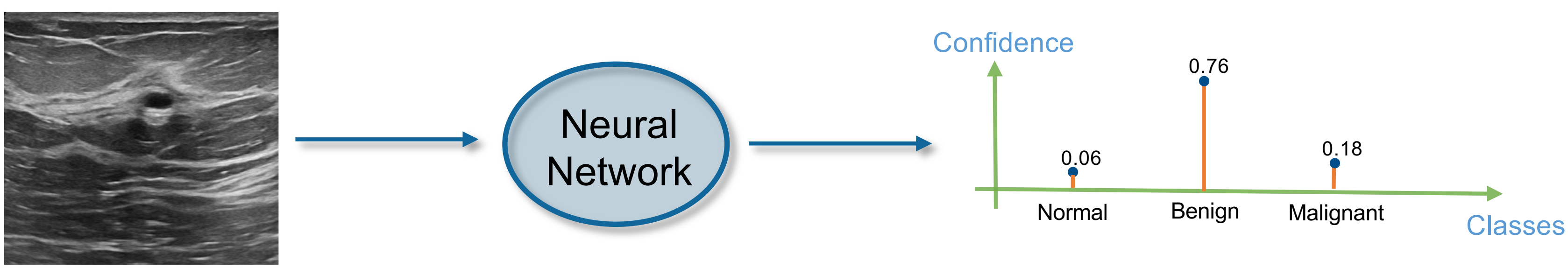
I. Introduction

- **Problem Statement**

✓ **Breast cancer** is the most frequent cause of cancer mortality among women.

✓ Medical **ultrasound (US)** imaging
  - ✓ Availability
  - ✓ Real-time display
  - ✓ Cost-efficiency
  - ✓ Non-invasive nature

✓ Medical images classification



IMPACT Laboratory
Est. 2014
Transforming Healthcare

II. Related Work

✓ A common approach in medical images Classification → **Convolutional Neural Networks (CNNs)**

- ResNet
- VGG
- AlexNet
- GoogLeNet
- NASNET

Pre-trained CNN models → Transfer Learning (TL) → transferring the pre-trained models based on a specific task

References	Pre-trained models	Application
Lazo *et al.* [1]	VGG-16 IncpectionV3	Lesion detection in breast US images.
Al-Dhabyani *et al.* [2]	VGG16, ResNet, Inception, and NASNet	Breast masses classification

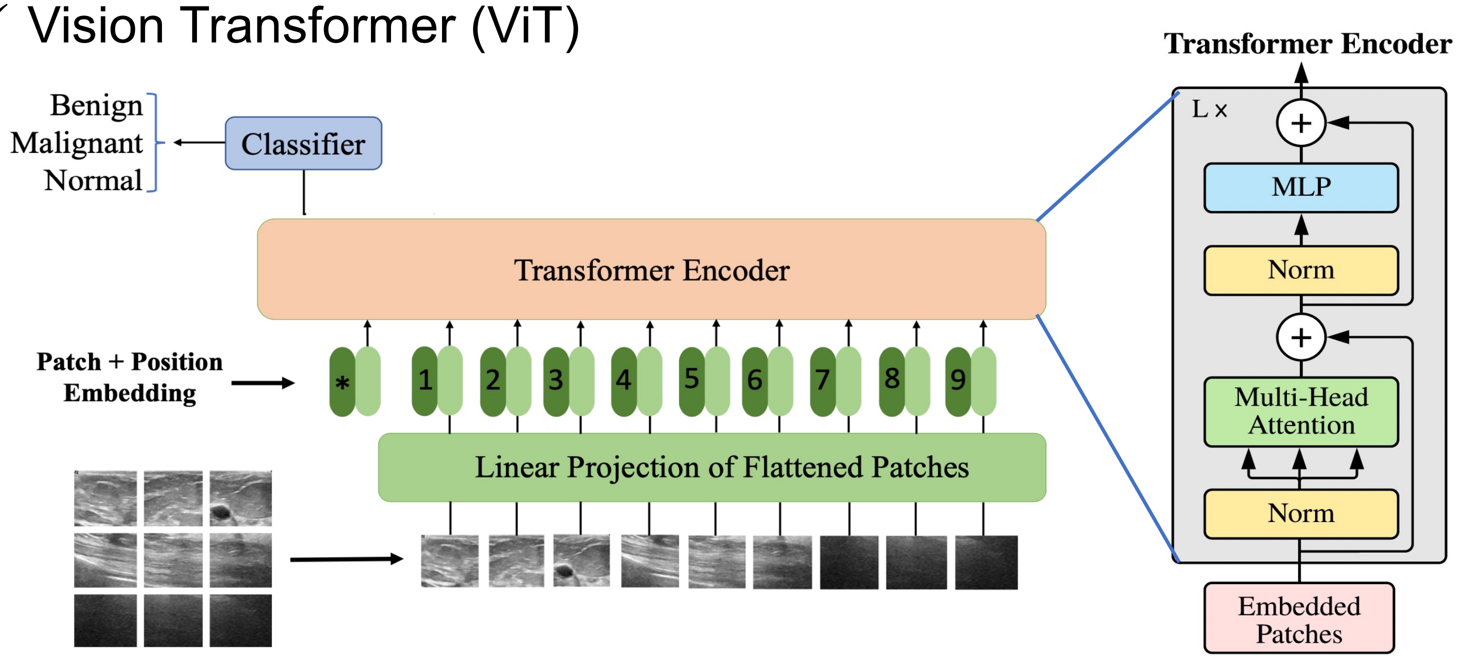
IMPACT Laboratory
Est. 2014
Transforming Healthcare

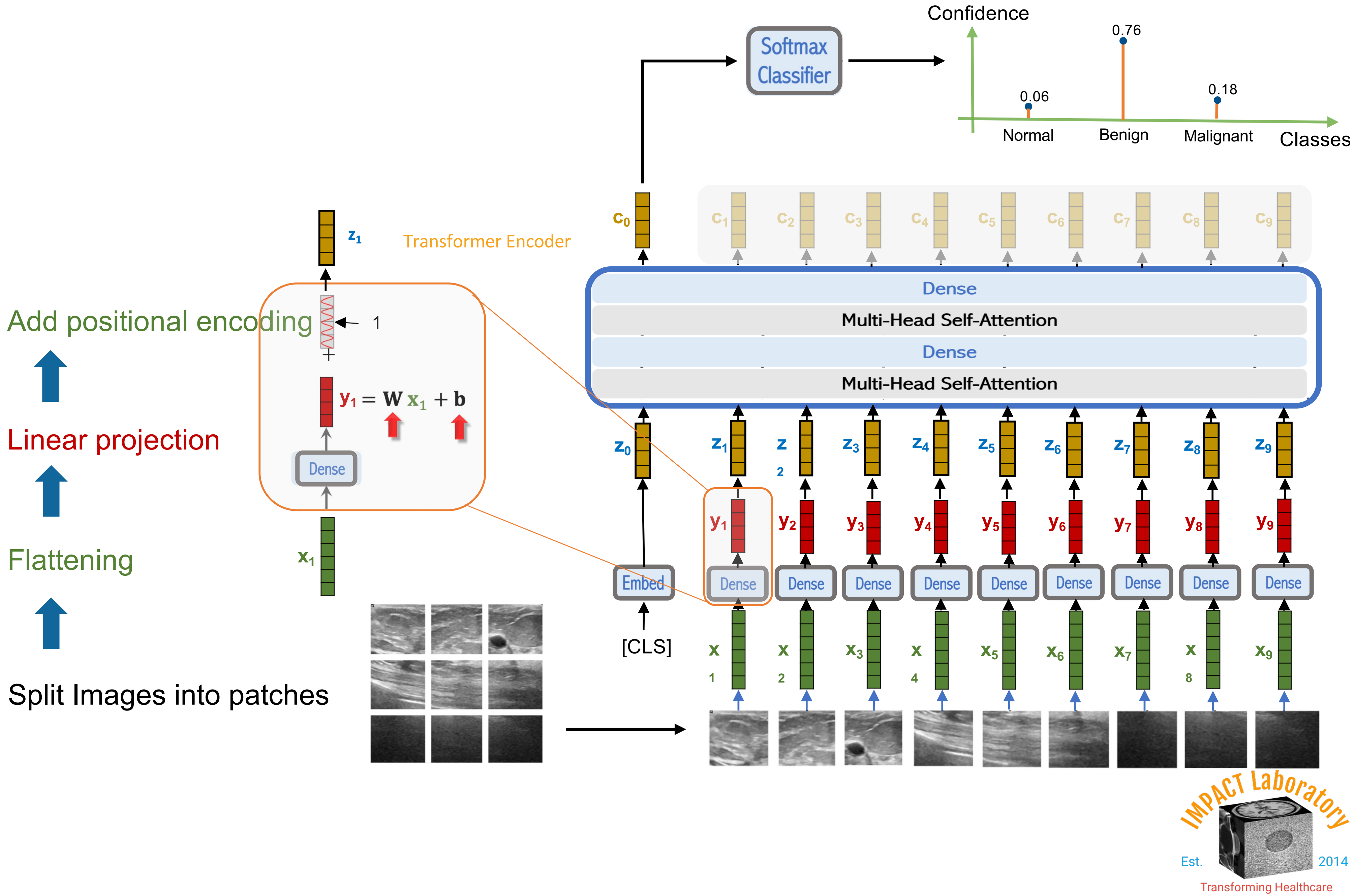
II. Related Work

- **Motivation**

  Recently, Vision Transformer (ViT)[3] designs have shown great potential to be an alternative to CNNs.

✓ Vision Transformer (ViT)





Confidence
Softmax Classifier
0.76
0.06
0.18
Normal
Benign
Malignant
Classes
Transformer Encoder
Dense
Multi-Head Self-Attention
Dense
Multi-Head Self-Attention
Add positional encoding
Linear projection
Flattening
Split Images into patches
$y_1 = W x_1 + b$
Embed
[CLS]
Dense
IMPACT Laboratory
Est. 2014
Transforming Healthcare

II. Related Work

- ViT applications
  - ✓ Image Classification
  - ✓ Image Segmentation
  - ✓ Object detection

Reference	Application
Karimi *et al.* [4]	3-D Medical Image Segmentation
Hatamizadeh *et al.* [5]	3-D Medical Image Segmentation
Dai *et al.* [6]	Multi-modal images classification
Carion *et al.* [7]	Object detection

IMPACT Laboratory
Est. 2014
Transforming Healthcare

III. Hypotheses

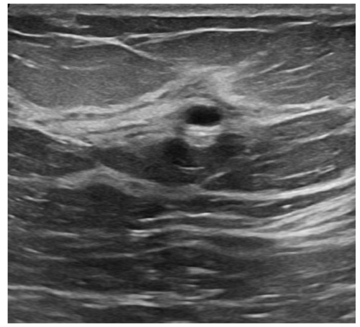
- Goal: The potential application of self-attention models to classify breast US images

- Contributions
  - ✓ Comparing different pre-trained ViT models.
  - ✓ Transferring pre-trained ViT models to the classification of breast US images.
  - ✓ Fine-tuning SOTA CNN networks to classify US images.
  - ✓ Adopting a weighted cross-entropy loss function.

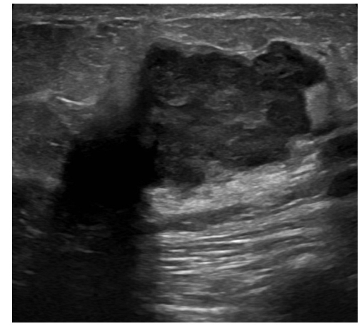
IMPACT Laboratory
Est. 2014
Transforming Healthcare

IV. Method

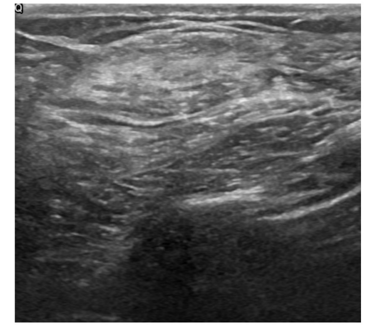
- **Dataset**



Benign



Malignant



Normal

- BUSI : 780 breast US images[8]

133 Normal
437 Malignant
210 Benign

- B : 163 images[9]

110 Benign
53 Malignant

133 Normal
320 Benign
490 Malignant

- **Evaluation Metrics**

Classification Accuracy
Area Under the ROC Curve (AUC)

IMPACT Laboratory
Est. 2014
Transforming Healthcare

IV. Method

- **Models**

  - CNN
    - VGG16
    - ResNet50
    - InceptionV3
    - NASNetLarge
  - ViT
    - ViT models (ViT-Ti, ViT-S, ViT-B)
    - hybrid ViT+ResNet models (R+Ti, R26+S)[10]

- **Fine-tuning details**

  - Splitting all datasets to 70%, 15% and 15% for training, validation and testing.
  - 5-fold cross-validation (CV) for the final evaluation.
  - Weighted cross-entropy loss (imbalanced training dataset).

IMPACT Laboratory
Est. 2014
Transforming Healthcare

IV. Method

A summary of our method is as follows:

- ✓ Classification of BUSI+B dataset using TL based on CNN models.
- ✓ Classification of BUSI+B dataset using TL based on ViT-based networks.
- ✓ Classification of augmented BUSI+B dataset using TL based on ViT-based networks.
- ✓ Classification of three B, BUSI, and BUSI+B datasets using TL based on two ViT-based and CNN networks.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

V. Results

- **Convolutional Neural Network (CNN) Models**

  - Dataset: BUSU + B

Evaluation	ResNet	VGG	Inception	NASNET
ACC	**85.3%**	82%	80%	79%
AUC	**0.94**	0.92	0.92	0.917

  - ResNet50 model has the best results with 85.3% Acc and 0.95 AUC.

- **ViT-based Models**

  - Dataset: BUSU + B

Evaluation	R+Ti/16	S/32	B/32	Ti/16	R26+S/16
ACC	85.7%	86%	**86.7%**	85%	86.4%
AUC	0.94	0.95	**0.95**	0.94	0.95

  - ViT models shows comparable or even better results comparing to the CNN models.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

VI. Conclusion

This study shows the potential of ViT models in US image classification and, therefore, the effectiveness of learning the global anatomical dependencies of US medical images.

VII. Acknowledgment

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), and thank NVIDIA for the donation of the GPU.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

VIII. References

[1] J. F. Lazo, S. Moccia, E. Frontoni, and E. De Momi, "Comparison of different cnns for breast tumor classification from ultrasound images," *arXiv preprint arXiv:2012.14517*, 2020.

[2] W. Al-Dhabyani, M. Gomaa, H. Khaled, and F. Aly, "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images," *Int. J. Adv. Comput. Sci. Appl*, vol. 10, no. 5, pp. 1–11, 2019.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[4] D. Karimi, S. Vasylechko, and A. Gholipour, "Convolution-free medical image segmentation using transformers," *arXiv preprint arXiv:2102.13645*, 2021.

[5] A. Hatamizadeh, D. Yang, H. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," *arXiv preprint arXiv:2103.10504*, 2021.

[6] Y. Dai, Y. Gao, and F. Liu, "Transmed: Transformers advance multi- modal medical image classification," *Diagnostics*, vol. 11, no. 8, p. 1384, 2021.

[7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision*. Springer, 2020, pp. 213– 229.

[8] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in brief*, vol. 28, p. 104863, 2020.

[9] M. H. Yap, G. Pons, J. Mart́ ı, S. Ganau, M. Sent́ ıs, R. Zwiggelaar, A. K. Davison, and R. Marti, "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.

[10] A. Steiner, A. Kolesnikov, X. Zhai, R. Wightman, J. Uszkoreit, and L. Beyer, "How to train your vit? data, augmentation, and regulariza- tion in vision transformers," *arXiv preprint arXiv:2106.10270*, 2021.

Ultrasound Imaging

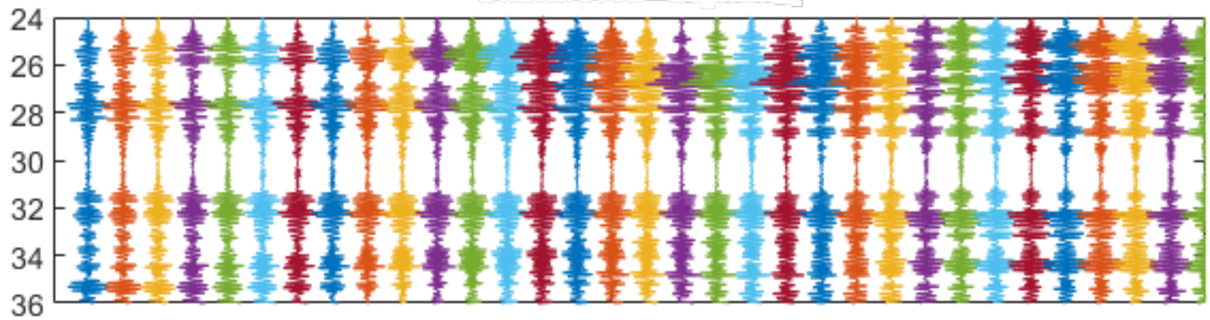
One of the most used imaging modalities

- Real-time
- Inexpensive
- Safe
- Portable
- Big ultrasound manufacturers have closely guarded raw data

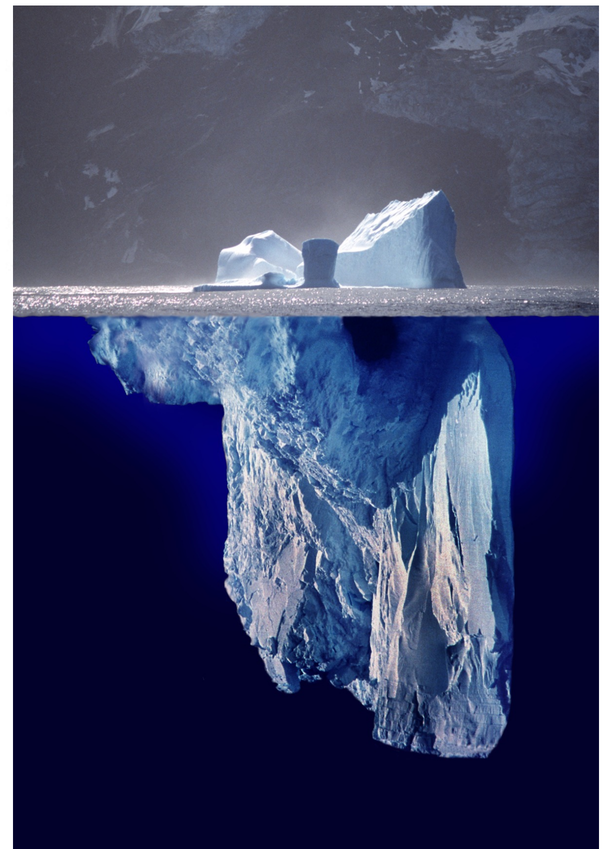
IMPACT Laboratory
Est. 2014
Transforming Healthcare

AI-Driven Ultrasound

- B-mode US (the familiar gray-scale images) only contains a small percentage of all data



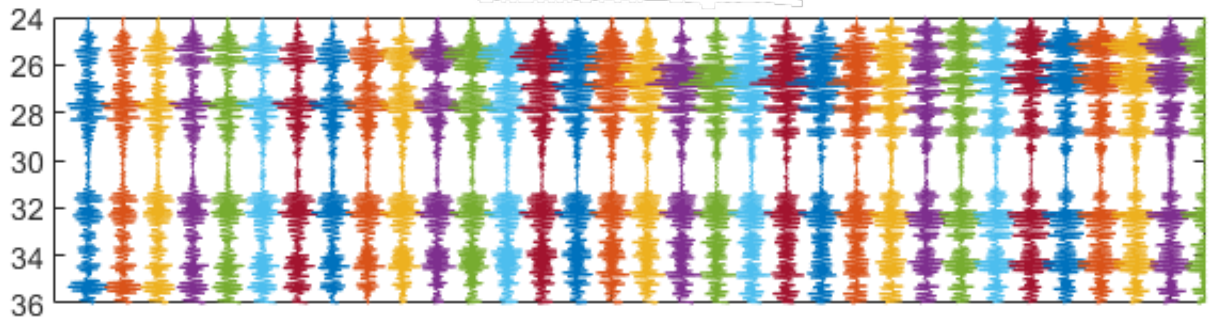
Raw pre-beamformed channel RF data
Visually not informative



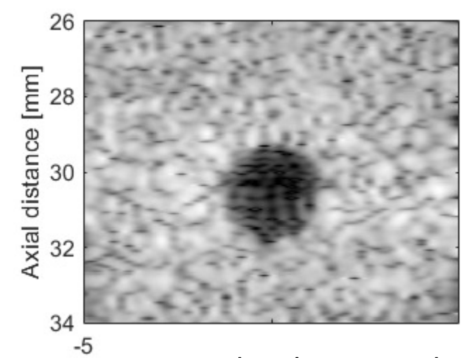
Est. 2014
Transforming Healthcare

AI-Driven Ultrasound

- B-mode US (the familiar gray-scale images) only contains a small percentage of all data

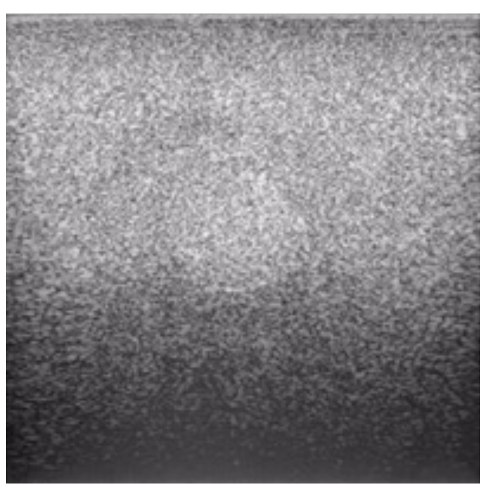


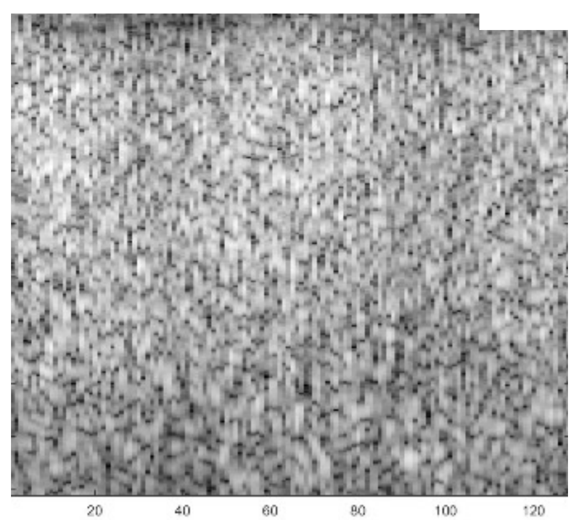
Raw pre-beamformed channel data



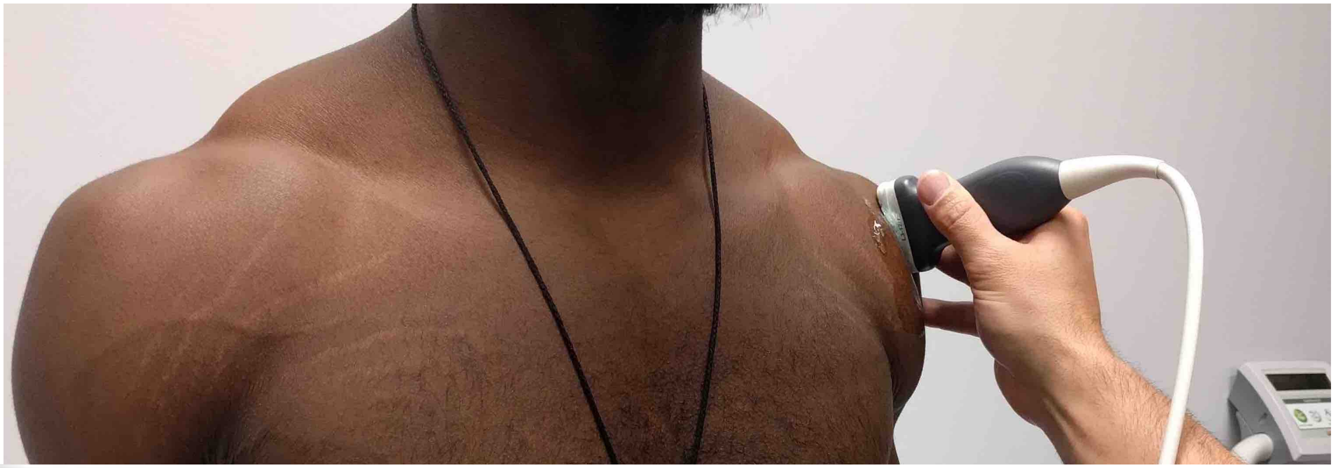
B-mode ultrasound

- US has a very high frame-rate

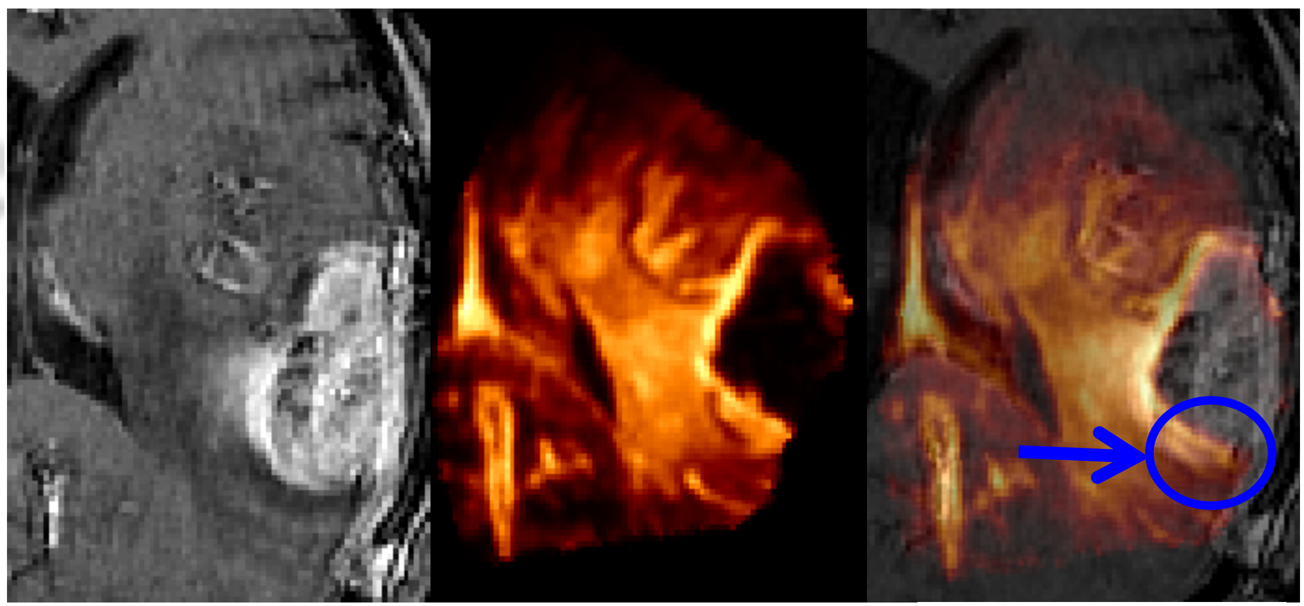




Collection and Interpretation is hard



US and rendered MR



Pre-operative MR

Post-resection US (tumour is resected)

Residual tumour

Rivaz, H, Chen, S, Collins, DL, IEEE TMI, 2015

Est. 2014

Transforming Healthcare

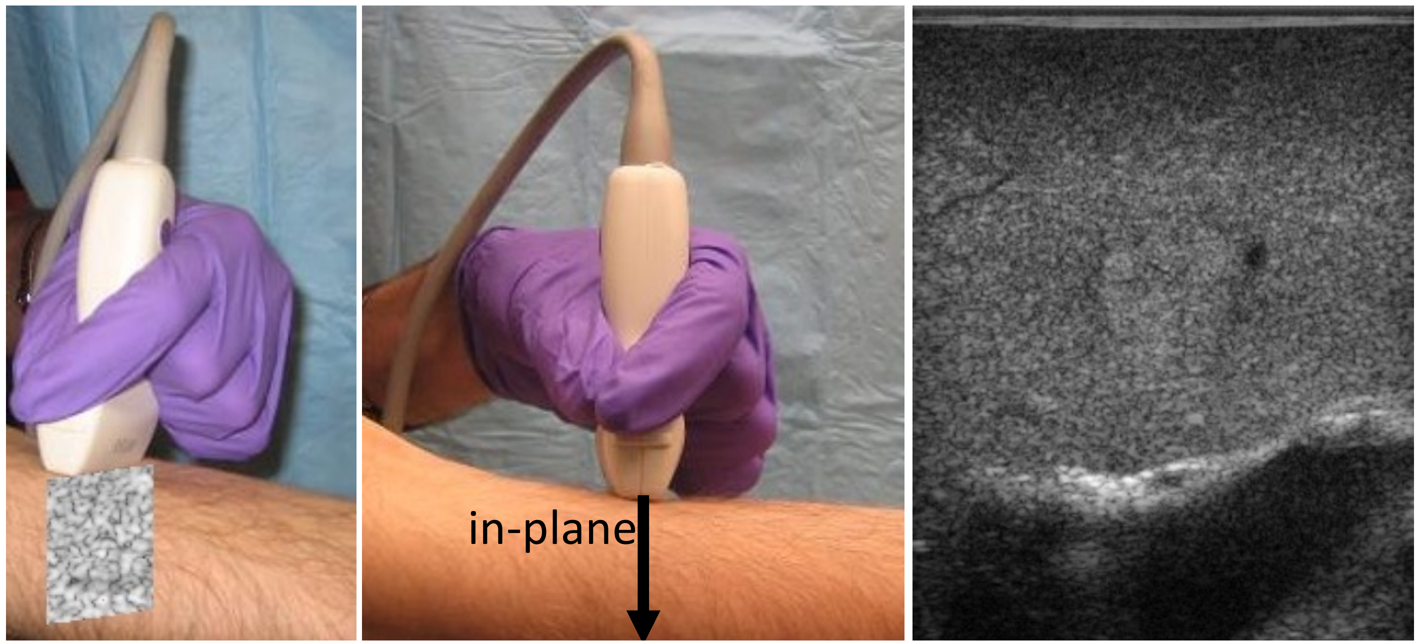
AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction for transcranial imaging & focusing
  - Deep beamforming
- Aid collection & interpretation
  - Image Registration
    - US to MRI
    - US to US
  - Image segmentation
    - Explainable AI (XAI)
  - Image selection

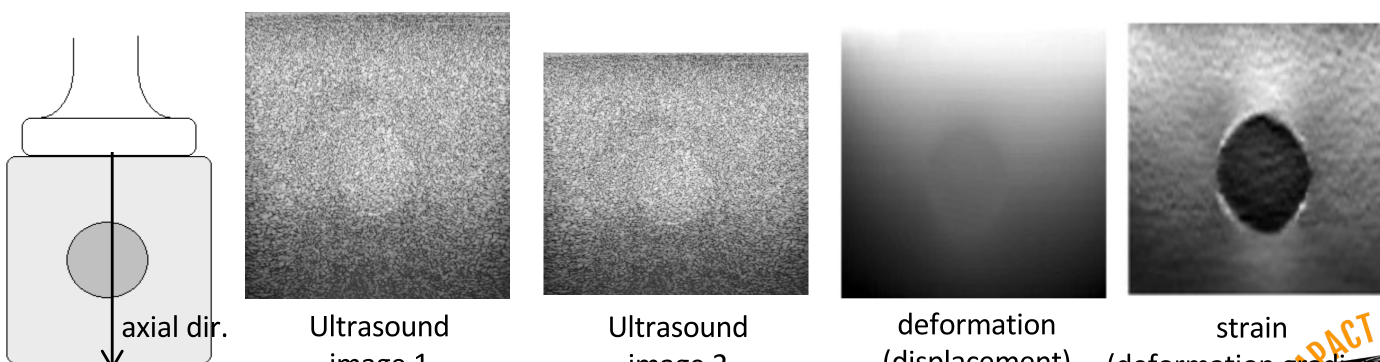
IMPACT Laboratory
Est. 2014
Transforming Healthcare

Ultrasound Elastography

Deformation can reveal mechanical properties of tissue

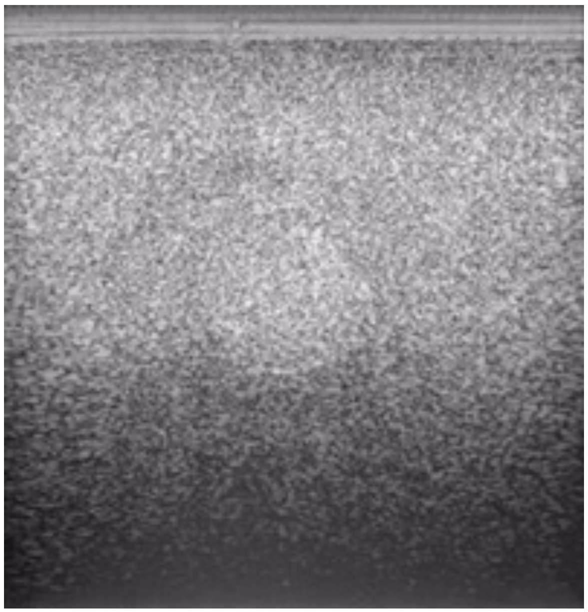


Estimating tissue deformation gives clues to the mechanical properties of tissue

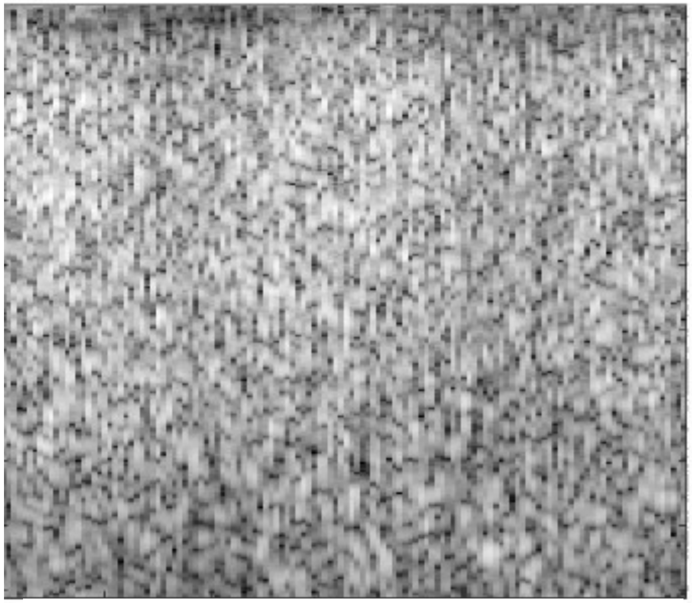


IMPACT Laboratory Est. 2014 Transforming Healthcare

Common Step of Elastography: Displacement Estimation



B-mode images from quasi-static compression
(focused imaging)



B-mode images after acoustic radiation force
(plane-wave imaging)

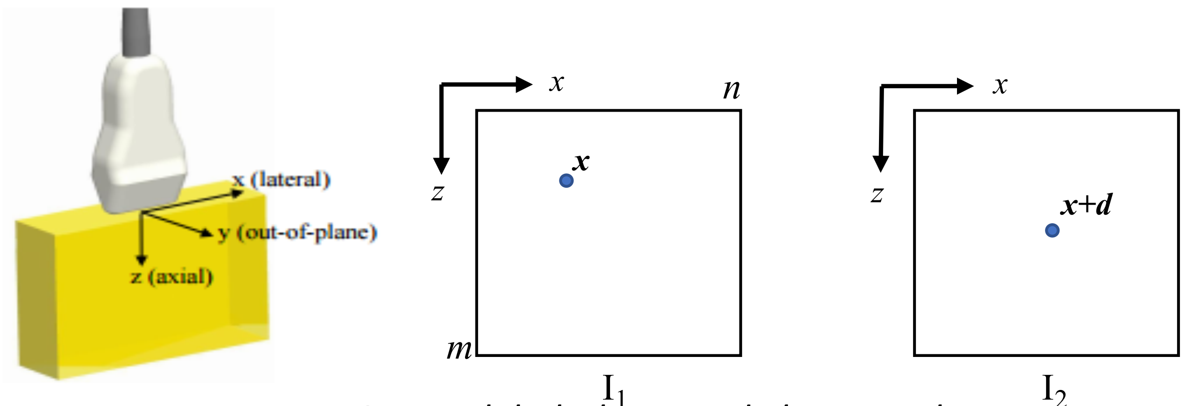


Real-Time Displacement Estimation

- Goal: minimize cost function $C(\mathbf{d})$:

$$C(\mathbf{d}) = \Sigma \ (D(\mathbf{d}) + R(\mathbf{d}))$$

- C is a nonlinear equation.
- There are >1 million variables.



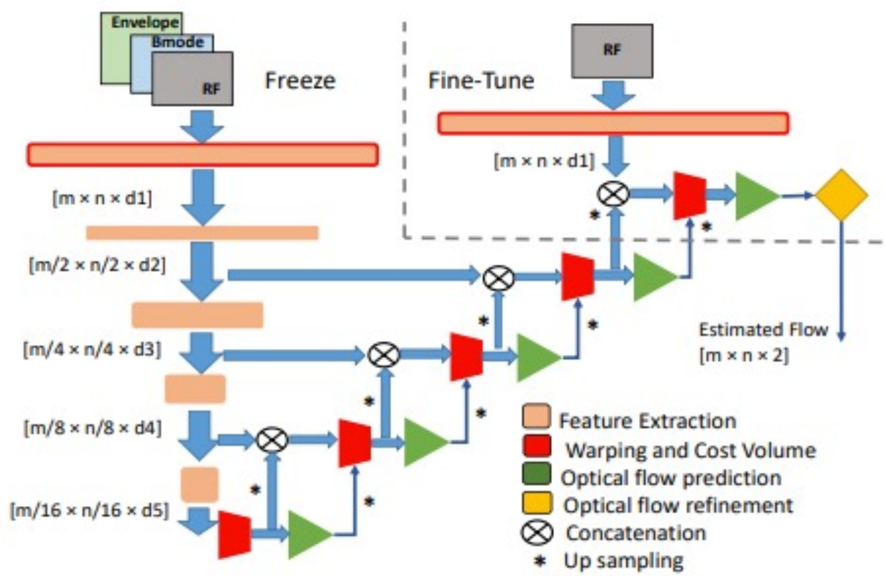
Hashemi & Rivaz, IEEE *TUFFC, 2017,* ***GLUE****: Global Ultrasound Elastography*
Mirzaei, Asif, Rivaz, IEEE *TMI, 1019,* ***OverWind****: tOtal Variation Regularization and WINDow-based Elastography*
Ashikuzzaman, Rivaz, IEEE *TUFFC, 2022,* ***L1-SOUL****: Second-Order Elastography*

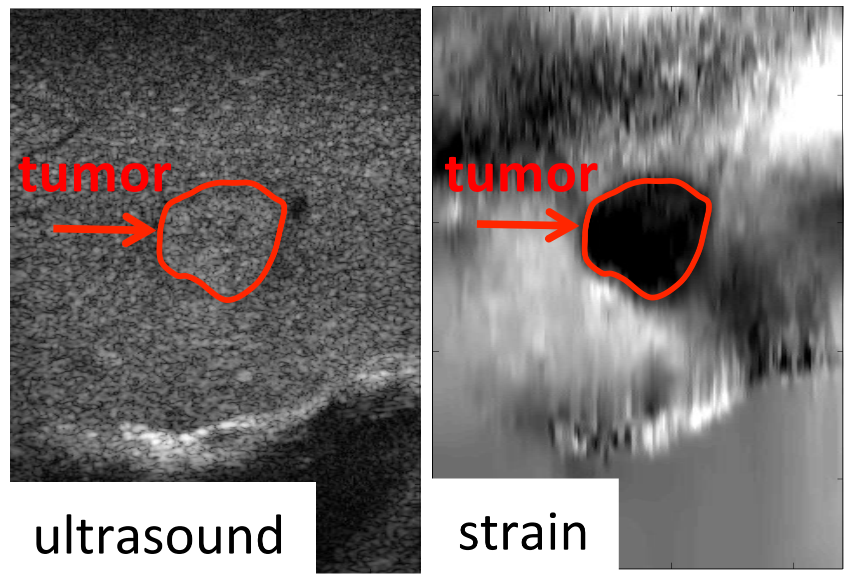
Code available online at code.sonography.ai

IMPACT Laboratory Est. 2014 Transforming Healthcare

CNN for displacement estimation

Radio-Frequency (RF) Modified PWC-Net





In collaboration with Johns Hopkins University

Tehrani*, Mirzaei*, Rivaz, MICCAI 2020
Kibria*, Rivaz, MICCAI POCUS 2018
Tehrani*, Rivaz, IEEE Trans. UFFC special issue on deep learning, 2020

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Supervised Learning

- There is no ground truth in real data
- Using simulated ultrasound images with known displacement

- Issues:
  - Simulation is simplistic (domain shift)
  - Many different imaging settings (domain shift)

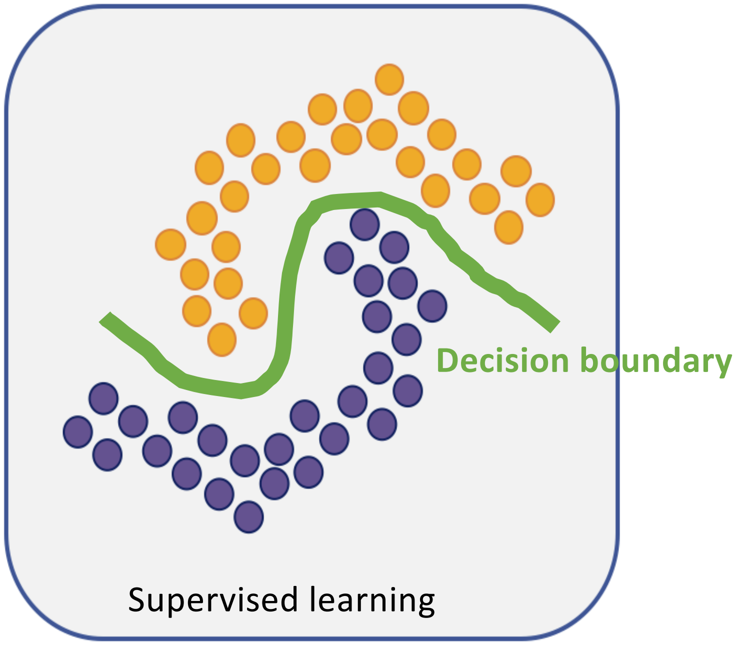
- We have thousands of real images, but **without** ground truth

Kibria*, Rivaz, MICCAI POCUS 2018
Tehrani*, Rivaz, IEEE Trans. UFFC special issue on deep learning, 2020

IMPACT Laboratory
Est. 2014
Transforming Healthcare

In collaboration with Johns Hopkins University

Semi-Supervised Learning (SSL)



- In this work, we fine-tuned the weights on real-data without knowing the ground truth

A Tehrani*, M Mirzaei*, H Rivaz, Bi-Directional Semi-Supervised Elastography, MICCAI 2020
A Tehrani* M Sharifzadeh, E Boctor, H Rivaz, IEEE TUFFC 2022

IMPACT Laboratory
Est. 2014
Transforming Healthcare

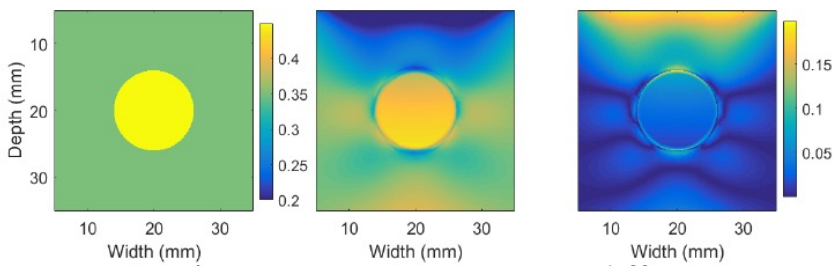
Incorporating Physics into Loss Function

- Homogeneous materials without boundary condition (Hook's law):

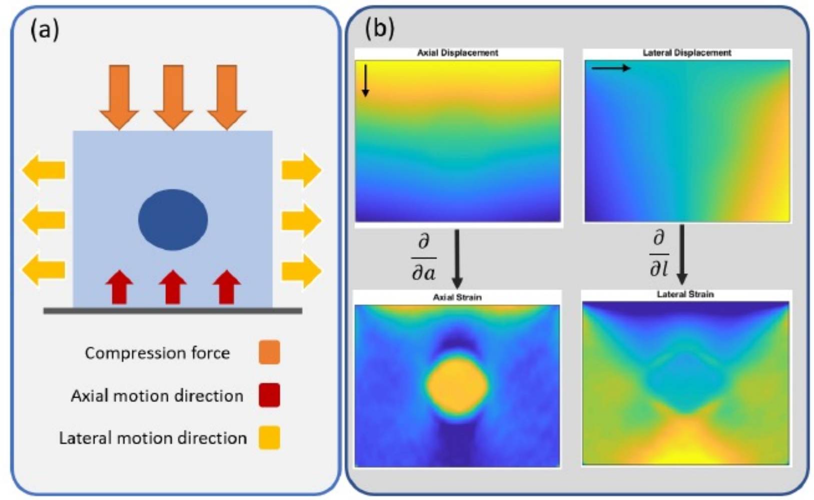
$$lateral\ strain\ = -Poisson's\ Ratio\ \times axial\ strain$$

- Inhomogeneous media with boundary condition:

$$\text{Effective } Poisson's\ Ratio\ (EPR): \frac{-lateral\ strain}{axial\ strain}$$



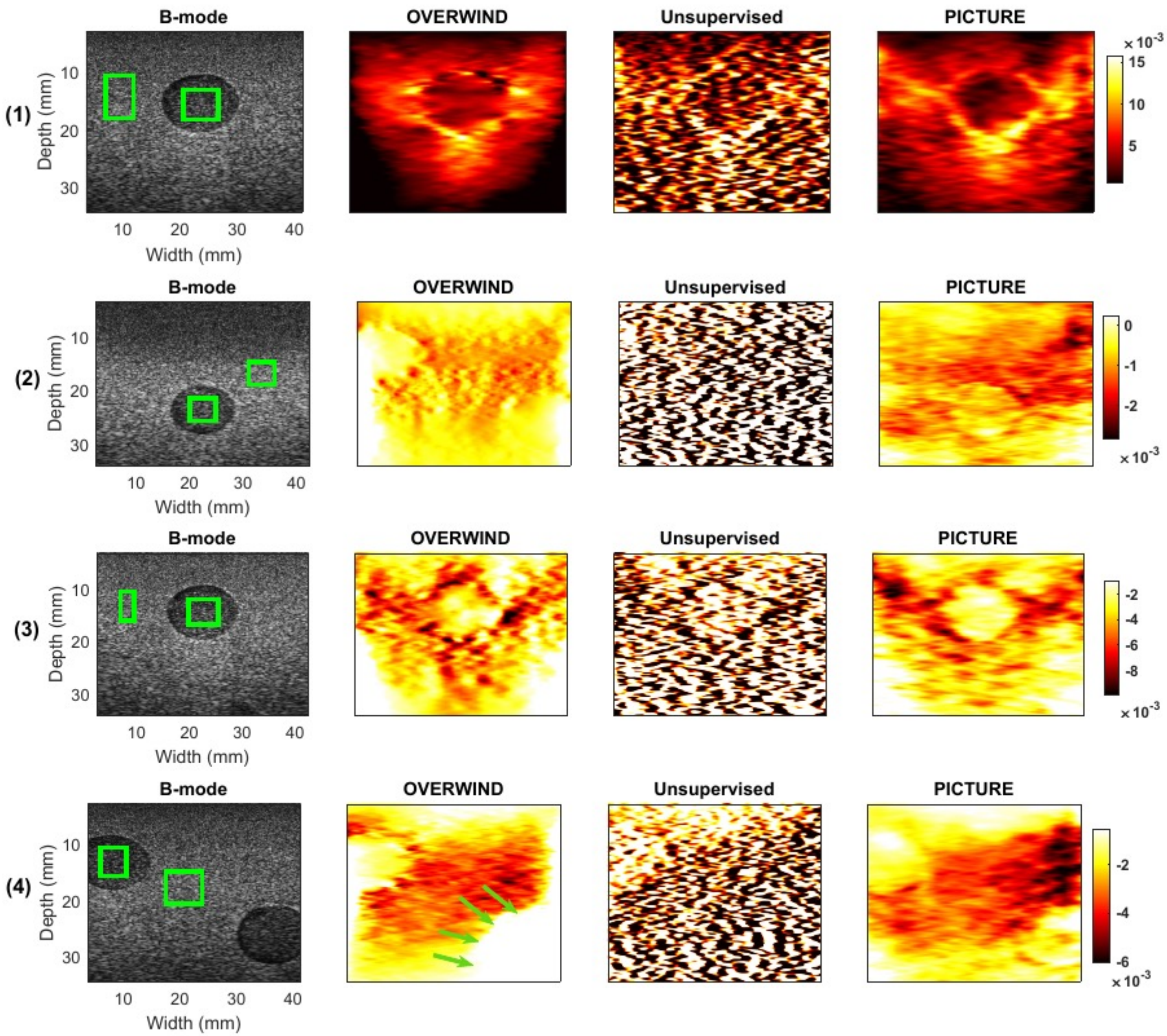
Poisson's Ratio EPR difference



- $EPR \neq Poisson's\ Ratio$ but it has the same range as Poisson's Ratio (0.2 – 0.5).
  We propose penalizing out of range EPRs.

Tehrani*, Rivaz, MICCAI 2022 (accepted)

IMPACT Laboratory
Est. 2014
Transforming Healthcare



PICTURE: Tehrani*, Rivaz, MICCAI 2022 (accepted)

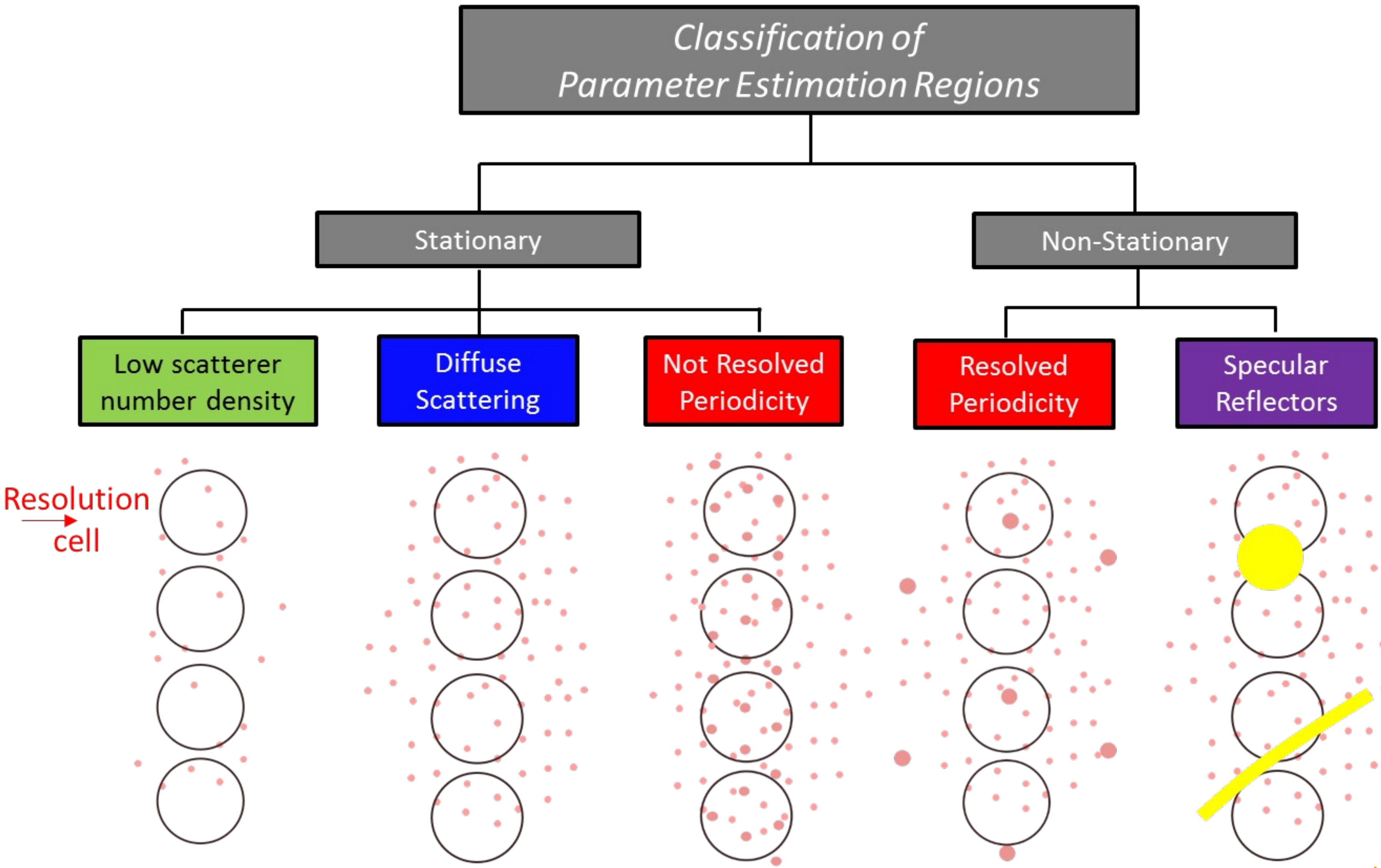
IMPACT Laboratory
Est. 2014
Transforming Healthcare

AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction for transcranial imaging & focusing
  - Deep beamforming

IMPACT Laboratory
Est. 2014
Transforming Healthcare

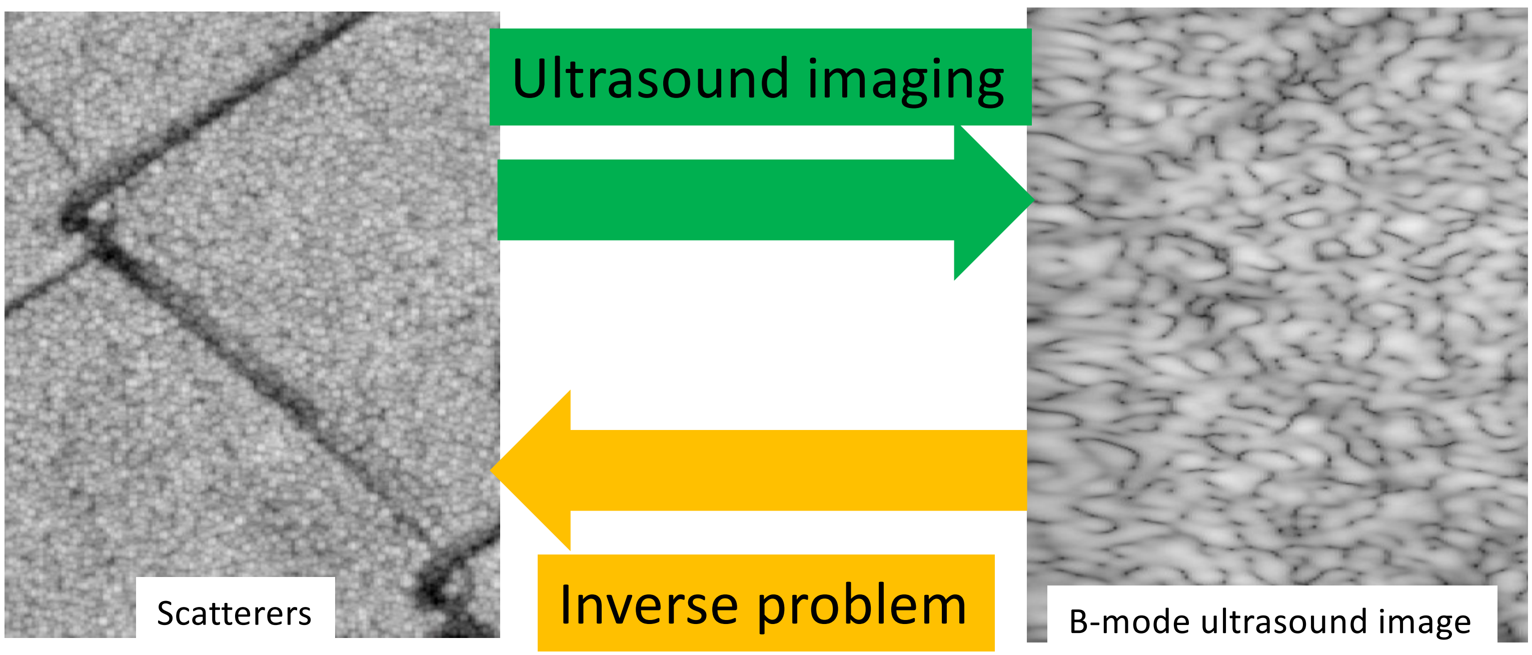
Quantitative US (qUS)



Slide courtesy of Tim Hall and Ivan Rosado-Mendez, UW-Madison

Inferring cell properties from RF data

- Estimate quantitative properties of tissue (cell size, density, scattering, attenuation)
- Open window into quantitative cells properties

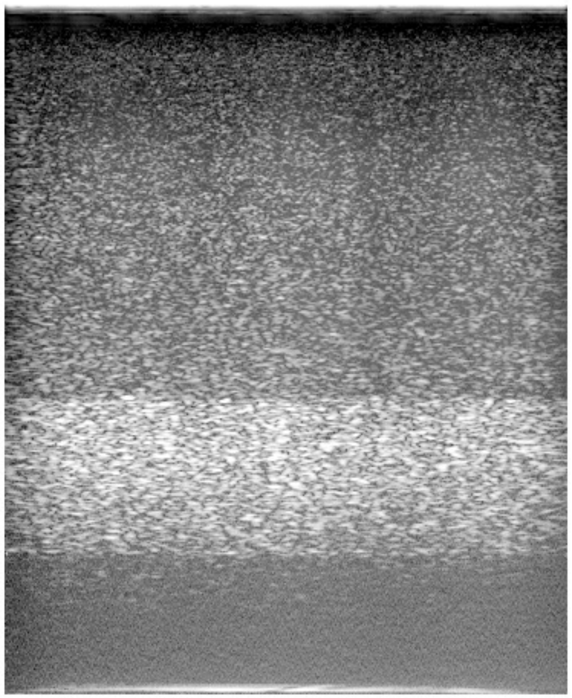


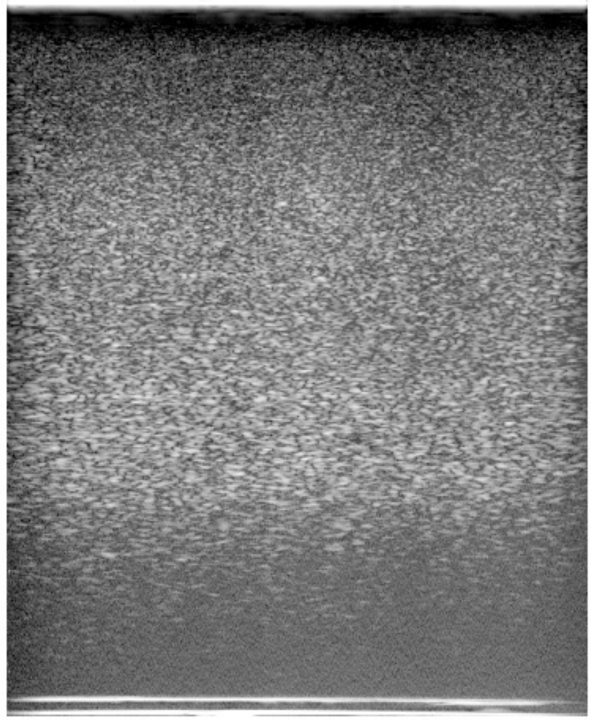
IMPACT Laboratory
Est. 2014
Transforming Healthcare

In collaboration with, UW-Madison

qUS

Layers **visible** in B-mode Layers **invisible** in B-mode





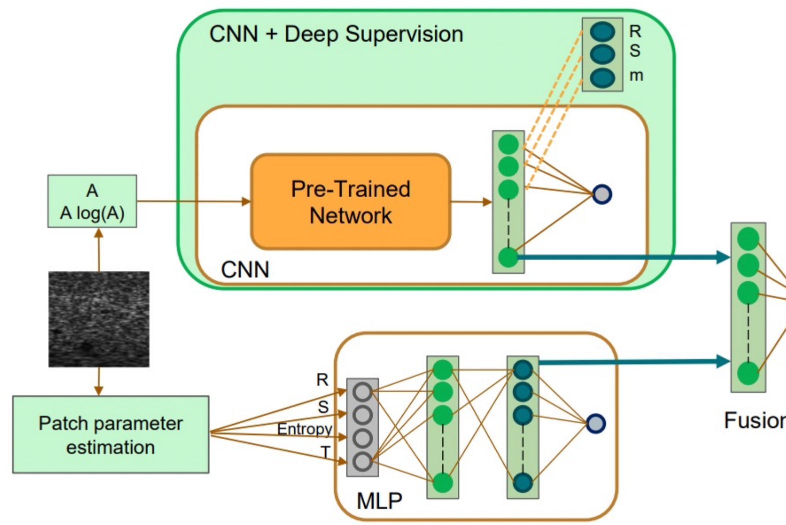
$$\sum_{l=1}^{N_F}\sum_{i=1}^{N_R} w_d(l,i)\left(X(f_l,z_i) - b_i - n_i \ln(f_l) + 4a_i f_l z_i\right)^2$$

Data

+

$$\sum_{i=2}^{N_R} w_a(a_i - a_{i-1})^2 + w_b|b_i - b_{i-1}| + w_n|n_i - n_{i-1}|$$

Physics-based regularization
ADMM Optimization



- Quantitative tissue properties (attenuation and backscattering) can be estimated with low variance from spectrum of RF data

Vajihi*, Rosado-Mendez, Hall, Rivaz, IEEE Trans. UFFC (TUFFC), 2018
Jafarpisheh*, Hall, Rivaz, Rosado-Mendez, IEEE Trans UFFC, 2021
A Tehrani*, M Amiri* I Rosado-Mendez, T J Hall, H Rivaz, IEEE Trans UFFC, 2021

Est. 2014
Transforming Healthcare

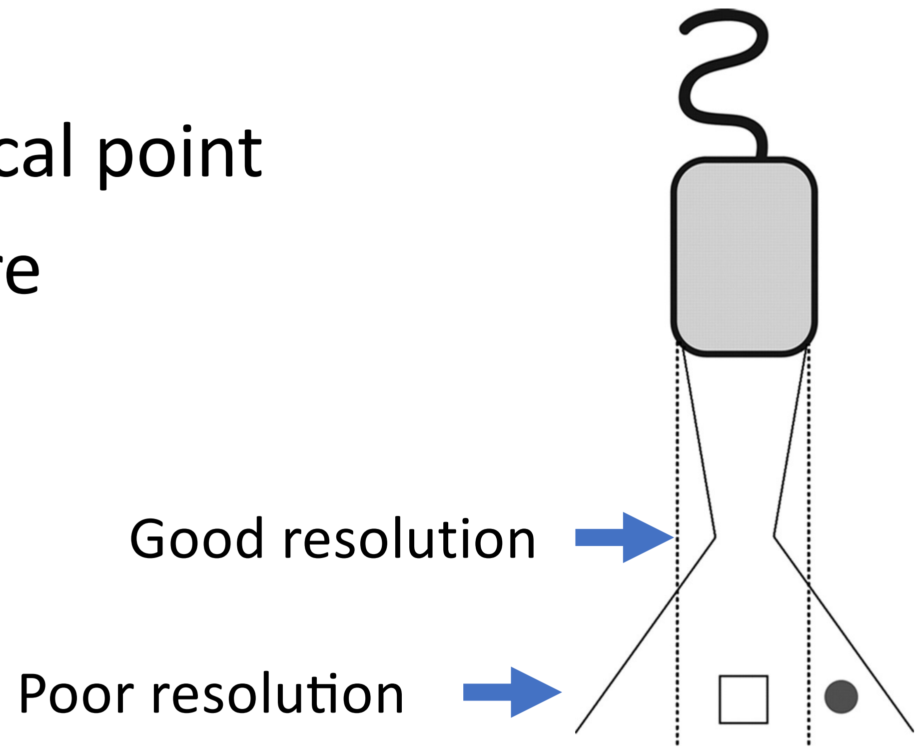
AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
    - Elastography
    - Quantitative ultrasound (qUS)
    - Faster US acquisition
    - Aberration correction for transcranial imaging & focusing
    - Deep beamforming

IMPACT Laboratory
Est. 2014
Transforming Healthcare

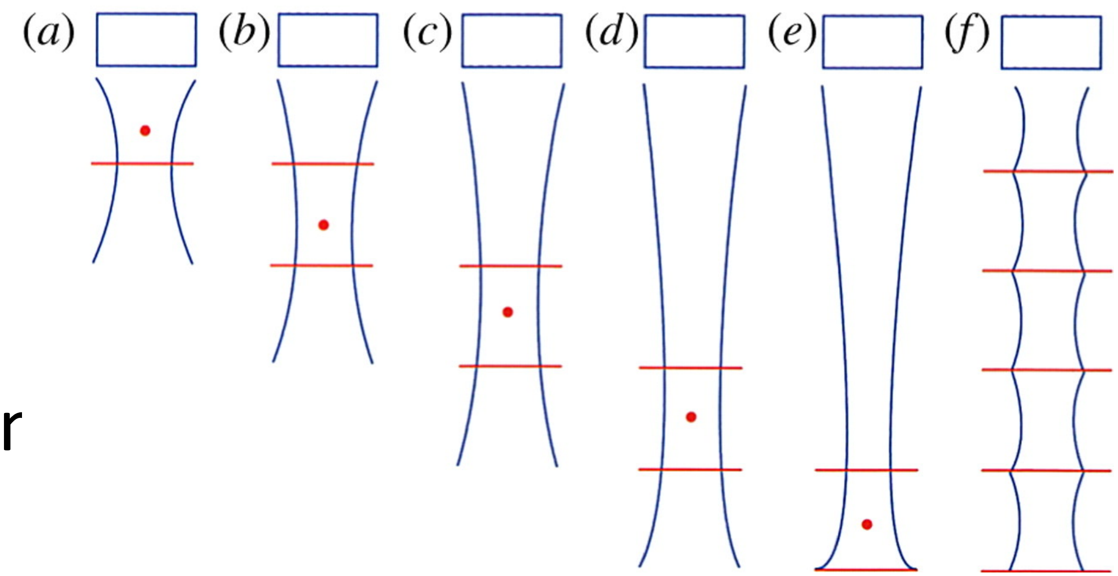
Focusing the Transmit Beam

- Transmit focusing
- Good resolution at focal point
- Not so good elsewhere



Multi-Focus US Imaging

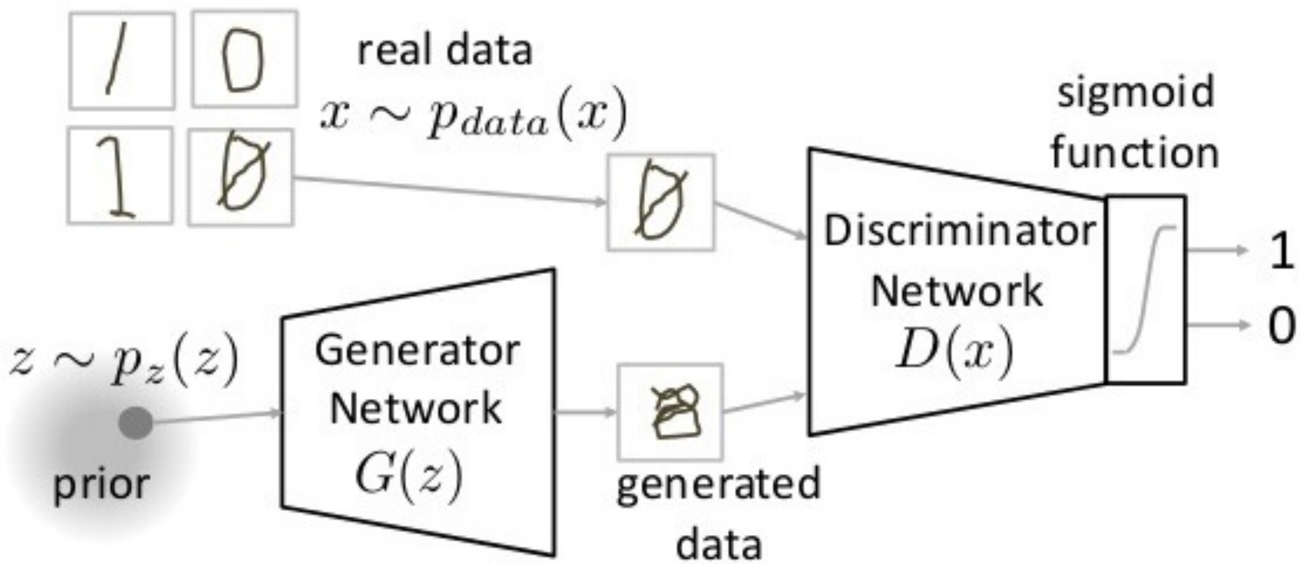
- Good spatial resolution
- Poor temporal resolution
- Susceptible to motior blurring



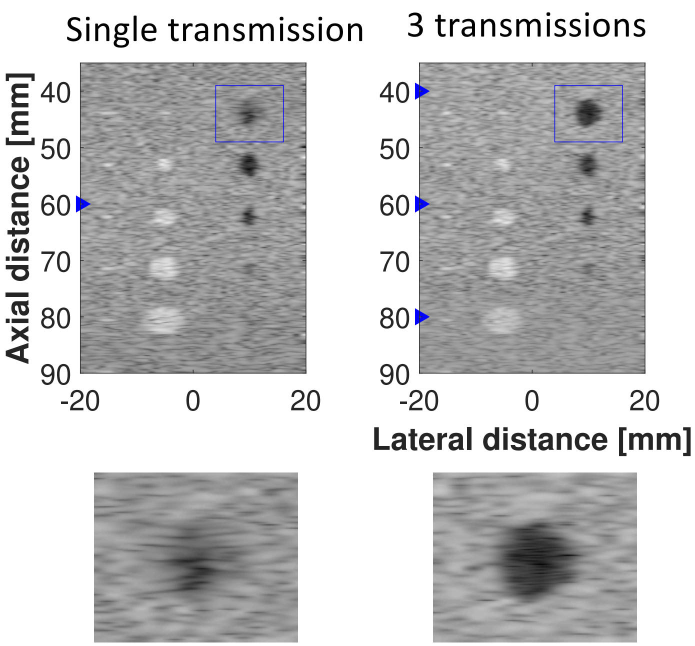
IMPACT Laboratory
Est. 2014
Transforming Healthcare

Generative Adversarial Network (GAN)

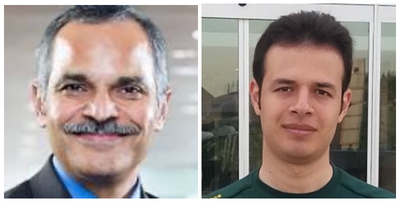
- What is a GAN?




Can We Obtain Multi-Focus US With GAN?



Goudarzi*, Asif, Rivaz, IEEE ISBI 2019
Goudarzi*, Asif, Rivaz, IEEE Transactions Computational Imaging, 2020



Can We Obtain Multi-Focus US With GAN?

Goudarzi*, Asif, Rivaz, IEEE Transactions Computational Imaging, 2020

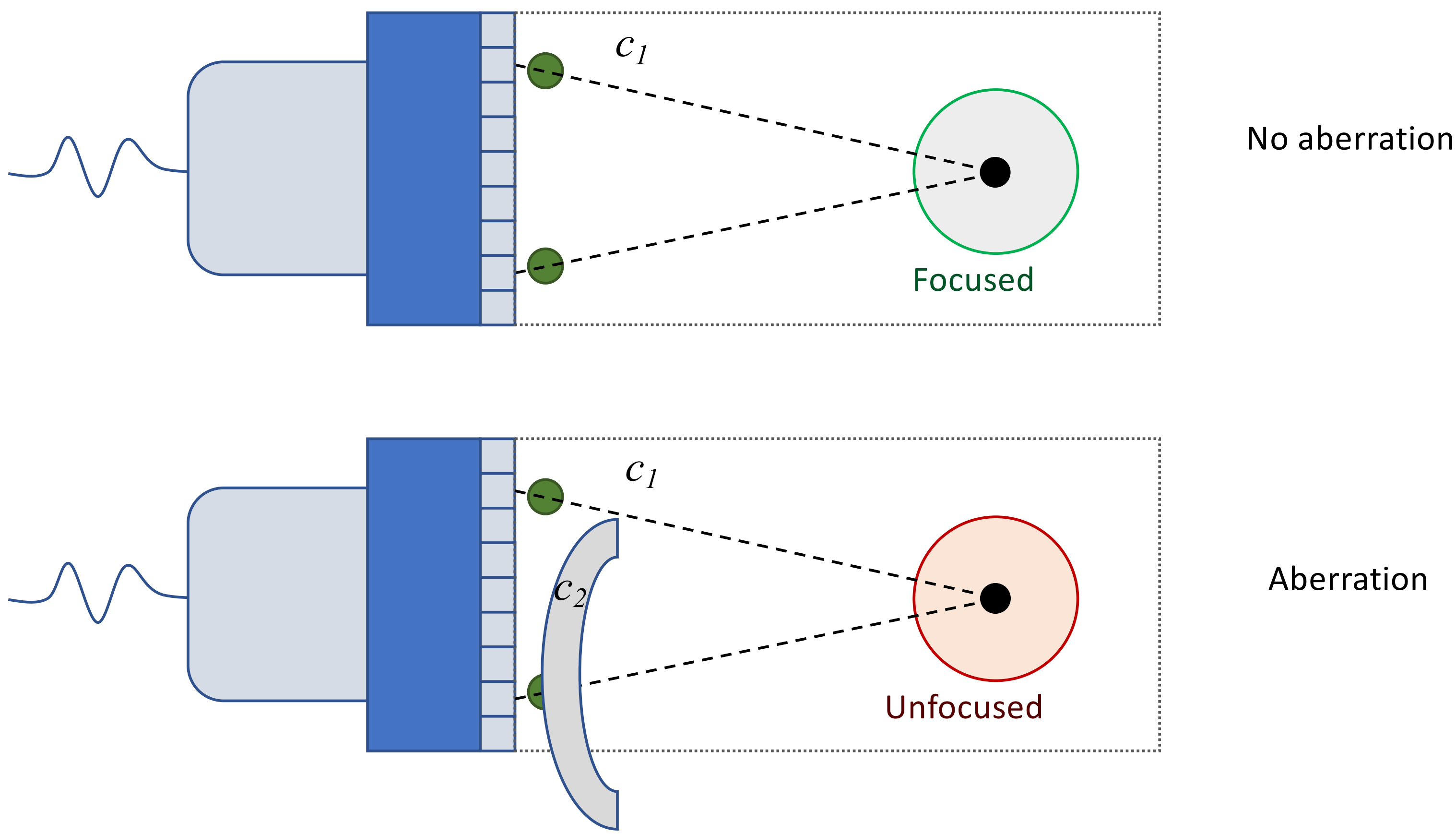
IMPACT Laboratory
Est. 2014
Transforming Healthcare

AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction for transcranial imaging & focusing
  - Deep beamforming

IMPACT Laboratory
Est. 2014
Transforming Healthcare

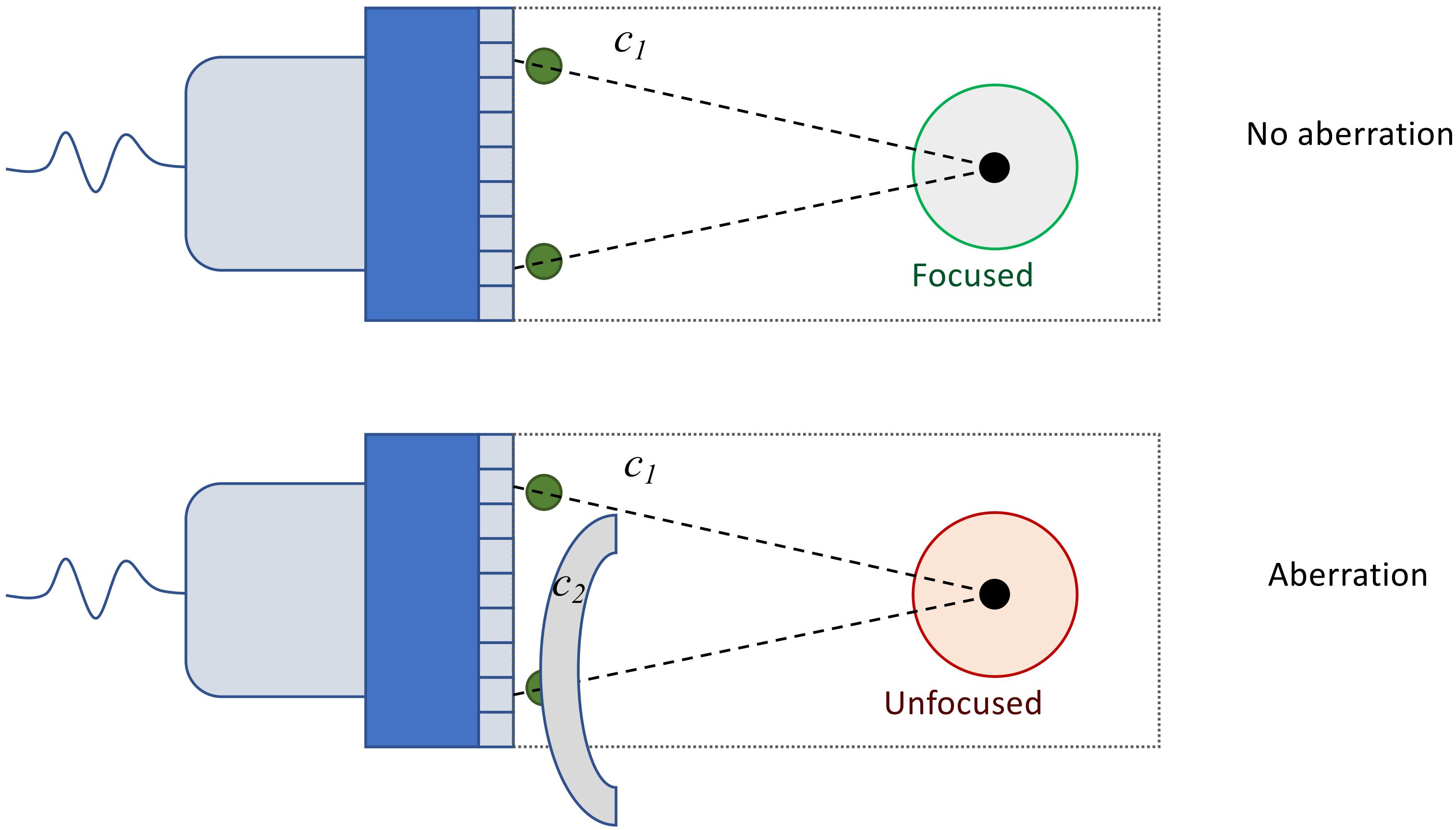
Phase aberration in US



The skull is inhomogeneous and anisotropic. Its speed of sound is also high.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Phase aberration in US

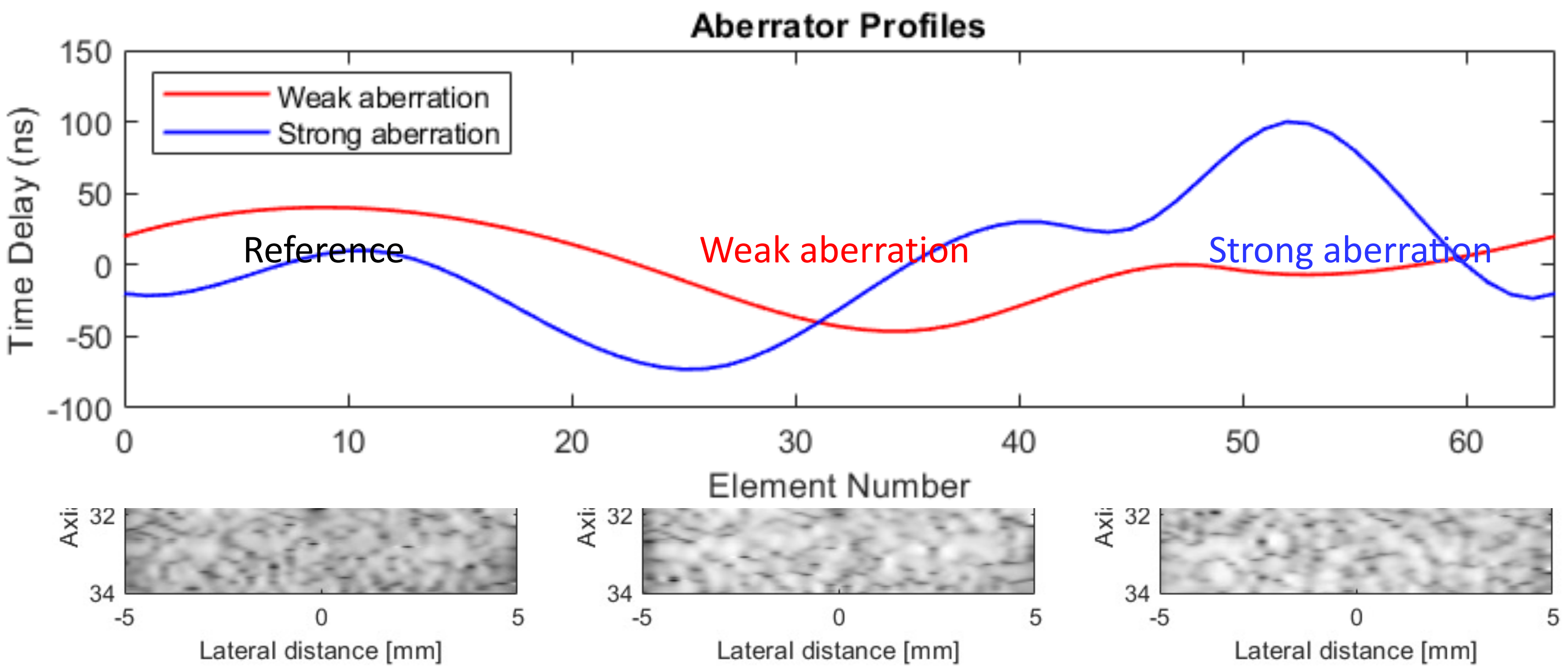


The skull is inhomogeneous and anisotropic. Its speed of sound is also high.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

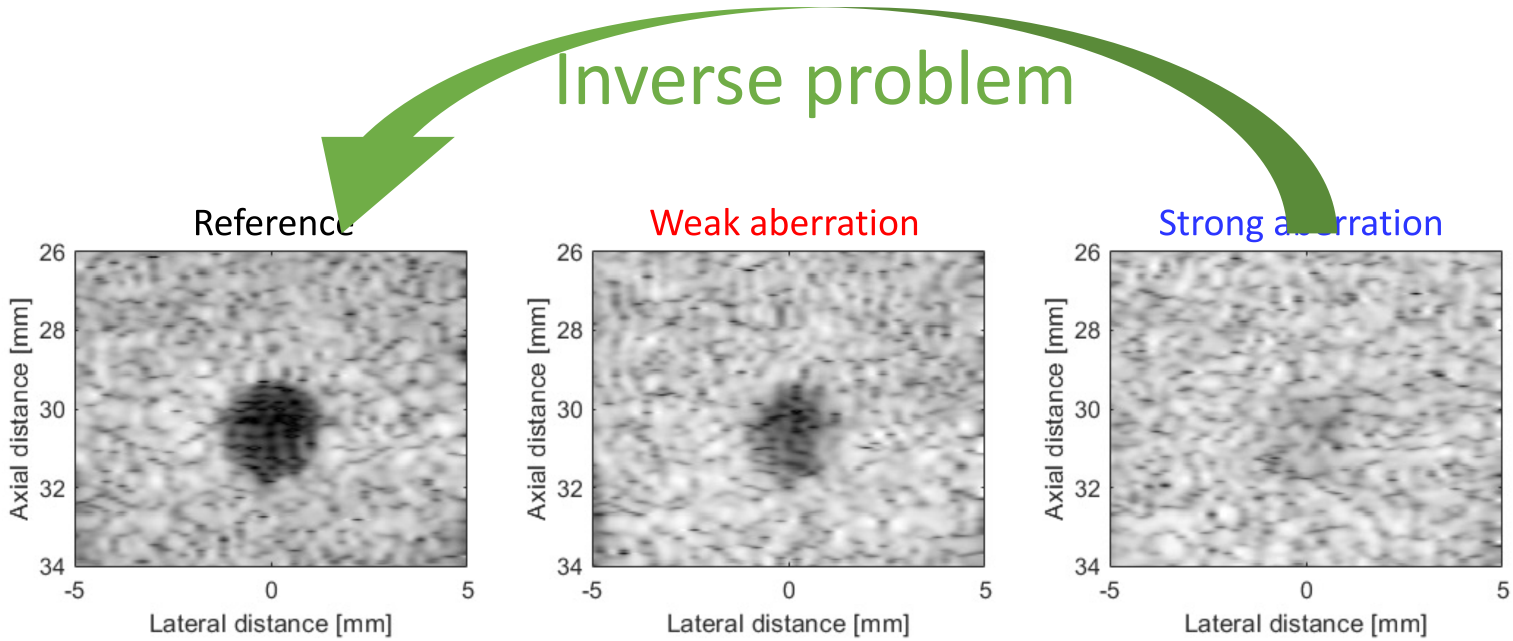
Phase aberration in transcranial US

- Apply an aberrator profile to the transducer’s elements in both transmit and receive cases.



Sharifzadeh*, Benali, Rivaz, IEEE Access 2020


Phase aberration in transcranial US



- Transcranial imaging
- LIFU for modulating brain function
- HIFU for drug delivery/ablation

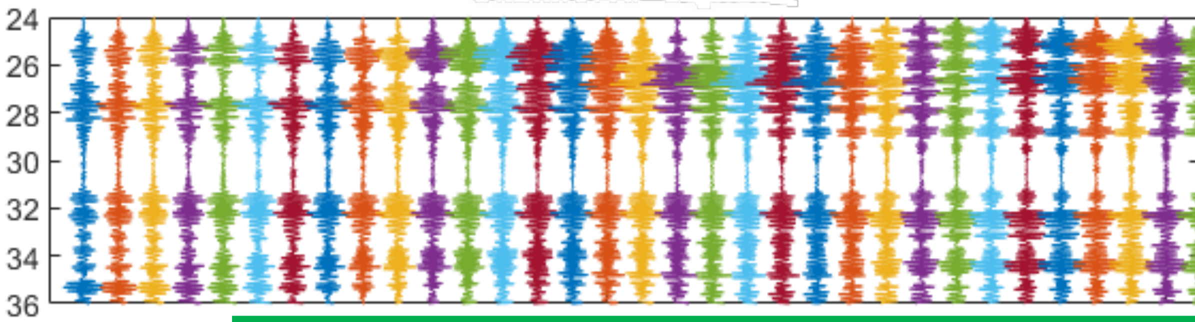
Sharifzadeh*, Benali, Rivaz, IEEE Access 2020

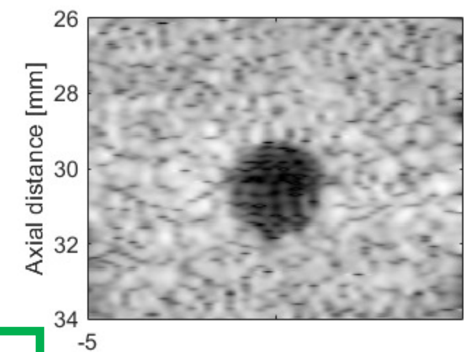
Making Sense of Data Trove in Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction in transcranial ultrasound
  - Deep beamforming

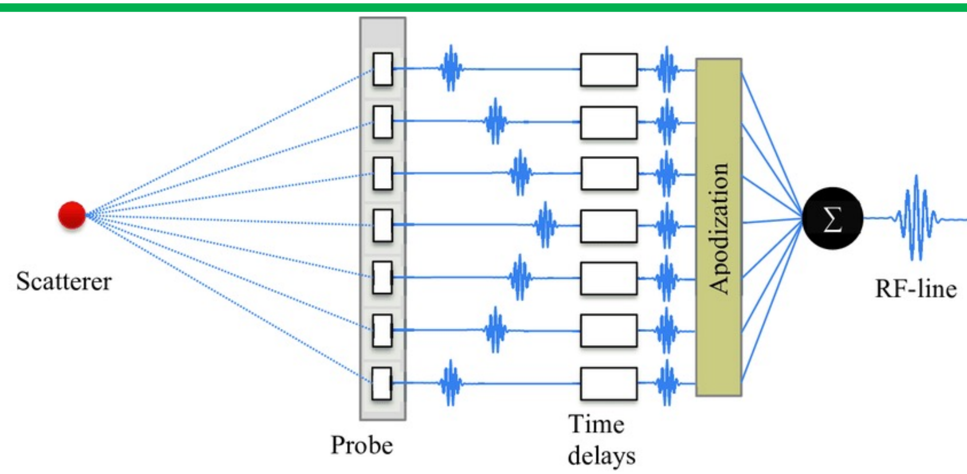
Deep Learning in Ultrasound Imaging

- B-mode US (the familiar gray-scale images) only contains a small percentage of all data





B-mode ultrasound
Visually informative



From: New beamforming strategy for improved ultrasound imaging: application to biological tissues nonlinear imaging, Toulemonde, PhD thesis, University Lyon 1, 2014

IMPACT Laboratory
Est. 2014
Transforming Healthcare

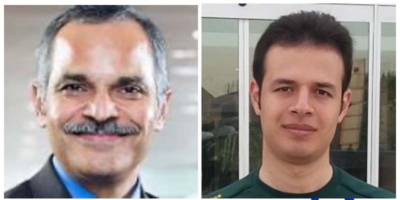

Challenges of deep beamforming

- Scarcity of training data
- Lack of ground truth in real data
- Domain shift between real and simulation data

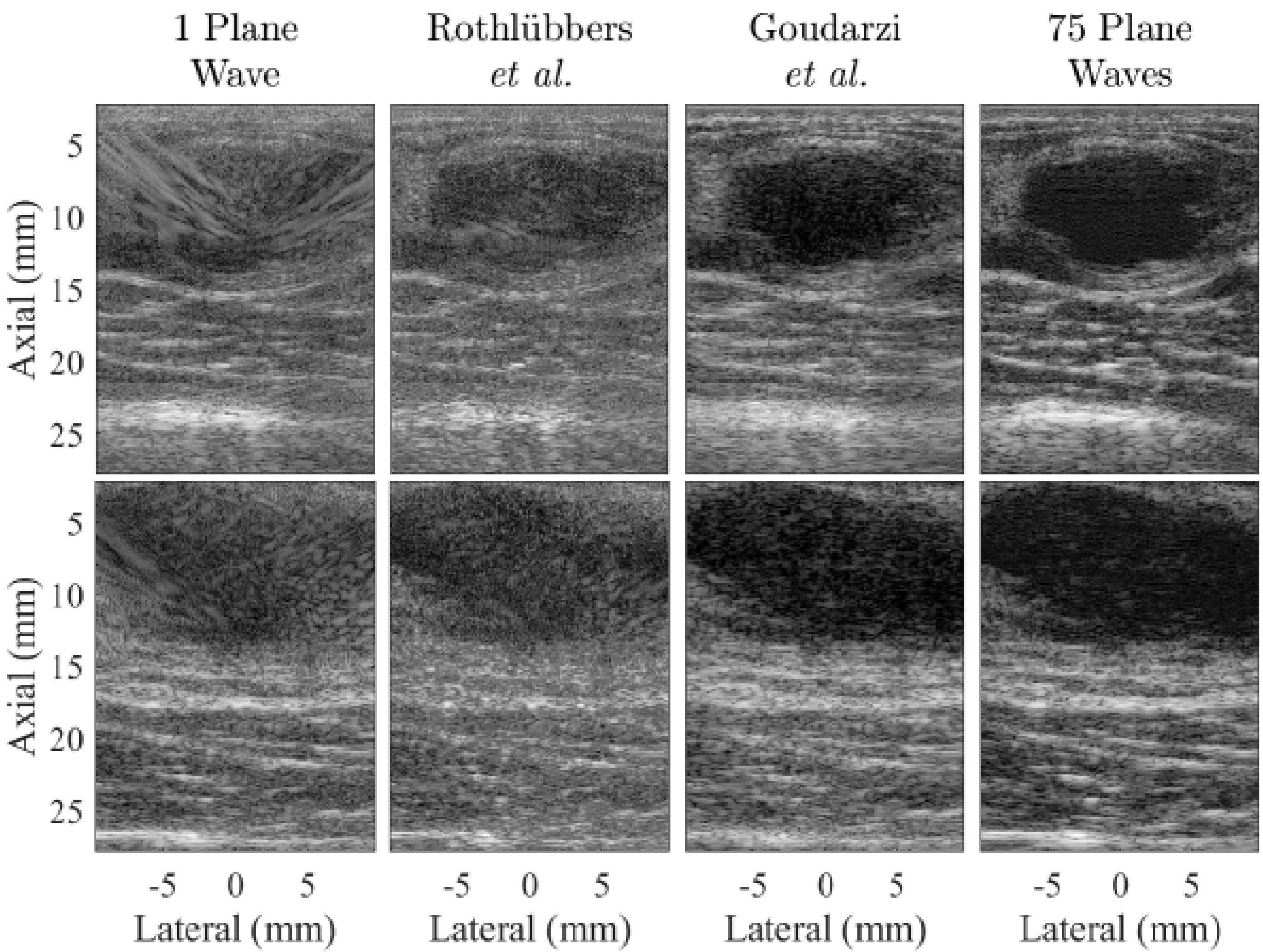
- Apply Minimum Variance Beamforming (MVB) to real data
  - A set of closed-form mathematical steps
  - Train on each pixel separately, so lots of training data
  - Training directly on real data

S. Goudarzi*, A. Asif, and H. Rivaz, Ultrasound Beamforming using MobileNetV2, IEEE IUS, 2020
Ranked First (tied) in IEEE IUS Challenge on Ultrasound Beamforming with Deep Learning (CUBDL)

IMPACT Laboratory
Est. 2014
Transforming Healthcare



Results on Breast Ultrasound



Hyun, D, et al. "Deep Learning for Ultrasound Image Formation: CUBDL Evaluation Framework & Open Datasets." *IEEE Transactions on UFFC* (2021).

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Clinical Applications

Breast Cancer Related Lymphedema

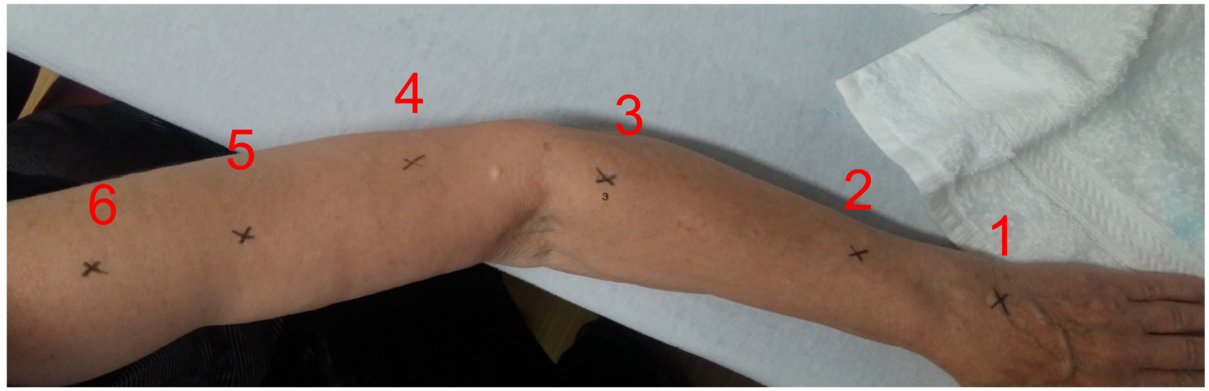
- Localized fluid retention and tissue swelling caused by a compromised lymphatic system
- Most frequently a complication of cancer treatment
- It's debilitating. Some patients refer to it as being worse than cancer
- Pitting test is commonly performed for staging
- Ultrasound cannot accurately stage it

IMPACT Laboratory
Est. 2014
Transforming Healthcare


In collaboration with McGill University


Staging Lymphedema

- Significant difference in strain between affected and unaffected arms
- Deep models can help clinicians

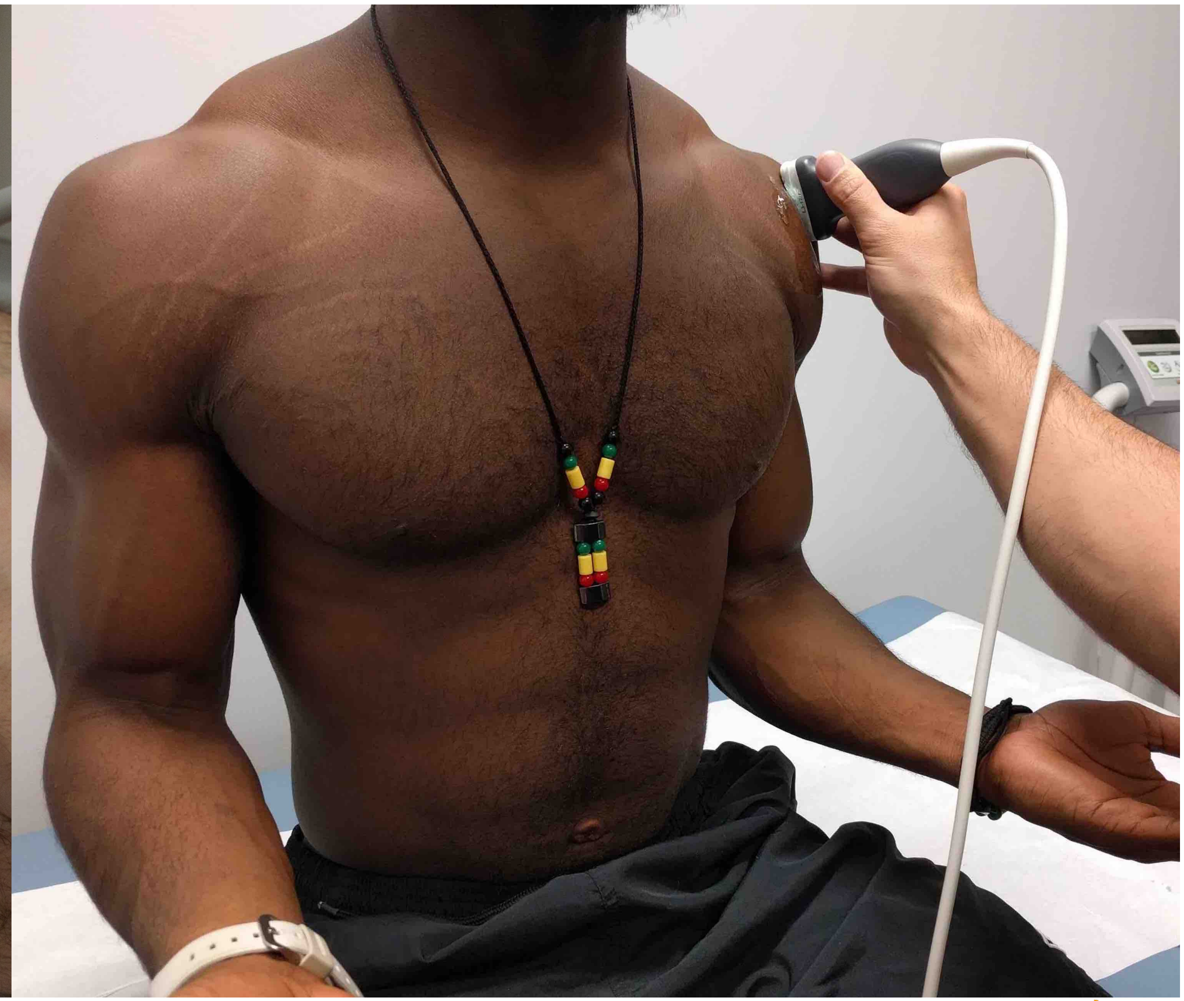


Hashemi*, Fallone, Boily, Towers, Kilgour, Rivaz, IEEE Transactions UFFC, 2019
Goudarzi*, Whyte, Boily, Towers, Kilgour, Rivaz, IEEE Transactions BME, under review

IMPACT Laboratory
Est. 2014
Transforming Healthcare


In collaboration with McGill University Health Centre


Tendon and Ligament Elastography

Hashemi*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2017

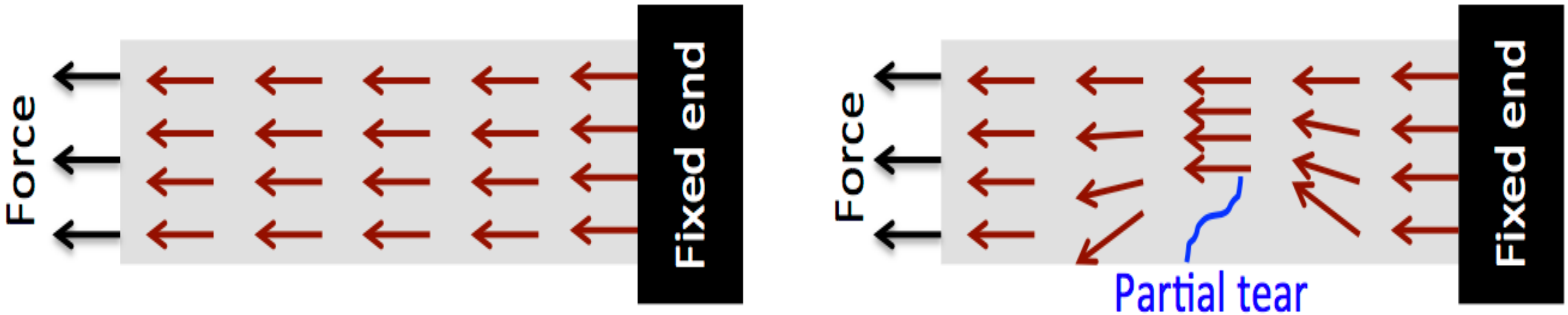
Shams*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2016, finalist for the Best Paper Award

IMPACT Laboratory

Est. 2014

Transforming Healthcare

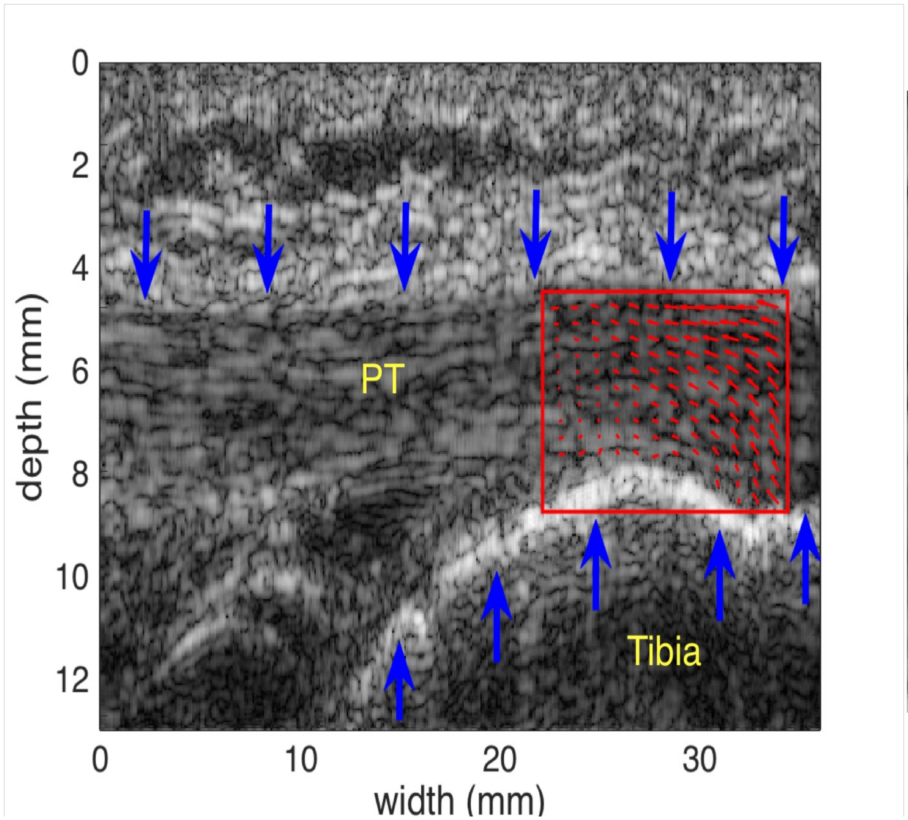
Diagnosis with Deformation Pattern

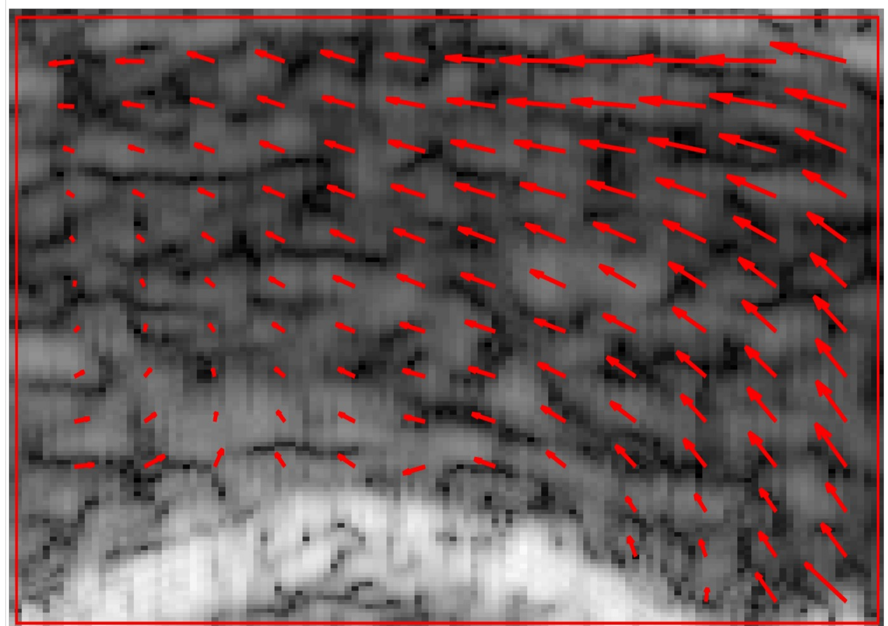


Hashemi*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2017
Shams*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2016, finalist for best paper award

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Diagnosis with Deformation Pattern

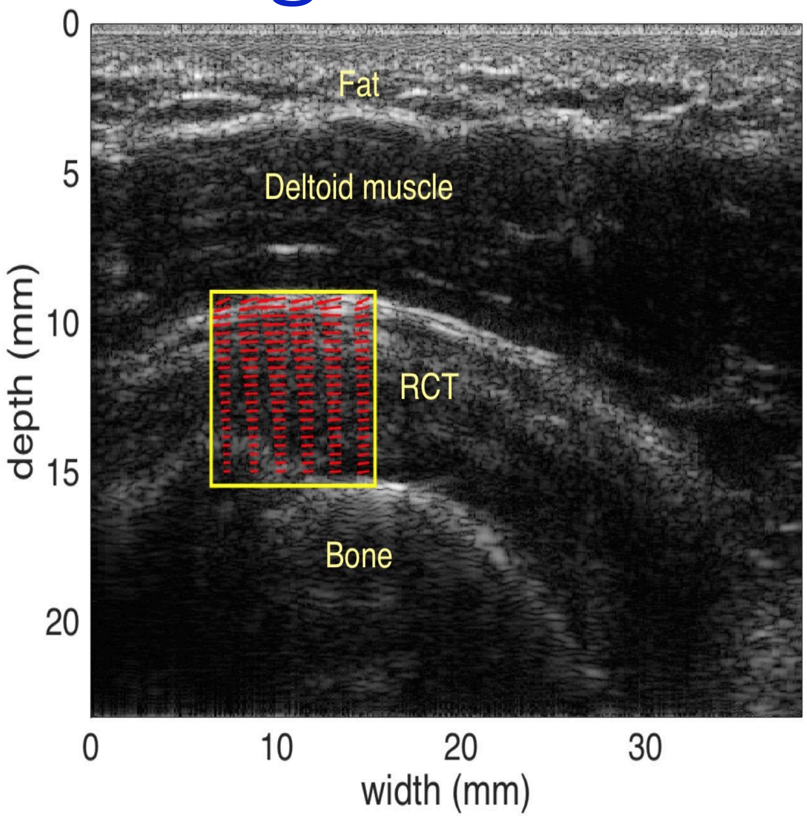


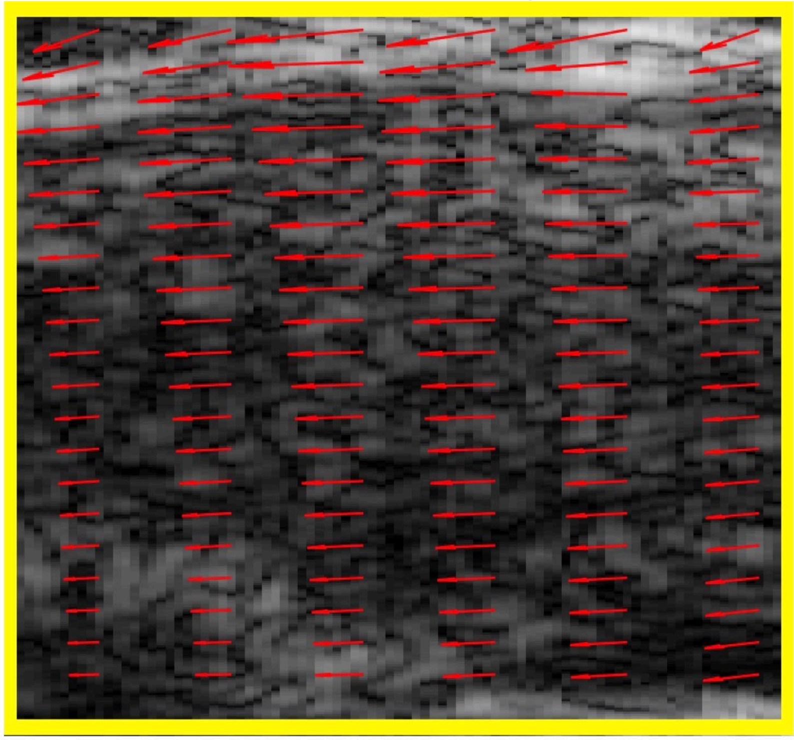


Hashemi*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2017
Shams*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2016, finalist for best paper award

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Diagnosis with deformation pattern





Hashemi*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2017
Shams*, Boily, Martineau, Rivaz, SPIE Med. Imag. 2016, finalist for best paper award

IMPACT Laboratory
Est. 2014
Transforming Healthcare

AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction for transcranial imaging & focusing
  - Deep beamforming
- Aid collection & interpretation
  - Image Registration
    - US to MRI
    - US to US
  - Image segmentation
    - Explainable AI (XAI)
  - Image selection

IMPACT Laboratory Est. 2014 Transforming Healthcare

AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction for transcranial imaging & focusing
  - Deep beamforming
- B-mode ultrasound is hard to collect & interpret
  - Image Registration
    - US to MRI
    - US to US
  - Image segmentation
    - Explainable AI (XAI)
  - Image selection

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Brainshift

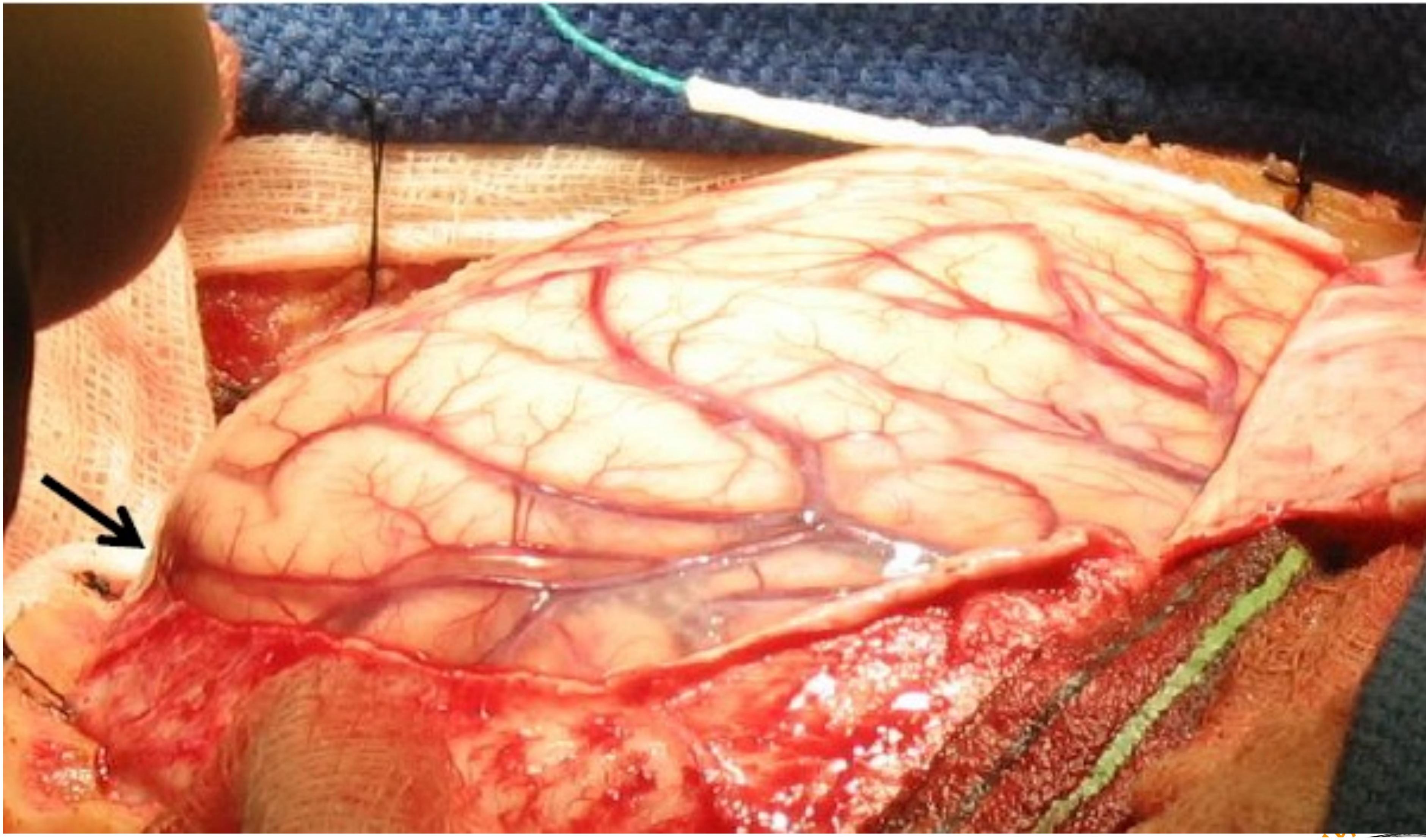


Photo courtesy of Sean Chen, MNI

2014

Transforming Healthcare

Image-Guided Neurosurgery

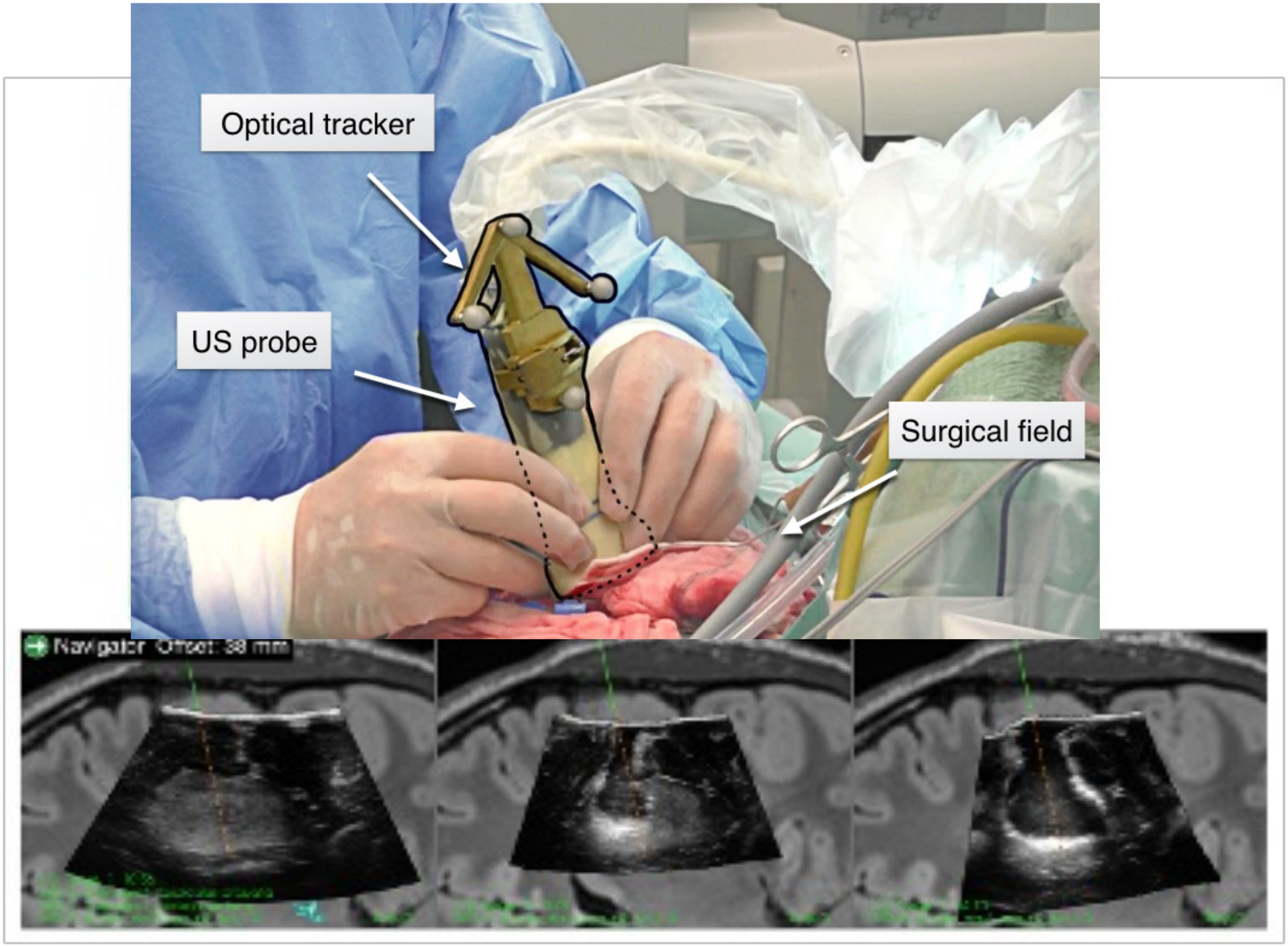
- Due to the brainshift, tracked US is used in the OR
- US images are hard to interpret and are noisy
- Pre-op MRI is available
- Deform the pre-op MRI to match intra-op US

Rivaz, Chen, Collins, IEEE Trans. Med. Imag. (TMI) 2015

drawing courtesy of Dr. Simon Drouin

IMPACT Laboratory
Est. 2014
Transforming Healthcare

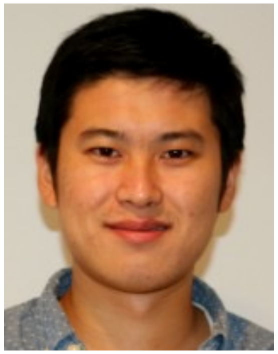
Not Just at the Beginning of Operation



Before During After

Y. Xiao, M. Fortin, G. Unsgard, H. Rivaz, I. Reinertsen, Med Phys 2017

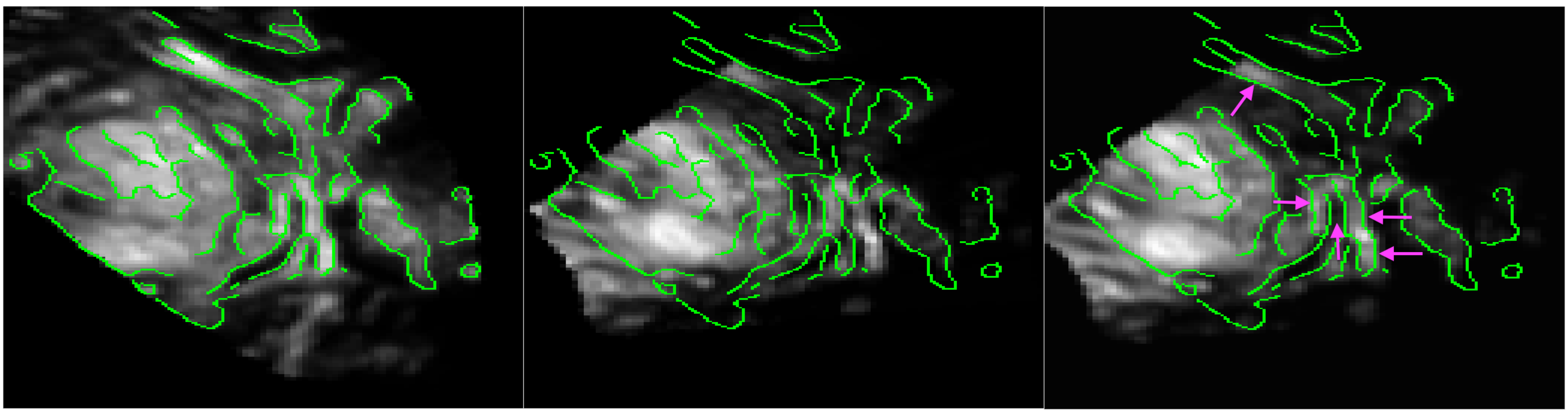
IMPACT Laboratory
Est. 2014
Transforming Healthcare



Registration of Pre- and Post-Resection US

- The quality of post-resection US image is poor due to the resection cavity
- Registration of pre- and post-resection US images is challenging due to missing tissue

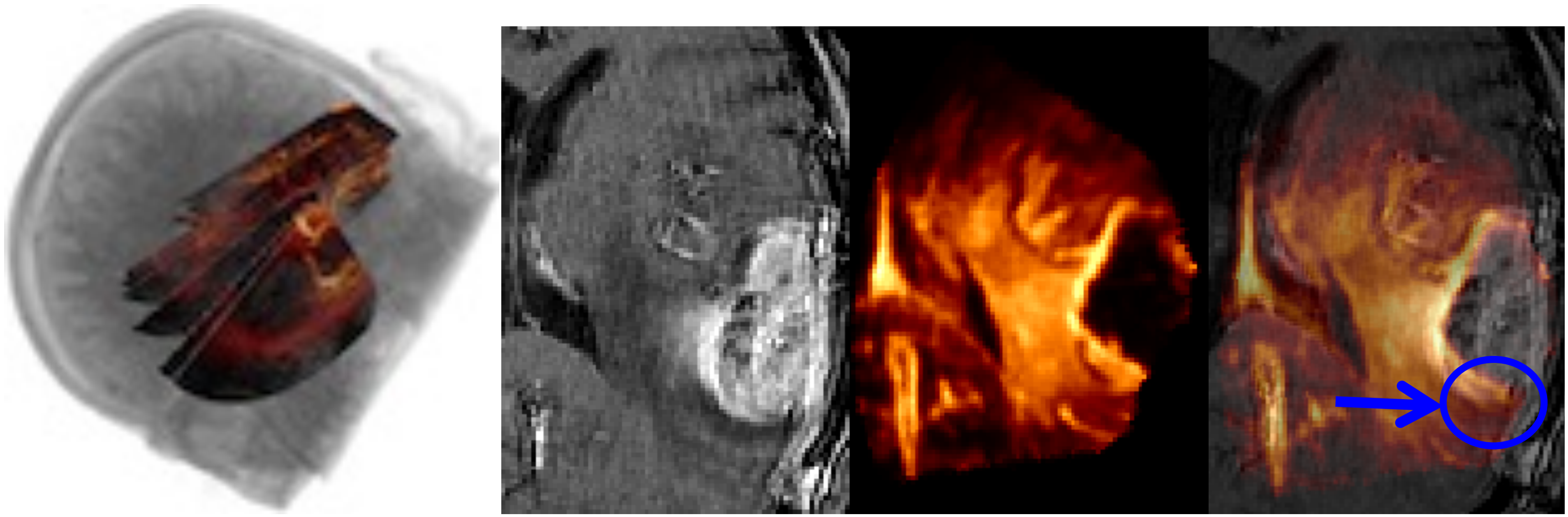
pre-resection | post-resection, initial alignment | post-resection, after registration



Zhou* & Rivaz, IEEE Journal Biomedical Health Informatics (JBHI), 2016. selected for submission of full paper to JBHI, **only 10 selections out of 2948 submissions to IEEE EMBC**

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Finding the Residual Tumor



US and rendered MR

Pre-operative MR

Post-resection US
(tumour is resected)

Residual tumour

Rivaz, H, Chen, S, Collins, DL, IEEE Trans. Med. Imag. 2015

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Brainshift in Tractography

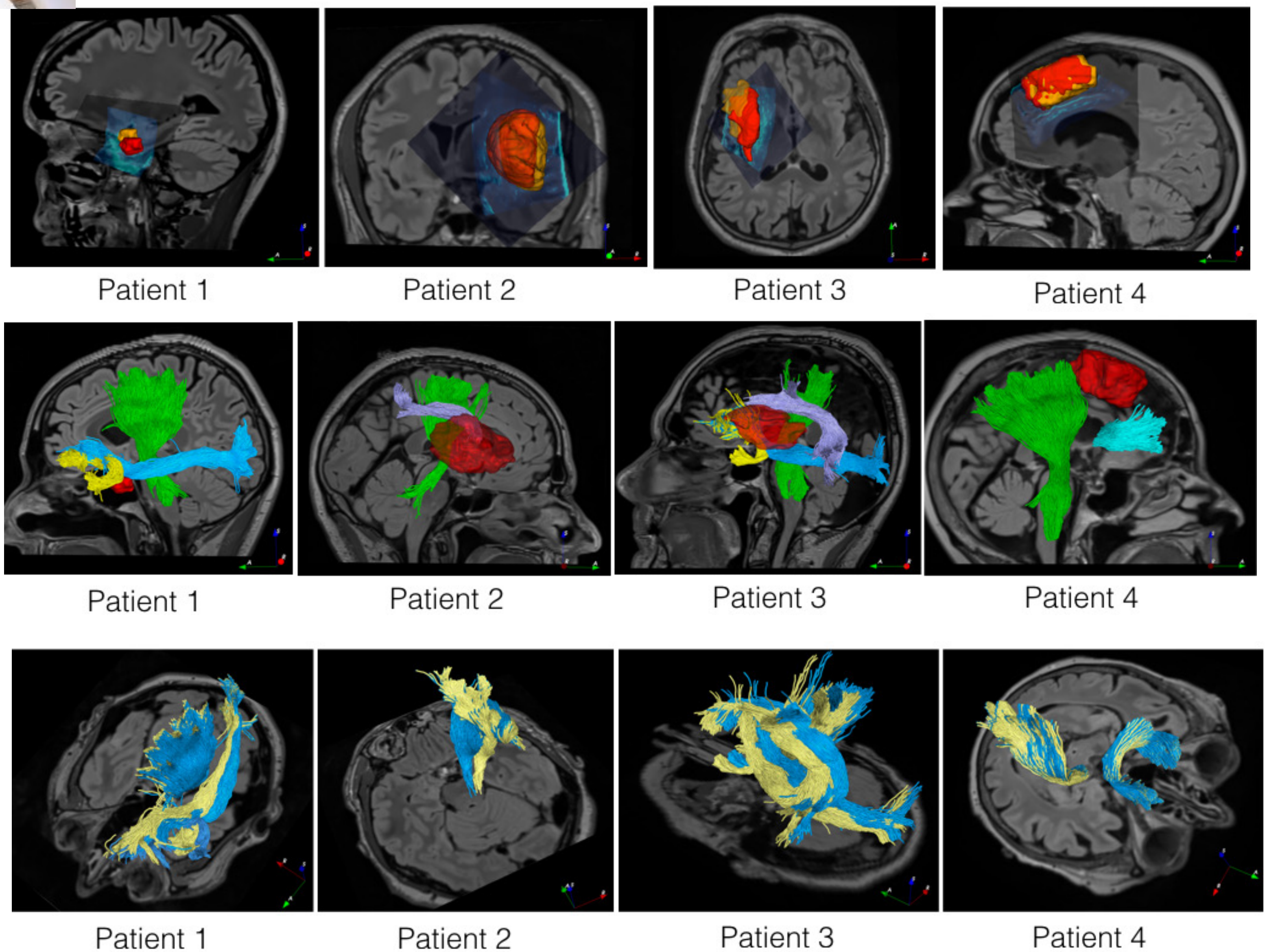
cortico-spinal tract

uncinate fasciculus

superior longitudinal fasciculus

inferior fronto-occipital tract

CC-genu



Xiao*, Eikenes, Reinertsen, Rivaz, IJCARS 2018

IMPACT Laboratory Est. 2014 Transforming Healthcare

AI-Driven Ultrasound

- Better use of raw radio-frequency (RF) data
  - Elastography
  - Quantitative ultrasound (qUS)
  - Faster US acquisition
  - Aberration correction for transcranial imaging & focusing
  - Deep beamforming
- B-mode ultrasound is hard to collect & interpret
  - Image Registration
    - US to MRI
    - US to US
  - Image segmentation
    - Explainable AI (XAI)
  - Image selection

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Transfer learning in ultrasound

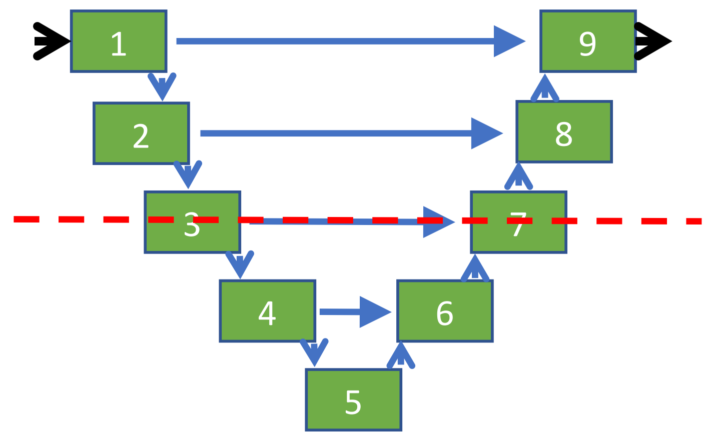
- Why: Low number of images available
- Shallow layers: Low-level features (usually common between different classes)
- Deep layers: High-level features (usually different in each class)

Image adapted from Deep Learning at Cloud Scale

- Speckle pattern/texture is important in ultrasound

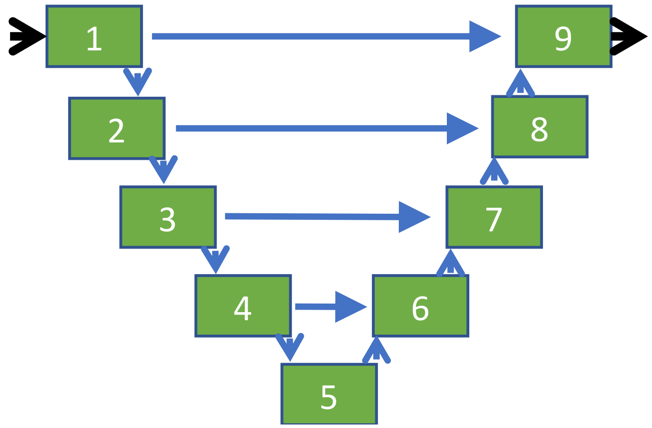
Transfer learning in ultrasound

- Because of skip connections, the notion of shallow/deep is not well-defined
- **Findings:**
- Shallow layers are important!
- Fine-tuning only the last layers is not a good practice in ultrasound image segmentation using U-Net



M. Amiri*, R. Brooks, H. Rivaz, IEEE Transactions UFFC, Special Issue on Deep Learning, 2020

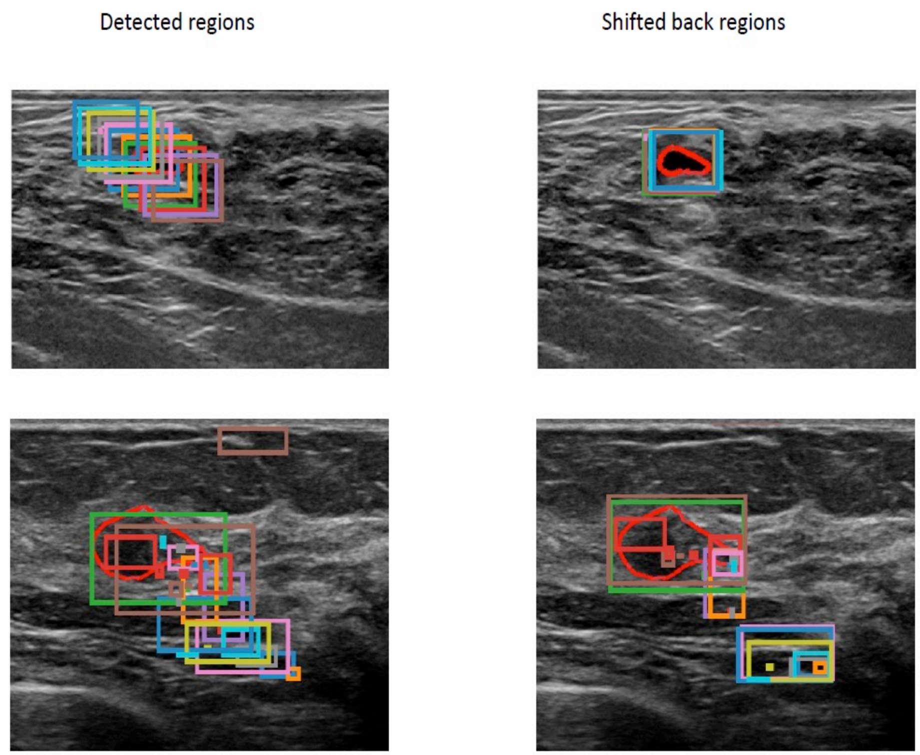
IMPACT Laboratory
Est. 2014
Transforming Healthcare



Est. 2014
Transforming Healthcare

Test-time augmentation

- Shifting
- Dropout uncertainty (dropout ON at test time)



- Valid detection is insensitivity to shift and dropout

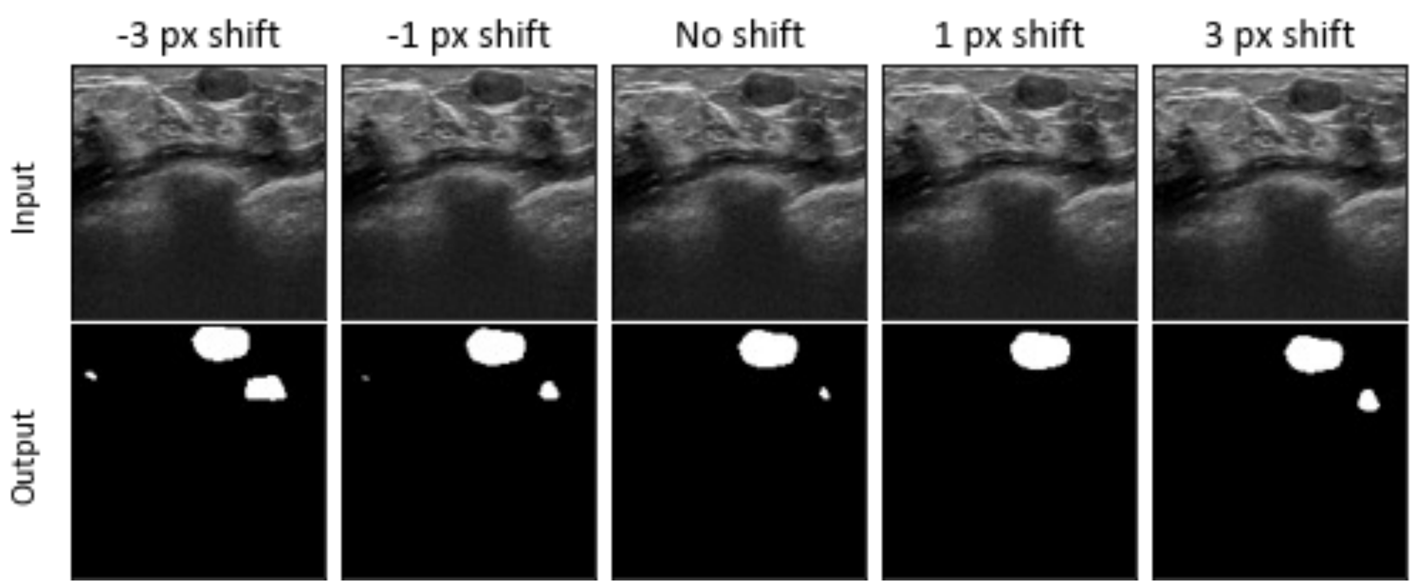
M. Amiri*, R. Brooks, B. Behboodi*, H. Rivaz, Springer IJCARS, 2020

IMPACT Laboratory
Est. 2014
Transforming Healthcare

The Shift-Variance Problem

- Convolutional neural networks (CNNs) have attracted growing interest in automatic ultrasound image segmentation recently.

- However, most modern CNNs are not shift-equivariant.



IMPACT Laboratory
Est. 2014
Transforming Healthcare

The Shift-Variance Problem

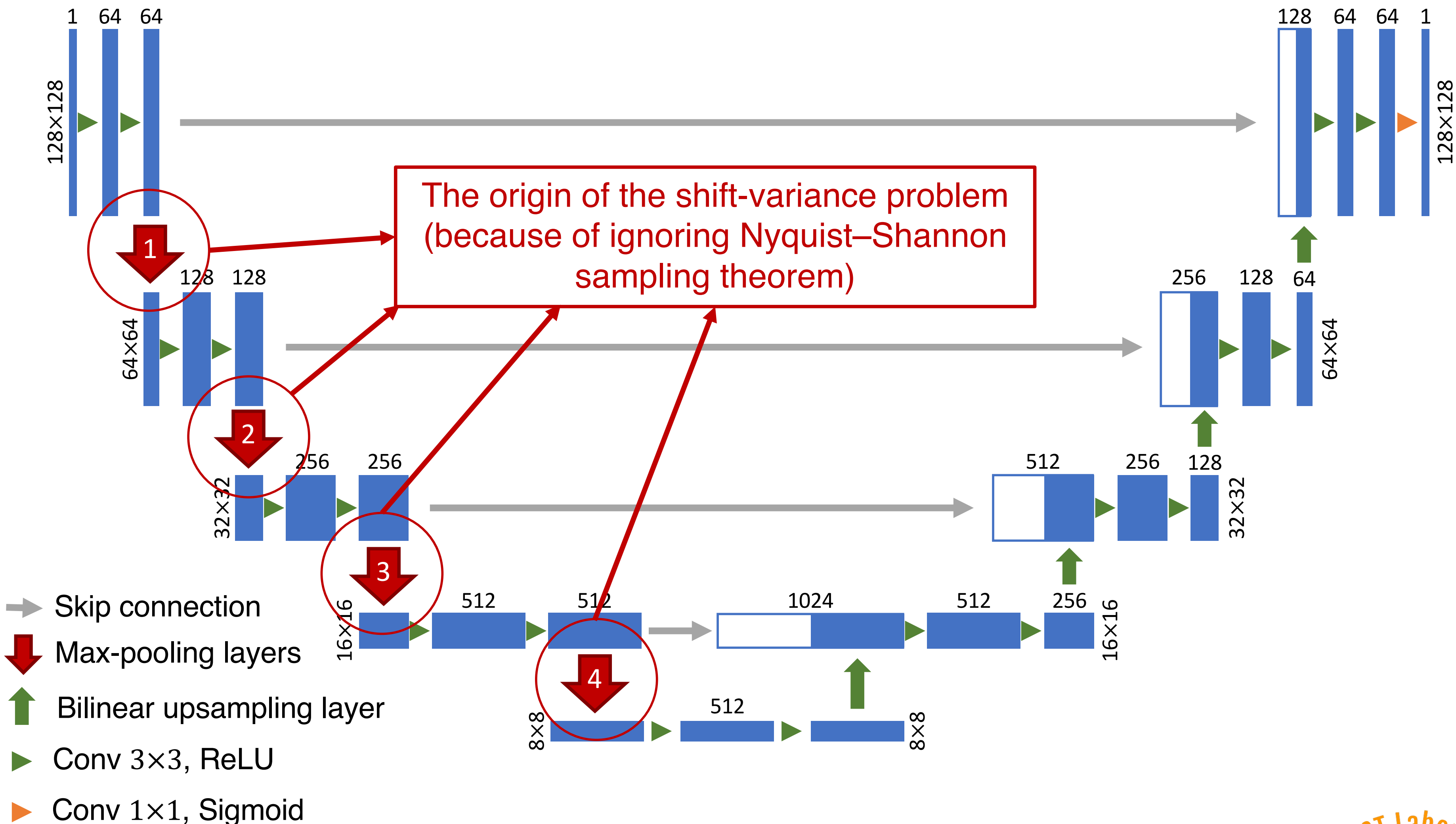
Output consistency is crucial in applications such as monitoring the patients' response to treatment, measuring the progression or regression of the disease, reaching a diagnosis, or treatment planning.

The main focus has primarily been on improving segmentation accuracy metrics.

The output robustness against input translations should not be overlooked, especially in medical applications.

IMPACT Laboratory
Est. 2014
Transforming Healthcare

U-Net as the Baseline



IMPACT Laboratory
Est. 2014
Transforming Healthcare

Methods

Baseline:
To investigate the shift-variance problem in this application for the first time.

1 , 2 , 3 , 4 Max-pool (kernel = 2×2, stride = 2)

BlurPooling[1] $m \times m$:
To evaluate the performance of a recently published method in this application. It integrates anti-aliasing filters to mitigate the shift-variance problem.

1 , 2 , 3 , 4 Max-BlurPool (blur kernel = $m \times m$)

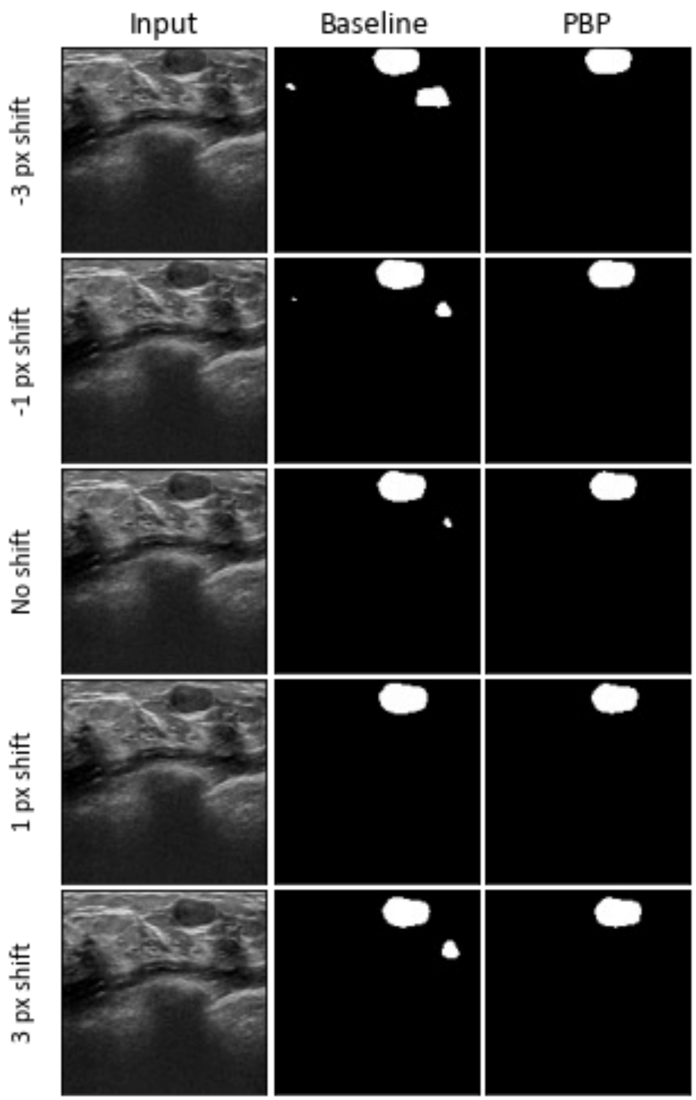
Pyramidal BlurPooling (PBP):
The proposed method.

1 Max-BlurPool 7×7 , 2 Max-BlurPool 5×5 , 3 Max-BlurPool 3×3 , 4 Max-BlurPool 2×2

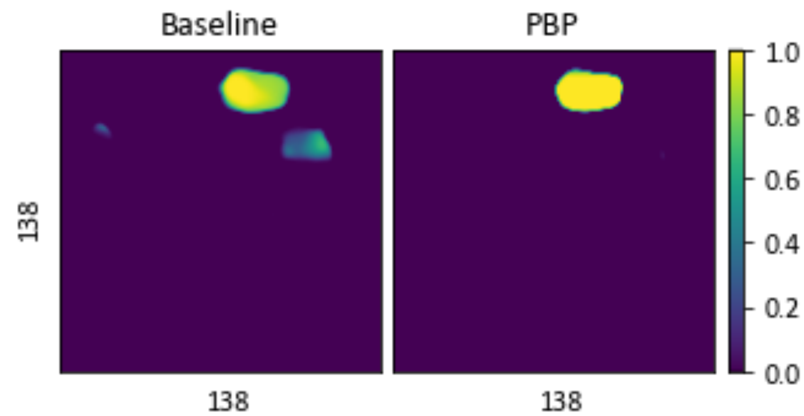
[1]R. Zhang, "Making convolutional networks shift-invariant again," 2019

IMPACT Laboratory
Est. 2014
Transforming Healthcare

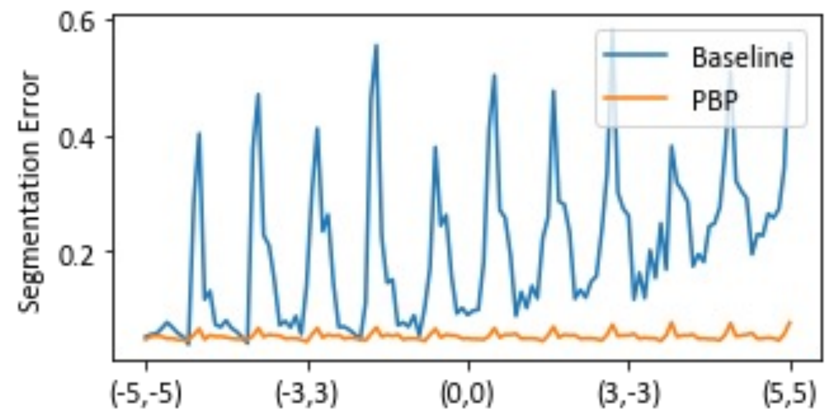
Results



Averaged outputs over 121 input translations (x, y), where $\{x, y \in N \mid -5 \leq x, y \leq 5\}$



Segmentation Error: $(1 - Dice\ similarity\ coefficient)$



(x, y) *shift in pixel*

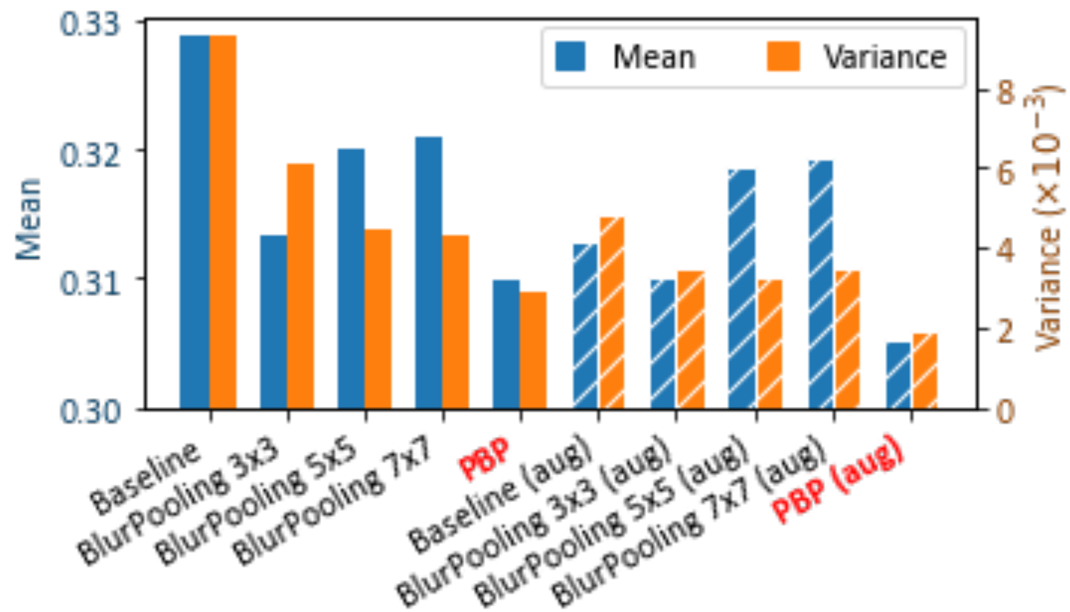
IMPACT Laboratory Est. 2014 Transforming Healthcare

Results

Lower error mean: Higher segmentation accuracy

Lower error variance: Higher output consistency

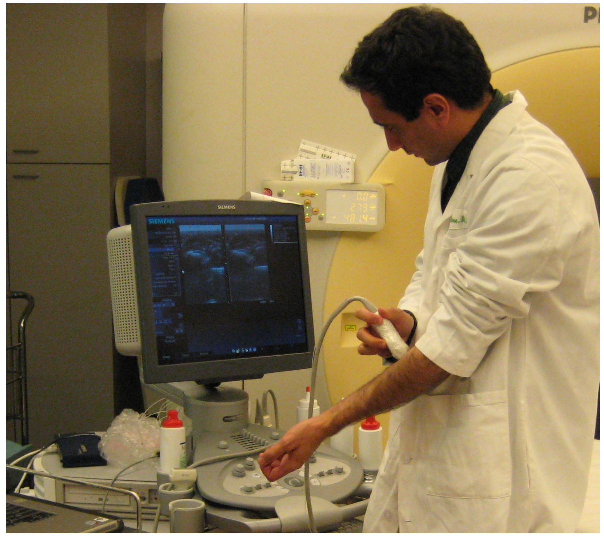
(aug): Repeated with data augmentation

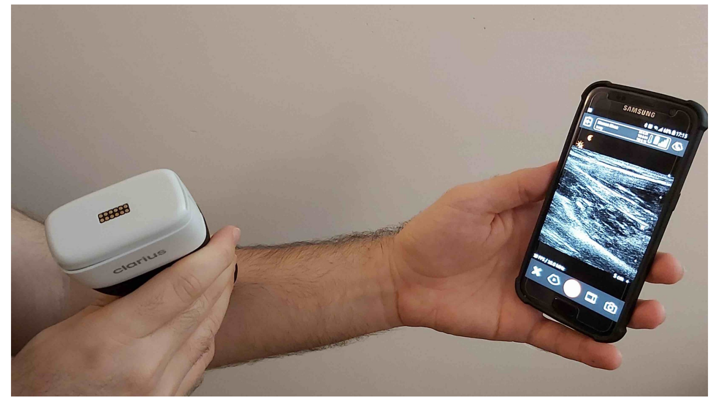


- More details and experiments on additional datasets:
  "Investigating Shift-Variance of Convolutional Neural Networks in Ultrasound Image Segmentation"
  IEEE TUFFC 2022
- Source Code:
  https://git.io/pbpunet

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Bright future for AI and ultrasound

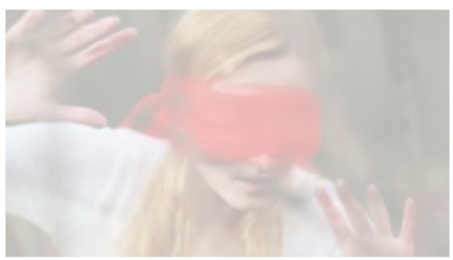




Trump nominee: 'I will not be bullied'

William Barr was also quizzed on why he last year called an aspect of the inquiry "fatally misconceived".

9h | US & Canada



YouTube bans dangerous or harmful pranks

The move comes in response to so-called "challenges" that sometimes resulted in death or injury.

1h | Technology

The ultrasound scan you can do yourself

A US company has produced an ultrasound scanner that plugs into an iPhone and costs $2,000 (£1,555).

3h | Business

Germany steps up monitoring of far right

The move will not yet amount to full-scale surveillance of the nation's main opposition party.

11h | Europe

IMPACT lab:

IMage Processing And Characterization of Tissue

THINK SURGICAL

Richard and Edith Strauss Foundation

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

NSERC CRSNG

Mitacs

Microsoft Azure

NVIDIA

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

Fonds de recherche sur la nature et les technologies Québec

Google

QBIN
QUEBEC BIO-IMAGING NETWORK

Est. 2014

Transforming Healthcare

Collaborators

H. Benali, C. Gauthier, M. Fortin, Y. Xiao, M. Kersten R. Kilgour (Concordia)

M. Boily, A. Towers, P. Martineau, L. Collins (McGill)

T. Hall, I. M. Rosado Méndez (Wisconsin Madison)

G. Cloutier (UdeM)

R. Brooks (Nuance, Microsoft)

P. Brisson, S. Rottoo, M Mirzaei (THINK Surgical)

I. Reinertsen (SINTEF, Norway)

Kris Dickie (Clarius)

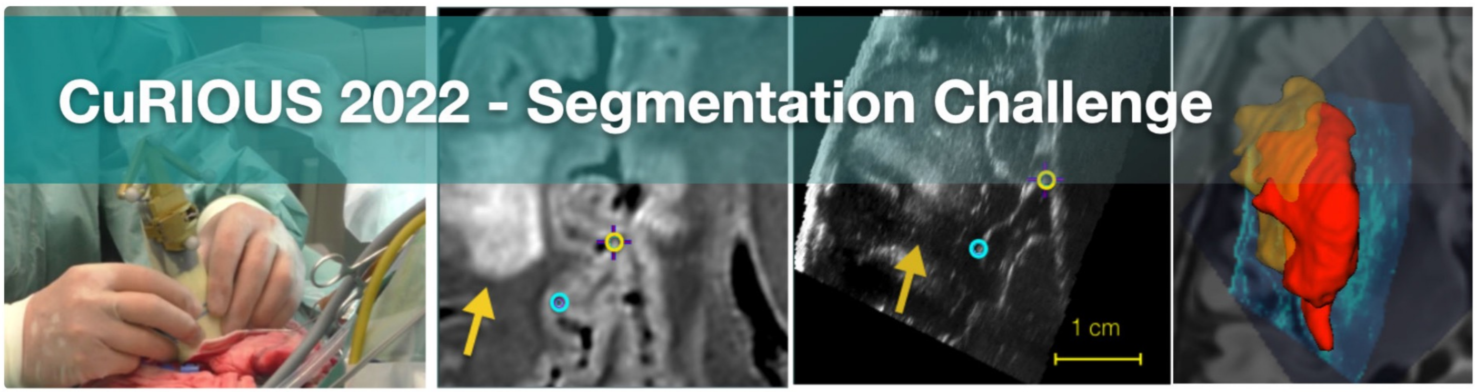
E. Boctor, M. Choti, G. Hager (Johns Hopkins)

IMPACT Laboratory
Est. 2014
Transforming Healthcare

Brain tumor segmentation in ultrasound, MICCAI 2022 Challenge

Grand Challenge Challenges Algorithms ••• Help Sign In Register

Challenges / Brain shift with Intraoperative Ultrasound - Segmentation tasks / Curious2022



Info Join

- Curious2022
- Organizers
- Email organizers

Motivation

Early brain tumor resection can effectively improve the patient's survival rate. However, resection quality and safety can often be heavily affected by intra-operative brain tissue shift due to factors, such as gravity, drug administration, intracranial pressure change, and tissue removal. Such tissue shift can displace the surgical target and vital structures (e.g., blood vessels) shown in pre-operative images while these displacements may not be directly visible in the surgeon's field of view.

https://curious2022.grand-challenge.org/

Est. 2014 IMPACT Laboratory Transforming Healthcare

Thanks!

Contact: sonography.ai

Code.sonography.ai
Data.sonography.ai

IMPACT Laboratory
Est. 2014
Transforming Healthcare