

Transformer Compressed Sensing via Global Image Tokens

Marlon Bran Lorenzana and Dr. Shekhar Chandra

Transformer Compressed Sensing via Global Image Tokens

The Importance of MRI

- No ionising radiation.
 - Measures the density of water molecules.
- Produces high quality images of soft tissue within the body from any imaging direction.
 - Assists in the clinical assessment of patients, used as a source of biomarkers.

Does MRI have Drawbacks?

- Relatively time- and monetarily-expensive.
 - Acquisition is limited by physical and physiological properties.
- MRI is constrained in terms of temporal and spatial resolution.
 - Patient movement can be difficult to account for and may result in artefacts within the image.

Compressed Sensing and MRI

- Data points collected in MRI are obtained from the frequency domain.
 - This constitutes "k-space".
- Sampling all points can be very slow.
 - Accelerated by under-sampling.
 - CS is a mapping from under-sampled to fully-sampled k-space.
- In the simplest configuration, CS solves the following optimisation problem:
 - $\min_x \|\Psi(x)\|_1 + \lambda \|F_u x - y\|_2^2$
 - x : is the reconstructed image.
 - y : are collected k-space samples.
 - F_u : is the under-sampled discrete Fourier transform (DFT).
 - $\Psi(\cdot)$: is image regularisation.
- Image regularisation controls and produces expected image features.

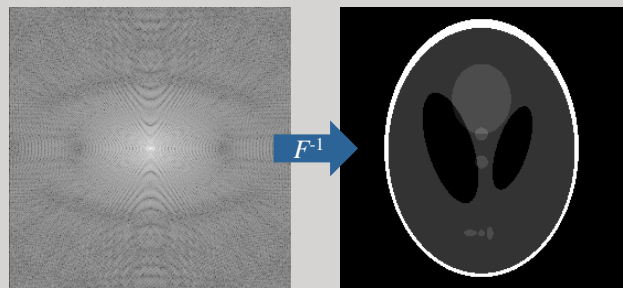
How does Deep Learning (DL) help?

- CS reconstruction can require several minutes of computation per-image.
 - Particularly crucial for dynamic MRI.
 - DL solutions compute in just fractions of a second for a fully reconstructed image.
- Reconstruction quality is limited by choice of reconstruction algorithm.
 - DL reconstruction performance scales with model and data-set size.
 - DL reconstruction has been shown to outperform traditional CS methods.
- There have been several advancements regarding computer vision (CV) and neural networks (NN).
 - Our research compares advantages and disadvantages of convolutional neural networks (CNN), and newer transformer neural networks (TNN) for CV.
 - TNN for CV applications are termed Vision Transformers (ViT).
- This research explores the potential benefits afforded by ViT for CS-MRI, and further, demonstrates a novel ViT architecture via Global Image Tokens.

Fully-Sampled k-space

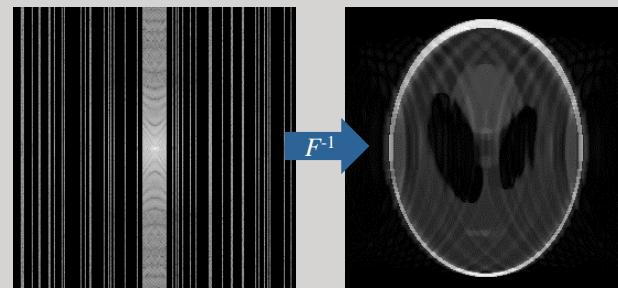
Scan time: 40 min

F^{-1} is Fourier transform

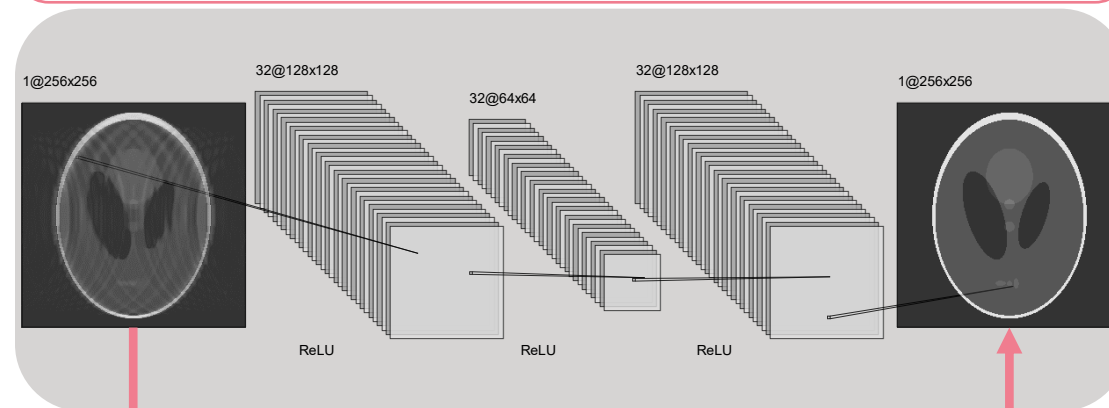


Under-Sampled k-space

Scan time: 10 min



Goal of CS is to develop this map.



Visualisation of DL-CNN mapping from under-sample to fully-sampled.

Transformer Compressed Sensing via Global Image Tokens

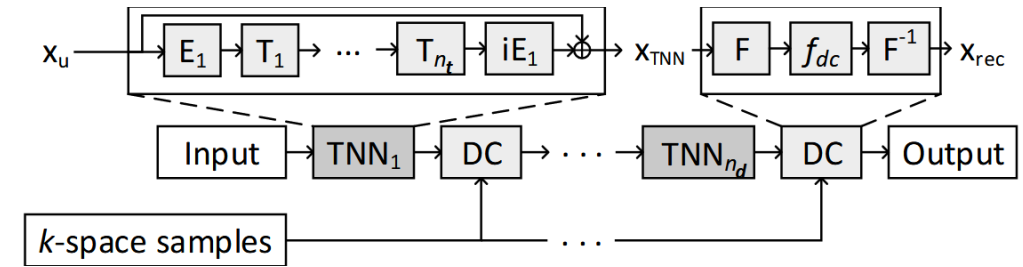
Why Not CNNs?

- Convolutional neural networks (CNN) have demonstrated astounding Compressed Sensing (CS) performance.
- However, they are fundamentally limited in terms of:
 - Generalisability
 - Inductive bias
 - Lack of spatial awareness
 - Difficulty to model long distance relationships.

Hurdles to ViT?

- Patch-Based approach has the following drawbacks:
 - High level of difficulty to capture edges and lines.
 - Sub-optimal model sizes and training time.
 - Presents only a single "scale" of vision.
 - Inherently local, high-frequency features.

Network Architecture



- There are n_d TNN denoising blocks, consisting of n_t ViT encoder layers.
- E is the choice of token embedding e.g. Patch, KD, DRT. iE is the inverse from token to image.
- f_{dc} Performs data consistency to ensure the image is faithful to collected data.

ViT Better than CNN?

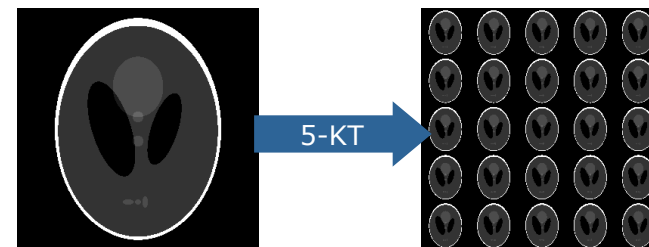
- Transformer networks overcome CNN maladies by implementing an attention mechanism designed to:
 - Capture global dependencies between inputs.
 - Dynamically adjust receptive fields.
 - Scale with large amounts of data.
- ViT has demonstrated state-of-the-art image classification performance.
 - Outperforms similarly sized CNN alternatives.
- Computation blocks are termed "Encoders".

Our ViT Solution

- Adapt the well-known Deep cascade of Convolutional Neural Networks (DcCNN) to use ViT blocks.
 - We call this the DcTNN.
- We employ two novel image token Embeddings:
 - Discrete Radon Transform (DRT) Projections.
 - Kaleidoscope Transform Embeddings (KD).
- We show the improvement to overall image quality compared to using a patch-based approach.
 - Further improve performance by combining multiple token-types.

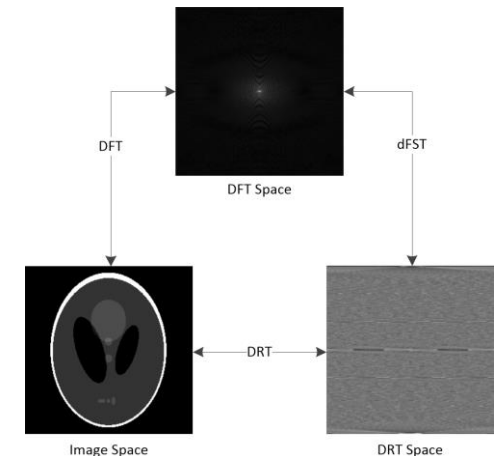
Kaleidoscope Transform

- The Kaleidoscope Transform (KT) provides a global context of the input image.
 - Inherently global, low frequency features.
- Objective is for patch and KD to attend to image features separately:
 - KD Attends to global, low frequencies.
 - Patch attends to local, high frequencies.
- A v -KT decomposes an image into v^2 down-sampled copies.



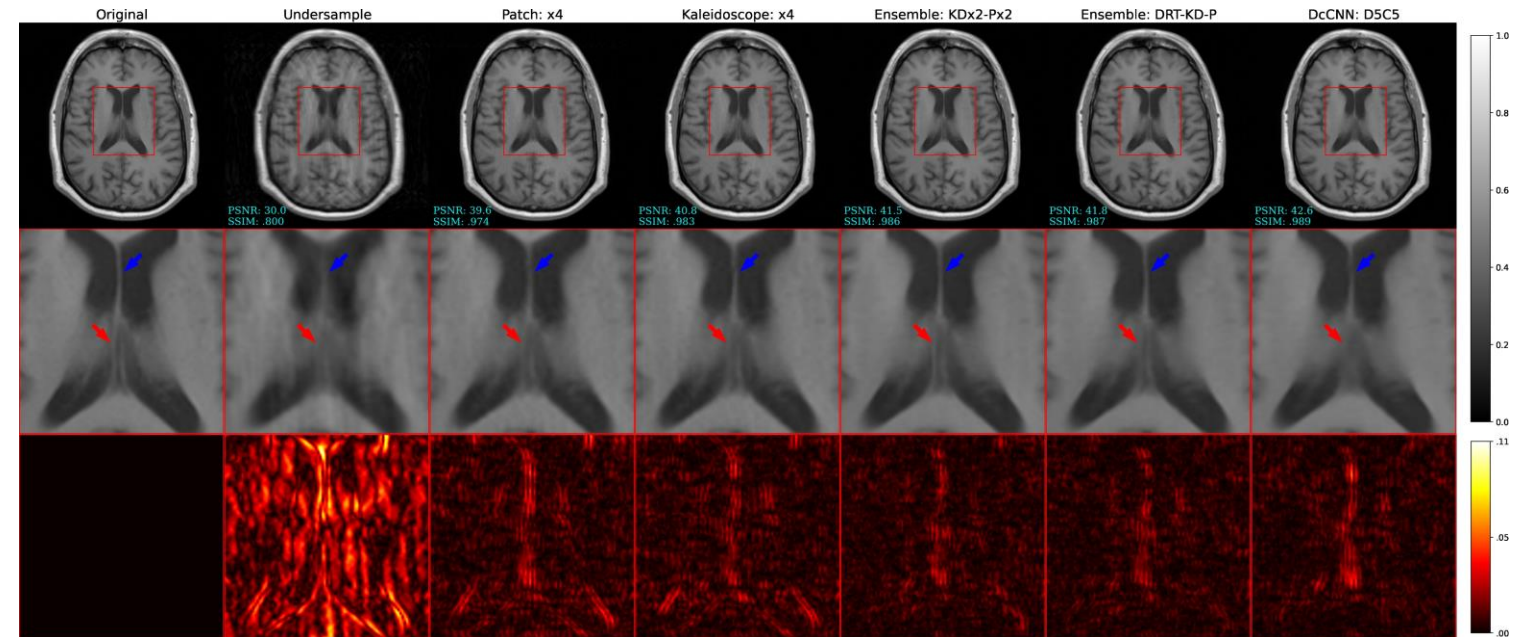
Discrete Radon Transform

- DRT produces approximately orthogonal tokens via a well-established image transform.
- Each token attends globally to the image, at varying frequency scales.
- Transform is similar to Radon transform.



MRI Reconstruction Quality

- Images to the right compare CS performance at 4-fold k-space under-sampling for representative brains.
 - Only 25% of k-space samples are collected.
 - Acquisition time is reduced by x4.
- Power Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) are used to quantify “closeness” to original image.
 - Higher is better.
- We see that transformer networks better understand the difference between under-sample artefacts and brain features.
 - Highlighted in the third row showing the error map.
- Performance significantly improves when using just image patches to KD tokens.
 - Improves again when combined with DRT tokens.



Reconstruction Characteristics

- CNN-based reconstructions demonstrate smoother overall reconstruction.
 - Helps to increase PSNR and SSIM scores.
 - Can flatten texture features within the image.
 - Confuses soft features as noise.
- Transformer-based reconstructions retain much more of the original image texture.
 - Better overall understanding of brain anatomy.
 - More realistic looking images.
- Images are large-scale DICOM from NYU fastMRI database.
 - fastmri.med.nyu.edu

