



170 BPM



Detecting MRI artifacts using deep convolutional neural networks

Reducing costs and improving imaging outcomes through deep learning

The challenge

Magnetic resonance imaging (MRI) is a widely used imaging modality, with almost 30 million scans performed each year throughout the US. Cranial MRI in particular is a mainstay of diagnosis and treatment of neurological disorders, such as stroke, brain tumors, arteriovenous and other vascular abnormalities, multiple sclerosis and encephalopathies, to mention only a few. In addition, magnetic resonance imaging is extensively used for pre-operative planning for neurosurgical interventions.

MRI creates images based on detecting the electromagnetic radiation from protons (hydrogen nuclei) emanated when the spin of protons in a magnetic field decays when the field is turned off. The resulting electromagnetic waves manifest as induction voltage in detector coils, which is then converted into images using inverse Fourier transforms. Due to the extreme sensitivity of this technique, images are often affected by artifacts – distortions or false signals that affect image quality or might even obscure, or masquerade as, clinically relevant signals that derive from the patient's own anatomy, issues with the scanner or issues in the processing software and hardware. Artifacts may adversely affect diagnostic quality, resulting in potential diagnostic errors and the need for costly and time-consuming repeat examinations that may delay timely treatment.

While some artifacts are generalized and can therefore be picked up by simple pixel-level statistics, many artifacts are more complex and require an understanding of the imaging parameters and image context. An example would be a so-called chemical shift artifact, which results from the difference between the resonant frequencies of fat and water, and manifests as a shift in the frequency-encode direction. Such an artifact might mimic a solid body pathology, such as a tumor. Detecting such artifacts has hitherto been a significant challenge.

Our approach

Through our collaboration with the Centre for Brain Imaging at the Hungarian Academy of Sciences and a range of academic medical providers with scanning facilities, a large number of MRI images were obtained, both from healthy volunteers and from clinical cases. Starschema designed a convenient and highly performant platform for submitting image series directly from a PACS connector, performing anonymization and removal of PHI for compliance, then committing the images to storage and, eventually, analysis.

A large training set was hand-annotated by qualified radiologists with experience in brain MR imaging, comprising a wide and balanced set of pathological and non-pathological images alike, across a wide range of MR submodalities (with the exception of magnetic resonance spectroscopy). Based on these images, a deep convolutional neural network was constructed in TensorFlow, which was then trained on NVIDIA TPUs. Processing code was initially written in OpenCV but optimized for fast and efficient execution in C++ and CUDA. Through the use of energy-aware pruning (Sze, Yang and Chen, 2017.) and Frankle-Carbin pruning (Frankle and Carbin, 2018.), the overall size of the network was reduced to a quarter of the equivalent fully-connected convolutional feed-forward architecture, all the while maintaining accuracy through training and combining successful subnetworks. The resultant network is small enough to deploy on devices with limited GPU memory, such as MRI workstations.

The scoring workflow was designed to provide a 'drill-down' capability. An overall image quality and diagnostic suitability index was calculated for each image. In addition to this, users could review both the types of artifacts present in the image, and their locations annotated on the image. This allows clinicians to understand not merely the issues with the image quality, but also permits them to assess whether an artifact affecting a part of an image would nonetheless retain the image's diagnostic suitability, e.g. where the region of diagnostic interest is not affected by the artifact. When used in conjunction with the MRI operator's workstation, artifacts can be detected at time of scan and advice can be provided to the operator as to possible approaches to avoid the artifact, such as

adjustments to the sequence parameters or ensuring that interfering signals are absent and magnet room shielding is intact. This reduces costly patient recalls, avoids unnecessary repeat contrast load in sensitive patients (e.g. patients with kidney disease) and facilitates timely access to appropriate treatment by getting the scan right the first time, every time.

Results

The final model provides an average IoU (intersection over union or Jaccard index) of 0.93 (average of all artifact types when weighted over their relative frequency in the clinical sample) and a classification ROC AUC of 0.90 for diagnostic suitability of images. The pruned model provides an IoU of 0.90 and a classification ROC AUC of 0.88, respectively. This rivals the accuracy of trained diagnosticians and is sufficient to serve as a valuable aid to the clinical radiology workflow. In particular in the emergency medicine setting, where time is of the essence, a rapid indication of the likelihood that the image might not be diagnostically suitable can save lives.

When run on a dual NVIDIA Tesla K80, the full model evaluates a 32-slice 384x512 matrix size series in approximately 45 seconds, which is a fraction of the scan times of even the fastest rapid parallel imaging scan with a 32-channel coil. Since images can also be submitted for evaluation individually, the evaluation can run contemporaneously with image acquisition, allowing the scan sequence to be stopped if artifacts are present and avoiding a wasted sequence.

The fully containerized and encrypted system can be deployed on a range of cloud vendors, including AWS, Azure and HIPAA-compliant specialist vendors as a client that integrates with radiology software, or deployed on-premise. The highly efficient pruned model means that on-premise deployment alongside the operator workstation is possible without the capital expenditure on expensive computing hardware and without significantly sacrificing accuracy: even on lower-end GPUs, rapid evaluation of image quality is possible using the pruned model, and with more powerful hardware, near real-time evaluation at high accuracy can be achieved using an entirely on-premise, HIPAA compliant solution that provides uncompromising quality without any patient data leaving the premises.

Technologies used

- Python
- TensorFlow
- NVIDIA CUDA running on NVIDIA TPUs
- OpenCV
- Orthanc

Skills used

- Image analysis and computer vision
- Image source data augmentation
- Deep convolutional neural networks