Wide-Range MRI Artifact Removal with Transformers
Lennart Alexander Van der Goten  Kevin Smith
KTH Royal Institute of Technology, SciLifeLab Stockholm

Abstract
Artifacts on magnetic resonance scans are a serious challenge for both radiologists and computer-aided diagnosis systems. Most commonly, artifacts are caused by motion of the patients, but can also arise from device-specific abnormalities such as noise patterns. Irrespective of the source, artifacts can not only render a scan useless, but can potentially induce misdiagnoses if left unnoticed. For instance, an artifact may masquerade as a tumor or other abnormality.

Retrospective artifact correction (RAC) is concerned with removing artifacts after the scan has already been taken. In this work, we propose a method capable of retrospectively removing eight common artifacts found in native-resolution MR imagery. Knowledge of the presence or location of a specific artifact is not assumed and the system is, by design, capable of undoing interactions of multiple artifacts. Our method is realized through the design of a novel volumetric transformer-based neural network that generalizes a window-centered approach popularized by the Swin transformer. Unlike Swin, our method is (i) natively volumetric, (ii) geared towards dense prediction tasks instead of classification, and (iii) uses a novel and more global mechanism to enable information exchange between windows. Our experiments show that our reconstructions are considerably better than those attained by ResNet, V-Net, MobileNet, DenseNet, CycleGAN, and BicycleGAN. Moreover, we show that the reconstructed images from our model imports the accuracy of FSL BET, a standard skull-stripping method typically applied in diagnostic workflows.

Overview
Artifacts are a topic of tremendous significance in radiology. For instance, motion artifacts can render an MRI scan totally useless, prompting a costly re-scan and causing further patient discomfort. Other artifacts are caused by the nature of the MRI device, e.g. due to inhomogeneities of the magnetic field, noise or perturbations in the frequency domain.

Naturally, removing artifacts retrospectively is of high practical interest. Modern artifact simulation frameworks (e.g. Sim4Lak) can synthesize a wide variety of highly-realistic MR artifacts. We apply said simulator on the highly-curated ADNI dataset and seek to find a transformer-based architecture that learns the inverse process. Within the paper, we suggest W-G2L-ART, a window-based vision transformer that rethink the window-to-window communication suggested by the Swin Transformer (Liu et al., 2021) to enable a more global communication between windows.

We refer to the suggested window-to-window attention mechanism as G2L-MSA. It can be implemented using the G2L function as depicted in Figure 3. Intuitively, it partitions an input matrix \( Z \) into non-overlapping windows of a fixed size (here \( 2 \times 2 \)) and collects values at equal positions over the set of all windows. Thus, "global" information is contracted into a local context. The resulting tensor can then easily be processed using a common window-based attention. The overall operation can be implemented efficiently with elementary operations such as transpositions and reshapings.

Related Work
- \textsc{DANN} (Hosseini et al., 2018): Assumes that patient motion is rigid, solves equations to remove degradation.
- \textsc{DeSim} (Liu et al., 2021): CycleGAN-inspired approach that translates between artifact-free and artifact-affected domains, considers three artifact types, slice-based.
- \textsc{DANN} (Yan et al., 2020): Synthesizes a full MRI scan from an undersampled one.
- Other papers (e.g. Liu et al., 2021) address inverse problems on planar (i.e. 2D) images.

Architectural Overview – W-G2L-ART

In this work, we propose W-G2L-ART, a retrospective artifact correction (RAC) system using volumetric vision transformers with a novel global window-to-window communication mechanism. Our experiments show that W-G2L-ART is able to effectively correct eight different artifact types, compares favorably against the benchmark methods according to PSNR/SSIM, and gives the best results when corrected images are used for skull-stripping. Future work will address larger and more diverse datasets, including different modalities such as CT scans.

Acknowledgements
This work was partially supported by the Swedish Research Council (VR) 2017-04609.

Experiments
- We use Sim4Lak (Pérez-García et al., 2021) to generate eight different artifact types.
- An artifact is applied with probability 1/9 for \( N = 9 \).
- Comparison: ResNet, DenseNet, V-Net, CycleGAN, BicycleGAN.
- Ablation: W-SW-MSA-ART (W-G2L-ART with SW-MSA attention).
- Model sizes are scaled to fully utilize memory of RTX 3090.

Conclusion & Future Work

British Machine Vision Conference 2022, London UK
https://bmvc2022.org