DOI: 10.1002/alz70862_109817

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

BIOMARKERS

POSTER PRESENTATION

NEUROIMAGING

AI Superresolution: Converting T1-weighted MRI from 3T to 7T resolution toward enhanced imaging biomarkers for Alzheimer's disease

Malo Gicquel[1] | Gabrielle Flood[2] | Ruoyi Zhao[1] | Anika Wuestefeld[3] | Nicola Spotorno[3] | Olof Strandberg[3] | Yu Xiao[1] | Kalle Åström[2] | Laura E.M. Wisse[4] | Danielle van Westen[4] | David Berron[5,6] | Oskar Hansson[3] | **Jacob W. Vogel**[1]

[1]Department of Clinical Sciences Malmö, SciLifeLab, Lund University, Lund, Sweden

[2]Centre for Mathematical Sciences, Lund University, Lund, Sweden

[3]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden

[4]Diagnostic Radiology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden

[5]Center for Behavioral Brain Sciences (CBBS), Otto-von-Guericke University Magdeburg, Magdeburg, Sachsen Anhalt, Germany

[6]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany

Correspondence
Jacob W. Vogel, Department of Clinical Sciences Malmö, SciLifeLab, Lund University, Lund, Sweden.
Email: jacob.vogel@med.lu.se

Abstract

Background: High-resolution (7T) MRI facilitates in vivo imaging of fine anatomical structures selectively affected in Alzheimer's disease (AD), including medial temporal lobe subregions. However, 7T data is challenging to acquire and largely unavailable in clinical settings. Here, we use deep learning to synthesize 7T resolution T1-weighted MRI images from lower-resolution (3T) images.

Method: Paired 7T and 3T T1-weighted images were acquired from 178 participants (134 clinically unimpaired, 48 impaired) from the Swedish BioFINDER-2 study. To synthesize 7T-resolution images from 3T images, we trained two models: a specialized U-Net, and a U-Net mixed with a generative adversarial network (U-Net-GAN) on 80% of the data. We evaluated model performance on the remaining 20%, compared to models from the literature (V-Net, WATNet), using image-based performance metrics and by surveying five blinded MRI professionals based on subjective quality. For $n = 11$ participants, amygdalae were automatically segmented with FastSurfer on 3T and synthetic-7T images, and compared to a manually segmented "ground truth". To assess downstream performance, FastSurfer was run on $n = 3{,}168$ triplets of matched 3T and AI-generated synthetic-7T images, and a multi-class random forest model classifying clinical diagnosis was trained on both datasets.

Result: Synthetic-7T images were generated for images in the test set (Figure 1A). Image metrics suggested the U-Net as the top performing model (Figure 1B), though blinded experts qualitatively rated the GAN-U-Net as the best looking images, exceeding even real 7T images (Figure 1C). Automated segmentations of amygdalae from the synthetic GAN-U-Net model were more similar to manually segmented

amygdalae, compared to the original 3T they were synthesized from, in 9/11 images (Figure 2). Classification obtained modest performance (accuracy~60%) but did not differ across real or synthetic images (Figure 3A). Synthetic image models used slightly different features for classification (Figure 3B).

Conclusion: Synthetic T1-weighted images approaching 7T resolution can be generated from 3T images, which may improve image quality and segmentation, without compromising performance in downstream tasks. This approach holds promise for better measurement of deep cortical or subcortical structures relevant to AD. Work is ongoing toward improving performance, generalizability and clinical utility.

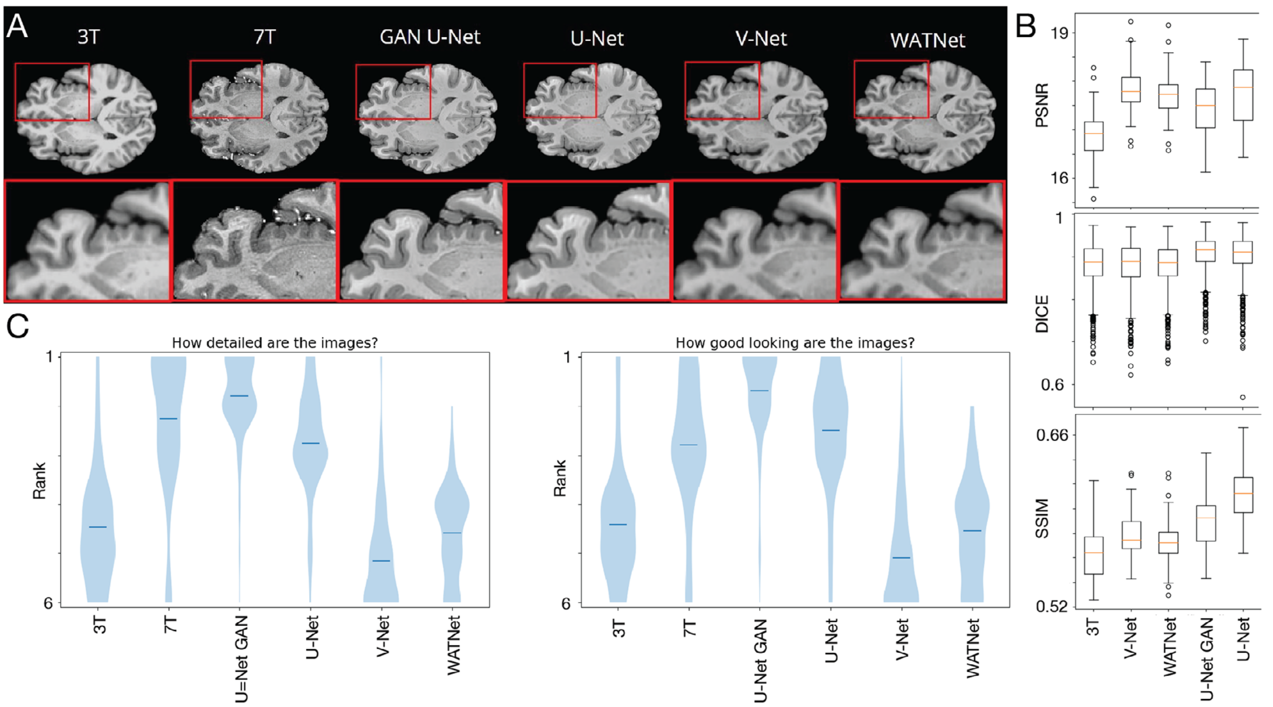

Figure 1: Objective and subjective evaluation of synthetic T1-weighted 7T images. Synthetic 7T images were generated from 34 3T images from the "test set" using four different deep learning models – two from this study (GAN U-Net and U-Net) and two from the literature (V-Net and WATNet). **A)** Example of synthetic images from each model. Images show a 3T and 7T image of a single subject (who was not used in model training), in addition to synthetic images from each of the four models. Top row shows a full axial slice, bottom row zooms in on the right frontal lobe. **B)** Image evaluation metrics compared the 3T and the synthetic 7Ts to the true 7Ts of subjects in the test set: peak signal-to-noise ratio (PSNR), Dice coefficient (DICE), structural similarity index (SSIM). Higher scores indicate better performance for each metric. The SSIM and PSNR were based on direct image-to-image comparison, whereas DICE was based on comparing segmentations. The median performance of the U-Net was best across all metrics, indicating it as the most similar to the true 7T. **C)** A set of five radiologists and MRI scientists looked at sets of six images from each of the 34 participants in the test set (the 3T, the 7T, and the synthetic 7T from each of the four models), and were asked to rank the images 1 to 6 based on how detailed they looked, and how "good" they looked (this question was left intentionally vague). The U-NET GAN consistently ranked best in both categories (likely due to 7Ts often featuring prominent artifacts). V-NET and WATNet images were ranked similarly to "true" 3T images.

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

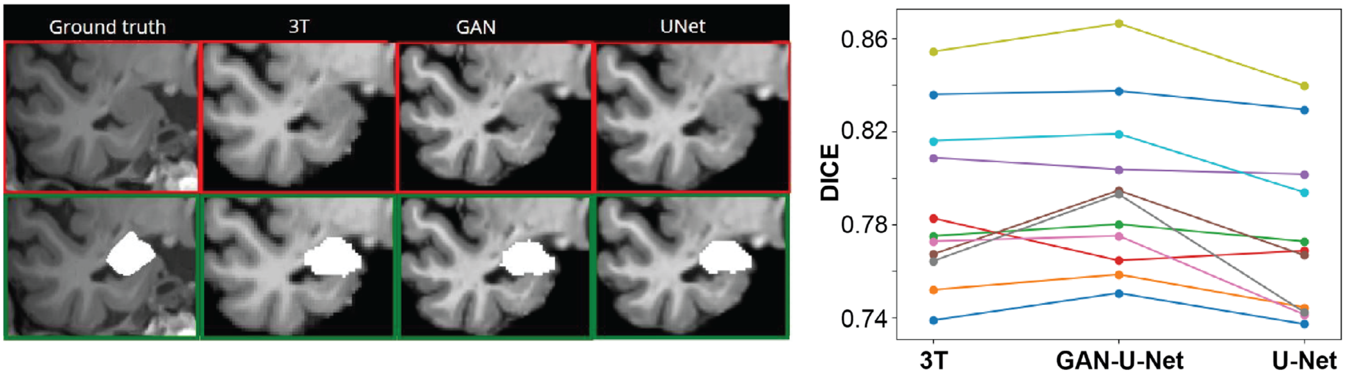

Figure 2: Synthetic translation to 7T improves automated segmentation of medial temporal lobe structures. Left and right amygdalae were manually segmented for 11 subjects, serving as a "ground truth" 3T segmentation. "FastSurfer" performed automated segmentation on each subject's 3T, and matched synthetic 7T images from the GAN and U-Net models. DICE coefficient compared the automated segmentation of the amygdala from each image to the "ground truth" manual segmentation. In 9 of 11 cases, the GAN-U-NET achieved the best DICE score (indicating the segmentation was closer to the manually segmented ground truth). The regular U-NET showed improvement over the 3T automated segmentation in 0 of 11 cases.

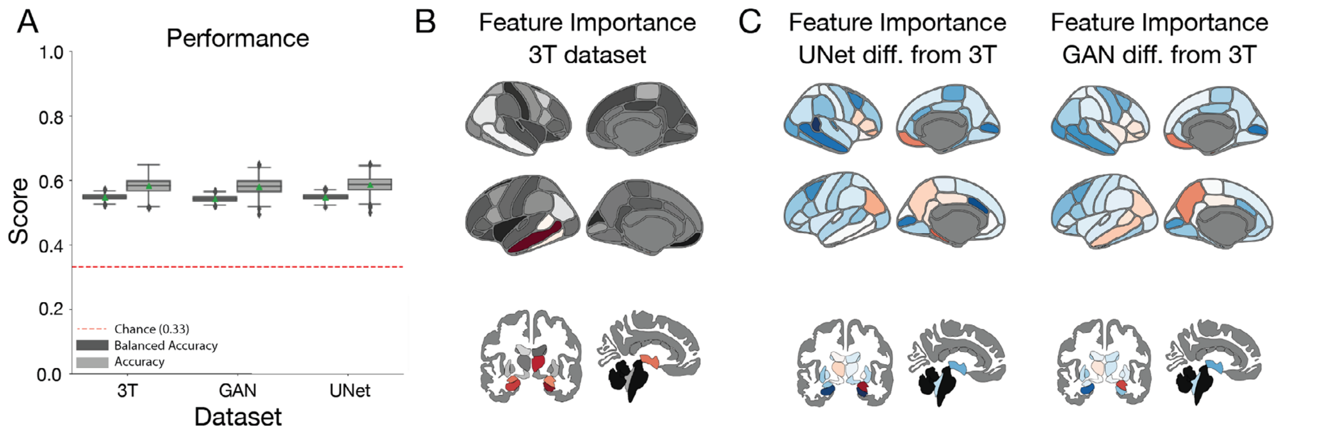

Figure 3: Synthesizing images does not compromise performance on downstream tasks but uses slightly different features. The U-Net and GAN-U-NET models were used to synthesize 7T images from 3,168 3T images, each of which had a clinical diagnosis of cognitively normal (CN), mild cognitive impairment (MCI) or Alzheimer's disease dementia (AD). Each dataset (real 3T, U-Net synthetic-7T, GAN-U-Net synthetic 7T) was automatically parcellated with FastSurfer and the cortical thickness derivatives and subcortical volumes were entered with age and sex into a multiclass random forest classifier predicting clinical diagnosis. **A)** Boxplots show accuracy and balanced accuracy scores ("Score" on y-axis) across 1000 train-test splits. Model performance was equivalent across the three datasets. **B)** Brains showing the regional mean feature importance across the 1000 models trained on the 3T dataset. Deeper red regions were more important for making predictions, whereas darker gray regions were less important. **C)** Plots show deviation in feature importance between models trained on synthetic data vs. models trained on the 3T data. Synthetic images seemed to rely more on medial temporal, temporal and parietal regions, especially in the left hemisphere.