Effects of defacing whole head MRI on neuroanalysis

Chenyu Gao, Linghao Jin, Jerry Prince, Aaron Carass

Chenyu Gao, Linghao Jin, Jerry L. Prince, Aaron Carass, "Effects of defacing whole head MRI on neuroanalysis," Proc. SPIE 12032, Medical Imaging 2022: Image Processing, 120323W (4 April 2022); doi: 10.1117/12.2613175

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Effects of Defacing Whole Head MRI on Neuroanalysis

Chenyu Gao[a,*], Linghao Jin[b,*], Jerry L. Prince[a,b,c], and Aaron Carass[c]

[a]Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD 21218
[b]Department of Computer Science, The Johns Hopkins University, Baltimore, MD 21218
[c]Department of Electrical and Computer Engineering,
The Johns Hopkins University, Baltimore, MD 21218

ABSTRACT

Recent advances in magnetic resonance (MR) scanner quality and the rapidly improving nature of facial recognition software have necessitated the introduction of MR defacing algorithms to protect patient privacy. As a result, there are a number of MR defacing algorithms available to the neuroimaging community, with several appearing in just the last five years. These various approaches have qualities that have been explored with respect to skull stripping masks or identifiability of the patient in previous works. However, to our knowledge there has been no evaluation of the subsequent impact of these defacing algorithms on a neuroimaging pipeline. In this work, we use six MR defacing algorithms on 179 subjects from the OASIS-3 cohort and 21 subjects from the Kirby 21 dataset, then apply a neuroimaging pipeline to the resultant defaced images. We compare the consistency of the output from the pipeline using the defaced images with the output of the same pipeline without defacing the MR data.

Keywords: MRI, neuroimaging, defacing, skull stripping

1. INTRODUCTION

In the last decade, magnetic resonance (MR) images (MRIs) have seen improvements in signal-to-noise ratio and resolution that have led to improved visualization of whole head images. These improvements, as well as the rapid development of deep learning and facial recognition software, have introduced the concern that patients may be identifiable from their MRIs. In conjunction with these privacy concerns, there has been a wider community effort towards data sharing of ever increasing amounts of medical studies. The removal of protected health information (PHI) is of great importance to preserve and respect an individuals privacy. Textual information (e.g. names, sex, etc.) are readily removed by several software packages or upon export from a Picture Archiving and Communication System (PACS), but the actual images are increasingly being considered to be PHI due to the potential to identify a study participant based on reconstruction of the facial features from MRI. Studies that have managed to identify study participants have been reported in the literature. For example, Schwarz et al.[1] demonstrated photographic images of participants could be matched to the 3D reconstructed MRIs with an 80+% accuracy. Given the ubiquity with which people post to FaceBook (f) and Instagram (◎), it may be reasonable to assume that all publicly available MRI data is identifiable given sufficient time and resources.

This paper will not address the question of how feasible such a data breach is. Instead, we explore the existing tools that are available to avoid such data leaks. We are focused on the defacing algorithms which are meant to remove such PHI from the MRIs, see Figure 1 for examples of defaced MRIs and the resultant 3D reconstructions. These defacing algorithms operate by obscuring or removing recognizable portions of the face from the MRI thus reducing the utility of the 3D reconstruction for identification purposes. There has been some initial work[1–3] concerned with the qualities of such defacing algorithms. We, however, are exclusively concerned with the impact such defacing algorithms have on subsequent data processing. Since that is ultimately the goal of acquiring the MRIs in the first place. As such this paper explores the impact of defacing on processing whole head MRIs through a standard neuroimaging pipeline.[4,5] To our knowledge, this is the first such paper to investigate the potential ramifications on data processing of such defacing algorithms.

*: These authors contributed equally to this work and should be considered co-first authors.
Further author information: (Send correspondence to Chenyu Gao)
E-mail: cgao17@jhu.edu

Medical Imaging 2022: Image Processing, edited by Olivier Colliot,
Ivana Išgum, Proc. of SPIE Vol. 12032, 120323W · © 2022 SPIE
1605-7422 · doi: 10.1117/12.2613175

We include six defacing algorithms in our experiments: Defacer,[6] Quickshear,[7] MRI_Deface,[8] Pydeface,[9] FSL_deface,[10] and mri_reface.[2] After applying these six defacing algorithms on T1-weighted MR images of 179 different subjects from OASIS-3 dataset, we feed the defaced image, as well as its corresponding original image, into the pipeline and quantify the difference in their outputs. We also use the Kirby 21 dataset, which contains scan-rescan MRI data from 21 subjects, to answer the question: are the effects of these defacing algorithms more significant than the effects of rescanning a subject followed by registration? Despite the fact that the effects of defacing on post-processing are unignorable, the effects of rescanning and registration are comparatively even more significant. We also note that all currently available defacing algorithms have the aim of preserving the whole brain. Unfortunately, they can remove portions of the cerebrum and more importantly they are destructive of MRI data around the nose, jaw, and tongue, making processing[11,12] of defaced images for related studies impossible.

2. METHODS AND DATA

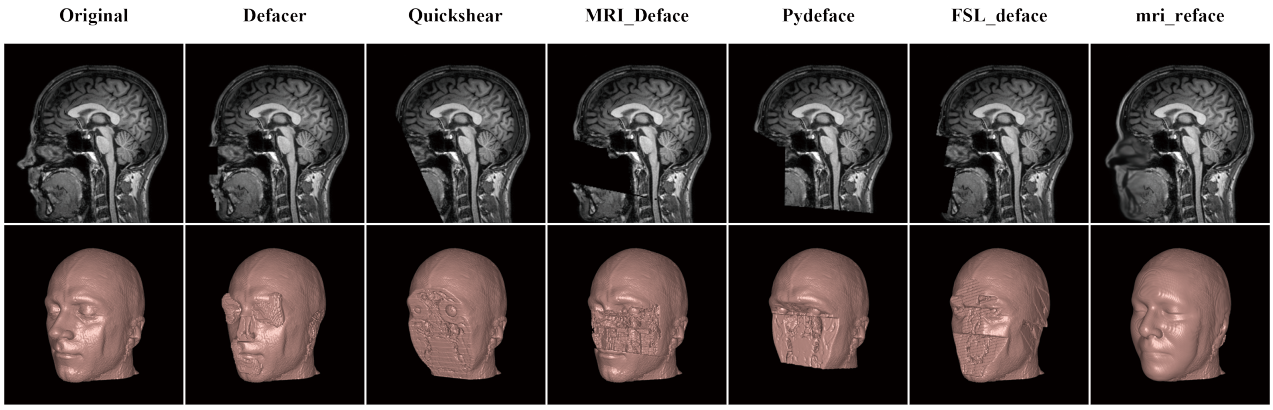

Figure 1. The top row shows a sagittal view of the magnetic resonance images (MRIs). The bottom row shows the 3D reconstruction of the corresponding MRIs in 3D Slicer.[13] The columns from left to right depict: the original MRI and its reconstruction, output of Defacer,[6] Quickshear,[7] MRI_Deface,[8] Pydeface,[9] FSL_deface,[10] and mri_reface.[2]

2.1 Methods

Here we outline the six defacing algorithms that are included in our experiments.

`Defacer`[6]: 3D U net-based network is trained to detect eyes, nose, mouth, and ears. After detection, the program proceed with the anonymization process by changing the values of the voxels of these facial features. We use the version distributed on November 19$^{\text{th}}$, 2020.

`QuickShear`[7]: Requires a brain mask as input. It creates an edge of the brain from the mask collapsed to the sagittal plane.[14] Based on the edge, a shearing plane is then identified. All voxels on the face side of the plane will be removed. The space between the plane and the brain mask is a tweakable parameter. We use version 1.1.0.

`MRI_Deface`[8]: For each input image, FLIRT[15] is used to register a template image and two corresponding template masks. The defacing mask is formed from the union of all voxels with nonzero probability of being brain and then morphologically dilating them. All voxels outside the mask with a nonzero probability of being a facial feature are set to zero. We use version 1.22.

`PyDeface`[9]: Uses FLIRT[15] to align an MRI template with the target structural scan and then applies the templates mask to remove face voxels. We use version 2.0.0.

`FSL_deface`[10]: Uses FLIRT[15] to align a template image with input image and then applies a pre-defined mask to remove face voxels on the input. We use the version released with FSL 6.0.3.

`mri_reface`[2]: Faces are replaced with a population average face, rather than removed, to better resemble a natural image. We use version 0.2.

We use the spatially localized atlas network tiles (SLANT)[16] method to segment the images pre- and post-defacing. SLANT deploys multiple spatially distributed networks, in which each network learned contextual information for a fixed spatial location. These independent networks allow for the fast segmentation of images and have been developed so that fusing the various network results after the fact is straightforward.

2.2 Data

We use two datasets in our research, OASIS-3 and Kirby 21.

`OASIS-3`[17]: OASIS-3 is the latest release in the Open Access Series of Imaging Studies (OASIS) that is aimed at making neuroimaging datasets freely available to the scientific community. By compiling and freely distributing this multi-modal dataset generated by the Knight ADRC and its affiliated studies, it is hoped to facilitate future discoveries in basic and clinical neuroscience. OASIS-3 is a longitudinal neuroimaging, clinical, cognitive, and biomarker dataset for normal aging and Alzheimer's Disease. We use 179 subjects randomly selected from the OASIS-3 cohort. From each subject, we pick one T1-weighted image for our experiments.

`Kirby 21`[18]: Kirby 21 contains scan-rescan imaging sessions on 21 healthy volunteers (no history of neurological disease). Imaging modalities include MPRAGE, FLAIR, DTI, resting state fMRI, B0 and B1 field maps, ASL, VASO, quantitative T1 mapping, quantitative T2 mapping, and magnetization transfer imaging. All data have been converted to NIFTI format. We use the MPRAGE images from the 21 subjects for our experiments.

3. EXPERIMENTS

3.1 Quantify the Effects of Defacing

We use 179 T1-weighted MR images from OASIS-3 for this part of the experiments. When applying defacing algorithms on MR images, we use the default parameters, if there is any option, for all algorithms. We ran SLANT on the original data (pre-defacing) and then again on the output of each of the defacing algorithms. We regard the pre-defacing SLANT result as the ground truth and compare the post-defacing results to that ground truth by computing the dice similarity coefficient (DSC)[19,20] for each of the 132 labels output by SLANT,[4,5] excluding the background.

3.2 Quality Check

After defacing the 179 MR images from OASIS-3 using all six algorithms, we perform a slice-by-slice quality check manually and categorize the outputs into three classes: i) Success: the algorithm processes the MR image as expected, although some facial voxels that are supposed to be removed may remain; ii) Failure I: the algorithm failed to detect facial features or run properly. Consequently, the face remains untouched and recognizable; iii) Failure II: a certain proportion of the brain is damaged due to excessive defacing.

3.3 Compare Defacing with Scan-Rescan

We use Kirby 21, which contains scan and rescan data from 21 subjects, for this part of the experiments. To avoid confusion, here we use 'session-1' to represent the first scan data, and 'session-2' to represent the rescan data. For each subject, we register session-1 to session-2, and register session-2 to session-1, using weighted affine transformation with binary brain mask as the weighting. We then apply six defacing algorithms on session-1 respectively. Subsequently, we ran SLANT on the original session-1 and session-2, defaced session-1, and registered session-1 and session-2. Similar to what we did in section 3.1, we quantify the effects of defacing by computing the dice similarity coefficient (DSC) between pre-defacing SLANT result and post-defacing SLANT results. In addition, we compute the DSC between SLANT results from session-1 and registered session-2, and between SLANT results from session-2 and registered session-1. We note that the DSCs are computed on the whole brain, excluding non-brain pixels.

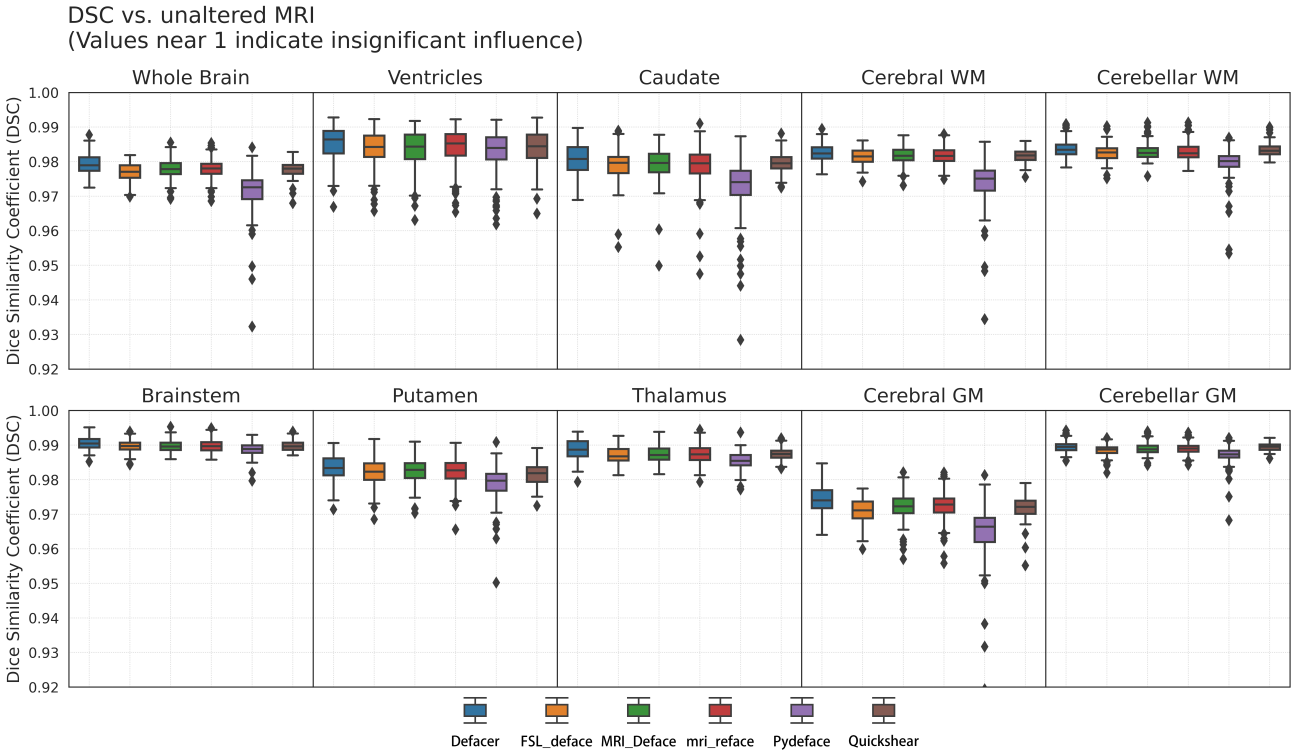

Figure 2. Dice similarity coefficient (DSC) between SLANT segmentation result from defaced MRI and SLANT segmentation result from original MRI.

Table 1. The results of the quality check. See section 3.2 for the definition of Success, Failure I, and Failure II.

Algorithm	Total	Success	Failure I	Failure II
`Defacer`	179	163	16	0
`FSL_deface`	179	142	0	37
`MRI_Deface`	179	136	7	36
`mri_reface`	179	179	0	0
`Pydeface`	179	179	0	0
`Quickshear`	179	78	0	101

4. RESULTS

4.1 Effects of Defacing: OASIS-3 Results

We quantify the effects of defacing MR images on the post-processing pipeline, here the SLANT segmentation, by computing the dice similarity coefficient (DSC) between outputs from defaced and unaltered MRI. For better visualization, we merge the 132 regions of interest (ROI) provided by SLANT into 9 larger regions: ventricles, caudate, cerebral white matter (WM), cerebral gray matter (GM), brainstem, putamen, thalamus, cerebellar gray matter (GM), and cerebellar white matter (WM). Figure 2 shows box plots that depict DSCs across ROIs, including the whole brain and the aforementioned 9 regions of the brain.

4.2 QC and Examples

We list the results of the manual quality check in Table 1. Note that here "*success*" does not mean perfect, or even good. We categorize the output to "*success*" as long as the algorithm performed its duty as designed and did not damage any brain voxel by mistake. Figure 3 shows one example of a type-II failure, in which brain voxels were accidentally removed by the defacing algorithm.

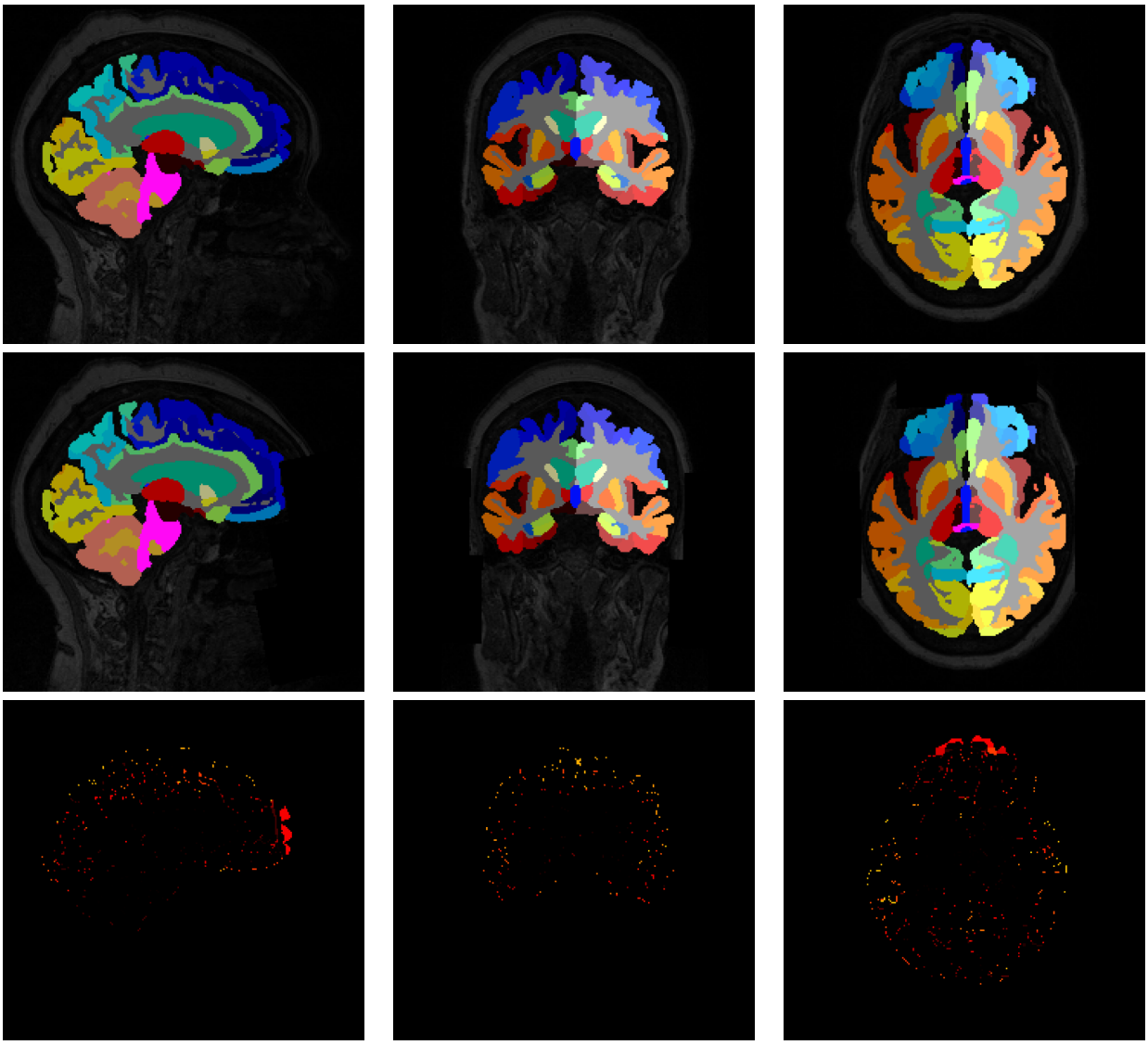

Figure 3. An example of a type-II failure case by FSL_deface[10] algorithm. The first row contains the original MR image overlaid by the segmentation result. The second row contains the defaced MR image by FSL_deface overlaid by the segmentation result. The third row contains the difference between segmentation results from the defaced and original MRI.

4.3 Defacing vs. Rescanning: Kirby 21 Results

We quantify the effects of defacing and rescanning by computing the dice similarity coefficient (DSC) between SLANT results from defaced (or rescanned) and original MRI. Figure 4 plots the DSCs from the various defacing algorithms and the DSCs from processing the rescanned subjects.

5. DISCUSSION AND CONCLUSIONS

5.1 Performances and Effects of Defacing Algorithms

It is controversial to give the definition of a good defacing algorithm. Here we narrow down the scope of comparison to the effects that each defacing algorithm has on the neural analysis pipeline. We argue that protection from

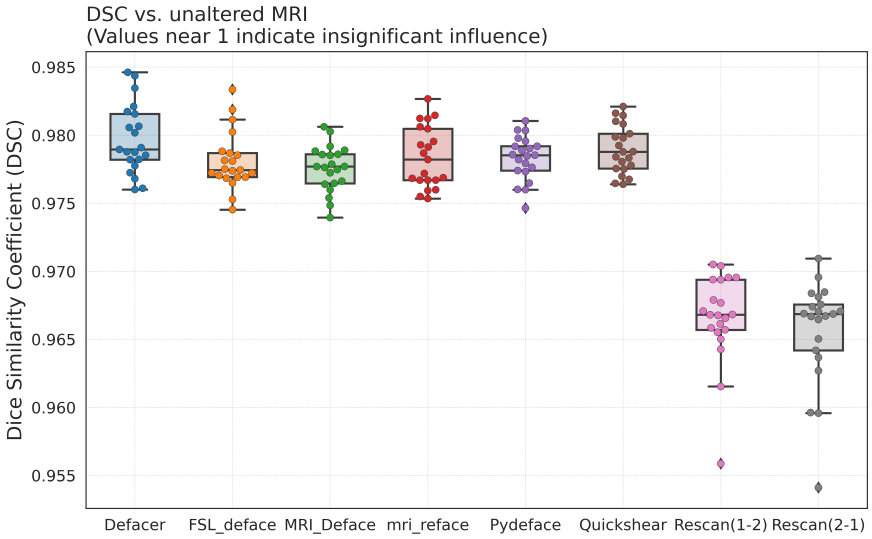

Figure 4. Dice similarity coefficient (DSC) between SLANT segmentation result from original MRI and SLANT segmentation result from defaced (or rescanned) MRI. In Rescan(1-2) group, session-2 MR image is registered to session-1 MR image. In Rescan (2-1) group, session-1 MR image is registered to session-2 MR image.

face recognition should be the essential ability of the defacing algorithm. However, we will still include those algorithms that failed to remove facial features thoroughly in the following discussion.

As indicated in Table 1, when used without fine-tuned parameters, Quickshear is more prone to remove brain voxels. Other algorithms, such as FSL_deface and MRI_Deface, also have this problem–though not to the same extent as Quickshear had on our cohorts. This is devastating to neuroimage analyses not only because a proportion of the cerebrum is missing, but also because the amount of the damaged voxels is unpredictable, thus handicapping both cross-sectional and longitudinal studies.

From Figure 2, we notice that for most of the regions of interest (ROIs), all six algorithms can achieve dice similarity coefficient (DSC) of over 0.96, but there are discernible differences among them, and such differences are fairly consistent in all ROIs. We also notice that all defacing algorithms tend to influence the cerebral gray matter more than other regions, as indicated by the low DSCs. The possible explanation for this is that cerebral gray matter locates mostly at peripheral area, thus is more likely to be tainted by defacing algorithms.

5.2 Defacing vs. Rescanning

In the experiment using 179 MR images from OASIS-3, we have quantified the effects of defacing on SLANT segmentation and compared algorithms with each other. However, we do not truly understand the meaning of a DSC of 0.96. Does it indicate that the influence of defacing is negligible? To address this question, we compared defacing with rescanning using the Kirby 21 dataset. Figure 4 shows the DSCs from the rescanning group as well as the DSCs from the defacing group. The 21 subjects in the Kirby 21 are displayed as points over the boxplots in Figure 4. We notice that the DSCs of the rescanning group is dramatically lower than the DSCs of the different defacing algorithms. This would seem to indicate that the effects of defacing are minor compared to the effects of rescanning a patient and then registering that scan to the previous scan. However, we point out that the registration method here is affine transformation, and there are other methods that can achieve more accurate registration. It is possible that a deformable registration algorithm may reduce this difference, but it does some excessive to use a deformable registration between scans taken one hour apart.

In this study, we analyzed the effects of defacing algorithms on a segmentation pipeline. We have found two important facts. First, many of the defacing algorithms have unstable performance, which is detrimental to

the consistency of neuroanalysis. In extreme cases, part of the brain can be removed due to excessive defacing. Second, defacing has a detrimental influence on brain segmentation. This influence varies from subject to subject and from ROI to ROI. Despite the fact that such influence is minor compared to rescanning and registration, it still requires future work to determine whether the effects of defacing can be neglected.

ACKNOWLEDGMENTS

This research was supported in part by the NIH/NINDS through grant R21-NS120286 (PI: J.L. Prince).

REFERENCES

[1] Schwarz, C. G., Kremers, W. K., Therneau, T. M., Sharp, R. R., Gunter, J. L., Vemuri, P., Arani, A., Spychalla, A. J., Kantarci, K., Knopman, D. S., Petersen, R. C., and Jack, C. R., "Identification of Anonymous MRI Research Participants with Face-Recognition Software," *New England Journal of Medicine* **381**(17), 1684–1686 (2019).

[2] Schwarz, C. G., Kremers, W. K., Wiste, H. J., Gunter, J. L., Vemuri, P., Spychalla, A. J., Kantarci, K., Schultz, A. P., Sperling, R. A., Knopman, D. S., Petersen, R. C., and Jack, C. R., "Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives," *NeuroImage* **231**, 117845 (2021).

[3] Theyers, A. E., Zamyadi, M., O'Reilly, M., Bartha, R., Symons, S., MacQueen, G. M., Hassel, S., Lerch, J. P., Anagnostou, E., Lam, R. W., et al., "Multisite comparison of mri defacing software across multiple cohorts," *Frontiers in psychiatry* **12**, 189 (2021).

[4] Huo, Y., Carass, A., Resnick, S. M., Pham, D. L., Prince, J. L., and Landmann, B. A., "Combining multi-atlas segmentation with brain surface estimation," in [*Proceedings of SPIE Medical Imaging (SPIE-MI 2016), San Diego, CA, February 27-March 3, 2016*], **9784**, 9784–9784–8 (2016).

[5] Huo, Y., Plassard, A. J., Carass, A., Resnick, S. M., Pham, D. L., Prince, J. L., and Landmann, B. A., "Consistent Cortical Reconstruction and Multi-atlas Brain Segmentation," *NeuroImage* **138**, 197–210 (2016).

[6] Jeong, Y. U., Yoo, S., Kim, Y.-H., and Shim, W. H., "De-identification of facial features in magnetic resonance images: software development using deep learning technology," *Journal of medical Internet research* **22**(12), e22739 (2020).

[7] Schimke, N., Kuehler, M., and J.Hale, "Preserving Privacy in Structural Neuroimages," in [*Data and Applications Security and Privacy XXV*], *Lecture Notes in Computer Science* **6818**, 301–308 (2011).

[8] Bischoff-Grethe, A., Ozyurt, I. B., Busa, E., Quinn, B. T., Fennema-Notestine, C., Clark, C. P., Morris, S., Bondi, M. W., Jernigan, T. L., Dale, A. M., Brown, G. G., and Fischl, B., "A technique for the deidentification of structural brain MR images," *Human Brain Mapping* **28**(9), 892–903 (2007).

[9] Gulban, O. F., Nielson, D., and Poldrack, R., "poldracklab/pydeface." https://github.com/poldracklab/pydeface (2019).

[10] Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., et al., "Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank," *Neuroimage* **166**, 400–424 (2018).

[11] Tolpadi, A. A., Stone, M. L., Carass, A., Prince, J. L., and Gomez, A. D., "Inverse biomechanical modeling of the tongue via machine learning and synthetic training data," in [*Proceedings of SPIE Medical Imaging (SPIE-MI 2018), Houstan, TX, February 10 – 15, 2018*], **10576**, 28–35 (2018).

[12] Shao, M., Carass, A., Zhuo, J., Liang, X., Ghunaim, D. H., Stone, M., , and Prince, J. L., "Dynamic palatogram generation from Cine MRI for normalized speech assessment," in [*Proceedings of SPIE Medical Imaging (SPIE-MI 2018), Houstan, TX, February 10 – 15, 2018*], **11596**, 115960T, International Society for Optics and Photonics (2018).

[13] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al., "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic resonance imaging* **30**(9), 1323–1341 (2012).

[14] Anbazhagan, P., Carass, A., Bazin, P.-L., and Prince, J.-L., "Automatic estimation of midsagittal plane and AC-PC alignment on nonrigid registration," in [*3rd International Symposium on Biomedical Imaging (ISBI 2006)*], 828–831 (2006).

[15] Jenkinson, M. and Smith, S. M., "A Global Optimisation Method for Robust Affine Registration of Brain Images," *Medical Image Analysis* **5**(2), 146–156 (2001).

[16] Huo, Y., Xu, Z., Xiong, Y., Aboud, K., Parvathanemi, P., Bao, S., Bermudez, C., Resnick, S. M., Cutting, L. E., and Landman, B. A., "3D whole brain segmentation using spatially localized atlas network tiles," *NeuroImage* **194**, 105–119 (2019).

[17] LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., et al., "Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease," *MedRxiv* (2019).

[18] Landman, B. A., Huang, A. J., Gifford, A., Vikram, D. S., Lim, I. A. L., Farrell, J. A., Bogovic, J. A., Hua, J., Chen, M., Jarso, S., et al., "Multi-parametric neuroimaging reproducibility: a 3-t resource study," *Neuroimage* **54**(4), 2854–2866 (2011).

[19] Dice, L. R., "Measures of the Amount of Ecologic Association Between Species," *Ecology* **26**(3), 297–302 (1945).

[20] Carass, A., Roy, S., Gherman, A., Reinhold, J. C., Jesson, A., Arbel, T., Maier, O., Handels, H., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Pham, D. L., Crainiceanu, C. M., Calabresi, P. A., Prince, J. L., Gray Roncal, W. R., Shinohara, R. T., and Oguz, I., "Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis," *Scientific Reports* **10**, 8242 (2020).