# Refacing Defaced MRI with PixelCNN

Yaorong Xiao
*Department of Computer Science*
*Georgia State University*
Atlanta, USA
yxiao11@student.gsu.edu

William Ashbee
*Department of Computer Science*
*Georgia State University*
Atlanta, USA
wsashbee@gmail.com

Vince D. Calhoun
*Center for Translational Research in Neuroimaging and Data Science*
*Georgia State University*
Atlanta, USA
vcalhoun@gsu.edu

Sergey Plis
*Department of Computer Science*
*Georgia State University*
Atlanta, USA
s.m.plis@gmail.com

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9891937

***Abstract*—Privacy protection is one of the most crucial factors when sharing MR images between researchers. There are many defacing software programs that can blur or remove the face of MR images. However, there are reasons to believe that the brain and other remaining features are not only identifiable but also can be used for facial reconstruction to fill the face of the subject back into the image, and it is possible to rebuild the facial part of the images using recently developed machine learning models. A demonstration of this is of practical significance as it could convince the community to adopt stricter data-sharing standards. Additionally, even if the reconstructed faces are not identifiable, smooth completion of the "damaged" MRI images may improve methods that depend on the head modeling, such as source localization approaches. Recent work has been focusing on using the generative adversarial networks (GAN) for this task, which is generally believed to be the best method. Because we hope the model can generate the face entirely depending on the information from the brain, we show here an alternative approach, pixel constrained CNN, which is a purely supervised facial reconstruction. We simulated the rebuild process and showed convincing results.**

***Index Terms*—Neuroimaging, Refacing, Privacy protection**

## I. Introduction

Privacy-preserving Deep learning of MRI (Magnetic Resonance Imaging) patient data is—like any security-related field—an ever-evolving game of cat and mouse. Like the adversarial atmosphere of security, generative adversarial networks by way of CycleGAN have been used to demonstrate refacing techniques that weaken anonymizing characteristics of defacing [1]. While striving to protect patient anonymity in the field of neuroscience, the white hats (neuroscientists, deep learning researchers, data scientists, security researchers, and regulators) must determine what parties to trust (e.g., patients, developers, researchers, software, and hardware providers) and reduce the number of trusted parties to as few as is reasonable. There is a balance that benevolent parties must discern between reasonable levels of privacy protections and added cost to the research process. Concomitantly, what historically once may have been a reasonable security measure will eventually be obsolete, as our work will show in the case of defacing MRIs to protect identity.

When a security practice is shown to be obsolete, it should be replaced with better methods to protect patient privacy. There is little value in continuing to use easily subverted countermeasures when they not only fail to protect privacy, as our research shows in the case of defacing, but also when those methods have the potential to increase the probability of incorrect analysis due to artifacts that may be introduced into the processed MRI [2]. MRI is an imaging technique that utilizes strong magnetic fields to align the orientation of protons in organic tissues. Additionally, radio frequencies (RF) are used to perturb protons from their magnetically induced alignment a finite amount. The photon from the RF wave deposits energy into the proton. When the radio frequencies are removed, another photon is emitted as the proton realigns with the strong magnetic field. The photon released during realignment of the proton with the strong magnetic field can be measured and used to create an image. MRI is useful in such fields as neuroscience, cardiology, and orthopedics. Essentially any time a physician would like to image internal structures of the human body, MRI is a significant tool. However, despite the usefulness of MRI, we will show that ensuring anonymity of the patient data is becoming increasingly difficult due to generative and transformational neural network models such as Pixel Constrained CNN, which we use to inpaint faces (reface) in zeroed out regions of an MRI that have been removed for the purpose of privacy–a practice known as defacing. Implicit in the practice of defacing and refacing is the premise that brain images cleaned of the face do not contain personally identifiable information or enough data to reconstruct that information. Our work demonstrates that the assumption of independence of face and brain may be flawed. We demonstrate that defaced MRIs can be reconstructed using freely available deep learning generative models such as pixel constrained CNN [1]. As generative models continue to evolve, the ability of these models to reface personally identifiable information

This work was funded in part by NIH RF1MH121885, R01MH123610, R01EB006841 and NSF 2112455 grants.

will only further make current defacing methods obsolete. The reasons why the brain and face are not independent in their structure are probably biological due to dependencies on DNA. Some neurological disorders have associated facial affects [3]. Both the face and the brain are a function of the DNA of a single individual, so there are reasons for potential correlation. More formally, one may consider the human body a non-linear dynamical system, and such systems often exhibit coupling and correlation in seemingly distinct modules [4].

Our work demonstrates that the face can be reconstructed after refacing. We used two methods to deface the MRI, and no matter which one is used, our model (Pixel Constrained CNN) can regenerate the face with a high SSIM. The regenerated image not only has the shape of a human being, but also has plausible details such as correct muscle structure and features.

The main contributions and advantages of defacing and refacing MRI can be summarized as:

- We demonstrate that current state-of-art defacing methods do not sufficiently protect patient's privacy.
- We establish that image orientation and mask construction can optimize the effects of Pixel Constrained CNN's receptive field on refacing results.
- We utilize image compression in a way that makes prediction computationally efficient without losing significant qualitative image fidelity of refacing results.

## II. Related work

The structure of the regions of the brain and the facial features exhibit measurable correlation. Facial features can be derived from DNA with varying levels of accuracy [3], [5]–[10]. DNA, brains and the human body behave as non-linear dynamical systems, and non-linear dynamical systems often exhibit high degrees of coupling [4], [11], [12]. Structures within the brain exhibit correlations inactivation during emotional changes [12]. Some genetic disorders are marked by facial as well as brain changes [3].

Another nail in the coffin of defacing is the phenomenon of covariance shift-related problems in deep learning. Covariance shift problems are a common problem in deep learning, yet they are infrequently known as such. This problem, also known as domain shift, is related to catastrophic forgetting in continuous learning. Covariance shift problems occur when there are changes to the distribution of training and test set data, such as a new MRI machine, different cameras, different lighting conditions, different camera angles, or increased noise in the input to the neural network. These will result in higher than expected error rates. Defacing or refacing presents a possible way to introduce a nontrivial covariance shift that could fail analysis models. Analysis algorithms trained with one face removal or replacement algorithm will be more likely to fail when presented with inputs that are dissimilar to training in the types of artifacts and noise such algorithms might introduce. Discerning how much defacing and refacing methods will impact the generalizability of neural network models during their lifecycle is difficult–except to say most likely negatively.

In a similar problem, Goodfellow proposes the famous adversarial example of adding engineered noise via the fast gradient sign method to an image and changing the predicted class from panda to gibbon [13]. Adversarial examples prove that neural networks are not always robust to covariance shift.

Rather than introducing a potential covariance shift that has become increasingly outmoded, one seeking to protect privacy has the opportunity to choose from myriad and growing existing privacy-preserving solutions that can replace or augment defacing in MRI. Homomorphic encryption that allows computation only on encrypted data may have been used in privacy preservation [14]–[23]. Federated learning that allows for data and models to be split among multiple parties could also be used to solve privacy issues [18], [20], [21], [24]–[32]. Blockchain is still another solution to patient privacy that allows tracking of computation and trusted parties [18], [33]–[37]. Trusted execution environments that implement hardware restrictions to access are yet another solution [18], [21]–[23], [29], [37]–[40]. With so many solutions, we can move to newer and better standards to protect the privacy and keep the number of trusted parties minimal.

## III. Image inpainting

Image inpainting is a problem where some portion of an image is masked by element-wise multiplication with a binary matrix of equivalent size. In general, given an Image $I_{m,n,c}$ and binary mask $M_{m,n,c}$, inpainting takes as input a value, $I \odot M$ where $\odot$ is element-wise multiplication with the binary mask.

When element-wise multiplication with 1 occurs, the original image pixel remains as it was. When element-wise multiplication with 0 occurs, the information about that pixel is destroyed. Thus, the mask tensor plays the role of a binary gate for each pixel in the original input image or MRI.

Image inpainting is an exciting problem in that it can be made arbitrarily simple (fill in a monochromatic image) or complex (fill in a masked out a tumor in the brain). One can imagine other challenging scenarios for reconstruction in sports and action photos requiring an understanding of the causality of the scene if key parts are removed. Reproducing engineering schematics that are redacted is approaching an AI-complete problem. The range of difficulty is worth pointing out because the history of image inpainting with deep neural networks has success somewhere between these extremes of arbitrarily simple and potentially AI-complete. However, each of these implementations can be made to have failure cases if the masks at the testing time are different from the masks at the training time or if there is a domain shift in the training data, such as shifting from celeb-a to ImageNet. Image inpainting on celeb-a is much closer to the type of problem of image inpainting on MRI because celeb-a is of a single super-class–much like MRI in neuroscience. However, often MRI datasets are much smaller than celeb-a. MRI datasets requiring defacing are generally similar in class cardinality and structure (there is a single class of similarly structured

images). This consistency increases the likelihood of learning success and decreases the training time.

Because one can see consistent improvements in the quality and difficulty of inpainting over time, the time to improve privacy-preserving measures in neuroscience approaches. Inpainting models will only make the premise of this paper that one can infer facial features from surrounding features more probable in its ability to compromise privacy. One can envision models with larger receptive fields, and more semantic understanding can learn from a large corpus of MRI and thus more accurately reconstruct inpainted facial features.

## IV. PixelCNN, PixenCNN++, and gated PixelCNN

Pixel by pixel image generation was effectively implemented in PixelRNN and PixelCNN formats as a generative model for unsupervised image modeling [41]. PixelCNN was not explicitly adapted for image inpainting at the outset, although being a generative model with decent scalability, speed, and effectiveness, PixelCNN was destined to be adapted for this purpose. The authors of PixelCNNs also created PixelRNNs, which are more accurate in generating models, but much slower. PixelCNNs have a bounded receptive field near the pixel being generated. A mask is used to bound the context for a generated pixel at index $i$ to only previously generated pixels–modeling the conditional probabilities of (1). Prior to pixel by pixel generation, image generation has completed a section or entire image at a time via convolution, dilated convolution, and deconvolution architectures [42], [43]. In PixelCNN, each unknown pixel is modeled as conditional probabilities on other known pixels, as one can see in (1). $x_i$ represents the $i_{th}$ pixel in the image, and $n$ is the width of the input image.

$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i|x_{i-1}, \ldots, x_1) \tag{1}$$

PixelCNN has numerous extensions, such as PixelCNN++ and gated pixel CNN. PixelCNN++ replaced several aspects of PixelCNN that were obsolete–replacing softmax, increasing the receptive field, adding skip connections, and using dropout regulation [44]. These now fairly standard model improvements created a much more robust generative model. Gated pixel CNN focuses on the conditional generation of images [45]. Additionally, as the name implies, a gating mechanism is added to the convolutions to help gradient propagation. However, the conditional image generation introduces the opportunity to apply pixel CNNs to inpainting, though the work focuses on more general conditional image generation from one hot encoding and portrait embeddings. In (2), $h$ represents the added conditions that help to generate an image based on a specific limitation.

$$p(\mathbf{x}|\mathbf{h}) = \prod_{i=1}^{n^2} p(x_i|x_{i-1}, \ldots, x_1, \mathbf{h}) \tag{2}$$

The gated pixel CNN can generate images from one-hot encodings of image classes as well as portrait embeddings (represented by h in (2)). Gated pixel CNN remarkably performs image inpainting, but it is conditional image generation from a latent space, a subtle but important difference. It is also demonstrated to perform well as an auto-encoder.

## V. Probabilistic Semantic Inpainting with Pixel Constrained CNNs

Pixel constrained CNNs introduce the capability to conditionally generate images in the domain of inpainting that are a function of the image x and conditional values $c$.

$$p(\mathbf{x}|\mathbf{c}) = \prod_{i:x_i \notin \mathbf{c}} p(x_i|x_{i-1}, \ldots, x_1, \mathbf{c}) \tag{3}$$

$c$ in the context of image inpainting is a function of mask $M$ and image $x$. The architecture of pixel constrained CNNs has two conceptual networks–a prior network that is a gated pixel CNN and a conditioning network that is a resnet. The prior network receives the ground truth but uses masked convolution to constrain the receptive field of the network to only previously generated pixels. The conditioning network takes two inputs–a masked image $X \odot M$ and the mask itself $M$–each of the same size. The inputs are concatenated together $(X \odot M) \| M$ before being fed into the conditioning network. The logits of both the prior network and the conditioning network are combined to predict the intensity of each generated pixel, and cross-entropy loss is used to teach the network based on the correctness of its answers. A prior network is fed the ground truth input during training to aid the network in learning representations. At inference time, the network is sampled to create unique images with different likelihoods from the test set. Pixel constrained CNN is effective at inpainting images of the celeb-a dataset when compared against gan and neural process [1].

## VI. Dataset

Our dataset was collected at the Mind Research Network (MRN) was used for this study. Data were collected on Siemens 3 T TIM Trio scanner, located at MRN, using a 12-channel head coil. High resolution T1-weighted images were acquired with a 5-echo multi-echo MPRAGE sequence [TE (echo times) = 1.64, 3.5, 5.36, 7.22, 9.08 ms, TR (repetition time) = 2.53 s, TI (inversion time) = 1.2 s, 7deg flip angle, number of excitations (NEX) = 1, slice thickness = 1 mm, field of view (FOV)= 256 mm, resolution = 256 × 256. There is a total of 946 samples in the dataset. We used 800 randomly selected samples as our training set and the remaining for the testing set.

## VII. Methods

### A. Methods for Defacing

We used two methods for defacing MRI. One is *fsl_deface* which is considered the best existing method at preventing face recognition in part because it removes the eyes and has minor effects on brain measurement [46]. Another method is *mri_deface*. *mri_deface* method has also been used as the defacing algorithm prior to refacing with CycleGAN [47]. Similarly as Abramian's and Eklund's work, We compare the structural

similarity index (SSIMs) between the defaced images (from both *fsl_deface* and *mri_deface*) to the ground truth image [48]. SSIM is used to measure the similarity between two images. Generally speaking, the higher the SSIM is, the more similar the two images are. The comparison is shown in figure 1. So one can clearly see that the similarity of MRIs defaced by *fsl_deface* is far lower to the original image than defaced by *mri_deface* which means *fsl_deface* is better in defacing.

In figure 2, we can see that *mri_deface* can mask out more pixels from the MRI. However, the problem is that a part of the eyes is still there, leaving important information for reconstructing the original face.

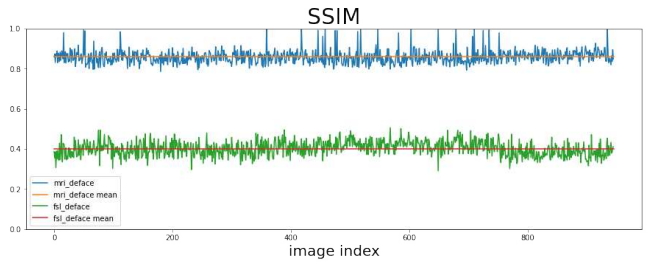


Fig. 1. SSIM Comparison Between Defacing *mri_deface* and *fsl_deface* on Whole Dataset

### B. Data Preprocessing

The data we have has a size of 256 by 256 pixels. We use OpenCV to resize the image to 128 by 128 pixels to save model space on GPU and allow for larger batch sizes. We reduced the pixel intensity from 8-bit (with a range of 0 to 255) to 3-bit(with a range of 0 to 7) so that our model (Pixel Constrained CNN) calculates the likelihood of each pixel within reasonable uncertainty bounds. From Figure 3 we can see that even when the processed data has a lower resolution and pixel intensity, the integrity of the original data is largely maintained by looking through the human eye.

### C. Model

The model we use is Pixel Constrained CNN described in section V. Pixel Constrained CNN has a receptive field at pixel $i$ dependent not only on the rows above the pixel and the partial row to the left of pixel $i$, but also on a global overview of each pixel in the image [1]. As a result, we orient our facial images face down, as seen in figure 4. The masked image and mask are fed into Pixel Constrained CNN as input to obtain refaced output–see figure 4. By applying Pixel Constrained CNN, our model can collect information from the face, brain, jaw, forehead, and any image part. This will better help to predict a facial shape.

### D. Training

We trained our model through backpropagation with 500 epochs in *pytorch*. We set our prior network with 19 layers and conditional network with 32 layers. 32 layers for the conditional network is because we want the model to have the largest receptive field of the image so that we can condition on as much information as possible. Also, we set a 5 by 5 kernel for each network and a cosine annealing schedule for a learning rate of 0.00005.

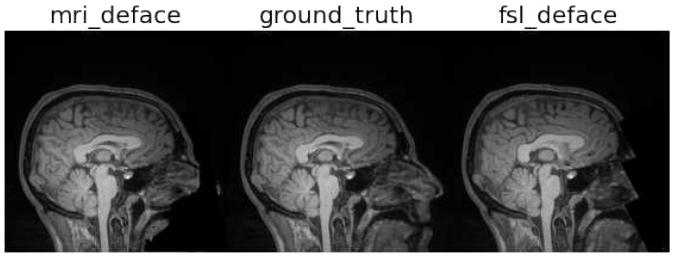


Fig. 2. Examples of *mri_deface*, Ground Truth and *fsl_deface*

In figure 4, the input of the model has two components. One component is for the input of the prior network, which should be the original image (4a). The other component for the conditioning network is a two-channel tensor consisting of one channel, the defaced image, and the second channel, the relative mask (4d). As we can see all the images in figure 4, we rotate them from face-right to face-down. Because PixelCNN starts to generate pixels from the top-left points. We want our model to estimate the likelihood of each pixel based on as much information as is available from the unmasked portion of the MRI.

## VIII. RESULTS

The structural similarity index measure(SSIM) is considered one of the best measures to evaluate the similarity between two images [49]. We use this method to evaluate the generated images' quality. We randomly selected five images from the held-out test dataset and demoed them with defacing by *MRI_dface* and *fsl_deface*. From the two demo plots on both experiments in figure 5 and figure 6, we can see that no matter what method is used to remove the face, our model can have a good prediction. One trend we observed through the 146 demos from the test data was that the MRI with a larger head shape and prominent neck will result in better reconstruction, such as in demos 1 and 4 in both figure 5 and figure 6. In demo 5, the neck part is missing from the MRI compared

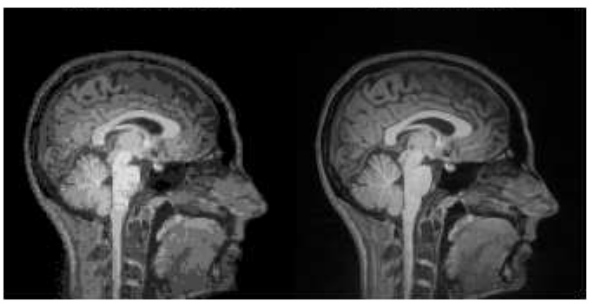


Fig. 3. Comparison between the processed image and ground truth. The left image is a processed image with lower resolution and lower intensity. The right image is the ground truth

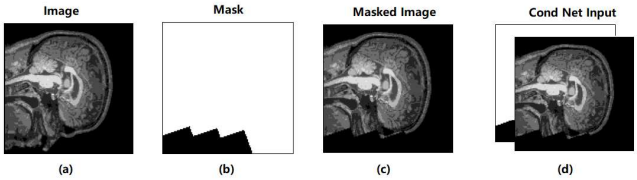


Fig. 4. From left to right. (a) is the processed image which will input to a prior network. (b) is the mask of relative image.(c) is masked out deface image. (d) is the concatenated of mask and the defaced image. This 2 channel image will input to the conditional network for feature extraction

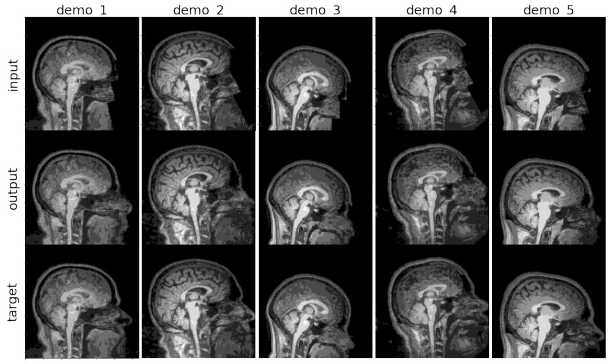


Fig. 5. Demo on Samples Defaced by *fsl_deface*. The first row is the input which are defaced image. The second row is the generated images. The third row is the ground truth.

to the other four demos, and the reconstruction is of lower quality. We anticipate that such anomalies would go away with larger data sets or more data augmentation, although these approaches would increase training time. The demo results in figure 5 and figures 6 are both generated from the held-out testing dataset.

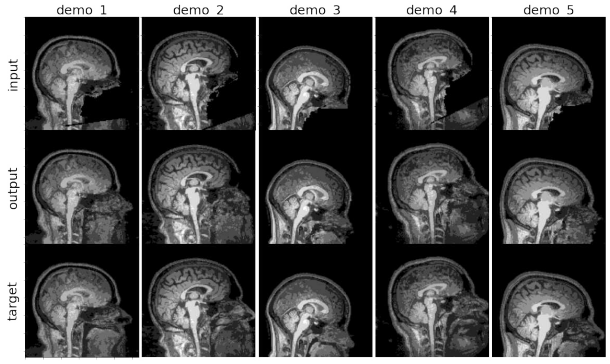


Fig. 6. Demo on Samples Defaced by *mri_deface*. The first row is the input which are defaced image. The second row is the generated images. The third row is the ground truth.

Looking through the plot of loss in figure 7, We can see a tendency towards overfitting in the most general sense of the term (loss on training would be better than the loss on the test set). There were diminishing returns on how much improvement in training resulted in improvement in the test set loss. Initially, we would always try to maintain parity between training loss and test loss and avoid any overfitting with early stopping. However, early stopping would result in

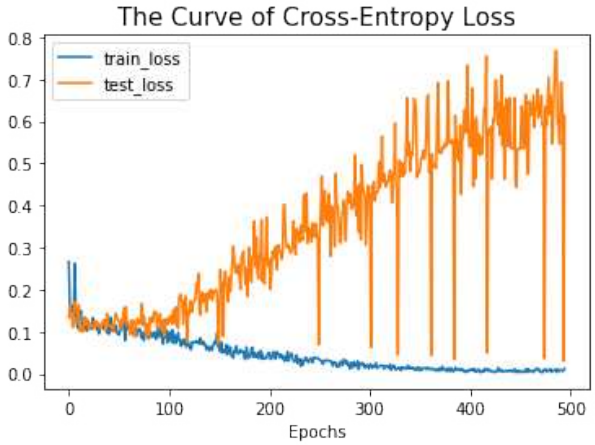


Fig. 7. The Cross-Entropy Loss in Epochs

a lack of precise qualitative results on training and test sets. For example, the shape of the face would be incorrect with early stopping, although the loss would be better. However, qualitative improvements in the test set are not reflected by the loss measures. Additionally, overfitting is quite difficult to avoid with such a small dataset. When we increase the iterations from 100 to 500 epochs, the model can have a tremendous qualitative performance not only on the trained 800 samples but also would have good generations on the held-out 146 samples which we have shown in figure 5 and figure 6. Although the loss of the test curve keeps going up, the generation process in the test set still performed well qualitatively. We check that there are no duplicated samples between the training and testing datasets. In general, a higher cross-entropy loss and qualitative accuracy will not be perfectly correlated. Some incorrect predictions can have a significantly larger effect than correct predictions. We speculate that the model is learning the face structure, as evidenced by the qualitative results, but cannot reproduce the exact face with as high of precision as indicated by the loss values.

In figure 8, the green line represents the SSIM between defaced image and ground truth. Compared with figure 1, the SSIM of the defaced image is numerically increased. This is because our prepossessing reduced the MRI resolution and pixel intensity. While the blue line represents the SSIM between refaced image and ground truth. The defacing process is based on *fsl_deface*. We can see that after prediction by Pixel Constrained CNN, the SSIM of the refaced image has increased. Although the increase is not significant, in our experiments, the pixel intensity is only 3-bit, and the changing of SSIM is only affected by the part of the masked pixels.

Figure 9 shows a comparison between image defaced by *fsl_deface* and *mri_deface*. From the plot, face reconstruction through Pixel Constrained CNN has a better performance on *fsl_deface*. We assume that it is because *mri_deface* can zero out more pixels at the direct back of the nose of MRI. We observed that the nose is the hardest part to generate through our experiments.

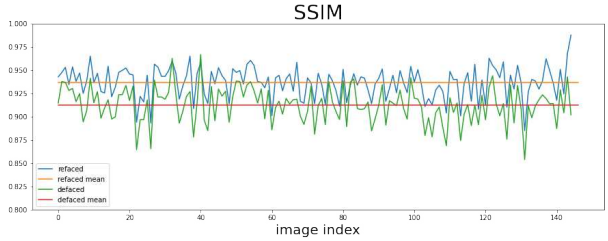


Fig. 8. Comparison of SSIM before and after predicted by Pixel Constrained CNN. The blue line is the SSIM of our newly generated images, compared with the green line, which is the SSIM before refacing

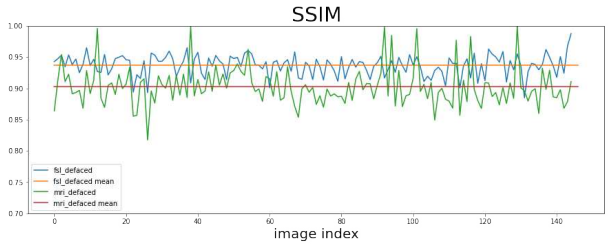


Fig. 9. SSIM of generated images between defacing by *mri_deface* and *fsl_deface*

## IX. Discussion and Conclusions

We can get reasonably good reconstructions of faces no matter what defacing method used. These results demonstrate that privacy still at risk despite defacing techniques. By reducing the resolution and intensity of the defaced images, we simplified the network's calculations and improved our likelihood predictions and meanwhile keeping the main feature of the subject. We observe that the most challenging part of reconstructing is the nose. We also observe that the larger the masked region, the more difficult it is to reconstruct based on our comparisons of *mri_deface* and *fsl_deface*.

Prior work in the relationship between DNA and facial features mentions that the nose is affected most by genes [50]. Because we can often reconstruct the nose from the defaced MRI, we surmise that these nose genes may also affect brain structures.

Increasing the size of our dataset would greatly help reconstruct masked regions. Additionally, 3d reconstruction is the next step of this research. The pipeline that an attacker could easily take to compromise security fully would require lifting images from 3d reconstructions of defaced MRI and searching images available in social media databases for nearest matches. Furthermore, we think our research has implications for criminology and anthropology. In anthropology, one could reconstruct ancient humans' brain structures or faces from their remains by masking out different regions and applying continuous learning (bringing what is learnable from modern MRI data into this field). In criminology, head trauma could be removed from data to identify the remains of victims.

Most importantly, we would like to inform the research community in neuroscience that current defacing measures are quickly becoming obsolete. We as a community have a responsibility to protect privacy. Organized crime and state-sponsored actors could easily afford to create the larger corpus of MRI necessary to reconstruct de-identified images using our methodology. Such methodologies could already be in place, and thus our community needs to implement more rigorous techniques to protect patient identities.

Based on our observation that larger masks are more challenging to reconstruct, it is clear that one should segment the brain and remove the nasal and throat structures in addition to only removing the face. Productive future research going forward might use brain segmentation to mask out these features in new defacing algorithms.

## References

[1] E. Dupont and S. Suresha, "Probabilistic semantic inpainting with pixel constrained cnns," 2019.

[2] A. De Sitter, M. Visser, I. Brouwer, K. Cover, R. van Schijndel, R. Eijgelaar, D. Müller, S. Ropele, L. Kappos, Á. Rovira *et al.*, "Facing privacy in neuroimaging: removing facial features degrades performance of image analysis methods," *European Radiology*, vol. 30, no. 2, pp. 1062–1074, 2020.

[3] S. Richmond, L. J. Howe, S. Lewis, E. Stergiakouli, and A. Zhurov, "Facial genetics: a brief overview," *Frontiers in genetics*, p. 462, 2018.

[4] W. Ashbee, L. Hively, and J. McDonald, "Nonlinear epilepsy forewarning by support vector machines," in *Epilepsy Topics*, M. D. Holmes, Ed. Rijeka: IntechOpen, 2014, ch. 3. [Online]. Available: https://doi.org/10.5772/57438

[5] J. D. White, K. Indencleef, S. Naqvi, R. J. Eller, H. Hoskens, J. Roosenboom, M. K. Lee, J. Li, J. Mohammed, S. Richmond *et al.*, "Insights into the genetic architecture of the human face," *Nature genetics*, vol. 53, no. 1, pp. 45–53, 2021.

[6] Z. Xiong, G. Dankova, L. J. Howe, M. K. Lee, P. G. Hysi, M. A. De Jong, G. Zhu, K. Adhikari, D. Li, Y. Li *et al.*, "Novel genetic loci affecting facial shape variation in humans," *Elife*, vol. 8, p. e49898, 2019.

[7] J. R. Shaffer, E. Orlova, M. K. Lee, E. J. Leslie, Z. D. Raffensperger, C. L. Heike, M. L. Cunningham, J. T. Hecht, C. H. Kau, N. L. Nidey *et al.*, "Genome-wide association study reveals multiple loci influencing normal human facial morphology," *PLoS genetics*, vol. 12, no. 8, p. e1006149, 2016.

[8] D. J. Crouch, B. Winney, W. P. Koppen, W. J. Christmas, K. Hutnik, T. Day, D. Meena, A. Boumertit, P. Hysi, A. Nessa *et al.*, "Genetics of the human face: Identification of large-effect single gene variants," *Proceedings of the National Academy of Sciences*, vol. 115, no. 4, pp. E676–E685, 2018.

[9] D. Tsagkrasoulis, P. Hysi, T. Spector, and G. Montana, "Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping," *Scientific reports*, vol. 7, no. 1, pp. 1–18, 2017.

[10] S. M. Weinberg, R. Cornell, and E. J. Leslie, "Craniofacial genetics: Where have we been and where are we going?" p. e1007438, 2018.

[11] G. Gaeta, C. Reiss, M. Peyrard, and T. Dauxois, "Simple models of non-linear dna dynamics." *Rivista del Nuovo cimento*, vol. 17, no. 4, pp. 1–48, 1994.

[12] L. I. Aftanas, N. V. Lotova, V. I. Koshkarov, and S. A. Popov, "Non-linear dynamical coupling between different brain areas during evoked emotions: an eeg investigation," *Biological Psychology*, vol. 48, no. 2, pp. 121–138, 1998.

[13] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[14] Y. Aono, T. Hayashi, L. Wang, S. Moriai *et al.*, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, 2017.

[15] P. Li, J. Li, Z. Huang, T. Li, C.-Z. Gao, S.-M. Yiu, and K. Chen, "Multi-key privacy-preserving deep learning in cloud computing," *Future Generation Computer Systems*, vol. 74, pp. 76–85, 2017.

[16] G. Kaissis, A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn *et al.*, "End-to-end privacy preserving deep learning on multi-institutional medical imaging," *Nature Machine Intelligence*, vol. 3, no. 6, pp. 473–484, 2021.

[17] M. S. Riazi and F. Koushanfar, "Privacy-preserving deep learning and inference," in *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE, 2018, pp. 1–4.

[18] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.

[19] P. Martins, L. Sousa, and A. Mariano, "A survey on fully homomorphic encryption: An engineering perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–33, 2017.

[20] D. Stripelis, J. L. Ambite, P. Lam, and P. Thompson, "Scaling neuroscience research using federated learning," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 1191–1195.

[21] Y. Chen, F. Luo, T. Li, T. Xiang, Z. Liu, and J. Li, "A training-integrity privacy-preserving federated learning scheme with trusted execution environment," *Information Sciences*, vol. 522, pp. 69–79, 2020.

[22] F. Mireshghallah, M. Taram, P. Vepakomma, A. Singh, R. Raskar, and H. Esmaeilzadeh, "Privacy in deep learning: A survey," *arXiv preprint arXiv:2004.12254*, 2020.

[23] T. Lee, Z. Lin, S. Pushp, C. Li, Y. Liu, Y. Lee, F. Xu, C. Xu, L. Zhang, and J. Song, "Occlumency: Privacy-preserving remote deep-learning inference using sgx," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–17.

[24] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso *et al.*, "Privacy-preserving federated brain tumour segmentation," in *International workshop on machine learning in medical imaging*. Springer, 2019, pp. 133–141.

[25] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen *et al.*, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[26] A. G. Roy, S. Siddiqui, S. Pölsterl, N. Navab, and C. Wachinger, "Braintorrent: A peer-to-peer environment for decentralized federated learning," *arXiv preprint arXiv:1905.06731*, 2019.

[27] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein *et al.*, "The future of digital health with federated learning," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–7, 2020.

[28] S. Silva, B. A. Gutman, E. Romero, P. M. Thompson, A. Altmann, and M. Lorenzi, "Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, 2019, pp. 270–274.

[29] G. A. Reina, A. Gruzdev, P. Foley, O. Perepelkina, M. Sharma, I. Davidyuk, I. Trushkin, M. Radionov, A. Mokrov, D. Agapov *et al.*, "Openfl: An open-source framework for federated learning," *arXiv preprint arXiv:2105.06413*, 2021.

[30] S. Pati, U. Baid, M. Zenk, B. Edwards, M. Sheller, G. A. Reina, P. Foley, A. Gruzdev, J. Martin, S. Albarqouni *et al.*, "The federated tumor segmentation (fets) challenge," *arXiv preprint arXiv:2105.05874*, 2021.

[31] S. M. Plis, A. D. Sarwate, D. Wood, C. Dieringer, D. Landis, C. Reed, S. R. Panta, J. A. Turner, J. M. Shoemaker, K. W. Carter *et al.*, "Coinstac: a privacy enabled model and prototype for leveraging and processing decentralized brain imaging data," *Frontiers in neuroscience*, vol. 10, p. 365, 2016.

[32] A. D. Sarwate, S. M. Plis, J. A. Turner, M. R. Arbabshirani, and V. D. Calhoun, "Sharing privacy-sensitive access to neuroimaging and genetics data: a review and preliminary validation," *Frontiers in neuroinformatics*, vol. 8, p. 35, 2014.

[33] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2438–2455, 2019.

[34] F. Zerka, V. Urovi, A. Vaidyanathan, S. Barakat, R. T. Leijenaar, S. Walsh, H. Gabrani-Juma, B. Miraglio, H. C. Woodruff, M. Dumontier *et al.*, "Blockchain for privacy preserving and trustworthy distributed machine learning in multicentric medical imaging (c-distrim)," *Ieee Access*, vol. 8, pp. 183 939–183 951, 2020.

[35] R. H. Hylock and X. Zeng, "A blockchain framework for patient-centered health records and exchange (healthchain): evaluation and proof-of-concept study," *J Med Internet Res*, vol. 21, no. 8, p. e13592, 2019.

[36] I. Abu-Elezz, A. Hassan, A. Nazeemudeen, M. Househ, and A. Abd-Alrazaq, "The benefits and threats of blockchain technology in healthcare: A scoping review," *International Journal of Medical Informatics*, vol. 142, p. 104246, 2020.

[37] Q. Yang, "Toward responsible ai: An overview of federated learning for user-centered privacy-preserving computing," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 11, no. 3-4, pp. 1–22, 2021.

[38] H. C. Tanuwidjaja, R. Choi, S. Baek, and K. Kim, "Privacy-preserving deep learning on machine learning as a service—a comprehensive survey," *IEEE Access*, vol. 8, pp. 167 425–167 447, 2020.

[39] S. Volos, K. Vaswani, and R. Bruno, "Graviton: Trusted execution environments on {GPUs}," in *13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18)*, 2018, pp. 681–696.

[40] F. Tramer and D. Boneh, "Slalom: Fast, verifiable and private execution of neural networks in trusted hardware," *arXiv preprint arXiv:1806.03287*, 2018.

[41] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," 2016.

[42] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 85–100.

[43] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5505–5514.

[44] T. Salimans, A. Karpathy, X. Chen, and D. P. Kingma, "Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications," *arXiv preprint arXiv:1701.05517*, 2017.

[45] A. van den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, "Conditional image generation with pixelcnn decoders," 2016.

[46] C. G. Schwarz, W. K. Kremers, H. J. Wiste, J. L. Gunter, P. Vemuri, A. J. Spychalla, K. Kantarci, A. P. Schultz, R. A. Sperling, D. S. Knopman, R. C. Petersen, and C. R. Jack, "Changing the face of neuroimaging research: Comparing a new mri de-facing technique with popular alternatives," *NeuroImage*, vol. 231, p. 117845, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811921001221

[47] D. Abramian and A. Eklund, "Refacing: Reconstructing anonymized facial features using GANS," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, apr 2019.

[48] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[49] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[50] S. Weinberg and J. Shaffer, "We scanned the dna of 8,000 people to see how facial features are controlled by genes," *THE CONVERSATION*, 2020.