Projection image-to-image translation in hybrid x-ray/MR imaging

Bernhard Stimpel, Christopher Syben, Tobias Würfl, Katharina Breininger, Jonathan M. Lommen, et al.

Bernhard Stimpel, Christopher Syben, Tobias Würfl, Katharina Breininger, Jonathan M. Lommen, Arnd Dörfler, Andreas Maier, "Projection image-to-image translation in hybrid x-ray/MR imaging," Proc. SPIE 10949, Medical Imaging 2019: Image Processing, 109492K (15 March 2019); doi: 10.1117/12.2512195

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Projection Image-to-Image Translation in Hybrid X-ray/MR Imaging

Bernhard Stimpel[a,b], Christopher Syben[a,b], Tobias Würfl[a], Katharina Breininger[a], Jonathan M. Lommen[a,b], Arnd Dörfler[b], and Andreas Maier[a]

[a]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[b]Department of Neuroradiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

ABSTRACT

The potential benefit of hybrid X-ray and MR imaging in the interventional environment is large due to the combination of fast imaging with high contrast variety. However, a vast amount of existing image enhancement methods requires the image information of both modalities to be present in the same domain. To unlock this potential, we present a solution to image-to-image translation from MR projections to corresponding X-ray projection images. The approach is based on a state-of-the-art image generator network that is modified to fit the specific application. Furthermore, we propose the inclusion of a gradient map in the loss function to allow the network to emphasize high-frequency details in image generation. Our approach is capable of creating X-ray projection images with natural appearance. Additionally, our extensions show clear improvement compared to the baseline method.

1. INTRODUCTION

Hybrid imaging exhibits high potential in diagnostic and interventional applications.[1] Future advances in research may leverage the combination of Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) to clinical applicability. Especially for interventional purposes, the gain from simultaneously acquiring soft- and dense-tissue information would yield great opportunities. Assuming the information of both modalities is present at the same time, numerous existing post-processing methods would become applicable. Image fusion techniques, e.g., image overlays, have proven useful in the past. Additionally, one can think about image enhancement techniques, e.g., image de-noising or super-resolution. To enable the latter methods, it is beneficial to have the data available in the same domain. Solutions to generate CT images from corresponding MRI data were presented previously,[2,3] mostly in order to generate attenuation maps for radiation therapy. However, all of these are applied to volumetric data, i.e., slice images. In contrast, interventional procedures rely heavily on line integral data from X-ray projection imaging. Projection images which exhibit the same perspective distortion can also be acquired directly using an MR device.[4] This avoids time-consuming volumetric acquisition and subsequent forward projection. The synthesis of the desired X-ray projections from the corresponding MRI signal is an inherently ill-posed problem. Large portions of the dominant signal in X-ray are obtained from bone which provides little to no signal in MRI. Furthermore, because air also provides no signal in MRI the intensity ranges of both materials overlap which accounts for an even more complicated differentiation. The information for the generation of accurate intensity values can, therefore, solely be drawn from the structural information that is present in the image. In case of volumetric imaging the materials may be unknown prior to the synthesis but they are resolved in distinct regions in the image. In contrast, in projection imaging this structural information diminishes by integration of the intensity or attenuation values on the detector. This corresponds to a linear combination of multiple slice images with unknown path length which further increases the difficulty of the synthesis task. To the best of our knowledge, no solution to this problem was proposed up to now. Therefore, we investigate a solution to generate X-ray projections from corresponding MRI views through image-to-image translation.

Further author information:
Bernhard Stimpel: bernhard.stimpel@fau.de

Medical Imaging 2019: Image Processing, edited by Elsa D. Angelini, Bennett A. Landman, Proc. of SPIE
Vol. 10949, 109492K · © 2019 SPIE · CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2512195

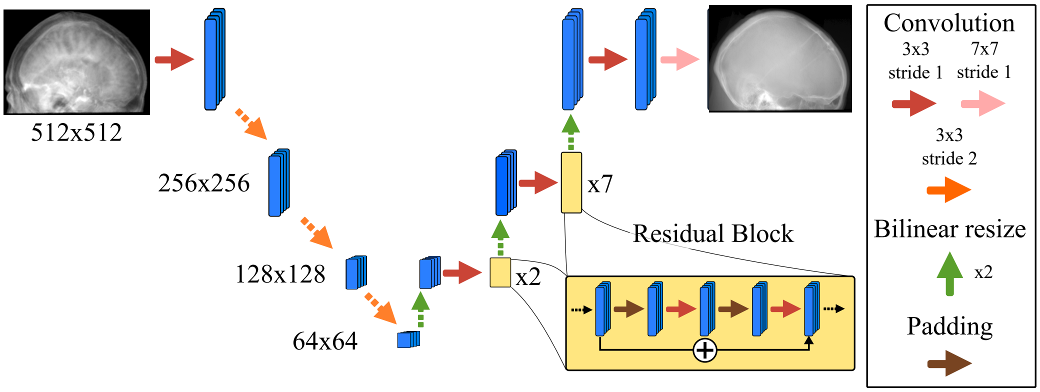


Figure 1: The proposed architecture for the generator network.

2. METHODS

Current methods for the synthesis of volumetric CT data from corresponding MR scans are typically atlas- or learning based. Atlas-based methods, however, are difficult to apply to projection imaging due to the high variance in appearance resulting from differences in the projection geometry. In contrast, learning-based solutions exploit image-level features only. Considering the aforementioned difficulties regarding the synthesis of X-ray projections from MR signal, the use of generative adversarial networks (GANs) is suitable. The adversarial design allows for a more extensive exploration of the possible solution space. Consequently, we propose a deep learning solution using a generative adversarial network for this image-to-image-translation task.

2.1 Network architecture

A feed-forward network is used as the image generator. Our network's design is based on the architecture proposed by Johnson et al.,[5] which was also adopted for various other applications, e.g.,.[6] The architecture is designed in an encoder-decoder fashion and is fully convolutional. In the original manuscript, multiple residual blocks are introduced at the lowest resolution level of the network. This results in the accumulation of a large portion of the available model capacity at these coarse resolution layers that determine the general structures and their arrangement in the image. Due to the vast amount of possibilities in natural imaging synthesis this may be intuitive. However, the underlying variance in medical projection images is only small compared to natural image scenes. Additionally, during interventional treatments, valuable information is largely drawn from high-frequency details such as contrast and clear edges. These indicate the outline of bones, medical devices, and similar structures. To shift the capabilities of the generator network in favor of these high-frequency structures, we distribute the residual blocks across higher resolution levels instead. Furthermore, bilinear upscaling is used in place of the transposed convolution operation, which was recently related to checkerboard artifacts.[7] A visualization of the final architecture is shown in Fig. 1.

2.2 Objective function

The main part of the proposed objective function is based on the discriminator network which is trained to separate the generated fake images from the real label images. For the discriminator we stick with the proven architecture proposed in.[8] Although an adversarial loss is powerful, it is also less constrained to the target image than conventional cost functions. Resulting from this, it is often used in combination with a second metric. The generator network is designed such that the focus is on generating accurate high-frequency structures. Consequently, the objective function for the optimization process must be chosen accordingly. Pixel-wise metrics, e.g., L1- or L2-norm, which are frequently used for attenuation map generation in radiotherapy, do not satisfy this requirement and are related to blurrier results in comparison.[9] Considering the importance of high-frequency structures, using a feature matching loss as proposed by[5] is suitable. This loss functions is based on the extraction and comparison of high-level image features between the generated and the reference image. In recently published work,[10] concluded that utilizing the VGG-19 network pre-trained on ImageNet for the computation of this feature matching loss is appropriate for medical projection images. The aforementioned high-level image features that are used to compute the feature matching loss exhibit many edges and similar structures. However, the majority of the projection images consist of homogeneous regions. Additional emphasis of high-frequency details is achieved by including an edge-weighting to the loss computation. First, a gradient map of the label image is computed using the Sobel filter.[11] Second, this gradient map is used to weight the loss such that the loss generated from edges is emphasized and that from homogeneous regions is attenuated. Starting with the GAN-loss, this can be formulated mathematically as

$$\ell_{\mathrm{GAN}}(\boldsymbol{L}, \boldsymbol{G}, D) = \mathbb{E}_{\boldsymbol{L},\boldsymbol{G}}\left[\log D(\boldsymbol{L}, \boldsymbol{G})\right] + \mathbb{E}_{\boldsymbol{L},\boldsymbol{G}}\left[1 - \log D(\boldsymbol{G})\right] \tag{1}$$

where D is the discriminator network and $\boldsymbol{L}$ and $\boldsymbol{G}$ are the label and generated image, respectively. The second part of the loss function is the feature matching loss described by

$$\ell_{\mathrm{FM}}(\boldsymbol{L}, \boldsymbol{G}) = \sum_{s}^{S} \left(\boldsymbol{V}_s(\boldsymbol{L}) - \boldsymbol{V}_s(\boldsymbol{G})\right) \ , \tag{2}$$

where $\boldsymbol{V}_s(\boldsymbol{L})$ and $\boldsymbol{V}_s(\boldsymbol{G})$ are the feature activation maps of the VGG-19 network at the layer $s \in S$. This leads to the final objective functions

$$\ell(\boldsymbol{L}, \boldsymbol{G}, D) = \left(\ell_{\mathrm{GAN}}(\boldsymbol{L}, \boldsymbol{G}, D) + \ell_{\mathrm{FM}}(\boldsymbol{L}, \boldsymbol{G})\right) \cdot \boldsymbol{E}_{\boldsymbol{L}} \ , \tag{3}$$

which is the combination of the GAN and feature matching loss weighted by the gradient map $\boldsymbol{E}_{\boldsymbol{L}}$ of the label image.

2.3 Data and Experiments

Both, MRI and CT scans of four individuals with different pathologies were provided (MR: 1.5 T MAGNETOM Aera / CT: SOMATON Definition, Siemens Healthineers, Erlangen / Forchheim, Germany). The tomographic data is registered using 3D Slicer and forward projections are generated using the CONRAD framework.[12] All projections are zero-centered and normalized prior to training. For the input data, i.e., the MRI projections, this preprocessing is applied on each subject individually and not on the whole dataset to account for differences in the MR protocols. Training was performed for a fixed number of 400 epochs using the ADAM optimizer. Evaluation of the proposed approach is performed quantitatively as well as qualitative. Because of the limited data available, a 4-fold cross validation is computed, i.e., for each evaluation the projections of three patient data sets are used for training and one for testing. To this end, projections from a 180° rotation in the transversal plane are created with a projection geometry that closely resembles common clinical X-ray systems. For evaluation, the mean squared error (MSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) are calculated. Furthermore, we investigate how the performance with respect to these metrics depends on the projection angle. All projections are normalized beforehand. For MSE and PSNR only pixel are considered that are nonzero in the label images to limit the optimistic bias caused by the large homogeneous air regions. We also compare our approach to the originally proposed architecture as used for example in,[5,6] which we will refer to as "baseline".

3. RESULTS

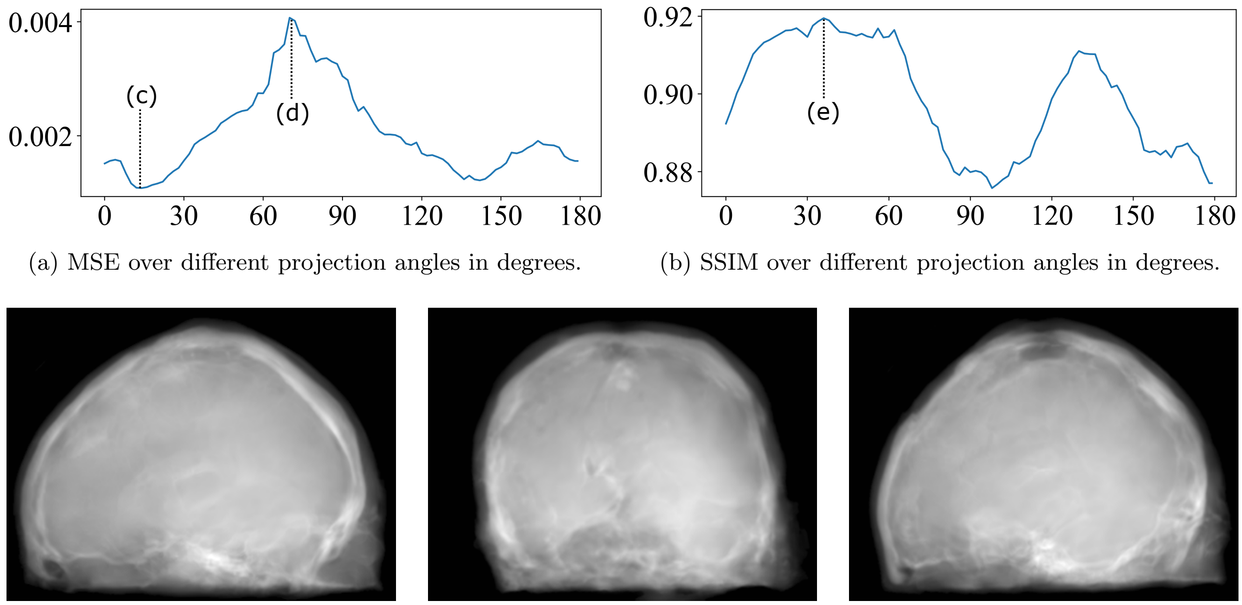
The proposed approach was successful in generating X-ray projections with a contrast similar to the one seen in true fluoroscopic X-ray images. Quantitative results of the generated projection images for all patients are presented in Tab. 1 and for the different network architectures in Tab. 2. In Fig. 2a to 2e the behavior of the MSE and SSIM w.r.t the projection angle is presented. Additional qualitative results of the proposed projection image-to-image translation pipeline for different patient data sets are shown in Fig. 3a to 3c. In Fig. 4 the influence of the modified network architecture, as well as the weighted loss w.r.t. to the edge map are presented.

Table 1: MSE, SSIM, and PSNR of our edge-weighted approach for all datasets.

	MSE	SSIM	PSNR
Patient 1	0.007 ± 0.001	0.894 ± 0.026	21.61 ± 0.93
Patient 2	0.006 ± 0.003	0.898 ± 0.015	22.49 ± 2.03
Patient 3	0.010 ± 0.002	0.892 ± 0.013	20.31 ± 0.99
Patient 4	0.017 ± 0.004	0.872 ± 0.014	17.79 ± 1.08

Table 2: MSE, SSIM, and PSNR of the different network architectures.

	MSE	SSIM	PSNR
Reference Architecture	0.010 ± 0.004	0.889 ± 0.015	20.22 ± 1.78
Ours w/o edge-weighting	0.009 ± 0.002	0.884 ± 0.011	20.50 ± 1.14
Ours w/ edge-weighting	**0.006 ± 0.003**	**0.898 ± 0.015**	**22.49 ± 2.03**



(a) MSE over different projection angles in degrees.

(b) SSIM over different projection angles in degrees.

(c)

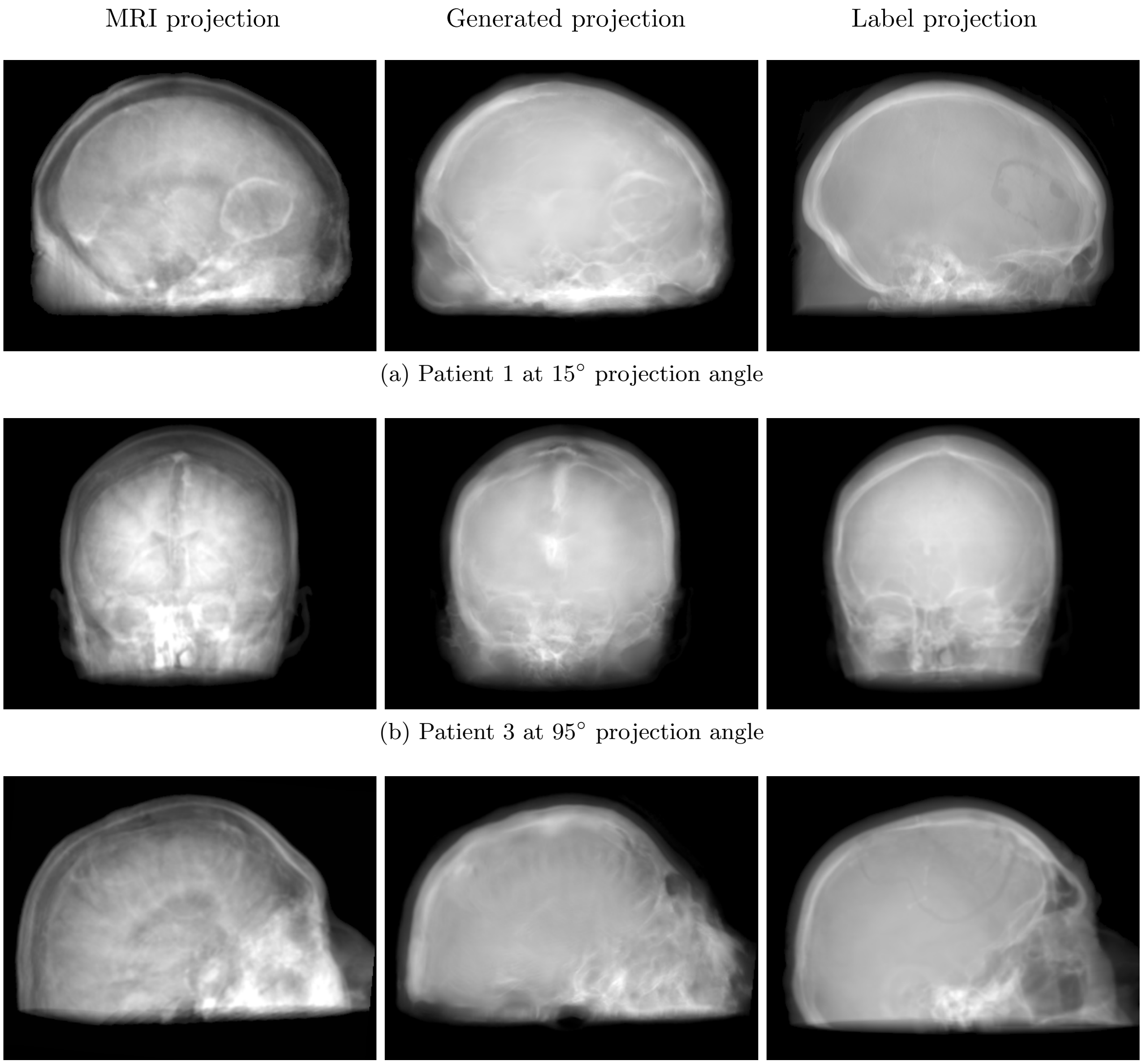
(d)

(e)

Figure 2: Evaluation metrics of the projections along a circular trajectory (a-b). Example projections at selected projection angles (c-e).

4. DISCUSSION

The improvement in our method compared to the baseline method is demonstrated by a decreased MSE, increased SSIM, and PSNR in Tab. 2. When examining Fig. 4a to 4c, improvements can be observed in the overall increased contrast of high-frequency details. Using the originally proposed architecture [5,6], which gathers the residual blocks at the lowest resolution level, results in overall blurrier results and missing bone structures as seen in Fig. 4a. In contrast, the projections generated with the edge-weighted loss resemble the label images more closely. This can especially be observed at the base of the head. The projections created without the weighting also produce many high-frequency details in this region, however, these are less specific in comparison with the edge-weighted results. This results in decreased MSE and increased SSIM and PSNR of the projections synthesized using our approach. In addition, unnatural holes in the brain are generated by the baseline architecture. A possible explanation for the fluctuations in the error measure shown in Fig. 2a and 2b is that in our trajectory in the angles around 45 and 135 degrees the projection rays are cast from the side through the brain while around 90 and 180 degrees the angle of incidence is from the front or back side of the skull. In the first case this results in projections that exhibit large homogeneous areas which are easier to synthesize. In the second case, however, the high-frequency edges from the eyes, jawbone, etc. are the dominant structures in the image. A limiting factor of this study is the low number of patient datasets available. However, the amount of variation introduced

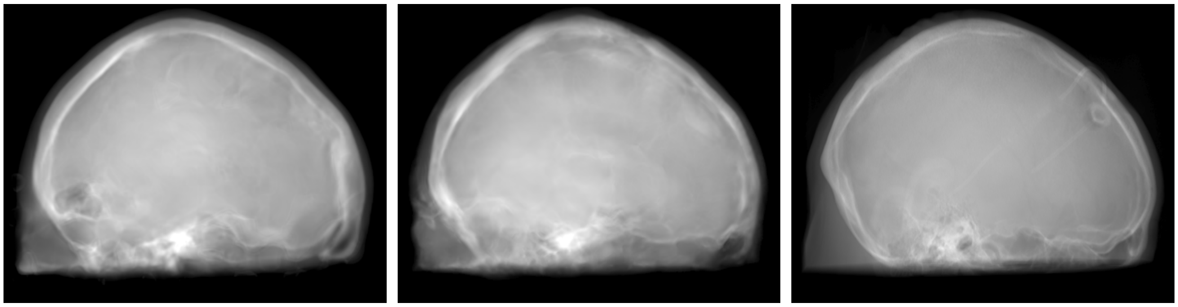


(a) Patient 1 at 15° projection angle

(b) Patient 3 at 95° projection angle

(c) Patient 4 at 360° projection angle

Figure 3: Representative examples of the projection image-to-image translation for different projection angles and patients.



(a) Baseline

(b) Our edge-weighted

(c) Label X-ray

Figure 4: Comparison of the projections generated by the different network architectures based on Patient 2.

by forward projecting the volumes is large. Varying projective geometries account for distinctively different structural appearance of the resulting projections. What is of course not covered by these transformations are unique characteristics of individual patients or different pathologies. To investigate the possible translation outcome of these properties larger datasets are required in the future. Also details that are not visible in the MRI projections can not be transferred to the generated images. An example would be interventional devices that are X-ray but not MR sensitive. Regarding subsequent post-processing applications, the question arises how this missing information in the generated projection images should be dealt with, which is subject to future work.

5. CONCLUSION

We presented an approach to synthesize X-ray projection images from corresponding MRI projections. The proposed redistribution of model capacity at higher resolution layers and the weighting of the computed loss by a gradient map show clear improvements over the baseline method derived from natural image synthesis. Increasing the dataset size in subsequent work could help to translate also patient specific details, e.g., pathologies, between the domains. With future advances in hybrid X-ray and MR imaging, this domain transfer can be used to apply valuable post-processing methods.

Acknowledgment: This work has been supported by the project P3-Stroke, an EIT Health innovation project. EIT Health is supported by EIT, a body of the European Union. Additional financial support for this project was granted by the Emerging Fields Initiative (EFI) of the Friedrich-Alexander University Erlangen-Nürnberg (FAU). Furthermore, we thank the NVIDIA Corporation for their hardware donation.

REFERENCES

[1] Fahrig, R., Butts, K., Rowlands, J. A., Saunders, R., Stanton, J., Stevens, G. M., Daniel, B. L., Wen, Z., Ergun, D. L., and Pelc, N. J., "A truly hybrid interventional MR/x-ray system: Feasibility demonstration," *J. Magn. Reson. Imaging* **13**(2), 294–300 (2001).

[2] Navalpakkam, B. K., Braun, H., Kuwert, T., and Quick, H. H., "Magnetic resonance-based attenuation correction for PET/MR hybrid imaging using continuous valued attenuation maps," *Invest. Radiol.* **48**(5), 323–332 (2013).

[3] Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., and Shen, D., "Medical Image Synthesis with Context-Aware Generative Adversarial Networks," *Med. Image Comput. Comput. Interv. - MICCAI* , 417–425 (2017).

[4] Syben, C., Stimpel, B., Leghissa, M., Dörfler, A., and Maier, A., "Fan-beam Projection Image Acquisition using MRI," in [*3rd Conf. Image-Guided Interv. Fokus Neuroradiol.*], 14–15 (2017).

[5] Johnson, J., Alahi, A., and Fei-Fei, L., "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," *arXiv:1603.08155* (2016).

[6] Wang, T.-C., Zhu, M.-Y. L. J.-Y., Tao, A., Kautz, J., and Nov, C. V., "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," *arXiv:1603.08155* (2017).

[7] Odena, A., Dumoulin, V., and Olah, C., "Deconvolution and Checkerboard Artifacts," *Distill* **1**(10) (2016).

[8] Zhu, J. Y., Park, T., Isola, P., and Efros, A. A., "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in [*Proc. IEEE Int. Conf. Comput. Vis.*], 2242–2251 (2017).

[9] Dosovitskiy, A. and Brox, T., "Generating Images with Perceptual Similarity Metrics based on Deep Networks," *Adv. Neural Inf. Process. Syst. 29* , 658–666 (2016).

[10] Stimpel, B., Syben, C., Würfl, T., Mentl, K., Dörfler, A., and Maier, A., "MR to X-Ray Projection Image Synthesis," in [*Proc. Fifth Int. Conf. Image Form. X-Ray Comput. Tomogr.*], (2017).

[11] Sobel, I. and Feldman, G., "A 3x3 Isotropic Gradient Operator for Image Processing," in [*Stanford Artif. Intell. Proj.*], (1968).

[12] Maier, A., Hofmann, H. G., Berger, M., Fischer, P., Schwemmer, C., Wu, H., Müller, K., Hornegger, J., Choi, J.-H., Riess, C., Keil, A., and Fahrig, R., "CONRAD - A software framework for cone-beam imaging in radiology," *Med. Phys.* **40**(11) (2013).