# Deep Learning-based Inpainting for Virtual DSA

Mathias Unberath, Jonas Hajek, Tobias Geimer, Frank Schebesch, Mario Amrehn, and Andreas Maier

## I. Introduction

Digital subtraction angiography (DSA) is a well-established imaging technique for background removal [1] in interventional X-ray-based angiography. For DSA, projections of the scene without contrast agent (mask scan) are digitally subtracted from projections acquired during contrast injection (fill scan), yielding images of the contrasted lumen only. Unfortunately, DSA can only be applied successfully to invariant anatomies as motion between the acquisition of mask and fill scan introduces misalignment artifacts that deteriorate diagnostic value. For cardiac imaging, this requirement is particularly problematic as the coronary arteries are subject to constant cardiac and respiratory motion [2].

Popular methods seek to avoid the problems induced by inter scan motion as a whole by estimating mask images from fill scans [2], [3], a technique that we will refer to as virtual DSA (vDSA). These methods require vessel segmentation to identify regions for image inpainting, i. e. background estimation. While the methods employed for vessel segmentation differ, the variation in inpainting algorithms is rather limited. Blondel *et al.* and Unberath *et al.* use morphological closure and spectral deconvolution, respectively, to estimate the background image [2], [3]. These methods work well for small regions of interest (ROIs) but usually perform poorly when large areas need to be estimated.

Recently, methods from machine learning based on denoising autoencoderes and convolutional neural networks (CNNs) have received increasing attention, and were applied successfully to image inpainting [4], [5]. State-of-the-art results on natural images are impressive, however, the applicability to medical images, particularly to vDSA has yet to be demonstrated.

In this paper, we adapt the U-net CNN architecture [6] to allow for regression, and apply it to image inpainting for background estimation in vDSA imaging.

## II. Materials and Methods

*Vessel Segmentation:* An image showing contrasted vessels $I$ is processed by Hessian-based segmentation algorithms [2] yielding a binary mask $M$ that is unity for pixels identified as contrast agent and zero else. Pixel-wise multiplication of $I$ and the inverse mask $M$ results in a defect image that serves as input to the U-net described below. A dilated version of the segmentation mask is denoted by $M^d$.

*U-net regression:* Previous approaches [4], [5] avoid downsampling because of the desired high resolution outcome. However, downsampling via pooling is beneficial as it yields more generic features. To combine downsampling and a high resolution output via skip-ahead connections, we employ the U-net [6]. Fig. 1 states all relevant parameters such as the number of features and their respective sizes. The U-net consists of a contracting path, where downsampling layers decrease the feature sizes with increasing depth of the network. Therefore, each level consists of two convolution layers followed by a max pooling downsampling layer. Within one max pooling operation, the size of the features is decreased by a factor of two. In return, the next convolution layer doubles the number of features. Therefore, the features lose spatial but gain contextual information. Moreover, the U-net has an expanding path that uses upsampling layers to propagate contextual information to higher resolution levels. An upward step halves the number of features while doubling their dimensions. The output of this layer is then combined with features from the downsampling layer via skip-ahead connections (blue arrows in Fig. 1). The structure of the horizontal level is similar to the downwards path, again with two convolution layers. The output $X(\theta)$ of the network depends on its parameters $\theta$ is compared to the ground truth of the input using a mean-squared-error loss.

We distinguish two stages: (i) Inpainting occurs only in a very small region defined by the segmentation mask; this approach will be referred to as *mask-guided*. For a batch of $N$ images, the batch loss $L_b$ reads

$$L_b(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{|M_i^d|}\sum_{\boldsymbol{p}\in\Omega} M_i^d(\boldsymbol{p})\big\| Y_i(\boldsymbol{p}) - [X_i(\theta)](\boldsymbol{p})\big\|_2\right),$$

where $|M|$ denotes the number of non-zero pixels in $M$, $\Omega$ is the image domain, and $Y$ is the no-contrast ground truth image. (ii) The network trained in (i) is refined to restore the complete image rather than segmented pixels only. The batch loss stated above is modified by omitting the weighting with

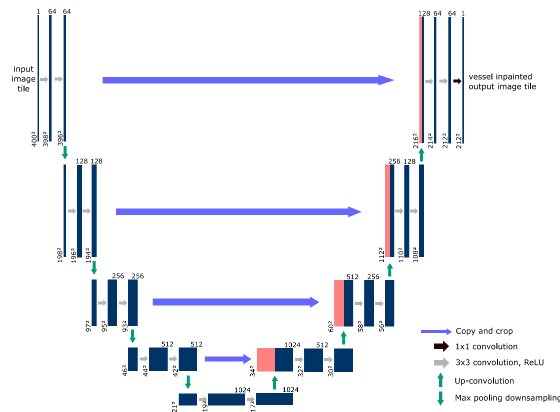


Fig. 1. U-net architecture employed here.

M. Unberath, J. Hajek, T. Geimer, F. Schebesch, M. Amrehn, and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, GER. E-mail: mathias.unberath@fau.de.

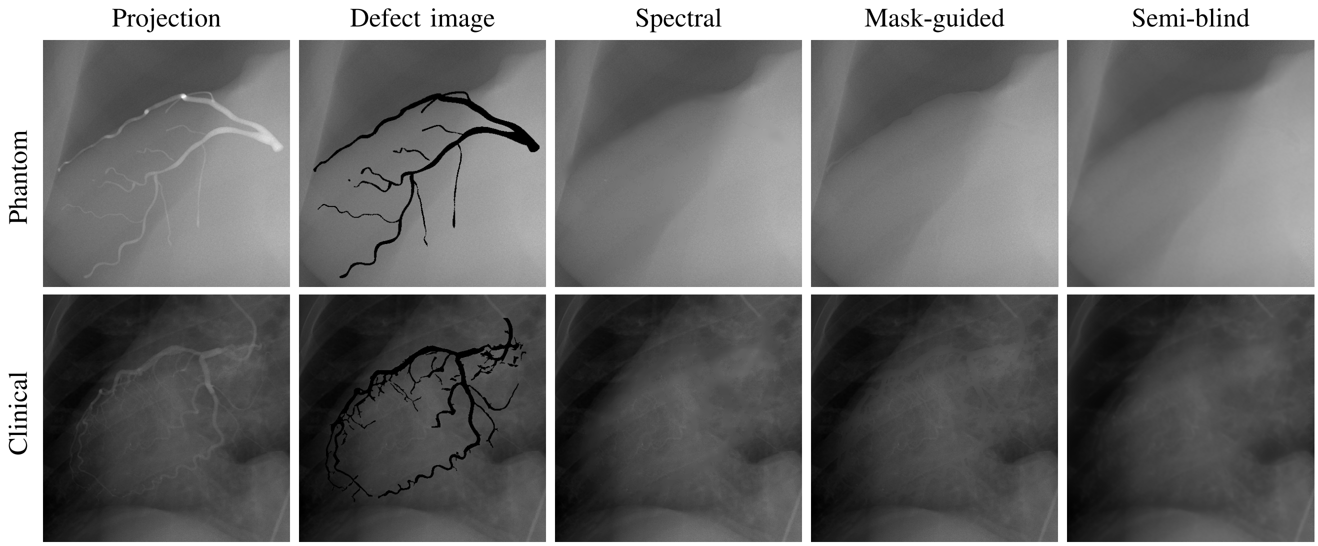


Fig. 2. Representative results of the inpaiting algorithms. From left to right, the columns show the input image, the defect image, and inpainting results achieved with spectral deconvolution as in [2], and the *mask-guided* and *semi-blind* U-net proposed here. We show a high noise realization for the phantom.

the mask $M^d(\boldsymbol{p})$. This approach is similar to *blind inpainting* as in [5], but still requires segmentations to create the defect image; thus, we call it *semi-blind*. The loss is optimized via stochastic gradient descent and back-propagation.

*Data and Experiments:* For training, we use the well-known numerical XCAT phantom [7] to simulate projection images at different viewing angles and noise realizations with and without contrasted arteries. The images and corresponding segmentation masks are separated in training, test, and validation sets, that are then partitioned into overlapping tiles of $400 \times 400$ pixels. Moreover, we use common data augmentation strategies, i. e. rotations, before partitioning, to artificially increase the training set size. We train U-nets for the *mask-guided* and *semi-blind* inpainting task and state the structural similarity (SSIM) [8] of our predictions with the ground truth within $M^d$, averaged over the test set. Moreover, we apply the networks trained on numerical phantom data to a clinical angiography data set. All results are compared to spectral deconvolution [2].

## III. Results and Discussion

Quantitative results on the phantom data are presented in Table I. Moreover, we show representative inpainting results on phantom and clinical data in Fig. 2. Objectively, performance of the *semi-blind* U-net is superior to *mask-guided* network and spectral deconvolution in the low noise case, and comparable to spectral deconvolution in the low noise realizations. Despite only being trained on phantom data, the performance on clinical data is promising. Due to the low complexity of U-net-based inpainting compared to the spectral method, we believe that these results encourage further research that should address possibilities to completely omit the need for segmentation masks and, hence, enable true *blind* inpaiting.

TABLE I
The average SSIM over all projections.

| | Spectral | Mask-guided | Semi-blind |
|---|---|---|---|
| **High noise** | 0.74 | 0.69 | 0.78 |
| **Low noise** | 0.97 | 0.95 | 0.96 |

## IV. Conclusion

We presented an image inpainting method for coronary arteries in X-ray projections using CNNs. The performance of the proposed methods was comparable to competing approaches both on phantom and clinical data and encourages further research targeted at enabling true *blind* inpainting.

## References

[1] W. A. Chilcote, M. T. Modic, W. A. Pavlicek, J. Little, A. J. Furlan, P. Duchesneau, and M. Weinstein, "Digital subtraction angiography of the carotid arteries: a comparative study in 100 patients." *Radiology*, vol. 139, no. 2, pp. 287–295, 1981.

[2] M. Unberath, A. Aichert, S. Achenbach, and A. Maier, "Consistency-based Respiratory Motion Estimation in Rotational Angiography," *Med Phys*, 2017, (in press).

[3] C. Blondel, G. Malandain, R. Vaillant, and N. Ayache, "Reconstruction of coronary arteries from a single rotational X-ray projection sequence." *IEEE Trans Med Imag*, vol. 25, no. 5, pp. 653–63, 2006.

[4] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Adv Neural Inf Process Syst*, 2012, pp. 341–349.

[5] N. Cai, Z. Su, Z. Lin, H. Wang, Z. Yang, and B. W.-K. Ling, "Blind inpainting using the fully convolutional neural network," *Visual Comput*, pp. 1–13, 2015.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*. Springer, 2015, pp. 234–241.

[7] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Med Phys*, vol. 35, no. 8, pp. 3800–3808, 2008.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans Image Process*, vol. 13, no. 4, pp. 600–612, 2004.