

PYRO-NN: Python Reconstruction Operators in Neural Networks

Christopher Syben,^{1,*} Markus Michen,¹ Bernhard Stimpel,¹
Stephan Seitz,¹ Stefan Ploner,¹ and Andreas K. Maier¹

¹*Pattern Recognition Lab, Friedrich-Alexander Universität Erlangen-Nürnberg, 91058 Erlangen, Germany*

Purpose: Recently, several attempts were conducted to transfer deep learning to medical image reconstruction. An increasingly number of publications follow the concept of embedding the CT reconstruction as a known operator into a neural network. However, most of the approaches presented lack an efficient CT reconstruction framework fully integrated into deep learning environments. As a result, many approaches are forced to use workarounds for mathematically unambiguously solvable problems.

Methods: PYRO-NN is a generalized framework to embed known operators into the prevalent deep learning framework Tensorflow. The current status includes state-of-the-art parallel-, fan- and cone-beam projectors and back-projectors accelerated with CUDA provided as Tensorflow layers. On top, the framework provides a high level Python API to conduct FBP and iterative reconstruction experiments with data from real CT systems.

Results: The framework provides all necessary algorithms and tools to design end-to-end neural network pipelines with integrated CT reconstruction algorithms. The high level Python API allows a simple use of the layers as known from Tensorflow. To demonstrate the capabilities of the layers, the framework comes with three baseline experiments showing a cone-beam short scan FDK reconstruction, a CT reconstruction filter learning setup, and a TV regularized iterative reconstruction. All algorithms and tools are referenced to a scientific publication and are compared to existing non deep learning reconstruction frameworks. The framework is available as open-source software at <https://github.com/csyben/PYRO-NN>.

Conclusions: PYRO-NN comes with the prevalent deep learning framework Tensorflow and allows to setup end-to-end trainable neural networks in the medical image reconstruction context. We believe that the framework will be a step towards reproducible research and give the medical physics community a toolkit to elevate medical image reconstruction with new deep learning techniques.

1.. INTRODUCTION

In recent years, major breakthroughs made deep learning increasingly prevalent in more and more fields. It revolutionizes the way of classification and regression tasks in speech and image recognition¹⁻³ and many other areas. Even in the medical domain, where interpretability and reliability are one of the most important driving forces, deep learning has led to astonishing results⁴. One of the most cited papers of recent years is the U-net⁵ which outperforms classical machine learning algorithms in segmentation tasks. In the subsequent time, the U-net architecture emerged to many more tasks, e.g., artifact correction, image fusion, image-to-image translation, and even into the context of medical image reconstruction⁶⁻⁸. However, this domain is fundamentally different from those in which the advent of deep learning began, and the question arises as to whether these learned signal reconstruction pipelines are reliable and stable enough for a critical area such as medical imaging⁹. Two special issues: 'Deep learning in medical imaging'¹⁰ and 'Machine Learning for Image Reconstruction'¹¹ in Transactions on

Medical Imaging (TMI) in 2016 and 2018 discuss the increasing relevance of deep learning methods in medical image reconstruction.

The presented approaches can be divided into either pre- or post-processing approaches or fully end-to-end trained methods. For the first type the actual reconstruction pipeline is based on well known signal reconstruction algorithms omitting the end-to-end capability due to its complexity. For the second type of approaches the modeling of the end-to-end pipeline can be realized under two different paradigms. One way is to learn the whole signal processing pipeline, an exceptionally clear representative of this paradigm is *AUTOMAP*¹². Directly in contrast to this is the emerging paradigm of embedding known operators¹³. This preserves the end-to-end learning capability but includes the known operations of the reconstruction chain to preserve the credibility of the signals, reduce the error bound of the learning process and decrease the number of parameters and thus the amount of necessary training data. This paradigm gets increasingly popular, with multiple publications following the way of embedding known operators in the Computed Tomography

(CT) context and successfully including the CT reconstruction as known operators into the network architecture to be able to benefit from the end-to-end training capability of deep learning¹⁴⁻²¹. However, the publications that follow this path are still less represented than those that use deep learning only as pre- or post-processing. We believe that a major reason for this is the non-trivial implementation of known operators in existing deep learning frameworks. Even publications that successfully take on this challenge often refer to their own implementations as prototypical¹⁵ or provide frameworks on abstract wrapped levels^{16,22}. An efficient and publicly usable solution integrated into one of the popular deep learning frameworks, however, remains pending.

To strengthen the paradigm of known operators, elaborate the research in the medical image reconstruction, and to avoid reimplementations and incompatibilities, we started to work on an open source software framework PYRO-NN, which allows an easy way to integrate known algorithms into the deep learning framework Tensorflow²³. We provide multiple forward and backward projectors for CT implemented in CUDA based on scientific publications supported with a high level Python API for simple use of state-of-the-art CT reconstruction, even from different setups of real CT scanners. The profound integration into Tensorflow on C++/Cuda level allows to handle occurring performance and memory issues and, additionally, allows an easy customization of the algorithms compared to a wrapper alternative like^{22,24}. Furthermore, the high level Python API offers an easy link between deep learning and community driven frameworks. For the CT domain this allows to use a wide range of tools (e.g. filter, redundancy weights, etc.)²⁴⁻²⁷.

We believe that this framework will help the community leverage the power of end-to-end training of machine learning algorithms directly from the data, while continuing to apply mathematically sound solutions to uniquely solvable problems.

II.. METHODS

The framework concept is designed to easily include C++ and CUDA based algorithms into the deep learning framework Tensorflow. In detail, PYRO-NN provides network layers as CUDA implementations to generate parallel-, fan-, and cone-beam X-ray projections and to reconstruct them within any neural network constructed with Tensorflow. Due to the nature of the projection and reconstruction operation we intrinsically provide the analytical gradi-

ents for all of these layers, which allows fully end-to-end trainable networks. Furthermore, with PYRO-NN we provide filters and weights based on scientific publications to allow proper filtered-backprojection (FBP) reconstructions. The PYRO-NN API is inspired by the CONRAD²⁶ framework to adapt the ability to reconstruct data from real clinical scanners and by using PyConrad²⁷ many more tools and phantoms can easily be used in the deep learning context. The current state of the framework features a CT reconstruction pipeline, while the basic design allows to transfer the whole concept to other signal reconstruction domains within one framework and, therefore, points out a direction to future development and community contribution.

This section is structured as follows: First we introduce the general design of the framework, followed by a description of how to create the link to the Tensorflow framework. In the second part we introduce the operator and algorithms necessary to conduct CT reconstruction and describe the two levels of the framework: the actual layer implementation on the C++/CUDA level and the high level Python API providing a convenient usage of the provided layers.

II.A.. Software Design / Rationale

The development speed in the deep learning community is tremendous. Like in the research itself, the toolkits and frameworks are developing in the same speed, which often causes conflicts in interoperability of self-developed solutions and version mismatches between different frameworks and toolkits. To ensure a robust version control, the framework is directly included into the building process of the Tensorflow sources. In Fig. 1 the process is shown. We provide a patch for all Tensorflow versions since 1.9 which adjust the build process of the Tensorflow sources such that all C++ and CUDA files in the framework directory are built and included into the Tensorflow Python package under the PYRO-NN namespace. This yields the benefit that the user can define specific version combinations for their purpose. Furthermore, the separation of the custom implemented algorithms and the information flow allows the possibility to extend support of the proposed framework towards other deep learning frameworks like PyTorch.

The architecture is mainly defined by the API to Tensorflow. In Fig. 2 the abstract concept is visualized. Known operators can be implemented as kernels consisting of a C++ class following the design of the Tensorflow API and the actual implementation of the operator as a kernel. The C++ class ensures

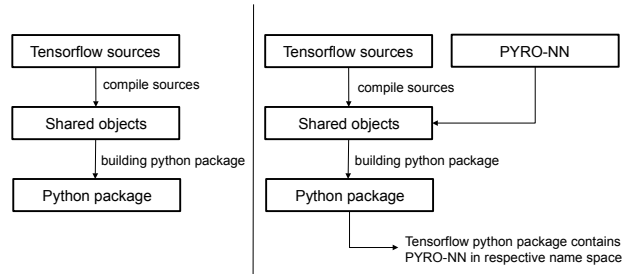


FIG. 1. The build process. Left the build process defined by Tensorflow, right our adjusted process.

the behavior of the kernel as a Tensorflow layer in the neural network. In contrast to other frameworks which wrap the implementation on the Python level, this gives full control over the device resources like memory usage.

The third part is the high level Python API. Due to the API defined by Tensorflow the user is able to use the known operators like normal Tensorflow layers with the defined input, output and attributes. A crucial aspect is the registration of the gradient to the respective layer. This allows to invoke the relevant algorithm to compute the gradient with respect to the layer in an efficient way. The provision of the gradient is a necessity to enable a gradient flow through the entire network and, thus, allow fully end-to-end trainable networks with known operators. The high level Python API allows to design helper structures for the respective operator. For the CT context, we provide wrapped layers with automated gradient registration and object oriented classes to feed the known operators the necessary attributes. The class design follows the high level abstraction of Tensorflow and wraps the implementation detail to provide the user an easy and convenient usage while preserving the possibility for the user to modify the wrapped methods.

All together, these rationals offer the community with a generic, version stable, framework to easily include known operators into neural networks. The source code is publicly available under the Apache 2.0 licence to be directly compatible with Tensorflow and to allow an uncomplicated integration of the community into existing projects. In the following the implementation of these rationals are shown on our case study of CT reconstruction, which provides a broad band of tools to allow end-to-end networks with reconstruction capability.

II.B.. CT Reconstruction in Neural Networks

Based on the generic design of the framework, the current state provides all necessary algorithms and tools for analytical parallel-, fan- and cone-beam reconstruction. The necessary algorithms are implemented within Tensorflow as an own layer, while the respective tools, e.g., filter, weights, etc. are provided on the Python level to supply a high level API for CT reconstruction. In the following, we introduce the mapping of the known operator to a layer, followed by a description of the provided algorithms and tools.

II.B.1.. The Known Operator

For the task of reconstructing object information from acquired X-ray projections, an efficient analytical method is well known and is called Filtered-Backprojection (FBP). To embed these methods into a neural network, the whole acquisition and reconstruction procedure of a CT system needs to be described with discrete linear algebra to embed them into a neural network. The acquisition of projection data of the object can be described with

$$\mathbf{Ax} = \mathbf{p} \quad , \quad (1)$$

where \mathbf{A} is the matrix describing the geometry, the so called system matrix which can be algorithmically implemented as the forward-projection operator. \mathbf{x} is the object and \mathbf{p} are the acquired projections of object \mathbf{x} under the system described by \mathbf{A} . The reconstruction can be obtained with

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{p} \quad , \quad (2)$$

where \mathbf{A}^{-1} is the inverse system matrix, which cannot be inverted because it is a tall matrix. Thus, the

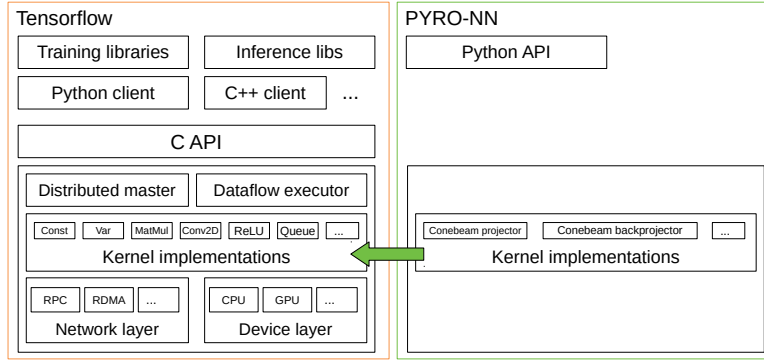


FIG. 2. The architecture of PYRO-NN beside to Tensorflow. Image adjusted from <https://www.tensorflow.org/guide/extend/architecture>.

reconstruction can be conducted using the Moore-Penrose Pseudo Inverse

$$\mathbf{x} = \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{p}, \quad (3)$$

where \mathbf{A}^\top is the adjoint system matrix which can algorithmically implemented as the back-projection operator. Following the analytical filtered back-projection approach, the inverse bracket with $\mathbf{A}\mathbf{A}^\top$ must be a filter operation with the ramp filter. The filtering step can be described by a multiplication with a diagonal matrix \mathbf{K} in the Fourier domain

$$\mathbf{x} = \mathbf{A}^\top \mathbf{F}^H \mathbf{K} \mathbf{F} \mathbf{p}, \quad (4)$$

where \mathbf{F} , \mathbf{F}^H is the Fourier transform and the respective adjoint, i.e. inverse operation. As Eq. 4 shows, the reconstruction of an object from projection data can be expressed completely as discrete linear algebra and, therefore, each matrix multiplication can be modeled as a layer in a neural network. The working horse of the neural network performance is based on the backpropagation algorithm to update the weights and, thus, train the network. To allow the integration of the presented operator, we need to ensure that the operator has at least a sub-gradient, to ensure the capability to propagate gradients through the whole network. The publications from Würfl and Syben et al. show that \mathbf{A} and \mathbf{A}^\top are their respective operators to calculate the gradient, thus the gradient flow through these layers can be ensured^{17,19}.

II.B.2.. The Operator as a Layer

From iterative reconstruction it is well known that the system matrix does not fit into memory, there-

fore, we compute the operator on the fly using ray-based algorithms. There are several ways for the computation. We introduce the *ray-driven forward-projection* and the *voxel-driven back-projection* algorithmically with respect to the integration into Tensorflow. Note that when using a ray-driven forward-projection algorithm to compute the result of the multiplication with \mathbf{A} , then the voxel-driven back-projection algorithm is not the respective adjoint operation \mathbf{A}^\top . They are a so called a unmatched projector-/back-projector pair. The implications of matched projectors and shear-warp projectors on the convergence and runtime are subject to future work and are briefly discussed in Section IV.

The *forward-projection* to generate projections from the input volume are implemented as CUDA kernels in a ray driven manner. For each detector pixel, a ray \vec{r} is cast through the scene, accumulating the absorption values along the line. We provide forward projectors for 2D parallel- and fan-beam geometry based on ray vectors and respective geometry parameters. Furthermore, a 3D cone-beam forward projector based on projection matrices is implemented according to Galigekere et al.²⁸. The projection matrices allow a simple generation of projection data under the geometry of real systems. The CUDA kernels are parallelized over the detector pixels computing the line integral along the ray. For the 3D cone-beam case the framework offers the possibility to choose between a texture and a kernel interpolation mode. While the texture interpolation of the graphics card leads to very short calculation times, the memory management of Tensorflow and the necessity of a pitched memory structure of CUDA for the data to be interpolated, which leads to the point

that the volume to be projected is kept in the memory twice. To provide the community with the possibility of including the projection operator into big networks and with the option for precise memory management, kernel interpolation is provided. The ray-casting algorithm using the kernel-based interpolation is much slower but directly works with the allocated tensors from Tensorflow.

The *back-projection* operators to reconstruct simulated or real projection data are implemented as CUDA kernels in a voxel-driven manner. For each pixel/voxel to be reconstructed, the projection of the point on all projection images is accumulated. The framework provides the respective 2D parallel- and fan-beam back-projection algorithms based on geometry parameters and ray vectors. Following the forward projection, the 3D cone-beam back-projection is based on projection matrices according to Scherl et al.²⁹. This allows to reconstruct data from simulations and real systems as shown in the CONRAD framework²⁶. The back-projection kernel is parallelized over the voxels projecting the respective position on the different detector coordinates interpolating the measured line integral. Like the 3D cone-beam forward-projection implementation, the back-projection kernel is implemented with a texture and kernel interpolation mode to give full control over the memory management to the user.

II.C.. High Level Python API

To supply the community with an easy-to-use version of the described layers, we provide the necessary structure and additional tools like filters, weights, phantoms, etc. within the Python framework. In the following, the outline of the necessary structure to evaluate the layers is shown, followed by a short introduction of the provided tools with their related publications to support a state-of-the-art reconstruction pipeline.

II.C.1.. Reconstruction & Geometry

The high level Python API wraps the provided reconstruction layers in Tensorflow. Thus, the framework registers the respective adjoint operation for the gradient computation automatically. All attributes necessary for the provided forward- and backward-projection layers are covered with a base geometry class and corresponding specialized derived classes. The geometry is defined by *volume shape*, *volume spacing*, *detector shape*, *detector spacing*, *number of projections*, *angular range* and for

fan- and cone-beam geometries the *source to iso-center distance* (SID) and *source to detector distance* (SDD). Additionally, the geometry contains the ray vectors (2D) or the projection matrices (3D) to describe the respective trajectory.

II.C.2.. Phantoms

PYRO-NN contains a set of simple geometric objects, e.g. circle, ellipsoid, sphere and rectangles to easily create a more complex numerical phantom. Furthermore, the framework provides an analytical description of the 2D Shepp-Logan phantom³⁰ as well as a 3D extension based on the CONRAD implementation²⁶.

II.C.3.. Trajectories

The trajectory describes the geometric scanner setup over the whole scan. For the 2D parallel- and fan-beam cases the trajectory is described by the central ray vector for each projection. For the 3D cone-beam case the trajectory is described by a set of projection matrices, which allows to use calibrated projection matrices from real scanner systems. Within the high level Python API we provide basic methods to compute the respective rays or projection matrices based on a given geometry. This general concept gives the user the option to easily use individual trajectories. The open-source concept of the whole framework allows to contribute to the diversity of provided trajectories.

II.C.4.. Filters

To allow a basic reconstruction in the context of neural networks, PYRO-NN provides the Ramp and Ram-Lak filter implemented according to Kak and Slaney³¹. The filters can be directly assigned as weights to a multiplication layer and, thus, can be interpreted as a multiplication with a diagonal matrix in the Fourier domain as shown in Eq. 4.

II.C.5.. Geometric Correction Weights

In order to support fan- and cone-beam reconstructions, the framework contains a cosine-filter to weight the each pixel in the sinogram with its distance to the X-ray source and, thus, correct fan- and cone-beam artifacts. It is implemented according to Kak and Slaney³¹. Furthermore, to support short

and super short scans redundancy weights are provided e.g. Parker³² and Riess³³ weights.

II.D.. Network Architectures

Following the paradigm of precision learning, different network architectures can be setup or even derived as shown by Syben et al.²¹. We provide various examples within the framework to assist users in using the framework. Therefore, we present a simple network able to reconstruct short-scan cone-beam CT according to the Feldkamp-Davis-Kress (FDK) algorithm³⁴. Additionally, an example of learning the correct discretization of the ramp filter is shown.

III.. EXAMPLE APPLICATIONS

We show three baseline experiments to outline the capability of the provided framework. This provides the code base for an easy entry into deep learning-based signal reconstruction and to support comparability within emerging methods.

III.A.. FDK Reconstruction Network

As a proof of concept and validation of the method we provide a FDK reconstruction with Parker weights³² for a short scan geometry formulated as neural network using the proposed Framework. On the one hand, this setup serves to demonstrate the ability of the presented framework to perform a cone-beam reconstruction. On the other hand, the basic network is a good starting point which makes it possible to integrate further deep learning tools in projection or reconstruction domain. The whole pipeline can then be trained in an end-to-end fashion. This allows general data-driven optimization as well as task specific solutions. Note that the error e.g. ℓ_2 -loss is computed in the reconstruction domain and can be used to update weights applied in the projection domain. The architecture is shown in Fig. 3. The operation of the FBP algorithm for cone-beam short-scan geometry can be described with

$$\mathbf{x} = \mathbf{A}^\top \mathbf{F}^\mathbf{H} \mathbf{K} \mathbf{F} \mathbf{W}_{red} \mathbf{W}_{cos} \mathbf{p}, \quad (5)$$

where \mathbf{W}_{cos} is a diagonal matrix describing the pixel-wise independent weighting of the projection data \mathbf{p} with cosine weights. For so called short-scan trajectories some rays are measured twice and some only once. These redundancies will lead to streaking artifacts in the reconstruction if the projection data

is not adequately weighted. \mathbf{W}_{red} weights the projection data appropriately, e.g., for a defined cone-beam short scan the weights by Parker³² can be used. \mathbf{A}^\top , \mathbf{K} , \mathbf{F} and $\mathbf{F}^\mathbf{H}$ are the system matrix, reconstruction filter and Fourier transform as described in Eq. 4. The described network architecture can reconstruct a 200 degree cone-beam CT short-scan with 248 projections properly. In Fig. 4 the reconstruction result of a 3D Shepp-Logan phantom is shown. The line profiles illustrate that the result follows the phantom and the reference FDK reconstruction using CONRAD²⁶.

This architecture can be seen as a baseline, which can be used to insert deep-learning features and nonlinearities in the projection as well as in the reconstruction domain while the gradient can be propagated through the whole network. For example, the redundancy weights can be learned for arbitrary short-scan angles. The update step for the weights according to the backpropagation algorithm is shown by Würfl et al.²⁰.

III.B.. Learn Reconstruction Filter

To highlight the capability to propagate the gradient through the whole network, from reconstruction domain to projection domain, we demonstrate that the framework can be used to learn the discretized version of the analytical derived ramp-filter for reconstruction in an end-to-end network approach.

The network architecture is defined by

$$\mathbf{x} = \mathbf{A}^\top \mathbf{F}^\mathbf{H} \mathbf{K} \mathbf{F} \mathbf{p}, \quad (6)$$

where \mathbf{K} is the reconstruction filter initialized with the ramp filter and the only trainable layer in the network. For an filtered back-projection algorithm the ramp filter applied in the projection domain is the continuous analytical derived solution. Without proper discretization of the filter, offset and cupping artifacts will occur in the reconstructed volume. The proper discretization is the so called RamLak filter³⁵.

We can show that we can use the backpropagation of the network to compute the error on the reconstruction domain and adjusting the filter weights of \mathbf{K} in the projection domain. The result is converging towards the analytical derived discretized RamLak filter as shown in Fig. 5. A detailed description and discussion of the experiment is given by Syben et al.¹⁹. Note that the training of the filter was done with purely numerical data and can directly applied on real data without loss of validity.

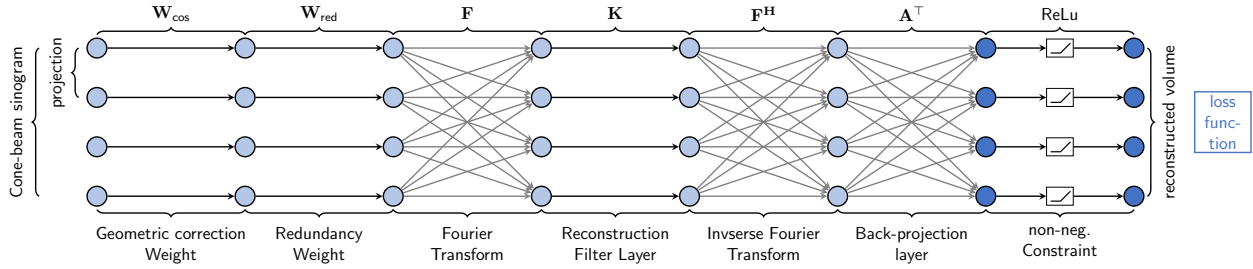


FIG. 3. FDK-Reconstruction Network. Light blue nodes represent the projection domain, while dark blue nodes stands for volume domain.

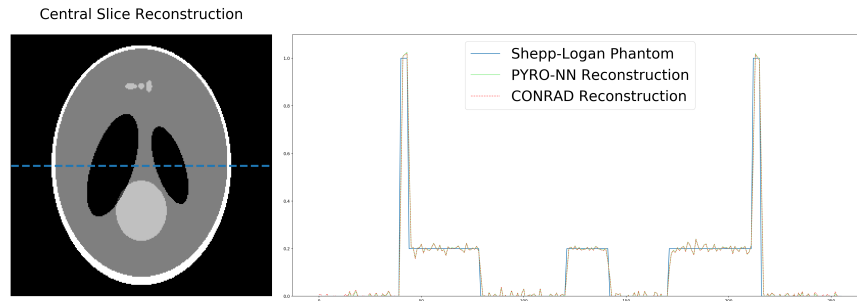


FIG. 4. Short scan FDK reconstruction. The central slice reconstruction shows the result of the network, comparing line plots from phantom (blue), PYRO-NN reconstruction (green) and reference CONRAD reconstruction (red). The gray-level window is $[0, 0.4]$.

III.B.1.. Iterative Reconstruction

With the third experiment we want to demonstrate the versatility and possibilities which are offered by the framework. We can model a network such that an iterative reconstruction algorithm with regularizers can be conducted. The network is designed such that the backpropagation algorithm mimics the iterative reconstruction based on a data fidelity term. Starting with the data fidelity term for iterative reconstruction with total variation (TV) regularizer

$$\min \|\mathbf{Ax} - \mathbf{p}\|_2^2 + \lambda \mathbf{TV}(\mathbf{x}) \quad (7)$$

where \mathbf{A} is the system matrix, \mathbf{x} is the volume to be reconstructed initialized with zero and \mathbf{p} is the measured sinogram. The data fidelity term can be also interpreted as a network architecture, where the input of the network is a zero-vector, followed by an additive layer and then forward projected using the provided forward projection layer as shown in Fig. 6. The loss function is then the ℓ_2 -loss w.r.t to the measured data \mathbf{p} . Training the network will update the weights of the additive layer. Thus, the seeked reconstruction is represented by the weights of the additive layer at the end of the learning process.

cess.

We demonstrate that the network setup can reconstruct a sparse view CT acquisition with 30 parallel projections over an angular range of 180 degree. 2% additive Gaussian noise where applied to the projection data. In Fig. 7 the reconstruction with the iterative PYRO-NN reconstruction is shown and compared to an FBP reconstruction and the Shepp-Logan phantom. Note that the actual TV-regularized iterative reconstruction using the proposed framework needs only few lines of code. The computation of TV is provided by Tensorflow. With this baseline example we provide an easy-to-use starting point for future investigations into learned regularizers and different loss functions.

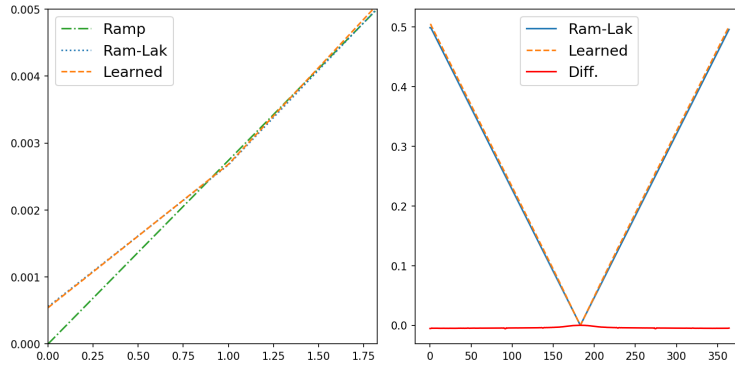


FIG. 5. Ramp, RamLak and Learned filter.

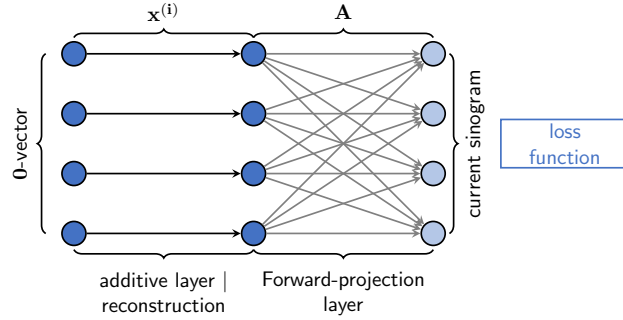
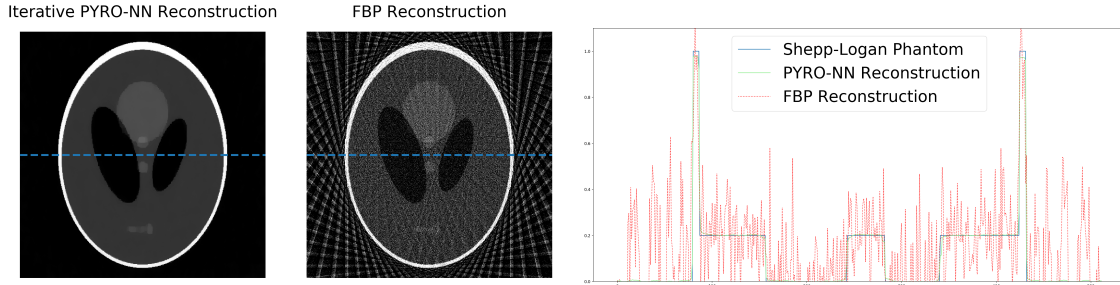


FIG. 6. Iterative reconstruction network. Light blue nodes represent the projection domain while dark blue nodes stand for volume domain.

FIG. 7. Sparse view reconstruction. The reconstruction shows the result of the iterative reconstruction network, comparing line plots from phantom (blue), iterative PYRO-NN reconstruction (green) and reference FBP reconstruction (red). The gray-level window is $[0,1.0]$.

IV.. DISCUSSION

Recently, there have been several different attempts to transfer the astonishing capability of deep learning into the field of medical image reconstruction. The approaches span from purely CNN based image-to-image translation task as pre- or post processing steps over end-to-end networks learning the complete signal reconstruction pipeline in a data-driven manner up to end-to-end networks which are

enriched with known operators to preserve interpretability and reliability of the signals. The different downsides of these approaches were discussed in the community and range from adversarial attacks and signal falsification^{9,36} over the need of huge amounts of data up to highly academic implementations not applicable on real scenarios due to huge memory requirements. In order to transfer deep learning towards medical image reconstruction and at the same time address these problems, the idea of embedding known operators into the neu-

ral networks is increasingly pursued as the growing number of publications shows.

In this paper, we present a framework providing the known operators for CT reconstruction and all necessary tools to conduct experiments on real scenarios. We believe that such an open-source framework will reduce the barriers of such approaches and will elevate the research in the medical image reconstruction domain. To encourage the research we provide the shown baseline experiments as example code within the framework, allow an starting point for own research ideas. To the best of our knowledge there exist no framework that allows simple use of CT reconstruction algorithms within neural networks and can handle and reconstruct data from real systems.

In order to show the capability of the provided framework, we presented a state-of-the-art FDK reconstruction using projection matrices from a real C-arm Cone-beam CT system and deliver this as one baseline example within the framework. This can directly be used to reproduce the results of Würfl et al.²⁰. Furthermore, we provide a reference implementation to learn CT reconstruction filters following Syben et al.¹⁹. We want to emphasize that the learning process is based on purely numerical data without loss of generality. Both experiments show the capability to compute the error in the reconstruction domain while the weight updates can be applied for operators in the projection domain. This gives the possibility to replace heuristically parameterized operators with learned data-optimal operators. With the last experiment we show the potential of the framework to elevate research, as we present a novel way to design a network with just a few lines of code that mimics a iterative reconstruction and, intrinsically, is a starting point to learn regularizers based on the capability of deep neural networks³⁷. We choose to implement the projector and back-projector as a unmatched projector pair. The implications of unmatched pairs are already analyzed in the context of iterative reconstruction by Zeng et al.³⁸. Zeng et al. concluded that unmatched-pairs can be beneficial due to the algorithmic speedup, while the convergence of the algorithm has to be kept in mind. While we have not noticed negative effects on the training process in our experiments^{19,21}, we want to investigate the implications of unmatched-projector pairs to the training procedure in future work.

As the combination of deep neural networks and CT reconstruction can, especially in the 3D case, easily exceeds the GPU's memory, the provided algorithms allow the user a trade-off choice between computational- and memory efficient implementations. Furthermore, the concept of the framework

enables a problem-specific solution, since the algorithms and the gradients can be changed by the user at any time. Additionally, as the core of PYRO-NN is an extension of the existing Tensorflow build process, every known operator which allows the calculation of sub-gradients can easily be modeled as a Tensorflow layer. Besides the actual CUDA implementation, there is only the need of an information flow control class following the Tensorflow API guidelines. Therefore, the setup allows an easy extension towards other frameworks like PyTorch as only the information flow control class has to be adjusted, while the core implementation of the known operator stays untouched.

We provide the known operators for CT reconstruction on CUDA level with the respective necessary tools like filters and weights on Python level. Nevertheless, the framework design allows an easy extension to other fields, e.g., magnetic resonance imaging (MRI) and many more. With the increasing amount of publications being supplemented by open-source reference implementations, we believe that with help of the community PYRO-NN can grow beyond the application on CT reconstruction.

V.. CONCLUSION AND OUTLOOK

PYRO-NN is an open-source software framework developed to elevate the use of known operators within neural networks to transfer the power of deep learning to medical image reconstruction. The framework provides state-of-the-art CT reconstruction algorithms within the Tensorflow deep learning environment, supported by the necessary tools for the reconstruction pipeline.

In this paper, we show the concept of embedding known operators by means of CT reconstruction and underline with experiments comparable to existing algorithms and the applicability of these algorithms in the deep learning framework Tensorflow. The generic design of the framework makes it very easy to extend it to other modalities. We hope that our open-source framework will encourage other groups to join these efforts making the framework a valuable element in the deep learning medical image reconstruction field.

The main objective of the framework is to enable the community to use CT reconstruction algorithms in end-to-end neural networks and to elevate the research in medical image reconstruction.

The software package is available under <https://github.com/csyben/PYRO-NN> and <https://github.com/csyben/PYRO-NN-Layers>.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research

and innovation program (ERC Grant No. 810316). Additional financial support for this project was granted by the Emerging Fields Initiative (EFI) of the Friedrich-Alexander University Erlangen-Nürnberg (FAU).

*Email: christopher.syben@fau.de

- ¹Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *nature* **521**, 436 (2015).
- ²A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, pages 1097–1105, Curran Associates, Inc., 2012.
- ³A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, Wavenet: A generative model for raw audio, *CoRR abs/1609.03499* (2016).
- ⁴A. Maier, C. Syben, T. Lasser, and C. Riess, A Gentle Introduction to Deep Learning in Medical Image Processing, *arXiv preprint arXiv:1810.05401* (2018).
- ⁵O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, Springer, 2015.
- ⁶B. Stimpel, C. Syben, T. Würfl, K. Mentl, A. Dörfler, and A. Maier, MR to X-ray projection image synthesis, *arXiv preprint arXiv:1710.07498* (2017).
- ⁷K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE Transactions on Image Processing* **26**, 4509–4522 (2017).
- ⁸A. Kofler, M. Haltmeier, C. Kolbitsch, M. Kachelrieß, and M. Dewey, A U-Nets Cascade for Sparse View Computed Tomography, in *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 91–99, Springer, 2018.
- ⁹V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, On instabilities of deep learning in image reconstruction-Does AI come at a cost?, *arXiv preprint arXiv:1902.05300* (2019).
- ¹⁰H. Greenspan, B. Van Ginneken, and R. M. Summers, Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique, *IEEE Transactions on Medical Imaging* **35**, 1153–1159 (2016).
- ¹¹G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, Image reconstruction is a new frontier of machine learning, *IEEE transactions on medical imaging* **37**, 1289–1296 (2018).
- ¹²B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, Image reconstruction by domain-transform manifold learning, *Nature* **555**, 487 (2018).
- ¹³A. Maier, F. Schebesch, C. Syben, T. Würfl, S. Steidl, J.-H. Choi, and R. Fahrig, Precision learning: Towards use of known operators in neural networks, in *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 183–188, IEEE, 2018.
- ¹⁴J. C. Ye, Y. Han, and E. Cha, Deep convolutional framelets: A general deep learning framework for inverse problems, *SIAM Journal on Imaging Sciences* **11**, 991–1048 (2018).
- ¹⁵H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, LEARN: Learned experts assessment-based reconstruction network for sparse-data CT, *IEEE transactions on medical imaging* (2018).
- ¹⁶J. Adler and O. Öktem, Learned primal-dual reconstruction, *IEEE transactions on medical imaging* **37**, 1322–1332 (2018).
- ¹⁷T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier, Deep learning computed tomography, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 432–440, Springer, 2016.
- ¹⁸K. Hammernik, T. Würfl, T. Pock, and A. Maier, A deep learning architecture for limited-angle computed tomography reconstruction, in *Bildverarbeitung für die Medizin 2017*, pages 92–97, Springer, 2017.
- ¹⁹C. Syben, B. Stimpel, K. Breininger, T. Würfl, R. Fahrig, A. Dörfler, and A. Maier, Precision learning: Reconstruction filter kernel discretization, in *Proceedings of the Fifth International Conference on Image Formation in X-Ray Computed Tomography*, pages 386–390, 2018.
- ²⁰T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems, *IEEE transactions on medical imaging* **37**, 1454–1463 (2018).
- ²¹C. Syben, B. Stimpel, J. Lommen, T. Würfl, A. Dörfler, and A. Maier, Deriving Neural Network Architectures using Precision Learning: Parallel-to-fan beam Conversion, in *Pattern Recognition, 40th German Conference*, edited by A. Bruhn, 2018.
- ²²J. Adler, H. Kohr, and O. Öktem, Operator discretization library (ODL), Software available from <https://github.com/odlgroup/odl> (2017).
- ²³M. Abadi et al., Tensorflow: a system for large-scale machine learning., in *OSDI*, volume 16, pages 265–283, 2016.
- ²⁴W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, Fast and flexible X-ray tomography using the ASTRA toolbox, *Optics express*

- 24**, 25129–25147 (2016).
- ²⁵D. Gürsoy, F. De Carlo, X. Xiao, and C. Jacobsen, TomoPy: a framework for the analysis of synchrotron tomographic data, *Journal of synchrotron radiation* **21**, 1188–1193 (2014).
- ²⁶A. Maier et al., CONRADA software framework for cone-beam imaging in radiology, *Medical physics* **40** (2013).
- ²⁷C. Syben, S. Seitz, and A. Maier, PyConrad, Software available from <https://git5.cs.fau.de/PyConrad/pyCONRAD> (2017).
- ²⁸R. R. Galigekere, K. Wiesent, and D. W. Holdsworth, Cone-beam reprojection using projection-matrices, *IEEE transactions on medical imaging* **22**, 1202–1214 (2003).
- ²⁹H. Scherl, B. Keck, M. Kowarschik, and J. Hornegger, Fast GPU-based CT reconstruction using the common unified device architecture (CUDA), in *Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE*, volume 6, pages 4464–4466, IEEE, 2007.
- ³⁰L. A. Shepp and B. F. Logan, The Fourier reconstruction of a head section, *IEEE Transactions on nuclear science* **21**, 21–43 (1974).
- ³¹A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, IEEE press New York, 1988.
- ³²D. L. Parker, Optimal short scan convolution reconstruction for fan beam CT, *Medical physics* **9**, 254–257 (1982).
- ³³C. Riess, M. Berger, H. Wu, M. Manhart, R. Fahrig, and A. Maier, TV or not TV? That is the question, *Proceedings Fully3D*, 341–344 (2013).
- ³⁴L. A. Feldkamp, L. Davis, and J. W. Kress, Practical cone-beam algorithm, *Josa* **1**, 612–619 (1984).
- ³⁵G. Ramachandran and A. Lakshminarayanan, Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms, *Proceedings of the National Academy of Sciences* **68**, 2236–2240 (1971).
- ³⁶Y. Huang, T. Würfl, K. Breininger, L. Liu, G. Lauritsch, and A. Maier, Some investigations on robustness of deep learning in limited angle tomography, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 145–153, Springer, 2018.
- ³⁷K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, Learning a variational network for reconstruction of accelerated MRI data, *Magnetic resonance in medicine* **79**, 3055–3071 (2018).
- ³⁸G. L. Zeng and G. T. Gullberg, Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, *IEEE transactions on medical imaging* **19**, 548–555 (2000).