Deep learning-based 2D-3D sample pose estimation for X-ray 3DCT

Alice Presenti, Shabab Bazrafkan, Jan Sijbers, Jan De Beenhouwer

imec-Vision Lab, Department of Physics, University of Antwerp, Universiteitsplein 1, Antwerp 2610, Belgium
e-mail: {alice.presenti, shabab.bazrafkan, jan.sijbers, jan.debeenhouwer}@uantwerpen.be

Abstract
3D X-ray Computed Tomography (CT) is increasingly being used for non-destructive inspection of objects. Conventional CT inspection requires many projections, typically spanning 360° to reconstruct a 3D image of the object, which is then segmented and subsequently compared with the reference computer-aided design (CAD) model. Such an inspection flowchart, however, is a time inefficient procedure, not suitable for inline inspection. To overcome this problem, we directly compare the measured projections with simulated ones from the CAD model. To do so, the simulated projections need to be created with the same acquisition geometry as the measured ones. When an object is inserted on a scanning system, its orientation may vary with respect to the default CAD model orientation. For this reason, 2D/3D registration between the CAD model and the measured projections of the real object is necessary. In this paper, we present a deep learning based method to accurately estimate the 3D orientation of an object from one projection image.

Keywords: pose estimation, X-ray CT, ResNet-50, deep learning

1 Introduction

X-ray Computed Tomography (CT) is a non-destructive technique widely used by industries for inspection of manufactured objects. Typically, product properties are specified on a reference computer aided design (CAD) model. A conventional 3D CT inspection workflow involves the reconstruction of a 3D image from a set of 2D projection images acquired at different view angles, segmentation, surface extraction, registration, and finally comparison with the reference CAD model. Such a conventional inspection workflow requires a time consuming acquisition and 3D reconstruction, which hinders real-time and inline industrial applications.
Other approaches directly compare measured radiographs of the objects with simulated ones from the corresponding CAD model. In [1], simulated CCD camera images are created and compared to the acquired ones. Comparison for quality control proposal is performed by using distance measures or feature descriptor based methods. Following this approach, we developed a CAD projector capable of simulating projection images from the CAD model. The CAD projector is efficiently implemented on the GPU and integrated with our flexible, open-source reconstruction software, the ASTRA Toolbox [2]. Regardless of the workflow adopted, it is possible to determine the optimal inspection configuration based on the specific task [3]. From our analysis in [4], it is shown that limiting the acquisition to an optimal set of angles, has an advantage both on the overall time consumption and on the performance of the final quality control. Moreover, by exploiting prior-knowledge of the inspected object (CAD model, material properties), it was shown that optimal projection angles as well as a suitable region of interest for inspection can be determined in advance.
In projection-based inspection, it is crucial that the simulated projections of the CAD model are properly aligned with the measured projection data. However, the 3D object pose is often only approximately known. In our previous work [5], 3D spatial information of the object (orientation and position) was estimated by mathematically joining 2D information from multiple projections, and the system was driven to acquire task specific projections. However, the method is limited to cylindrical objects. In order to generalize our method, we now present a deep learning scheme to estimate the 3D orientation of a non-symmetric object from a single radiograph. Methods based on convolutional neural networks (CNNs) have been largely applied in pose estimation of objects for robotic manipulation, scene understanding and augmented reality [6–9]. For these methods, the objective is often to jointly classify the objects in the scene and estimate their pose in complex, realistic scenarios. State-of-the-art models for such challenging tasks reach average errors in the order of a few degrees [10–12].
Contrary to these methods, we assume a known object to be inserted in the X-ray scanning system, and that, after calibration, the variation of its pose is limited, e.g. by a sample holder. Although these restrictions lead to a simpler task than those mentioned above, for inspection purposes higher accuracy is required. A comparable approach to ours is presented in [13], where synthetic 2D images are created from a CAD model to train a CNN for camera pose estimation. To make their images look more realistic, scene parameters lightening and material reflectance are modified for each point of view. The background is also allowed to vary in color or by applying a texture. In their experiments, simulated data is generated with evenly distributed points on a sphere. Their experiments on simulated test data show an average accuracy of 5°. In our paper, 3D orientation of an object is retrieved from a 2D X-ray projection image by using a modified ResNet-50 [14], pretrained on ImageNet dataset. To evaluate our method, preliminary experiments on synthetic data are performed.

2 Methods

2.1 The geometry of the system

In what follows, the object is assumed to be stationary and the source and detector to rotate around it. We also assume that the source and detector are aligned so that the optical axis is perpendicular to the detector plane. The method can however easily be adapted to other acquisition geometries.

Let $\mathscr{S} = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ be a reference system, with the $\mathbf{y}$ and $\mathbf{z}$ axis parallel to the detector plane, and $\mathbf{x}$ in the source-detector direction, with respect to the initial source and detector center positions. The center $\mathbf{O} = \{0,0,0\}$ of $\mathscr{S}$ coincides with the rotation center of the system and with the ideal barycenter of the object (see Fig. 1). When the sample is placed in between the source and the detector, the position and orientation generally slightly deviates from the default position of the CAD model in the reference system. The orientation of the object is defined by the rotation angles φ, δ and γ around $\mathbf{x}, \mathbf{z}'$ and $\mathbf{y}''$, respectively, with $\mathbf{z}' = \mathbf{R}_x(\varphi)\mathbf{z}$ and $\mathbf{y}'' = \mathbf{R}_{z'}(\delta)\mathbf{R}_x(\varphi)\mathbf{y}$ (see Fig. 2). The three rotation angles define a new reference system $\mathscr{S}'' = \{\mathbf{x}'', \mathbf{y}'', \mathbf{z}''\}$ centered in $\mathbf{O}$.

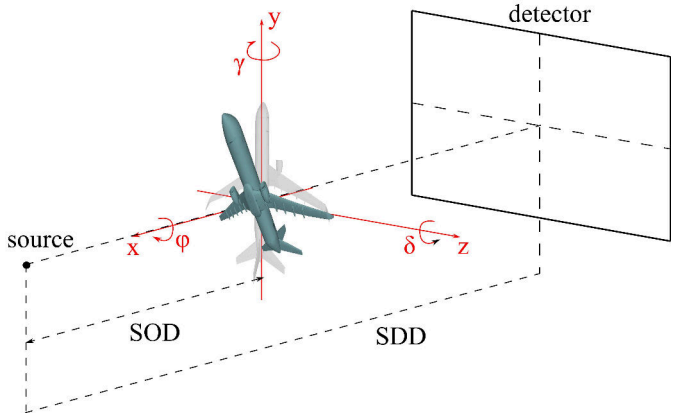

Figure 1: The system's geometry. In gray, an object in the default CAD model orientation, in green, the same object in a random orientation.

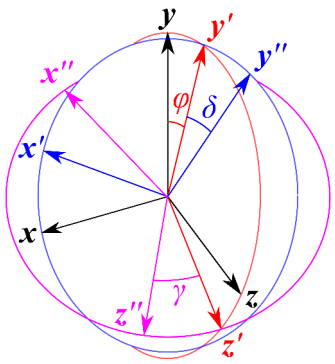

Figure 2: The Euler angles φ, δ and γ.

2.2 Creation of the synthetic dataset

Polychromatic synthetic projections from CAD models are created by using the CAD projector described in [15]. When the object is positioned in the system, small deviations from its rotation angles δ and φ may occur. For the purpose of this study, the rotation rotation angle γ of the sample is assumed to be unknown and within the range $(0,360)^\circ$, while δ and ϕ are considered to vary on a limited interval, $\delta \in (\delta_{min}, \delta_{max})$,$\varphi \in (\varphi_{min}, \varphi_{max})$. The dataset used for training and testing the neural network is composed of projection images simulated with uniformly distributed parameters in the respective interval range.

2.3 The Neural Network structure and training

Large convolutional neural networks are powerful tools, but require big datasets and consequently a long training time. Pre-trained networks allow to partly overcome these problems, by applying features learned on large datasets to different tasks and different labels [16]. To be able to adapt a pre-trained network to a different task from the one it was trained for, the last layer is replaced by a new one (for classification, for example, with a number of neurons corresponding to the number of classes), with randomly initialized weights. The weights of the remaining layers are fixed to those of the pre-trained network, thus, during fine-tuning, only the last layer is trained again [17]. Motivated by the good performances obtained with pre-trained networks, we used a convolutional neural network (ResNet-V2-50) [14, 18], pre-trained on the ImageNet database [19] for image classification, with the aim of estimating the object pose from a single projection image. In order to adapt the network to our specific task, we replaced its final fully connected layer with three pooling layers each of them followed by flatten and activation layers, in order to output three continuous values: the estimates of the parameters φ, δ and γ (see Fig.3). Since we are transferring learning from a classification network to a regression one, we propose to first run few epochs by fixing the weights of all the layers except from the new ones, and then continue training the whole network.

3 Experiments and discussion

In our experiments, we created labeled synthetic projections of a CAD model of the body of a car lamp by randomly varying $\delta, \varphi \in (-10,10)^\circ$,$\gamma \in (0,360)^\circ$. The dataset is composed of 2.88×10^5 200×200 noiseless synthetic projections, with pixel size 0.2 mm. In Fig. 4, views of the CAD model and the synthetic projections are shown. The data was subdivided in 90% training and 10% validation, using the mean absolute error as the loss function. The last layer of the ResNet-V2-50 pre-trained on the ImageNet database, was modified to output $\hat{\gamma}, \hat{\delta}$ and $\hat{\varphi}$, and the corresponding weights were randomly initialized. During the first

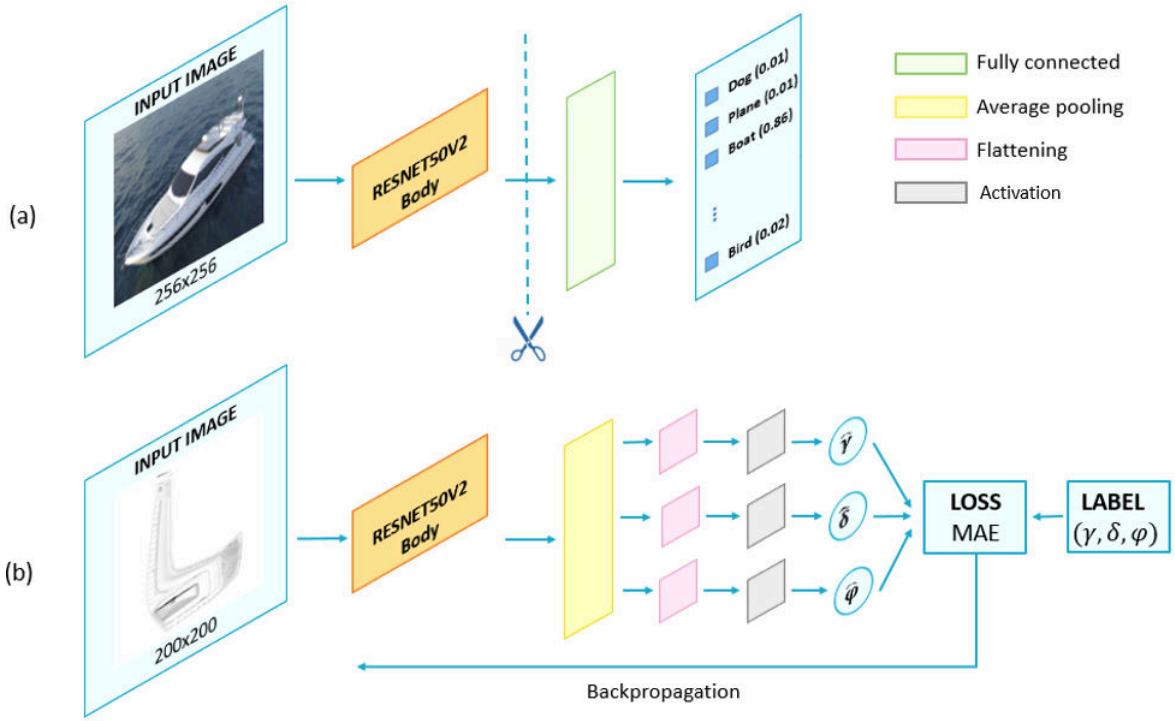

Figure 3: The ResNet-V2-50 network was pre-trained on ImageNet dataset for image classification. The network was cut before the last fully connected layer and adapted for our regression task (b) to output the estimates $\hat{\gamma}, \hat{\delta}$ and $\hat{\varphi}$.

10 epochs only these last new layers were trained, subsequently the whole network was trained for 100 more epochs. To take into account the difference in amplitude among the angles, weights were given to compute the loss ($0.9, 0.05$ and 0.05 for γ, δ and φ, respectively).
For testing, 5000 synthetic projections of the same CAD model were simulated by varying the parameters randomly in the same range as training and validation data. The absolute difference between the labels and the network output is displayed in Fig. 5. The average absolute error on φ,δ and γ is $0.0207°$, $0.0268°$ and $0.1100°$ (SD=$0.0163°$, $0.0204°$ and $0.0687°$), respectively. Figure 8 shows the difference between one of the images in the test dataset and the corresponding image simulated with the parameters estimated by the network. The network was also tested on the same test images with Poisson noise (intensity beam = 15×10^4). Results shown in Figs. 6,7show worse accuracy for the estimation of γ, compared to δ and φ. One reason for it could be that the network, which was trained on noiseless images, learns to recognize a uniform background and thus is not robust to noise. As future work, we plan to improve our results by training the network with synthetic data with different levels of noise.

4 Conclusions

As we already showed in [4], limiting the acquisition to an optimal set of angles, has an advantage both on the overall time consumption and on the performance of the inspection method. A crucial prerequisite for this to work is the quick determination of the orientation of the sample so that the right set of images can be acquired. In this paper, we presented a deep learning scheme to estimate the 3D orientation of an object in relation to a reference CAD model from a single radiograph. By using the pretrained ResNet-V2-50 adapted to our regression task, we showed promising results on synthetic data. Further study will be done to test the network with other models and improve the current results for noisy data. Also, we plan to extend our method to estimate both the 3D orientation and position of a sample from a single projection.

Acknowledgements
This research is funded by the FWO SBO project MetroFlex (S004217N). This work is partially supported by the European Commission through the INTERREG Vlaanderen Nederland program project Smart*Light (0386).

References

[1] E. Hirsch, U. Lübbert, Vision based online inspection of manufactured parts: Comparison of CCD and CAD images, in: L. Faria, W. Van Puymbroeck (Eds.), Computer Integrated Manufacturing, Springer London, 1990, pp. 76–90.
[2] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. Batenburg,

Figure 4: The CAD model of the object used in this study and the respective synthetic projections, created with the CAD projector, at the orientation $(\gamma, \delta, \varphi)$.

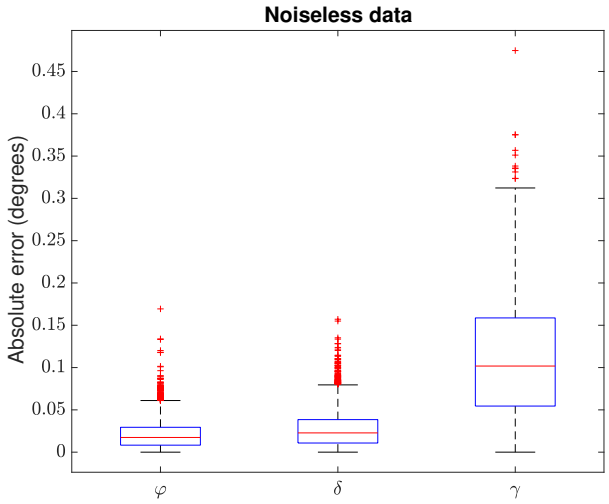

Figure 5: The absolute difference between the ground truth angles and the outputs of the network for 5000 test data.

J. Sijbers, Fast and flexible x-ray tomography using the ASTRA toolbox, Optics Express 24 (2016) 25129–25147.

[3] A. Fischer, T. Lasser, M. Schrapp, J. Stephan, P. B. Noël, Object specific trajectory optimization for industrial x-ray computed tomography, Scientific Reports 6 (2016). doi:10.1038/srep19135.

[4] A. Presenti, J. Sijbers, A. J. den Dekker, J. De Beenhouwer, CAD-based defect inspection with optimal view angle selection based on polychromatic X-ray projection images, in: 9th Conference on Industrial Computed Tomography, 2019. doi:10.1117/12.2534894.

[5] A. Presenti, J. Sijbers, J. De Beenhouwer, Dynamic angle selection for few-view x-ray inspection of CAD based objects, in: 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Fully3D), Vol. 11072, 2019. doi:https://doi.org/10.1117/12.2534894.

[6] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, S. Birchfield, Deep object pose estimation for semantic robotic grasping of household objects, CoRR abs/1809.10790 (2018).
URL http://arxiv.org/abs/1809.10790

[7] A. Collet, S. S. Srinivasa, Efficient multi-view object recognition and full pose estimation, in: 2010 IEEE International Conference on Robotics and Automation, 2010, pp. 2050–2055. doi:10.1109/ROBOT.2010.5509615.

[8] W. Li, Y. Luo, P. Wang, Z. Qin, H. Zhou, H. Qiao, Recent advances on application of deep learning for recovering object pose, in: 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2016, pp. 1273–1280. doi:10.1109/ROBIO.2016.7866501.

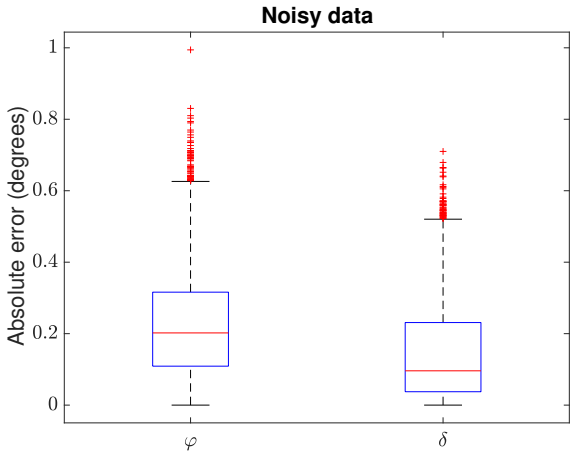

Figure 6: The absolute difference between the ground truth angles φ and δ and the outputs of the network for 5000 test data with simulated Poisson noise (beam intensity=15×10^4).

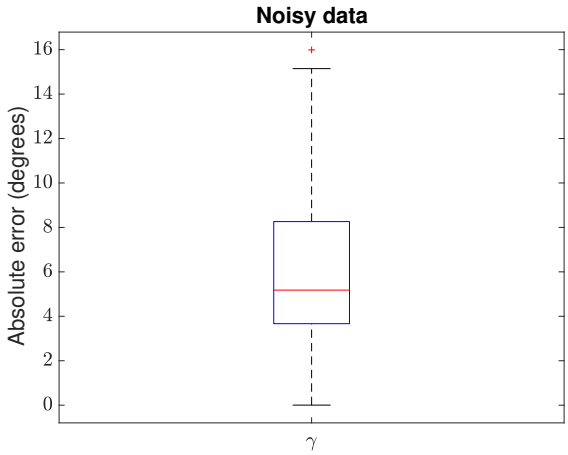

Figure 7: The absolute difference between the ground truth angle γ and the output of the network for 5000 test data with simulated Poisson noise (beam intensity=15×10^4).

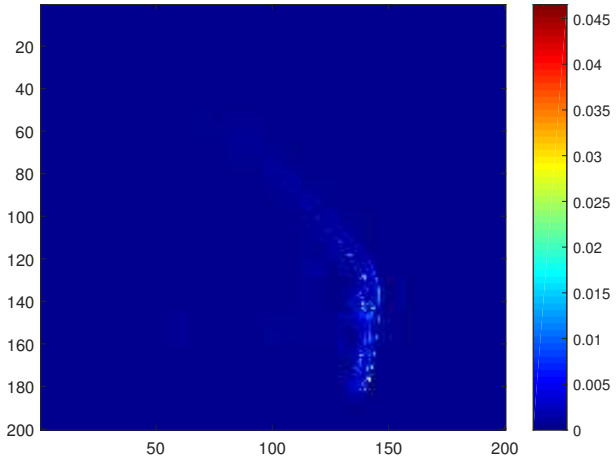

Figure 8: Absolute difference image between one image on the database ($\varphi = 3.4801°, \delta = 5.1796°, \gamma = 299.6289°$) and the image obtained with the values output from the network ($\hat{\varphi} = 3.4393°, \hat{\delta} = 5.1301°, \hat{\gamma} = 299.3931°$).

[9] Y. Xiang, T. Schmidt, V. Narayanan, D. Fox, Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes, CoRR abs/1711.00199 (2017).
URL http://arxiv.org/abs/1711.00199

[10] J. M. Wong, V. Kee, T. Le, S. Wagner, G. Mariottini, A. Schneider, L. Hamilton, R. Chipalkatty, M. Hebert, D. M. S. Johnson, J. Wu, B. Zhou, A. Torralba, Segicp: Integrated deep semantic segmentation and pose estimation, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 5784–5789. doi:10.1109/IROS.2017.8206470.

[11] L. Haochen, Z. Bin, S. Xiaoyong, Z. Yongting, Cnn-based model for pose detection of industrial pcb, in: 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA), 2017, pp. 390–393. doi:10.1109/ICICTA.2017.93.

[12] W. Li, Y. Luo, P. Wang, Z. Qin, H. Zhou, H. Qiao, Recent advances on application of deep learning for recovering object pose, in: 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2016, pp. 1273–1280. doi:10.1109/ROBIO.2016.7866501.

[13] J. Langlois, H. Mouchère, N. Normand, C. Viard-Gaudin, 3D Orientation Estimation of Industrial Parts from 2D Images using Neural Networks, in: International Conference on Pattern Recognition Applications and Methods, Madeira, Portugal, 2018.
URL https://hal.archives-ouvertes.fr/hal-01681124

[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[15] A. Marinovszki, J. De Beenhouwer, J. Sijbers, An efficient cad projector for x-ray projection based 3D inspection with the ASTRA toolbox, in: 8th Conference on Industrial Computed Tomography, 2018.

[16] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, Decaf: A deep convolutional activation feature for generic visual recognition, CoRR abs/1310.1531 (2013). arXiv:1310.1531.
URL http://arxiv.org/abs/1310.1531

[17] V. Campos, B. Jou, X. G. i Nieto, From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction, Image and Vision Computing 65 (2017) 15 – 22, multimodal Sentiment Analysis and Mining in the Wild Image and Vision Computing. doi:https://doi.org/10.1016/j.imavis.2017.01.011.
URL http://www.sciencedirect.com/science/article/pii/S0262885617300355

[18] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, CoRR abs/1611.05431 (2016). arXiv:1611.05431.
URL http://arxiv.org/abs/1611.05431

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database, in: CVPR09, 2009.