CNN-based pose estimation of manufactured objects during inline X-ray inspection

Reference:
Presenti Alice, Liang Zhihua, Alves Pereira Luis Filipe, Sijbers Jan, De Beenhouwer Jan.- CNN-based pose estimation of manufactured objects during inline X-ray inspection
2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), 6-9 September, 2021, Naples, Italy- ISSN 2687-6809 - IEEE, 2021, p. 388-393
Full text (Publisher's DOI): https://doi.org/10.1109/RTSI50628.2021.9597330
To cite this reference: https://hdl.handle.net/10067/1859240151162165141

CNN-based pose estimation from a single X-ray projection for 3D inspection of manufactured objects

Alice Presenti[1], Zhihua Liang[1], Luis Filipe Alves Pereira[1], Jan Sijbers[1], Jan De Beenhouwer[1]

[1]imec-Vision Lab, Department of Physics, University of Antwerp, Belgium,
e-mail: {alice.presenti, zhihua.liang, luisfilipe.alvespereira, jan.sijbers, jan.debeenhouwer}@uantwerpen.be

Abstract
X-ray Computed Tomography (CT) is widely used in inspection of manufactured objects, but typically requires hundreds of radiographs over a wide angular range, making it unsuitable for real-time applications and limited angular view imaging. Fortunately, inspection can also be performed by directly comparing measured radiographs with simulated ones from a reference model. For an effective comparison, an accurate alignment between the radiographs is crucial. In this work, we propose a deep learning-based 3D pose estimation of the object to be inspected from a single radiograph.

Keywords: X-ray inspection, defects detection, 2D/3D pose estimation, CAD models, CNN

1 Introduction

For inspection of manufactured objects, a Computer Aided Design (CAD) reference model of the object could be compared with a triangular mesh extracted from the CT volumetric reconstruction from the X-ray radiographs of the measured sample. This method, however, is not suitable for real-time applications, as hundreds to thousands of radiographs are usually required for a high resolution reconstruction. Also, in case a 360 degrees rotation around the object is not possible, the limited angular distribution of the radiographs may introduce strong artifacts in the reconstructed volume, that could hinder the defect inspection. However, for some applications, a full CT reconstruction and volumetric comparison can be avoided. Indeed, it may suffice to acquire only few radiographs to be compared to simulated ones from the CAD model. For this purpose, it is necessary to first accurately align the 2D projections. Few works have been recently published on Convolutional Neural Network (CNN) - based 3D pose estimation from X-ray radiographs [1, 2]. In this work, we employ a pre-trained ResNet to estimate the pose of an object from a single radiograph. After alignment, the measured radiograph is compared to a simulated one to investigate the presence of defects. Preliminary results on simulated data are presented to detect spherical flaws.

2 Method

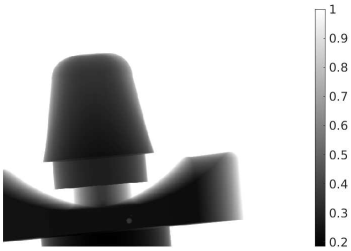
For an effective comparison between measured and simulated radiographs, an accurate alignment method is crucial. We estimate the parameters defining the pose of the object (three Euler angles and the translations along the principal axes) from one radiograph by using a ResNet-50-V2 network [3] pre-trained on the ImageNet dataset [4]. To regress the 3D pose, the last layer of the Resnet-50-V2 body is followed, for every output, by an average pooling layer with pooling size 3, a flattening layer and, finally, a fully connected layer to the output. This last layer needs an activation function to map an extracted tensor to one or multiple continuous values, generally in $[0,1]$ or $[-1,1]$. However, in case of an angle in $[0,360°]$ this mapping would introduce a high penalty between angles close to the opposite interval boundaries. To overcome this problem, we decided to output the sine and cosine of the angle, that are in $[-1,1]$ by definition. To reduce the size of the dataset, we supposed the object to be placed inside an holder, thus introducing constraints to its pose. The only parameter we considered unconstrained was the rotation γ around the vertical axes. The network has 7 outputs (the sine and cosine of γ, the other two rotations and the three translations) and, as activation functions, we used tangent hyperbolic for the sine and cosine of γ and sigmoid for the other parameters, that are mapped in $[0,1]$. The weighted mean absolute error (MAE) between labels and predictions is used as a loss function, with weights $w = \{0.35, 0.35, 0.05, 0.05, 0.1, 0.05, 0.05\}$ for the sine and cosine of $\gamma, \delta, \varphi, t_x, t_y$ and t_z, respectively. The sine and cosine of γ and the translation t_x are attributed a higher weight being out of plane parameters (x is the axis perpendicular to the detector) and thus more difficult to regress.

3 Experiments

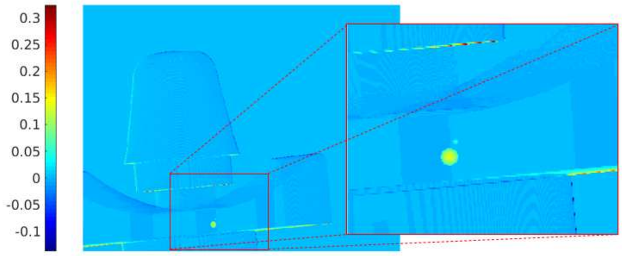
We created simulated images from a CAD model of a snap rivet for printed circuit board support by using a polychromatic CAD projector integrated with the ASTRA toolbox [5, 6]. For fine-tuning the pre-trained ResNet, a dataset of 288×10^3 labeled images was created by randomly varying the object pose, and split in 80% training and 20% validation. A real acquisition was simulated by setting the SOD=108.33mm, SDD=650.00mm and detector pixel size of 0.15mm. To use high resolution images, only a region of interest of 520×670 pixels was used as input for the network (see Fig. 1a). For each image, the CAD model was randomly translated by $t_x, t_y, t_z \in [-4,4]mm$, and rotated by $\gamma \in [0,360)°$ around the vertical axis and $\delta, \varphi \in [-7,7]°$ around the other two axes. 5×10^3 test images of the same object were simulated by varying the pose parameters in the same range as

		γ (deg)	δ (deg)	φ (deg)	t_x (μm)	t_y (μm)	t_z (μm)
Train and test without defects	**Mean**	0.05	0.02	0.03	13	8	11
	STD	0.22	0.02	0.02	14	9	15
Test with defects	**Mean**	1.07	0.15	0.11	175	51	44
	STD	1.59	0.13	0.10	220	41	40
Train and test with defects	**Mean**	0.11	0.04	0.05	60	20	20
	STD	0.29	0.04	0.05	76	22	22

Table 1: The mean and standard deviation of the absolute error between the ground truth labels and network's predictions in case of: 1) Network trained and tested on non-defective objects, 2) Network trained on non-defective objects and tested on defective ones, 3) Network trained and tested on defective objects.



(a)



(b)

Figure 1: (a) A test image with two flaws (two spheres with diameter of $0.1mm$ and $0.35mm$) and (b) Difference image with a simulated radiograph with the CAD model's pose as estimated from the network after re-training.

for the training. Another test was performed on 5×10^3 images where flaws were simulated and displaced randomly as spheres of various diametes (500 of $0.2mm$, 500 of $0.3mm$, 500 of $0.1mm$, 500 of $0.35mm$, 2000 with two spheres of $0.2mm$ and $0.3mm$ and 1000 with two spheres of $0.1mm$ and $0.35mm$). Results are summarized in Table 1. From the comparison between these experiments, it is evident the the network is not robust to such small deviations. For this reason, we re-trained the network with 4×10^4 images with flaws simulated as spheres of various diameters (10^4 of $0.2mm$, 10^4 of $0.3mm$ and 2×10^4 with two spheres with diameters of $0.2mm$ and $0.3mm$). Results for the same test dataset introduced before are shown in Table 1. These findings demonstrate that deviations in the CAD model have to be taken into account in the training process. In Figure 1b, a difference image between a test image with two flaws and a simulated one (without defects) with the pose parameters as estimated with the proposed network. The spheres in this image are of unseen diameter from the network during training.

4 Conclusion

In this work, we present a CNN to estimate the 3D pose from a single 2D radiograph for inspection of manufactured objects. From experiments on simulated data, we demonstrate that defects must be included in the training data to make the network robust to deviations. In our experiments on simulated data, we demonstrate that non only an accurate 3D pose estimation and successive alignment is possible from only one image, but also that, after alignment, inspection becomes feasible from the comparison with images simulated from the reference CAD model.

Acknowledgements

This research is funded by the FWO SBO project MetroFlex (S004217N), the FWO SBO project (S003421N) and the European Commission through the INTERREG Vlaanderen Nederland program project Smart*Light (0386).

References

[1] S. Miao e al., A Convolutional Neural Network Approach for 2D/3D Medical Image Registration, in: CoRR, abs/1507.07505, 2015.

[2] Kügle et al. i3PosNet: instrument pose estimation from X-ray in temporal bone surgery. Int J CARS 15, 1137–1145, 2020.

[3] K. He e al., Identity mappings in deep residual networks, in: Computer Vision – ECCV 2016, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, Eds., Cham, 2016, pp. 630–645, Springer International Publishing.

[4] J. Deng et al., ImageNet: A large-scale hierarchical image database, in: CVPR09, 2009.

[5] W. van Aarle et al., Fast and flexible X-ray tomography using the ASTRA toolbox, in: Optics Express, vol. 24, pp. 25129–25147, 2016.

[6] A. Marinovszki et al., An efficient CAD projector for X-ray projection based 3D inspection with the ASTRA toolbox, in: 8th Conference on Industrial Computed Tomography, 2018, Wels, Austria.