



CycleGAN for interpretable online EMT compensation

Henry Krumb¹ · Dhritimaan Das² · Romol Chadda¹ · Anirban Mukhopadhyay¹

Received: 20 January 2021 / Accepted: 15 February 2021 / Published online: 14 March 2021
© The Author(s) 2021

Abstract

Purpose Electromagnetic tracking (EMT) can partially replace X-ray guidance in minimally invasive procedures, reducing radiation in the OR. However, in this hybrid setting, EMT is disturbed by metallic distortion caused by the X-ray device. We plan to make hybrid navigation clinical reality to reduce radiation exposure for patients and surgeons, by compensating EMT error.

Methods Our online compensation strategy exploits cycle-consistent generative adversarial neural networks (CycleGAN). Positions are translated from various bedside environments to their bench equivalents, by adjusting their z-component. Domain-translated points are fine-tuned on the x–y plane to reduce error in the bench domain. We evaluate our compensation approach in a phantom experiment.

Results Since the domain-translation approach maps distorted points to their laboratory equivalents, predictions are consistent among different C-arm environments. Error is successfully reduced in all evaluation environments. Our qualitative phantom experiment demonstrates that our approach generalizes well to an unseen C-arm environment.

Conclusion Adversarial, cycle-consistent training is an explicable, consistent and thus interpretable approach for online error compensation. Qualitative assessment of EMT error compensation gives a glimpse to the potential of our method for rotational error compensation.

Keywords Electromagnetic tracking · Hybrid navigation · Generative adversarial networks · Adversarial domain adaptation

Introduction

In minimally invasive surgery, electromagnetic tracking (EMT) has the potential to partially replace continuous X-ray navigation [9], reducing the radiation exposure to both patients and surgeons. Such procedures are traditionally performed under X-ray only (for example laparoscopy [1], endovascular aneurysm repair (EVAR) [2]). Our vision is to enable hybrid navigation in the clinical setting, where EMT can replace X-ray as the primary continuous tracker, and X-ray snapshots are only acquired intermittently. However, in current practice EMT is susceptible to metallic distortion caused by the C-arm, such that the surgeon can put little trust in EMT in between snapshots.

Traditional error-compensating algorithms to increase trust in EMT are offline in nature, resulting in tedious cali-

bration and impractical clinical translation. We thus advocate learning-based online error compensation where a general purpose learning model is trained *only once*. Online compensation of EMT error can be realized by implementing data-driven models, which generalize among data from different environments. In previous works, we investigated learning-based online EMT compensation by employing a series of increasingly complex learning models (polynomial regression [10], artificial neural networks (ANNs) [9]).

Despite giving increasingly better results, the interpretation of the failure modes for these complex models is a growing concern in learning literature [16]. Our major focus is to develop an *interpretable* error compensation technique, which imposes two more constraints beyond mere error reduction: *explicability* and *consistency*. Firstly, predictions need to be *explicable*; that is we need to know *why* a certain point from C-arm domain is mapped to a specific compensated point. Otherwise, compensation results could be arbitrary and still fulfill topological constraints, giving a false sense of reliability.

✉ Henry Krumb
henry.john.krumb@gris.tu-darmstadt.de

¹ Technische Universität Darmstadt, Darmstadt, Germany

² Indian Institute of Technology (IIT-BHU), Varanasi, India

Our second constraint is *consistency* among distorted environments. Consistency is important for *online* error compensation, where training data stem from environments with varying distortion characteristics. If output is inconsistent among input environments, changing distortion characteristics of the electromagnetic field (e.g., by moving the C-arm in the course of a surgical procedure) affect the coordinate frame of compensated points. Such changes jeopardize the X-ray-to-EMT registration, rendering our envisioned hybrid setting infeasible.

To impose these two constraints, we approach the problem as domain adaptation with cyclic-consistent generative adversarial networks (GANs) [4,17]. The main purpose of our approach is to learn a mapping from points of domain C , which is the domain of C-arm-distorted points from environments similar to bedside, to domain L , which is the bench domain with points collected in a distortion-free laboratory setting. Our approach is both explicable and consistent by design and thus more interpretable than our ANN [9] approach. While domain adaptation using GANs is quite popular in the medical imaging setting [5], it has never been studied before in the context of EMT error compensation.

Related work

Offline compensation methods, such as interpolation schemes or polynomial fits, are extensively described in the comprehensive reviews by Kindratenko et al. [6] and Franz et al. [3]. For the sake of brevity, we refer the reader to these papers for a detailed comparison of offline methods.

An *online* compensation framework is proposed by Sadjadi et al. in 2016 [14]. Their approach combines Simultaneous Localization and Mapping (SLAM) and an extended Kalman filter to rectify EMT measurements by 67%. Although this approach is promising, it presumes a multi-sensor setup, which is hardly feasible in minimally invasive applications with narrow access canals.

Neural networks for error compensation are employed by Kindratenko et al. in 2005 [7]. The authors use two-layer feedforward networks to compensate EMT error in an offline setup. In our previous work [9], we show that ANNs not only work for offline compensation, but are capable of generalizing among different distortion scenarios (online) as well. The ANNs learn a nonlinear transform which reduces distance error between measured points. Effectively, this method adjusts the topology of input points to minimize their distance loss. As illustrated in Fig. 6, correcting only the topology can lead to *inconsistent* behavior.

In this paper, compensation is performed by translating individual points from C-arm to laboratory domain using generative adversarial neural networks (GANs). This adversarial domain adaptation approach is based on the CycleGAN

[17] architecture, which is originally developed for the application of image-to-image translation. While the topology-optimizing ANN fails silently (Fig. 3), predictions made by our proposed online approach are interpretable by design.

Materials

Our data-driven compensation approach uses measurements we acquired during our previous work [9] and was collected using an Ascension 3D Guidance trakSTAR system (stated accuracy of 1.4 mm), a 180-type sensor (6 DOF, 1.8 mm in diameter) and mid-range field generator. Custom C++ software is used for data collection with the trakSTAR system. We use a calibrated Lego board to collect data in three positional degrees of freedom (DoF), as proposed by us earlier [10]. The calibrated phantom, placed in three different elevations, allows us to measure varying positions on x-, y- and z-axes of the EMT system. For each measuring point, 500 samples are collected and averaged to reduce random noise. Displacement distances between points on the Lego board are then used as ground truth to calculate displacement error. This metric, based on relative displacements, eliminates the need for an additional measurement standard (e.g., a ruler). Positional error of displacements we use for training, validation and evaluation is listed in Table 1.

Similarly to various AC EMT systems, the DC-based trakSTAR provides a quality estimate with each measuring point, indicating the amount of metallic distortion. Each measuring point is constituted by $(x, y, z, q, \phi_x, \phi_y, \phi_z)$, where q denotes the system's quality estimate and ϕ_x , ϕ_y and ϕ_z denote rotation around x, y and z axes.

In our phantom study (Section “Bedside evaluation on aortic phantom”), we place the sensor at different positions inside an acrylic glass phantom, manufactured at Universitätsklinikum Würzburg, Germany. The phantom resembles a human aorta in life size. A 3D-printed Lego adapter holds the sensor cable (Fig. 1) in place. Measurements for the phantom study and other bedside measurements are taken in vicinity of a Ziehm Vision RFD C-arm device. In the bench setting, measurements are taken in an office room without significant sources of metallic distortion.

Our learning models are implemented in Python using the PyTorch [13] framework.

Methods

To fulfill the explicability constraints we identified for interpretable online error compensation, we modify a domain adaptation approach that is originally used for image translation tasks. Instead of translating between two image domains (e.g., photorealistic vs. abstract), our goal is to translate EMT

Table 1 Datasets collected in varying distances to c-arm and in a laboratory setup. Bold values indicate lowest RMSE and std. dev. among all datasets

	Scenario	#points	Displacement RMSE (mm)	Displacement std. dev. (mm)
Training	Laboratory ¹	60	0.367	0.202
	c-arm 8 cm	60	1.292	1.264
	c-arm 11 cm	60	1.064	0.917
	c-arm ² 50 cm	60	0.639	0.309
Validation	c-arm 10 cm	60	1.101	0.989
	c-arm ³ 30 cm	60	0.743	0.372
Evaluation	c-arm 7 cm	60	1.386	1.389
	c-arm 9 cm	60	1.192	1.139
	c-arm 12 cm	60	1.025	0.833

Number of displacements, RMSE and standard deviation are noted for each dataset. Distances to c-arm are measured from X-ray source to base board center. ¹: only 2 of 3 z-layers used for training, c-arm²: gantry rotated at 60°, c-arm³: gantry rotated at 30°

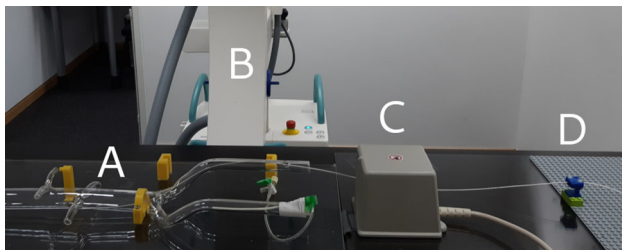


Fig. 1 Hybrid navigation experiment with Aortic phantom. **a** phantom with EMT sensor inside, **b** C-arm, **c** EMT field generator, **d** sensor cable fixture on Lego board

measurements from the bedside (high error, C-arm) to the bench (low error, laboratory) domain. Since the objective of this translation task is intuitive, we expect this compensation to be *explicable*.

In the following, we detail the modified domain adaptation architecture, the protocol and parameters we use for training. Subsequently, we describe how error in laboratory domain is compensated by a post-processing step, which uses a simple linear regression model.

Domain adaptation by adversarial training

We employ cycle-consistent adversarial training for interpretable EMT compensation. Similar to the work of Zhu and Park et al. [17], we make use of two different GAN models, one for each direction of domain translation (C-arm to laboratory, laboratory to C-arm). As illustrated in Fig. 2, the training process connects both GANs to achieve cyclic consistency. Since input data from laboratory and C-arm (Table 1) are unpaired and translation is thus under-constrained, adversarial training benefits from this additional cycle-consistency constraint.

Each of the two GANs consists of a generator network and a discriminator neural network. The generator receives an input point and generates a domain-adapted point, whereas

the discriminator judges whether the generated point stems from the target domain. The generator's objective is to trick the discriminator by generating points close to its target domain, given an input point from the original domain. For instance, G_{CL} takes a point from C-arm environment and tries to generate a corresponding laboratory point, and D_{CL} learns to judge whether this point actually is a *valid* laboratory point. Ideally, both parties in this adversarial two-player game achieve the Nash equilibrium [4,12].

Our two generator models (and discriminator models, respectively) share an identical structure. Generators receive a $(x, y, z, q, \phi_x, \phi_y, \phi_z)$ point (normalized to $[0, 1]$) as input and produce a vector $(\hat{z}, \hat{q}, \hat{\phi}_x, \hat{\phi}_y, \hat{\phi}_z)$, where $\hat{z}, \hat{q}, \hat{\phi}_x, \hat{\phi}_y, \hat{\phi}_z$ denote domain-translated values for z , quality and orientation.

To ensure that points from C-arm domain are not translated to *arbitrary* points in the laboratory domain, the generators do not alter the x and y positional components. We focus only on compensating the z component, since of all positional components, it is the most susceptible to error (for the trakSTAR system). Error in the x - y plane is compensated in a fine-tuning step further described in section “[Fine-tuning](#)”.

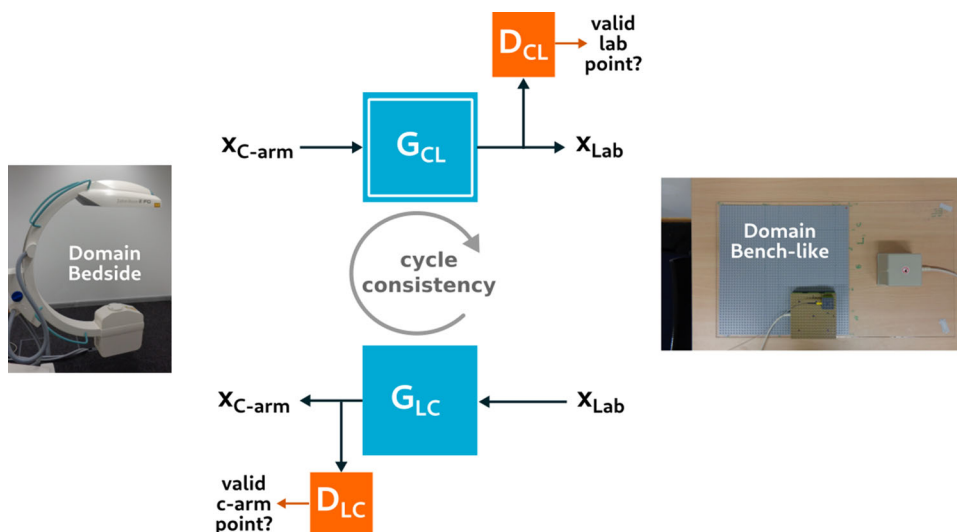
Training protocol

Our training protocol is similar to that of CycleGAN, but includes additional loss terms tailored to the problem of EMT error compensation. In particular, we compute the generator loss as weighted sum of individual penalties, which are described in the following:

Adversarial Loss L_{adv} is a binary cross-entropy (BCE) loss term, which reflects how well each generator can fool its corresponding discriminator. It is computed as

$$L_{adv} = BCE(D_{CL}(G_{CL}(x_C)), l_{valid}) + BCE(D_{LC}(G_{LC}(x_L)), l_{valid})$$

Fig. 2 Cycle-consistent GAN architecture for unsupervised domain adaptation. G_{CL} translates points from C-arm to laboratory domain and, once trained, is used for compensation



where l_{valid} denotes the discriminator label we assign to valid points, that is the label the discriminator assigns to points that are deemed to originate from the target domain. l_{fake} is the label our discriminator assigns to points that do not stem from the target domain.

Cycle loss As described in the original CycleGAN paper [17], we enforce cycle-consistency by adding a loss term L_{cycle} :

$$L_{recov,L} = |G_{CL}(G_{LC}(x_L)) - x_L|$$

$$L_{recov,C} = |G_{LC}(G_{CL}(x_C)) - x_C|$$

$$L_{cycle} = L_{recov,L} + L_{recov,C}$$

$L_{recov,L}$ indicates how well an input point is recovered after translating it from domain L to domain C and back to L ($C \rightarrow L \rightarrow C$ in $L_{recov,C}$, respectively).

Compensation loss In addition to the CycleGAN losses, we also penalize distance error only in the laboratory domain (we cannot enforce error to be low in generated C-arm samples) as a means of regularization. Compensation loss is computed as $L_{comp} = MSE(d_{G_{CL}}, d_{true})$ where $d_{G_{CL}}$ and d_{true} are distances between pairs of points, which stem from G_{CL} or ground truth data, respectively.

The total generator loss is $L_{total} = \lambda_{adv} \cdot L_{adv} + \lambda_{cycle} \cdot L_{cycle} + \lambda_{comp} \cdot L_{comp}$ with coefficients $\lambda_{adv} = 0.5$, $\lambda_{cycle} = 10$ and $\lambda_{comp} = 10^{-5}$, which we determine empirically.

Prediction uncertainty

Although neural networks are known to generalize well to unseen data, predictions made under a lack of knowledge are uncertain. Fortunately, this uncertainty can be approximated by training the same architecture multiple times, but initialized with different random seeds (deep ensembling [11]). Computing the standard deviation among the resulting

predictions yields an approximation of model-inherent (epistemic) uncertainty, and averaging the predictions is expected to yield better prediction accuracy, since we combine the knowledge of multiple models. We choose to sequentially train 10 different initializations in an ensemble as a compromise between training time and accuracy.

Network and training parameters

Our discriminator models have three layers each, with 16 nodes per layer. All layers use LeakyReLU activations with a leak of 0.2, except for the last layer, which is Sigmoid-activated. The discriminators are trained with soft labels (uniform distribution of 0.0..0.2 and 0.8..1.0, respectively) [15].

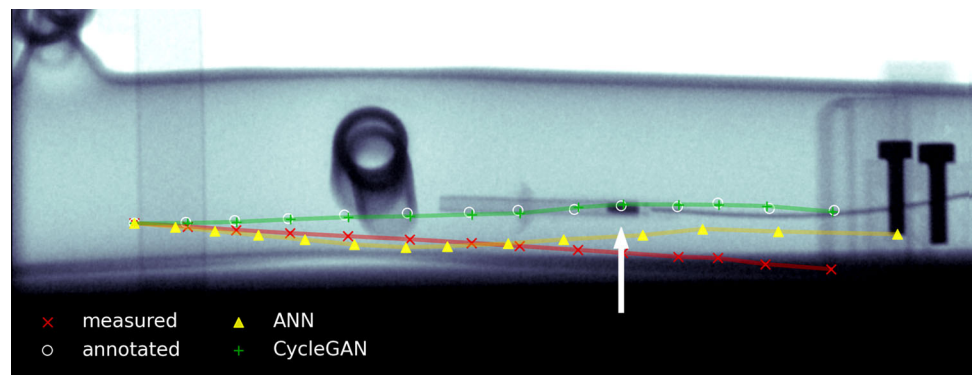
The generators also have four layers each, with 16 nodes per layer. Similar to the discriminators, the generator's layers use LeakyReLU activations, but with a leak of 0.01. The last layer uses linear activation.

Generators and discriminators are optimized under the use of Adam optimizer [8] both with a learning rate of 0.0005, which linearly decays to 0 after 100 epochs. Whereas the generators share a common optimizer, the discriminators are both trained by individual optimizers. During training, we use minibatches with a batch size of 16. The whole training for each model in the ensemble lasts 200 epochs. This training protocol is similar to that of the original CycleGAN implementation [17].

Fine-tuning

As the cycle-consistent GAN model does not affect the x and y components of input points, there still exists positional error on the x - y plane. Assuming that the points compensated by G_{CL} always lie in the laboratory domain, we can apply

Fig. 3 X-ray snapshot of aortic phantom with sensor inside. (Please find an animated video in supplementary material.) Arrow points to EMT sensor tip (dark rectangle). White circles are manually annotated sensor center points. Red, yellow and green overlays show uncompensated, ANN-compensated and domain-adapted EMT points, respectively



a compensation approach tailored to the laboratory domain. For the sake of simplicity, we choose to fit a linear regression model that compensates distance error similarly to our previous work [10], using input features (x, y, z, q) .

Results

We first perform a bedside phantom evaluation of online error compensation. Afterward, we compare our domain adaptation approach to the ANN we have previously proposed [9].

Bedside evaluation on aortic phantom

To assess the quality of our online compensation method in a realistic hybrid bedside setting, we combine EMT and X-ray imaging in a pilot phantom study. Figure 1 shows the measuring setup used in this experiment, in which the EMT sensor is placed inside an aortic phantom which is positioned close to the C-arm. Since the C-arm gantry is rotated to 90° , this setup is different from anything the CycleGAN has seen during training. Using a custom 3D-printed Lego fixture, we pull out the sensor in 13 steps of 8 mm. For each individual step, an X-ray snapshot is created in the median plane, which corresponds to the x - z plane of the EMT coordinate system.

The EMT sensor can be clearly distinguished from its background on the X-ray (Fig. 3). We could therefore annotate the EMT sensor's center points in all 13 snapshots by hand. Points measured by EMT are scaled to pixel dimensions and translated to match the annotated point (leftmost in Fig. 3). Rotation angle of the whole trajectory is estimated from compensated ϕ_y at the first measuring point. We use the same transform for all three sets of points (uncompensated, ANN and CycleGAN) and omit the fine-tuning step, to allow for better comparison between uncompensated and compensated trajectories. Figure 3 shows that our compensated points are close to the annotations, indicating that our domain adaptation model generalizes well to the unseen environment.

Since our compensations alter rotation angles ϕ_x , ϕ_y and ϕ_z as well, the phantom study in the C-arm environment also allows for a pilot qualitative assessment of rotational error compensation. Although we cannot directly measure the actual orientation of the sensor inside the phantom, we can make three assumptions:

1. The sensor tip is heading in positive direction of the tracker's x -axis throughout the whole experiment. We expect azimuth angle to be constant and close to zero.
2. Elevation is almost constant and near zero. Since the aortic phantom is slightly curved, elevation should decrease with higher x .
3. Roll angle is hard to determine absolutely (only relatively). However, we know that it does not change substantially during the experiment, as the cable is not twisted and is rigidly attached to the Lego block.

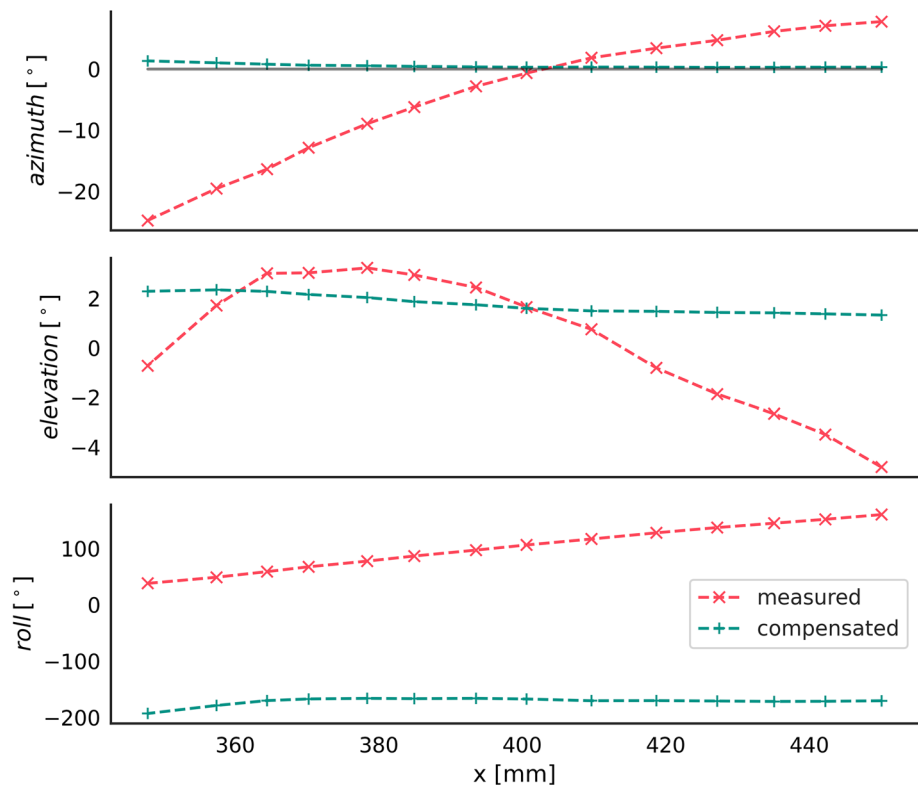
Figure 4 shows orientations over positions on the x -axis, which are measured in our phantom study. It illustrates that all three assumptions hold for compensated values: (1) Azimuth is constant and close to zero degrees (gray line), (2) elevation is almost constant at about 2° , and (3) roll is nearly constant. Contrary to this, raw measurements violate all three assumptions.

Quantitative comparison

In Fig. 5, we see that the domain adaptation approach is more consistent among distorted environments and yields results close to laboratory points. Even C-arm points without a corresponding laboratory point in the training set are matched to their corresponding point in the laboratory domain, indicating that our method generalizes well.

On the other hand, topology-based compensation by our previously proposed ANN [9] does not yield consistent output among input environments. Actually, ANN-compensated points are still close to the input points on the z -axis and several millimeters away from corresponding laboratory points.

Fig. 4 Measured and CycleGAN-compensated rotation angle (azimuth, elevation, roll) over x position, during sensor retraction in aorta phantom study



Ablation experiment

Adversarial training with cycle consistency is beneficial for compensation performance, compared to training a single GAN translating C-arm to laboratory points. Vanilla GAN without cycle loss worsens overall compensation performance, as we show in Table 2.

Our fine-tuning step brings an additional boost in accuracy to the proposed CycleGAN setup (Table 2). However, this step comes with the cost of introducing another source of predictive uncertainty.

Discussion

Comparing our domain adaptation approach to previously proposed topology-based compensation [9], we see that online compensation performance is not only a matter of RMSE, but needs to be assessed with interpretability in mind. Predictions made by the topology-based method are hardly explicable (as seen in Fig. 5). We can only hypothesize that the ANN tries to fulfill topological constraints to minimize distance error, without developing an understanding of what makes a plausible compensated point. Figure 6 illustrates how domain adaptation is explicable and consistent by design, whereas topological compensation is not.

We train a modified CycleGAN on positions from various environments, achieving a translation of distorted (C-arm) to undistorted (laboratory) points. The CycleGAN-based setup is capable of reducing error in different unseen C-arm scenarios, while still producing consistent results. Since we can show that points are mapped to their correct laboratory counterparts—regardless of these being directly present in the training set—our approach is also explicable. Hence, our domain adaptation approach is interpretable by design.

Conclusion and future work

In this paper, we present an approach for online positional error compensation for EMT, which focuses on interpretability. Interpretable online error compensation raises trust in EMT, making it suitable for hybrid navigation, where the surgeon has to rely on EMT until the next X-ray image is taken. With reliable, online-compensated EMT navigation, radiation exposure can be reduced to a minimum, thus bringing less harm to patient and surgeon.

Our cycle-consistent adversarial domain adaptation approach for EMT error compensation is interpretable by design and generalizes to unseen scenarios, as we demonstrate in a prototype hybrid scenario. As our method was originally designed to only compensate positional error, it is surprising to see that it has potential to correct rotational inaccuracies as

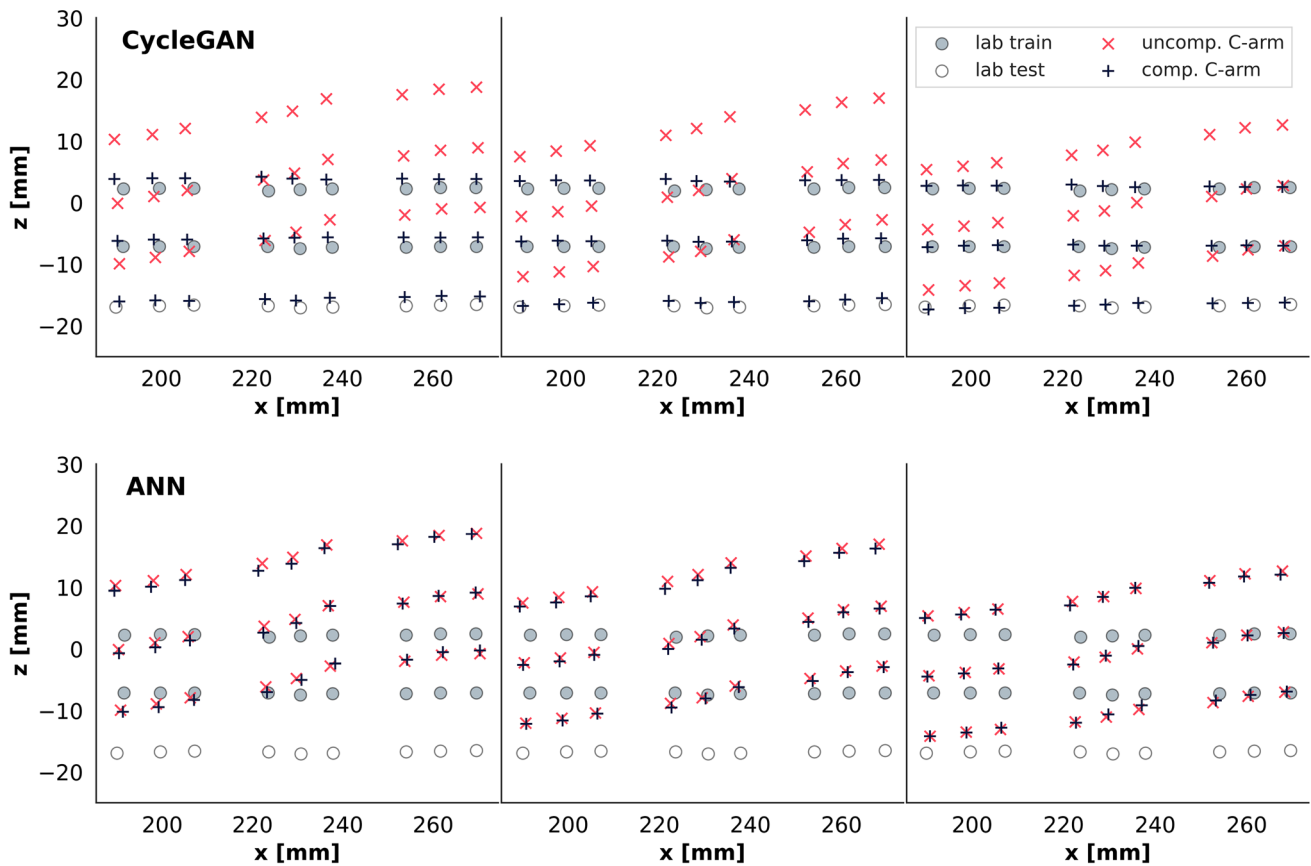


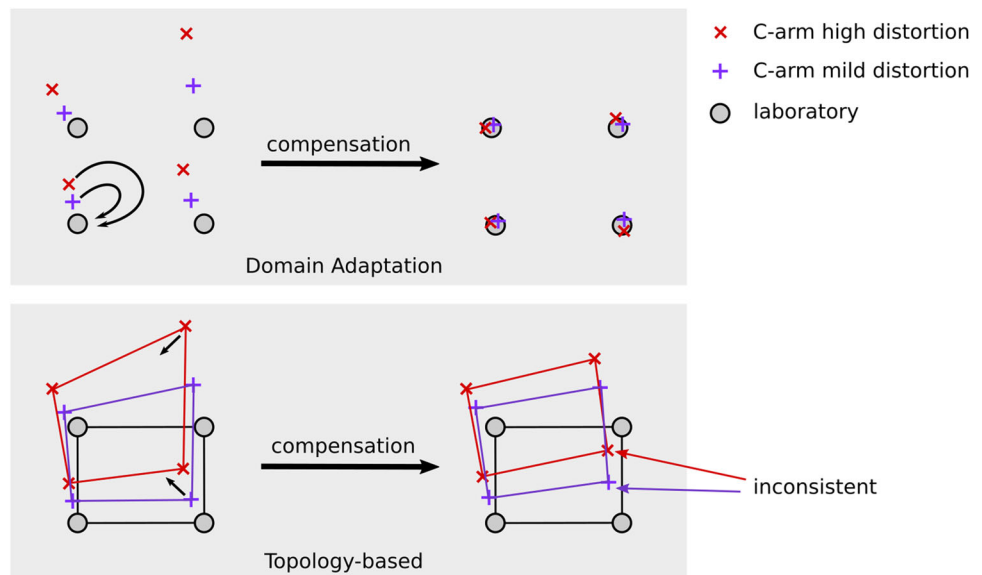
Fig. 5 CycleGAN-adapted (top) and ANN-compensated (bottom) measuring points from C-arm at 7 cm (left), 9 cm (center), 12 cm (right), compared to corresponding laboratory points. Although laboratory points at z close to -18 mm are not present in the training set, all CycleGAN-compensated C-arm points still map to their laboratory counterparts

Table 2 Comparison of tracking error (RMSE & standard deviation) and prediction uncertainty σ_{pred} for different online architectures in evaluation setups from Table 1. Bold values indicate lowest error on individual datasets

Method	Dataset	RMSE (mm) ↓	σ_{error} (mm) ↓	σ_{pred} (mm) ↓
CycleGAN	7 cm	1.295	1.264	
	9 cm	1.090	0.949	0.370
	12 cm	1.007	0.722	
CycleGAN	7 cm	1.100	1.117	
+ Fine	9 cm	0.811	0.759	0.768 ¹
Tuning	12 cm	0.622	0.530	
GAN	7 cm	1.400	2.744	
	9 cm	1.176	1.890	0.654
	12 cm	1.030	1.256	

¹Linear regression only: $\sigma_{pred} = 0.673$ mm

Fig. 6 Schematic comparison of domain adaptation (top) versus topological (bottom) compensation schemes



well. However, specialized measuring standards need to be developed to verify the results on rotational error compensation. Further, we plan to investigate the applicability to other EMT devices, many of which feature a metallic distortion (quality) estimate similar to the trakSTAR's.

In the future, we plan to investigate rotational compensation. Once our approach yields verifiable results on the rotational axes, it is ready for further bedside evaluations incorporating feedback of surgeons who use the hybrid system. Finally, more evaluations on the bedside bring us closer to make hybrid EMT navigation clinical reality.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11548-021-02324-1>.

Acknowledgements We would like to thank Prof. Dr. med. Wolfram Voelker at Universitätsklinikum Würzburg for providing us the aortic phantom, which added significant value to our paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Code availability Code and data are available at <https://github.com/ai-health-care/EMT-CycleGAN> under MIT license.

Conflict of interest This research was partially funded by the German Research Foundation. The authors declare that they have no conflict of interest.

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain patient data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Aoki T, Mansour DA, Koizumi T, Wada Y, Enami Y, Fujimori A, Kusano T, Matsuda K, Nogaki K, Tashiro Y, Hakozaiki T, Shibata H, Tomioka K, Hirai T, Yamazaki T, Saito K, Goto S, Watanabe M, Otsuka K, Murakami M (2020) F. *J Gastrointest Surg* 230(3):1–8
2. Dijkstra ML, Eagleton MJ, Greenberg RK, Mastracci T, Hernandez A (2011) Intraoperative c-arm cone-beam computed tomography in fenestrated/branched aortic endografting. *J Vasc Surg* 53(3):583–590
3. Franz AM, Haidegger T, Birkfellner W, Cleary K, Peters TM, Maier-Hein L (2014) Electromagnetic tracking in medicine—a review of technology, validation, and applications. *IEEE Trans Med Imaging* 33(8):1702–1725
4. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: *NEURIPS*, pp 2672–2680
5. Kazemina S, Baur C, Kuijper A, van Ginneken B, Navab N, Albarqouni S, Mukhopadhyay A (2020) Gans for medical image analysis. *Artif Intell Med* 109:101938
6. Kindratenko VV (2000) A survey of electromagnetic position tracker calibration techniques. *Virtual Reality* 5(3):169–182

7. Kindratenko VV, Sherman WR (2005) Neural network-based calibration of electromagnetic tracking systems. *Virtual Reality* 9(1):70–78
8. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
9. Krumb H, Hofmann S, Kügler D, Ghazy A, Dorweiler B, Bredemann J, Schmitt R, Sakas G, Mukhopadhyay A (2020) Leveraging spatial uncertainty for online error compensation in emt. *IJCARS*, pp 1–9
10. Kügler D, Krumb H, Bredemann J, Stenin I, Kristin J, Klenzner T, Schipper J, Schmitt R, Sakas G, Mukhopadhyay A (2019) High-precision evaluation of electromagnetic tracking. *IJCARS* 14(7):1127–1135
11. Lakshminarayanan B, Pritzel A, Blundell C (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. In: *Advances in neural information processing systems*, pp 6402–6413
12. Nash J (1951) Non-cooperative games. *Ann Math* 54(2):286–295
13. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) Pytorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché Buc F, Fox E, Garnett R (eds) *NEURIPS* 32. Curran Associates, Inc, pp 8024–8035
14. Sadjadi H, Hashtrudi-Zaad K, Fichtinger G (2016) Simultaneous electromagnetic tracking and calibration for dynamic field distortion compensation. *TBME* 63(8):1771–1781
15. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2818–2826
16. Tjoa E, Guan C (2020) A survey on explainable artificial intelligence (xai): towards medical xai. *IEEE Trans Neural Netw Learn Syst*
17. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *ICCV*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.