

# Automated C-Arm Positioning via Conformal Landmark Localization

Ahmad Arrabi<sup>1</sup> Jay Hwasung Jung<sup>1</sup> Jax Luo<sup>2</sup> Nathan Franssen<sup>3</sup> Scott Raymond<sup>2</sup> Safwan Wshah<sup>1</sup>  
 University of Vermont, Department of Computer Science, Burlington VT, USA<sup>1</sup>  
 Cleveland Clinic, Neurological Institute, Cleveland OH, USA<sup>2</sup>  
 University of Vermont Medical Center, Burlington VT, USA<sup>3</sup>

{aarrabi, jjjung2, swshah}@uvm.edu, {luoj2, raymons3}@ccf.org, {nathan.franssen}@uvmhealth.org

## Abstract

*Accurate and reliable C-arm positioning is essential for fluoroscopy-guided interventions. However, clinical workflows rely on manual alignment that increases radiation exposure and procedural delays. In this work, we present a pipeline that autonomously navigates the C-arm to predefined anatomical landmarks utilizing X-ray images. Given an input X-ray image from an arbitrary starting location on the operating table, the model predicts a 3D displacement vector toward each target landmark along the body. To ensure reliable deployment, we capture both aleatoric and epistemic uncertainties in the model’s predictions and further calibrate them using conformal prediction. The derived prediction regions are interpreted as 3D confidence regions around the predicted landmark locations. The training framework combines a probabilistic loss with skeletal pose regularization to encourage anatomically plausible outputs. We validate our approach on a synthetic X-ray dataset generated from DeepDRR. Results show not only strong localization accuracy across multiple architectures but also well-calibrated prediction bounds. These findings highlight the pipeline’s potential as a component in safe and reliable autonomous C-arm systems.*

## 1. Introduction

C-arm machines are used in interventional procedures that require fluoroscopy. These machines allow real-time radiographic projections during operations such as vascular access and orthopedic repairs. Before beginning a procedure, clinicians manually position the C-arm over a region of interest (ROI) on the patient [29]. This manual positioning step is often guided by repeated fluoroscopy exposures, which can introduce delays and unnecessary radiation for both patients and clinicians. Moreover, the need for quick and precise positioning becomes critical in urgent procedures, e.g., stroke and trauma, which may be performed by less experienced personnel in low-resource settings [32]. These challenges highlight the need for assistive

technologies that enable safe, efficient, and reliable fluoroscopy setup and operation.

Automated fluoroscopy positioning can help imaging workflows by utilizing predefined anatomical ROIs that are frequently targeted in procedures. For example, during a stroke thrombectomy, the ROI involves cerebral vessels, such as the internal (ICA) and middle cerebral artery (MCA) [27]. In orthopedic trauma surgeries, the ROI would be around the area of injury, e.g., specific bones or joints [4]. Aligning the C-arm machine with the target ROI can be formalized as a localization task, where the system needs to maneuver the C-arm to a given anatomical landmark [17].

Introducing any system of this kind requires providing both accuracy and a confidence measure. However, deep learning methods often produce deterministic point estimates that lack any measure of uncertainty. To address this gap, we propose a pipeline designed for automatic C-arm positioning. Our method models both the aleatoric uncertainty (inherent variability in the data) and epistemic uncertainty (model’s lack of knowledge) [15, 31]. Aleatoric uncertainty is learned by training the model to output Gaussian-distributed predictions and minimizing the negative log-likelihood loss. Epistemic uncertainty is estimated using Monte Carlo Dropout (MCD), by averaging predictions from multiple stochastic forward passes.

Additionally, we integrate conformal prediction, a distribution-free, post-hoc method that constructs prediction intervals with formal coverage guarantees under the assumption of data exchangeability. In our context, these intervals take the form of 3D spheres around predicted landmarks, defining a region where the true location is expected to lie with a predefined probability ( $1 - \alpha$  in Sec. 4). The radius of each sphere adapts dynamically based on the total predicted uncertainty, enabling the system to express per-sample confidence. This offers a step toward clinically viable, risk-aware systems in fluoroscopy.

We evaluate our method on a synthetic X-ray dataset generated using DeepDRR [34] from CT scans collected from the New Mexico Decedents Image Database (NM-

DID) [7]. Our results demonstrate not only strong localization performance but also accurate and calibrated uncertainty estimates. We also present a stroke thrombectomy use case, where the model navigates to the skull using different paths. The main contributions of this work can be summarized as follows:

- We introduce a general uncertainty-aware pipeline for automated C-arm positioning, specifically to predict anatomical landmarks with quantified confidence.
- Utilizing conformal prediction, our method provides formal statistical guarantees, ensuring that true landmark locations lie within calibrated bounds.
- We validate our approach through extensive experiments, including a practical evaluation on a simulated stroke thrombectomy use case.

## 2. Related Work

Computer vision has been applied in medical imaging for tasks such as organ segmentation, lesion detection, and image registration [9, 22, 37]. While these approaches have shown strong performance in diagnostic tasks, they are often unsuitable for time-sensitive clinical applications due to their limited interpretability and lack of uncertainty understanding [9, 19, 22, 37].

**Image-Guided C-arm Automation:** Image-guided C-arm control has been explored in interventional procedures, where accurate positioning to ROIs is critical. To overcome the limitations of traditional approaches, such as manual or VR- and AR-based control [29, 33], deep learning-based methods have emerged.

[18] proposed a model for automatic C-arm positioning, targeting anatomy-specific views in orthopedic surgery. Their approach uses a two-stage network to predict 5 Degrees of Freedom (DoF) pose updates from a single synthetic DRR image. They focus on localizing the proximal femur (PF) and the fourth lumbar vertebra (LV4), demonstrating robust performance across anatomical variability using synthetic training data. However, its limited anatomical ROI and reliance on multi-stage predictions limit its broader clinical applicability. Another study in [8] introduced C-arm pose estimation from a single view. Deep-DRR [34] was used to produce realistic X-ray images of the pelvis. Their approach achieved competitive results, but the use of a limited dataset raises concerns about the model’s generalizability, not only to other ROIs but also across diverse patient anatomies.

To move beyond region-specific models, [3] proposed a self-supervised approach for C-arm landmark classification and regression. Their method uses a pretext regression task to predict the location of input X-ray images across the entire body and a downstream landmark classification task. However, no navigation or control was implemented. These simplifications raise questions about the method’s reliability

when applied to real-world settings.

**Uncertainty Quantification in Medical Vision:** Artificial intelligence (AI) has become a powerful assistive tool for decision-making across various domains [16, 35]. Despite its efficiency, the reliability of AI predictions remains questionable due to the underlying uncertainty derived from data variability, model miscalibration [1]. As a result, many researchers argue that trust in AI systems should be supported by proper uncertainty quantification [14, 16, 24, 25, 30]. We summarize the most common uncertainty quantification approaches below [1]:

- **Bayesian inference:** These approaches quantify uncertainty by treating the network weights as probability distributions rather than point estimates. They capture epistemic uncertainty by defining approximations over the weight posterior integral. These methods include variational inference or Markov Chain Monte Carlo methods [1, 2, 20].
- **Ensemble techniques:** Ensembles estimate uncertainty by training multiple models independently and combining their predictions. The diversity among the models reflects epistemic uncertainty [1, 20].
- **Monte Carlo dropout:** Monte Carlo dropout estimates uncertainty by applying dropout during both training and testing [11]. Dropout randomly deactivates parts of the network, creating slightly different models each time. By running multiple forward passes with different dropout masks at test time, the method generates varying predictions [1, 20].

In medical imaging tasks such as segmentation and registration, uncertainty quantification provides confidence estimates that can be used as additional information in the decision-making process. This has motivated the integration of uncertainty estimation into deep learning models across various imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), and X-ray [9, 22, 37]. The integration of uncertainty quantification into C-arm automation remains largely unexplored, highlighting the novelty of our approach.

## 3. Dataset

Training data for C-arm positioning is limited due to the practical and ethical constraints of obtaining large-scale fluoroscopic images from real patients. As a result, existing datasets [12, 36] are often focused on specific anatomical regions, e.g., chest, hands, pelvis, making it challenging to train deep learning models that continuously generalize across the entire body. Thus, we constructed a synthetic X-ray dataset covering the head, neck, and upper extremities from computed tomography (CT) data using Digitally Reconstructed Radiographs (DRRs). The collected data is summarized in Tab. 1.

**CT data Acquisition:** To construct a dataset that enables

Table 1. Dataset summary. The dataset was randomly split into 70% for training, 15% for calibration, and 15% for testing, on the patient level.

Property	Value
# CT scans	260
Samples per CT volume	1024 X-ray images
Total samples	266,240 X-ray images
# landmarks (ROIs)	14
Vertical sampling	Uniform
Horizontal sampling	Gaussian ( $\sigma = 47.5\text{mm}$ )
Depth sampling	Gaussian ( $\sigma = 100\text{mm}$ )

learning across the continuous anatomical structure of the human body, we utilize CT scans from the publicly available New Mexico Decedent Image Database (NMDID) [7]. For this study, we selected 260 CT scans to generate synthetic DRR images. To ensure comprehensive anatomical coverage, we use scans covering the head, neck, and upper extremities, captured with 3 mm slice thickness and 3 mm spacing. Each slice was stored in DICOM (.dcm) format, but for processing efficiency, we converted these slices into one unified file with NIfTI (.nii) format.

To ensure consistent spatial representation, all CT scans were normalized by mapping their physical coordinates into the unit cube  $[0, 1]^3$  using each scan’s own dimensions. This transformation preserves the relative spatial relationships between anatomical structures while allowing the model to learn from scale-invariant spatial patterns.

**DRR Image Generation:** A DRR is a synthetic 2D projection of a 3D CT volume. To ensure the realism and clinical applicability of our training data, we use DeepDRR [34] - a deep learning-based framework designed to accurately simulate the C-arm machine. It takes a 3D CT Volume and isocenter (a fixed reference point in 3D space that defines the center of the projection) as inputs, and outputs a synthetic X-ray image.

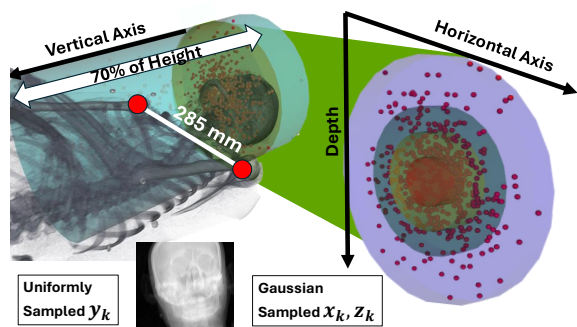


Figure 1. Overview of the sampling process. We densely sample each CT volume by defining independent distributions across each spatial axis.

**Sampling Process:** We systematically generate synthetic X-ray images from multiple viewpoints using DeepDRR. Each view is simulated by positioning the C-arm system around a fixed anatomical isocenter, indicated by the red dot in Fig. 1.

To introduce spatial variability and create a rich training dataset, we densely sampled isocenters within the CT volume to accurately simulate the practical workflow of a typical C-arm. Specifically, the vertical (superior-inferior) position is sampled uniformly within 70% of the anatomical height. The horizontal (left-right) position is sampled with a Gaussian spread centered around the anatomy, with a standard deviation chosen to reflect the average width (285 mm) of the annotated left and right humeral heads. The depth (anterior-posterior) also follows a Gaussian distribution, with a broader spread ( $\sigma = 100\text{ mm}$ ).

These sampling strategies not only capture natural variations in views but also ensure that projections avoid empty regions and focus on the torso where the ROIs are most likely to be found. Each CT scan was sampled 1,024 times, resulting in a total of 266,240 synthetic images.

**Landmark Annotations:** To create ROI annotations, we define 14 landmarks that span the 3D volume and provide substantial anatomical coverage, shown in Fig. 2. This enables us to assess per-landmark uncertainty when guiding C-arm machines from arbitrary initial positions to predefined ROIs. To facilitate landmark annotations from each CT scan, we utilized the annotation tool introduced in [3]. In total, we collected 3,643 landmark coordinates from 260 patient CT scans.

**Dataset split:** We partition the dataset into 70% for train-

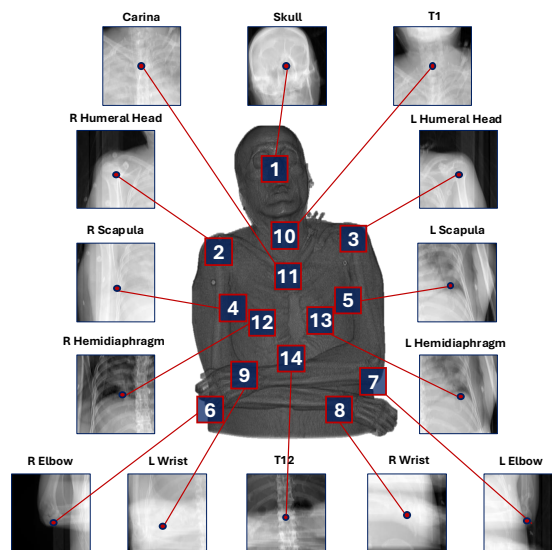


Figure 2. Visualization of landmark annotations. These landmarks represent ROI used across different clinical applications.

ing and 15% for each calibration and testing. This corresponds to 182 CT volumes used in training, and 39 each for calibration and testing. The need for a calibration set comes from our application of split conformal prediction, which requires held-out data to construct prediction intervals with formal statistical coverage guarantees in testing.

## 4. Methodology

We formalize our problem as a regression task aimed at predicting C-arm control. Given an input X-ray image  $I_{Xray} \in \mathbb{R}^{3 \times h \times w}$  (where  $h$  and  $w$  denote height and width, respectively) captured from an arbitrary location  $\mathbf{P} = (x, y, z) \in [0, 1]^3$  within a normalized 3D space. We predict the needed displacements in all positional axes to reach a set of predefined target landmarks  $\{\Delta \hat{\mathbf{P}}_k = (\Delta \hat{x}_k, \Delta \hat{y}_k, \Delta \hat{z}_k) : k \in \{1, \dots, m\}\}$ , where  $m$  is the number of landmarks. These landmarks represent ROIs where operators typically maneuver the C-arm (see Sec. 3). *Note that for readability, we will consider a single landmark in our notation, but our pipeline processes multiple independent landmark displacements.*

Fig. 3 illustrates the proposed pipeline. The input image is first projected into a latent representation  $f \in \mathbb{R}^{c \times h_f \times w_f}$  by a backbone network. This representation is then passed to a positioning encoding module that concatenates it with a positional embedding vector that encodes the C-arm pose  $f_p = \text{cat}(f, \text{linear}(\mathbf{P}))$ . Finally, the regression head (sequence of linear layers) outputs the predicted displacements  $\{\Delta \hat{\mathbf{P}}_k\}$ . Note that in all linear layers, SiLU activation was used, followed by a dropout layer.

### 4.1. Uncertainty Modeling

In deep learning literature, it is widely accepted that the total predictive uncertainty can be approximated as the sum of aleatoric and epistemic uncertainties [15, 31]. These two arise from fundamentally different sources, so following prior work, we assume independence between the two, and define the total uncertainty as follows,

$$\sigma_{total}^2 = \sigma_{epistemic}^2 + \sigma_{aleatoric}^2 \quad (1)$$

To estimate  $\sigma_{aleatoric}^2$ , we model the regression outputs as independent Gaussian distributions, replacing point estimates with predictive distributions. For each landmark, the network predicts the parameters for its mean vector  $\boldsymbol{\mu}_k = (\mu_{\Delta x}, \mu_{\Delta y}, \mu_{\Delta z})$  and its diagonal covariance matrix  $\boldsymbol{\Sigma}_k = \text{diag}(\sigma_{\Delta x}^2, \sigma_{\Delta y}^2, \sigma_{\Delta z}^2)$ , so the predicted displacement is modeled as in Eq. (2). This formulation captures the inherent variance of the data, e.g., anatomical variability and landmark annotation inconsistencies.

$$\Delta \hat{\mathbf{P}}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (2)$$

To have a comprehensive understanding of uncertainty, we account for  $\sigma_{epistemic}^2$ , which reflects the uncertainty

over the model’s parameters. We adopt Monte Carlo Dropout (MCD), a lightweight Bayesian approximation that interprets dropout as a form of variational inference [10]. In MCD, dropout layers are activated in both training and testing, which transforms any deterministic network into a probabilistic one, where each forward pass samples a different subnetwork by randomly deactivating weights with probability  $p$ .

In Fig. 3, dropout layers are added after all linear layers, making our pipeline stochastic. Every forward pass yields a different output due to the random masking of weights. This variability across multiple passes for the same input is used to estimate the variance of the model’s parameters. Therefore, we run  $T$  forward passes for each input. The final prediction is the mean value of those runs, while their variance is  $\sigma_{epistemic}^2$ , which is added with the predicted variance as in Eq. (1).

### 4.2. Conformal Prediction

Conformal prediction is a post-hoc uncertainty quantification framework that produces prediction sets with guaranteed statistical coverage, with no assumptions about the underlying model [28]. We utilize a split conformal prediction method where we use 15% of the dataset to calibrate the nonconformity scores [15]. We use the Euclidean error, defined as the distance between predicted and ground-truth landmark positions, to construct prediction regions. Specifically, we compute the empirical distribution of these errors over the calibration set and extract quantiles corresponding to the desired  $1 - \alpha$  confidence levels. These quantiles define the radii of fixed spherical regions around each prediction. In our proposed pipeline, conformal prediction complements the model’s uncertainty estimates, as it produces statistically valid confidence regions around predictions.

**Nonconformity Score:** To adapt the prediction region to the predicted uncertainty of each test sample, inspired by [5], we develop a dynamic nonconformity score based on the model’s predicted uncertainty,  $\sigma_{total}^2$  in Eq. (1). Since our task involves 3D landmark localization, we define the nonconformity score using a normalized Euclidean error between the predictions and ground-truth landmark displacements. For a sample  $i$  and landmark  $k$ , the nonconformity score is given by,

$$s_k^{(i)} = \frac{d(\Delta \hat{\mathbf{P}}_k^{(i)}, \Delta \mathbf{P}_k^{(i)})}{\sigma_{k\ total}^{2(i)}} \quad (3)$$

Where  $\Delta \hat{\mathbf{P}}_k^{(i)}$  denotes the predicted 3D displacement to reach landmark  $k$ ,  $\Delta \mathbf{P}_k^{(i)}$  is the ground truth displacement,  $\sigma_{k\ total}^{2(i)}$  is the total predicted uncertainty of the model,  $d(\cdot, \cdot)$  is the Euclidean distance. After computing these scores for all samples in the calibration set, we extract the empirical quantiles  $Q_{1-\alpha}^k$  for predefined  $\alpha$  values.

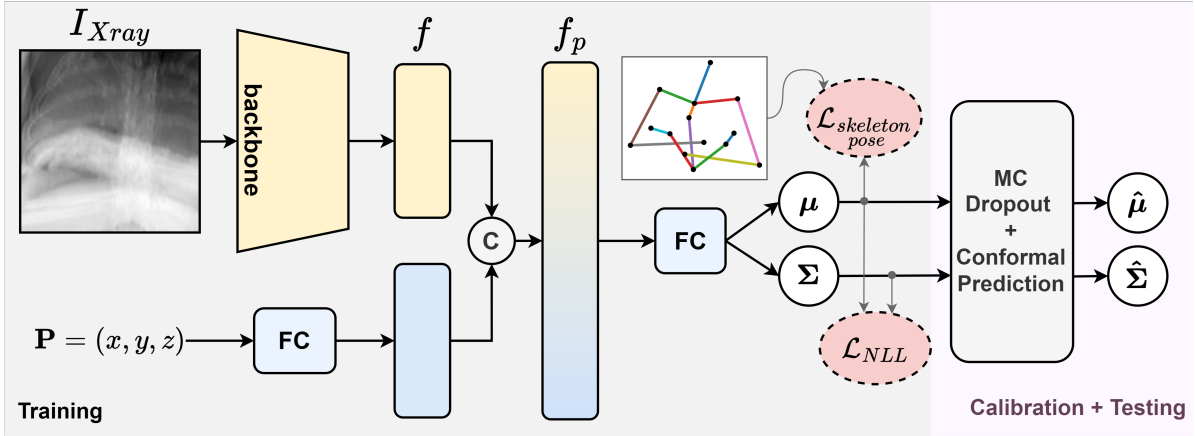


Figure 3. Overview of the proposed pipeline. The input X-ray is encoded using a backbone network, and the C-arm position is embedded into the image features. For each of the 14 anatomical landmarks, the model outputs the parameters of a 3D Gaussian: a mean vector  $\mu = (\mu_{\Delta x}, \mu_{\Delta y}, \mu_{\Delta z})$  and a diagonal covariance matrix  $\Sigma = \text{diag}(\sigma_{\Delta x}^2, \sigma_{\Delta y}^2, \sigma_{\Delta z}^2)$ , leading to  $14 \times 3 \times 2$  output neurons. A skeleton pose loss further regularizes predictions based on the patient’s prior anatomical topology. In Calibration and testing, we apply Monte Carlo dropout with  $T$  stochastic forward passes to estimate the total uncertainty.

**Prediction Region:** At test time, the prediction region for each landmark  $k$  is defined using the model’s predicted uncertainty and the precomputed quantile  $Q_{1-\alpha}^k$ , as follows,

$$R_k^i = \{ \mathbf{P} \in \mathbb{R}^3 : d(\mathbf{P}, \Delta \hat{\mathbf{P}}_k^{(i)}) \leq \sigma_{k\ total}^2 \cdot Q_{1-\alpha}^k \} \quad (4)$$

This region corresponds to a 3D sphere centered at the predicted displacement  $\Delta \hat{\mathbf{P}}_k^{(i)}$ , whose radius scales with the model’s uncertainty. Ensuring that the true landmark location lies within this region with probability  $1 - \alpha$ , which provides interpretable and risk-aware predictions.

### 4.3. Training Objectives

**Negative Log-Likelihood Loss:** As each regression target is modeled as a Gaussian, it is possible to perform Maximum Likelihood Estimation (MLE) through minimizing the negative log-likelihood (NLL) loss. Note that with our assumption of independence between landmarks and positional axes, the predicted covariance matrix is diagonal, which simplifies the NLL derivation. For a given landmark  $k$ , the NLL loss is given by,

$$\begin{aligned} \mathcal{L}_{NLL} &= -\log p(\Delta \hat{\mathbf{P}} | \mu, \Sigma) \\ &= \frac{1}{2} \beta \log |\Sigma| + \frac{1}{2} (\Delta \hat{\mathbf{P}} - \mu)^T |\Sigma| (\Delta \hat{\mathbf{P}} - \mu) + \gamma \\ &= \frac{1}{2} \sum_{i \in \{x, y, z\}} \left( \beta \log \sigma_i^2 + \frac{(\Delta \hat{P}_i - \mu_{\Delta i})^2}{\sigma_i^2} \right) + \gamma \end{aligned} \quad (5)$$

Where  $\Delta \hat{\mathbf{P}} = (\mu_{\Delta x}, \mu_{\Delta y}, \mu_{\Delta z})$ ,  $\beta$  is a constant that regularizes the variance prediction, and  $\gamma$  is a constant term

from the Gaussian likelihood function which can be omitted during optimization.

**Skeleton Pose Loss:** Landmarks are spatially correlated and constrained by skeletal geometry. To incorporate this structural prior, we introduce a skeleton pose loss that regularizes the model’s predictions by enforcing consistency with a predefined skeletal topology. For each CT scan, a custom skeletal graph is constructed from its landmark annotations. An example topology is illustrated in Fig. 3, capturing intuitive dependencies among landmarks in a humanoid form (e.g., wrists connected to shoulders, and the Carina, T1, and skull forming a central axis). We represent the skeleton pose as an undirected graph, where each node is a landmark, and edges represent connections. The loss is defined as the distance between the predicted graph and the ground truth skeleton graph. The loss is formally introduced as follows,

$$\mathcal{L}_{skeleton\ pose} = \sum_{(i,j) \in G} \left( d(\hat{\mathbf{P}}_i, \hat{\mathbf{P}}_j) - d(\mathbf{P}_i, \mathbf{P}_j) \right) \quad (6)$$

Where  $G$  is the prior skeleton graph,  $\hat{\mathbf{P}}$  is the location of the C-arm after the displacement, and  $\mathbf{P}$  is the actual landmark location. The total training loss used in our experiments is given by,

$$\mathcal{L} = \mathcal{L}_{NLL} + \lambda \mathcal{L}_{skeleton\ pose} \quad (7)$$

Where  $\lambda$  is a hyperparameter that determines how much influence the skeleton loss has on training.

### 4.4. Data Augmentation

In clinical practice, patients’ initial positioning varies due to time constraints and workflow differences across procedures. To account for this variability and improve the

Table 2. Pipeline performance with different backbones. In all backbones, the given calibration intervals are reliable, as PRCP values are close to their  $1 - \alpha$  levels.

Backbones	Mean distance from GT (mm) ↓	NLL ↓		PRCP		
		Calibration	Test	90%	95%	97%
ConvNeXt Base	<b>38.60</b>	<b>-2.18</b>	<b>-2.27</b>	90.18%	95.03%	96.92%
Resnet101	40.91	<b>-2.18</b>	-2.21	89.74%	94.71%	96.69%
Resnet34	44.04	-2.11	-2.13	89.18%	94.58%	96.83%
ViT Base/16	42.61	-2.13	-2.15	89.71%	94.58%	96.58%

generalizability of the model, we incorporate a patient-position augmentation method. This method simulates realistic shifts in patient placement by randomly perturbing their position within a controlled range. The augmentation is applied with probability  $p_o$  to the patient’s  $(x, y)$  coordinates as follows,

$$(x, y) = \begin{cases} (x, y) + shift & p \leq p_o \\ (x, y) & p > p_o, \end{cases} \quad (8)$$

Where  $shift \sim \mathcal{U}(-\eta, \eta)$  as  $\eta$  controls the augmentation strength: weak, medium, or strong.

## 5. Experiments

### 5.1. Implementation Detail

All models were implemented using PyTorch [26] and trained on a single NVIDIA V100 GPU. We used a batch size of 128 and trained each model for 50 epochs, where the loss converged to a stable value. Input X-ray images were resized to  $224 \times 224$  pixels, and optimization was performed using the AdamW optimizer [23].

For conformal prediction, we evaluated three significance levels  $\alpha \in \{0.1, 0.05, 0.03\}$ , corresponding to prediction confidence levels of 90%, 95%, and 97%, respectively. These values determine the desired coverage probability for the prediction regions generated during inference. For Monte Carlo Dropout, we perform  $T = 20$  stochastic forward passes per sample to estimate the epistemic uncertainty, and the dropout probability was set to  $p = 0.3$ . Unless otherwise noted, all reported results are computed on the test set described in Sec. 3, with calibration-specific metrics evaluated on the calibration set.

### 5.2. Evaluation Metrics

**Mean Euclidean Distance:** For a quantitative evaluation of the C-arm positioning, we use the Euclidean distance between the predicted and ground-truth landmark locations. This measures how far the predicted C-arm displacement places the system from the true landmark. We compute the mean distance across all landmarks and test samples to get a comprehensive evaluation. The measure is defined as follows,

$$\text{mean}(d_k^{(i)}) = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{k=1}^m d(\mathbf{P}_k^{(i)}, \hat{\mathbf{P}}_k^{(i)}) \quad (9)$$

Where  $n$  is the number of test samples,  $m$  is the number of anatomical landmarks. These distances are calculated between the ground-truth position of the landmarks  $\mathbf{P}$  and the corresponding predicted positions after applying the model’s estimated displacements  $\hat{\mathbf{P}}$ . For more fine-grained analysis, we also report per-landmark mean distances by omitting the inner average across  $m$  landmarks. Ideally, this metric should approach zero, indicating perfect alignment between predicted and actual C-arm positions.

**Negative-log Likelihood:** As we trained probabilistic models that represent a Gaussian distribution over displacements (Eq. (5)), it is natural to include the NLL loss for evaluation. This loss not only quantifies the deviation of the predicted means from the ground truth targets but also assesses the quality of the predicted variance, as it penalizes overconfident predictions. While the NLL may lack an intuitive interpretation compared to spatial errors, it provides a reliable quantitative measure of the overall model fit. For comparisons, lower NLL values indicate better predictive performance and more reliable variance estimates.

**Prediction Region Coverage Probability (PRCP):** We evaluate the reliability of the prediction regions by defining the Prediction Region Coverage Probability (PRCP) metric. PRCP quantifies the proportion of test-time predictions where the ground-truth landmark lies within the calibrated error sphere defined around the model’s output (see Eq. 4). Since these radii are computed from empirical quantiles of calibration errors, PRCP directly reflects how well the learned error bounds generalize to unseen data.

As the number of test samples increases, the empirical PRCP should converge to the target coverage level  $1 - \alpha$ . For instance, with  $\alpha = 0.05$ , the ideal PRCP would approach 95%. Formally, the PRCP is given as follows,

$$PRCP = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{k=1}^m \mathbb{I}\{d(\Delta \mathbf{P}_k^{(i)}, \Delta \hat{\mathbf{P}}_k^{(i)}) \in R_k^{(i)}\} \quad (10)$$

Where  $R_k^{(i)}$  is the spherical region from Eq. (4) that depends on the chosen  $1 - \alpha$  probability, and  $\mathbb{I}$  is the indicator function. For more detailed analysis, per-landmark PRCP can also be computed.

### 5.3. Quantitative results

**Backbone Comparison:** To give a comprehensive analysis of the proposed pipeline, we compare multiple different backbones. We test ResNet34 and ResNet101 [13] (standard backbones in medical imaging), ConvNeXt Base [21] (a modern convolutional architecture), and ViT Base/16 [6] (transformer-based model). For comparisons, we rely on the Euclidean distance, NLL, and PRCP. These metrics show both positioning accuracy and confidence modeling.

Tab. 2 presents the quantitative results. ConvNeXt Base achieves the best performance, with a mean distance of 38.6

Table 3. Performance per landmark. Regions with less movement tend to produce more accurate C-arm positioning, while moving ones are less stable. Most empirical PRCP values match the designed  $1 - \alpha$  confidence levels.

landmark	ConvNeXt Base			Resnet101			Resnet34			ViT Base/16						
	Distance from GT (mm) ↓	PRCP			Distance from GT (mm) ↓	PRCP			Distance from GT (mm) ↓	PRCP			Distance from GT (mm) ↓	PRCP		
		90%	95%	97%		90%	95%	97%		90%	95%	97%		90%	95%	97%
1: Skull	<b>29.25</b>	89.55%	94.44%	96.55%	31.49	87.63%	93.24%	95.62%	35.60	87.89%	93.55%	96.07%	33.62	89.70%	94.58%	96.39%
2: R Humeral Head	<b>29.52</b>	88.92%	94.46%	96.78%	32.54	87.55%	93.34%	95.80%	35.81	86.61%	92.70%	95.42%	34.54	88.97%	94.29%	96.36%
3: L Humeral Head	<b>30.04</b>	93.73%	98.04%	99.52%	32.99	93.75%	98.06%	99.51%	36.42	92.49%	97.66%	99.61%	35.31	92.31%	96.90%	98.82%
4: R Scapula	<b>38.70</b>	86.12%	92.79%	95.55%	40.96	85.40%	91.55%	94.56%	43.03	84.94%	92.23%	95.86%	42.91	86.80%	92.78%	95.47%
5: L Scapula	<b>36.07</b>	86.08%	92.40%	95.00%	38.09	86.64%	92.58%	94.99%	40.52	85.57%	92.85%	96.05%	40.28	85.43%	91.66%	94.62%
6: R Elbow	<b>45.94</b>	94.61%	98.39%	99.52%	47.53	95.35%	98.81%	99.64%	51.80	93.88%	98.01%	99.21%	49.52	94.28%	98.04%	99.22%
7: L Elbow	<b>48.61</b>	89.15%	93.61%	95.70%	50.48	89.26%	94.44%	96.57%	54.23	89.88%	94.92%	97.03%	51.56	91.08%	95.14%	96.92%
8: R Wrist	<b>58.02</b>	93.94%	96.84%	97.85%	60.27	92.94%	96.14%	97.54%	65.99	92.88%	95.87%	97.24%	61.10	92.55%	96.06%	97.50%
9: L Wrist	<b>63.14</b>	92.49%	96.46%	98.06%	65.66	93.79%	97.31%	98.56%	67.68	92.16%	96.81%	98.52%	65.51	94.21%	97.31%	98.36%
10: T1	<b>26.09</b>	89.07%	94.70%	96.77%	28.89	87.00%	93.53%	96.02%	32.31	87.87%	93.57%	95.98%	31.01	88.93%	94.55%	96.46%
11: Carina	<b>34.49</b>	83.06%	89.57%	92.33%	36.81	81.77%	88.55%	91.49%	39.53	81.90%	88.36%	91.37%	38.56	81.67%	88.22%	91.50%
12: R Hemidiaphragm	<b>33.39</b>	92.01%	95.84%	97.34%	34.89	91.77%	95.78%	97.37%	38.24	90.09%	95.25%	97.37%	37.29	90.13%	94.56%	96.45%
13: L Hemidiaphragm	<b>34.42</b>	90.99%	95.97%	97.81%	36.44	91.33%	96.09%	97.82%	39.04	90.78%	95.96%	97.90%	38.89	89.38%	94.55%	96.60%
14: T12	<b>32.78</b>	92.79%	96.85%	98.16%	35.73	92.24%	96.54%	98.16%	36.39	91.64%	96.44%	98.02%	36.42	90.55%	95.46%	97.50%

mm and NLL of -2.18 (calibration) and -2.27 (test) across all landmarks. Other benchmarks show comparative performance with differences not exceeding 6 mm. All models offer impressive PRCP values that closely match the calibrated confidence levels  $1 - \alpha$ . This signifies the generality of our pipeline regardless of the chosen backbone network.

**Per landmark Performance:** For a more detailed analysis, we report per-landmark results. These results are relevant in practice, as different use cases may prioritize different regions. From Tab. 3, the performance varies across landmarks, where moving ones are generally more challenging to track, e.g., wrists and elbows. However, the PRCP values are consistent across all landmarks, except the Carina. We hypothesize this is due to overconfident bounds in that region. We leave a more detailed analysis of this phenomenon for future work.

**Skeleton Pose Loss:** We introduced the skeleton pose loss in Eq. (7) to regularize the model, encouraging anatomically plausible landmark predictions. To evaluate the effect of this regularization, we vary the loss weight  $\lambda$ , which controls the influence of the skeleton term in the total loss function. Table 5 presents results across different  $\lambda$  values, from  $\lambda = 0$  (no regularization) to  $\lambda = 10$  (strong regularization). As  $\lambda$  increases, model performance improves consistently across all metrics, which shows that utilizing anatomical priors enhances both localization and confidence accuracy.

**Data Augmentation:** As described in Sec. 4, we introduced a data augmentation scheme that randomly positions the patient on the operating table. This mimics a real clinical setting where the initial position of the patient is not constant. We compare three levels of augmentation: weak, medium, and strong, where each gradually increases the magnitude

Table 4. All augmentation strengths improved ResNet101 performance compared to not applying it.

Augmentation Strength	Mean distance from GT (mm) ↓	NLL ↓		PRCP		
		Calibration	Test	90%	95%	97%
none	42.28	-2.14	-2.18	89.86%	95.02%	96.99%
weak	40.91	-2.18	-2.21	89.74%	94.71%	96.69%
mid	<b>39.14</b>	<b>-2.22</b>	<b>-2.28</b>	90.04%	94.79%	96.73%
strong	41.04	-2.17	-2.20	89.33%	94.57%	96.64%

of a patient’s shift across the table. Tab. 4 shows our comparison, where we can see that including data augmentation, regardless of its strength, improves performance. The medium level is the empirically optimal one.

#### 5.4. Qualitative Results

Qualitative results of our pipeline are presented in Fig. 4. The left images show randomly selected initial test X-rays, while the right images visualize the simulated C-arm position after applying the predicted displacements. The model successfully aligns the C-arm above all ROIs, demonstrating its localization ability in a simulated environment.

#### 5.5. Application: Stroke Thrombectomy

To demonstrate a practical application of our C-arm pipeline, we design an experiment to explore the following question: *Does taking multiple steps improve localization accuracy?* We simulate a clinical scenario involving stroke thrombectomy, where the ROI is the skull (Landmark 1), and intermediate steps involve targeting the T1 (10) and the Carina (11). These landmarks were chosen due to their rigid anatomical alignment, making them relatively fixed with respect to the skull. We design a multi-step prediction pipeline, where the output of each stage serves as the input for the subsequent stage, simulating iterative repositioning toward the target. In this experiment, we compare four pre-defined paths: direct prediction to landmark 1, and three multi-stage paths: **10**→**1**, **11**→**1**, and **11**→**10**→**1**.

As shown in Table Tab. 6, utilizing intermediate land-

Table 5. Changing the weight of the skeleton pose loss with ResNet101. As  $\lambda$  increases, the performance improves across all metrics.

$\lambda$	Mean distance from GT (mm) ↓	NLL ↓		PRCP		
		Calibration	Test	90%	95%	97%
0	41.47	-2.15	-2.20	89.67%	94.85%	96.88%
0.5	42.05	-1.99	-2.02	89.60%	94.99%	96.97%
1	40.91	-2.18	-2.21	89.74%	94.71%	96.69%
5	40.23	-2.23	-2.27	89.24%	94.46%	96.56%
10	<b>39.28</b>	<b>-2.28</b>	<b>-2.35</b>	90.20%	95.04%	96.94%

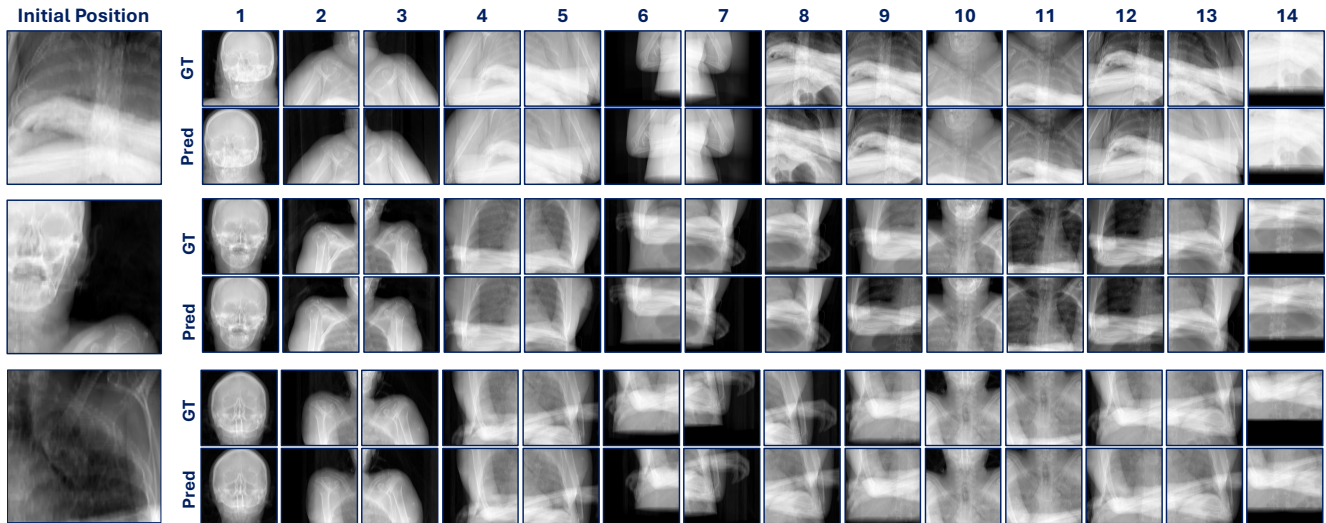


Figure 4. C-arm movement visualization. In all landmarks, the model was successful in localizing the C-arm onto the corresponding ROI.

Table 6. MAE and variance of 2D absolute error in the stroke thrombectomy application.

Metric	Landmark Path			
	1	10→1	11→1	11→10→1
MAE ↓	20.81	<b>18.97</b>	21.30	19.27
Absolute Error Variance	211.24	130.89	144.34	137.32

mark predictions reduces the 2D Mean Absolute Error (MAE) compared to direct prediction. It also influences the convergence quality, which is seen in the error variance values. A single-shot prediction to the skull leads to a much higher error variance value than all other multi-step predictions, meaning that the latter approach is more stable. Fig. 5 illustrates the test error distribution across the vertical and horizontal axes. Compared to single-shot prediction, the multi-stage path 11→10→1 achieves better error convergence, closer to 0. Further investigation is needed to explore different intermediate landmark combinations to reach different ROIs, which we leave for future work.

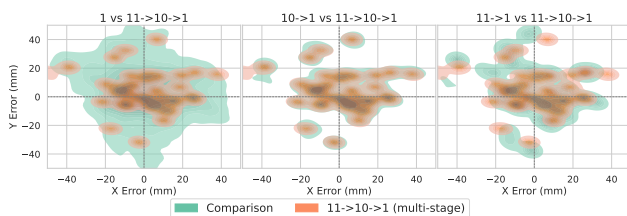


Figure 5. Kernel Density Estimates (KDE) of prediction errors in  $x$  and  $y$  directions. Long tails, due to outliers, were clipped for visibility.

## 6. Conclusion and Future Work

This work proposed a pipeline for C-arm control to achieve safe and reliable automation in interventional fluoroscopy.

We see this contribution as a step toward fully autonomous C-arm systems capable of moving from any initial point to any desired target. We focused on confidence and uncertainty quantification, which we view as potential sub-modules that can be integrated into broader autonomous systems. Our method demonstrated strong localization accuracy and reliable confidence calibration across a range of backbone architectures. However, as shown in Tab. 3, the errors vary across landmarks, with larger errors observed for anatomically dynamic regions. This suggests that a deeper analysis of landmark-specific variability is needed for clinical deployment, which we leave for future work.

Although the C-arm operates with 6 DoF, our current implementation controls only 3 translational axes. Extending the framework to include rotational control will be a key area of future research. Moreover, validating the applicability of our approach across various interventional procedures beyond Stroke Thrombectomy would enhance its generalizability. We plan to investigate broader procedures that benefit from uncertainty-aware automated C-arm positioning.

Another important step is to validate the framework with real X-ray data in place of synthetic DRRs. Real X-rays contain noise and artifacts that are absent in synthetic data, which better reflect operating settings. Finally, transferring the learned policies from simulation to a physical C-arm will be a natural extension, requiring engineering challenges and hardware interfacing.

## 7. Acknowledgments

This work was supported by the National Science Foundation under Grants No. 2218063.

## References

- [1] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul

- Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76: 243–297, 2021. 2
- [2] Fabio Arnez, Huáscar Espinoza, Ansgar Radermacher, and François Terrier. A comparison of uncertainty estimation approaches in deep learning components for autonomous vehicle applications. *CoRR*, abs/2006.15172, 2020. 2
- [3] A. Arrabi, J. Jung, J. Le, A. Nguyen, J. Reed, E. Stahl, N. Franssen, S. Raymond, and S. Wshah. C-arm guidance: A self-supervised approach to automated positioning during stroke thrombectomy. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, 2025. 2, 3
- [4] Cleveland Clinic. Orthopedic surgeon: Definition, expertise & specialties. <https://my.clevelandclinic.org/health/articles/orthopedic-surgeon-orthopedist>, 2022. Accessed: 2025-06-19. 1
- [5] Mathieu Cochetoux, Julien Moreau, and Franck Davoine. Uncertainty-aware online extrinsic calibration: A conformal prediction approach. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6167–6176, 2025. 4
- [6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. 6
- [7] Heather J.H. Edgar, S. Daneshvari Berry, E. Moes, N.L. Adolphi, P. Bridges, and K.B. Nolte. New mexico decedent image database, 2020. 2, 3
- [8] Hooman Esfandiari, Seyed Sadegh Mohseni Salehi, and Ali Gholipour. A deep learning approach for single shot c-arm pose estimation. In *EPiC Series in Health Sciences*, pages 162–169, 2020. 2
- [9] Shahriar Faghani, Mana Moassefi, Pouria Rouzrokh, Bardia Khosravi, Francis I. Baffour, Michael D. Ringler, and Bradley J. Erickson. Quantifying uncertainty in deep learning of radiologic images. *Radiology*, 308(2):e222217, 2023. PMID: 37526541. 2
- [10] Yarín Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 1050–1059, New York, New York, USA, 2016. PMLR. 4
- [11] Yarín Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2016. 2
- [12] Carlos González, Mauricio Escobar, Laura Daza, Fernando Torres, Gabriel Triana, and Pablo Arbeláez. SIMBA: Specific Identity Markers for Bone Age Assessment. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 749–759. Springer, Cham, 2020. 2
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 6
- [14] Ziyi Huang, Henry Lam, and Haofeng Zhang. Quantifying epistemic uncertainty in deep learning, 2023. 2
- [15] Ziyi Huang, Henry Lam, and Haofeng Zhang. Quantifying epistemic uncertainty in deep learning, 2023. 1, 4
- [16] Heinrich Jiang, Been Kim, Melody Y. Guan, and Maya Gupta. To trust or not to trust a classifier, 2018. 2
- [17] A Kashkoush, A Arrabi, J Jung, N Franssen, S Wshah, and S Raymond. E-151 deep learning model predicts biplane trajectory from pelvis to the head. *Journal of NeuroInterventional Surgery*, 17(Suppl 1):A193–A194, 2025. 1
- [18] Lisa Kausch, Simon Thomas, Holger Kunze, Maximilian Privalov, Sven Vetter, Jochen Franke, Andreas H. Mahnken, Lena Maier-Hein, and Klaus Maier-Hein. Toward automatic c-arm positioning for standard projections in orthopedic surgery. *International Journal of Computer Assisted Radiology and Surgery*, 15(7):1095–1105, 2020. 2
- [19] Barret Kompa, Jasper Snoek, and Andrew L. Beam. Second opinion needed: communicating uncertainty in medical machine learning. *npj Digital Medicine*, 4(1):4, 2021. 2
- [20] Atul Kumar, Siddharth Garg, and Soumya Dutta. Uncertainty-aware deep neural representations for visual analysis of vector field data, 2024. 2
- [21] Zhuang Liu, Hanzi Mao, Chao Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings - 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022*, pages 11966–11976. IEEE Computer Society, 2022. Publisher Copyright: © 2022 IEEE.; 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022 ; Conference date: 19-06-2022 Through 24-06-2022. 6
- [22] Timothy J. Loftus, Benjamin Shickel, Matthew M. Ruppert, Jacob A. Balch, Tezcan Ozrazgat-Baslanti, Patrick J. Tighe, Philip A. Efron, William R. Hogan, Parisa Rashidi, Gilbert R. Upchurch, and Azra Bihorac. Uncertainty-aware deep learning in healthcare: A scoping review. *PLOS Digital Health*, 1(8):e0000085, 2022. 2
- [23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. 6
- [24] J. Luo, A. Sedghi, K. Popuri, D. Cobzas, M. Zhang, F. Preiswerk, M. Toews, A.J. Golby, M. Sugiyama, W.M. Wells, and S. Frisken. On the applicability of registration uncertainty. 11765:410–419, 2019. 2
- [25] J. Luo, S. Frisken, D. Wang, A.J. Golby, M. Sugiyama, and W.M. Wells. Are registration uncertainty and error monotonically associated? 12263:264–274, 2020. 2
- [26] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 6
- [27] Christopher A. Potter et al. Ct for treatment selection in acute ischemic stroke: A code stroke primer. *RadioGraphics*, 39(6):1717–1738, 2019. 1

- [28] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(12): 371–421, 2008. [4](#)
- [29] Zhenglong Shao, Yukun Guan, and Jian Tan. Virtual reality aided positioning of mobile c-arms for image-guided surgery. *Advances in Mechanical Engineering*, 6:1–10, 2014. [1](#), [2](#)
- [30] Freddie Bickford Smith, Jannik Kossen, Eleanor Trollope, Mark van der Wilk, Adam Foster, and Tom Rainforth. Rethinking aleatoric and epistemic uncertainty, 2024. [2](#)
- [31] Freddie Bickford Smith, Jannik Kossen, Eleanor Trollope, Mark van der Wilk, Adam Foster, and Tom Rainforth. Rethinking aleatoric and epistemic uncertainty, 2024. [1](#), [4](#)
- [32] Laura K Stein, J Mocco, Johanna Fifi, Nathalie Jette, Stanley Tuhim, and Mandip S Dhamoon. Correlations between physician and hospital stroke thrombectomy volumes and outcomes: A nationwide analysis. *Stroke*, 52(9):2858–2865, 2021. [1](#)
- [33] Mathias Unberath, Javad Fotouhi, Johannes Hajek, Andreas Maier, Greg Osgood, Russell Taylor, Mehran Armand, and Nassir Navab. Augmented reality-based feedback for technician-in-the-loop c-arm repositioning. *Healthcare Technology Letters*, 5(5):143–147, 2018. [2](#)
- [34] Mathias Unberath, Jan-Nico Zaeck, Sing Chun Lee, Bastian Bier, Javad Fotouhi, Mehran Armand, and Nassir Navab. Deepdr – a catalyst for machine learning in fluoroscopy-guided procedures, 2018. [1](#), [2](#), [3](#)
- [35] Ke Wang, Chongqiang Shen, Xingcan Li, and Jianbo Lu. Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications. *IEEE Transactions on Intelligent Transportation Systems*, 26(3):2880–2896, 2025. [2](#)
- [36] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald Summers. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, 2017. [2](#)
- [37] Ke Zou, Zhihao Chen, Xuedong Yuan, Xiaojing Shen, Meng Wang, and Huazhu Fu. A review of uncertainty estimation and its application in medical imaging. *Meta-Radiology*, 1(1):100003, 2023. [2](#)