ID:56

A Handful of Data: Evaluating Few–shot Incremental Landmark Detection

Anonymous

Abstract. While automated landmark detection in medical imaging has achieved remarkable accuracy, it still requires sufficiently annotated datasets. This remains a significant barrier to clinical adoption. This paper investigates the effect of the number of annotations available on the model performance without relying on overly specialised few-shot configurations. We explore two practical scenarios: landmark detection with limited annotated data (≤ 60), and the incremental addition of new landmarks to existing models. Through experiments on hand radiographs, we demonstrate that models trained on just a fraction of the full dataset can achieve an accuracy comparable to that of other methods. Furthermore, we show that new landmarks can be effectively learnt through fine-tuning with as few as five examples, though performance varies with landmark variance. Also, we validate weight initialisation performance and find fine-tuning from prior landmark models tend to under-perform. Our findings suggest that the relationship between the amount of annotated training data and detection accuracy is nonlinear, with diminishing accuracy gains beyond certain thresholds. This insight has important implications for clinical practice, suggesting that label-efficient, modular landmark detection systems are valuable options, particularly when sub-millimetre precision is not critical.

Keywords: Landmark detection, Keypoints, X-Ray, Few-Shot

1 Introduction

Landmark annotations are used for clinical diagnostics and treatment planning, and require different levels of precision depending on the task. For sensitive localisation tasks (e.g. osteotomy), sub-millimetre predictions are critical. For other tasks, such as less sensitive localising or sketching an anatomical feature, detecting a landmark around the contour of the feature can have large inter-annotator variance; therefore, a regional probability heatmap is an acceptable substitute.

Automated methods for landmark-based detection in orthopaedic tasks have shown performance within *mm* of clinicians' annotations [1–4]. Due to the difficulty in curating expert landmark annotations for medical imaging tasks, few-shot learning methods have been explored [5–9].

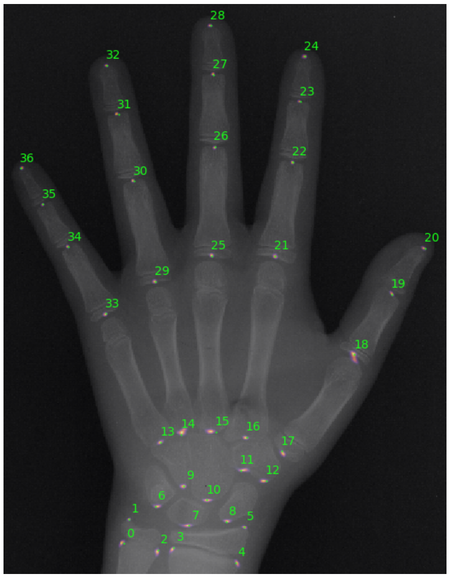


Fig. 1: Hand X-ray annotated with ground truth landmarks (green) and predicted heatmaps, as per Section 4.1.

Landmark detection methods have been shown to reach sub-millimetre precision [3, 10–13]. As such, this paper does *not* propose to reduce a few *mm* from the current best performance. Rather, we empirically investigate the relationship between the need for human annotations and the model's detection ability.

In particular, we investigate two scenarios:

- the number of full image annotations required to detect landmarks convincingly,
- the number of single landmark annotations required to add a fresh, unseen landmark to a known set without overly specialised few-shot methods.

We show that hundreds of annotations may not always be required. If an accuracy within a couple of *mm* is acceptable as a precursor to another task, then a few annotations should suffice.

Moreover, we show that the incremental few-shot procedure, which introduces an additional landmark with only a few examples, yields better performance when using ImageNet-1k weight initialisation over reusing a model already trained on a large number of landmark annotations. This exploration into model initialisation allows us to assess whether the model's performance on the handful of extra examples is both reliable and not overly confident.

2 Related Work

Landmark detection is typically approached via heatmap regression [3, 14], as direct coordinate regression suffers from spatial ambiguity. However, care must be taken for evaluating heatmap regression as this introduces quantisation issues due to discretised output resolution. Most works optimise these heatmaps using either binary cross-entropy (BCE) loss [1] or negative log-likelihood (NLL) loss [3].

More recently, one-shot methods have been developed to reduce the need for multiple annotations [15, 16]. Zhu et. al. [16] developed a two-stage model which includes: a contrastive learning approach for multi-domain data to generate pseudo-landmark labels and secondary supervised learning method for domain-adaptation leveraging transformers. A recent one-shot method exceeded these results by leveraging foundation models for one-shot landmark detection [6]. They utilise a bi-directional feature matching scheme between a template image and a query image that aids in the massive reduction of data. Although the model learns an improved generalisation, outperforming other one-shot methods, this work does not perform as well as existing methods with multiple annotations [1].

These methods focus on adapting few or one-shot learning strategies to learn a set of landmarks for a clinical application. This work does not aim to remove all annotations. Instead, we consider an incremental scenario where, after fully training a landmark model, an additional landmark is later introduced. We ask: How many annotations are required for this new landmark to achieve performance comparable to the originally trained landmarks? This, therefore, extends annotation labelling procedures to instead focus on more difficult landmarks. To our knowledge, no work has evaluated the effect of an additional landmark that has not been seen by the network.

3 Setup

3.1 Dataset

The Digital Hand Atlas dataset [17] contains 909 X-rays of left hands with an average resolution of 1560×2167 px. Each image is annotated with 37 landmarks (or keypoints, KP): wrist KP0-KP17 and fingers KP18-KP36. An example is shown in Figure 1. As with the original proponents [10], the pixel size is approximated by making the inconsistent assumption that the wrist diameter is $50mm$. In practicality, scans where precise localisation is required, a calliper is used to calculate the scale. Following other prior work, we evaluate in mm, but additionally report px as the assumed wrist diameter will naturally change between patients. During training, the images are padded to $1:1$ aspect ratio on the bottom-right, downsampled to 512×512, and augmented the same way as in [10]. The predicted landmark pixel coordinates are kept in 512×512 space. The other popular public landmark detection dataset has been taken off-line [18].

In literature, there are two different commonly used splits for the hand dataset. One with 895 images, with three cross-validation splits - which we will refer to as the *Payer* split [10]. Another uses all 909 images, split into 546 images for training and 63 images for validation, and the remaining 300 for testing - which we will refer to as the *Zhu* split [1].

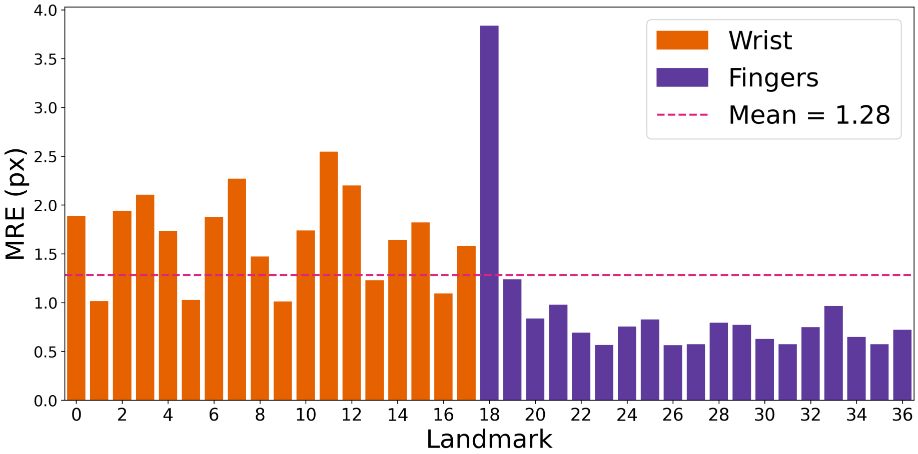


Fig. 2: MRE (in px) per-landmark results for the fully supervised version on the test set, using the *Zhu* split. The average MRE across all landmarks is shown as the red dashed line. Best KP23 at 0.57 px. Worst KP18 at 3.84 px.

3.2 Architecture

We use a UNet[1] following the same heatmap regression protocol from [3], with the same decoder channels for each level and a detection head which outputs a 512×512 image with N channels (N is the number of landmarks to predict). Conventionally, the network is initialised with a ResNet-34 encoder pre-trained on ImageNet-1k [19].

Each experiment is run for 10 epochs, with an NLL loss, Adam optimizer [20] , and MultiStepLR scheduler starting from a learning rate of 10^{-3}. Continuing from McCouat *et al.*'s [3] implementation, we apply a 2D softmax over each output channel, producing a probability distribution focusing on a single-pixel activation (Figure 1). The coordinates of each predicted landmark are those of the hottest point in its output heatmap.

For inference, the results from the paper are generated by taking M groups of 3 UNet models with different random seeds. For each group, the final heatmap is computed by averaging the outputs from the three models. Metrics are then calculated from each group-averaged heatmap and finally averaged across the M groups to produce the final results. Unless specified, we set $M = 3$.

3.3 Evaluation metrics

As with other landmark detection work [1, 4, 10], Mean Radial Error (MRE) and Successful Detection Rate (SDR) were calculated to evaluate the landmark localisation. MRE is the average Euclidean distance between the ground truth

---

[1] `segmentation-models-pytorch.readthedocs.io`

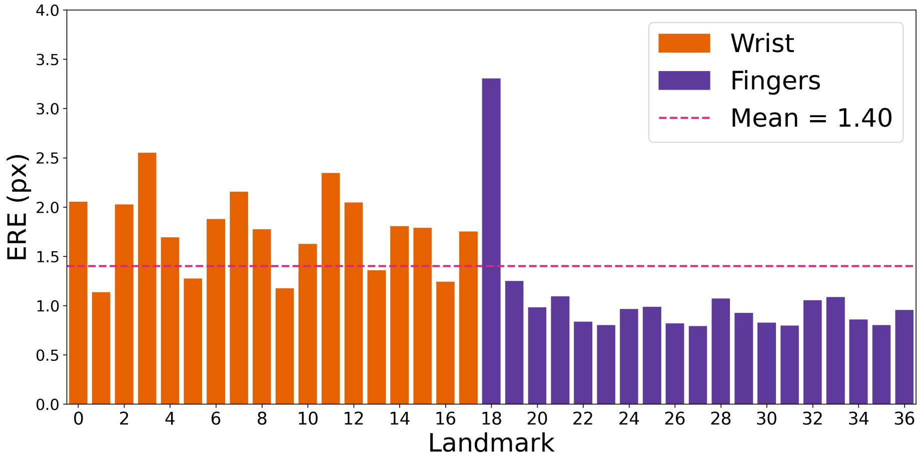


Fig. 3: ERE (in px) per-landmark results for the fully supervised version on the test set, using the *Zhu* split. The average EREs (per-landmark) is reported and shown as the red dashed line. Best KP23 at 0.80 px. Worst KP18 at 3.31 px.

and the predicted landmark. SDR can then be calculated as the percentage of predicted landmarks that fall within a specified threshold of the MRE (usually $2\,mm$) [21]. As previously mentioned, we report in mm for easy backwards comparisons and in px (out of 512).

The Expected Radial Error (ERE) is reported to evaluate the uncertainty in each landmark prediction corresponding to the spread of each heatmap, and can help assess model confidence in few-shot learning scenarios [3].

4 Experiments and Results

4.1 Baseline

We first evaluated the performance of a fully supervised method, where we trained on all 37 landmarks, and compare to previous literature. We report an average over the three cross-validations in the *Payer* split, as in [10, 12, 13] (see Table 1).

Next, we evaluate the per-landmark performance of a fully supervised method on the *Zhu* split. This method shows an MRE of 1.28 px for the wrist and fingers, with per-landmark results in Figure 2 and an ERE of 1.40 px (out of 512), with per-landmark results in Figure 3. As such, this indicates near-optimal performance given the quantisation limits at this resolution.

As these experiments did not require multiple stages of training and fine-tuning, the computational costs were lower so we temporarily increase M to 5.

Table 1: Comparison of existing methods on the Hand dataset. The first half uses the *Zhu* split, the second half and the current method uses the *Payer* split. † indicates the model is trained on mixed data.

Method	**Label/ Unlabel**	**MRE (mm)**	**SDR (%)**		
			2mm	**4mm**	**10mm**
Zhu et al. [1] †	988/0	0.84	95.40	99.35	99.75
Miao et al. [6]	1/0	1.41	86.66	96.66	99.11
Han et al. [22]	1/608	1.43	75.90	96.93	99.6
Zhu et al. [16] †	3/985	2.52	53.37	84.27	97.59
Di Via et al. [2]	909/0	**0.65**	**96.90**	**99.84**	**99.95**
Urschler et al. [12]	895/0	0.80	92.19	98.46	99.95
Lindner et al. [13]	895/0	0.85	93.68	98.95	99.94
Payer et al. [10]	895/0	**0.66**	94.99	99.27	**99.99**
McCouat et al. [11]	895/0	0.67	**95.97**	**99.48**	**99.99**
Current method	895/0	0.75	93.83	98.83	99.91

Table 2: Fully supervised 37-landmark detection with varying randomised training/validation splits (70/10%, 20/40%, 15/5%) of the 895 images used in the *Payer* split, listed by image count.

Train / Val (#)	**MRE (mm)**	**SDR (%)**			**MRE (px)**	**SDR (%)**			**ERE (px)**
		2mm	**4mm**	**10mm**		**2px**	**4px**	**10px**	
626 / 90	**0.76 ± 1.37**	**94.01**	**98.92**	**99.87**	**1.38 ± 2.49**	**77.13**	**94.23**	**99.67**	**1.45**
179 / 358	0.89 ± 2.29	92.31	98.41	99.71	1.61 ± 4.02	73.61	92.69	99.31	1.77
134 / 45	0.97 ± 2.74	91.59	98.10	99.62	1.77 ± 5.09	71.89	91.98	99.10	2.13

4.2 Reduced training data

Since annotation scarcity remains the predominant shortcoming in medical image analysis, we explore the reduction in training data. We vary the number of images used for training and validation on the 895 images used in the *Payer* split, using 70/10%, 20/40%, and 15/5% of the data for training/validation, allocating the remaining data for testing.

We randomly generate splits and show results in Table 2. This transfer learning approach gives sub-millimetre accuracy and some less sensitive applications would function well with the sub-millimetre accuracy achieved (training and validating with only 20% of annotations).

4.3 Few-shot fine-tuning on additional landmarks

Training a model with only a handful of fresh example annotations can reduce the clinical annotation burden. To understand whether this is possible we train our model on 36 of the landmarks, and add the additional landmark at a later stage. In the Appendix (Tables 5 and 6), we extend such analysis to include the addition of two landmarks, following the same strategy.

Table 3: Fine-tuning results for KP18 and (separately) KP23, on the *modified Zhu* split.

# Fine Tune Images	**MRE (mm)**	**SDR (%)**			**MRE (px)**	**SDR (%)**			**ERE (px)**
		2mm	**4mm**	**10mm**		**2px**	**4px**	**10px**	
				KP18					
60	$\mathbf{1.95 \pm 1.07}$	54.98	**96.15**	**100.00**	$\mathbf{3.58 \pm 2.00}$	**21.12**	57.50	**100.00**	**3.08**
30	2.13 ± 2.40	**56.18**	93.89	99.73	3.89 ± 4.03	20.32	**59.10**	98.54	3.35
5	3.08 ± 8.06	55.11	93.23	97.88	5.57 ± 14.04	16.07	58.43	97.34	7.16
				KP23					
60	$\mathbf{0.36 \pm 0.32}$	**100.00**	**100.00**	**100.00**	$\mathbf{0.66 \pm 0.57}$	**98.94**	**100.00**	**100.00**	**1.76**
30	1.00 ± 5.79	98.41	98.41	98.41	1.79 ± 9.95	97.74	98.41	98.41	2.86
5	0.58 ± 2.56	99.73	99.73	99.73	1.02 ± 4.17	98.14	99.73	99.73	2.31

Table 4: ImageNet weight initialisation results for KP18 and (separately) KP23, on the *modified Zhu* split.

# Training Images	**MRE (mm)**	**SDR (%)**			**MRE (px)**	**SDR (%)**			**ERE (px)**
		2mm	**4mm**	**10mm**		**2px**	**4px**	**10px**	
				KP18					
60	$\mathbf{1.93 \pm 1.52}$	**59.10**	94.69	99.87	$\mathbf{3.57 \pm 3.00}$	21.51	62.02	**99.47**	3.16
30	1.99 ± 1.20	58.30	94.42	**100.00**	3.66 ± 2.28	**31.87**	**72.11**	99.20	**3.04**
5	2.18 ± 2.52	49.27	**96.28**	99.73	4.01 ± 4.93	12.75	53.25	99.34	3.71
				KP23					
60	$\mathbf{0.41 \pm 0.33}$	**100.00**	**100.00**	**100.00**	$\mathbf{0.74 \pm 0.59}$	**97.08**	**100.00**	**100.00**	**1.09**
30	0.65 ± 2.86	99.47	99.73	99.73	1.16 ± 4.84	93.09	99.60	99.73	1.73
5	2.13 ± 5.22	75.30	95.09	98.94	3.91 ± 9.52	40.50	77.29	98.41	9.80

Based on per-landmark error values in Section 4.1, we select the best– and worst–performing keypoints in the baseline model to be used (separately) for this experiment. KP23 has lowest, and KP18 has the highest MRE and ERE. KP18 is the hardest to predict due to inconsistent ground truth placements - either on the inferior or superior bone in the thumb's metacarpophalangeal (MCP) joint [11].

These form the basis of separate investigations. A base model, whose encoder is initialised with ImageNet-1k weights, is trained on the remaining 36 keypoints; it is then fine–tuned using only a few (5, 30 or 60) instances of the new keypoint. For a second approach, we consider a model, again whose encoder is initialised the same, but trained only on the new keypoint with few instances.

When fine-tuning on 60 and 30 images, we change the starting learning rate to 1×10^{-5}, to reduce catastrophic forgetting from the previous 36 landmarks. We do not change the learning rate with 5 training images as we found the models not to converge with the same number of epochs.

The data split is the same as the *Zhu* split, except the first 50 images from the test set and 30 images from the validation set were instead used for fine-tuning (removing them from the testing/validation pool). The models to fine-tune were first trained on the same 546 images as the previous experiment and validated on the reduced validation set of 33 images. We then varied the number of images for the training set when fine-tuning and used a validation set of 20 images of the fresh landmark. All models trained only on this new landmark used the same fine-tune set. Finally, all models were tested on the same test set of 250 images. We refer to this split as the *modified Zhu* split.

This uses the same model procedure as Sec. 3.2. The results for these tests are shown in Tables 3 and 4.

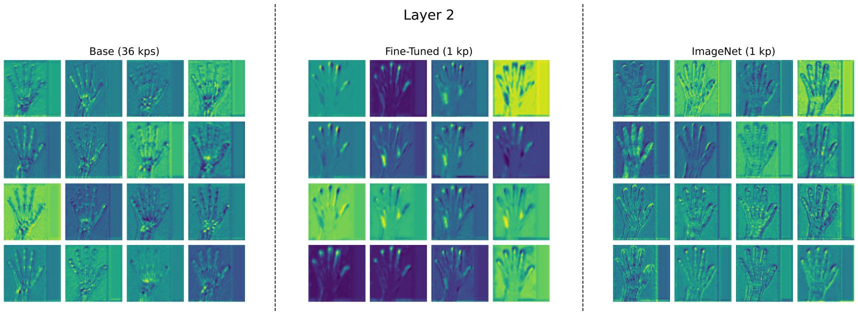


Fig. 4: First 16 feature maps from the second layer output of the ResNet encoder (128 channels total, each of shape 64x64), on a single template image. (Left) Base model trained on 36 landmarks (excluding KP23). (Middle) Fine-tuned from the base model on the additional KP23. (Right) Trained from ImageNet-1K weights on the additional KP23.

5 Analysis

It is evident from Tables 3 and 4 that fine-tuning the model on an additional 30 or 60 examples of the extra landmark is sufficient for a sub-millimetre detection rate. With an 'easy' landmark such as KP23, the model gets quite low MRE and standard deviation. With the 'harder' KP18 the error is marginally higher, but still acceptable for some tasks, when $2mm$ error is viable.

The network's confidence—measured inversely via ERE—increases as ERE decreases with the number of examples. It is encouraging to see that the network is not confident when trained with only 5 fresh examples, and then its confidence increases as we expand the number of examples. As expected, its confidence is higher with KP23 than it is with KP18.

A clinical user of such a scheme would want to consider the correlation between such a measure of confidence and the amount of fresh data presented to the network. Therefore, aiding the clinician in deciding the quantity of annotations when adding additional landmarks.

The difference with Tables 3 and 4 is the weight initialisation. Interestingly, the ImageNet-1k initialisation performance is generally better, this can be explained from the learned feature maps in Figure 4. The fine-tuned maps at the second layer of ResNet-34 retain too much of the base model weights and, therefore, lose fine-detail signals which are required for translating to new landmarks. We also show layers 1 and 3 in the Appendix (Figure 5).

When evaluating few-shot scenarios like the one described here, as well as incremental learning in general, it is also important to ensure that the initial landmarks have not been ‘forgotten.’ In our case, incremental forgetting is mitigated by modularising the original and new landmark models and ensembling them, thereby preserving the performance of both on their respective landmarks - as seen in Table 7.

6 Conclusions

Whilst high precision clinical techniques will continue to need fully supervised models relying on thousands of annotations, we show that less training data can still be used to generate practicable landmark predictions. These can be used as an intermediate step in localisation or downstream diagnostic tasks. More investigations are necessary in order to extend these preliminary results to other landmark detection datasets (e.g. CeLDA [4] and CEPHA29 [23]). Future work should evaluate out-of-distribution performance, given that these scenarios are typically those that require clinical intervention.

Unsurprisingly, all model variants are successful at detecting a single landmark. As such, expensive and bespoke few-shot training and diverse augmentations are only worth using when high precision is required (e.g. surgical planning and intra-operative guidance).

7 Ethical considerations

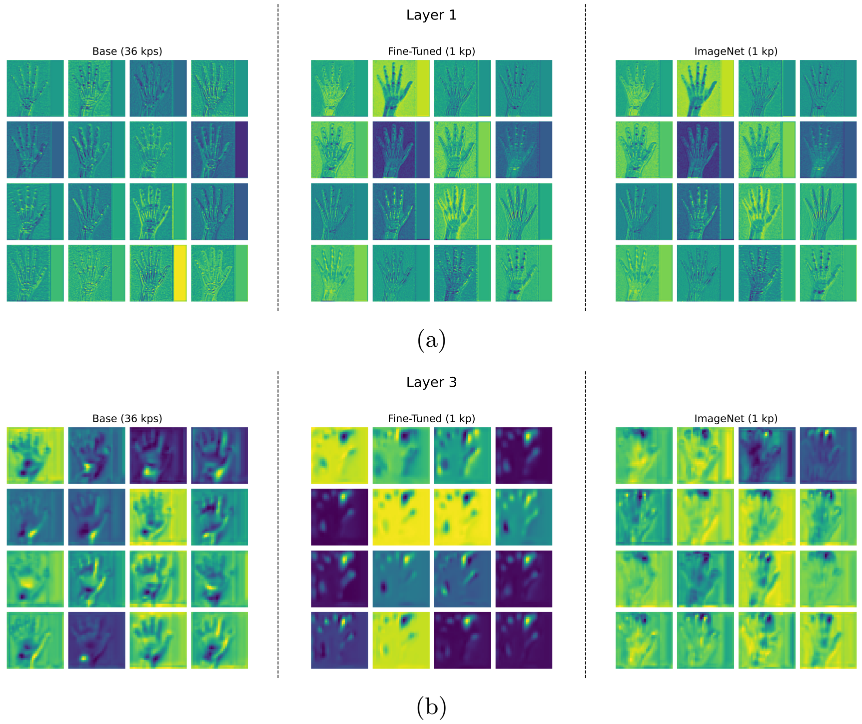
This study was conducted retrospectively using ethically acquired publicly available human subject data. The authors have no interests to disclose.

References

1. Zhu et al., “You only learn once: Universal anatomical landmark detection,” in *MICCAI*, 2021, p. 85–95.
2. Di Via et al., “Is in-domain data beneficial in transfer learning for landmarks detection in x-ray images?,” in *ISBI 2024*, 2024, pp. 1–5.
3. James McCouat and Irina Voiculescu, “Contour-hugging heatmaps for landmark detection,” in *Proceedings of the IEEE/CVF CVPR*, 2022, pp. 20597–20605.
4. Wu et al., “Cephalometric landmark detection across ages with prototypical network,” in *MICCAI*. Springer, 2024, pp. 155–165.
5. Roberto Di Via, Francesca Odone, and Vito Paolo Pastore, “Self-supervised pretraining with diffusion model for few-shot landmark detection in x-ray images,” in *WACV*, February 2025, pp. 3886–3896.
6. Juzheng Miao, Cheng Chen, Keli Zhang, Jie Chuai, Quanzheng Li, and Pheng-Ann Heng, “Fm-osd: Foundation model-enabled one-shot detection of anatomical landmarks,” in *MICCAI*. Springer, 2024, pp. 297–307.

7. Heqin Zhu, Quan Quan, Qingsong Yao, Zaiyi Liu, and S Kevin Zhou, "Uod: Universal one-shot detection of anatomical landmarks," in *MICCAI*. Springer, 2023, pp. 24–34.
8. Zihao Yin, Ping Gong, Chunyu Wang, Yizhou Yu, and Yizhou Wang, "One-shot medical landmark localization by edge-guided transform and noisy landmark refinement," in *ECCV*. Springer, 2022, pp. 473–489.
9. Quan Quan, Qingsong Yao, Jun Li, and S Kevin Zhou, "Which images to label for few-shot medical landmark detection?," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20606–20616.
10. Christian Payer et al., "Integrating spatial configuration into heatmap regression based cnns for landmark localization," *Medical Image Analysis*, pp. 207–219, 2019.
11. James McCouat, *Quantifying uncertainty in landmark detection for medical imaging*, Ph.D. thesis, University of Oxford, 2023.
12. Urschler et al., "Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization," *MedIA 2017*, 2017.
13. Lindner et al., "Robust and accurate shape model fitting using random forest regression voting," in *ECCV 2012*. 2012, pp. 278–291, Springer.
14. Chen et al., "Cephalometric landmark detection by attentive feature pyramid fusion and regression-voting," in *MICCAI*, 2019, pp. 873–881.
15. Qingsong Yao, Quan Quan, Li Xiao, and S Kevin Zhou, "One-shot medical landmark detection," in *MICCAI*. Springer, 2021, pp. 177–188.
16. Heqin Zhu, Quan Quan, Qingsong Yao, Zaiyi Liu, and S Kevin Zhou, "Uod: Universal one-shot detection of anatomical landmarks," in *MICCAI*. Springer, 2023, pp. 24–34.
17. Christian Payer et al., "Regressing heatmaps for multiple landmark localization using cnns," in *MICCAI 2016*. 2016, pp. 230–238, Springer.
18. Wang et al., "Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge," *IEEE transactions on medical imaging*, vol. 34, no. 9, pp. 1890–1900, 2015.
19. Kaiming He et al., "Deep residual learning for image recognition," in *2016 IEEE CVPR*, 2016, pp. 770–778.
20. Diederik P Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
21. Yueh-Peng Chen et al., "Automatic and human level Graf's type identification for detecting DDH," *Biomed J*, vol. 47(2), pp. 100614, 2024.
22. Lu Han, Boyu Chen, Zherui Zhang, Li Guo, and Shibiao Xu, "Scs: Spatially consistent self-supervised approach for one-shot anatomical landmark detection," in *ICASSP 2025*. IEEE, 2025, pp. 1–5.
23. Muhammad Anwaar Khalid, Kanwal Zulfiqar, Ulfat Bashir, Areeba Shaheen, Rida Iqbal, Zarnab Rizwan, Ghina Rizwan, and Muhammad Moazam Fraz, "Cepha29: automatic cephalometric landmark detection challenge 2023," *arXiv preprint arXiv:2212.04808*, 2022.

8 Appendix



(a)

(b)

Fig. 5: First 16 feature maps from (a) the first layer output of the ResNet encoder (64 channels total, each of shape 128x128); (b) the third layer output of the ResNet encoder (256 channels total, each of shape 32x32), on a single template image. (Left) Base model trained on 36 landmarks (excluding KP23). (Middle) Fine-tuned from the base model on the additional KP23. (Right) Trained from ImageNet-1K weights on the additional KP23.

Table 5: Fine-tuning results for KP23 and (separately) KP26, when fine-tuning on both landmarks simultaneously, on the *modified Zhu* split.

# Fine Tune Images	MRE (mm)	SDR (%)			MRE (px)	SDR (%)			ERE (px)
		2mm	4mm	10mm		2px	4px	10px	
KP23									
60	**0.32 ± 0.31**	**100.00**	**100.00**	**100.00**	**0.59 ± 0.57**	**99.20**	**100.00**	**100.00**	**1.33**
30	0.73 ± 3.54	99.34	99.34	99.34	1.27 ± 5.78	98.54	99.34	99.34	2.43
5	0.46 ± 0.80	99.87	99.87	99.87	0.82 ± 1.34	97.34	99.87	99.87	3.15
KP26									
60	**0.79 ± 4.30**	**99.07**	**99.07**	**99.07**	**1.47 ± 8.35**	**98.01**	**99.07**	**99.07**	**2.30**
30	1.33 ± 6.47	97.88	97.88	97.88	2.45 ± 12.18	97.34	97.88	97.88	3.86
5	2.38 ± 10.40	96.15	96.15	96.15	4.29 ± 18.47	92.30	96.15	96.15	9.04

Table 6: ImageNet weight initialisation results for KP23 and (separately) KP26, when learning both landmarks simultaneously, on the *modified Zhu* split.

# Fine Tune Images	MRE (mm)	SDR (%)			MRE (px)	SDR (%)			ERE (px)
		2mm	4mm	10mm		2px	4px	10px	
KP23									
60	**0.37 ± 0.32**	**100.00**	**100.00**	**100.00**	**0.67 ± 0.58**	**98.54**	**100.00**	**100.00**	**1.17**
30	1.52 ± 7.05	98.01	98.01	98.01	2.81 ± 13.15	92.30	98.01	98.01	3.52
5	1.70 ± 2.01	76.49	93.36	99.34	3.07 ± 3.50	41.97	77.03	98.14	6.42
KP26									
60	**0.61 ± 3.14**	**99.60**	**99.60**	**99.60**	**1.14 ± 6.09**	**96.55**	**99.60**	**99.60**	**2.00**
30	1.52 ± 7.05	98.01	98.01	98.01	2.81 ± 13.15	92.30	98.01	98.01	3.52
5	5.75 ± 12.66	71.71	86.72	90.04	10.52 ± 23.17	35.46	72.78	89.64	12.58

Table 7: Overall 37 landmark performance, using the base models trained on 36 landmarks, and the corresponding fine-tuned models for the new landmark prediction for KP18 and (separately) KP23, on the *modified Zhu* split.

# Fine Tune Images	MRE (mm)	SDR (%)			MRE (px)	SDR (%)			ERE (px)
		2mm	4mm	10mm		2px	4px	10px	
KP18									
60	**0.70 ± 0.78**	94.82	**99.29**	**99.97**	**1.26 ± 1.40**	**78.44**	95.09	**99.82**	**1.39**
30	0.70 ± 0.87	**94.85**	99.22	**99.97**	1.27 ± 1.53	78.42	**95.14**	99.78	1.40
5	0.75 ± 2.26	94.80	99.19	99.89	1.36 ± 3.93	78.29	95.11	99.72	1.55
KP23									
60	**0.71 ± 0.85**	**94.64**	**98.93**	**99.97**	**1.29 ± 1.56**	**79.04**	**94.90**	**99.76**	**1.42**
30	0.72 ± 1.28	94.59	98.89	99.93	1.32 ± 2.27	79.01	94.86	99.72	1.45
5	0.71 ± 1.04	94.63	98.92	**99.97**	1.30 ± 1.85	79.02	**94.90**	99.75	1.44