# Landmarks Are Alike Yet Distinct: Harnessing Similarity and Individuality for One-Shot Medical Landmark Detection

Xu He[1,2*], Zhen Huang[3,4], Qingsong Yao[5], Xiaoqian Zhou[1,2], and S. Kevin Zhou[1,2✉]
[1]School of Biomedical Engineering, University of Science and Technology of China (USTC)
[2]Suzhou Institute for Advanced Research, USTC
[3]School of Computer Science and Technology, USTC
[4]School of Information Science and Technology, Eastern Institute of Technology (EIT)
[5]Stanford University

## Introduction

Landmark detection in medical images is essential for surgical planning, disease diagnosis, and treatment evaluation.

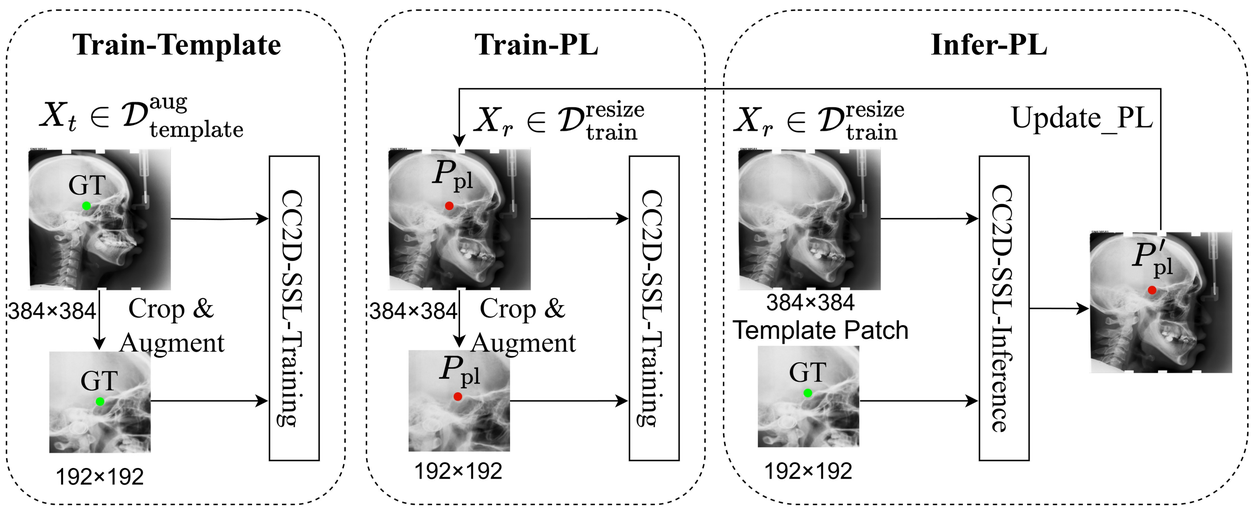
Although deep learning–based landmark detection under the fully supervised learning paradigm has shown excellent performance, it relies heavily on large-scale annotated datasets, which are often difficult to acquire in medical applications.

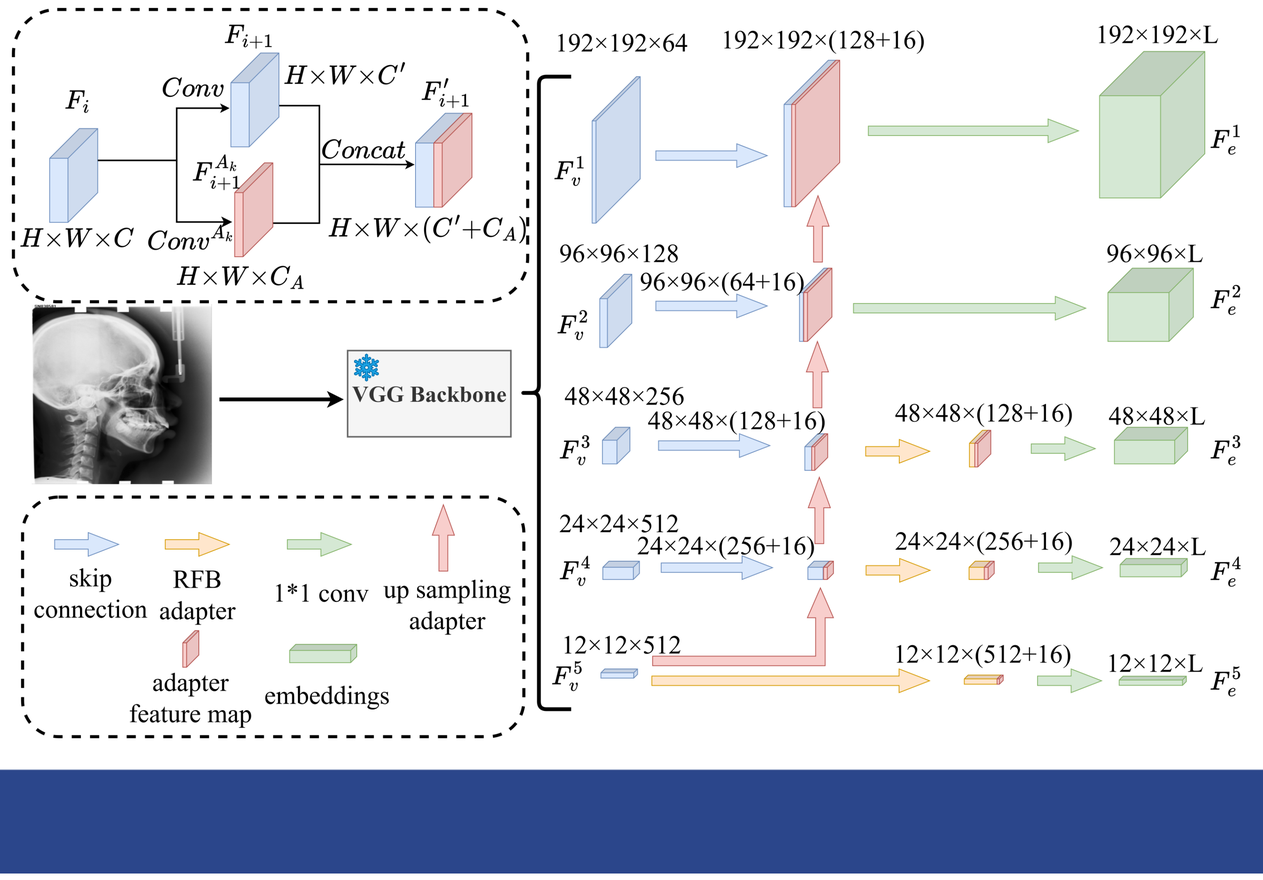
To alleviate this limitation, recent studies have explored one-shot landmark detection. However, most existing methods train multiple landmarks jointly, which often leads to a seesaw phenomenon, where the improvement of some landmarks degrades the performance of others.
In this work, we address these challenges by:

- Independently training each landmark model, thereby improving detection accuracy.
- Employing template augmentation, which further enhances robustness and precision.
- Introducing an adapter mechanism that combines shared parameters with landmark-specific parameters, enabling an end-to-end model. This design reduces both computation and memory overhead while maintaining high-accuracy landmark detection.

## Method

- CC2D-SLA: Train-PL + Infer-PL
- C-ATD: Train-Template + CC2D-SLA
- C-Adapter: Using adapter in C-ATD
- C-F2: C-Adapter + FM-OSD(fine stage)




## Results

- C-ATD achieves the best accuracy, obtaining **72.02%** SDR@2mm and 1.79mm MRE, significantly outperforming all previous state-of-the-art (SOTA) one-shot MLD methods.
- C-Adapter compresses 19 single-landmark models into one unified model, with only a slight performance drop compared to C-ATD.
- By combining C-Adapter with FM-OSD's fine stage (C-F2), performance improves to 70.48% SDR@2mm, exceeding the previous SOTA under high-resolution inference.

Table 1: Performance comparison of different methods on the Head [20] dataset.

| Method | Model Count | MRE(↓) (mm) | SDR(↑)(%) | | | |
|---|---|---|---|---|---|---|
| | | | 2mm | 2.5mm | 3mm | 4mm |
| SAM [22] | 1 | 2.56 | 54.11 | 63.66 | 70.25 | 80.84 |
| UOD [29] | 1 | 2.43 | 51.14 | 62.37 | 74.40 | 86.49 |
| CC2D [24] | 1 | 2.04 | 62.46 | 71.62 | 80.00 | 89.45 |
| FM-OSD(coarse) [13] | 1 | 1.93 | 63.60 | 75.43 | 83.03 | 91.94 |
| FM-OSD(fine) [13] | 2 | 1.82 | 67.35 | 77.92 | 84.59 | 91.92 |
| CC2D-SLA(ours) | 19 | 1.82 | 69.73 | 76.69 | 84.04 | **92.17** |
| C-ATD(ours) | 19 | **1.79** | **72.02** | **78.02** | **84.72** | 92.00 |
| C-Adapter(ours) | 1 | 1.96 | 67.83 | 75.33 | 81.89 | 90.82 |
| C-F2(ours) | 3 | 1.83 | 70.48 | 77.35 | 83.96 | 91.43 |

Table 2: Performance of different methods on the 19 landmarks.

| Landmark | CC2D | | FM-OSD | | C-ATD | | C-F2 | |
|---|---|---|---|---|---|---|---|---|
| | MRE | SDR(2mm) | MRE | SDR(2mm) | MRE | SDR(2mm) | MRE | SDR(2mm) |
| 1 | 1.35 | 85.2 | 1.55 | 84.0 | **0.98** | **97.6** | 1.26 | 92.0 |
| 2 | 1.60 | 74.0 | **1.49** | 73.2 | 1.51 | **78.0** | 1.54 | 73.6 |
| 3 | 1.58 | 72.4 | 1.66 | 68.4 | **1.37** | **84.4** | 1.46 | 78.4 |
| 4 | 1.93 | 66.4 | 2.28 | 55.6 | **1.80** | **70.8** | 2.00 | 67.2 |
| 5 | 1.86 | 62.8 | 1.72 | 65.6 | **1.54** | **76.0** | 1.61 | 75.2 |
| 6 | 2.50 | **50.0** | **2.41** | 48.4 | 2.47 | 49.2 | 2.43 | 48.4 |
| 7 | 1.42 | 81.6 | 1.05 | 88.0 | 0.96 | **94.0** | **0.94** | 93.6 |
| 8 | 1.46 | 78.4 | **0.99** | 91.6 | 1.10 | **93.2** | 1.93 | 89.6 |
| 9 | 1.08 | 88.8 | **0.83** | 94.8 | 0.84 | **96.8** | 0.84 | 95.2 |
| 10 | 4.19 | 20.4 | 3.23 | 33.6 | 3.59 | 24.8 | **3.17** | **37.6** |
| 11 | 2.58 | 42.8 | **2.44** | **52.0** | 2.85 | 48.8 | 2.66 | 48.8 |
| 12 | 2.60 | 48.8 | 1.91 | 68.0 | **1.43** | **86.4** | 1.50 | 80.8 |
| 13 | 1.63 | 70.0 | 1.60 | 68.4 | 1.56 | 78.4 | **1.52** | **79.2** |
| 14 | 1.69 | 71.2 | **1.43** | 80.0 | 1.54 | 79.6 | 1.47 | **80.0** |
| 15 | 1.73 | 65.6 | **1.69** | **68.8** | 2.50 | 50.0 | 1.89 | 59.6 |
| 16 | 3.54 | 26.4 | 2.61 | 47.2 | **2.52** | **51.2** | 2.56 | 48.8 |
| 17 | 1.73 | 67.6 | 1.55 | 74.8 | **1.24** | **84.0** | 1.51 | 77.6 |
| 18 | 1.77 | 65.6 | **1.67** | **67.6** | 2.00 | 63.2 | 1.82 | 66.4 |
| 19 | 2.44 | 48.8 | 2.49 | 49.6 | **2.25** | **62.0** | 2.75 | 47.2 |
| Mean | 2.04 | 62.5 | 1.82 | 67.4 | **1.79** | **72.0** | 1.83 | 70.5 |




## Conclusion

In this paper, we present a progression from exploiting each landmark's individuality—through single-landmark training—to utilizing inter-landmark similarity by incorporating adapters into a unified model. This two-fold approach demonstrates a feasible and effective strategy for improving MLD accuracy. The proposed C-Adapter represents an initial endeavor toward jointly learning multiple landmarks via shared and landmark-specific weights. We believe that continued exploration of this balance between individuality and similarity will yield more robust, efficient, and accurate solutions for one-shot MLD.