

CT-ScanGaze: A Dataset and Baselines for 3D Volumetric Scanpath Modeling

Supplementary Material

Ethical Statement

This research follows all relevant ethical guidelines for medical research. All patient data used in this study was properly anonymized and de-identified following HIPAA guidelines. The radiologists who participated in the eye-tracking study provided informed consent. No personal or identifying information is included in the dataset or results. Our study aims to augment, not replace, clinical expertise and maintains the central role of human medical professionals in diagnostic decisions.

Summary

The appendix is organized as follows:

- Appendix A describes additional dataset statistics including gaze data split, number of slices distribution, and duration of data collection recording videos.
- Appendix B presents additional visualizations of our synthetic training data.
- Appendix C provides additional implementation details.
- Appendix D describes the 3D scanpath similarity metrics.
- Appendix E describes implementation details of baseline methods adapted for 3D scanpath prediction.
- Appendix F provides additional qualitative results and visualizations.
- Appendix G compares different CT Visual Encoder backbones.
- Appendix H discusses the MultiMatch simplification algorithm and analysis.

A. Additional Dataset Details

A.1. Gaze Data Splits

Our dataset consists of 909 CT-gaze pairs, split into training, validation, and test sets with a ratio of 70:10:20 respectively. This translates to:

- Training set: 636 pairs
- Validation set: 90 pairs
- Test set: 183 pairs

A.2. Radiological Finding Distribution

CT-ScanGaze has a total of 9,332 findings with 60 unique finding names. The distribution of our dataset is shown in Fig. I.

A.3. Number of CT Slices

Fig. II shows the distribution of slice counts for all CT volumes. While all CT slices have a fixed resolution of 512×512 pixels in the axial plane, the number of slices

varies across volumes with a total number of slices is 131,618 and 186 slices per CT in average.

A.4. Recording Duration

In our dataset, we prioritize maintaining natural workflow by allowing radiologists to read CT scans following their standard clinical practice. Table I summarizes the duration statistics of our video recordings.

Table I. Recording Duration Statistics.

	Duration (minutes)
Total recording time	4722
Average session time	5.36
Minimum session time	1.27
Maximum session time	9.68

A.5. Additional Visualization of CT-ScanGaze

We visualize a CT volume with its eye tracking data from an alternative point of view, showing all scanpaths across slices in Fig. III and temporal navigation patterns in a line chart in Fig. IV. To create Figs. III and IV, we select a CT volume such that its fixation sequence has only 48 unique slices (48 unique z values) to maintain the simplicity of the visualization. An animated video is provided, named `vis_gt.mp4`. The observed scanpath pattern demonstrates a natural progression from peripheral regions inward to areas of diagnostic significance. The timestamped report is:

```
[00:00.000 --> 00:29.960] there's a left
upper chest pacemaker or ICD
[00:29.960 --> 00:40.240] with leads in the
right atrium right ventricle and a
pericardial lead along
[00:40.240 --> 00:43.240] the left
ventricle
[00:43.240 --> 00:57.960] there are no
enlarged axillary or supraclavicular
lymph nodes
[00:57.960 --> 01:19.320] mildly enlarged
paratracheal lymph nodes are present
there's a mildly enlarged
[01:19.320 --> 01:28.320] large lymph node
in the anterior mediastinum the left
ventricle appears
[01:28.320 --> 01:34.680] mildly dilated
with fatty metaplasia in the left
ventricular apex and
[01:34.680 --> 01:43.320] interventricular
septum there's no pericardial effusion
the great vessels
```

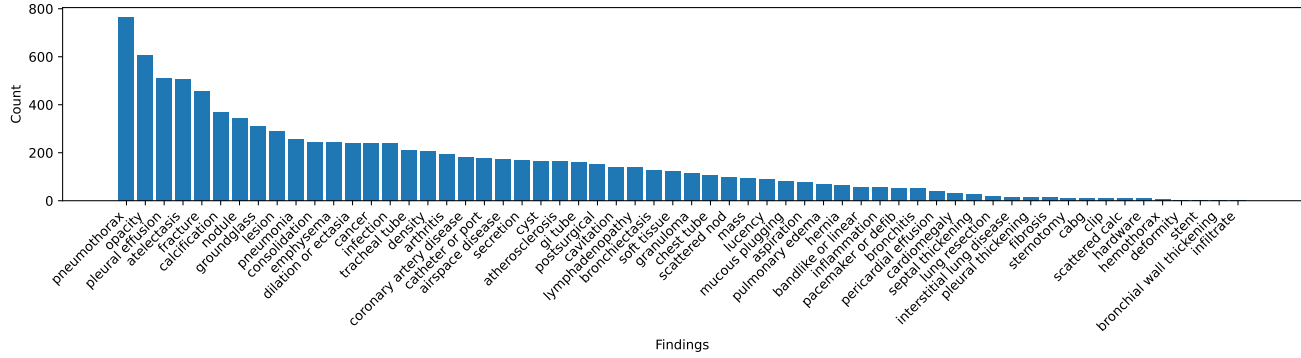


Figure I. Extracted radiological finding histogram. The y-axis represents the findings. The x-axis represents the occurrence frequency (number of samples). From SARLE [20], we extract a total of 9,332 findings with 60 unique finding names.

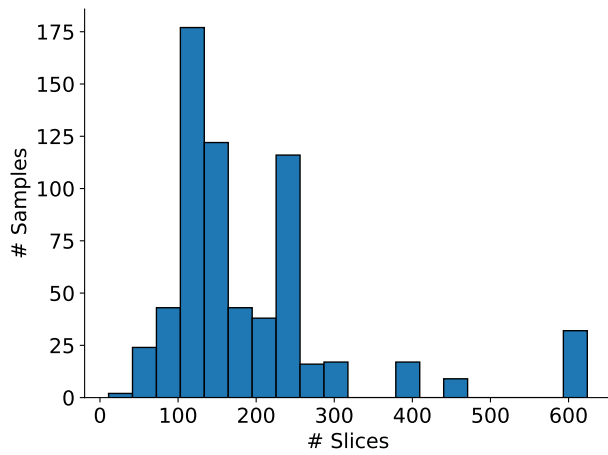


Figure II. Number of slices histogram. The y-axis represents the number of slices. The x-axis represents the occurrence frequency (number of samples). CT-ScanGaze has 131,618 slices in total and 186 slices per CT volume in average.

946 [01:43.320 --> 01:49.920] are normal in
 947 diameter there's mild aortic
 948 atherosclerotic calcification
 949 [01:49.920 --> 01:55.920] there is a stent
 950 in the LAD
 951 [02:01.720 --> 02:04.720] there's no
 952 pleural effusion
 953 [02:04.720 --> 02:19.840] there is
 954 cholelithiasis a low density nodule is
 955 present in the left adrenal
 956 [02:19.840 --> 02:25.800] gland likely
 957 representing an adenoma there's a
 958 calcified granuloma in the
 959 [02:25.800 --> 02:28.800] spleen
 960 [02:34.720 --> 02:37.720] there's no
 961 pericardial effusion

[02:37.720 --> 02:40.720] there's no
 pericardial effusion
 [02:40.720 --> 03:06.200] a small right
 pneumothorax is present
 [03:10.720 --> 03:17.720] the trachea and
 central airways are clear
 [03:17.720 --> 03:31.720] a calcified
 granuloma is present in the right upper
 lobe
 [03:31.720 --> 03:38.720] a small area of
 ground glass opacity is present in the
 right lower lobe
 [03:38.720 --> 03:50.720] there's a right
 lower lobe nodule measuring
 approximately 15 millimeters
 [04:08.720 --> 04:24.440] impression number
 one there's a small right pneumothorax
 number two a small
 [04:24.440 --> 04:27.840] area of ground
 glass opacity in the peripheral right
 lower lobe is likely
 [04:27.840 --> 04:37.440] infectious
 inflammatory there may be a cavitory
 component which could be the
 [04:37.440 --> 04:43.240] cause of the
 right pneumothorax number three there's
 a solid right lower lobe
 [04:43.240 --> 04:47.960] nodule measuring
 15 millimeters the differential includes
 infection slash
 [04:47.960 --> 04:50.960] inflammation and
 malignancy
 [04:54.360 --> 04:57.360] number four
 [04:57.360 --> 05:00.360] mildly enlarged
 lymph nodes in the mediastinum are
 nonspecific

By removing the timestamps (e.g.,
 [00:00.000 --> 00:29.960]), we obtain a
 free-text radiology report: *There's a left upper chest
 pacemaker or ICD with leads in the right atrium right*

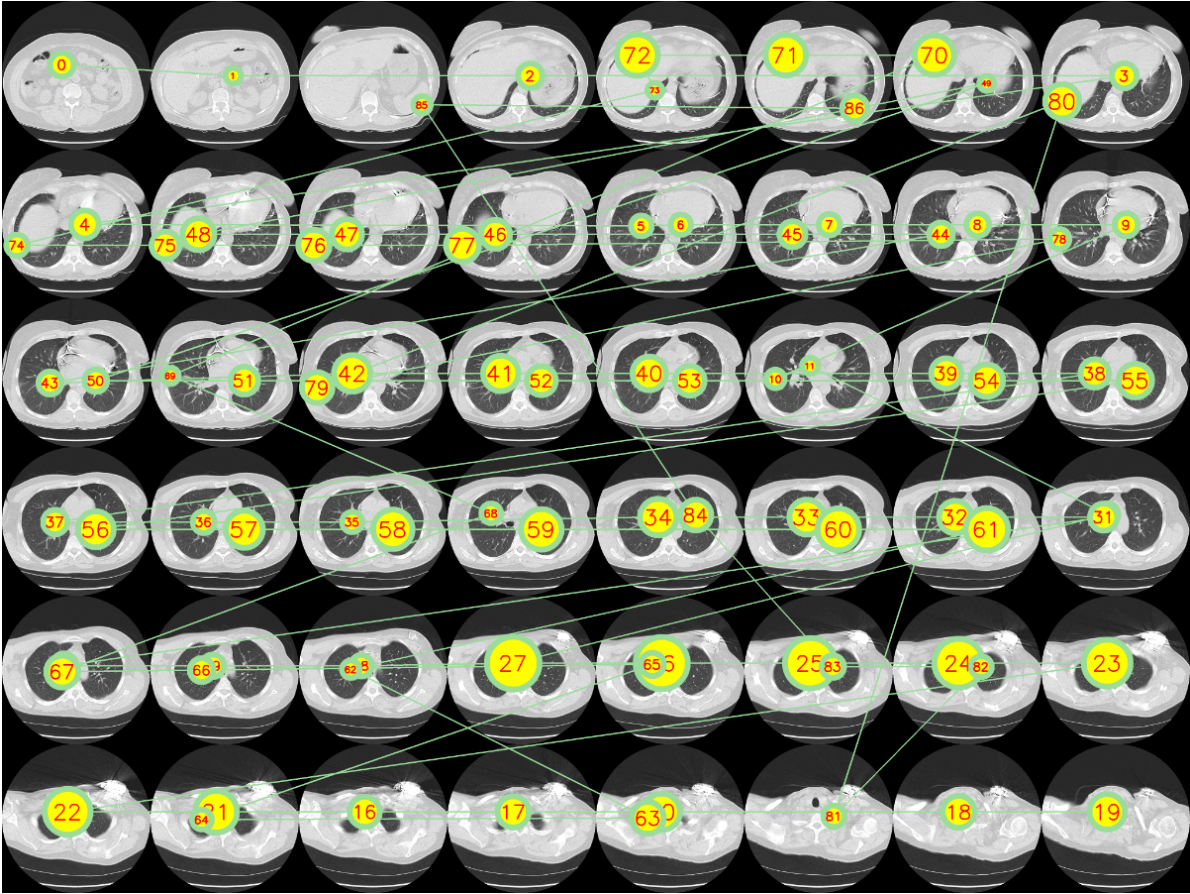


Figure III. Illustration of fixation sequence across sequential CT slices, following a left-to-right and top-to-bottom order. The fixation start from 0 in the top left corner. The numbered annotations indicate the order of fixation points. In this figure, we observe the radiologist’s systematic viewing pattern: first scanning through all slices before returning to central regions for detailed examination. This navigation pattern is also demonstrated in Fig. IV. We suggest the readers to watch `vis_gt.mp4` to see the animated version of this figure.

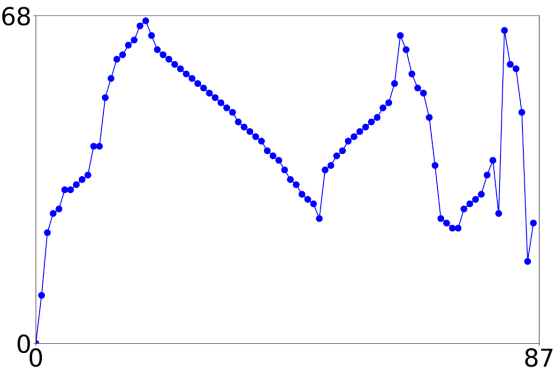


Figure IV. Visualization of temporal navigation patterns from the gaze data in Fig. III. The scanpath reveals that the radiologist follows a systematic approach: first traversing from start to end slices, then returning to central regions for detailed examination.

ventricle and a pericardial lead along the left ventricle. There are no enlarged axillary or supraclavicular lymph nodes. Mildly enlarged paratracheal lymph nodes are present. There’s a mildly enlarged large lymph node in the anterior mediastinum. The left ventricle appears mildly dilated, with fatty metaplasia in the left ventricular apex and interventricular septum. There’s no pericardial effusion the great vessels are normal in diameter there’s mild aortic atherosclerotic calcification. There is a stent in the LAD there’s no pleural effusion, there is cholelithiasis. A low density nodule is present in the left adrenal gland, likely representing an adenoma. There’s a calcified granuloma in the spleen. There’s no pericardial effusion. There’s no pericardial effusion. A small right pneumothorax is present. The trachea and central airways are clear. A calcified granuloma is present in the right upper lobe. A small area of ground glass opacity is present in the right lower lobe. There’s a right lower lobe nodule measuring approximately 15 millimeters. IMPRESSIONS. Number

1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020

one. There’s a small right pneumothorax. Number two. A small area of ground glass opacity in the peripheral right lower lobe is likely infectious inflammatory. There may be a cavitory component, which could be the cause of the right pneumothorax. Number three. There’s a solid right lower lobe nodule measuring 15 millimeters, the differential includes infection slash inflammation and malignancy. Number four. Mildly enlarged lymph nodes in the mediastinum are nonspecific.

Using CheXbert [60] to extract the 13 CheXpert findings [33], we identify the following positive findings: ‘Enlarged Cardiomeastinum’, ‘Lung Lesion’, ‘Pleural Effusion’, and ‘Support Devices’.

A.6. Example of CT-ScanGaze

We also provide one example of our data in example_data.zip.

- ct_id9.nii.gz is the CT scan.
- finding_id9.csv is the finding annotations.
- fixation_id9.json is the original fixations (without being simplified).
- recorded_video_id9.mp4 is the recorded video session.
- report_id9.txt is the report created by speech to text software.

B. Additional Visualization of Synthetic Data

Fig. V illustrates samples from our synthetic dataset, comprising temporal slice navigation patterns and fixation heatmaps. Overall, the synthetic data exhibits similarities with real eye movement for 3D, particularly in temporal characteristics when transitioning slices along the depth dimension.

C. Additional Implementation Details

Due to GPU memory constraint, directly using a full CT volume as input to extract complete feature maps and train end-to-end is often not feasible. In our implementation, we follow the common practice of sliding windows and merging windows. CT-Searcher uses Swin UNETR [29] encoder with $96 \times 96 \times 96$ input window. The output shape is reduced to $3 \times 3 \times 3$ with feature dimension $C = 768$. With a down-sampling ratio $r = 32$, we merge all features into a single feature map with shape $(W', H', D') = (W/r, H/r, D/r)$. For technical convenience, we interpolate all feature maps to a standardized size of $16 \times 16 \times 16$ with $C = 768$ channels, ensuring uniformity across varying feature map shapes. This standardization is reasonable since our CT scans have fixed dimensions of 512×512 in height and width (with only depth D varying), and the feature size of $16 \times 16 \times 16 \times 768$ keeps the model within our GPU memory (48 GB VRAM).

D. 3D Scanpath Metrics

Due to the complexity of 3D scanpath metrics, we describe them at a high level and point out the modifications from 2D to 3D. For detailed implementation, we encourage readers to examine the source code directly:

- visual_attention_metrics.py: Contains implementations of:
 - Saliency metrics Linear Correlation Coefficient (CC), Normalized Scanpath Saliency (NSS), Kullback-Leibler divergence (KLDiv).
 - String-edit-distance (SED) metric.
- scanmatch3d.py: Contains the 3D-adapted version of ScanMatch.
- multimatch_3dgaze.py: Contains the 3D-adapted version of MultiMatch.

D.1. Saliency Metrics

All three saliency metrics, CC (Correlation Coefficient), NSS (Normalized Scanpath Saliency), and KLDiv (Kullback-Leibler Divergence), are based on heatmaps and can be used directly without modification. The CC metric is defined as:

$$\begin{aligned}\hat{S} &= \frac{S - \mu_S}{\sigma_S} \\ \hat{G} &= \frac{G - \mu_G}{\sigma_G} \\ CC(S, G) &= \frac{\sum_{i,j,k} \hat{S}_{ijk} \hat{G}_{ijk}}{\sqrt{\sum_{i,j,k} \hat{S}_{ijk}^2 \sum_{i,j,k} \hat{G}_{ijk}^2}}\end{aligned}\quad (11)$$

where $S \in [0, 1]^{H \times W \times D}$ is the saliency map, $G \in \{0, 1\}^{H \times W \times D}$ is the ground truth fixation map, and μ and σ are mean and standard deviation. Given the fixation sequence $\{(x_l, y_l, z_l)\}_{l=1}^N$, where N is the fixation length, the ground truth map is defined as:

$$G_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) \in \{(x_l, y_l, z_l)\}_{l=1}^N \\ 0 & \text{otherwise} \end{cases}\quad (12)$$

Higher CC scores indicate better matching between sequences, with an upper bound of 1.0.

The NSS metric is defined as:

$$\begin{aligned}\tilde{S} &= \begin{cases} \frac{S}{\max(S)} & \text{if } \max(S) \neq 0 \\ S & \text{otherwise} \end{cases} \\ \hat{\tilde{S}} &= \begin{cases} \frac{\tilde{S} - \mu_{\tilde{S}}}{\sigma_{\tilde{S}}} & \text{if } \sigma_{\tilde{S}} \neq 0 \\ \tilde{S} & \text{otherwise} \end{cases} \\ NSS(\hat{\tilde{S}}, F) &= \frac{1}{N} \sum_{i,j,k} \hat{\tilde{S}}_{ijk} G_{ijk}\end{aligned}\quad (13)$$

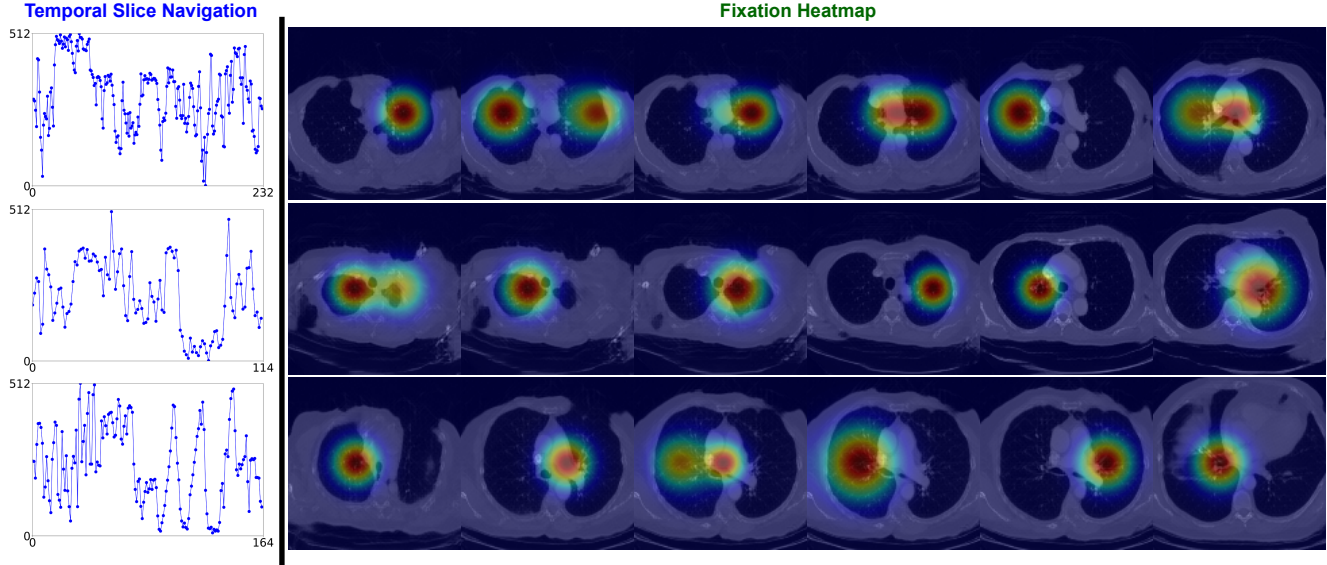


Figure V. Visualization of our synthetic training data. The Temporal Slice Navigation demonstrates the temporal transition of slice. The Fixation Heatmap column displays representative CT slices with corresponding fixation heatmaps.

where \hat{S} is the normalized saliency map. Higher NSS scores indicate better matching between sequences, with an upper bound that depends on the ground truth.

The KLDiv metric is defined as:

$$\begin{aligned}\tilde{S}_{ijk} &= \frac{S_{ijk}}{\sum_{i,j,k} S_{ijk}} \\ \tilde{G}_{ijk} &= \frac{G_{ijk}}{\sum_{i,j,k} G_{ijk}} \\ KLDiv(S, G) &= \sum_{i,j,k} \tilde{G}_{ijk} \log \left(\epsilon + \frac{\tilde{G}_{ijk}}{\tilde{S}_{ijk} + \epsilon} \right)\end{aligned}\quad (14)$$

where $\epsilon = 2.2204 \times 10^{-16}$ is a small constant to prevent divide-by-zero and log-zero. Lower KLDiv scores indicate better matching between sequences, with a lower bound of 0.0.

D.2. String-edit-distance

String-edit-distance (SED) has two main steps:

1. Converts gaze sequences into strings:
 - a) Dividing the 3D volume into discrete cells. In the original 2D version, this step divides the image into patches.
 - b) Assigning unique characters to each cell.
 - c) Mapping fixation points to these characters in sequence.
2. Compares two sequences using Levenshtein distance by counting minimum number of operations (insertions, deletions, substitutions).

In our 3D adapted SED, we change the step 1.a from 2D into 3D, the other steps are left as is. Lower SED scores indicate better matching between sequences, with a lower bound of 0.0.

D.3. ScanMatch

Calculating ScanMatch (SM) score between predicted and ground truth fixations consists of 3 main steps:

1. Convert fixation sequences into letter strings. This step is similar to the first step of SED. In addition, when considering duration (ScanMatch w/ Dur.), each character is repeated n times, where n is the duration in milliseconds. This repetition is not performed when duration is not considered (ScanMatch w/o Dur.).
2. Create a substitution matrix with scores for all possible letter pairs. The original score function uses 2D Euclidean distance, which we extend to 3D Euclidean distance.
3. Sequence comparison:
 - a) Create comparison matrix:
 - Columns: letters from first sequence
 - Rows: letters from second sequence
 - Cell values: costs from substitution matrix
 - b) Apply Needleman-Wunsch algorithm to find optimal alignment path
 - c) Calculate normalized similarity score (0-1 scale)

We adapt to 3D by modifying both step 1 and the substitution matrix calculation at step 2, replacing 2D Euclidean distance with 3D Euclidean distance. Higher SM scores indicate better matching between sequences, with an upper bound of 1.0.

D.4. MultiMatch

Different from ScanMatch and SED, MultiMatch (MM) measures scanpath similarity regarding shape, direction, length, position, and duration. Higher MM scores indicate better sequence matching, with an upper bound of 1.0 for all aspects. Given the predicted fixations $\{(\hat{x}_l, \hat{y}_l, \hat{z}_l, \hat{t}_l)\}_{l=1}^N$ and ground truth fixations $\{(x_l, y_l, z_l, t_l)\}_{l=1}^N$, we calculate MultiMatch scores with 3 main steps:

1. Temporal alignment:
 - a) Calculate how similar each element i in one scanpath is compared to each element j in the other scanpath based on a similarity metric. Collect all pairs (i, j) to create a similarity matrix $M(i, j)$ between elements.
 - b) From $M(i, j)$, build adjacency matrix A with connection weights like a graph.
 - c) Find the shortest path from i to j using Dijkstra's algorithm.
 - d) Align scanpaths along shortest path.
2. For every align pair of fixation (i, j) , we compute similarity across five dimensions:
 - a) **Vector (shape):** shape difference between fixation vectors $(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{t}_i) - (x_j, y_j, z_j, t_j)$.
 - b) **Direction:** difference in direction (angle) between fixation vectors. We measure angles using spherical coordinates in 3D, analogous to the original authors' use of polar coordinates in 2D.
 - c) **Length:** difference in amplitude (length) between fixation vectors $|\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{t}_i - (x_j, y_j, z_j, t_j)|$.
 - d) **Position:** 3D Euclidean distance between fixations.
 - e) **Duration:** difference in duration between fixations.
3. Score normalization:
 - a) Vector, Length, and Position scores are normalized by volume diagonal.
 - b) Direction is normalized by π .
 - c) Duration is normalized by maximum duration.

E. 3D Scanpath Prediction Baselines

E.1. PathGAN

Similar to original PathGAN [4], our CT-adapted PathGAN architecture has two major components: a Discriminator D and a Generator G . The Generator takes a CT volume V as input and produces a fixation sequence $G(V) = \{(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{t}_i)\}_{i=1}^N$. The Discriminator aims to assign low scores to N predicted fixations $(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{t}_i)_{i=1}^N$ and high scores to N_{gt} ground truth fixations $gt = \{(x_i, y_i, z_i, t_i)\}_{i=1}^{N_{gt}}$, taking both the fixation sequence and CT volume features as input. The PathGAN architecture is illustrated in Fig. VI. Note that we share a frozen Swin UNETR module between G and D during training while optimizing other modules.

For the loss functions, we maintain PathGAN's default implementation using two main components: conditional

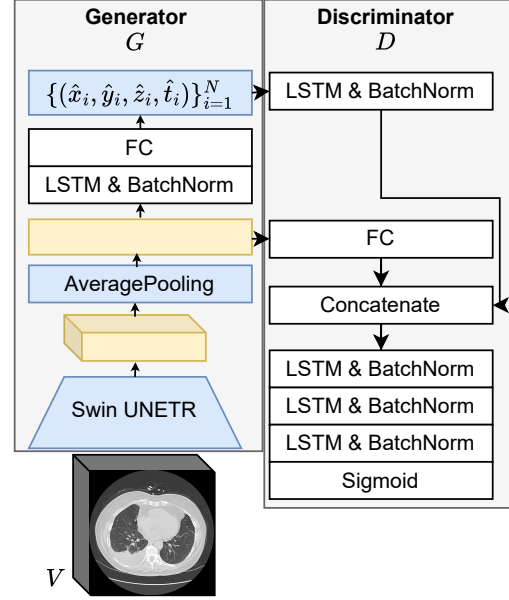


Figure VI. Our adapted version of PathGAN for CT scanpath prediction maintains the core architecture while altering components: the visual encoder module to Swin UNETR, Average Pooling to 3D, and predicted fixations (highlighted in blue).

GAN loss and L^2 loss between ground truth and predicted fixations. Specifically, the conditional GAN loss is defined as:

$$\mathcal{L}_{cGAN} = \mathbb{E}_{V, gt} [\log D(V, gt)] + \mathbb{E}_V [\log(1 - D(V, G(V)))], \quad (15)$$

The L2 loss is defined as:

$$\mathcal{L}_{L^2} = \mathbb{E}_{V, gt} [\|gt - G(V)\|^2] \quad (16)$$

The final formulation of the loss function for the generator during adversarial training is:

$$\mathcal{L} = \mathcal{L}_{cGAN} + \alpha \mathcal{L}_{L^2} \quad (17)$$

Following the original PathGAN implementation, we set the hyperparameter $\alpha = 0.05$.

E.2. HAT

The detailed adapted HAT [77] architecture is shown in Fig. VII. Similar to HAT's FPN visual encoder [44] that generates features at both bottleneck and high-resolution levels, we extract features from two Swin UNETR layers: $P1 \in \mathbb{R}^{H/32 \times W/32 \times D/32 \times C}$ from the bottleneck layer as low resolution feature and $P4 \in \mathbb{R}^{H/4 \times W/4 \times D/4 \times C_4}$ from the 4th layer of Swin UNETR's decoder as high resolution feature. We extend HAT's original loss to handle 3D fixation maps and use a single query as the class query. Note that Fig. VII shows one-step prediction because HAT predicts fixations step by step, with the working memory being updated after each fixation heatmap prediction [77].

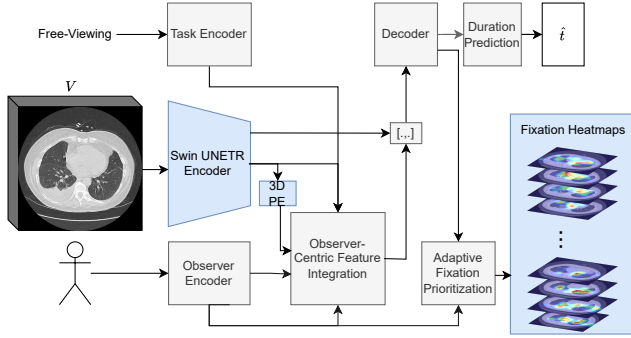


Figure IX. Our adapted version of GazeformerISP for CT scanpath prediction maintains most of the original architecture while modifying three key components (highlighted in blue): the input features, 3D Position Encoding, and prediction output. Here, [...] denotes the concatenation operator, and ‘3D PE’ represents 3D Positional Encoding. Similar to HAT’s freeview mode, we employ a single query embedding to predict fixations autoregressively.

where N is the maximum length of fixations, Y_t and t_i represent the ground-truth fixation maps and fixation duration, respectively. Finally, we train GazeformerISP with Self-Critical Sequence Training (SCST) set up using ScanMatch as reward, and the Consistency Divergence loss as originally described in [12, 13].

F. Additional Qualitative Results

Fig. X presents an additional comparison of the temporal slice navigation and fixation heatmaps across multiple CT slices between CT-Searcher with state-of-the-art scanpath prediction methods. CT-Searcher outperforms others by capturing a balance between realism and variability, avoiding excessive noise or oversimplification in the temporal slice navigation comparison. Additionally, CT-Searcher achieves more visually faithful heatmaps compared to the ground-truth, outperforming other approaches in detail and accuracy. For baseline methods, we observe several limitations. Some methods produce no heatmaps (N/A) in certain positions, showing their inability to generate meaningful outputs. Similar to PathGAN, Gazeformer covers limited CT slices, indicating a constraint in handling 3D fixation tasks. Both GazeformerISP and HAT can produce heatmaps for most CT slices, however their temporal slice navigation appears noisy and inconsistent, deviating from the ground-truth pattern. In conclusion, our method outperforms other approaches and mimics ground truth scanpath in both temporal slice navigation and fixation heatmap generation.

G. Comparison of CT Backbone Architectures

Tab. II demonstrates the comparative performance of CT-ViT and Swin UNETR backbones as CT-Searcher’s Visual Encoder across multiple scanpath similarity metrics:

Table II. Ablation: Comparison of backbone architectures across scanpath metrics. Arrows (↑/↓) indicate whether higher or lower scores are better. Bold values indicate the best performance.

Visual Encoder	SM ↑	MM ↑	SED ↓	KLDiv ↓
CT-ViT	0.1287	0.6934	193	3.665
Swin UNETR	0.1318	0.7002	174	3.645

Table III. Fixation count comparison between the original gaze data and the simplified gaze data.

Version	Original	Simplified	Reduction (%)
Number of Fixations	2,234,920	954,311	57.3%

Table IV. MultiMatch similarity scores between the original gaze data and the simplified gaze data.

Dimension	Vector	Direction	Length	Position	Average
Score	0.993	0.853	0.989	0.944	0.945

ScanMatch (SM), MultiMatch (MM), String-Edit Distance (SED), and Kullback-Leibler Divergence (KLDiv). We freeze both backbones during training. Tab. II shows that Swin UNETR outperforms CT-ViT across all metrics, achieving higher scores in pattern-based measures (SM: 0.1318 vs. 0.1287, MM: 0.7002 vs. 0.6934) and lower values in distance-based metrics (SED: 174 vs. 193, KLDiv: 3.645 vs. 3.665). Based on the empirical results, we adopt Swin UNETR as our Visual Encoder.

H. MultiMatch Simplification Analysis

Due to the original gaze data containing scanpaths with numerous fixations that are dense and complex with sequences averaging 543 fixations and reaching up to 2,708 fixations per CT, we employ the simplification algorithm to make this data more manageable while preserving essential gaze patterns, from the MultiMatch toolbox [18] with default settings: an angular threshold of 45° and an amplitude threshold of 10% of the volume resolution diagonal. This reduces sequences to an average of 222 fixations with a maximum of 1,507 fixations. Both original and simplified versions will be made available.

To demonstrate the effectiveness of simplification, Figs. XI to XIII presents a comparative illustration between original and simplified scanpaths. While radiologists’ eye movements on a single CT slice generally focus on the image center, movement along the slice dimension exhibits more complexity. Fig. XI reveals continuous and intricate radiologist navigation through depth. Nevertheless, the MM simplification approach introduces only minor changes to the movement landscape while preserving the overall pattern. This consistency in pattern preservation is also evident along the x-axis (Fig. XII) and y-axis (Fig. XIII). In sum-

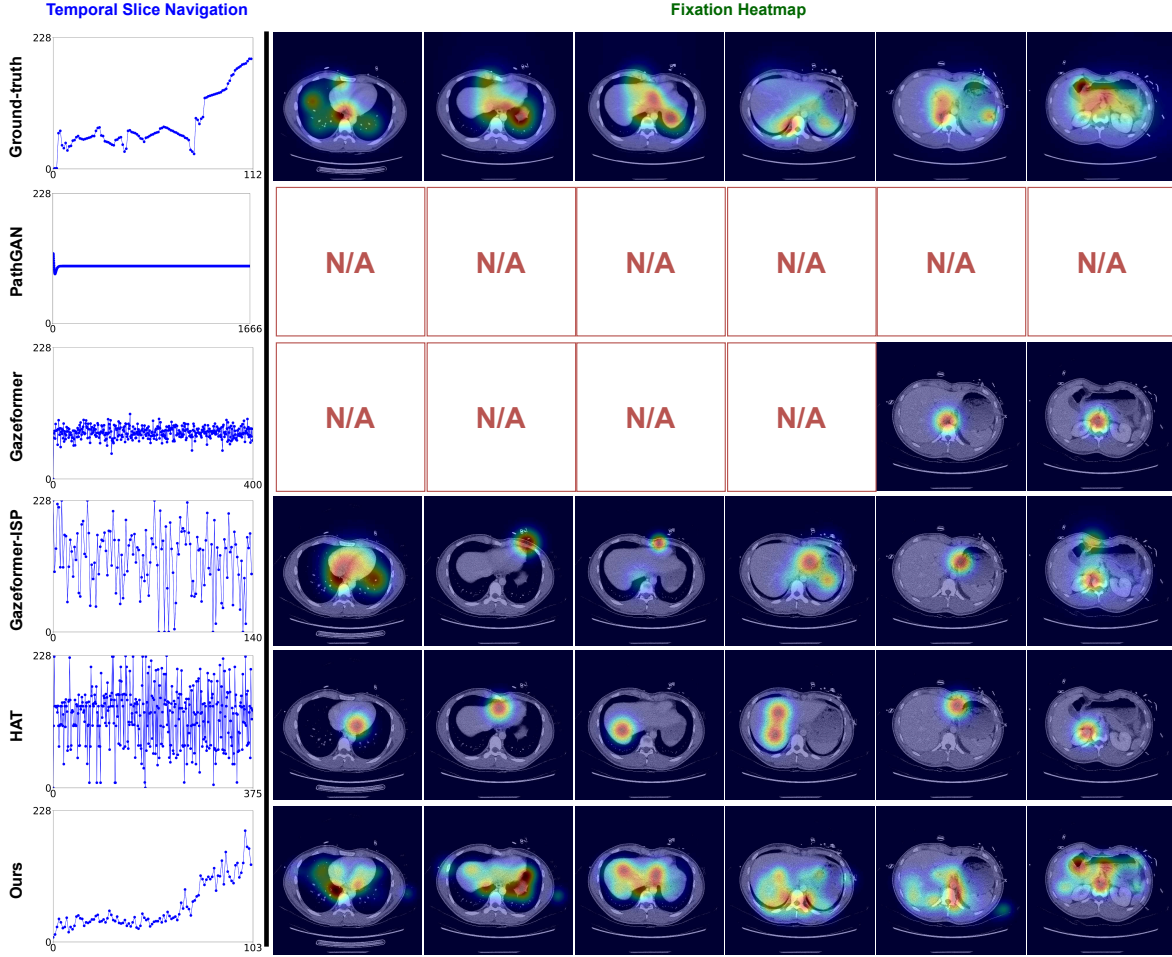


Figure X. Additional qualitative results between CT-Searcher and state-of-the-art scanpath prediction methods. N/A in a particular slice position (column) means that the corresponding model (row) fails to predict scanpath for that slice, thus no heatmap can be created. The heatmap images in the Fixation Heatmap column show the eye gaze fixation patterns across different CT slices. The left columns show Temporal Slice Navigation patterns, illustrating how scanpath traverses through different slices over time.

mary, the line charts demonstrate that while significantly reducing the number of points (Tab. III), the simplification process maintains the essential scanpath characteristics, as evidenced by minimal changes in MultiMatch similarity scores across all spatial dimensions (Tab. IV).

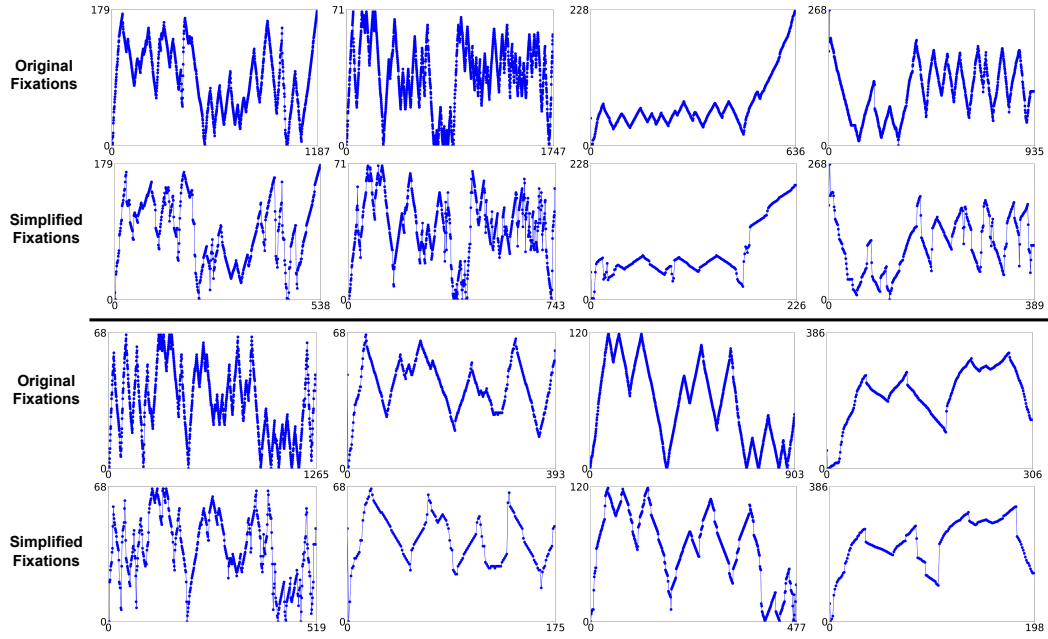


Figure XI. The effect of MM simplification algorithm on the z (slice) dimension.

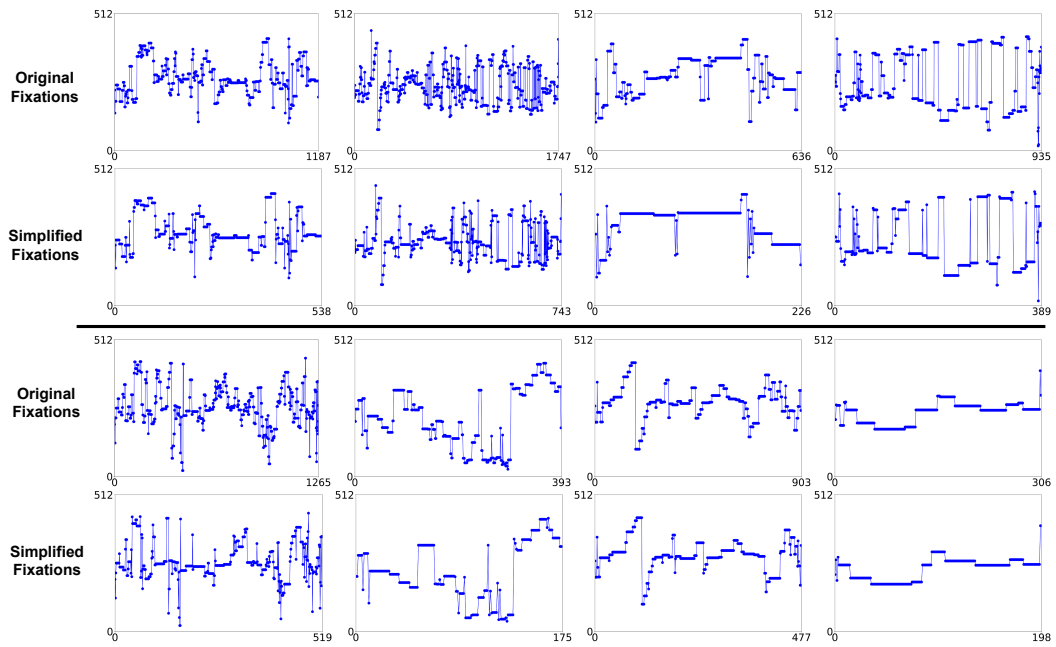


Figure XII. The effect of MM simplification algorithm on the x (width) dimension.

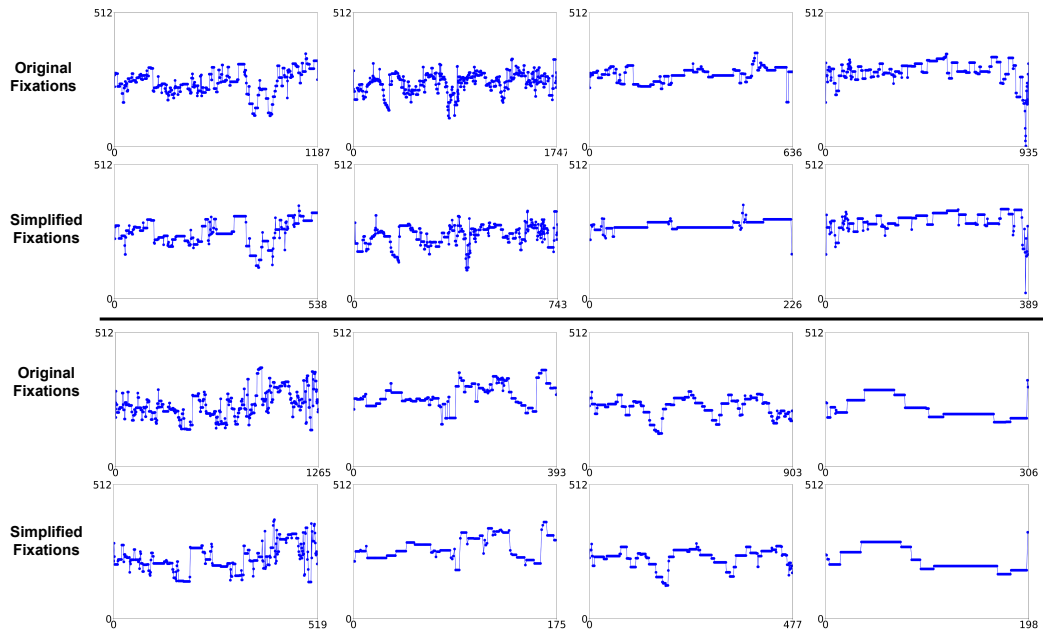


Figure XIII. The effect of MM simplification algorithm on the y (height) dimension.