ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images

Xiangjie Sui[1], Yuming Fang[1], Hanwei Zhu[2], Shiqi Wang[2], Zhou Wang[3]

[1]Jiangxi University of Finance and Economics, [2]City University of Hong Kong, [3]University of Waterloo

TUE-PM-273

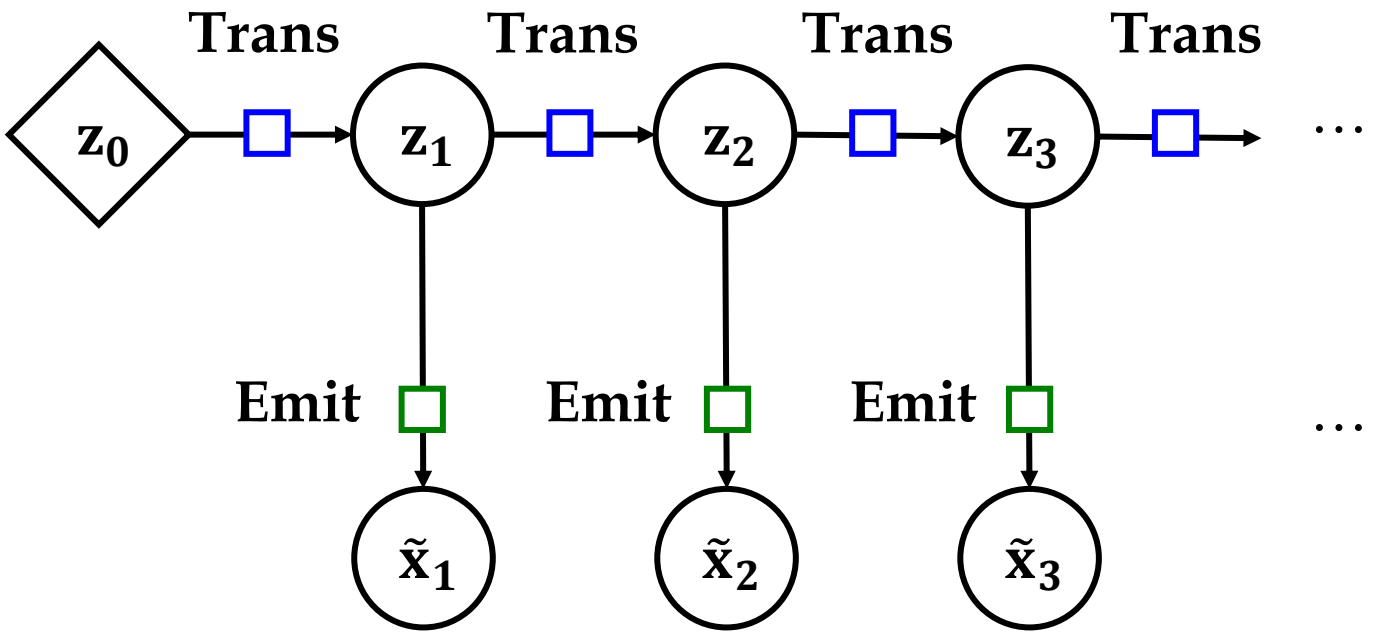
»» Quick Preview

Viewing behavior **Observation** **Scanpath**

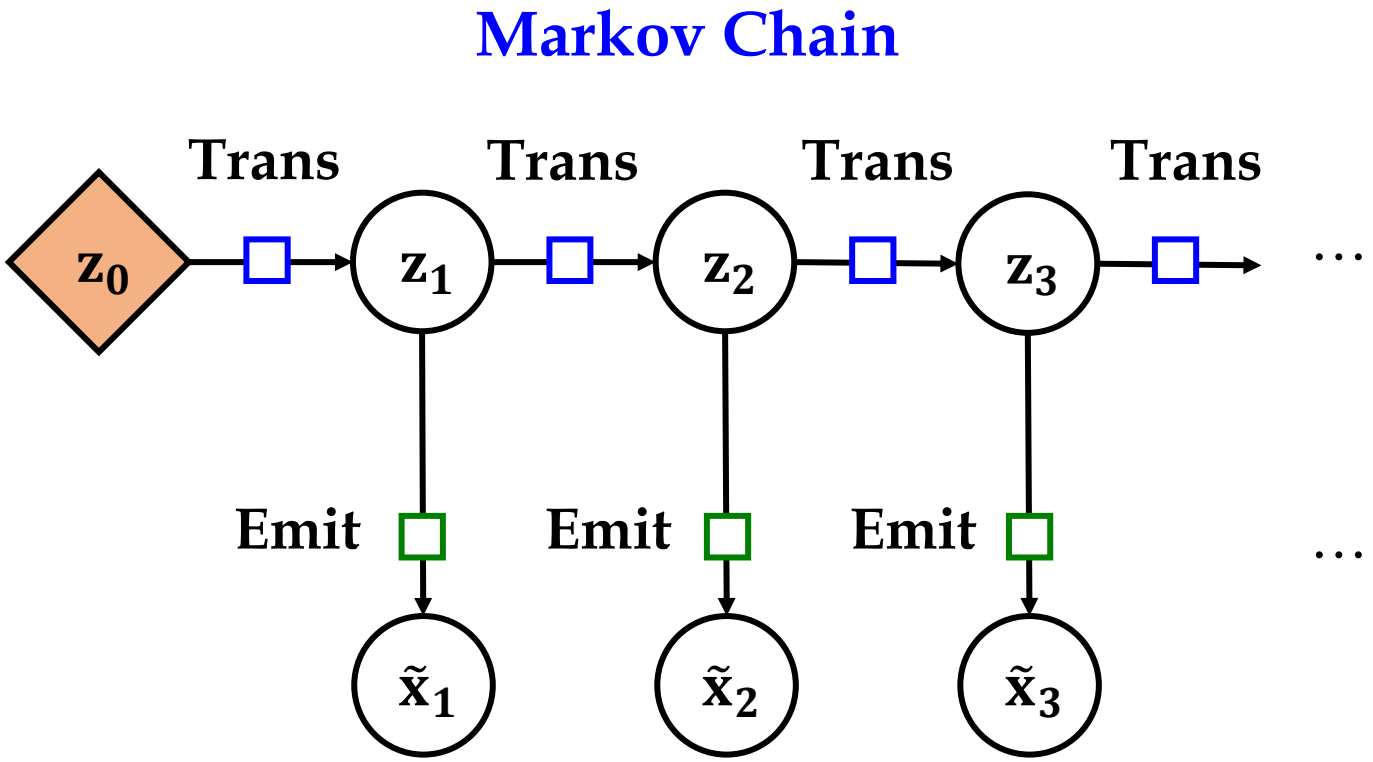
X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

$\mathbf{z_t}$: the visual state $\tilde{\mathbf{x}}_\mathbf{t}$: the produced coordinate of gaze point

Markov Chain



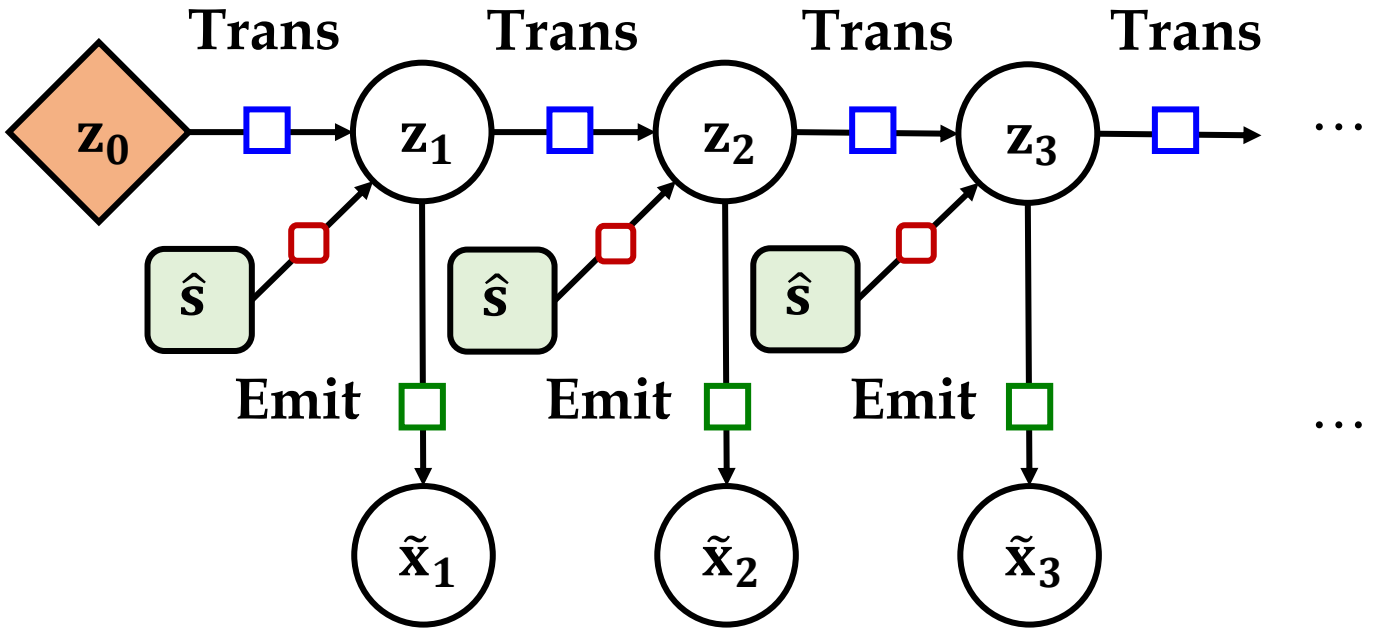
X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.



- **State initialization**
- ☑ Facilitate our model to focus on learning the dynamics of states with correct "launcher".
- ☑ Assign a specific starting point for scanpath generation.

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Markov Chain



- **Semantic-guided transition**
- ☑ Model the mechanism of visual working memory by maintaining and updating the visual state in the Markov chain.
- ☑ Model the interventions of scene semantics on visual working memory.

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Realistic

360° Image | Ground Truth | ScanDMM (Ours)

Room

Museum

Party

Park

Time

Efficient

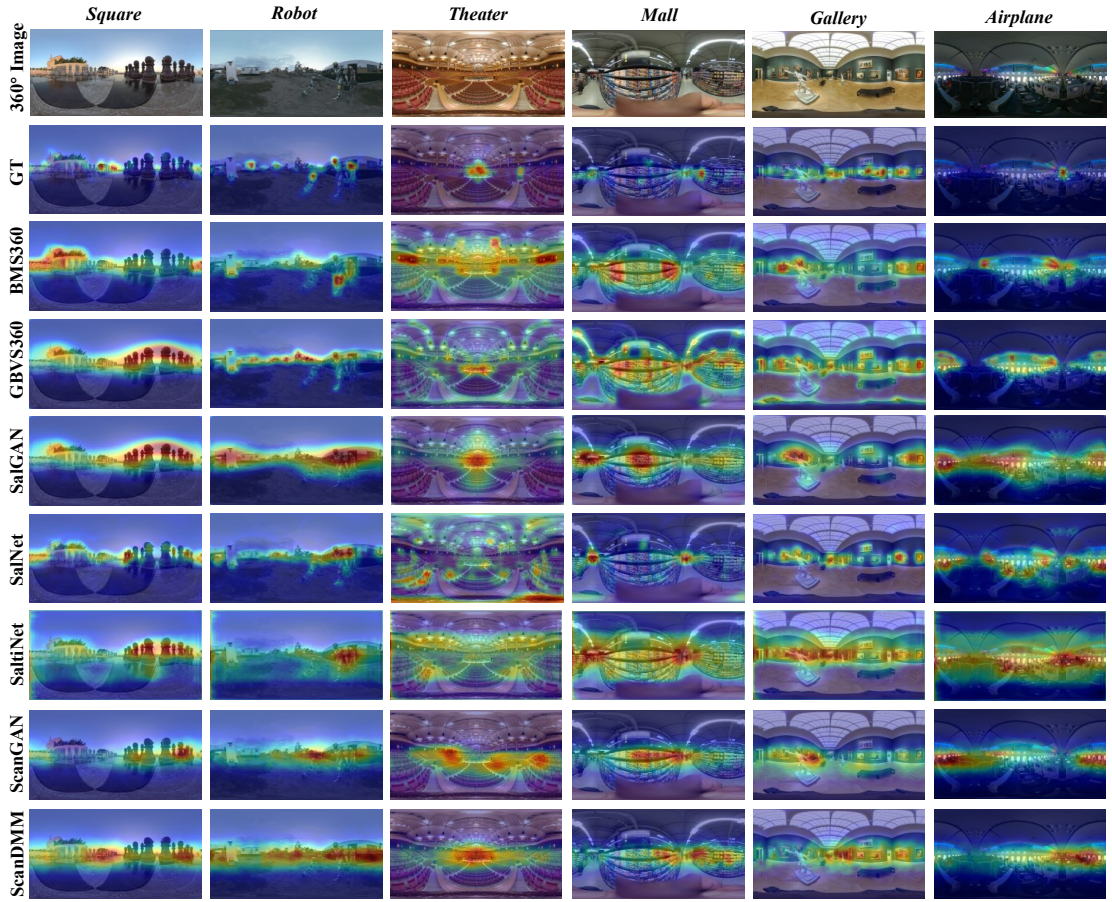
Producing 1,000 scanpaths within 1-second

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

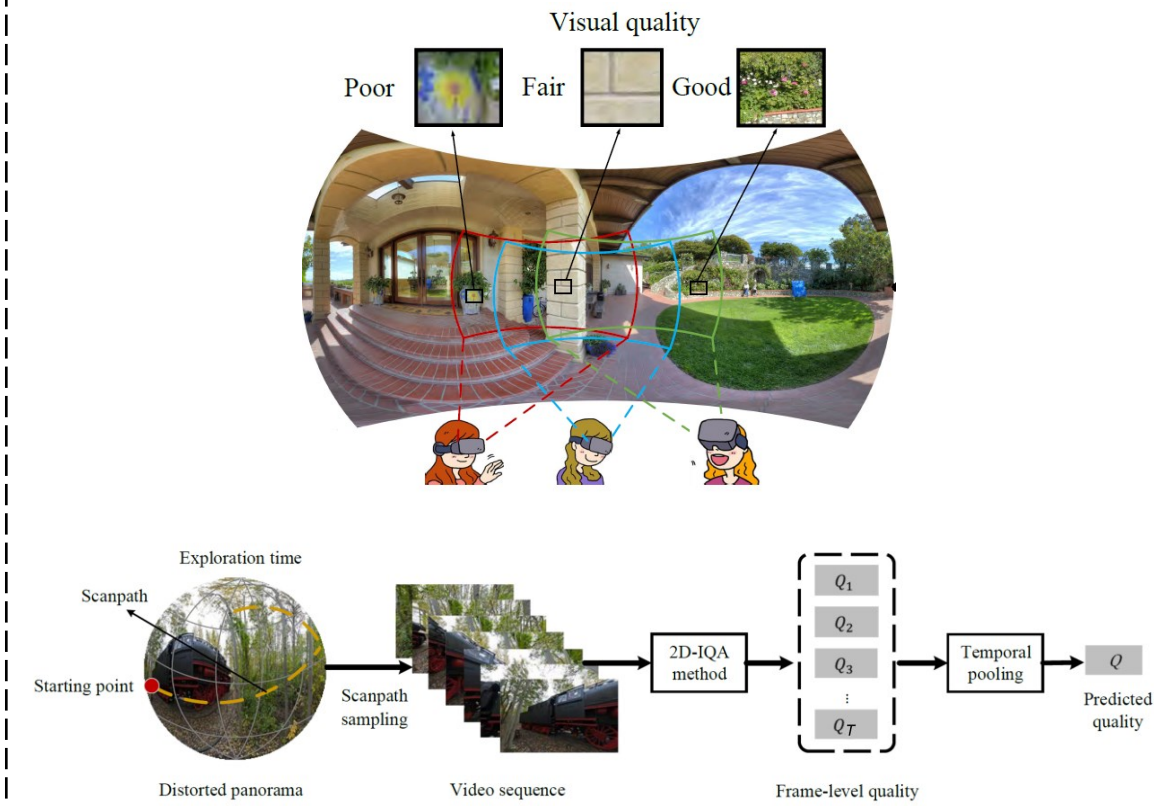
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Generalizable

Saliency Detection



Visual Quality Assessment



X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Background and Motivation

2D Scanpath

Source from: G. Boccignone, V. Cuculo, and A. D'Amelio. How to look next? A data-driven approach for scanpath prediction. In International Symposium on Formal Methods, pages 131–145. Springer, 2019.

Field of viewing

Full **Local**

Boundary

Disconnected

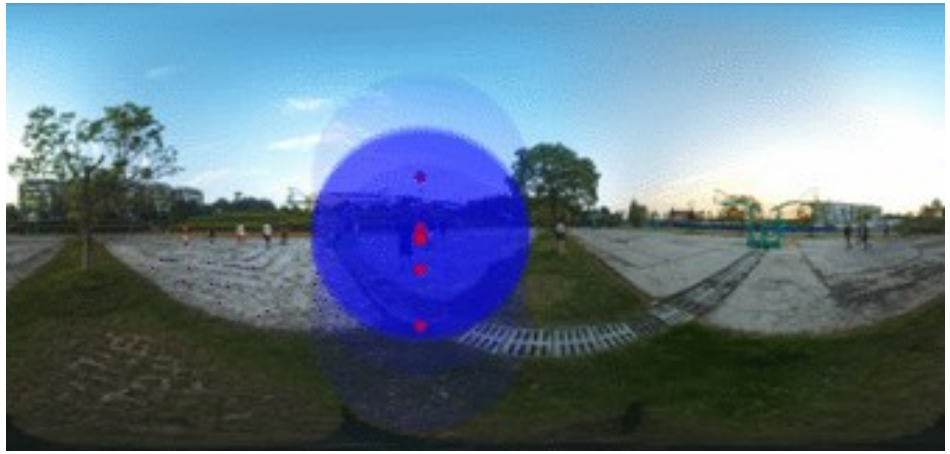
Continuous

360° Scanpath

Source from: G. Y. Rai, J. Guti´errez, and P. Le Callet. A dataset of head and eye movements for 360 degree images. In ACM on Multimedia Systems Conference, page 205–210. Association for Computing Machinery, 2017.

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Diverse Viewing Conditions



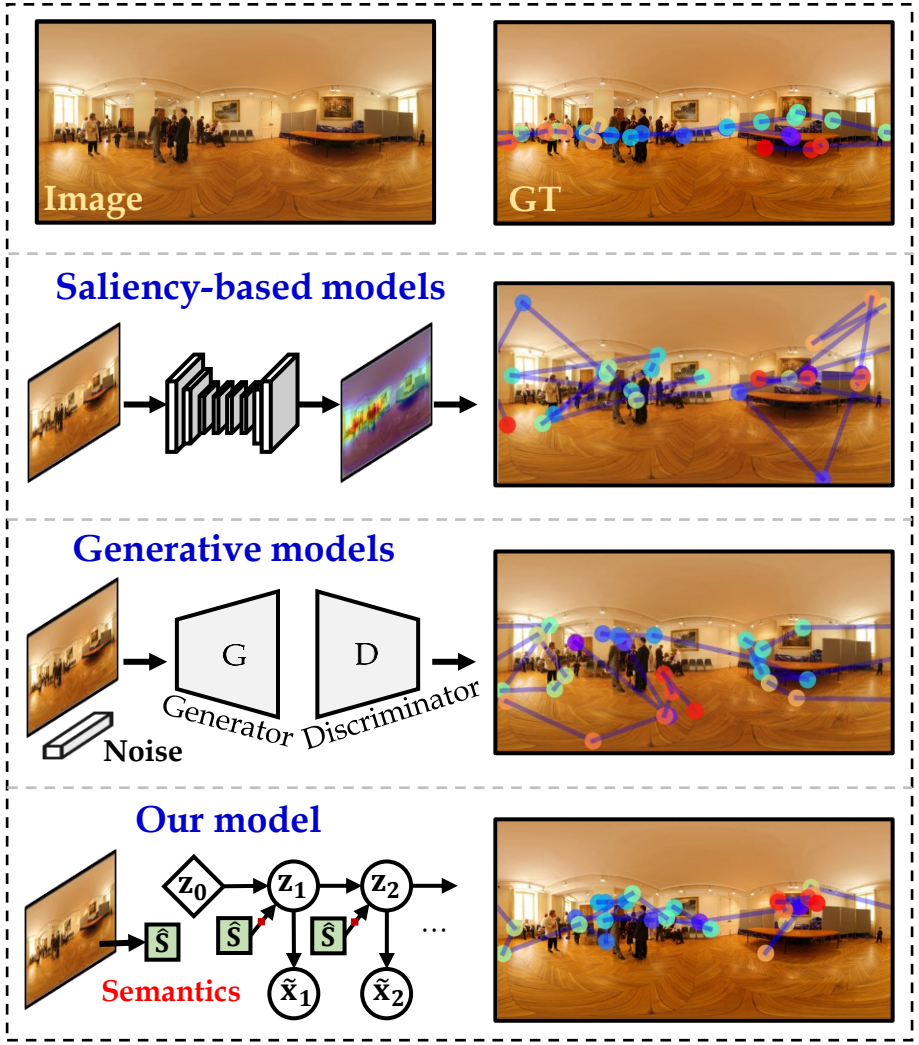
Starting Point 1

Starting Point 2

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Background and Motivation



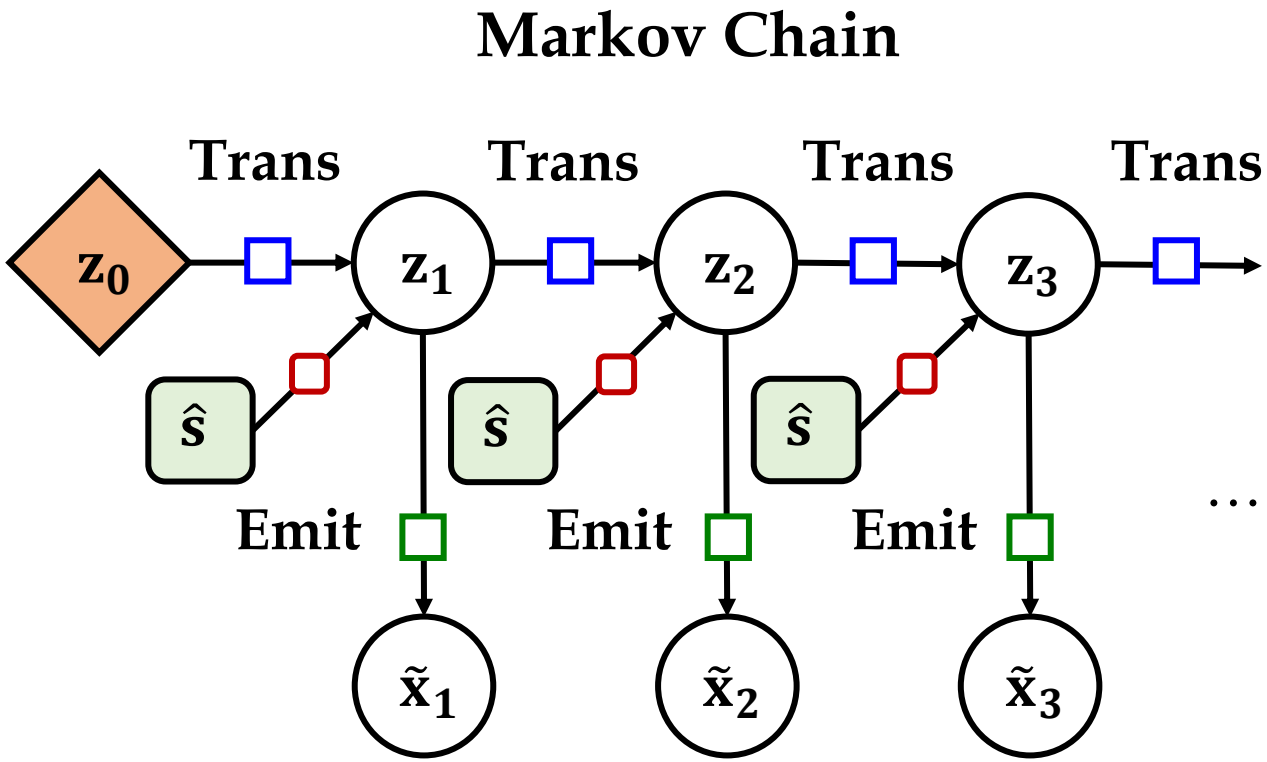
Unstable behavior with large displacements and scarce focal regions

Show less attention to regions of interest.

Time-dependent Visual Attention Modeling

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Method



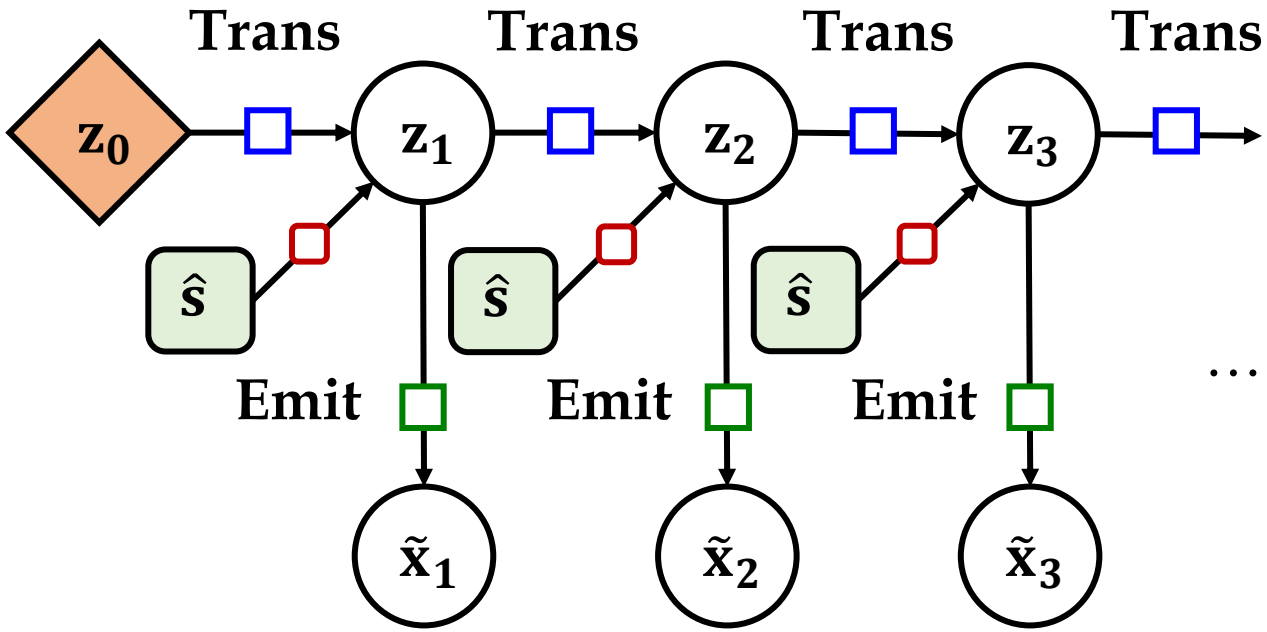
$\mathbf{z_t}$: the visual state
(n-dimensional Gaussian function)

$\tilde{\mathbf{x}}_\mathbf{t}$: the produced coordinate of gaze point
(3D coordinate – (x, y, z))

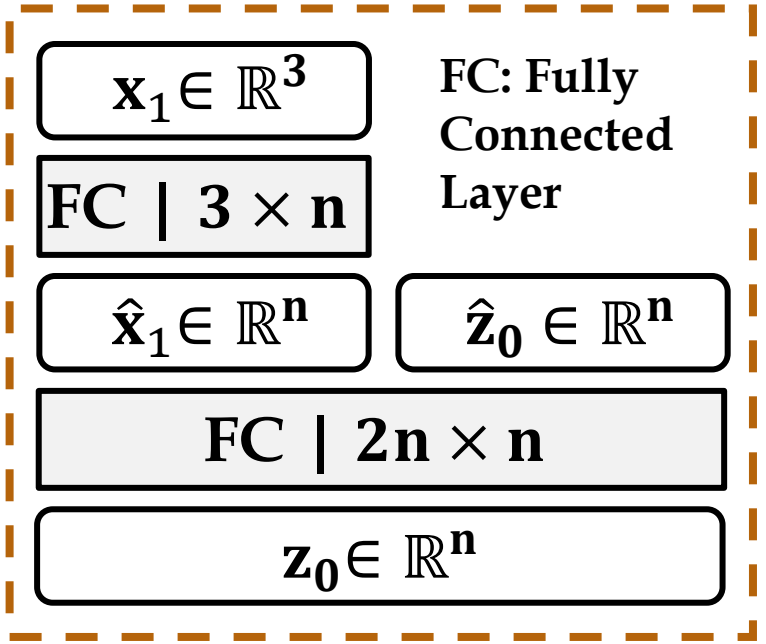
$\hat{\mathbf{s}}$: the encoded scene semantics

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Markov Chain

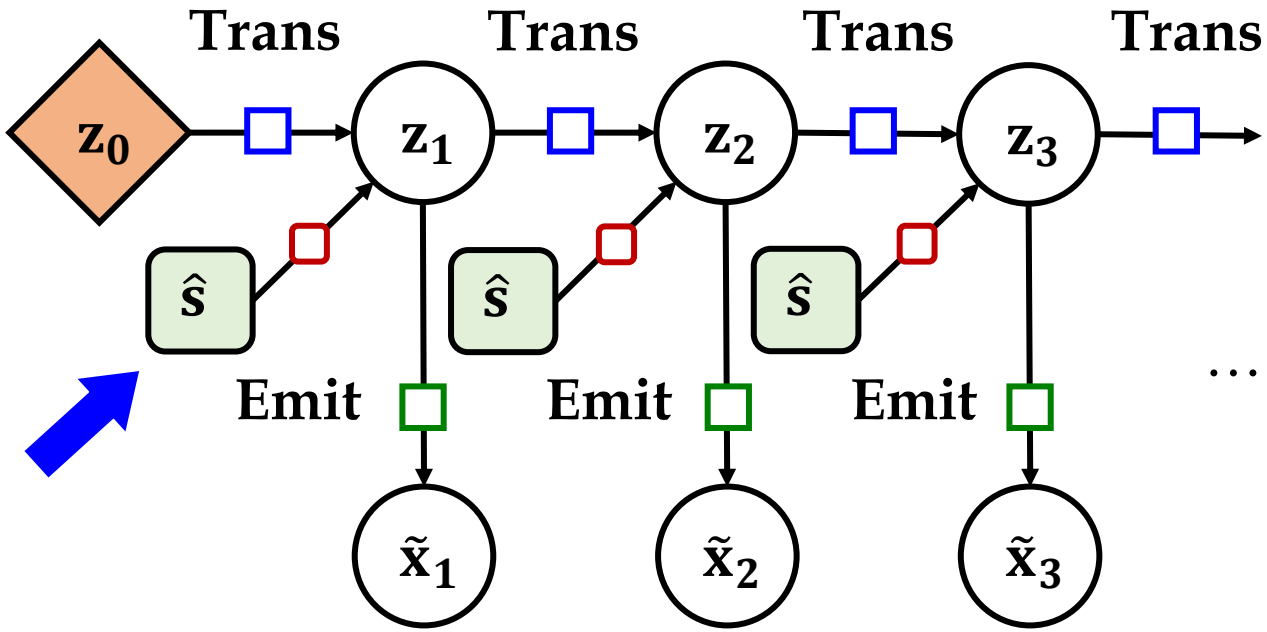


State Initialization

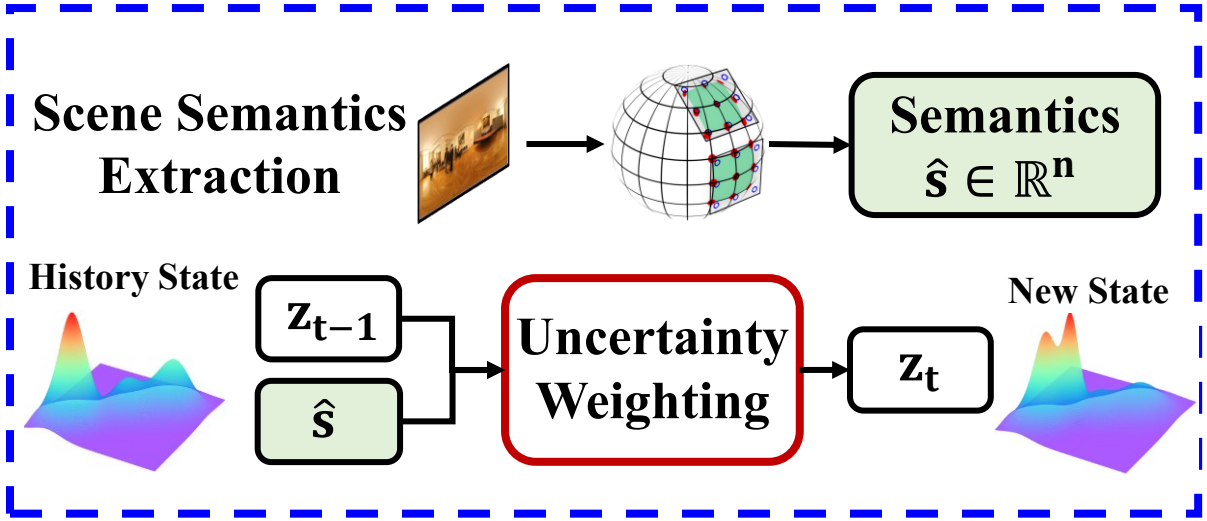


X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Markov Chain

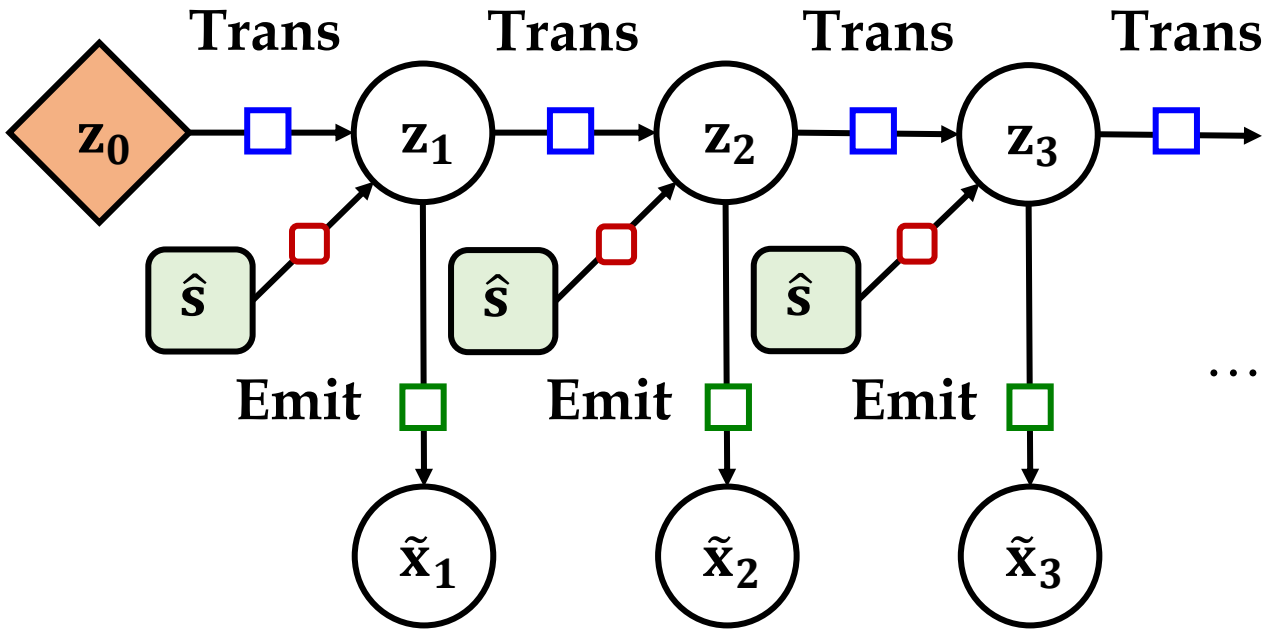


Transition Function

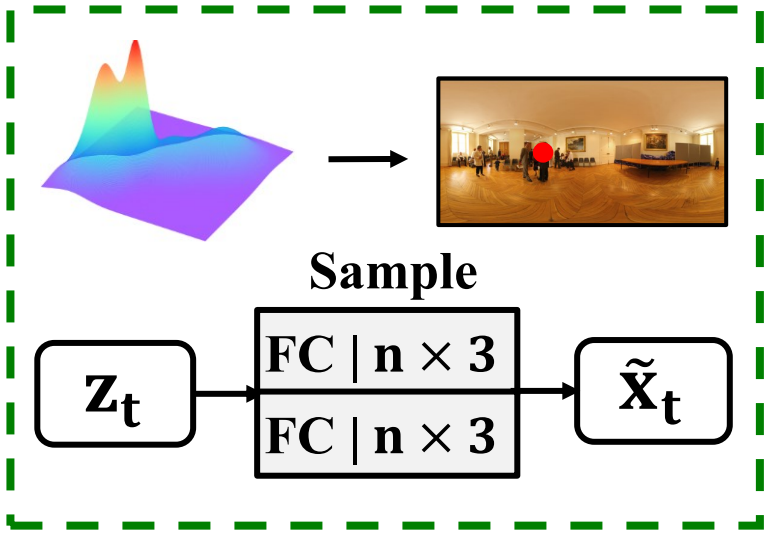


X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Markov Chain



Emission Function



X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

State Initialization

$x_1 \in \mathbb{R}^3$

FC | $3 \times n$

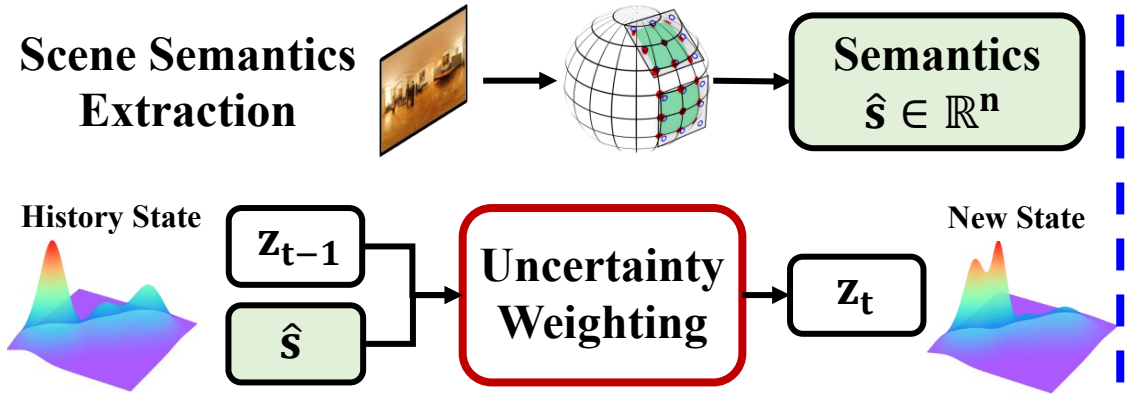
$\hat{x}_1 \in \mathbb{R}^n$ $\hat{z}_0 \in \mathbb{R}^n$

FC | $2n \times n$

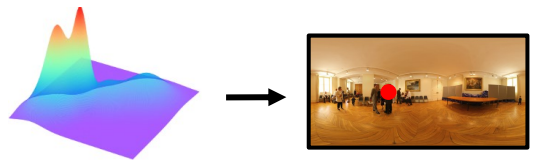
$z_0 \in \mathbb{R}^n$

FC: Fully Connected Layer

Transition Function

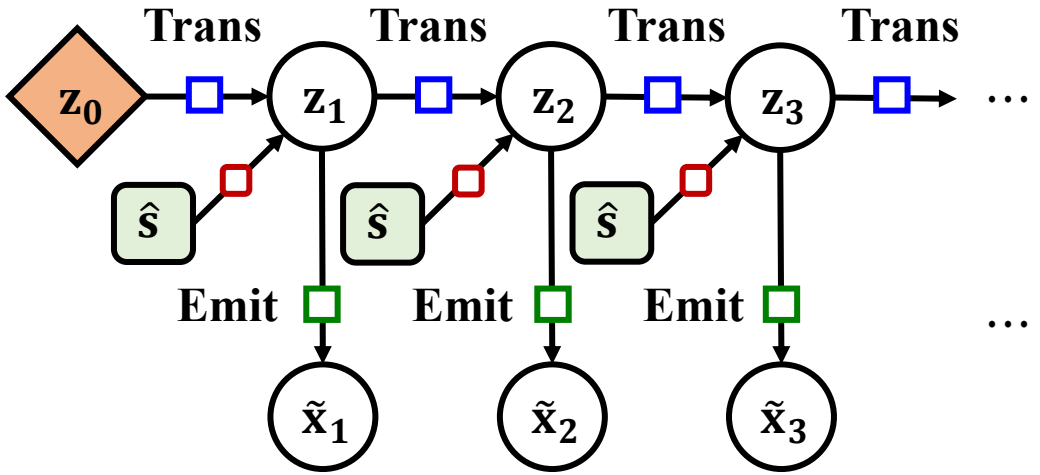


Emission Function



Sample

z_t → FC | $n \times 3$ → FC | $n \times 3$ → $\tilde{x}_t$

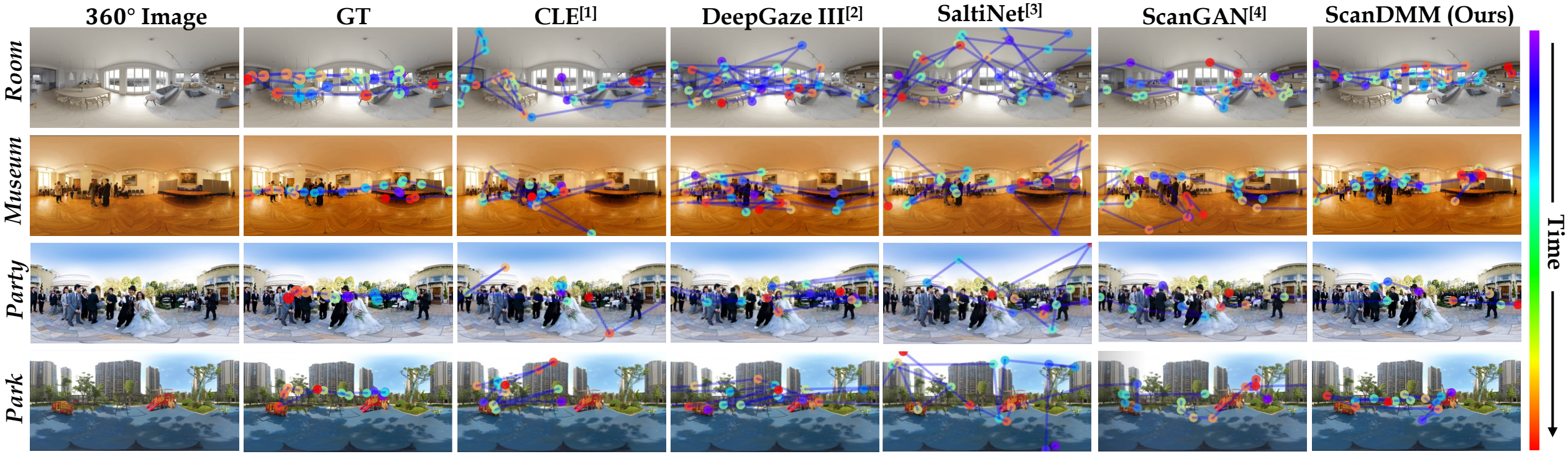


- Spherical characteristics
- Time-dependent attention modeling
- Flexible – arbitrary starting point of viewing and arbitrary exploration time.

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

»» Experiments

Experiments



1. G. Boccignone, V. Cuculo, and A. D'Amelio. How to look next? A data-driven approach for scanpath prediction. In International Symposium on Formal Methods, pages 131–145. Springer, 2019.
2. M. K¨ummerer, M. Bethge, and T. S. A. Wallis. DeepGaze III: Modeling free-viewing human scanpaths with deep learning. Journal of Vision, 22(5):7–7, 2022.
3. M. Assens, X. Giro-i-Nieto, K. McGuinness, and N. E. O'Connor. SaltiNet: Scan-path prediction on 360 degree images using saliency volumes. In IEEE International Conference on Computer Vision Workshops, pages 2331– 2338, 2017.
4. D. Martin, A. Serrano, A. W. Bergman, G. Wetzstein, and B. Masia. ScanGAN360: A generative model of realistic scanpaths for 360◦ images. IEEE Transactions on Visualization and Computer Graphics, 28(5):2003–2013, 2022.

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Table 1. Performance of scanpath prediction models on different databases. The best performance is highlighted.

Database	Method	LEV ↓	DTW ↓	REC ↑
Sitzmann[7]	*Random walk*	48.942	2232.987	2.669
	CLE	45.176	1967.286	3.130
	DeepGaze III	46.424	1992.859	3.082
	SaltiNet	51.370	2305.099	1.564
	ScanGAN	45.270	**1951.848**	3.241
	ScanDMM	**44.966**	1965.427	**3.475**
	Human	41.188	1836.986	6.345
AOI[9]	*Random walk*	13.696	711.516	2.993
	CLE	12.865	547.892	3.617
	DeepGaze III	13.155	558.445	2.892
	SaltiNet	14.695	596.544	2.244
	ScanGAN	12.889	552.446	3.750
	ScanDMM	**12.127**	**537.504**	**4.024**
	Human	9.243	389.477	6.228

Database	Method	LEV ↓	DTW ↓	REC ↑
Salient360! [8]	*Random walk*	40.802	2231.681	2.744
	CLE	39.774	1714.409	3.323
	DeepGaze III	40.006	1742.351	2.588
	SaltiNet	40.848	1855.477	2.305
	ScanGAN	38.932	1721.711	3.099
	ScanDMM	**37.272**	**1528.592**	**3.576**
	Human	35.084	1382.590	5.202
JUFE[10]	*Random walk*	24.039	1193.725	3.109
	CLE	24.844	1172.150	3.013
	DeepGaze III	24.129	1104.848	2.774
	SaltiNet	26.074	1287.144	1.540
	ScanGAN	24.209	1094.978	3.075
	ScanDMM	**23.091**	**1086.014**	**4.329**
	Human	18.306	1038.880	7.745

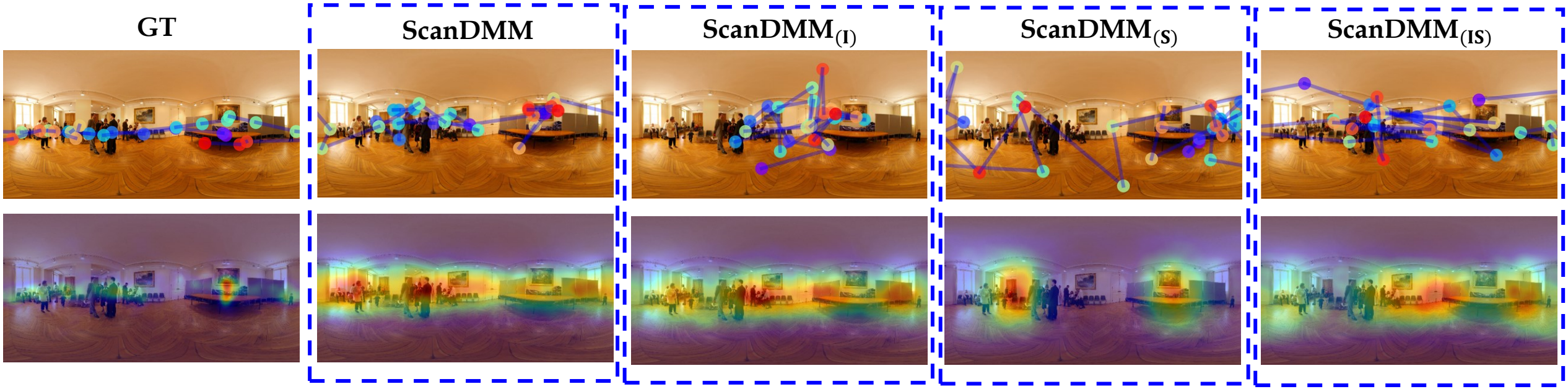
7. V. Sitzmann, A. Serrano, A.y Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. Saliency in VR: How do people explore virtual environments? IEEE Transactions on Visualization and Computer Graphics, 24(4):1633–1642, 2018.
8. Y. Rai, J. Guti´errez, and P. Le Callet. A dataset of head and eye movements for 360 degree images. In ACM on Multimedia Systems Conference, page 205–210. Association for Computing Machinery, 2017.
9. M. Xu, L. Yang, X. Tao, Y. Duan, and Z. Wang. Saliency prediction on omnidirectional image with generative adversarial imitation learning. IEEE Transactions on Image Processing, 30:2087–2102, 2021.
10. Y. Fang, L. Huang, J. Yan, X. Liu, and Y. Liu. Perceptual quality assessment of omnidirectional images. In AAAI Conference on Artificial Intelligence, volume 36, pages 580–588, 2022.

Table 2. Efficiency comparison in terms of the model size and running time (computed by the time cost of producing 1, 000 scanpaths).

Method	Parameters	Running Time
CLE	-	≈ 39 seconds
DeepGaze III	78.9MB	≈ 11 minutes
SaltiNet	103.6MB	≈ 49 minutes
ScanGAN	33.9MB	0.987 seconds
ScanDMM	**18.7MB**	**0.737 seconds**

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Ablation Study



ScanDMM$_{(I)}$: initializing the state using learnable parameter

ScanDMM$_{(S)}$: removing scene semantics

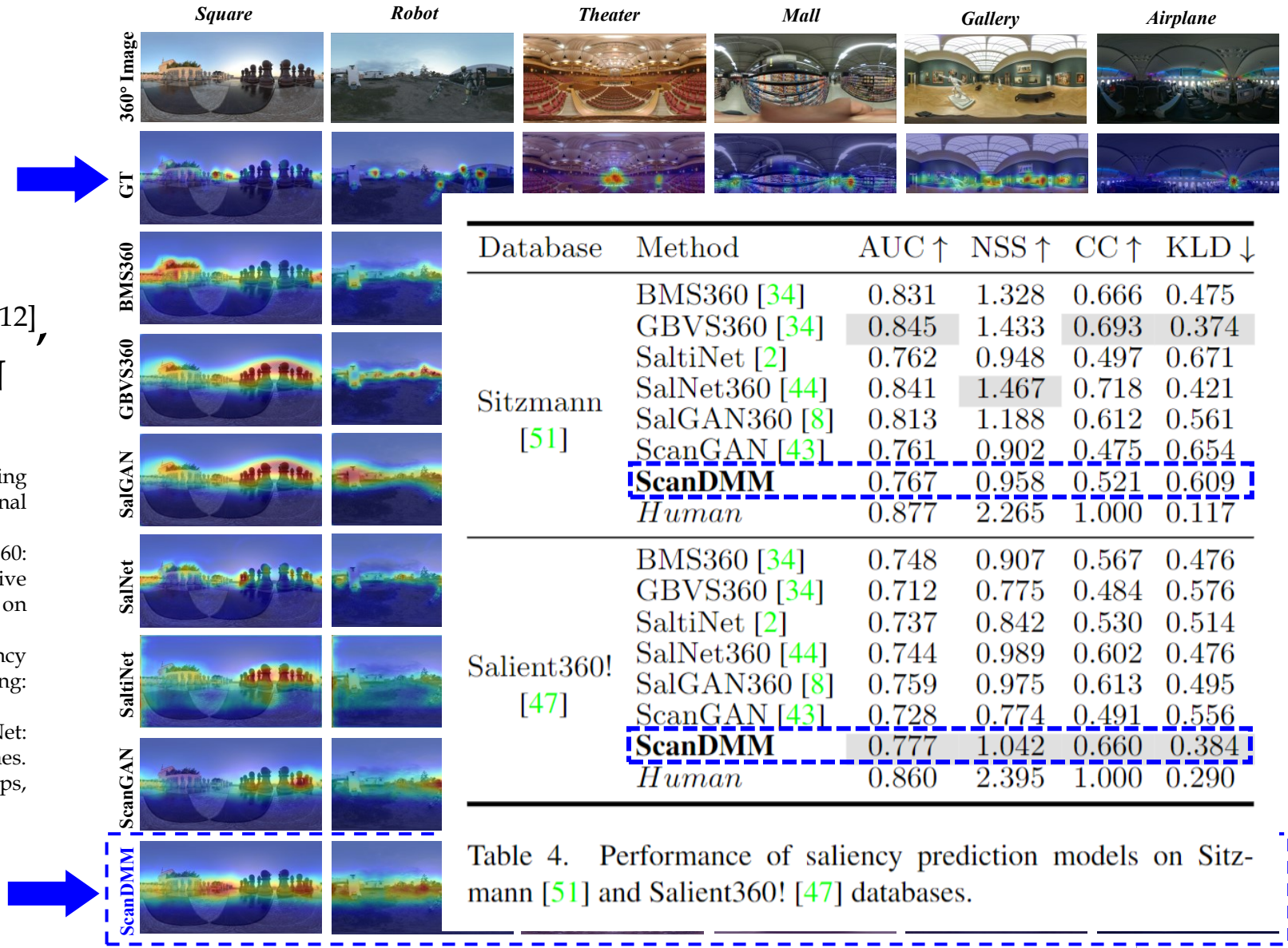
ScanDMM$_{(IS)}$: performing both of the two modifications

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Application

Saliency Detection

Competitive models:
BMS360[11], GBVS360[11], SalGAN[12], SalNet[13], SaltiNet[14], ScanGAN

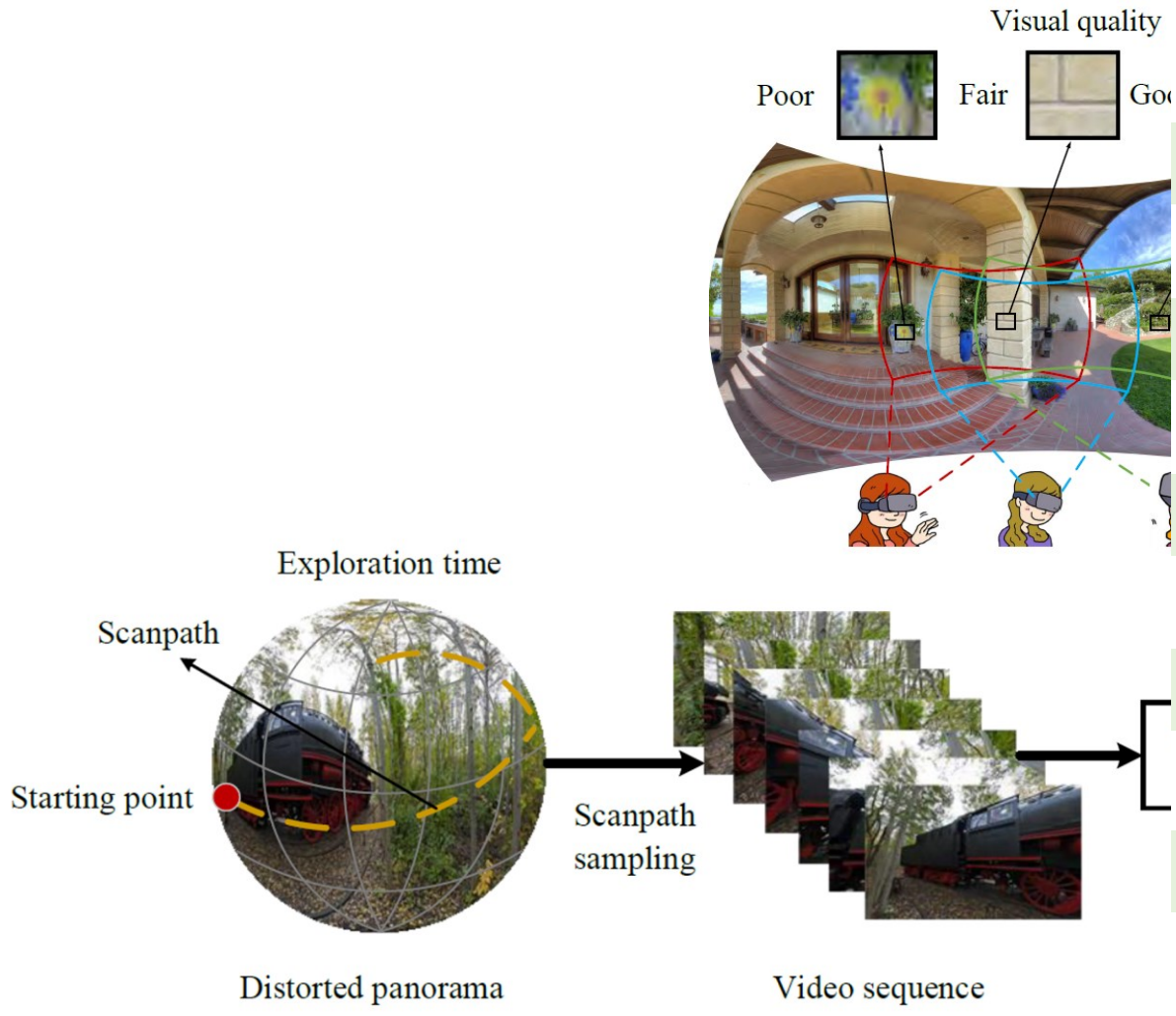


Database	Method	AUC ↑	NSS ↑	CC ↑	KLD ↓
Sitzmann [51]	BMS360 [34]	0.831	1.328	0.666	0.475
	GBVS360 [34]	0.845	1.433	0.693	0.374
	SaltiNet [2]	0.762	0.948	0.497	0.671
	SalNet360 [44]	0.841	1.467	0.718	0.421
	SalGAN360 [8]	0.813	1.188	0.612	0.561
	ScanGAN [43]	0.761	0.902	0.475	0.654
	ScanDMM	0.767	0.958	0.521	0.609
	Human	0.877	2.265	1.000	0.117
Salient360! [47]	BMS360 [34]	0.748	0.907	0.567	0.476
	GBVS360 [34]	0.712	0.775	0.484	0.576
	SaltiNet [2]	0.737	0.842	0.530	0.514
	SalNet360 [44]	0.744	0.989	0.602	0.476
	SalGAN360 [8]	0.759	0.975	0.613	0.495
	ScanGAN [43]	0.728	0.774	0.491	0.556
	ScanDMM	0.777	1.042	0.660	0.384
	Human	0.860	2.395	1.000	0.290

Table 4. Performance of saliency prediction models on Sitzmann [51] and Salient360! [47] databases.

11. P. Lebreton and A. Raake. GBVS360, BMS360, ProSal: Extending existing saliency prediction models from 2D to omnidirectional images. Signal Processing: Image Communication, 69:69–78, 2018.
12. F.-Y. Chao, L. Zhang, W. Hamidouche, and O. Deforges. SalGAN360: Visual saliency prediction on 360 degree images with generative adversarial networks. In IEEE International Conference on Multimedia and Expo Work- shops, pages 1–4, 2018.
13. R. Monroy, S. Lutz, T. Chalasani, and A. Smolic. SalNet360: Saliency maps for omni-directional im- ages with CNN. Signal Processing: Image Communication, 69:26–34, 2018.
14. M. Assens, X. G. Nieto, K. McGuinness, and N. E. O'Connor. SaltiNet: Scan-path prediction on 360 de-gree images using saliency volumes. In IEEE International Conference on Computer Vision Workshops, pages 2331– 2338, 2017.

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.



Method	CC ↑	SRCC ↑	RMSE ↓
PSNR	0.156	0.016	0.787
$PSNR_{GAN}$	0.135	0.014	0.790
$PSNR_{DMM}$	0.563	0.546	0.659
$PSNR_{Human}$	0.585	0.583	0.646
SSIM	0.148	0.046	0.788
$SSIM_{GAN}$	0.162	0.046	0.790
$SSIM_{DMM}$	0.519	0.509	0.681
$SSIM_{Human}$	0.527	0.532	0.677
VIF	0.163	0.096	0.786
VIF_{GAN}	0.147	0.092	0.788
VIF_{DMM}	0.597	0.572	0.639
VIF_{Human}	0.616	0.601	0.628
DISTS	0.162	0.081	0.786
$DISTS_{GAN}$	0.176	0.100	0.784
$DISTS_{DMM}$	0.662	0.675	0.597
$DISTS_{Human}$	0.700	0.711	0.569
DeepWSD	0.160	0.044	0.786
$DeepWSD_{GAN}$	0.162	0.072	0.786
$DeepWSD_{DMM}$	0.635	0.628	0.616
$DeepWSD_{Human}$	0.668	0.667	0.593

Table 5. Performance of quality assessment models on JUFE [20] database.

15. X. Sui, K. Ma, Y. Yao, and Y. Fang. Perceptual quality assessment of omnidirectional images as moving camera videos. IEEE Transactions on Visualization and Computer Graphics, 28(8):3022–3034, 2022.

»» Summary

Summary

- Deep Markov Model. Time-dependent viewing behavior modeling
- Flexible and Fast. ScanDMM can produce 1,000 variable-length scanpaths from arbitrary starting points within 1-second.
- Generalizability. We apply ScanDMM to two other computer vision tasks, which demonstrates our model equips with strong generalizability.

Models :
https://github.com/xiangjieSui/ScanDMM

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.

Thank you for your attention!

Email: fa0001ng@e.ntu.edu.sg

X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, ScanDMM: A Deep Markov Model of Scanpath Prediction for 360° Images, IEEE Conference on Computer Vision and Pattern Recognition, 2023.