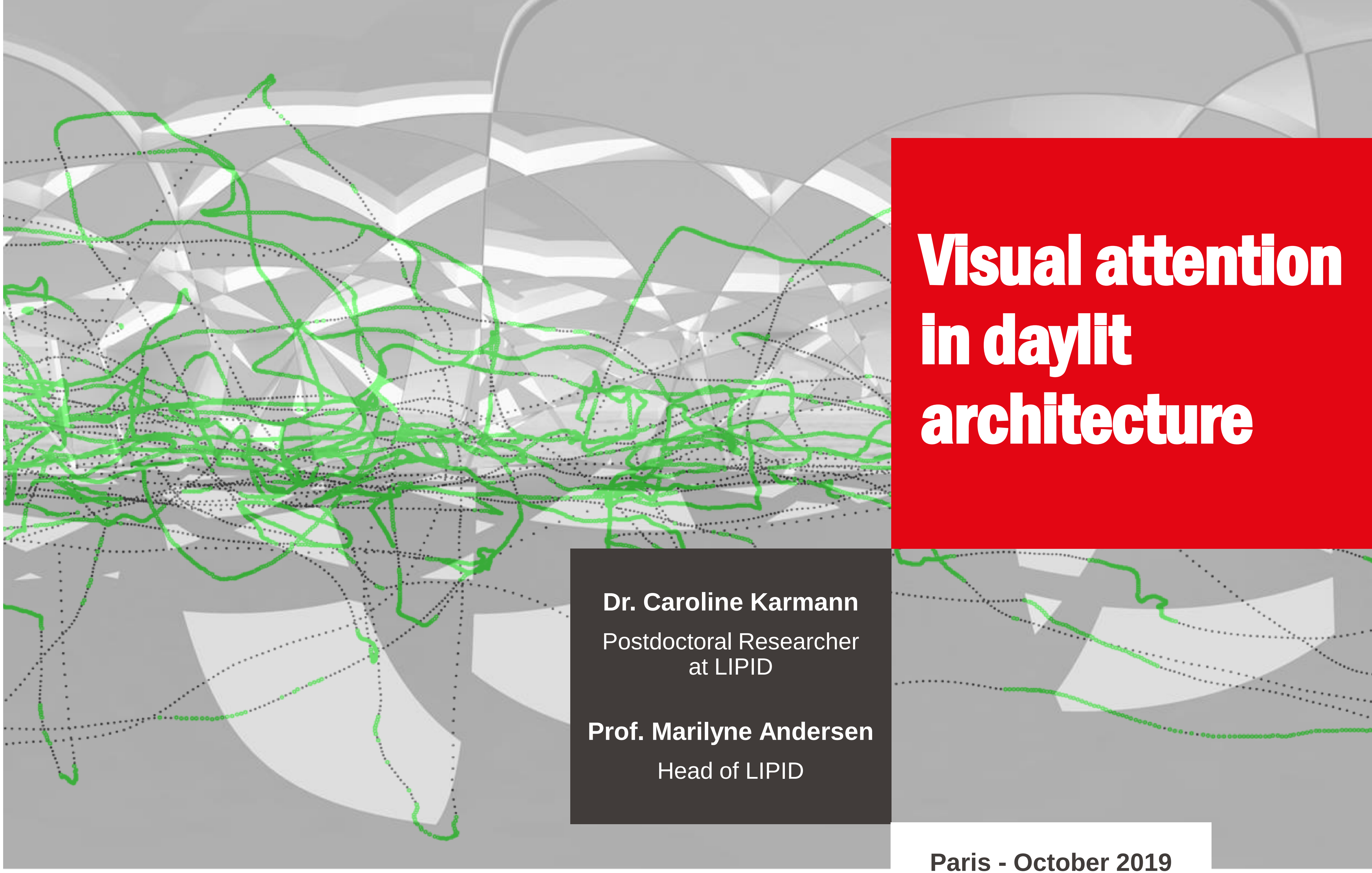
Visual attention in daylit architecture



Dr. Caroline Karmann

Postdoctoral Researcher
at LIPID

Prof. Marilyne Andersen

Head of LIPID

Paris - October 2019

EPFL

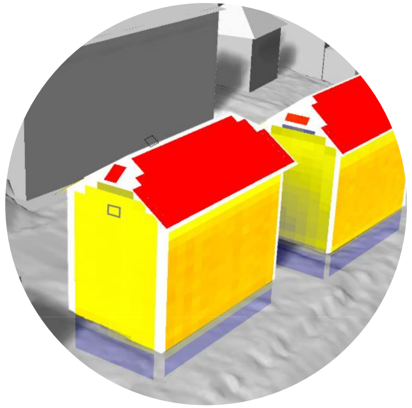
LIPID
Laboratory of
Integrated
Performance
in design

Laboratory of Integrated Performance in Design (LIPID)

Comfort

Perception

Health



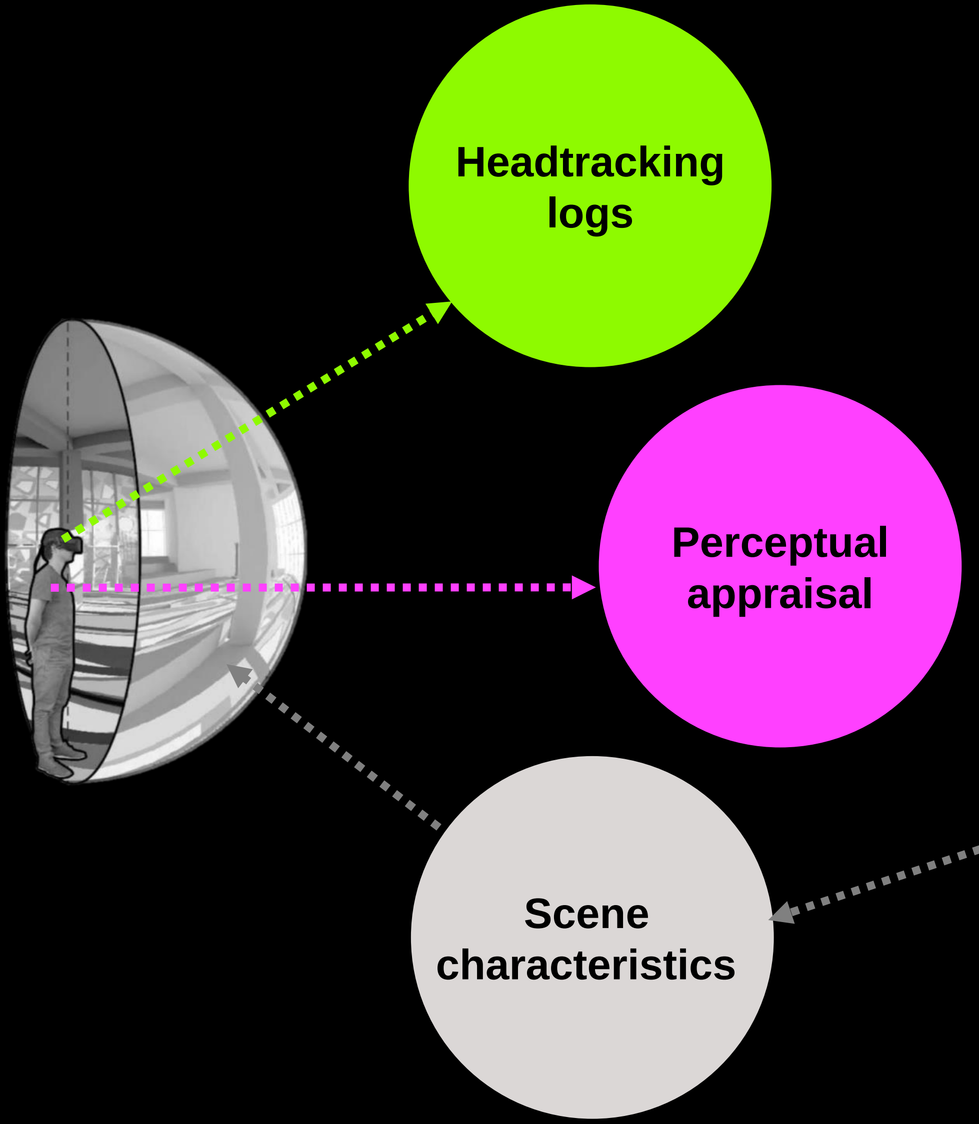
Energy

Example:

Ryerson Student Learning Centre, Toronto, CA

Zeidler Partnership Architects + Snøhetta

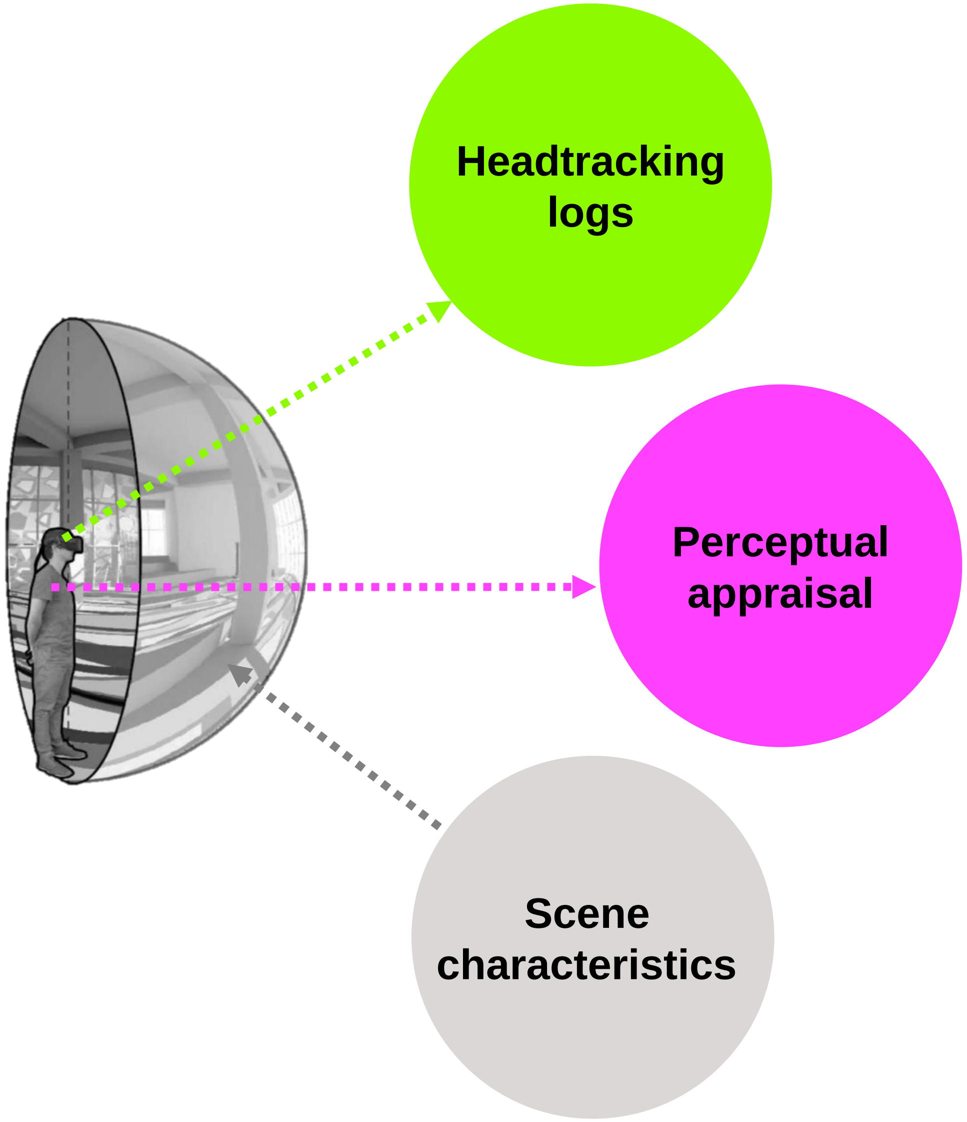
Images: L. Bridgman, Doublespace Photography



Example:
Ryerson Student Learning Centre, Toronto, CA
Zeidler Partnership Architects + Snøhetta

Illustration from (Rockcastle et al.,2017) (adapted) | Photograph from L. Bridgman, Doublespace Photography

LIPID previous work



Study by (Rockcastle et al., 2017)

Objectives

- Evaluate perceptual responses
- Effect of sky condition / space
- Develop a contrast-based metric (mSC5)

Procedure

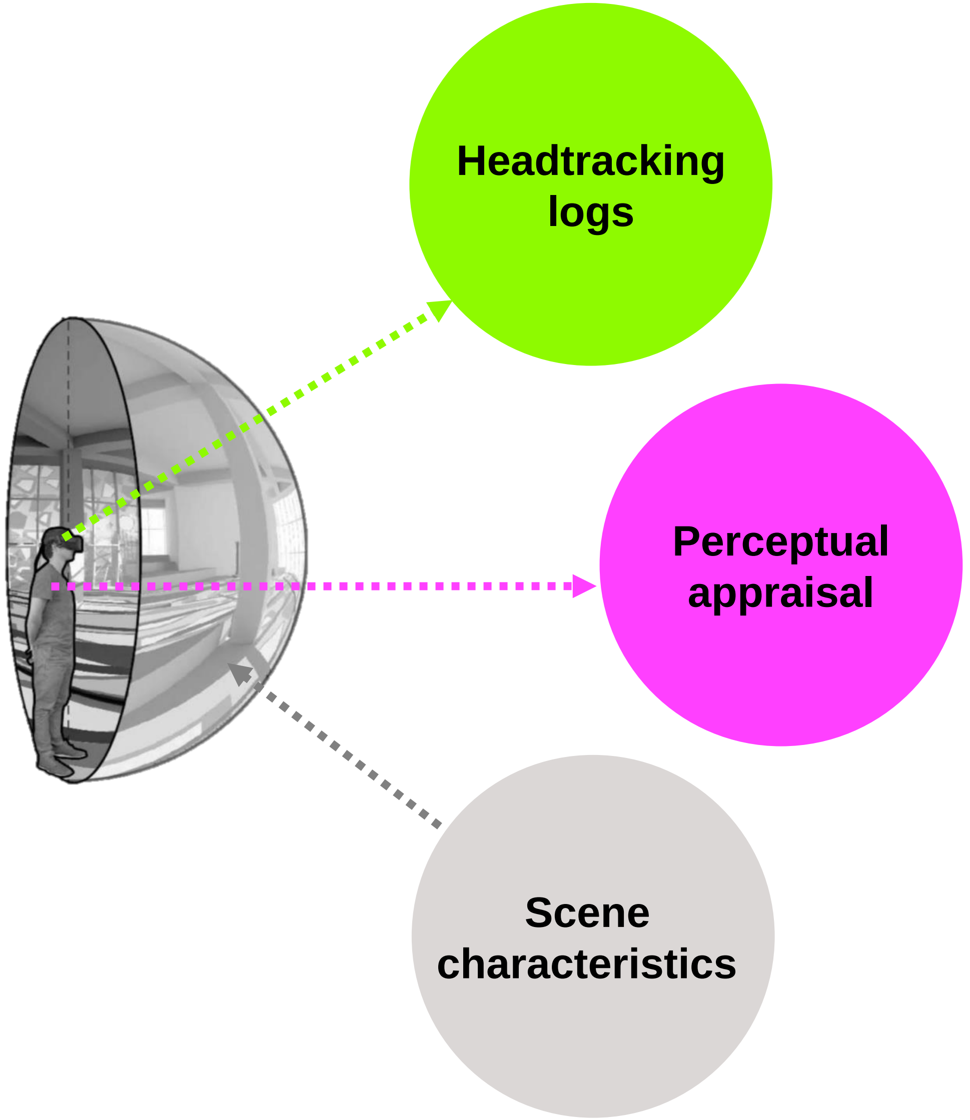
- VR-based human-subject experiment
- 360° fully immersive (Oculus rift, cubemaps)
- Exposure to stimuli: participant's choice

Dataset

- Visual stimuli: 16 scenes (B/W renderings)
- Headtracking logs: ~12 participants/scene

Rockcastle, S.F., 2017. Perceptual Dynamics of Daylight in Architecture. EPFL, Thesis No. 7677.
Rockcastle, S. F., Chamilothori, K., & Andersen, M., 2017. An Experiment in Virtual Reality to Measure Daylight-Driven Interest in Rendered Architectural Scenes. Proceedings of Building Simulation 2017.

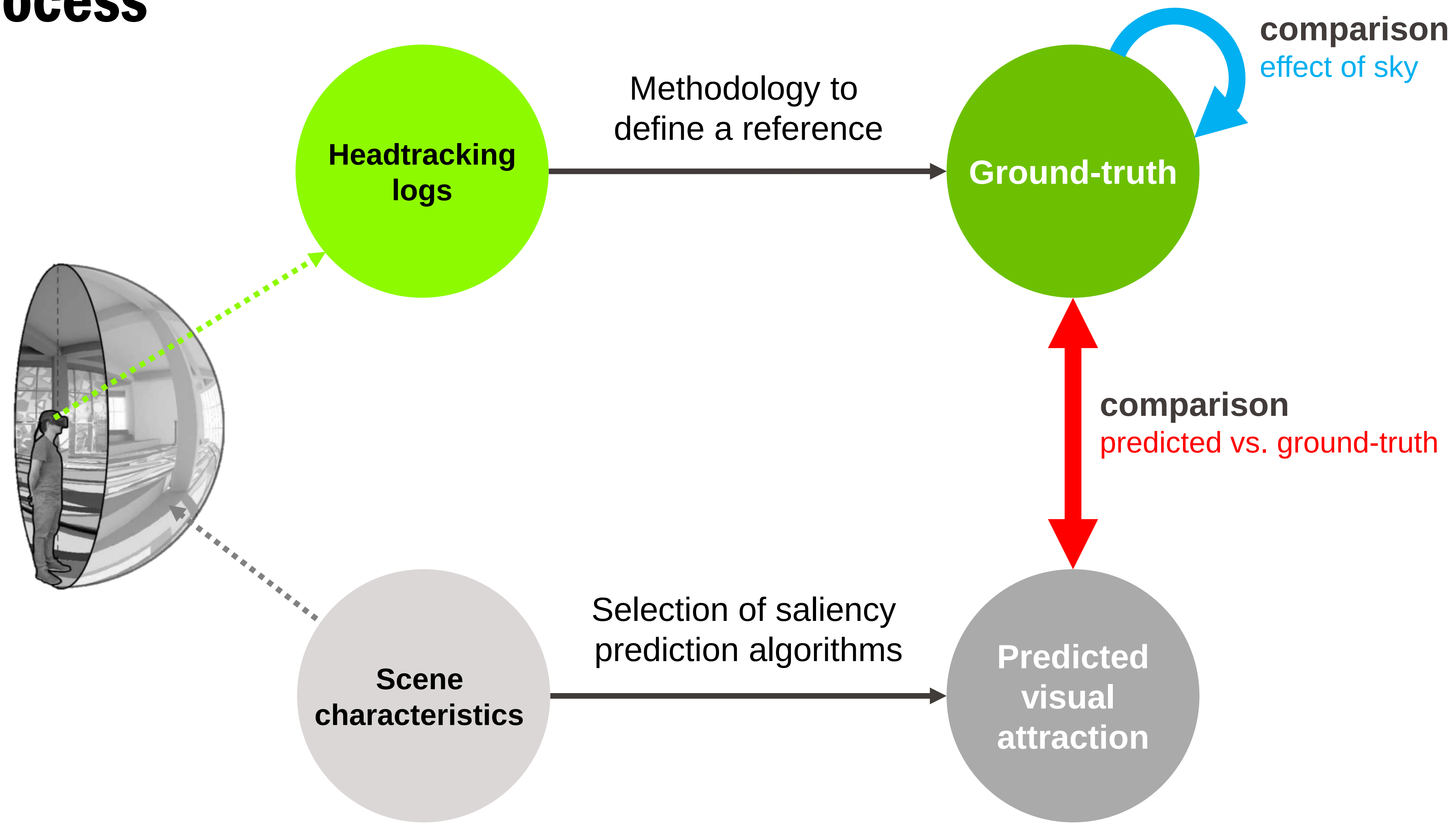
LIPID previous work



Study by (Rockcastle et al., 2017)

Rockcastle, S.F., 2017. Perceptual Dynamics of Daylight in Architecture. EPFL, Thesis No. 7677.
Rockcastle, S. F., Chamilothori, K., & Andersen, M., 2017. An Experiment in Virtual Reality to Measure Daylight-Driven Interest in Rendered Architectural Scenes. Proceedings of Building Simulation 2017.

Process



Headtracking logs
Methodology to define a reference
Ground-truth
comparison
effect of sky
comparison
predicted vs. ground-truth
Scene characteristics
Selection of saliency prediction algorithms
Predicted visual attraction

EPFL

Saliency prediction

Do existing saliency models accurately predict visual attention in daylit architectural spaces?

Establishing ground truth

Coordinate systems

Equi-projection

Account for distortions

Exposure to visual stimuli

5 sec, 25 sec, no limit

Fixations vs. saccades

Angular velocity threshold

15°/sec, 60°/sec, no limit

Gaussian filter

Standard deviation (Rai et al., 2017)

Corrected filter (Upenik et al., 2017)

Rounding and flattening (Upenik et al., 2017)

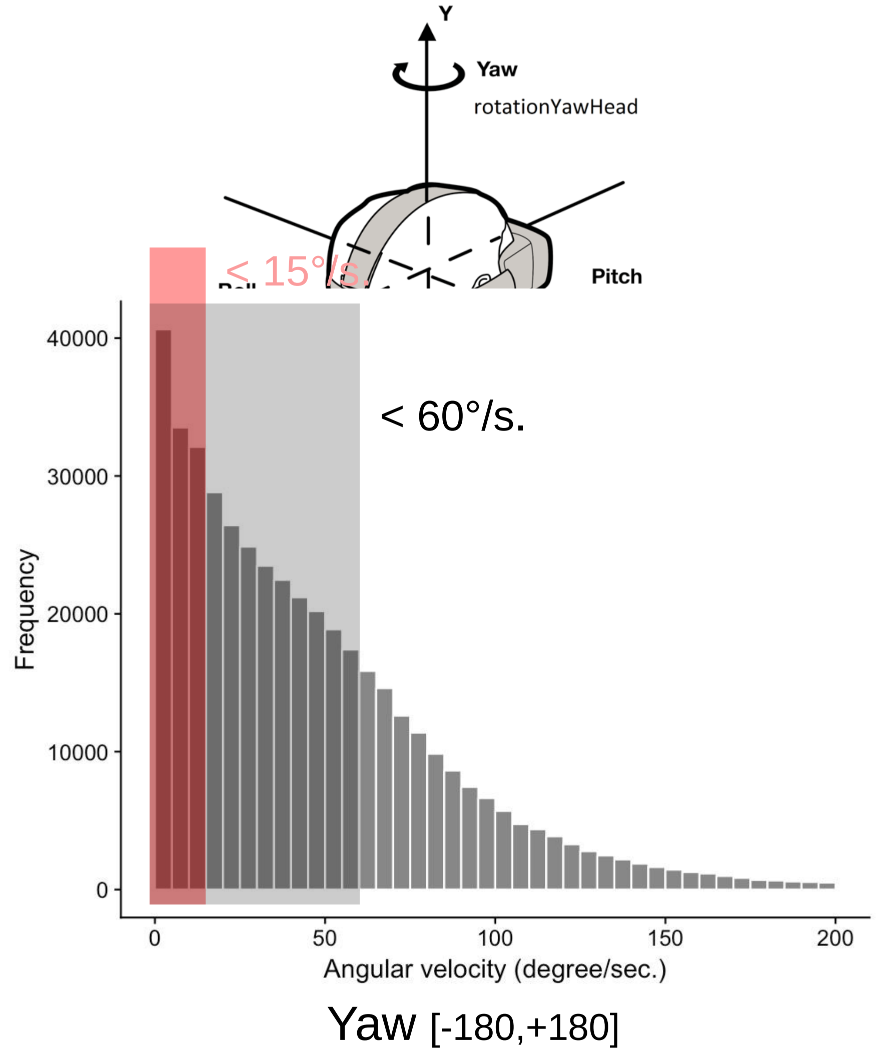
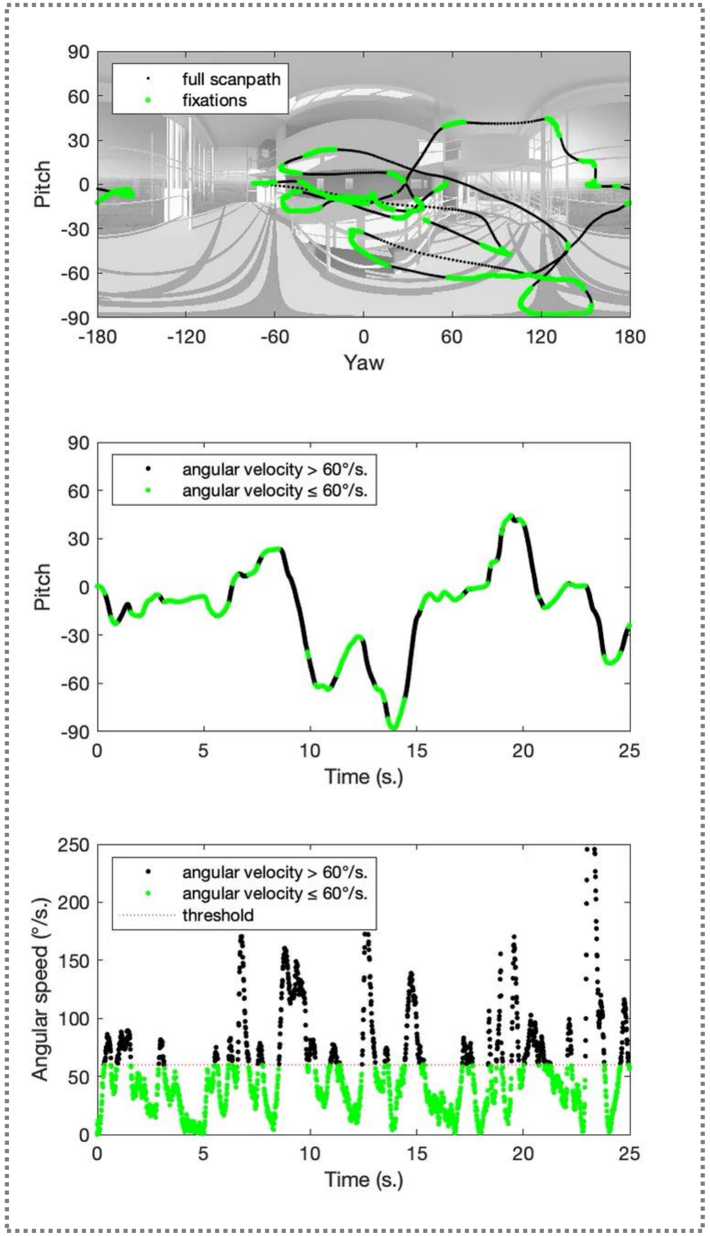


Illustration from Oculus Rift Developer Guide

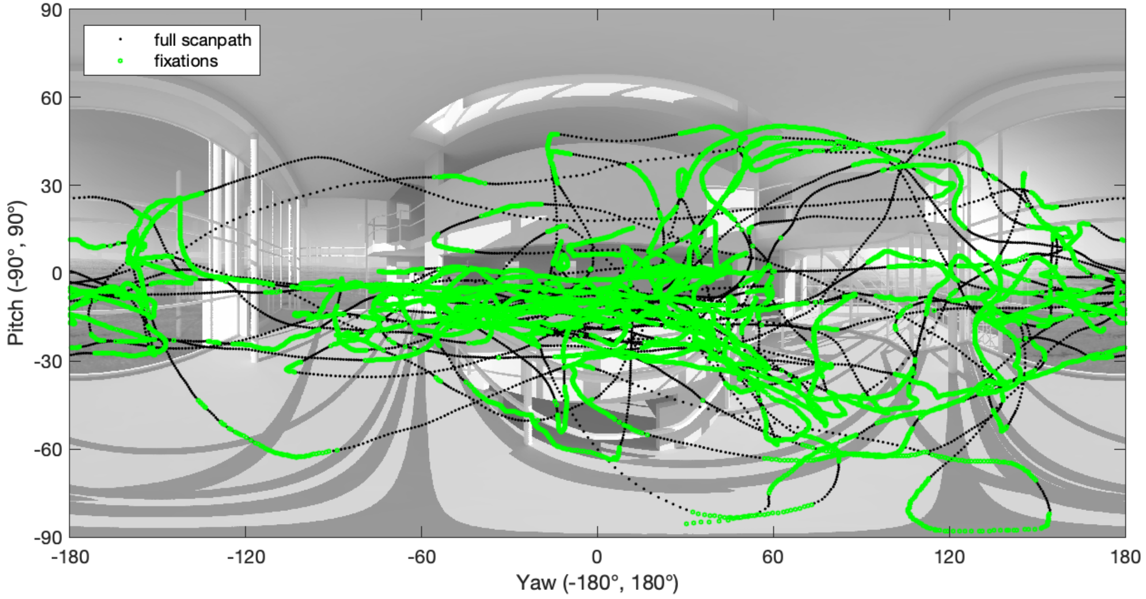
Upenik et al.. 2017, A simple method to obtain visual attention data in head mounted virtual reality

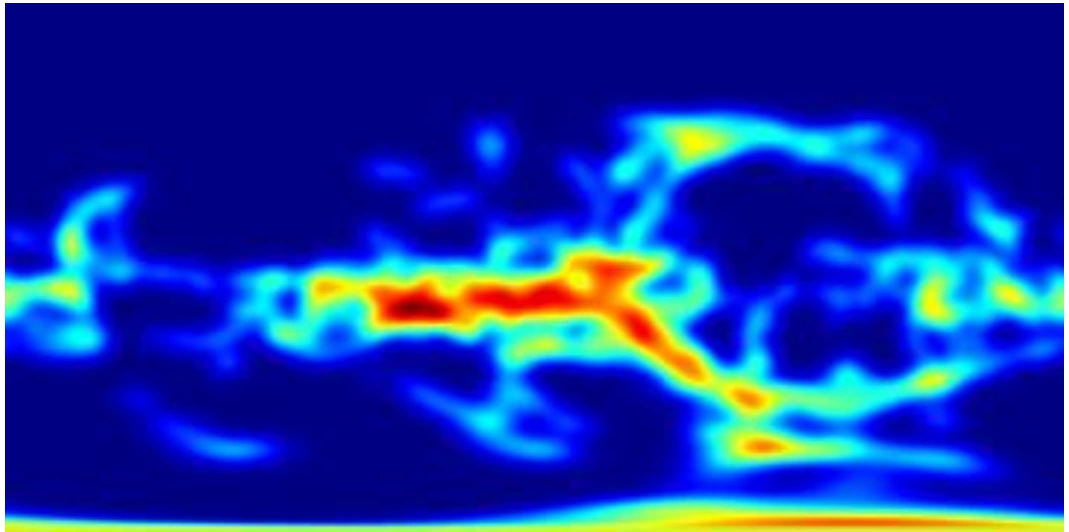
Rai et al., 2017, A Dataset of Head and Eye Movements for 360 Degree Images

Establishing ground truth

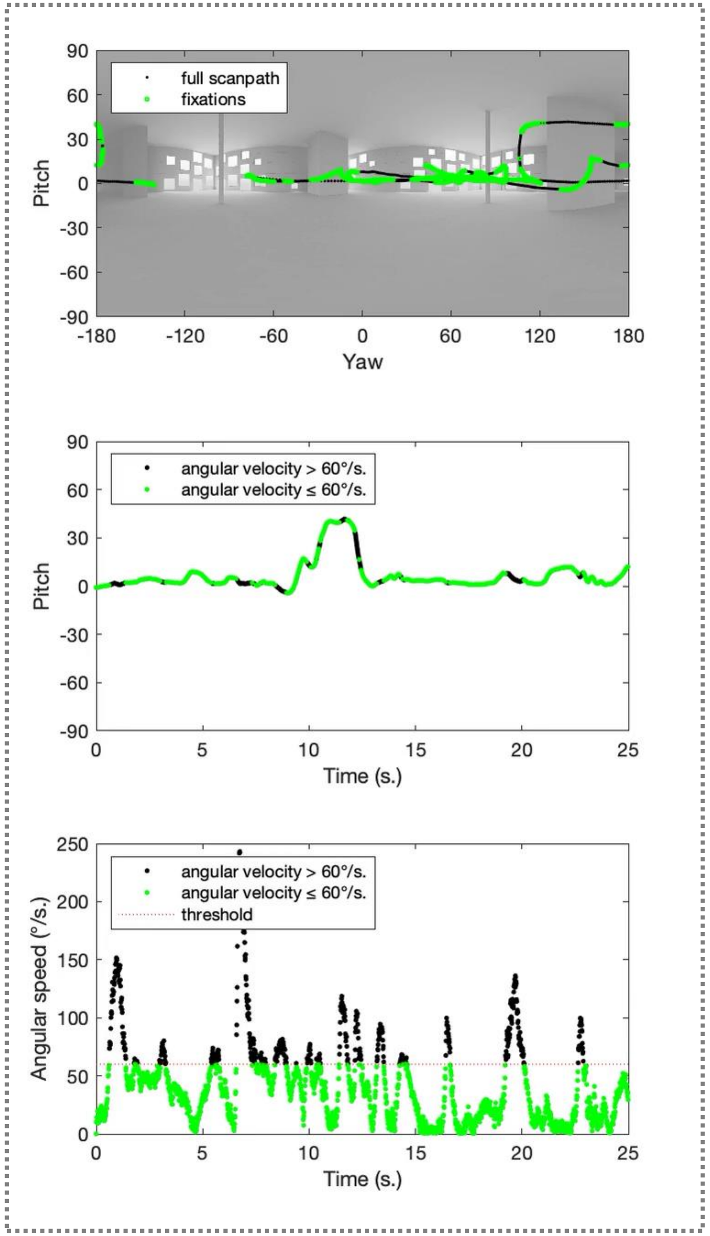


Douglas House | Clear

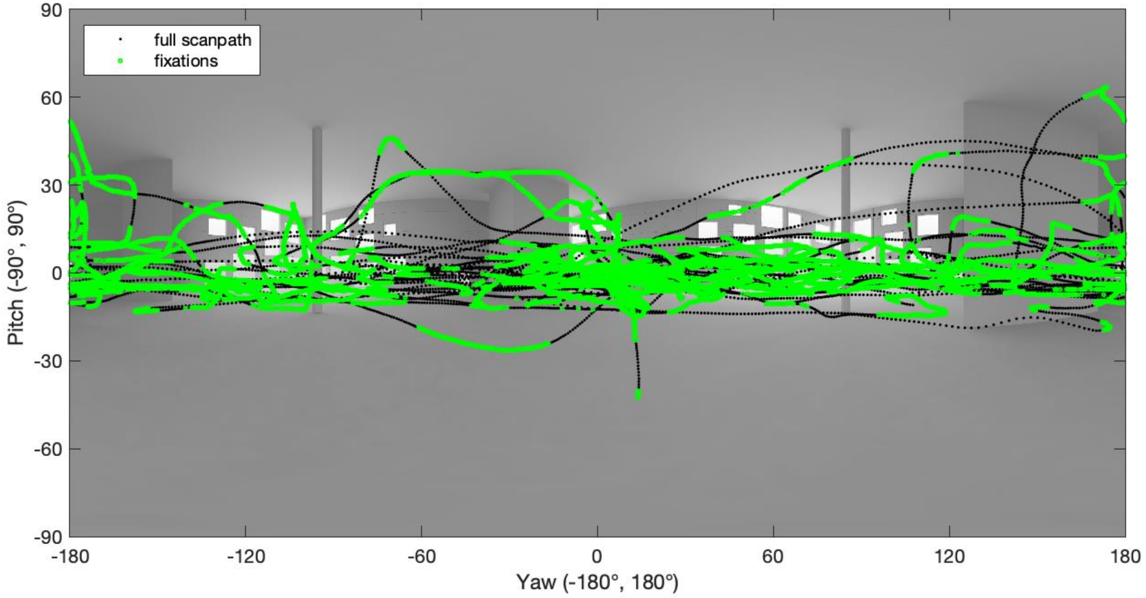


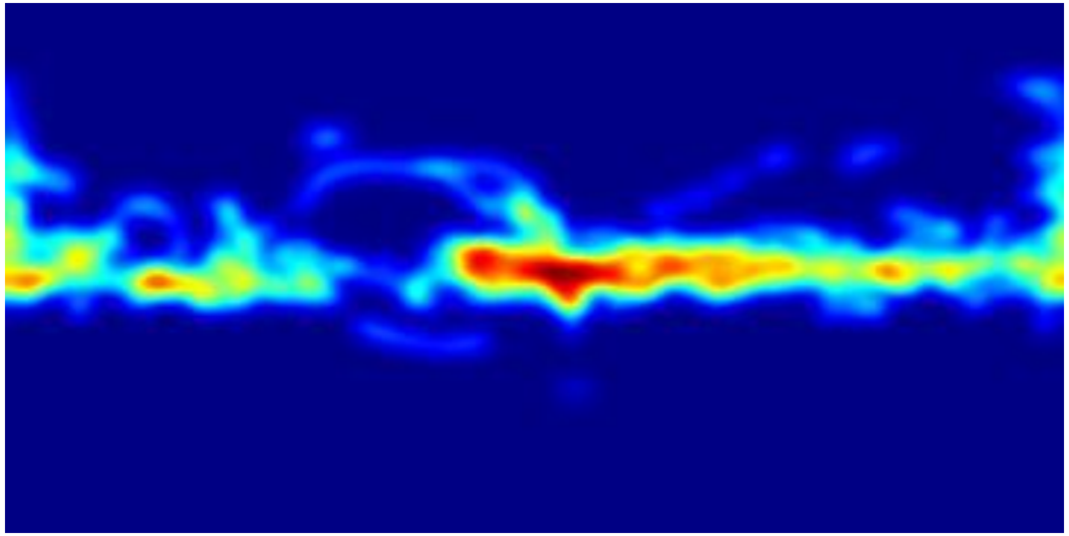


Establishing ground truth



Zollverein | Overcast





Saliency prediction

Visual attentional mechanism

Bottom-up attention

low-level visual features

e.g., intensity, color, orientation, texture, directions

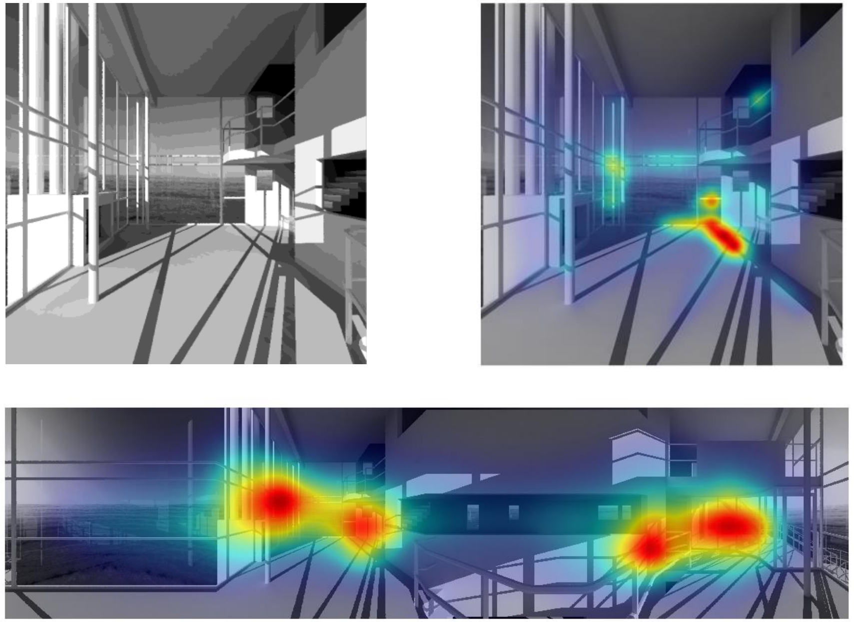
Top-down attention

high-level features (recognition)

e.g., faces, cars, objects, furniture, etc.

Early models:

- Itti-Koch model (2001)
- Graph-based Visual Saliency (GBVS) (Harel et al., 2006)



Test with GBVS

Itti L., 2001. Computational modelling of visual attention, Nat. Rev. Neurosci.
Harel J. et al., 2006. Graph-based visual saliency, in: Proceedings of Neural Information Processing Systems, NIPS.

Saliency prediction

Visual attentional mechanism

Bottom-up attention

low-level visual features

e.g., intensity, color, orientation, texture, directions

Top-down attention

high-level features (recognition)

e.g., faces, cars, objects, furniture, etc.

Advanced computational capabilities

Deep learning

Convolutional neural networks (CNN)

Learn from images / multiple layers (input > output)
Allows large scale object recognition corpses
Perform better than traditional saliency models

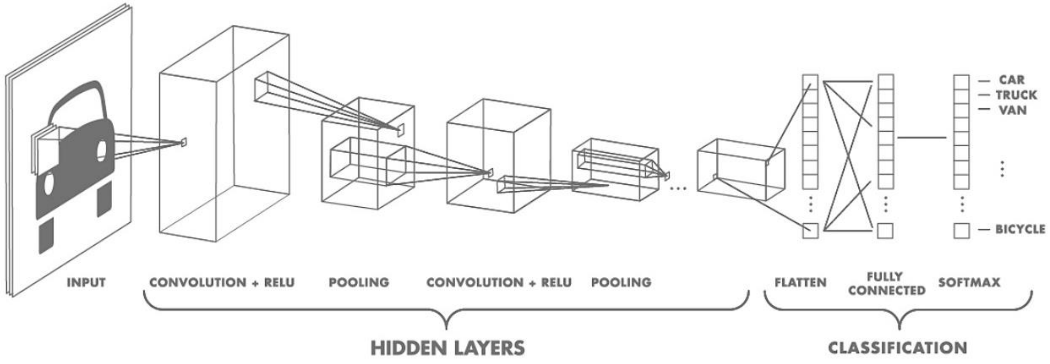


Image: https://ch.mathworks.com/fr/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

Saliency prediction

Visual attentional mechanism

Bottom-up attention

low-level visual features

e.g., intensity, color, orientation, texture, directions

Top-down attention

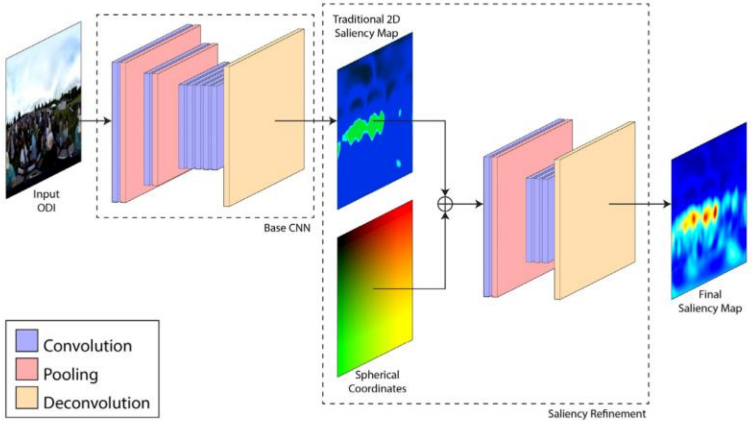
high-level features (recognition)

e.g., faces, cars, objects, furniture, etc.

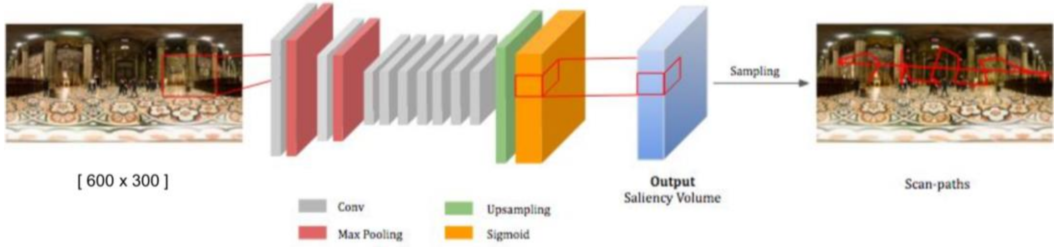
Advanced computational capabilities

Publicly available and pre-trained CNN-based models from VR data

SalNet360 (Monroy et al. 2018)



SaltiNet (Assens et al. 2017)



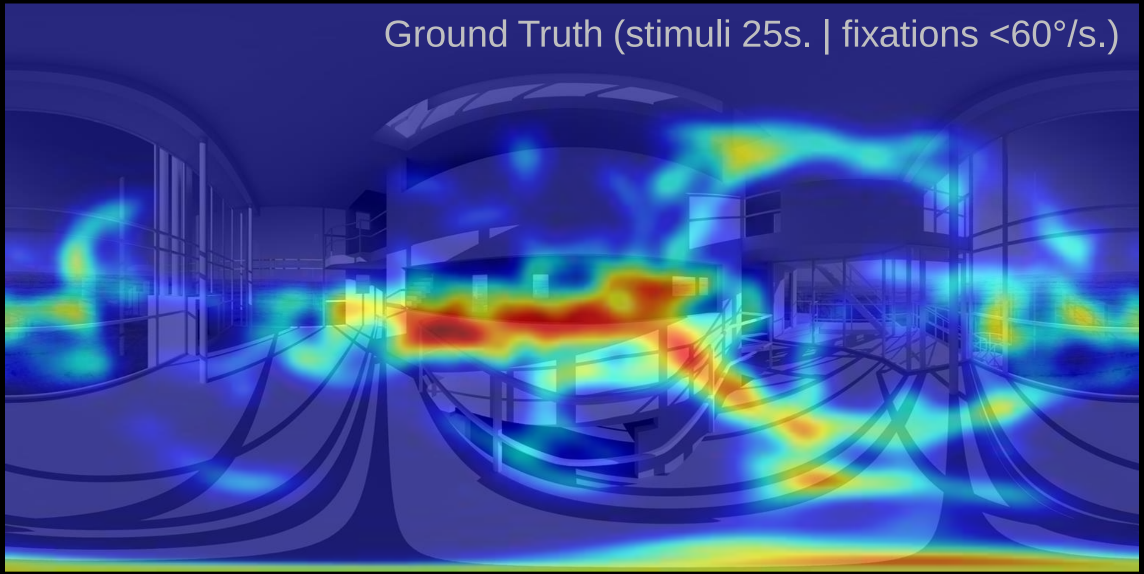
Monroy et al., 2018. SalNet360: Saliency Maps for Omni-Directional Images with CNN. Signal Processing: Image Communication
Assens, M. et al.. 2018. Scanpath and Saliency Prediction on 360 Degree Images. Signal Processing: Image Communication

Results

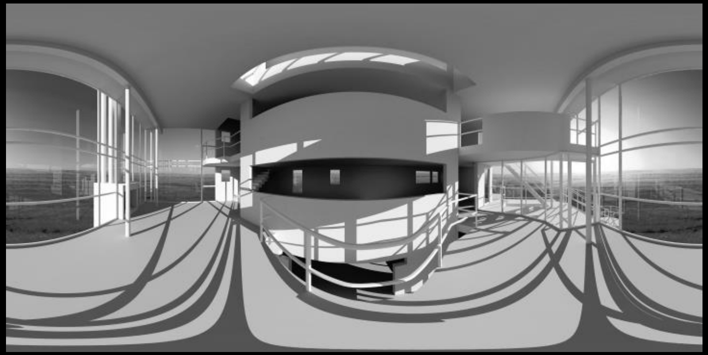
Scene 1A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.03, KLdiv 5.5
SaltiNet (Assens Reina et al. 2017) - CC 0.05, KLdiv 4.4
Scene 1C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.16, KLdiv 4.4
SaltiNet (Assens Reina et al. 2017) - CC 0.16, KLdiv 4.1
Scene 2A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.15, KLdiv 4.5
SaltiNet (Assens Reina et al. 2017) - CC 0.26, KLdiv 3.6
Scene 2C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.06, KLdiv 4.7
SaltiNet (Assens Reina et al. 2017) - CC 0.24, KLdiv 3.7
Scene 3A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.19, KLdiv 5.3
SaltiNet (Assens Reina et al. 2017) - CC 0.36, KLdiv 3.6
Scene 3C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.04, KLdiv 6.9
SaltiNet (Assens Reina et al. 2017) - CC 0.29, KLdiv 5
Scene 4A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.25, KLdiv 3.6
SaltiNet (Assens Reina et al. 2017) - CC 0.35, KLdiv 2.9
Scene 4C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.06, KLdiv 5
SaltiNet (Assens Reina et al. 2017) - CC 0.15, KLdiv 4.5
Scene 5A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.13, KLdiv 5.7
SaltiNet (Assens Reina et al. 2017) - CC 0.39, KLdiv 3.9
Scene 5C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.14, KLdiv 6.4
SaltiNet (Assens Reina et al. 2017) - CC 0.39, KLdiv 5.1
Scene 6A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.23, KLdiv 3.7
SaltiNet (Assens Reina et al. 2017) - CC 0.37, KLdiv 2.7
Scene 6C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.36, KLdiv 4.8
SaltiNet (Assens Reina et al. 2017) - CC 0.37, KLdiv 3.7
Scene 7A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.09, KLdiv 6.2
SaltiNet (Assens Reina et al. 2017) - CC 0.26, KLdiv 5
Scene 7C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.09, KLdiv 6.9
SaltiNet (Assens Reina et al. 2017) - CC 0.31, KLdiv 5.4
Scene 8A
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.3, KLdiv 5.1
SaltiNet (Assens Reina et al. 2017) - CC 0.43, KLdiv 4.2
Scene 8C
Ground Truth
SalNet360 (Monroy et al. 2018) - CC 0.33, KLdiv 5.7
SaltiNet (Assens Reina et al. 2017) - CC 0.41, KLdiv 4.7

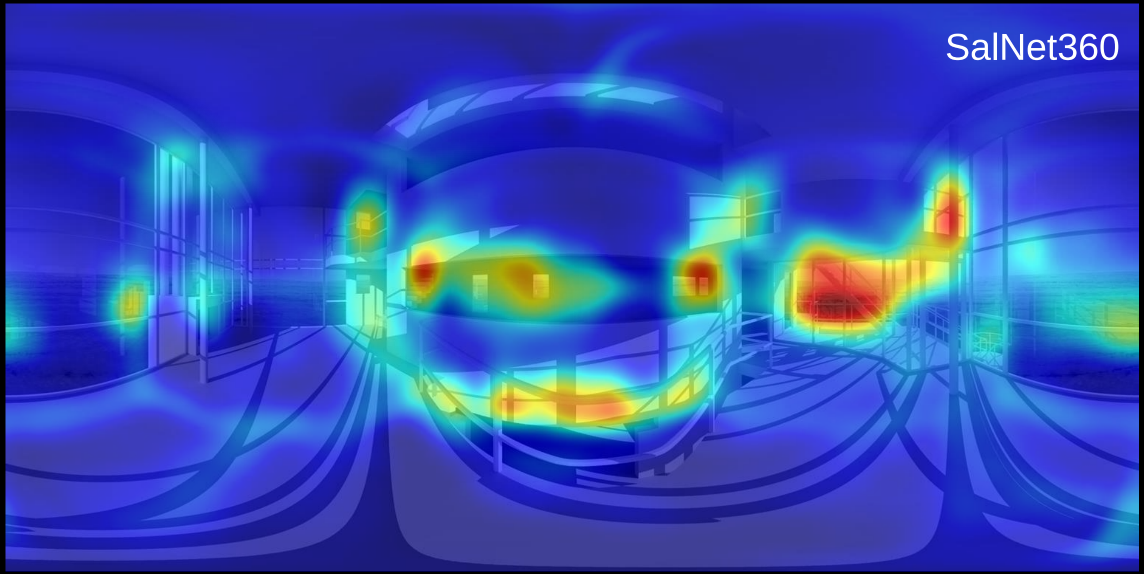
EPFL

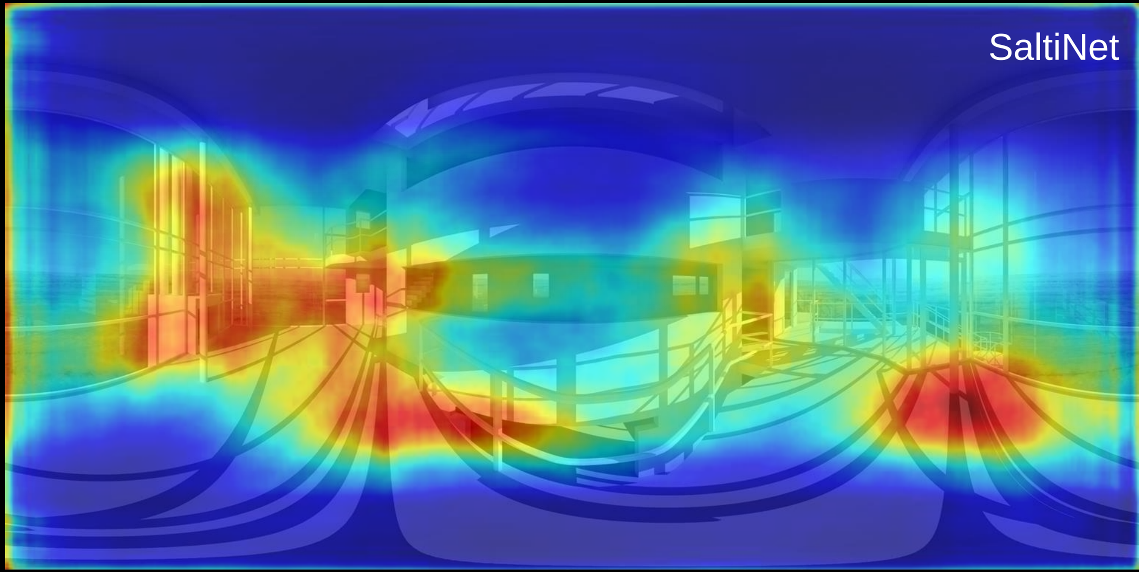
Saliency prediction vs. ground truth



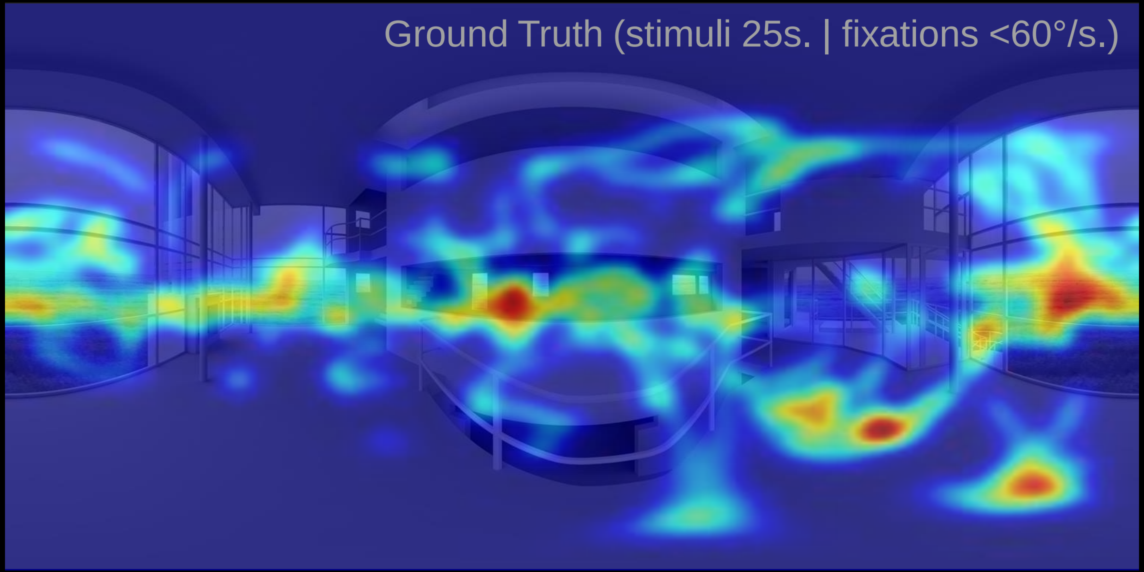
Douglas House | Clear



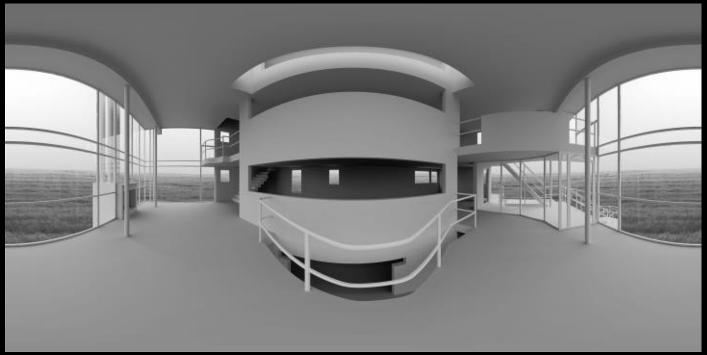


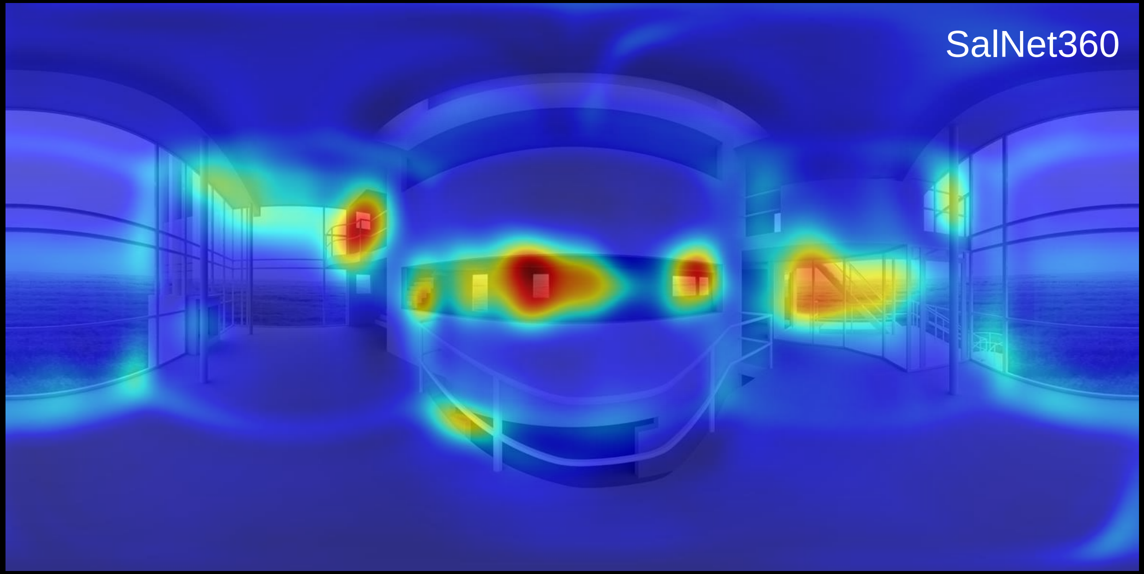


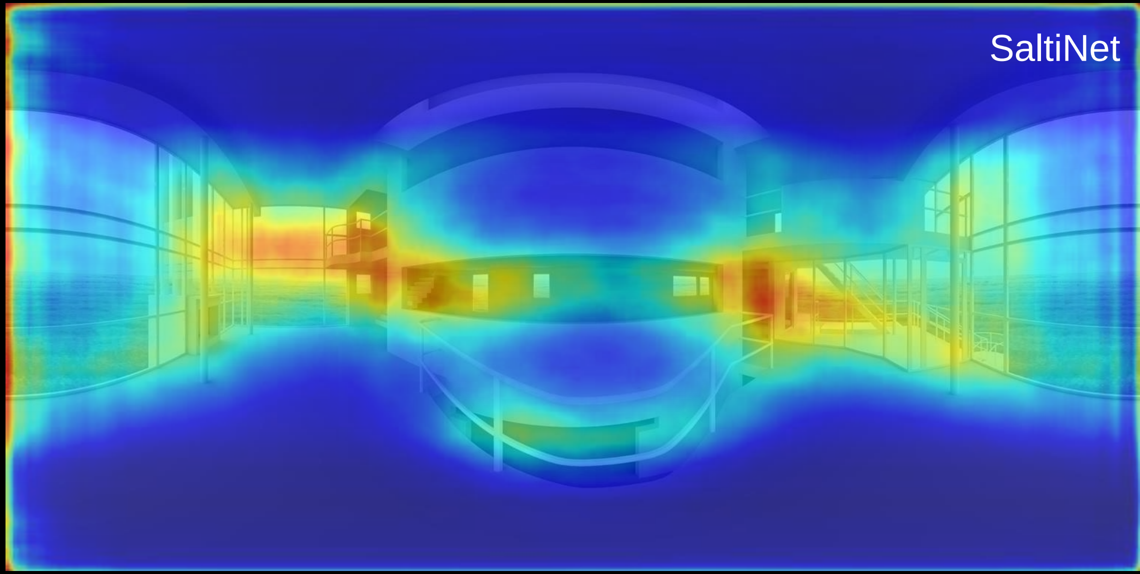
Saliency prediction vs. ground truth



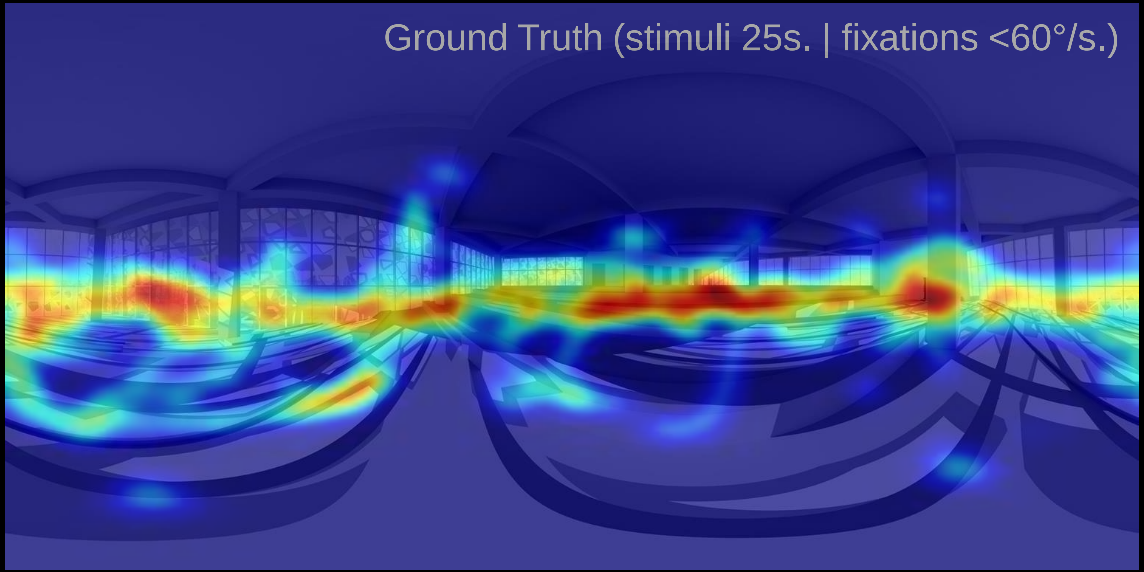
Douglas House | Overcast



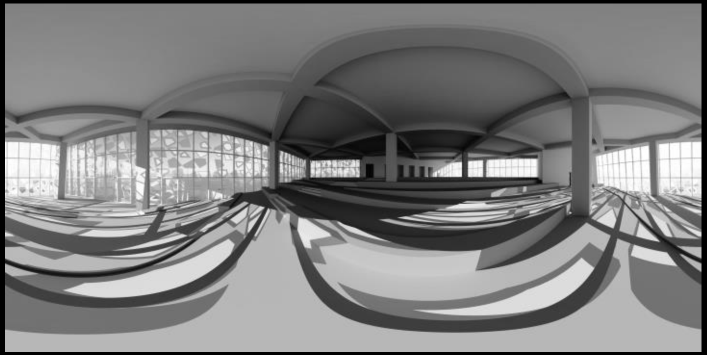


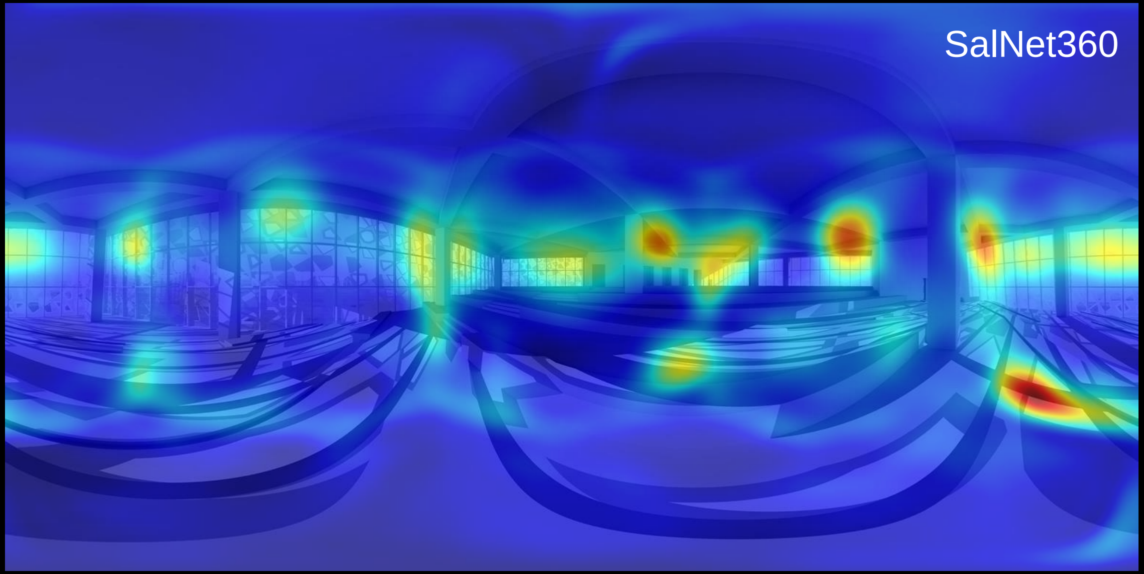


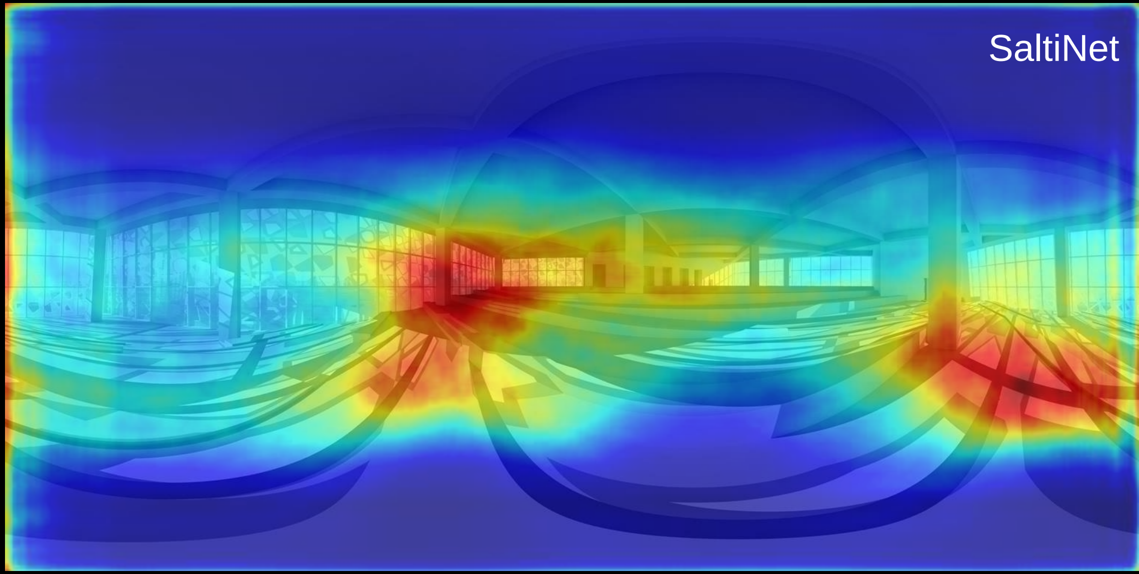
Saliency prediction vs. ground truth



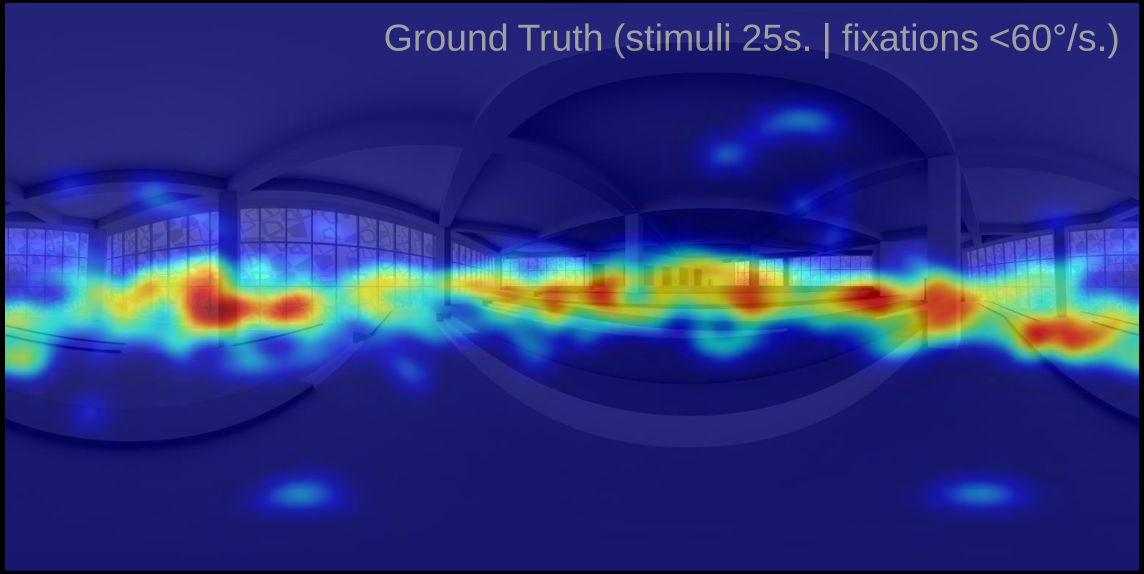
Ryerson | Clear



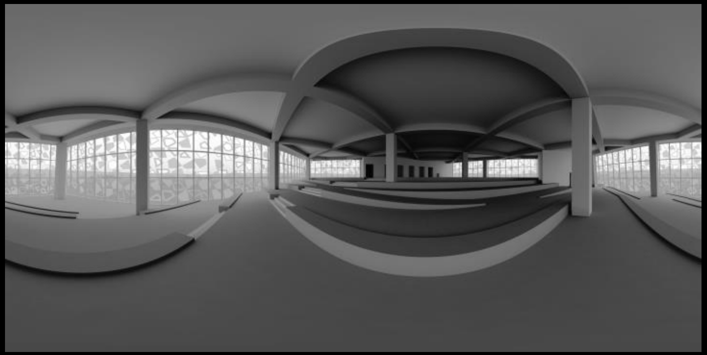


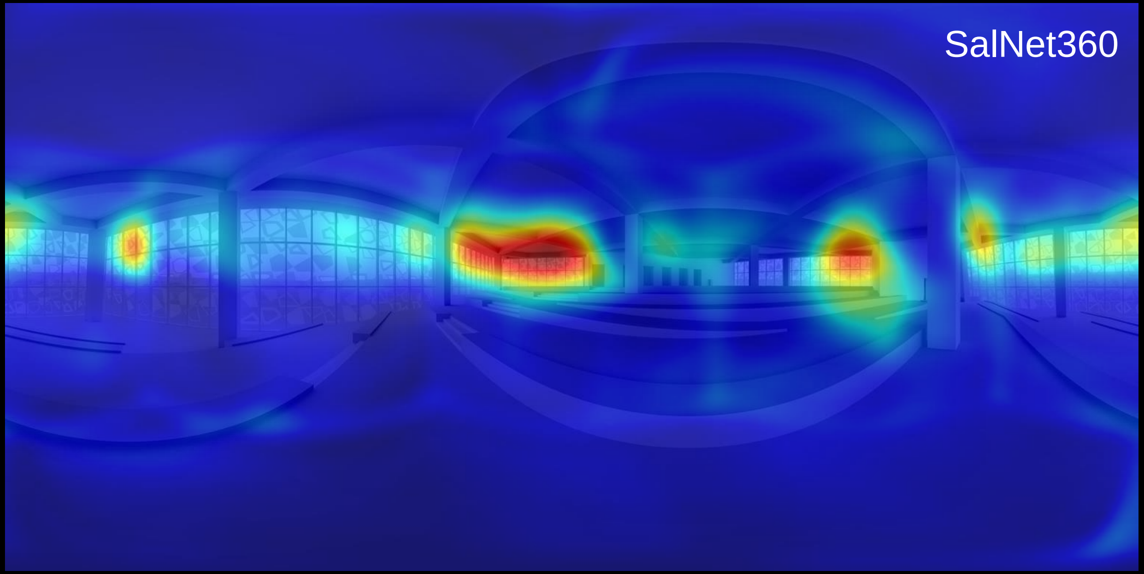


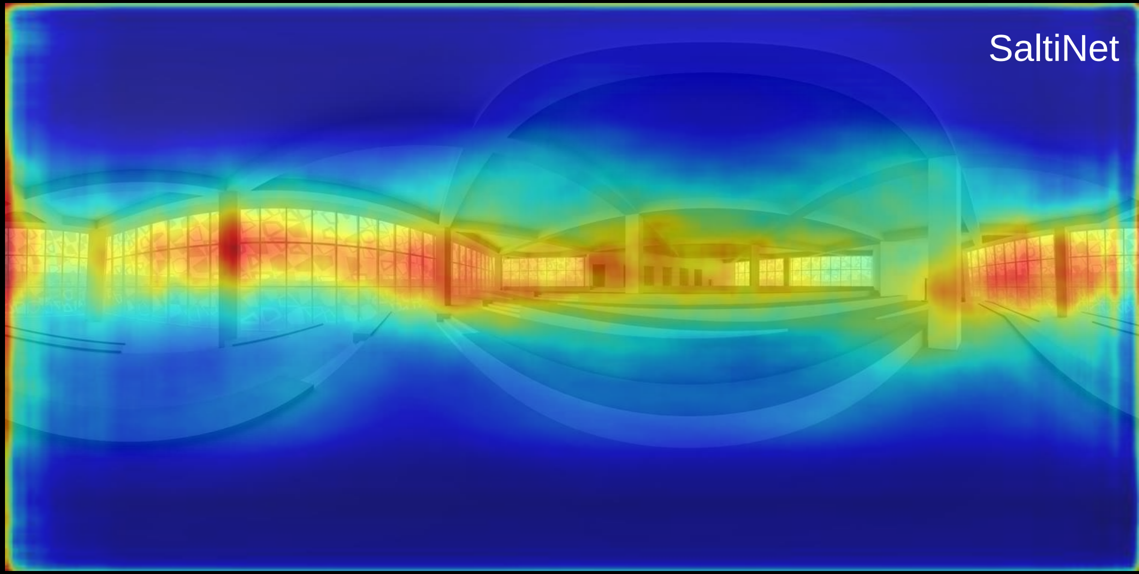
Saliency prediction vs. ground truth



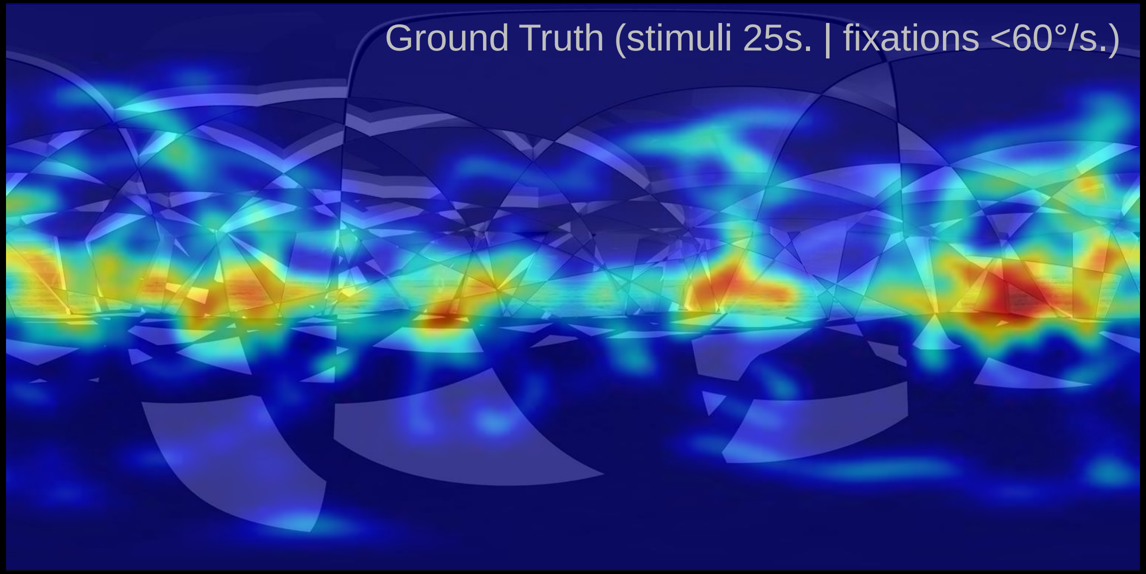
Ryerson | Overcast



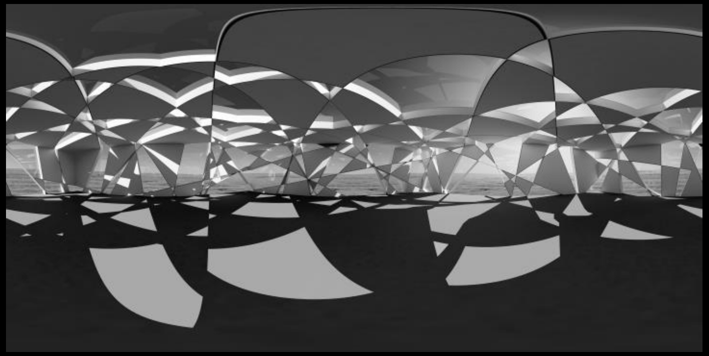


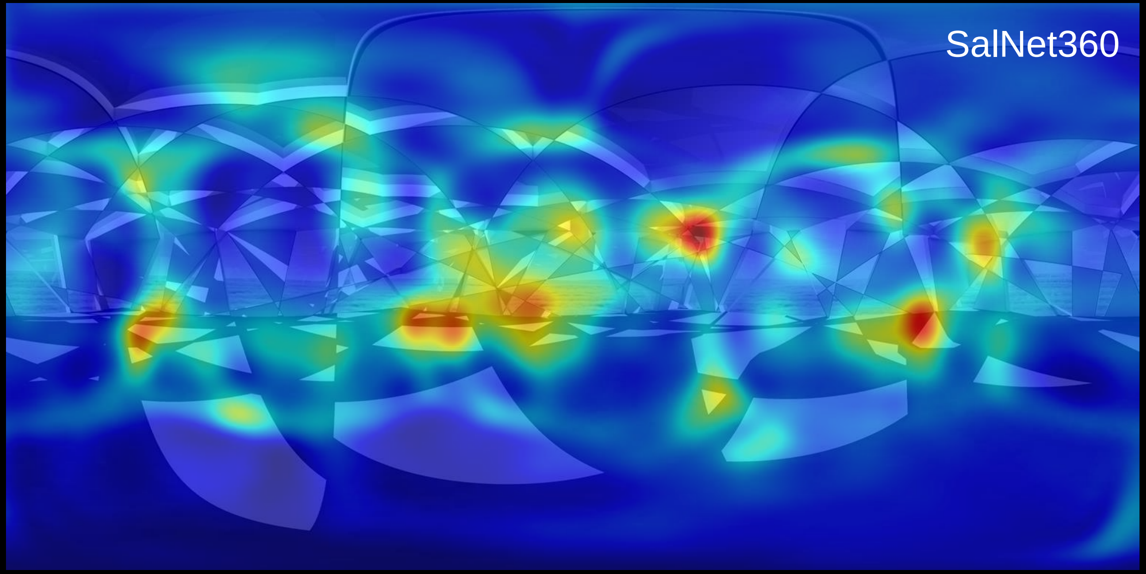


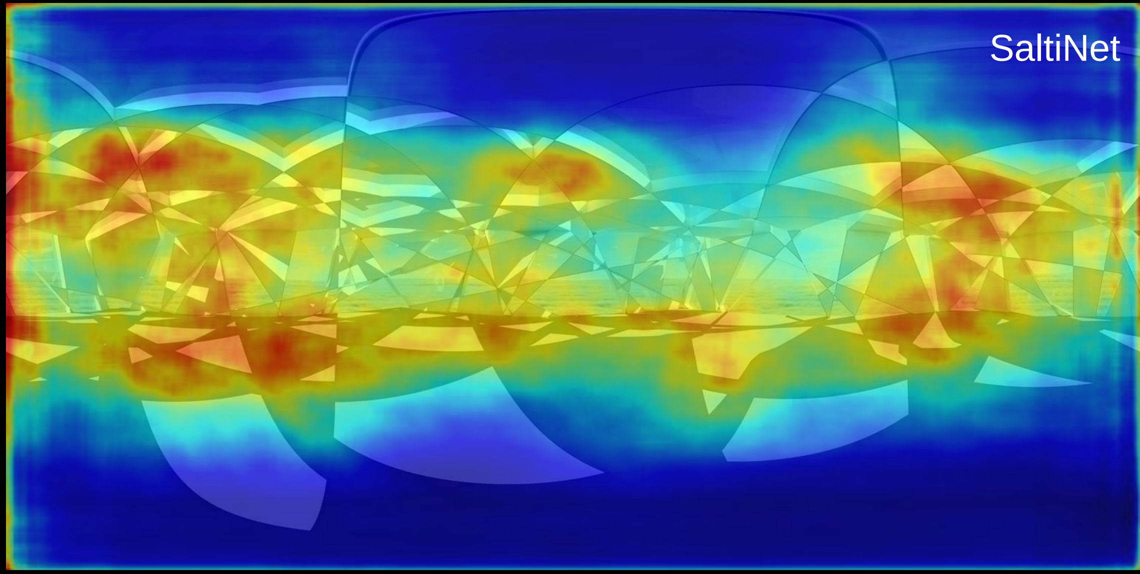
Saliency prediction vs. ground truth



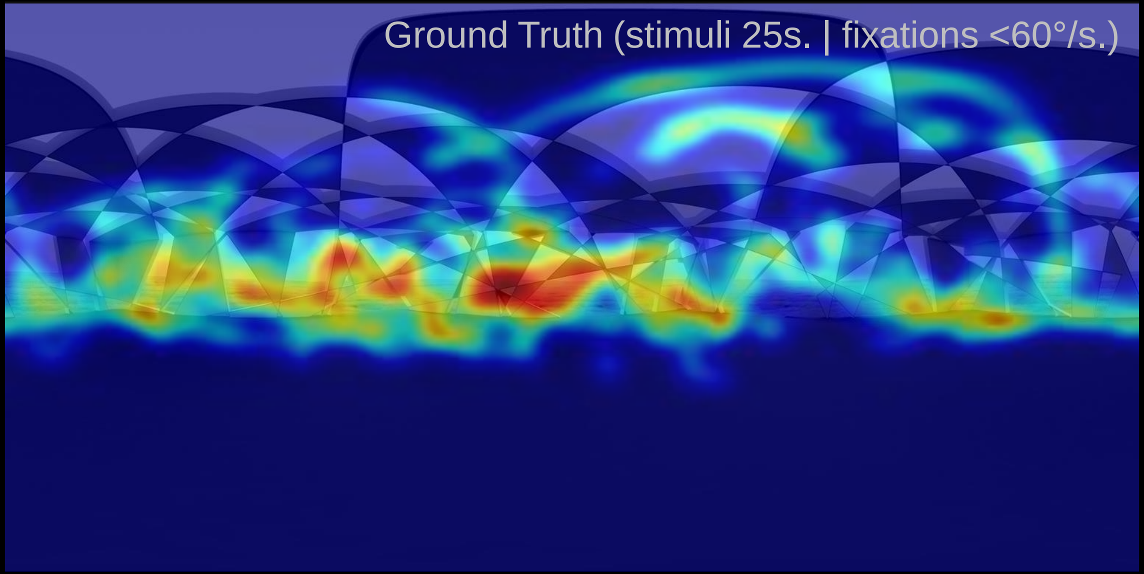
Serpentine pavilion | Clear



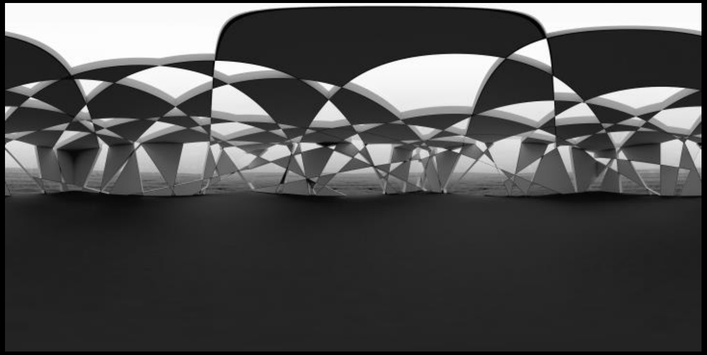


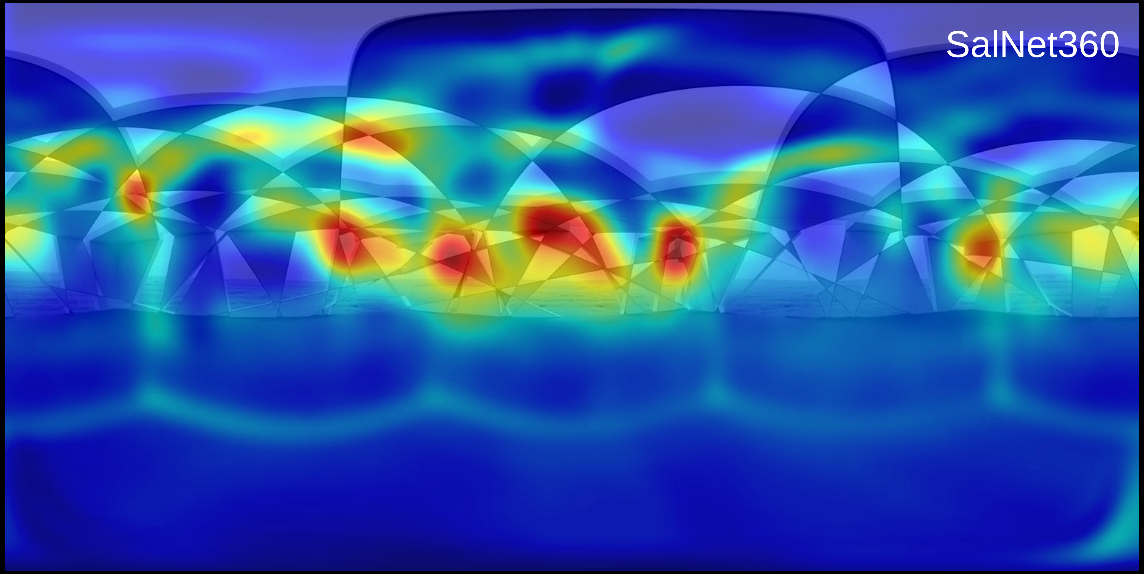


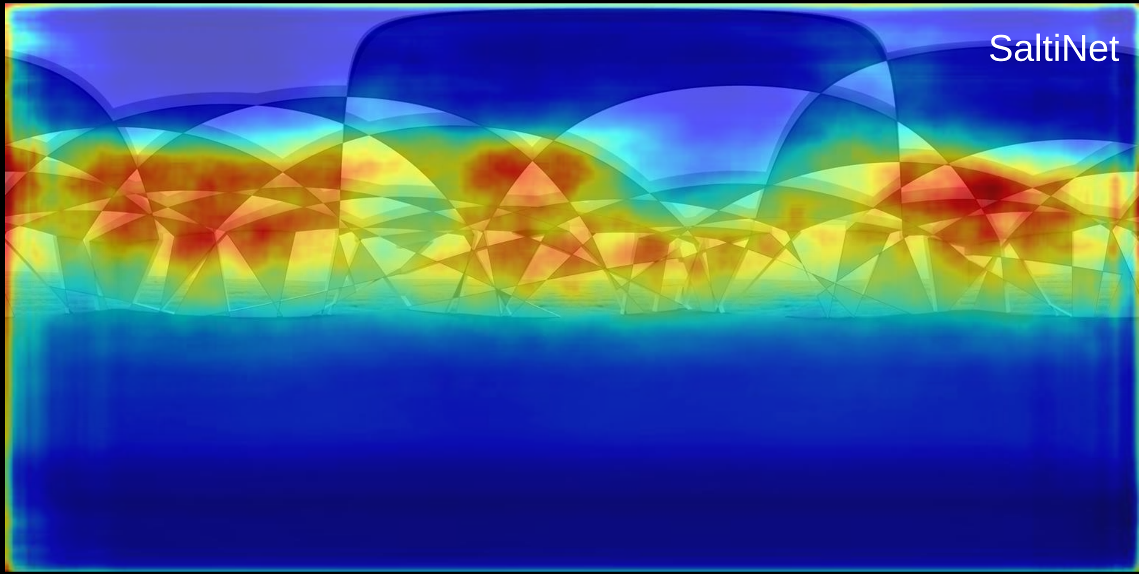
Saliency prediction vs. ground truth



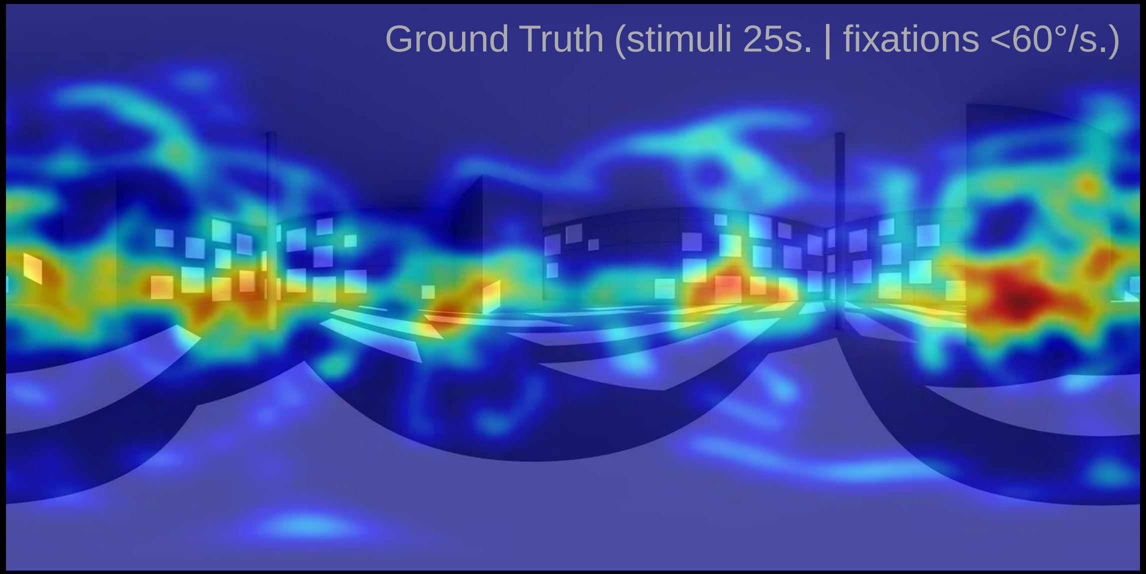
Serpentine pavilion | Overcast



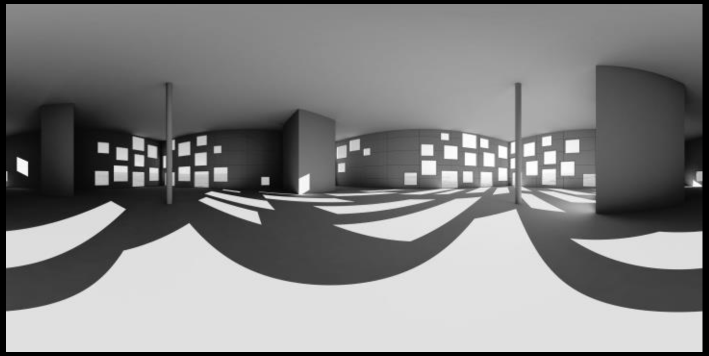


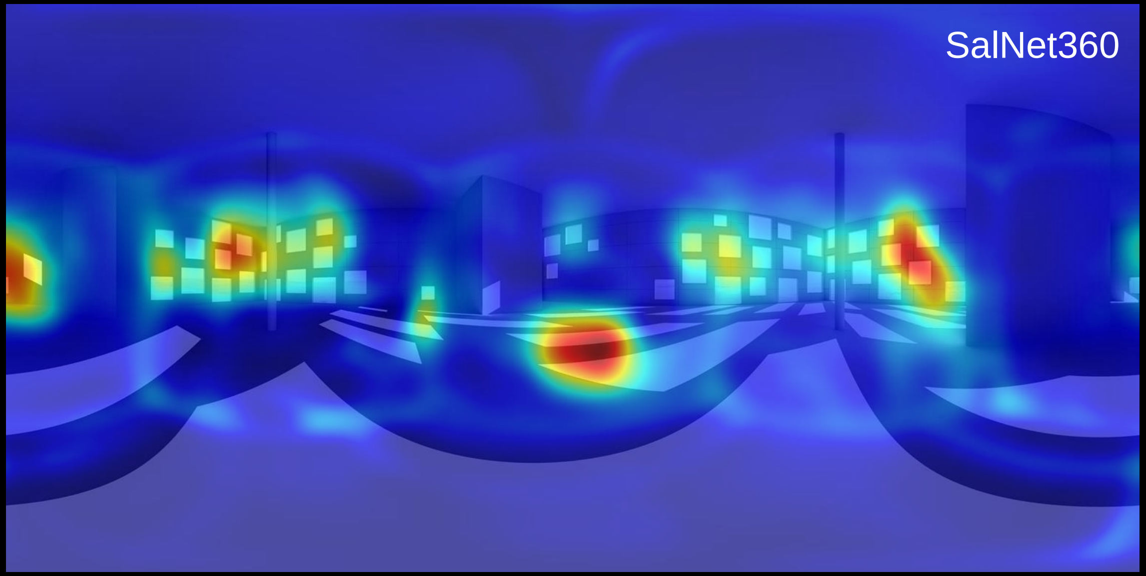


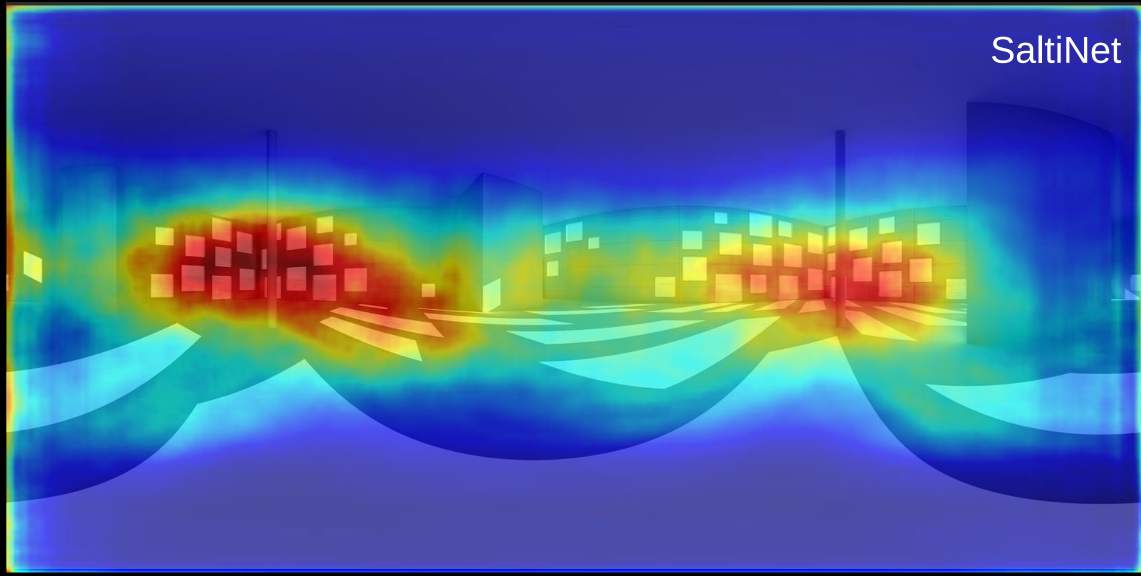
Saliency prediction vs. ground truth



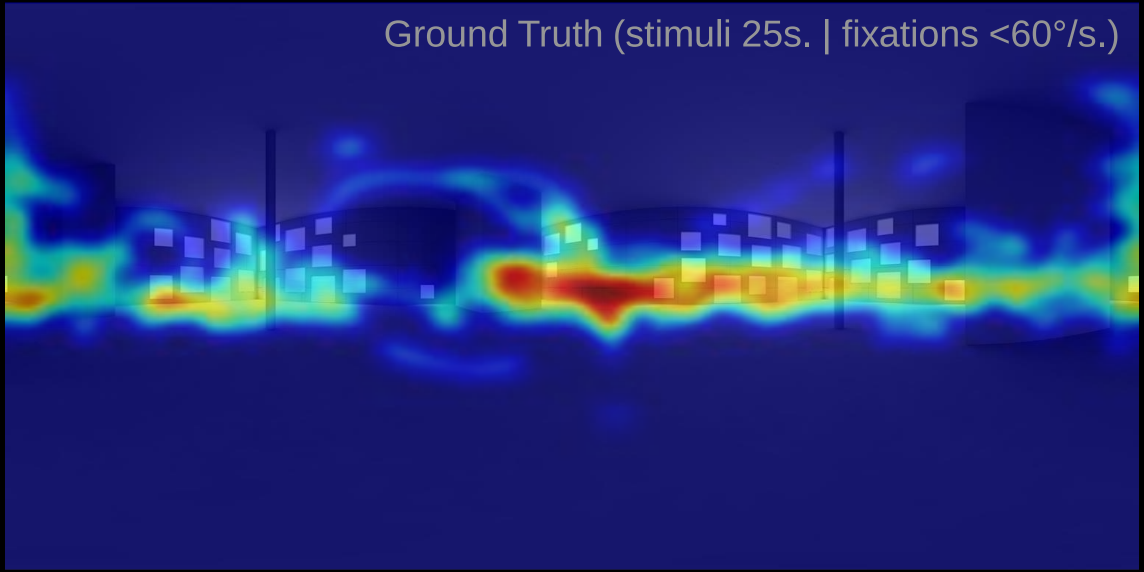
Zollverein | Clear



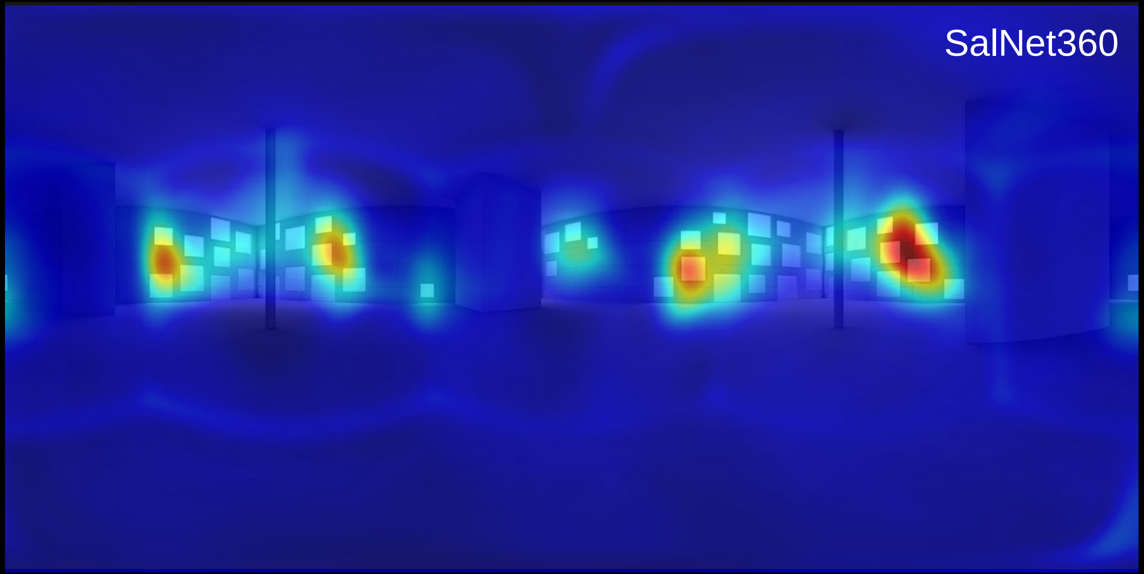


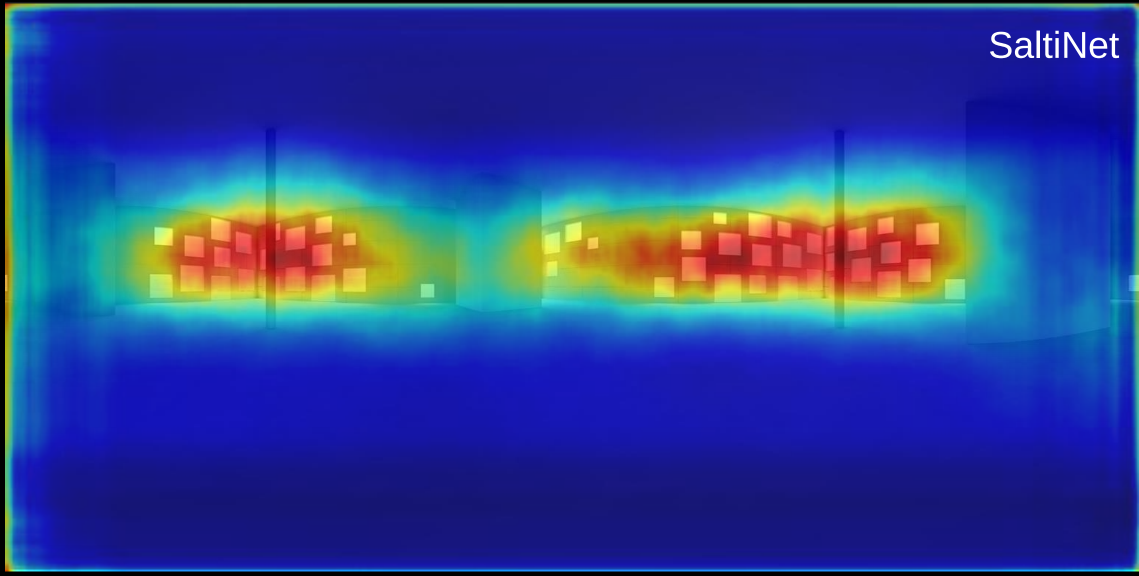


Saliency prediction vs. ground truth



Zollverein | Overcast





Insights and limitations

Qualitative analysis

SalNet360 finds **salient spots**
SaltiNet identifies **larger zones**

Work in different ways

Both model embed equator bias
Artefacts not corrected (e.g., cubes, border)
Low-level visual features identified

Quantitative analysis

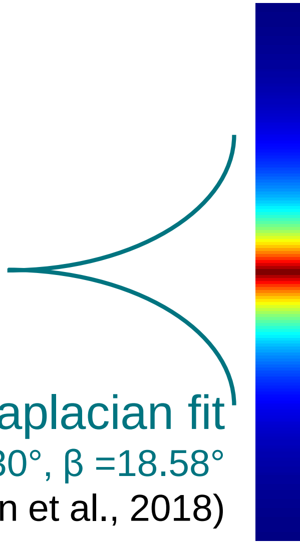
Distribution-based statistical metrics

Linear correlation coefficient
Range: [-1;+1] (abs. [0;+1])
1 means perfect correlation

Mean $CC_{SalNet360}$	= 0.25	(0.03 - 0.44)
Mean $CC_{SaltiNet}$	= 0.46	(0.23 - 0.58)
Mean $CC_{Laplacian}$	= 0.66	(0.26 - 0.80)

Kullback-Leibler divergence
Range: [0;inf.]
0 means density functions are equal

Mean $KL_{SalNet360}$	= 3.60	(2.04 - 5.99)
Mean $KL_{SaltiNet}$	= 2.58	(1.30 - 4.35)
Mean $KL_{Laplacian}$	= 1.77	(0.63 - 2.83)



Laplacian fit
$\mu = -1.30°$, $\beta = 18.58°$
(Sitzman et al., 2018)

Sitzman et al., 2018, Saliency in VR: How do people explore virtual environments? IEEE TVCG

Insights and limitations

How are saliency algorithms trained?

How does this compare to our data?

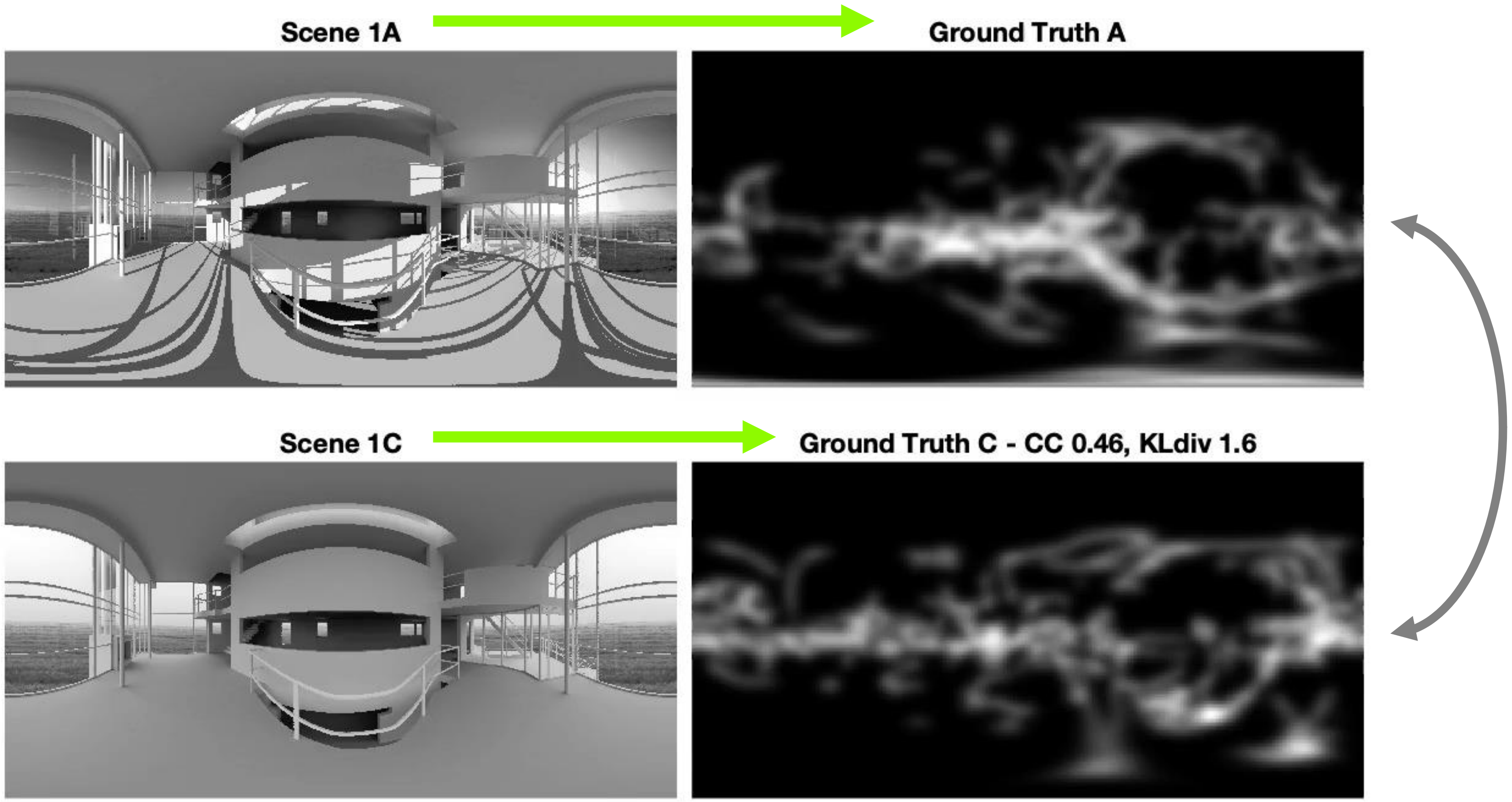
EPFL

Effect of sky conditions

Does our viewing behavior change with sky conditions?

EPFL

Comparing ground truth output for different sky conditions



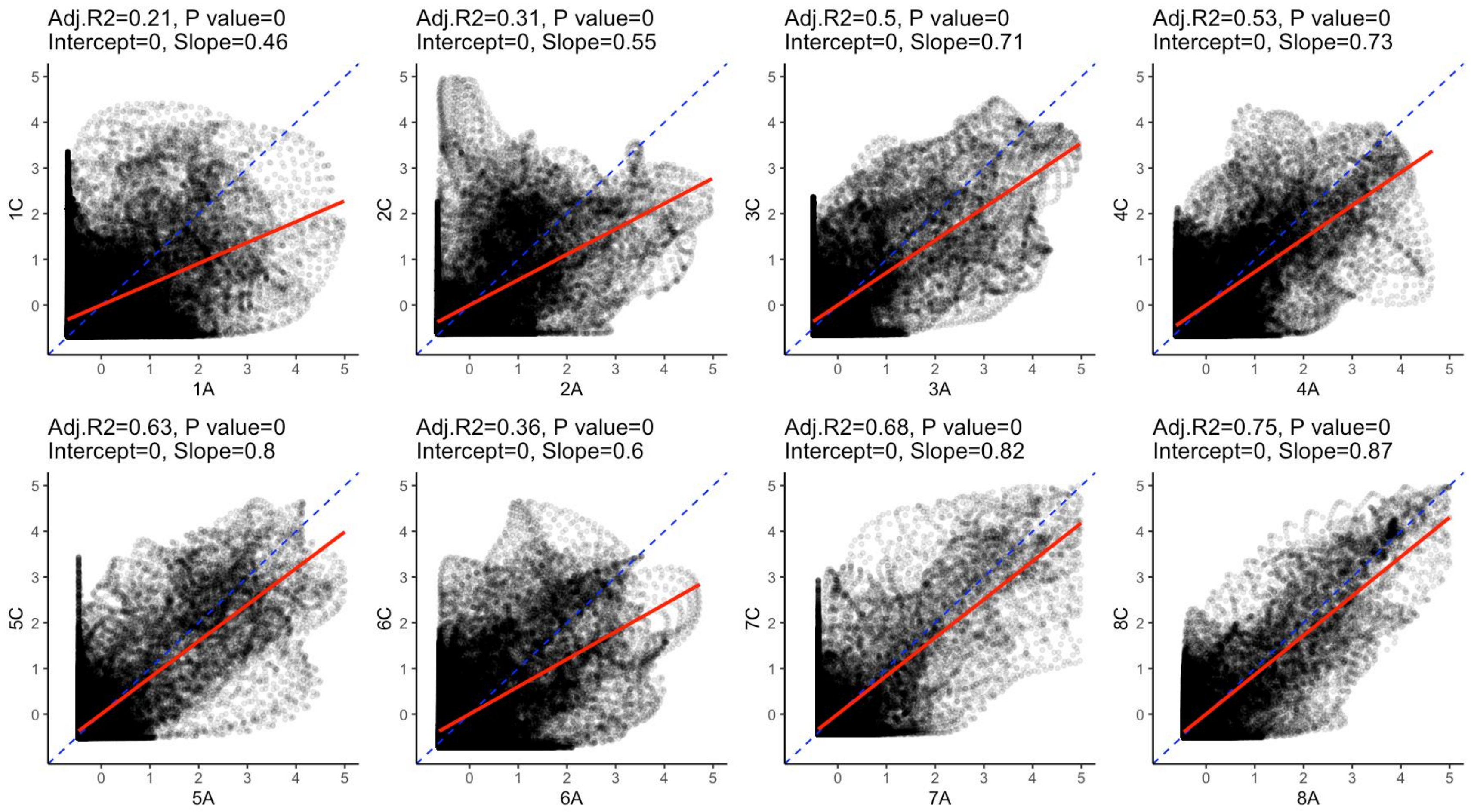
Clear vs. **overcast**
(same scene)

$$\text{Mean } CC_{\text{Clear-Overcast}} = 0.69 \quad (0.46 - 0.87)$$
$$\text{Mean } KL_{\text{Clear-Overcast}} = 1.96 \quad (1.17 - 2.97)$$

Viewing patterns presents similarities despite changing sky conditions

Comparing ground truth output for different sky conditions

Linear model output and plots



Higher correlations for the most 'horizontal' scenes

Conclusions

Existing saliency models could not accurately predict visual attention in rendered black and white architectural scenes

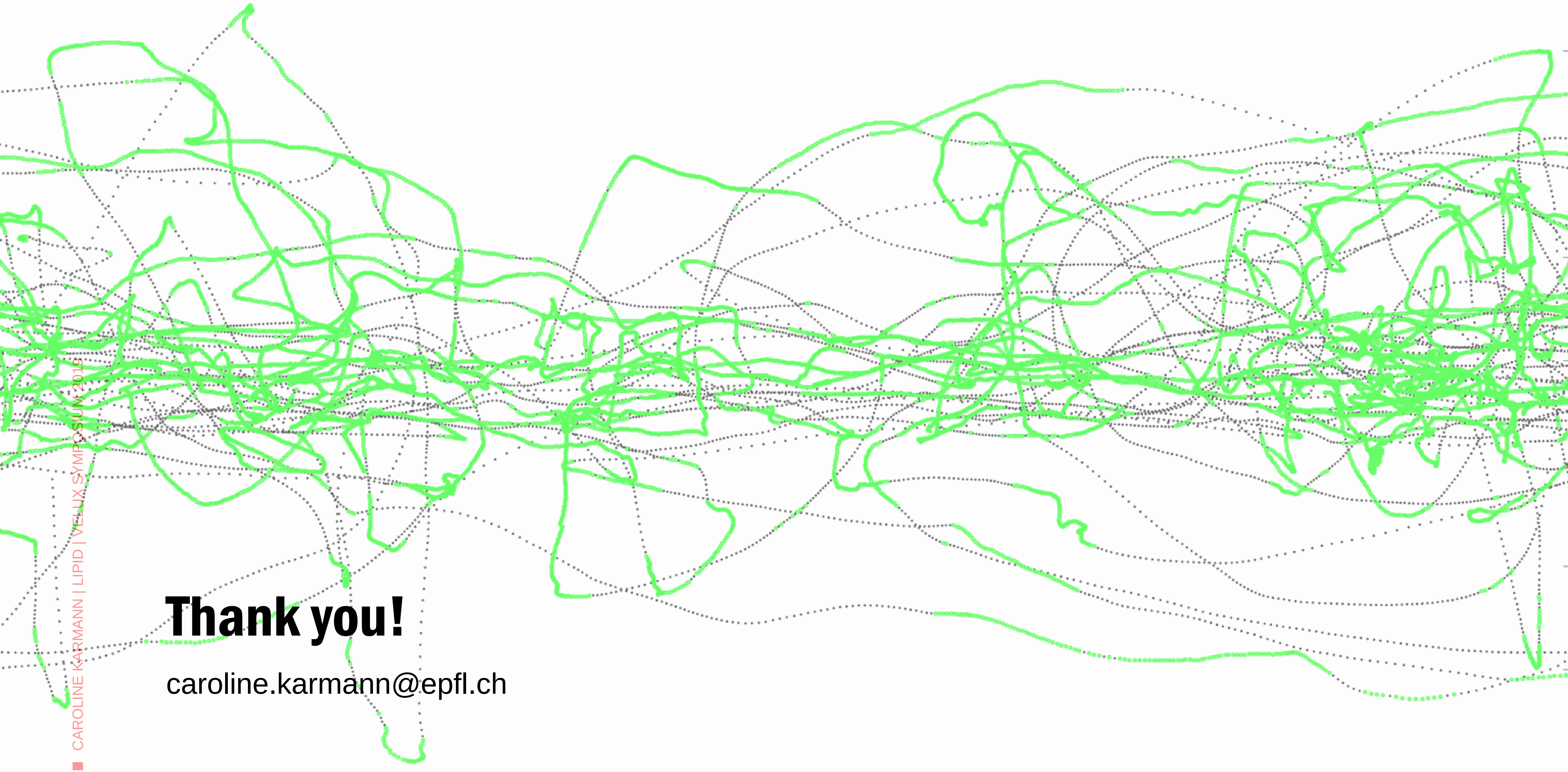
Good insights to be gained

Discrepancies in the experimental protocol (limitation)
→ Adapt our protocol if we like to further test/use saliency models

Tendency to look outside

Viewing patterns in a space remains somehow consistent despite varying sky conditions

Validation under real conditions (missing)



Thank you!

caroline.karmann@epfl.ch

EPFL

Acknowledgments

Data collection

Kynthia Chamilothori
Assistant Professor at the Eindhoven University of Technology (TU/e)
Laboratory of Integrated Performance in design, EPFL (previously)

Siobhan Rockcastle
Assistant Professor at University of Oregon
Laboratory of Integrated Performance in design, EPFL (previously)

Data processing and expertise

Seungryong Kim
Image and Visual Representation Lab, EPFL

Evgeniy Upenik
Multimedia Signal Processing Group, EPFL