Improving saliency models' predictions of the next fixation with humans' intrinsic cost of gaze shifts

- Supplementary Material -

Florian Kadner Tobias Thomas David Hoppe Constantin A. Rothkopf
Centre for Cognitive Science & Institute of Psychology, TU Darmstadt
{firstname.lastname}@tu-darmstadt.de

S1. Optimization of the model parameters

S1.1. Hyperparameters for the Limited-memory BFGS-B algorithm

All optimizations in this paper were done with the L-BFGS-B algorithm, and specifically it's *scipy*[1] implementation. The following hyperparameters were used:

- Maximum number of variable metric corrections: $m_{\text{cor}} = 10$
- Tolerance limit for stopping criterion $\frac{f^k - f^{k+1}}{\max(|f^k|,|f^{k+1}|,1)} \leq f_{\text{tol}} \cdot \varepsilon$ where ε is the machine precision: $f_{\text{tol}} = 10^{-7}$
- Tolerance limit for stopping criterion $\max |\text{proj}(g_i)| i = 1, ..., n <= p_{\text{tol}}$ with $\text{proj}(g_i)$ the i-th component of the projected gradient: $p_{\text{tol}} = 10^{-5}$
- Gradient step size: $\epsilon = 10^{-8}$
- Maximum number of function evaluations: $n_{\text{fun}} = 15000$
- Maximum number of iterations: $n_{\text{iter}} = 15000$
- Maximum number of line search steps per iteration: $n_{\text{ls}} = 20$

S1.2. Estimated parameters

Below are the concrete values for all estimated parameters, for all models. Table S1 shows the parameters for the experiments, where ϕ_i values where estimated for each model individually and Table S2 for the experiments with fixed ϕ_i values. The ϕ_i values for the second experiment where the weighted averages from the first one.

	w_1	w_2	σ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
DeepGaze II	0.345	2.893	34.158	1.737	2.087	2.022	2.462	3.319	3.376	4.744	5.219	5.218	4.374
SAM-ResNet	0.007	0.003	93.337	0.410	0.097	0.031	0.165	0.201	0.237	0.407	0.333	0.952	-2.17
EML-NET	0.095	0.481	18.296	0.155	0.790	0.427	0.748	1.081	1.104	1.449	-0.22	2.553	4.523
CASNet II	0.157	0.851	22.328	0.580	1.408	1.142	1.419	1.930	1.608	2.592	1.616	3.185	5.331

Supplementary Table S1. Estimated model parameters with individual exploration values.

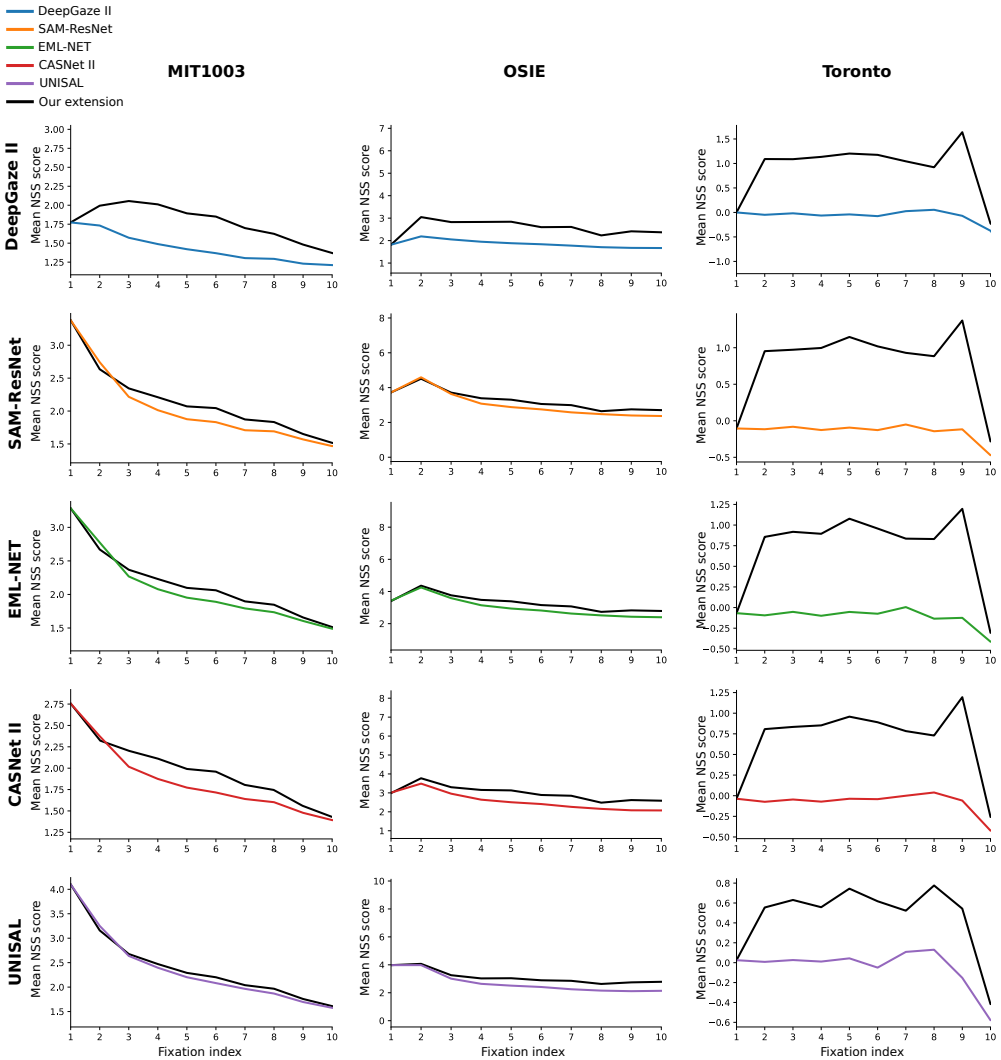
[1] https://docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html#optimize-minimize-lbfgsb

	w_1	w_2	σ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
DeepGaze II	0.351	1.989	33.632	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
SAM-ResNet	0.110	0.510	26.742	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
EML-NET	0.095	0.619	21.553	0.720	1.095	0.906	1.198	1.633	1.581	2.298	1.737	2.977	3.014
CASNet II	0.160	1.134	25.961	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
UNISAL	0.061	0.483	12.643	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Supplementary Table S2. Estimated model parameters with fixed exploration values.

S2. Quality of the predictions depending on the ordinal position in the gaze sequence

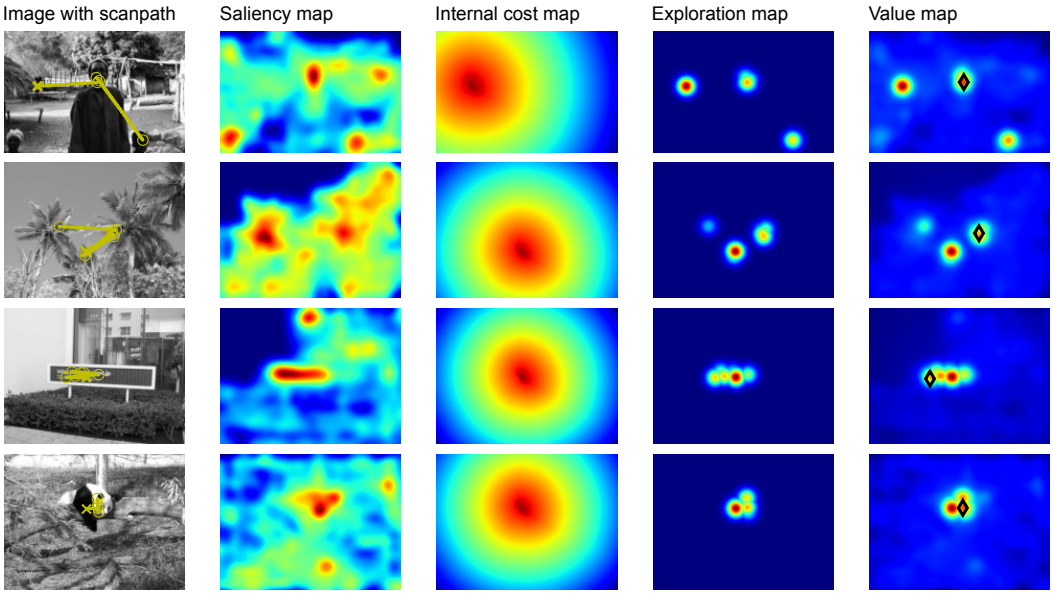
Here we show a detailed comparison between all baseline models and our one-step ahead extension on all three datasets, based on the ordinal position in the scanpath. Only the first 10 fixations were included because for positions larger than 10, the number of fixations was very low. The first position is always equal because our model uses only saliency information when there is no prior fixation information available.



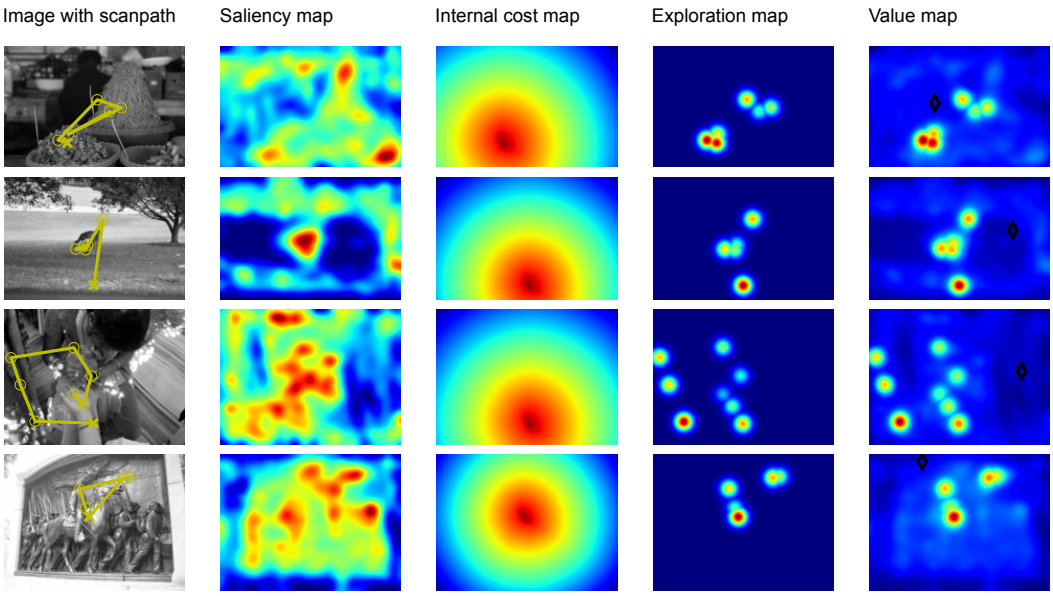
Supplementary Figure S1. NSS scores for the one-step ahead prediction depending on the ordinal position in the gaze sequence.

S3. Example predictions with best and worst NSS

4 of the best and worst single fixations, with respect to NSS score of the OSIE/MIT 1003 dataset. Shown is the image, with the scanpath up to that point (left), all three internal parts of our model (middle), and the resulting value map with the next fixation highlighted (right).



Supplementary Figure S2. Examples of predictions of the next fixations with highest NSS score.



Supplementary Figure S3. Examples of predictions of the next fixations with lowest NSS score.

S4. Three-step ahead predictions

Similar to Table 1, we evaluated also the three-step ahead predictions of the model, using only information up to timestep $t-3$. Now our model is only marginally better than the respective baseline and is even worse on some datasets/baseline model combinations.

	MIT 1003		OSIE		Toronto	
	AUC	NSS	AUC	NSS	AUC	NSS
DeepGaze II	0.844	1.506	0.906	1.867	0.497	-0.031
Our extension	**0.850**	**1.662**	0.885	1.784	**0.565**	**0.394**
SAM-ResNet	0.864	2.222	0.905	3.088	0.477	-0.105
Our extension	0.857	**2.271**	0.886	2.813	**0.557**	**0.317**
EML-NET	0.864	2.255	0.902	3.050	0.490	-0.073
Our extension	0.864	**2.304**	0.867	2.851	**0.564**	**0.307**
CASNet II	0.860	1.993	0.898	2.587	0.515	-0.059
Our extension	**0.861**	**2.062**	0.888	2.406	**0.576**	**0.354**
UNISAL	0.889	2.612	0.890	2.755	0.542	0.020
Our extension	0.885	**2.646**	0.887	**2.786**	**0.582**	**0.243**

Supplementary Table S3. Evaluation results. AUC and NSS scores for the three-step ahead prediction of gaze targets based on sequential value maps compared to the respective saliency model's baseline.