FastSal: a Computationally Efficient Network for Visual Saliency Prediction

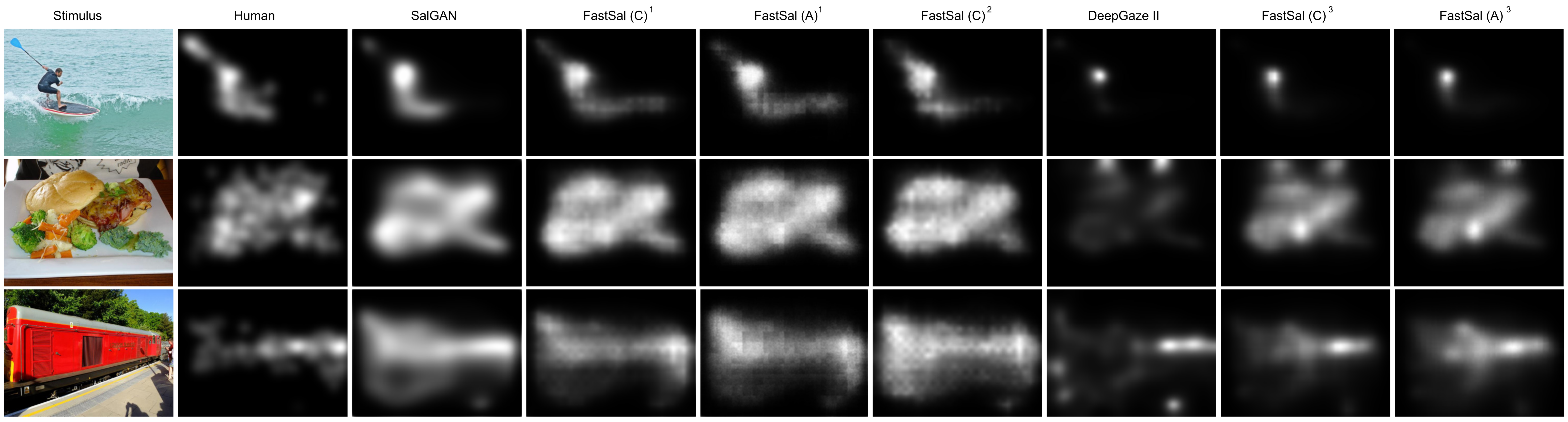
Feiyan Hu and Kevin McGuinness

Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

A World Leading SFI Research Centre

Science Foundation Ireland sfi For what's next



Introduction

Efforts to create effective models of human visual attention have made rapid progress. Early approaches, largely based on engineered features and heuristics, have been replaced by more accurate and more complex deep neural models. Effective computational visual attention models have many potential applications. For example, visual saliency has been used to improve the performance of image retrieval engines, provide cues to improve object detection, used in video survellience systems, video compression and image retargetting etc.

Computational constrains of SoA saliency models

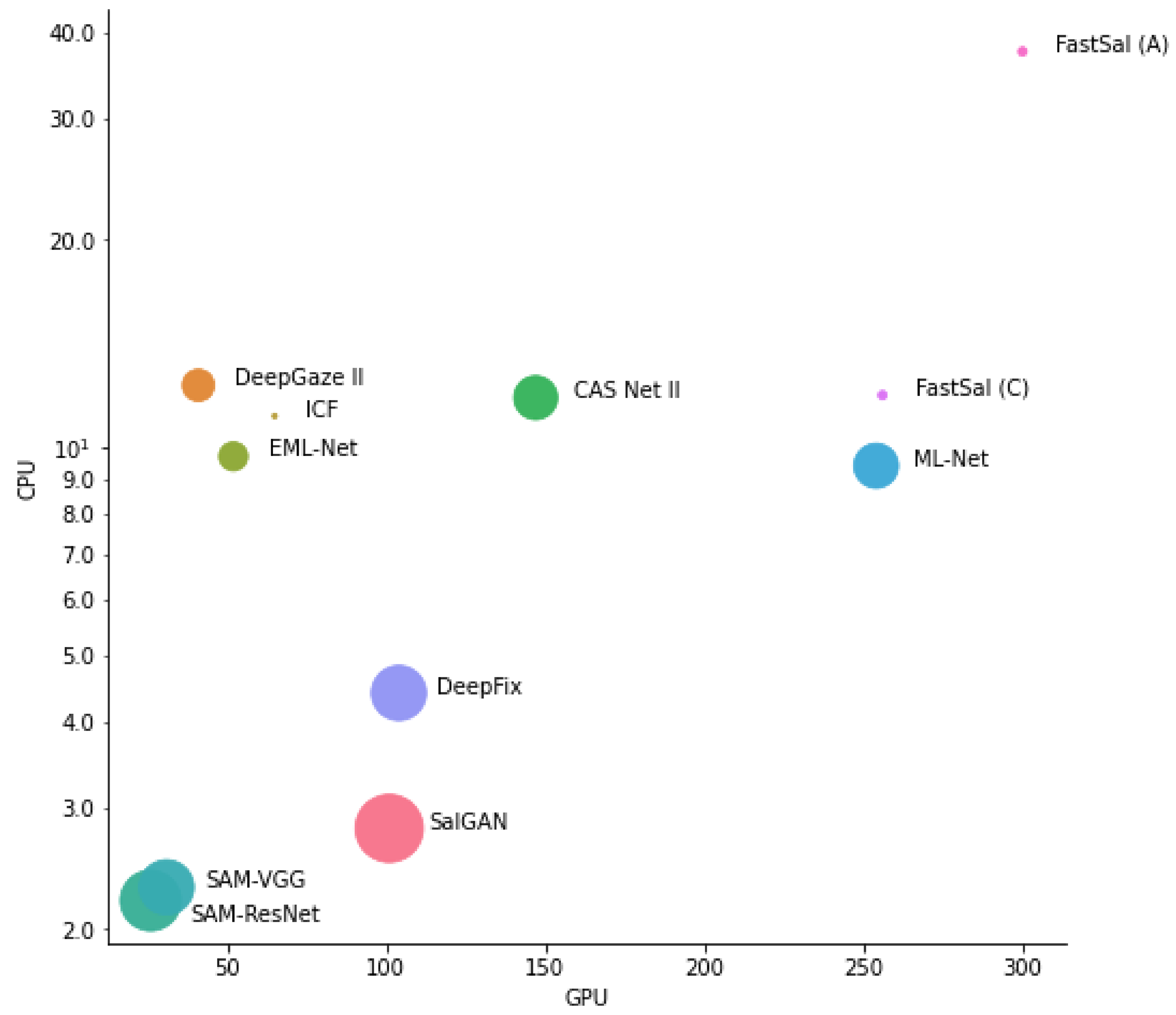
One of the most promising applications of computational visual attention models is in their potential to reduce the complexity of scene analysis and speed up downstream tasks. This requirement has been identified as one of the key motivations of many works on computational visual attention. Unfortunately, many existing deep saliency models are as computationally demanding, if not more so, than the subsequent visual analysis steps which has limited their adoption in downstream tasks.

	GFLOPs	Size (M)	Output size
SalGAN	91.46	31.92	192 × 256
DeepGaze II	20.22	20.44	192 × 256
EML-Net	16.24	47.09	192 × 256
CASNet-II	37.62	42.01	6 × 8
SAM-ResNet	72.9	70.04	192 × 256
SAM-VGG	59.4	51.83	192 × 256
ML-Net	39.42	15.45	24 × 32
DeepFix	59.82	28.73	24 × 32
FastSal (C)	**1.32**	**2.57**	192 × 256
FastSal (A)	**1.32**	3.65	192 × 256

FastSal Architecture

- Light weight MobileNet V2 Backbone as encoder to extract image representation.
- Hierarchical features adaptation and aggregation.
- Concatenation and Addition version of light weight decoder to aggregate hierarchical features.
- Knowledge distillation to learn better intermediate representations from teachers in pretrain process.
- Binary cross entropy for each pixel is used to compute loss.
- Open source code and weights are available on Github.

We tested the 2 versions of FastSal on CPU and GPU along with other SoA approached. FastSal addition version can achieve more than 30FPS on CPU and 260FPS on GPU.



HOST INSTITUTIONS

DCU · NUI Galway OÉ Gaillimh · UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh · UCD Dublin

PARTNER INSTITUTIONS

Maynooth University · Tyndall National Institute · Trinity College Dublin · University of Limerick

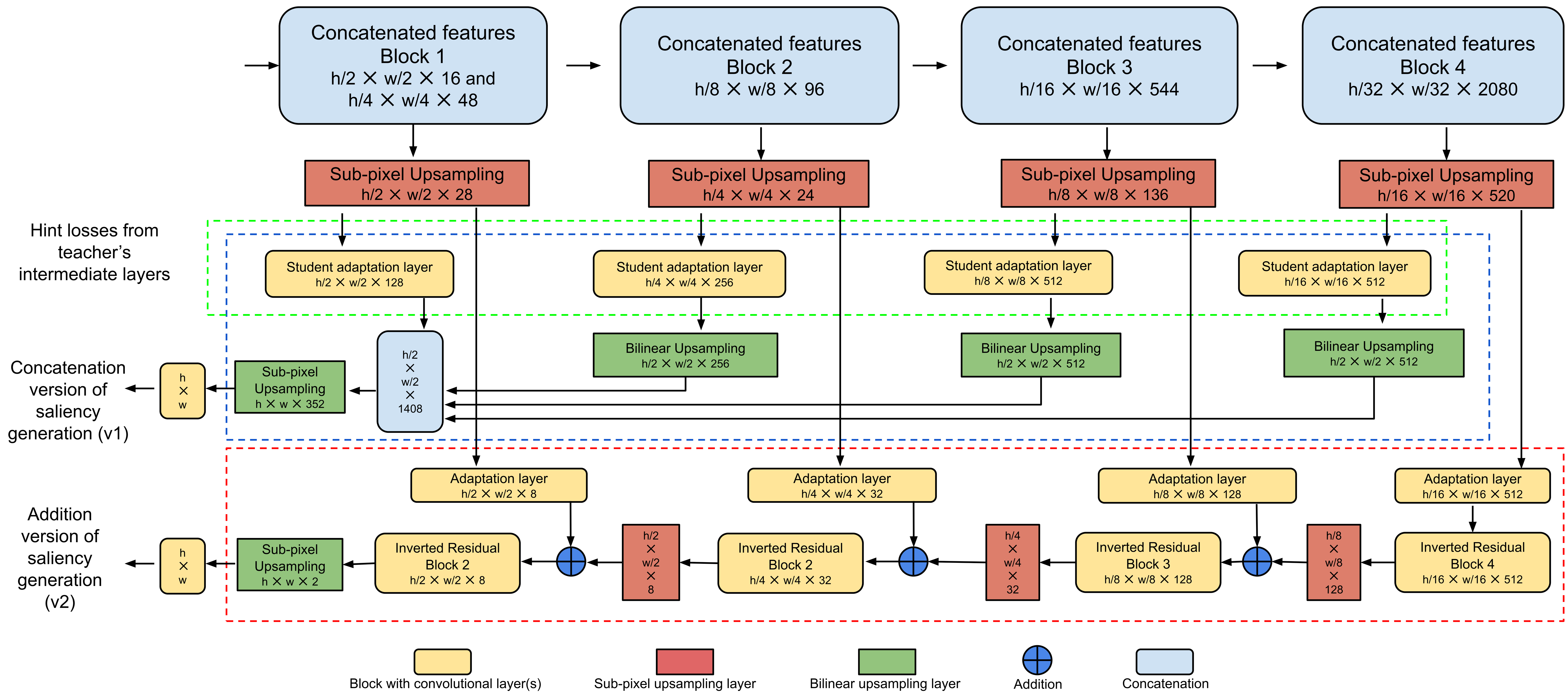
FUNDED BY:

European Union European Regional Development Fund · Science Foundation Ireland sfi For what's next

This work has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/15/SIRG/3283 and SFI/12/RC/2289 P2.

FastSal: a Computationally Efficient Network for Visual Saliency Prediction

Feiyan Hu and Kevin McGuinness



Comparing performance with SoA methos on MIT300 and SALICON 2017.

	AUC↑	sAUC↑	NSS↑	CC↑	KLDiv↓	IG↑
EML-NET	0.8762	0.7469	2.4876	0.7893	0.8439	N/A
DeepGaze II	0.8733	0.7759	2.3371	0.7703	0.4239	0.9247
GazeGAN	0.8607	0.7316	2.2118	0.7579	1.3390	N/A
SAM-ResNet	0.8526	0.7396	2.0628	0.6897	1.1710	N/A
SalGAN	0.8498	0.7354	1.8620	0.6740	0.7574	N/A
ML-Net	0.8386	0.7399	1.9748	0.6633	0.8006	N/A
FastSal (C)[1]	0.8635	0.7261	2.1158	0.7448	0.7086	N/A
FastSal (C)[2]	0.8684	0.7701	2.1913	0.7507	0.4665	0.8355

Table: Performance comparison on MIT300 (FastSal concatenation version).

	AUC↑	sAUC↑	NSS↑	CC↑	KLDiv↓	SIM↑	IG↑
EML-NET	0.866	0.746	2.050	0.886	0.520	0.780	0.736
SAM-ResNet	0.865	0.741	1.990	0.899	0.610	0.793	0.538
SalGAN	0.864	0.732	1.861	0.880	0.288	0.772	0.775
FastSal (C)	0.863	0.732	1.845	0.874	0.288	0.768	0.770
FastSal (A)	0.862	0.731	1.828	0.870	0.291	0.764	0.760

Table: Performance comparison on the SALICON 2017 test set.

Comparing performance with different backbones.

	sAUC↑	bAUC↑	NSS↑	CC↑	hint loss↓
EfficientNet (b0)	0.7275	0.8421	1.7623	0.8579	0.0617
MobileNetV2	**0.7365**	**0.8450**	**1.8163**	**0.8751**	**0.0526**

Table: Results on SALICON 2017 (validation) with different backbones.

Conclusion

We proposed FastSal, a new fast saliency model suitable for inference on constrained computing devices. The proposed model is significantly smaller than other state-of-the-art with only 2.57×10^6 parameters, which amounts to less than 10MB uncompressed single precision floating point memory. The computational complexity of the model, approx 1.32×10^6 FLOPs for a 192×256 image, is orders-of-magnitude lower than comparable state-of-the-art (e.g. $45\times$ lower than DeepFix) while remaining competitive with top models on most metrics. Tests on a downstream task show that the model can be used in place of more complex models like SalGAN without deteriorating performance.

This work has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/15/SIRG/3283 and SFI/12/RC/2289 P2.