KDSalBox:
A toolbox of efficient knowledge-distilled saliency models

Ard Kastrati 1&2, Zoya Bylinskii 2, Eli Schechtman 2
1: ETH Zurich, Department of Electrical Engineering, Distributed Computing Group, 2: Adobe Research

Motivation

Dozens of saliency models have been designed over the last few decades, targeted at diverse applications ranging from image compression and retargeting to robot navigation, surveillance, and distractor detection. Barriers to their use include the different and often incompatible software environments that they rely on, as well as the computational inefficiency of older implementations.
To facilitate the evaluation and selection of saliency models for different applications, we present KDSalBox - a toolbox of 10 knowledge-distilled saliency models. KDSalBox allows these 10 models to be efficiently run, compared, and be practically applied.

Results:

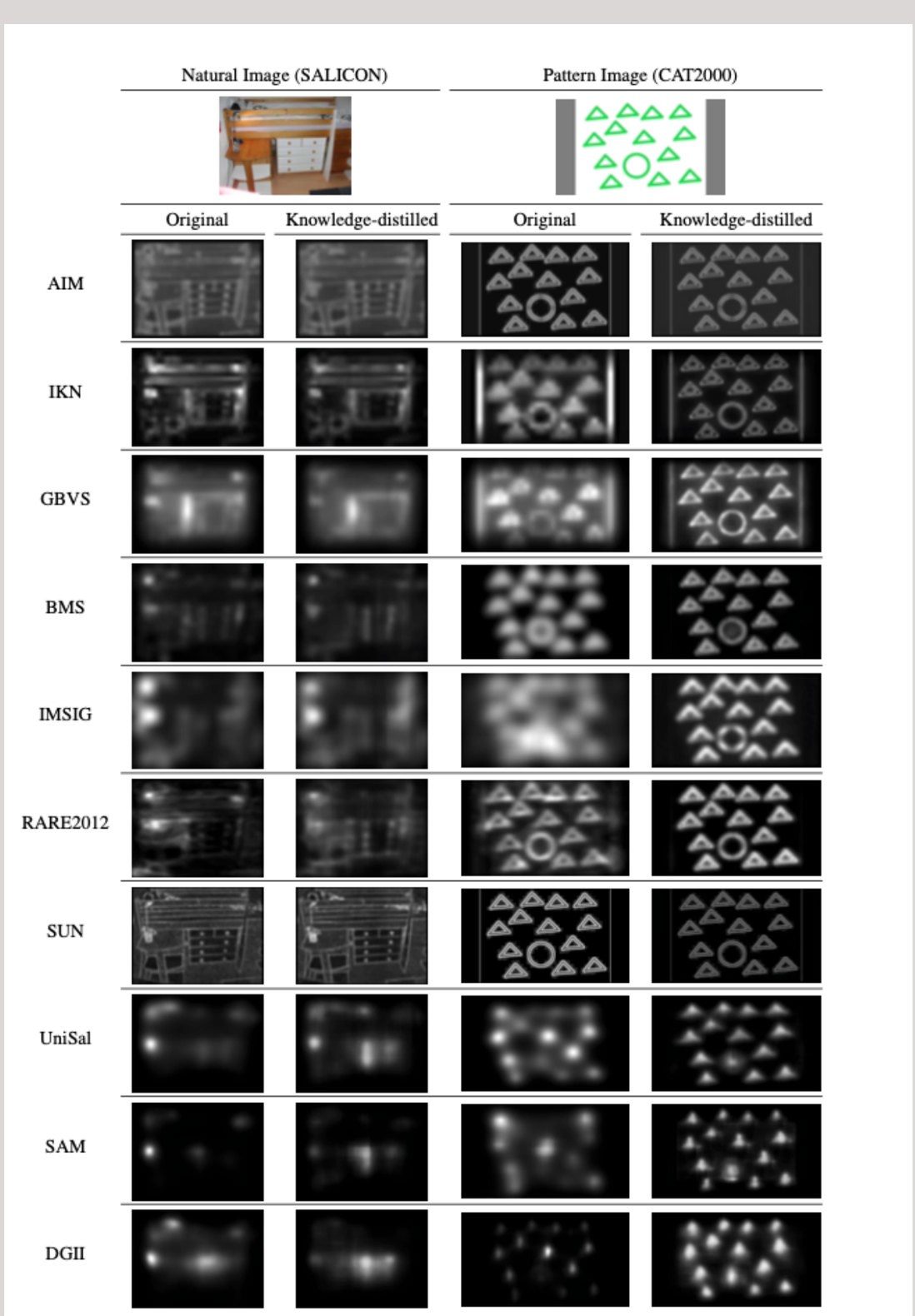


Figure: The performance of ten knowledged-distilled models in our KDSalBox shown for a sample image from both datasets: SALICON and CAT2000

Model	Original implementation			(Dis)similarity between student & original models			
				Natural Images (SALICON)		Patterns (CAT2000)	
	Codebase	Size	Time (CPU)	CC ↑	KL ↓	CC ↑	KL ↓
AIM	Matlab	24MB	10sec	0.99 ±0.00	0.00 ±0.00	0.96 ±0.04	0.18 ±0.23
IKN	Matlab	1MB	6.0sec	0.97 ±0.01	0.02 ±0.01	0.78 ±0.09	0.14 ±0.06
GBVS	Matlab	1MB	6.1sec	0.98 ±0.01	0.01 ±0.01	0.75 ±0.09	0.17 ±0.07
BMS	Matlab	5MB	0.4sec	0.89 ±0.07	0.05 ±0.02	0.78 ±0.11	0.30 ±0.26
IMSIG	Matlab	20KB	5.7sec	0.96 ±0.04	0.02 ±0.02	0.71 ±0.15	0.16 ±0.13
RARE2012	Matlab	20KB	6.3sec	0.87 ±0.10	0.10 ±0.06	0.76 ±0.12	0.30 ±0.16
SUN	Matlab	1MB	12sec	0.94 ±0.04	0.02 ±0.01	0.95 ±0.02	0.21 ±0.25
UniSal	Python	30MB	0.4sec	0.92 ±0.05	0.11 ±0.06	0.68 ±0.14	0.46 ±0.36
SAM	Python	561MB	7.3sec	0.82 ±0.13	0.28 ±0.17	0.48 ±0.15	0.87 ±0.40
DGII	Python	330MB	4.8sec	0.90 ±0.08	0.14 ±0.09	0.68 ±0.16	0.48 ±0.42

KDSalBox

Compute Saliency

Upload an image to compute the saliency.
You can use the graphical user interface to compute the saliency of any image for all 10 knowledge-distilled models, you can perform histogram matching and other post-processing operations. You may also provide ground truth saliency data for the image to get each of the models scored.

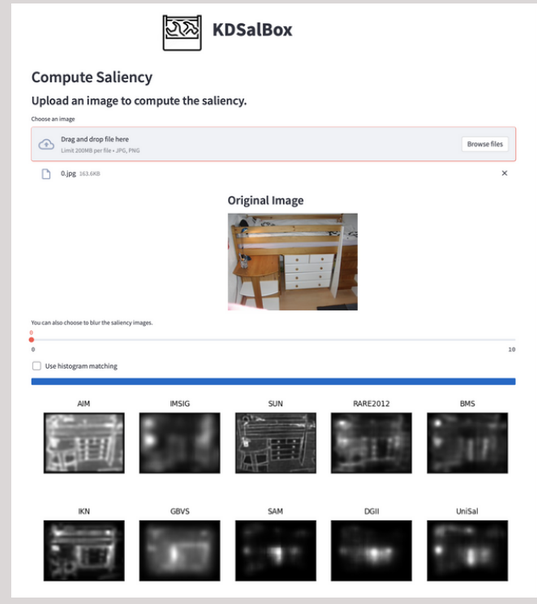
Task Evaluation

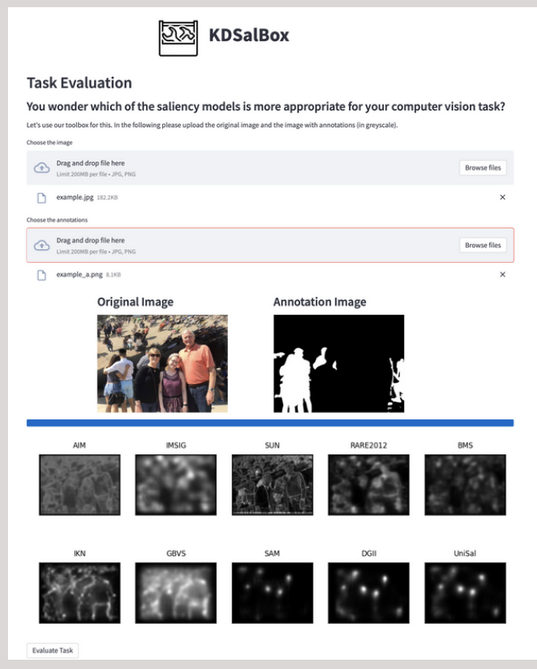
Which of the saliency models is best suited for your computer vision task?
You can use our toolbox to upload an image and corresponding annotations (as a mask). KDSalBox will evaluate the given models for the task in terms of Precision, Recall, Accuracy and F1 score. Use this to find which of the saliency models does best at detecting the main subject, image distractors, etc.

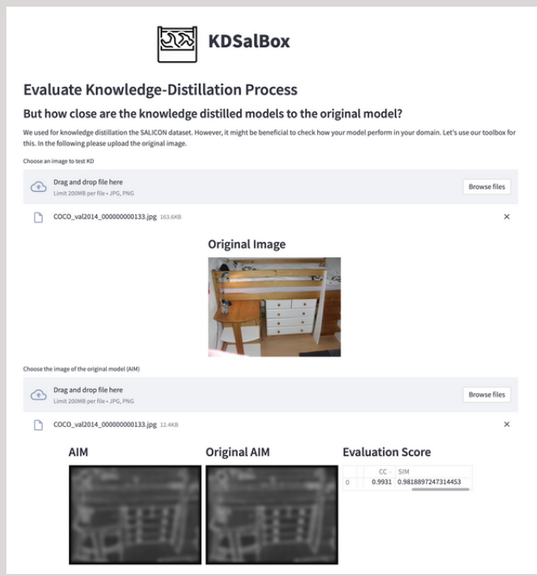
Knowledge-Distillation

How close are the knowledge distilled versions to the original models? We knowledge distilled the models using the SALICON dataset. However, you might want to check how your model performs on the target domain.

User Interface







Table(left): The ten models currently part of the KDSalBox. We include scores comparing the knowledge-distilled student models to the original saliency model implementations. We report the average inference time by using the SMILER (Wloka et al. (2018)) framework and 640×480 sample images. We note that in comparison to the original implementations, all our resulting student models are 41.2 MB in size with an average inference time of 0.4 sec on CPU and 0.01 sec on GPU.

Outlook

In our work, we have used a unified architecture for all models for simplicity and easier maintainability.The current limitation is that our models are trained only on natural images of the SALICON dataset, and the included technical report provides an analysis of how they perform on out-of-distribution datasets, including 20 different image categories from CAT2000 and 5 graphic design categories from Imp1K. We recommend referring to these results before applying a model to a new task/dataset. Extensions of this work call for training on other domains and with larger datasets.