

SMILER: Saliency Model Implementation Library for Experimental Research

Calden Wloka¹, Toni Kunić¹, Iuliia Kotseruba¹, Ramin Fahimi², Nicholas Frosst³, Neil D.B. Bruce⁴,
and John K. Tsotsos¹

¹Department of Electrical Engineering and Computer Science, York University, Toronto, Canada

²Department of Computer Science, University of Manitoba, Winnipeg, Canada

³Google Brain, Toronto, Canada

⁴ Department of Computer Science, Ryerson University, Toronto, Canada

¹{calden, tk, yulia, tsotsos}@eecs.yorku.ca, ²ramin@cs.umanitoba.ca, ³frosst@google.com,
⁴bruce@ryerson.ca

Abstract

The Saliency Model Implementation Library for Experimental Research (SMILER) is a new software package which provides an open, standardized, and extensible framework for maintaining and executing computational saliency models. This work drastically reduces the human effort required to apply saliency algorithms to new tasks and datasets, while also ensuring consistency and procedural correctness for results and conclusions produced by different parties. At its launch SMILER already includes twenty three saliency models (fourteen models based in MATLAB and nine supported through containerization), and the open design of SMILER encourages this number to grow with future contributions from the community. The project may be downloaded and contributed to through its GitHub page: <https://github.com/tsotsoslab/smiler>

1. Introduction

Many aspects of modern scientific research are heavily dependent on software. This dependence raises a number of challenges, including the fact that software developed primarily for research is often difficult or time consuming to set up and execute, and may include undocumented assumptions, parameters, or conflicting requirements which present a major impediment to research sharing and reproducibility [23, 4]. The field of saliency research is an area in which many of these challenges may be seen: over the past two decades there has been a dramatic increase in both the number and nature of computational saliency models.

Not only does this volume make it increasingly difficult for researchers to effectively explore and test the landscape of different approaches to saliency modeling, the lack of a standard interface to each model increases the likelihood that any given model may be incorrectly or erroneously configured, leading to mistaken or inconsistent results in the saliency literature.

For example, Table 1 shows scores computed by the similarity (SIM) metric [35] as calculated in three different studies (Vig *et al.* [55], Wang and Shen [56], and Berga and Otazu [5]) on the Toronto dataset [10]. Note that not only do the scores not match for even a single algorithm across any two studies, but also the rank order of performance shifts. For instance, Vig *et al.* find that their eDN model [55] outperforms the CAS model [24] both with and without added center bias, whereas Wang and Shen find that CAS outperforms eDN. Similarly, Vig *et al.* find that with center bias added, AWS [22] outperforms AIM [10] which outperforms GBVS [28], and without center bias added the order shifts from best to worst to be GBVS, AIM, AWS. Neither ordering, however, agrees with the results of Berga and Otazu, who find a ranking of best to worst for these three models to be GBVS, AWS, and then AIM.

Note that we are not accusing any authors of impropriety or misconduct, but rather are simply highlighting that without standardization three different studies give rise to three different sets of scores and rankings. Why this happens is best explained through an example: several saliency algorithms have been shown to operate best over colour spaces alternative to RGB, including the Covariance-based Saliency (CVS) model [18], the Image Signature (IMSIG) model [30], and the Saliency Detection by Self-

Resemblance (SSR) model [50] which all operate best in the CIELAB colour space, and the Quaternion-based Spectral Saliency (QSS) model [49] which performs best using YUV colour. However, the original model code released by each method’s authors handles image input completely differently. CVS expects as input a string argument specifying an image path, then loads the image and converts it to CIELAB space internally. SSR expects as input an image variable in RGB format, which is converted to CIELAB space internally. IMSIG expects as input an image variable in RGB format, which is converted to CIELAB or DKL colour space when provided with an optional parameter setting (when no parameter is provided, IMSIG will process the image in RGB, despite the recommendation of the authors to use CIELAB space). QSS expects an image variable as input, but provides no internal image conversions; the model authors recommend that YUV format be used, and the conversion is expected to be performed by the user before calling the QSS model code. No one approach is any more correct than any other, but this lack of standardization places a non-negligible burden on users and can easily lead to errors or oversights in which a user believes they have configured the models to operate in the desired colour spaces but some subset of models is actually not being applied as expected. When one takes into account the number of additional parameters which must be controlled (such as numerical scaling of saliency map output, post-processing smoothing, the application of a center prior, or any model-specific settings), the burden of use and chance for error is only further compounded. Likewise, for any hope of reproducibility, these parameters must all be exhaustively documented (which, unfortunately, has not always been the case within the literature).

Recent years have also seen a shift in model development toward methods which rely on deep learning networks. While many of these methods achieve very high benchmark performance, they also introduce a new practical challenge for the dissemination and sharing of code. In order to operate in a reasonable timeframe, most deep learning algorithms require a significant share of their computation to take place on a graphical processing unit (GPU). This necessitates that a user not only has access to GPU hardware, but also has the appropriate libraries installed which will allow access to the GPU for calculations. Unfortunately, the setup processes of different GPU scientific computing libraries as part of the same development environment are often fairly involved and daunting for non-experts. Likewise, there is a lack of standardization, and frequently the libraries necessary to run one model will be incompatible with the libraries required for another model. For example, oSALICON [53] is implemented in Caffe [34], while DeepGaze II [36] is implemented in TensorFlow [3]; as of the time of this writing, following the official setup docu-

mentation for one project interferes with the setup of the other. While running both libraries on the same system is possible, it requires knowledge that goes beyond the official documentation. Therefore, not only will the installation of even a single model likely be a large barrier of entry for a user who is not actively pursuing work with deep learning-based development, but also providing simultaneous access to a general-purpose library of saliency models is extremely difficult without isolating incompatible model dependencies from each other. Due to potentially frail assumptions regarding backward compatibility, there is a significant risk that important contributions may be lost to time or not explored in sufficient detail owing to the need to operate within a specific ecosystem.

Our work aims to facilitate research efforts in computational saliency by addressing these software challenges. We do so by introducing the Saliency Model Implementation Library for Experimental Research (SMILER). SMILER provides library-like functionality for saliency models, standardizing the input, output, and parameter specifications for each model, and isolating incompatible model components from each other. At the time of this publication SMILER supports twenty three models: Attention by Information Maximization (AIM) [10], Adaptive Whitening Saliency (AWS) [22], Boolean Map Saliency (BMS) [61], Context Aware Saliency (CAS) [24] based on an open implementation [54], Covariance-based Saliency (CVS) [18], DeepGaze II (DGII) [36], Deep Visual Attention Prediction (DVAP) [56], Dynamic Visual Attention (DVA) [31], Ensemble of Deep Networks (eDN) [55], Fast and Efficient Saliency Detection (FES) [52], Graph-based Visual Saliency (GBVS) [28], Intensity Contrast Features (ICF) [37], the Itti-Koch-Niebur Saliency Model (IKN) [33], Image Signature (IMSIG) [30], Learning Discriminative Subspaces (LDS) [20], a Deep Multi-Level Network (MLNet) [15], an open implementation [53] of Saliency in Context [32] (oSALICON), Quaternion-Based Spectral Saliency (QSS) [49], RARE2012 [47], Saliency Attentive Model (SAM) [16], Saliency Detection by Self-Resemblance (SSR) [50], Saliency using Generative Adversarial Networks (SalGAN) [45], and Saliency Using Natural statistics (SUN) [62]. This set provides a broad representative sample of models popular in the saliency research community focused on fixation prediction, and the system is designed to be easily extensible with additional models.

1.1. Related Work

The rapid expansion in saliency model numbers has been met in at least one area by a concerted effort at consolidation and standardization: performance benchmarking. Starting with a number of isolated benchmark surveys (*e.g.* see [7, 8, 35, 47]), this effort eventually culminated with the establishment of the MIT Saliency Benchmark [12], a con-

Reference	[55]	[56]	[5]
eDN [55]	0.573/0.487	0.40	-
CAS [24]	0.555/0.427	0.44	-
AWS [22]	0.558/0.407	-	0.352
GBVS [28]	0.534/0.496	0.49	0.397
AIM [10]	0.549/0.426	0.36	0.314
IKN [33]	-	0.45	0.366

Table 1: An example of inconsistent model results. Here we show the SIM [35] scores for five algorithms over the Toronto dataset [10] as reported by three recent publications (note that [55] report two scores, one with added center bias and one without). - indicates that a model was not run in that particular study. Note that we are not claiming any wrong-doing on the parts of these studies, but rather pointing out that each study likely executed these models in slightly different ways, leading to inconsistent results and a substantial challenge for reproducibility in the literature.

tinually updated ranking of saliency algorithm performance over a pair of curated benchmark datasets.

While these benchmarking efforts have provided an important overview of progress in the field of fixation prediction and an impartial ranking of models, the scope of this effort has remained predominantly focused on comparative performance evaluation. The MIT Saliency Benchmark does helpfully provide an index of links to code which has been released by model authors, but the onus of handling setup and operation of each different model’s code remains with the user. SMILER, therefore, provides a complementary service to the saliency community; rather than focus on standardizing evaluations of performance, SMILER seeks to standardize the execution of model code and thereby enable exploration of additional research avenues not encapsulated by current benchmarks.

It should be noted that the current collection of models supported by SMILER consists of models which focus on pixel-wise assignment of conspicuity values and which have been predominantly applied to the domain of human fixation prediction. There are, however, other branches of saliency model research, such as salient object detection (*e.g.* see [14] for an early example, and [26] for an overview and recent survey). Likewise, the models included are predominantly focused on saliency prediction over static scenes, but there is nevertheless significant interest in saliency over dynamic stimuli (*e.g.* see [42, 40, 60]). This focus on models which are more representative of fixation prediction over static images is not intended to dismiss or ignore these other research avenues, but rather is meant to form a solid base for the SMILER platform.

1.2. Our Contributions

SMILER provides two primary benefits to the saliency research community: reducing the burden of use for code execution, and promoting the consistency and reproducibility of experimental results. The first step in achieving these goals is the establishment of a common model-independent application programming interface (API). In order to make this API as effective as possible in facilitating a wide range of research, we ensure the following qualities:

- It should be possible to run each algorithm in a default mode which requires minimal user input or selection of settings, providing an intuitive mode which can be used without expert-level algorithmic familiarity. At the most basic level of function, a model should expect only that an input image is specified, and it should return as output a saliency map corresponding to that image. By default, this saliency map should be the same height and width as the input image.
- As much as possible, there should be no loss in the flexibility of parameter options available for each individual model. While it is not possible to have a common set of parameters for each algorithm, to reduce the complexity of operation as much as possible it should be possible to selectively choose which parameters to manually specify, with unspecified parameters automatically populated with default values (thereby allowing for a smooth transition from fully default mode through to fully user-specified operations).

By creating a standard interface for model execution, we allow users to learn a single API rather than one for each model. The flexible method for parameter specification allows researchers to engage with models at a variety of levels of depth, from novel benchmarking work using default settings through to the analysis of model behaviour over a range of parameter settings.

By standardizing model execution, we also ensure that when researchers run a given model with particular settings, they are sure to get the same results as when another researcher runs the same model with the same settings. If both researchers were expected to independently set up and execute the model using their own custom scripts, it is entirely possible for unintentional bugs or oversights to lead to inconsistencies between them. Of course, it is entirely possible for SMILER to contain bugs, but by fostering an open and centralized repository for saliency model code, we ensure that when bugs are found and corrected this correction is distributed to all users.

With a straightforward and flexible code base for easily executing a large ecosystem of saliency models, we envision a number of research directions which SMILER can support, including but not limited to:

- Performance benchmarking on applications outside of fixation prediction for which saliency may be applicable, but extensive performance testing is not currently available. Examples include:
 - Anisotropic image or video compression (e.g. [17, 58, 25, 27])
 - Defect detection (e.g. [41, 6])
 - Image cropping (e.g. [51]) or retargeting (e.g. [63])
 - Image domains outside the natural images which form the bulk of fixation datasets, such as websites [43] or satellite imagery [21]
 - Image quality assessment (e.g. [39, 59, 38])
 - Robotic navigation (e.g. [13, 48]) or search (e.g. [46])
- Saliency model evaluation for other attentional aspects beyond fixation prediction, such as the psychophysical evaluations proposed in Bruce *et al.* [11].
- Increasing the robustness of conclusions for research which compares experimental findings in psychology or neuroscience to saliency algorithms (e.g. [44, 29, 57]) by allowing comparison against many saliency models rather than a single one.

2. Design Overview

In order to leverage as wide a range of existing saliency model implementations as possible, as well as to support researchers with different degrees of computational and software resources available to them, SMILER is comprised of two major programming language components: a MATLAB component and a command-line interface (CLI) implemented in Python. The MATLAB component comes with a subset of available models and is fully cross-platform so long as the computer supports the MATLAB environment and the user has access to the appropriate licenses for MATLAB and any specific toolboxes required by a given model. The CLI is currently only supported for the Linux operating system, but provides access to the full suite of SMILER models, both MATLAB and deep learning. In order to foster open software development and move away from proprietary software systems, the CLI will be the primary focus of future development for the SMILER project, with an emphasis on adding models that do not depend on a MATLAB license. To minimize code drift across multiple interfaces, all models included in SMILER contain a configuration and information file described in Section 2.1.

Prior to the shift in algorithm development toward deep learning models, the majority of saliency models were released for the MATLAB programming environment. As a consequence, early development of SMILER was also based in MATLAB. However, the need to handle deep learning models which are predominantly implemented in

languages other than MATLAB necessitated a shift to another language. Nevertheless, it was felt that it would not be desirable to drop the MATLAB specific structure which is already in place, as there are many users who would prefer to operate within the MATLAB environment (for example, many researchers are already familiar with MATLAB through the use of tools such as the PsychToolbox [9], and may prefer to keep their research efforts in the same programming environment). Therefore, the design of SMILER retains MATLAB functionality for all algorithms available in the MATLAB environment, as well as a functional MATLAB interface for executing these models. The SMILER CLI utilizes the MATLAB's Python API to allow invocation of MATLAB models in the background, without using the full MATLAB graphical user interface.

Whether one is working through MATLAB or the SMILER CLI, the general principles of SMILER operation remain the same, and the details of operation are kept as close as possible given the different nature of the MATLAB Integrated Development Environment (IDE) and the CLI. An overview of operation for the MATLAB interface is given in Section 2.2, and for the SMILER CLI in Section 2.3. Due to the more extensive support of saliency models and support for YAML-based experiment specification (discussed more thoroughly in Section 2.3), we would encourage users to preferentially use the CLI.

SMILER attempts as closely as possible to maintain the originally intended functionality of each model. However, there are times when this is not possible. For example, although expecting the output map to be the same height and width as the input image seems like a straightforward assumption, it is not the default behaviour of all algorithms. Some models automatically resize input images to a specified size, and return this size as output, whereas others such as SUN [62] make a point of returning only the portion of the image for which output is valid without image padding (trimming the half-width of the feature kernels from the image border). Although this inconsistency between input and output size may be a distinct choice by the model designers with a clear justification, for the purposes of SMILER it was felt that ensuring a common behaviour across algorithms was the more important consideration and therefore SMILER will re-scale or pad as appropriate the saliency maps to be the same dimensions as the original input image.

Models for which the full source code has been released are preferred candidates for inclusion in SMILER, as this allows for more robust crowd-sourced bug checking, access to the full range of algorithm parameters (particularly for post-processing steps such as smoothing), and aids in future code maintenance (for example, the use of deprecated functions which are no longer supported by MATLAB or third party libraries). It should be noted, however, that several

models are nevertheless included despite only having access to a pre-compiled version, namely AWS [22], FES [52], and RARE2012 [47]. The pre-compiled version of CAS [24] provided by the original study authors is not compatible with SMILER, and therefore an open source implementation [54] has been used. In a similar vein, code for the SALICON model [32] is not available at the time of this writing, but we include the oSALICON [53] implementation which is based on the original model.

2.1. A Common Format for Information and Configuration

There are a number of parameters and controls for pre- and post-processing of saliency maps which are common to many or all models. As well, each model in SMILER requires several important attributes to be associated with it, including citation information and model-specific parameters. In order to provide this information in a manner which is extensible to the inclusion of future model properties or specifications and independent of the specific programming interface accessing the model, several JavaScript Object Notation (JSON) configuration files are used. JSON is an open standard for providing human-readable attribute-value pairs, and provides an effective format for storing model information in SMILER.

Parameters which affect the execution of a majority of SMILER models are referred to as *global* and are described in a `config.json` file in the root of the SMILER directory. These parameters and their default values are shown in Table 2. Listing 1 shows an example JSON parameter specification. The user shouldn't need to modify these JSON files directly, as they contain specifications for the SMILER system. The user should specify parameters at run-time via the SMILER CLI or MATLAB interface, or via YAML experiment files.

Listing 1: The global parameter dictionary contained in `config.json`.

```
"do_smoothing": {
  "default": "default",
  "description": "Specification for post-
  processing smoothing.
  'default' uses whatever smoothing step
  is provided by the originally released
  model code.
  'none' turns off post-processing
  smoothing (though it should be noted
  that some implicit smoothing (eg.
  through image resizing) will remain).
  'custom' smooths the map with a
  specified kernel.
  'proportional' smooths the map with a
  kernel sized to the major dimension of
  the image.",
```

```
  "valid_values": ["default", "none", "
  custom", "proportional"]
}
```

As can be seen, parameters are defined as a nested dictionary. Each parameter is populated with three fields: `default`, `description`, and `valid_values`. The `default` field is used when no other source of parameter specification is available. The `description` and `valid_values` fields are intended for human consumption; each SMILER interface provides a method for accessing and displaying this information to a user (detailed in Section 2.2 for the MATLAB interface and Section 2.3 for the CLI). The `description` field provides a brief explanation for the role the parameter plays in the calculation of a saliency map, while the `valid_values` field provides either an explicit set of available parameter assignments (e.g. for the `scale_output` parameter there are three options: `min-max`, `none`, or `normalized`) or a specified range (e.g. an “Integer greater than 0” for `smooth_size`).

Listing 2: An example `smiler.json` file showing model-specific information for the AIM algorithm. [10]

```
{
  "name": "AIM",
  "long_name": "Attention by Information
  Maximization",
  "version": "1.0.0",
  "citation": "N.D.B. Bruce and J.K.
  Tsotsos (2006). Saliency Based on
  Information Maximization. Proc. Neural
  Information Processing Systems (NIPS)",
  "model_type": "matlab",
  "model_files": [],
  "parameters": {
    "AIM_filters": {
      "default": "21jade950.mat",
      "description": "The feature filter
      set to be used by the AIM algorithm.
      In the form [size][name][info], where
      each filter is size by size in
      dimension, name is the ICA algorithm
      used to derive the filters, and info
      provides a measure of the retained
      information (higher numbers
      correspond to more filters).",
      "valid_values": [
        "21infomax
        [900,950,975,990,995,999].mat",
        "21jade950.mat",
        "31infomax[950,975,990].mat",
        "31jade[900,950].mat"
      ]
    }
  }
}
```

Parameter	Default Value	Valid Values	Description
color_space	“default”	“RGB”, “gray”, “YCbCr”, “LAB”, “HSV”	Specification for pre-processing conversion of the image color channels.
center_prior	“default”	”default”, ”none”, ”proportional_add”, ”proportional_mult”	Specification for the addition of a priori bias toward the center.
center_prior_prop	0.2	Float between 0 and 1.	Specifies the sigma of the Gaussian as a proportion of image dimensions. Only used when center_prior is set to proportional_add or proportional_mult.
center_prior_weight	0.5	Float greater than 0.	Specifies the weight given to the prior in the form of multiplicative gain. Only used when center_prior is set to proportional_add or proportional_mult.
center_prior_scale_first	true	Boolean.	If true, first scale the saliency map to be between 0 and 1 before adding the prior.
do_smoothing	“default”	“default”, “none”, “custom”, “proportional”	Specification for post-processing smoothing.
smooth_size	9	Integer greater than 0.	Custom smoothing kernel size, only used when do_smoothing is set to custom.
smooth_std	3.0	Float greater than 0.	Custom smoothing kernel standard deviation, only used when do_smoothing is set to custom.
smooth_prop	0.05	Float greater than 0.	Proportional smoothing kernel parameter, only used when do_smoothing is set to proportional.
scale_output	“min-max”	“min-max”, “none”, “normalized”	Specification for rescaling saliency map values into a specified range.
scale_min	0.0	Float less than scale_max.	Minimum saliency value in the map, only used when scale_output is set to min-max.
scale_max	1.0	Float greater than scale_min.	Maximum saliency value in the map, only used when scale_output is set to min-max

Table 2: Default values for global SMILER parameters, as defined in `config.json`.

Each model includes additional model-specific information in a `smiler.json` file included in the root of its subfolder. An example `smiler.json` file is shown in Listing 2.

As can be seen, each file contains information providing both the SMILER shortened designation for the model (in this case, AIM) as well as its full name and citation information. `model_type` allows the code to easily check whether pre-requisites are available to execute the code (for example, if the MATLAB engine is not installed, SMILER will skip MATLAB-based models with a warning rather than an execution error). `model_files` provides SMILER with a list of any files required for the model execution (e.g. network weights for a CNN-based model). Model-specific parameters are specified using the same system as the global parameters in `config.json`. Additionally, some models contain a `notes` field which includes human-readable information pertinent to the specific model (such as recommendations by the original model authors or additional in-

formation which may be of use to a user).

SMILER is programmed to take a flexible approach to parameter specification, populating parameter fields according to a priority order. This order is, from greatest to least precedence: user specified values (provided at runtime, or via YAML experiment file), model-specific default values (defined in model’s `smiler.json` specification), and global default values (defined in SMILER’s internal `config.json`).

2.2. Overview of MATLAB Interface

In order to help users navigate and use the code base provided by SMILER, a number of helper functions are provided. This section describes the supporting code base for the MATLAB portion of SMILER; the suite of tools which support the CLI are described in Section 2.3.

The primary helper file is the installation file, `iSMILER.m`. This file adds all other helper functions and all bundled MATLAB-based models to the MATLAB

path. By default, this installation will not save the changes to the path beyond the current session, but a user may optionally specify that path changes should be permanent by calling:

```
iSMILER(true);
```

Should users have permanently modified the path and later change their mind, SMILER path changes may be undone by using the `uninstall` function provided in the file `unSMILER.m`.

The function `smiler_info` provides a text interface in MATLAB for a user to query parameter information. This may be called without any arguments or using the string argument `'global'` to receive information about global parameters, or a specific MATLAB-based model may be specified as the input argument and the model-specific parameter and citation information for that model will be displayed.

In order to bring each included algorithm into compliance with the common API of SMILER, the code for each model is encapsulated in a wrapper function with the format `[model_name]_wrap.m`, where `[model_name]` is a string selected from the following available list of included algorithms:

- AIM: Attention by Information Maximization [10]
- AWS: Adaptive Whitening Saliency [22]
- CAS: Context Aware Saliency [24], using the implementation by [54]
- cG: A centered Gaussian prior
- CVS: Covariance-based Saliency [18]
- DVA: Dynamic Visual Attention [31]
- FES: Fast and Efficient Saliency [52]
- GBVS: Graph-Based Visual Saliency [28]
- IKN: The Itti-Koch-Niebur Saliency Model [33]
- IMSIG: Image Signature [30]
- LDS: Learning Discriminative Subspaces [20]
- QSS: Quaternion-Based Spectral Saliency [49]
- RARE2012: A multi-scale rarity-based saliency model [47]
- SSR: Saliency Detection by Self-Resemblance [50]
- SUN: Saliency Using Natural statistics [62]

Each function operates with the following function call:

```
output_map = [MODEL_NAME]_wrap(input_image,
                               params)
```

where the `input_image` is either a string specifying the file path of an image or is a variable containing image data, and `output_map` is a single-channel saliency map with the same height and width as the image specified by `input_image`. `params` is an optional input variable in the MATLAB structure format which provides a mechanism for specifying parameter values. As mentioned in Section 2.1, every model's behaviour is governed by a set

of parameters which are specified as key-value pairs. If no parameter structure is provided, the wrap function will automatically populate the parameter settings with default values appropriate for the given model. If some but not all parameters are specified in the input, then the wrap function will likewise operate with default values for any unspecified structure elements.

A basic example showing the explicit calculation of four models (AIM, AWS, IKN, QSS) on an input image specified by a path string is given in Listing 3.

Listing 3: Example MATLAB script showing the calculation of saliency maps for the AIM, AWS, IKN, and QSS models.

```
img_path = 'path/to/example.png';
AIM_map = AIM_wrap(img_path);
AWS_map = AWS_wrap(img_path);
IKN_map = IKN_wrap(img_path);
QSS_map = QSS_wrap(img_path);
```

This can be written more conveniently as a loop, iterating over the same set of models and executing each in turn and saving the saliency map as a separate image.

Listing 4: Example MATLAB script using SMILER to calculate saliency maps for the AIM, AWS, IKN, and QSS models.

```
models = {'AIM', 'AWS', 'IKN', 'QSS'};
img_path = 'path/to/example.png';
for i = 1:length(models)
    salmap = feval([models{i}, '_wrap'],
                  img_path);
    imwrite(salmap, [models{i}, '_saliency_map.png']);
end
```

Note that the code makes use of the MATLAB `feval` function to dynamically execute code based on a string argument, which allows for a simple interface for scripting and batch execution.

Sample 4 can be easily extended if specific parameters for some models are desired. For example, if a user wanted to use one of the other learned ICA filter bases for the AIM algorithm and wanted QSS to operate over the HSV colour space (but were otherwise fine with all other default parameters), then the modified version of the script shown in Listing 5 could be used.

Listing 5: Example MATLAB script using SMILER to calculate saliency maps for the AIM and QSS models with customized parameters.

```
models = {'AIM', 'AWS', 'IKN', 'QSS'};
```

```

img_path = 'path/to/example.png';

for i = 1:length(models)
    params = struct();
    switch(models{i})
        case 'AIM'
            params.AIM_filters = '21infomax999.
            mat';
        case 'QSS'
            params.color_space = 'hsv';
    end
    salmap = feval([models{i}, '_wrap'],
    img_path, params);
    imwrite(salmap, [models{i}, '_
    _saliency_map.png']);
end

```

Note that whether the parameter is model-specific or global, the method of user specification is the same (in this example the user specifies AIM's model-specific parameter `AIM_filters`, whereas for QSS it is the global parameter `color_space` which is specified).

All the above samples assumes that the `iSMILER` function has already been run, and therefore all wrapper functions are available on the MATLAB `path`. Additional examples are available as part of the SMILER GitHub repository.

2.3. Overview of SMILER CLI

Although the MATLAB interface is fully functional and supports all MATLAB-based models, the CLI is the recommended method of use, and future extensions to the SMILER library will likely be focused in this direction. Not only does this help migrate SMILER away from software requiring a proprietary license (MATLAB), but it also provides a more flexible platform for extension and experiment design which better supports protocol documentation.

The SMILER CLI is based on a core structure of functions which provide an interactive text-based interface to users. This includes commands to manage the containerized images for the available non-MATLAB models (see Section 3.2 for more details on model isolation and containerization), which at the time of this writing include the following models:

- BMS: Boolean Map Saliency [61]
- DVAP: Deep Visual Attention Prediction [56]
- DGII: DeepGaze II [36]
- eDN: Ensemble of Deep Networks [55]
- ICF: Intensity Contrast Features [37]
- MLNet: Deep Multi-Level Network [15]
- oSALICON: Open-source Saliency in Context [53], based on the original model by [32]
- SAM: Saliency Attentive Model [16]

- SalGAN: Saliency using Generative Adversarial Networks [45]

SMILER's CLI is designed to function in a Linux environment. The library is interacted with using commands with the following pattern:

```
smiler COMMAND [OPTIONS] [ARGS]
```

where [OPTIONS] and [ARGS] are command-specific options and arguments to modify program behaviour. The SMILER commands available are as follows:

- `clean`: Deletes downloaded files and docker images.
- `download`: Downloads model files and docker images.
- `info`: Provides information on SMILER models.
- `run`: Runs model(s) on images in a directory.
- `shell`: Runs a shell interface appropriate to the model environment.
- `version`: Displays SMILER version information.

Further information about the usage of any command can be obtained by appending the `--help` flag.

Although users may directly use the CLI to conduct experiments and generate saliency maps with SMILER, the CLI additionally supports experiment specification using YAML. This is the recommended method of operation, as it allows a user to maintain explicit records of experimental settings and protocols through stored YAML specification files.

YAML is a data serialization language designed to be easily written, read, and understood by humans. SMILER uses YAML files to specify experiments. These YAML specification files are composed of two sections: an `experiment`, which provides global specification details, and one or more `experimental runs`, which provide details for a specific algorithm call. An example is presented in Listing 6.

Listing 6: An example YAML specification file

```

experiment:
  name: Example 1
  description: An illustrative example of
  how to set up SMILER YAML experiments.
  input_path: /tmp/test_in
  base_output_path: /tmp/test_out
  parameters:
    do_smoothing: none

runs:
  - algorithm: AIM
    output_path: /tmp/AIM_smoothing
    parameters:
      do_smoothing: default

  - algorithm: AIM

```

```

output_path: /tmp/AIM_no_smoothing
- algorithm: DGII
- algorithm: oSALICON
  parameters:
    color_space: LAB

```

The name and description fields are primarily for user records, and facilitate organization and sharing of experimental protocols by providing a lightweight document which can easily be created and stored for each experiment conducted and run on any system with SMILER installed. `input_path` is the folder which contains the images to be processed in this particular experiment. `base_output_path` provides a root location for output maps to be saved, which by default will be placed in a subfolder at this location named for the algorithm that produced it (e.g. in listing 6 DGII and oSALICON will be saved in `/tmp/test_out/DGII` and `/tmp/test_out/oSALICON` respectively).

YAML specification introduces an additional layer to parameter precedence. The `parameters` field within the `experiment` field provides a way to set customized values which will be used for all runs, but these may be overridden for a specific run by adding a `parameters` field to that run. This is demonstrated in the example shown in Listing 6; all runs are set to be performed without smoothing based on the parameter specification under the `experiment` field, but the first run using AIM overrides this specification and instead uses default smoothing parameters. In this case, both AIM runs include `output_path` fields which will override the default behaviour using `base_output_path`. In the provided example, DGII will be run without any additional specifications beyond those provided in the `experiment` field, while oSALICON will be run with an additional specification of the `color_space` parameter (since there is no `color_space` specification under `experiment`, all other runs will use the built-in SMILER default: RGB).

3. Further Implementation Details and Requirements

3.1. MATLAB Implementation

In addition to the interface functions mentioned in Section 2.2, a number of support functions are provided which are used internally by either the function wrappers or user-level helper functions (such as image loading functions which will work with either a path specification or an array variable containing image data). These functions are primarily intended for SMILER’s internal use, and therefore users should not expect to interact with them directly. This includes the `jsonlab` toolbox [19] for interacting with the

configuration files.

We recommend using 2016a-2017b versions of MATLAB. Other versions may not fully support all available MATLAB models or SMILER code (for example, the pre-compiled AWS model, for which no source code is available, does not function in newer versions of MATLAB due to the deprecation of the `princomp` function).

3.2. Command Line Interface (CLI)

As mentioned in Section 1, deep learning libraries are not always compatible on the same system, which presents a challenge for executing deep learning-based saliency models which rely on incompatible libraries. To solve this issue, SMILER makes use of *containerization*, which is also known as *operating system level virtualization*. This method creates isolated user-space instances called *containers* that share the same OS kernel and drivers, but are otherwise separated. Thus, each model may be fully encapsulated within its own container, isolating any system-level libraries which may interfere with those used by other model implementations. In addition to granting this isolation, the container may be designed to provide a full specification of a model’s software requirements which will be downloaded and installed upon container instantiation without the necessity of user input. This alleviates the (sometimes significant) challenge of installing all required dependencies for a given model, and allows any model encapsulated in this way to be called using a common format, analogous to the functional wrapping described in Section 2.2.

SMILER accomplishes containerization using Docker [1], and its extension `nvidia-docker` [2] which supports GPU computing. The only other dependencies for using the SMILER CLI are Python and the `click` and `yaml` Python modules.

Note that a GPU is required to efficiently run the deep learning-based models supported by SMILER. It may be possible to run all models with a compatible graphics card with at least 4GB of memory, though it is highly recommended that one use a system with 6GB or more of GPU memory.

It should be mentioned that, although all efforts have been made to eliminate code duplication in order to avoid implementation drift between the MATLAB and Python code bases, there are some support functions in the SMILER MATLAB suite that had to be re-implemented in Python in order for the CLI portion of SMILER to be able to operate independently of a MATLAB license. To ensure that these processing steps are equivalent, we maintain a set of unit tests that can be used to ensure these processing steps remain equivalent in face of future improvements to the SMILER code or new MATLAB versions.

4. Discussion and Future Directions

We have presented here an overview of the SMILER software package, which provides an open, standardized, and extensible framework for maintaining and executing computational saliency models. The contributions of SMILER are two-fold: a drastic reduction in human effort to set up and run saliency algorithms, and an improvement in the consistency and procedural correctness of results and conclusions produced by different research parties. SMILER is implemented and provided as an open source software project, and it is intended to foster a collaborative research community among researchers interested in exploring computational models of visual salience.

As a continually developing project, it is recommended that users familiarize themselves with SMILER documentation supplied through the GitHub project page to be made aware of any changes or updates not reflected in this document. We encourage researchers to contribute their own saliency models to SMILER, and have included a set of ‘skeleton’ models in both MATLAB and dockerized container formats to provide a template and guidance for contributors.

References

- [1] Docker Community Edition. <https://github.com/docker/docker-ce>. 9
- [2] NVIDIA Container Runtime for Docker. <https://github.com/NVIDIA/nvidia-docker>. 9
- [3] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org. 2
- [4] Y. Alnoamany and J. A. Borghi. Towards computational reproducibility: researcher perspectives on the use and sharing of software. *PeerJ Computer Science*, 4:e163, Sept. 2018. 1
- [5] D. Berga and X. Otazu. A neurodynamic model of saliency prediction in V1. *arXiv*, abs/1811.06308, 2018. 1, 3
- [6] O. Boiman and M. Irani. Detecting irregularities in image and in video. *International Journal of Computer Vision*, 2007. 4
- [7] A. Borji, D. N. Sihite, and L. Itti. Salient object detection: A benchmark. In *European Conference on Computer Vision (ECCV)*, Oct 2012. 2
- [8] A. Borji, D. N. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2013. 2
- [9] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997. 4
- [10] N. D. B. Bruce and J. K. Tsotsos. Saliency based on information maximization. In *Advances in Neural Information Processing Systems (NIPS)*, volume 18, pages 155–162, 2006. 1, 2, 3, 5, 7
- [11] N. D. B. Bruce, C. Wloka, N. Frosst, S. Rahman, and J. K. Tsotsos. On computational modeling of visual saliency: Examining what’s right, and what’s left. *Vision Research*, In Press, 2015. 4
- [12] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba. Mit saliency benchmark. <http://saliency.mit.edu/>. 2
- [13] C.-K. Chang, C. Siagian, and L. Itti. Mobile robot vision navigation and localization using gist and saliency. In *IEEE Conference on Intelligent Robots and Systems (IROS)*, 2010. 4
- [14] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai. Fusing generic objectness and visual saliency for salient object detection. In *ICCV*, 2011. 3
- [15] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. A deep multi-level network for saliency prediction. In *IEEE International Conference on Pattern Recognition (ICPR)*, pages 3488–3493, 2016. 2, 8
- [16] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting human eye fixations via an lstm-based saliency attentive model. *arXiv*, abs/1611.09571, 2016. 2, 8
- [17] N. Dhavale and L. Itti. Saliency-based multifoveated mpeg compression. In *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, volume 1, pages 229–232 vol.1, July 2003. 4
- [18] E. Erdem and A. Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, 13(2013):11, 2013. 1, 2, 7
- [19] Q. Fang. jsonlab toolbox. 9
- [20] S. Fang, J. Li, Y. Tian, T. Huang, and X. Chen. Learning discriminative subspaces on random contrasts for image saliency analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 28(5):1095–1108, 2017. 2, 7
- [21] K. Fu, J. Li, H. Shen, and Y. Tian. How Drones Look: Crowdsourced Knowledge Transfer for Aerial Video Saliency Prediction. *arXiv*, Nov. 2018. 4
- [22] A. Garcia-Diaz, X. R. Fdez-Vidal, X. M. Pardo, and R. Dosi. Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing*, 30(1):51–64, 2012. 1, 2, 3, 5, 7
- [23] C. Goble. Better software, better research. *IEEE Internet Computing*, 18(5):4–8, Sept 2014. 1
- [24] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(10):1915–1926, 2012. 1, 2, 3, 5, 7
- [25] C. Guo and L. Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Transactions on Image Processing*, 19:185–198, 2010. 4
- [26] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu. Advanced deep-learning techniques for salient and category-specific object detection: A survey. *IEEE Signal Processing Magazine*, 35(1):84–100, Jan 2018. 3

- [27] P. Harding and N. Roberston. Task-based visual saliency for intelligent compression. In *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 2009. 4
- [28] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, volume 19, pages 545–552, 2007. 1, 2, 3, 7
- [29] J. M. Henderson, G. L. Malcolm, and C. Schandl. Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin and Review*, 2009. 4
- [30] X. Hou, J. Harel, and C. Koch. Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34:194–201, 2012. 1, 2, 7
- [31] X. Hou and L. Zhang. Dynamic visual attention: Searching for coding length increments. *Neural Information Processing Systems*, 21:681–688, 2008. 2, 7
- [32] X. Huang, C. Shen, X. Boix, and Q. Zhao. SALICON: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In *ICCV*, 2015. 2, 5, 8
- [33] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 20:1254–1259, 1998. 2, 3, 7
- [34] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv*, 2014. 2
- [35] T. Judd, F. Durand, and A. Torralba. A benchmark of computational models of saliency to predict human fixations. Technical report, Massachusetts Institute of Technology, 2012. 1, 2, 3
- [36] M. Kümmerer, T. S. A. Wallis, and M. Bethge. DeepGaze II: reading fixations from deep features trained on object recognition. *arXiv*, abs/1610.01563, 2016. 2, 8
- [37] M. Kümmerer, T. S. A. Wallis, L. A. Gatys, and M. Bethge. Understanding low- and high-level contributions to fixation prediction. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 2, 8
- [38] J. Y. Lin, T. J. Liu, W. Lin, and C. C. J. Kuo. Visual-saliency-enhanced image quality assessment indices. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–4, Oct 2013. 4
- [39] Q. Ma and L. Zhang. Saliency-based image quality assessment criterion. In *Proceedings of the 4th International Conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications - with Aspects of Theoretical and Methodological Issues*, ICIC '08, pages 1124–1133, Berlin, Heidelberg, 2008. Springer-Verlag. 4
- [40] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):171–177, Jan 2010. 3
- [41] M. Mancas, D. Unay, B. Gosselin, and B. Macq. Computational attention for defect localization. In *Proc. of ICVS Workshop on Computational Attention and Applications*, 2007. 4
- [42] S. Marat, T. HoPhuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guérin-Dugué. Modelling spatio-temporal saliency to predict gaze direction for short videos. *International Journal of Computer Vision*, 82(3):231–243, 2009. 3
- [43] C. M. Masciocchi and J. D. Still. Alternatives to eye tracking for predicting stimulus-driven attentional selection within interfaces. *Human-Computer Interaction*, 28(5):417–441, 2013. 4
- [44] A. Nuthmann and J. M. Henderson. Object-based attentional selection in scene viewing. *Journal of Vision*, 10(8):20, 2010. 4
- [45] J. Pan, C. Canton-Ferrer, K. McGuinness, N. E. O’Connor, J. Torres, E. Sayrol, and X. Giró i Nieto. Salgan: Visual saliency prediction with generative adversarial networks. *arXiv*, abs/1701.01081, 2017. 2, 8
- [46] A. Rasouli and J. K. Tsotsos. Visual saliency improves autonomous visual search. In *Canadian Conference on Computer and Robot Vision (CRV)*, 2014. 4
- [47] N. Riche, M. Mancas, M. Duvinage, M. Mibulumukuni, B. Gosselin, and T. Dutoit. RARE2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, 28(6):642–658, 2013. 2, 5, 7
- [48] R. Roberts, D.-N. Ta, J. Straub, K. Ok, and F. Dellaert. Saliency detection and model-based tracking: A two part vision system for small robot navigation in forested environments. In *The Proceedings of SPIE 8387*, 2012. 4
- [49] B. Schauerte and R. Stiefelhagen. Quaternion-based spectral saliency detection for eye fixation prediction. In *European Conference on Computer Vision (ECCV)*, pages 116–129, 2012. 2, 7
- [50] H. J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of vision*, 9(12):15–15, 2009. 2, 7
- [51] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs. Automatic thumbnail cropping and its effectiveness. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST ’03, pages 95–104, New York, NY, USA, 2003. ACM. 4
- [52] H. R. Tavakoli, E. Rahtu, and J. Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *Proceedings of Scandinavian Conference on Image Analysis (SCIA)*, 2011. 2, 5, 7
- [53] C. L. Thomas. OpenSalicon: An open source implementation of the salicon saliency model. Technical Report TR-2016-02, University of Pittsburgh, 2016. 2, 5, 8
- [54] J.-F. Tsai and K.-J. Chang. Opensource implementation of context-aware saliency detection. <https://sites.google.com/a/jyunfan.co.cc/site/opensource-1/contextsaliency>. 2, 5, 7
- [55] E. Vig, M. Dorr, and D. Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2798–2805, 2014. 1, 2, 3, 8

- [56] W. Wang and J. Shen. Deep visual attention prediction. *IEEE Transactions on Image Processing*, 27(5):2368–2378, 2018. [1](#), [2](#), [3](#), [8](#)
- [57] B. J. White, J. Y. Kan, R. Levy, L. Itti, and D. P. Munoz. Superior colliculus encodes visual saliency before the primary visual cortex. *Proceedings of the National Academy of Sciences*, 114(35):9451–9456, 2017. [4](#)
- [58] S. X. Yu and D. A. Lisin. *Advances in Visual Computing*, chapter Image Compression Based on Visual Saliency at Individual Scales, pages 157–166. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. [4](#)
- [59] T. Yubing, H. Konik, F. A. Cheikh, and A. Tremeau. Full reference image quality assessment based on saliency map analysis. *Journal of Imaging Science and Technology*, 54, 2010. [4](#)
- [60] A. Zaharescu and R. P. Wildes. Spatiotemporal saliency via centre-surround comparison of visual spacetime orientations. In *Asian Conference on Computer Vision (ACCV)*, 2012. [3](#)
- [61] J. Zhang and S. Stan. Exploiting surroundedness for saliency detection: A Boolean map approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 5(38):889–902, 2016. [2](#), [8](#)
- [62] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7:32):1–20, 2008. [2](#), [4](#), [7](#)
- [63] T. Zhu, W. Wang, P. Liu, and Y. Xie. Saliency-based adaptive scaling for image retargeting. In *2011 Seventh International Conference on Computational Intelligence and Security*, pages 1201–1205, Dec 2011. [4](#)