Article

S|S|S

Social Studies of Science
2017, Vol. 47(6) 811–840

DOI: 10.1177/0306312717730428
journals.sagepub.com/home/sss

We get the algorithms of our ground truths: Designing referential databases in digital image processing

Florian Jaton
Institut des Sciences Sociales, Université de Lausanne, Lausanne, Switzerland

Abstract

This article documents the practical efforts of a group of scientists designing an image-processing algorithm for saliency detection. By following the actors of this computer science project, the article shows that the problems often considered to be the starting points of computational models are in fact provisional results of time-consuming, collective and highly material processes that engage habits, desires, skills and values. In the project being studied, problematization processes lead to the constitution of referential databases called 'ground truths' that enable both the effective shaping of algorithms and the evaluation of their performances. Working as important common touchstones for research communities in image processing, the ground truths are inherited from prior problematization processes and may be imparted to subsequent ones. The ethnographic results of this study suggest two complementary analytical perspectives on algorithms: (1) an 'axiomatic' perspective that understands algorithms as sets of instructions designed to solve given problems computationally in the best possible way, and (2) a 'problem-oriented' perspective that understands algorithms as sets of instructions designed to computationally retrieve outputs designed and designated during specific problematization processes. If the axiomatic perspective on algorithms puts the emphasis on the numerical transformations of inputs into outputs, the problem-oriented perspective puts the emphasis on the definition of both inputs and outputs.

Keywords

algorithms, computational science, computer science, ground truths, image processing

Correspondence:
Florian Jaton, Institut des Sciences Sociales, Université de Lausanne, CH-1015 Lausanne, Switzerland.
Email: florian.jaton@unil.ch

Introduction: Bringing algorithms down to the ground

Reacting against the depiction of algorithms as both powerful and inscrutable,[1] recent publications in Science and Technology Studies (STS) have portrayed them as uncertain socio-material configurations. Inspired by approaches for which relations do not connect but '*enact* entities in the flow of becoming' (Introna, 2016: 23), case studies such as those of the computerization of the Arizona Stock Exchange in the 1990s (Muniesa, 2011), the performativity of a text-matching program for academic essays (Introna, 2016) and the implementation of an airport security system (Neyland, 2016) participate in deflating the 'algorithmic drama' (Ziewitz, 2016) and challenging it with more nuanced and grounded narratives. Instead of questioning the effects algorithms may have on society – thus excluding these technical artefacts from the common world (Simondon, 2017) – these studies explore the co-constitution of algorithms and society (Ananny, 2016; Crawford, 2016; Zarsky, 2016). By understanding algorithms as assemblages embedded in mundane practices (the surveillance habits of airport security crews, the academic duties of undergraduate students in Modern History, the doubts of a EU Data Protection officer), these case studies around algorithms bring to light new suggestions for innovative modes of design and cooperation (Jackson et al., 2014; Knobel and Bowker, 2011). In short, by examining how algorithms are brought into existence, this STS research program helps to better compose (Latour, 2010) common worlds with algorithms.

In order to participate in these stimulating efforts to bring algorithms 'down to earth' (Bogost, 2015) and know them better (Seaver, 2013), this exploratory article reports on an ethnographic inquiry into the constitution of an image-processing algorithm. Inspired by 1980s laboratory studies and their attempts to deflate elusive aspects of scientific facts by accounting for mundane practices of scientists (Fujimura, 1987; Knorr-Cetina, 1981; Latour and Woolgar, 1986; Lynch, 1985), this article examines a computer science laboratory where researchers manufacture algorithms. In an attempt to learn more about computation on the ground, this study accounts for instruments, habits, intuitions, desires, duties and skills that participate in the shaping of algorithms.

Specifically, this case study accounts for the constitution of an image-processing algorithm for saliency detection (described in detail below). It documents the practical efforts of a group of young researchers designing and publishing a new algorithm or, as computer scientists usually put it, a new computational model. By following the actors of this project, we see that the problems computational models are intended to solve are in fact provisional results of time-consuming and highly material processes that engage habits, desires, skills and values. During such *problematization processes* (Callon, 1986), the inputs upon which a desired algorithm will work and the outputs that it is supposed to produce are both manually shaped and gathered in databases that computer scientists called 'ground truths'. These ground truths are used both to define the numerical features of the algorithm and to evaluate its performances. Working as expansive common touchstones for research communities in computer science, these ground truths also inherit from prior problematization processes and engender subsequent ones. The centrality of ground truths for the design and evaluation of algorithms strongly suggests that, to a certain extent, we get the algorithms of our ground truths.

Building upon these ethnographic insights, I tentatively propose two complementary analytical perspectives on algorithms. A first perspective would see algorithms as sets of instructions designed to computationally solve problems in the best possible way (Cormen et al., 2009). By considering problems to be given, this *axiomatic* way of seeing algorithms effectively facilitates inquiries into the capabilities of numbers – a fascinating research topic. Yet, a second and equally legitimate perspective could consider algorithms to be sets of instructions designed to computationally retrieve in the best possible way what have been designed as outputs during specific problematization processes. I assume that if this *problem-oriented* way of seeing algorithms manages to coexist with its axiomatic counterpart, it could pave the way for refreshing human-algorithm reconfigurations (Suchman, 2007). Indeed, if ground truths define the problems that algorithms are supposed to solve computationally, the construction processes of these ground truths might be important situations to be investigated critically and creatively.

The lab

The main setting of my case study is a European technical institute with a quite renowned computer science (CS) faculty. On the third floor of the CS faculty's main building, scattered in six offices along one of the four hallways that surround the central patio, lies what I will call 'the Lab': a well-respected laboratory of computational photography. Computational photography is a broad research area linked to the Charge-Coupled Device (CCD) developed in the late 1960s (Seitz and Einspruch, 1998). Through the translation of electromagnetic photons into electron charges that can be amplified and digitalized, CCDs enable the production of pixel images constituted of discrete elements. Organized in grids, these discrete signals possess the ability to be processed automatically by computer programs that are themselves non-trivial expressions of mathematical algorithms.[2] In the Lab, colors, shadows, smiles or elephants can also be considered 'two-dimensional digital signals'[3] upon which automated calculations can be processed, mostly by means of linear algebra. The creation of new algorithms and their translation into computer programs able to compute the constitutive elements of digital photographs (often called 'natural images') is one of the research foci of the Lab. This area of practice is also called 'two-dimensional digital signal processing' or, more succinctly, 'image processing'.

Some image-processing algorithms designed by computational photography laboratories are quite specialized and intended for specific purposes (e.g. superpixel segmentation algorithms), whereas others are widespread and industrially implemented in broader assemblages such as digital cameras (e.g. red-eye removal programs), expensive software, and large information systems (e.g., text-recognition programs, compression schemes and feature clustering). Whether widespread or specialized, these algorithms first need to be trained, nurtured, evaluated, and compared in places such as the Lab.

The Lab was the setting, with the support of its interdisciplinary collaborative director, of my two-year ethnographic inquiry into the constitution of image-processing algorithms. In order to conduct the investigation and collect data, I stayed with the Lab between November 2013 and December 2015, participating in its projects, seminars, meetings and social events. This paper draws upon data I collected while participating in

a project run by three members of the Lab – two computer science PhD students (GY and BJ) and a post-doc (SL), who collectively I call 'the Group' – from November 2013 to June 2014. As we will see, the Group ultimately tried to introduce new research directions in a subfield of image processing called 'saliency detection'. Yet in order to propose an innovative algorithm for saliency detection, the Group had to construct a new referential database – what computer scientists call a 'ground truth' – that could both provide the numerical features of their algorithm and attest to its reliability. Insofar as the Group's final paper on the algorithm was rejected by the committee of an important European conference on image processing, we can assume that the project – at least in June 2014 – failed to fulfil its initial ambitions.

My analysis is based mainly on excerpts of discussions recorded in the Lab's cafeteria during weekly 'Group meetings', where the Group and I discussed the project's framing, progress and issues. From these excerpts, as well as some other collected documents, I try to account for how the Group designed a new ground truth for saliency detection, and why this move was considered an essential step for the success of their project. These empirical elements further allow me to propose broader propositions about the relations between ground truths and algorithms.

Saliency detection and digital image processing, 1970–2013

'Saliency' for computer scientists in image processing is a blurry term with a difficult-to-track history involving different – yet closely related – research areas. One possible point of departure in the 1970s is when explicative models in cognitive psychology and neurobiology started to schematize how the human brain could handle an amount of visual data far larger than its estimated processing capabilities. After many disputes and controversies,[4] a rough agreement about the overall process of humans' 'selective visual attention method' has emerged, distinguishing between two neuronal processes of selecting and gating visual information (Heinke and Humphreys, 2005). On the one hand, there is a task-independent and rapid 'bottom-up visual attention process' that selects conspicuous stimuli such as color contrasts, feature orientations or spatial frequency. On the other hand, there is a slower, selectively operating, task-based 'top-down visual attention process'. The term 'saliency map' was proposed by Koch and Ullman (1985) to define the final result of the brain's bottom-up visual attention process.

In the 1980s, the two theorized different 'paths' for the brain to process analogical light signals – one fast and generic, the other slower and task-specific – inspired scientists in computer vision whose machines face a similar problem, the stream of sampled digital signals that emanate from CCDs being too large to be processed all at once. Thus computer scientists have progressively shaped two different classes of image-processing detection algorithms. The first class is supposed to detect 'low-level features' inscribed within the pixels of a given image, such as intensity, color, orientation and texture. Through the efforts of Laurent Itti and Christof Koch in the 2000s[5] the term 'saliency' was progressively assimilated into this first class of algorithms, which became 'saliency-detection algorithms'. The second class of image-processing detection algorithms is based on 'high-level features' that have to be learned by machines according to specific metrics (e.g. face or car detection). This often involves automated learning procedures and the management of increasingly large databases (Lowe, 1999).

Despite differences in terms of substratum, both high-level and low-level detection algorithms were (and still are) bound to the same construction workflow, which consists of five interrelated and problematic steps: (1) the acquisition of a finite dataset, (2) on the data of this dataset, the manual labelling of clear *targets*, defined here as the elements (faces, cars, salient regions) that the desired algorithm will be asked to detect, (3) the construction of a database, usually called a 'ground truth' by the research community, gathering the unlabelled data and their manually labelled counterparts, (4) the design of the algorithm's calculating properties and parameters based on a statistically representative part of the ground-truth database, and (5) the evaluation of the algorithm's performances based on the rest of the ground-truth database. Thus the very existence of a standard detection algorithm depends upon a finite set of digital images for which some human workers have previously labelled targets (e.g. faces, cars or salient regions). The unlabelled images and their manually labelled counterparts are then gathered together within a database to form the ground truth. In order to design and code the algorithm, the ground truth is randomly split into two parts: the 'training set' and the 'evaluation set'. The designers would use the training set to extract formal information about the targets and translate them into mathematical expressions. Once formalized and implemented in machine-readable code, the algorithm is tested on the evaluation set to see how well it detects targets that were not used to design its properties. It produces a precise number of outputs that can be qualified as 'true positives', 'false negatives' or 'false positives', thanks to the previous human labelling work. Out of this comparison between manually designed targets and automatically produced outputs, statistical measures in terms of precision (the fraction of detected items that were previously defined as targets) and recall (the fraction of targets among the detected items) can be obtained in order to compare and rank competing algorithms[6] (see Figure 1).

One drawback of high-level detection algorithms is that they are task-specific and cannot by themselves detect different types of targets: a face-detection algorithm will detect faces, a car-detection algorithm will detect cars, etc.[7] Yet, one of the benefits of such high-level detection algorithms is that the definition of their targets often involves only minor ambiguities for those who design them: cars and faces have rather unambiguous characteristics that facilitate agreement. Ground truths can then be manually shaped by computer scientists in order to train high-level detection algorithms. Moreover, these ground truths can also serve as referees between competing high-level detection algorithms since they provide precision and recall metrics. The sub-field of face-detection with its numerous ground truths and algorithmic propositions provides a paradigmatic example of a highly developed and competitive topic in image processing (Hjelmås and Low, 2001; Zhang and Zhang, 2010).

In the 2000s, unlike research in high-level detection, low-level saliency detection had no obvious ground truth allowing the design and evaluation of computational models.[8] At that time, if the task-independent and adaptive character of saliency detection was theoretically interesting for automatic image cropping (Santella et al., 2006), adaptive display on small devices (Chen et al., 2003), advertising design and image compression (Itti, 2000), the absence of any ground truth that could allow the training and evaluation of computational models prevented saliency detection from being an active topic in digital image processing. As Itti et al. (1998) confessed when they tested the very first saliency-detection algorithm on natural images:

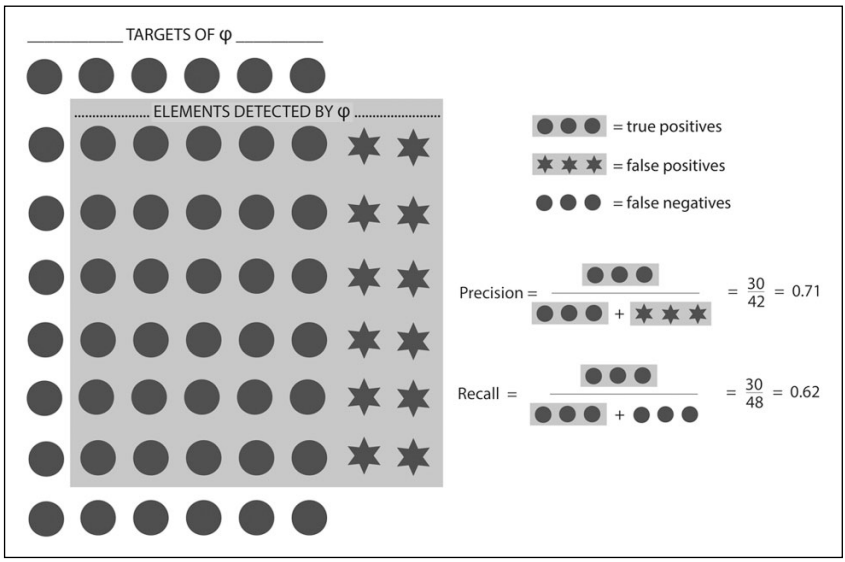

Figure 1. Schematic of precision and recall measures on ϕ. In this hypothetical example, ϕ (grey background) managed to detect 30 targets (true positives) but missed 18 of them (false negatives). This performance makes ϕ have a recall score of 0.62. ϕ also detected 12 elements that are not targets (false positives) and this makes it have a precision score of 0.71. From this point, other algorithms intended to detect the same targets can be tested on the same ground truth and may have better or worse precision and recall scores than ϕ.

> With many such [natural] images, it is difficult to objectively evaluate the model, because no objective reference is available for comparison, and observers may disagree on which locations are the most salient. (Itti et al., 1998: 1258)

Saliency detection in natural images is an equivocal topic, not easily expressed in a univocal ground truth. If it is usually straightforward (yet time consuming) to define univocal targets for training and evaluating high-level face or car detection algorithms, it is far more complex to do so for saliency-detection algorithms, because what is considered as salient in a natural image tends to change from person to person. While in the 2000s saliency-detection algorithms might have been promising for many industrial applications, nobody in the field of image processing had found a way to design a ground truth for natural images.

In 2007, Liu et al. proposed an innovative solution to this problem and created the very first ground truth for saliency detection in natural images. Their shift was smart, costly and contributed greatly to framing and establishing the subfield of saliency detection in the image-processing literature. Liu et al.'s first move was to propose one possible scope of saliency detection by incorporating concepts from high-level detection: Instead of trying to highlight salient areas within digital images, computational models for saliency should detect 'the most salient object' within a given digital image. They thus framed the saliency problem as binary and object-related:

> We incorporate the high-level concept of salient object into the process of visual attention in each respective image. We call them salient objects, or foreground objects that we are familiar with. … [W]e formulate salient object detection as a binary labelling problem that separates a salient object from the background. Like face detection, we detect a familiar object; unlike face detection, we detect a familiar yet unknown object in an image. (Liu et al., 2007: 1–2)

Thanks to this refinement of the concept of saliency (from 'anything that first attracts attention' to 'the one object in a picture that first attracts attention'), Liu et al. could organize an experiment in order to construct legitimate targets to be retrieved by computational models. They first collected 130,099 random high-quality natural images from Internet forums and search engines. Then they manually selected 20,840 images that fit their definition of the saliency problem, images that, according to them, contained only one salient object. This initial selection operation was crucial, since it excluded images with several potentially salient objects.

For each image, three human workers then manually drew rectangles on what they thought was the most salient object. Liu et al. had thus obtained three different rectangles for each image, whose consistencies could be measured by the percentage of shared pixels. For a given image, if its three rectangles were more consistent than a chosen threshold (here, 80% of pixels in common), the image was considered to contain a 'highly consistent salient object' (Liu et al., 2007: 2). After this first selection step, their dataset called α contained around 13,000 images.

Liu et al. then randomly selected 5000 highly consistent salient-object images from α to create a second dataset called β. They then asked nine other human workers to label the salient object of every image in β with a rectangle. This time, Liu et al. obtained for every image nine different yet highly consistent rectangles whose average surface was considered their 'saliency probability map' (Liu et al., 2007: 3). Thanks to this constructed social agreement, the 5000 saliency probability maps – from a computer-science perspective, tangible *matrices* constituted by specific numerical values – could then be considered the best solutions to the saliency problem as they framed it. The database gathering the natural images and their corresponding saliency probability maps became the material base upon which the desired algorithm could be developed. By constructing this ground truth, Liu et al. defined the terms of a new problem whose solutions could be retrieved by means of calculating methods.

By organizing this survey, inviting people into their laboratory, welcoming them, explaining the topic to them, writing the appropriate programs to make them label the images, and gathering the results in a proper database in order to statistically process them, Liu et al. transformed their initial reduced conception of saliency detection into workable and unambiguous targets with specific numerical values. At the end of this laborious process, Liu et al. could randomly select 2000 images from set α and 1000 images from set β to construct a training set (Liu et al., 2007: 5–6) in order to analyze the shared features of their targets. Once the adequate numerical features were extracted from the targets of the training set and implemented in machine-readable language, they used the 4000 remaining images from set β to measure the performances of their algorithm. Furthermore, and for the very first time, they also could compare the detection

performances of their algorithm with two competing algorithms already proposed by other laboratories but that could not have been evaluated on natural images, due to the lack of any 'natural' targets related to saliency (see Figure 2). Besides the actual completion of their saliency-detection algorithm, the innovation of Liu et al. was to redefine the saliency problem so that it could allow performance evaluations.

By publishing their paper and also publicly providing their ground truth online, it is not an exaggeration to say that Liu et al. established a newly assessable research direction in image processing. A costly infrastructure had been put together, ready to be reused in order to support competing algorithmic propositions. Their publication was more than a paper: It was a paper that allowed other papers to be published, as they had provided a ground truth that could be used by other researchers who would quote the seminal paper and accept the ground truth's restricted – yet operational – definition of saliency.[9]

Another important paper for saliency detection – and therefore also for the Group's project we shall soon start to follow – was published in 2008 by Wang and Li. To them, even though Liu et al. were right to frame the saliency problem as a binary problem, their bounding-box ground truth remained unsatisfactory, since it could well evaluate inaccurate results (see Figure 3). In order to refine the measures of Liu et al.'s first ground truth for saliency detection, Wang and Li randomly selected 300 images from the β dataset and used a segmentation tool to manually label the *contours* of each of the 300 salient objects. What they proposed and evaluated then was a saliency detection algorithm that 'not only captures the rough location and region of the salient objects, but also roughly keeps the contours right' (Wang and Li, 2008: 965).

From this point, saliency detection in image-processing was almost set: Even though many algorithms exploiting different low-level pixel information were later proposed,[10] they were all bound to the saliency problem as defined by Liu et al. (2007). And even though other ground truths have since been proposed in published papers (Judd et al., 2012; Movahedi and Elder, 2010), in order to widen the scope of saliency detection (notably by proposing images with two objects that could be decentered), Liu et al.'s framing of saliency detection as a binary object-related problem remained unchallenged. And when the Group started their project in November 2013, Liu et al.'s problematization of the saliency problem was continuing to support a competition between algorithms that differentiated themselves by speed and accuracy (see Figure 4).

With this brief history in mind, we are now ready to follow the Group as it tries to constitute its own innovative saliency-detection algorithm.

Reformulating the saliency problem

Around 3 pm on November 7, 2013, I entered the Lab's cafeteria for the first Group meeting. Previous discussions in the Lab had led to an agreement to work on a new collective publication on saliency detection, and had identified the particular expertise of CL, GY, and BJ as relevant. The day before the Group meeting, I had attempted to read some papers on saliency detection that GY had sent me earlier, but I was confused by their tacit postulates. How would it be possible to detect saliency, since what is important in an image certainly varies from person to person? And what is this strange notion of

Figure 14. Comparison of different algorithms. From left to right: FG, SM, our approach, and ground-truth.

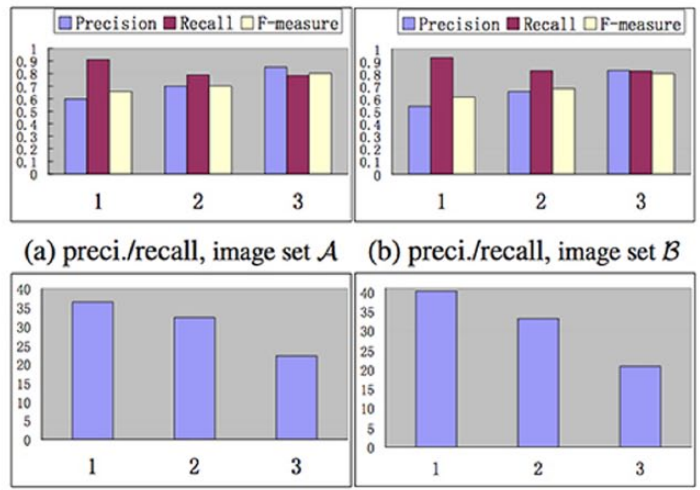

Figure 12. Comparison of different algorithms. (a-b) and (c-d) are region-based (precision, recall, and F-measure) and boundary-based (BDE - boundary displacement error) evaluations. 1. FG. 2. SM. 3. our approach.

Figure 2. Taken from Liu et al. (2007: 7). On the left, a visual comparison between three different saliency-detection algorithms according to the ground truth created by Liu et al. (2007: 7). Note that the pictures of the ground truth contain one centered and contrastive salient object. On the right, histograms that summarize the statistical performances of the three algorithms. In these histograms, the ground truth corresponds to the 'y' coordinate, thus instituting a referential space for "precision," "recall," and "Fmeasure" (the weighted average of precision and recall values) evaluations.

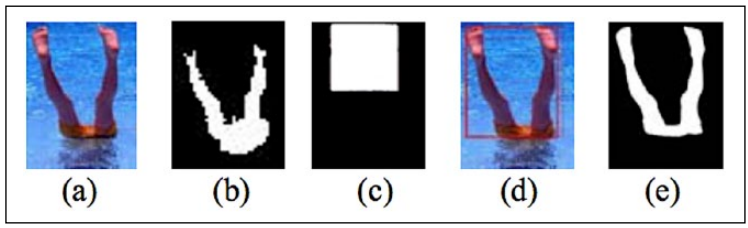

Figure 3. Taken from Wang and Li (2008: 968): (a) is an unlabelled image of Liu et al.'s ground truth database; (b) is the result of Wang & Li's saliency-detection algorithm; (c) is the imaginary result of some other saliency-detection algorithm on (a); (d) is the bounding-box target as provided by Liu et al.'s ground truth database. As we can see, even though (b) is more accurate than (c), it will obtain a lower statistical evaluation if compared to (d). That is why Wang & Li propose (e), a binary target that matches the contours of the already-defined salient object.
Source: reproduced with permission from Elsevier, Feb 02, 2015, 3560760282261, Zheshen Wang; Baoxin Li, March-April 2008.

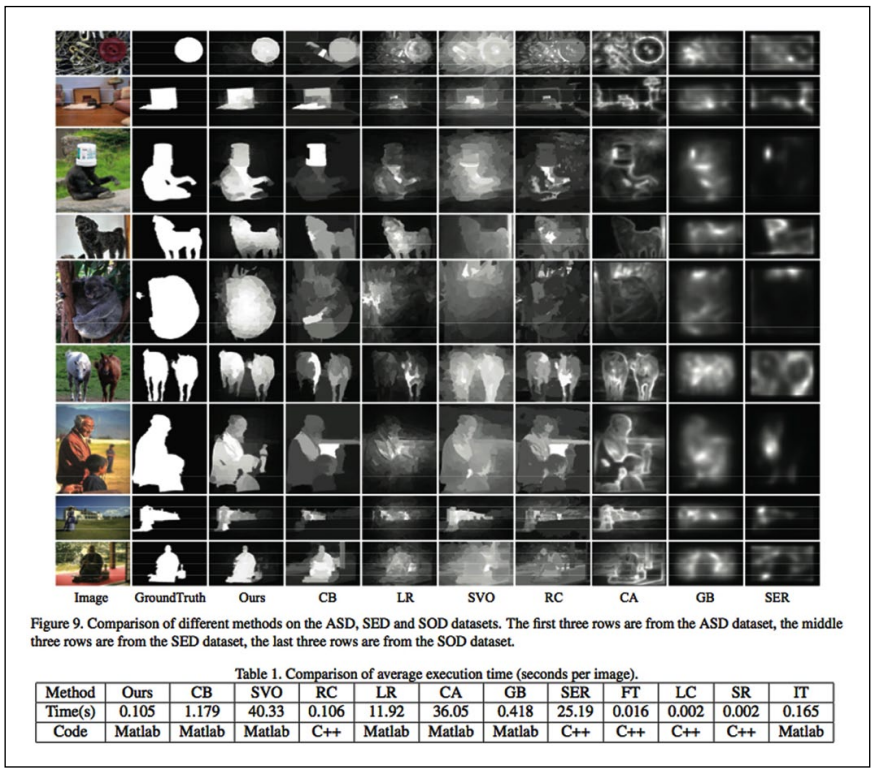

Figure 9. Comparison of different methods on the ASD, SED and SOD datasets. The first three rows are from the ASD dataset, the middle three rows are from the SED dataset, the last three rows are from the SOD dataset.

Table 1. Comparison of average execution time (seconds per image).

Method	Ours	CB	SVO	RC	LR	CA	GB	SER	FT	LC	SR	IT
Time(s)	0.105	1.179	40.33	0.106	11.92	36.05	0.418	25.19	0.016	0.002	0.002	0.165
Code	Matlab	Matlab	Matlab	C++	Matlab	Matlab	Matlab	C++	C++	C++	C++	Matlab

Figure 4. Comparison table taken from Jiang et al. (2013: 1672). The number of competing algorithms has increased since 2007. Here, three ground truths are used for performance evaluations: ASD (Achanta et al., 2009), SED (Alpert et al., 2007) and SOD (Movahedi and Elder, 2010). Beneath, a table comparing the execution time of each implemented algorithm.
Source: reproduced with permission from Elsevier, Feb 02, 2015, 35609059883, Bowen Jiang; Lihe Zhang; Huchuan Lu; Chuan Yang; Ming-Hsuan Yan, Dec. 2013.

'ground truth' on which the papers' algorithms seem to rely? For an STS scholar, the notion sounded highly problematic.

As soon as I entered the Lab's cafeteria, the Group presented me with an overview of the ambitions of the project and how it intended to run it.

Group meeting, Lab's cafeteria, 7 November 2013. After saluting the Group, FJ sits at its table:

CL: Have you heard about saliency?
FJ: Well, I've read some stuff.
CL: Huge topic but basically, when you look at an image, not everything is important usually, and you focus only on some elements. … What we try to do basically, it's like a model that detects elements in an image that should attract attention. … GY's worked on a model that uses contrasts to segment objects and BJ has a model that detects faces. We'll use them as a base. … For now, most saliency models only detect objects and don't pay attention to faces. But what we say is that faces are also important and usually attract directly the attention. … And that's the point: We want to include faces to saliency, basically.
GY: And segment faces. Because face-detectors output only rectangles. … There can be many applications [for the model], like in display or compression for example.

According to the Group, saliency detection models should also take human faces into account as faces are important in human attention mechanisms. Moreover, investing in this project within saliency detection would be a good opportunity to merge some of the Group's research on both low-level segmentation and high-level face detection. The idea to combine high-level face detection with low-level saliency detection has already been proposed in image-processing papers (Borji, 2012; Karthikeyan et al., 2013). But the Group's ambition here is to go further in the saliency direction as framed by Wang and Li (2008), after Liu et al. (2007), by proposing an algorithm capable of detecting and segmenting the *contours* of faces. In order to accomplish such subtle results, the previous work done by GY on segmentation and BJ on face detection constitutes capital with which to work.

The Group also wanted to construct a saliency-detection model that could effectively process a larger range of natural images:

Group meeting, Lab's cafeteria, November 7, 2013:

GY: But you know [to FJ], we hope the algorithm could detect multiple objects and faces. Because in saliency detection, models can only detect like one or two objects on simple images. They don't detect multiple salient objects in complex images. … But the problem is that there's no ground truth for that. There's only ground truth with like one or two objects, and not that many faces.

Pictures produced by users of digital cameras – according to the Group – are generally more cluttered than those used to train and evaluate saliency-detection algorithms. Indeed, at least in November 2013, saliency detection is a research area in which algorithms are increasingly efficient only on those rare natural images with clear and untangled features. But the Group also knew that this issue is intimately related to the current

ground truths for saliency detection, which are all bound to Liu et al's initial definition of the saliency problem. If the Group wants to propose a model that could detect a different and more subtle saliency, it must construct the targets of such saliency. If it wants to propose a model that can calculate and detect multiple salient features (objects and faces) in more complex and 'realistic' images, it must construct a new ground truth that would gather complex images and their corresponding multiple salient features.

The Group's desire to redefine the terms of the saliency problem did not come *ex nihilo*. When Liu et al. did their research on saliency in 2007, it was difficult for computer scientists to organize a large social survey on complex images. But in November 2013, the wide distribution of crowdsourcing services enabled new possibilities:

Group meeting, Lab's cafeteria, 7 November 2013:

GY: But we want to use crowdsourcing to do a new ground truth and ask people to label features they think are salient. … And then we could use that for our model and compare the results, you see?

As defined by Estellés-Arolas and González-Ladrón-de-Guevara (2012), crowdsourcing is

> a type of participative online activity in which an individual, an institution, a non-profit organization, or a company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. (p. 195)

In November 2013, there were multiple available crowdsourcing services, such as ShortTasks and Amazon's Mechanical Turk.[11] For the Group, the estimated benefits were huge: Once the desired web application was coded and set with a clear instruction, such as 'please highlight the features that directly attract your attention', the Group would be able to pay a crowdsourcing company that will take charge of linking the application to dozens of paid workers. In turn, these workers would feed the Group's server with labelling coordinates that can be processed on basic yet powerful software such as Matlab.[12] For our story, crowdsourcing – as a rather easily available paid service – creates a difference (Latour, 2005): The gathering of many manually labelled salient features would become more manageable for the Group than it had been for Liu et al. and an extension of the notion of saliency to multiple features would become doable (Fujimura, 1987).

Another difference effected by crowdsourcing was a potential redefinition of the saliency problem as *continuous*:

Group meeting, Lab's cafeteria, 7 November 2013:

FJ: So basically you want many labels?

GY: Yeah because you know, in the state-of-the-art face detection or saliency, models only detect things in a binary way, like face/no face, salient/not salient. What we also try to do is a model that evaluates the importance of faces and objects and segments them. Like 'this face is more important than this other face which is more important than that object' and so on. … But anyways, to do that [a ground truth based on the results of a crowdsourcing experiment], we first need a dataset with many images with different contents.

CL: Yeah, we thought about something like 1000 images at least, to train and evaluate. But it has to be images with different objects and faces with different sizes.

GY: And we have to select the images; good images to run the survey. … We'll try to propose a paper in Spring so it would be good to have finished crowdsourcing in January I guess.

If the images used to construct the ground truth contained only one or two objects and were labelled by only several individuals, no relational values between the labelled features could be calculated. In this situation, defining saliency as a binary problem in the manner of Liu et al. makes complete sense. Yet if the Group could afford to launch a social survey that asked for many labels on a dataset with complex images containing many features, it would become methodologically possible to assign relative importance values to the different labelled features. This was a question of arithmetic values: If one feature were manually labelled as salient, one could only obtain a binary value (foreground and background). But if several features were labelled as more or less salient by many workers, one could obtain a continuous subset of results. For the Group, crowdsourcing once again creates a difference, by making it possible to create new types of targets with relatively continuous values. It was difficult at this point to predict if the Group's algorithm would be effectively able to approach these subtle results. Nevertheless, the ground truth the Group wanted to constitute would enable the development of such an algorithm by providing targets that the model should try to retrieve in the best possible way.

Even though the Group had managed to build upon previous work in saliency detection and related fields in order to reformulate the saliency problem, it still lacked the ground truth that could effectively establish the terms of this new problem. Both the inputs on which the desired algorithm should work and the outputs (the 'targets') it should retrieve needed to be constructed. The Group was only at the beginning of the *problematization* process that could enable the construction of a new computational model. The Group's reformulation of the saliency problem still needed to be equipped (Vinck, 2011) with tangible elements (a new set of complex images, a crowdsourcing experiment, continuous values, segmented faces) in order to form a referential database that would in turn constitute the material base of the new computational model. The new ground truth was an obligatory passage point (Callon, 1986) for the Group, and the Group also hoped that it would become an obligatory passage point for the research community. Without a new ground truth, saliency-detection models would still operate on unrealistic images, they would still be one-off object-related, they would still ignore the detection and segmentation of faces, and they would still, therefore, be irrelevant for real-world applications. With the help of a new ground truth, these shortcomings the Group attributed to saliency detection might be overcome. In a similar vein, we can say that saliency detection was at this point doable (Fujimura, 1987) only at the level of the laboratory. Without a new ground truth, the Group had no tangible means to articulate this 'laboratory level' with the research community in image processing. It was only by constructing a database gathering 'input data' and 'output targets' that the Group would be able to propose and publish an algorithm capable of solving the newly formulated saliency problem.

Constructing a new ground truth

In addition to working on the coding of the crowdsourcing Web application, the Group also dedicated November and December 2013 to the selection of images that echoed the

algorithm's three expected performances: (1) detecting and segmenting the contours of salient features, including faces, (2) detecting and segmenting these salient features in complex images, and (3) evaluating the relative importance of the detected and segmented salient features. These specifications led to several Group meetings specifically organized to discuss the content and distribution of the selected images:

Group meeting, Lab's cafeteria, 21 November 2013:

BJ: Well we may avoid this kind of basketball photo because these players may be famous-like. They are good because the ball contrasts with faces but at least I know some of the players. And if I know, we include other features like 'I know this face' so I label it.
CL: I think maybe if you have somebody that is famous, the importance of the face increases and then we just want to avoid modelling that in our method.
…
CL: OK. And the distributions are looking better?
FJ: Yes definitely. BJ just showed me what to improve.
CL: OK. So what other variables do we consider?
GY: Like frontal and so on. But equalizing them is real pain.
CL: But we can cover some of them; maybe not equalize. So there should be like the front face with images of just the front of the face and then there is the side face, and a mixture in between.

The Group's anticipated capabilities for the algorithm oriented this manual selection process. As with Liu et al.'s efforts, but in a manner that made the Group include more complex 'natural' situations, the creation of a dataset was driven by the algorithm's future tasks. By December 2013, 800 high-resolution images were gathered – mostly from Flickr – and stored in the Lab's server. Because the Group considered the inclusion of faces the most significant contribution of the project, 632 of the selected images included human faces.

Running parallel to this problem-oriented selection of images, organizational work on the selected images had to be defined so that the Group would not be overloaded by the number of files and labelled results to be gathered through the crowdsourcing experiment. This kind of organizational procedure was very close to data management and implied the realization of a whole new database for which information could be easily retrieved and anticipated. Moreover, the shaping of the crowdsourcing survey also required coordination and adjustments: What questions would be asked? How would answers be collected? How would answers be processed to fulfil the ambitions of the project? Those were crucial issues because the 'raw' labelled answers obtained via crowdsourcing could only be rectangles, not precise contours.

Group meeting, Lab's cafeteria, 12 December 2013:

CL: But for the database, do we rename the images so that we have a consistency?
BJ: Hum… I don't think so because now we can track the files back to the website with their ID. And with Matlab you can like store the jpg files in one folder and retrieve all of them automatically

…

CL: What do you think, GY? Can we ask people to select a region of the image or to do something like segmenting directly on it?

GY: I don't think you can get pixel-precision answers with crowdsourcing. We'll need to do the pixel-precision [in the Lab] because if we ask them, it's gonna be a very sloppy job. Or too slow and expensive anyway.

CL: So what do you want? There is your Matlab code to segment features, right?

GY: Yes but that's low-level stuff, pixel-precision [segmentation]. It's gonna be for later, after we collect the coordinates I guess. I still need to finish the scripts [to collect the coordinates] anyway. Real pain… But what I thought was just like ask people to draw rectangles on the salient things, then collect the coordinates with their ID and then use this information to deduce which feature is more salient than the other on each image. Location of the salient feature is a really fuzzy decision but cutting up the edges it's not that dependent. … You know where the tree ends, and that's what we want. Nobody will come and say 'No! The tree ends here!' There is not so many variance between people I guess in most of the cases.

CL: OK, let's code for rectangles then. If that's easy for the workers, let's just do that.

The IDs of the selected images made it easy for the Group to put the images in a Matlab database. But within the images, the salient features labelled by the participants of the crowdsourcing experiment were more difficult to handle, since GY's interactive tool to get the precise boundaries of image-contents was based on low-level information. As a consequence, segmenting the boundaries of low-contrast features such as faces could take several minutes, whereas affordable crowdsourcing (crowdsourcing services cost the Lab approximately US $950) is about small and quick tasks. The labelled features would thus have to be post-processed within the Lab in order to obtain precise contours.

Moreover, another potential point of failure of the project resided in the development of the crowdsourcing application. Getting people to draw rectangles around features, translating these rectangles into coordinates and storing them in files in order to process them statistically require non-trivial programming skills. By January 2014, when the crowdsourcing application was made fully operational, it comprised seven different scripts (around 1000 lines of code) written in HTML, PHP and JavaScript, that responded to each other depending on the workers' inputs (see Figure 5). Yet if computer scientists in image processing tend to be at ease with numerical computing and programming languages such as Matlab, C or C++, web designing and social pooling are not competencies for which they are necessarily trained.

Once coded and debugged, the different scripts were stored in one public section of the Lab's server whose address was made available in January 2014 to 30 external paid workers of a crowdsourcing company. By February 2014, tens of thousands of rectangles' coordinates were stored in the Group's database as TXT files, ready to be processed thanks to the previous preparatory steps. At this point, each image of the

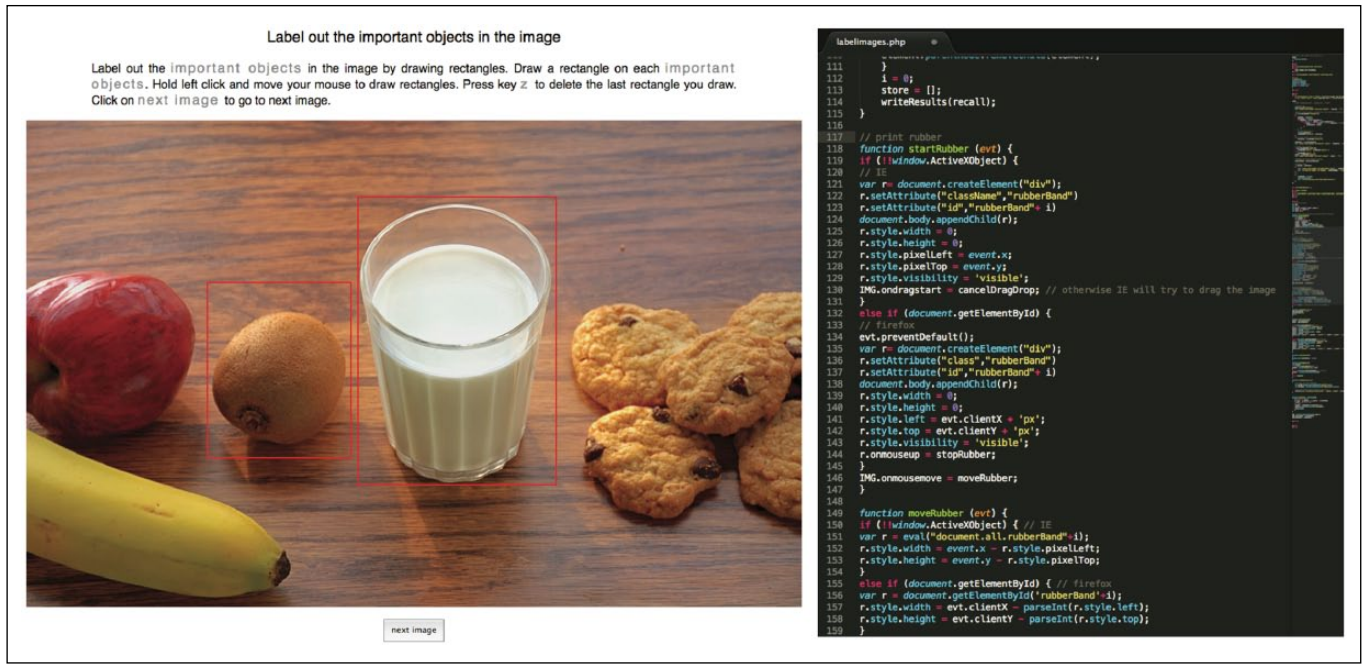

Figure 5. Screen captures of the Web application designed for the Group's crowdsourcing experiment. On the left, the application when run by a Web browser. Once a worker creates a username, he/she can start the experiment and draw rectangles. When the worker clicks on the "Next Image" button, the coordinates of the rectangles are stored in TXT files on the Lab's server. On the right, one excerpt of one of the seven scripts required to realize such interactive labels and data storage.

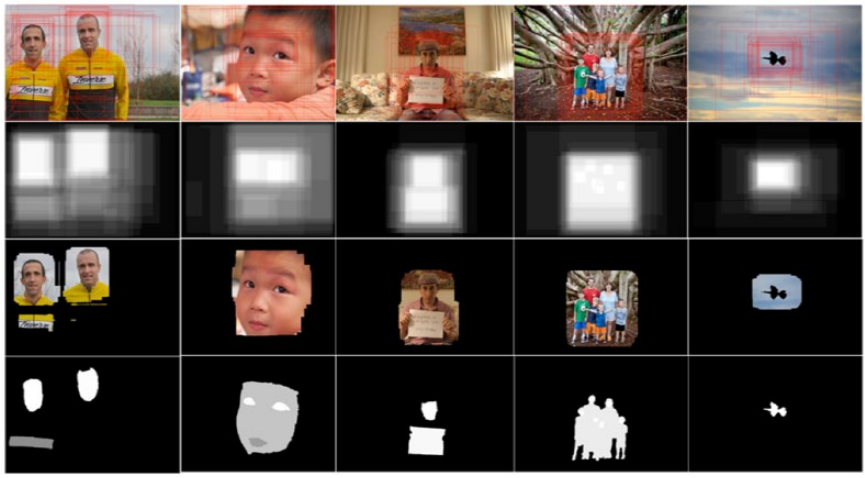

Figure 6. Matlab table summarizing the different steps required for the processing of the coordinates produced by the participants in the crowdsourcing experiment. The first row shows examples of images and rectangular labels collected from the crowdsourcing experiment. The second row shows the weight maps obtained from the superposition of the labels. The third row shows the salient regions produced by using Otsu's (1979) threshold. The last row presents the final targets with relative saliency values. If the first three steps can be automated, the last segmentation step must be done manually. At the end of this process, the images (first row, without the labels) and their corresponding targets (last row) are gathered in a single database that constitutes the Group's ground truth.

previously collected dataset was linked with many different rectangles drawn by the workers. By superimposing all the coordinates of the different rectangles on Matlab, the Group created for each image a 'weight map' with varying intensities that indicated the relative consensus on salient regions (see Figure 6). The Group then applied to each image a widely used threshold taken from Otsu (1979) – included in Matlab's library – to keep only weighty regions considered salient by the workers. In a third step that took an entire week, the Group manually segmented the contours of the salient elements within the salient regions to obtain 'salient features'. Finally, the Group assigned the mean value of the salient regions' map to the corresponding salient features in order to obtain the final targets capable of defining and evaluating a new class of saliency-detection algorithms. This laborious process took place between February and March 2014; almost a month was dedicated to the post-processing of the coordinates produced by the workers and collected by the HTML-JavaScript-PHP scripts and database.

By March 2014, the Group had successfully managed to create targets with relative saliency values. The selected images and their corresponding targets could then be gathered in a single database that finally constituted the new ground truth. At this point, the Group had managed to redefine the terms of the saliency problem: The transformations that the desired algorithm should conduct were materially defined. Thanks to the definition of inputs (the

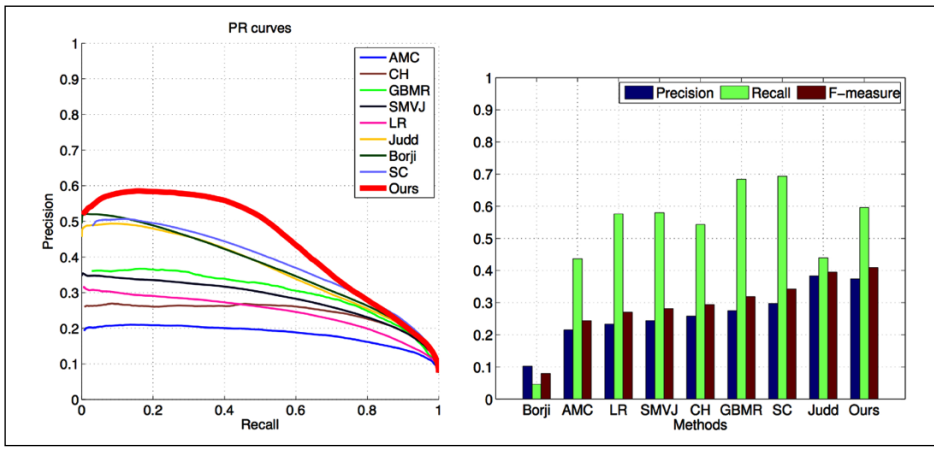

Figure 7. Two Matlab graphs that compare the performances of the Group's algorithm ('Ours') to already published ones ('AMC', 'CH', etc.). The evaluation set of the new ground truth defines the referential space of both graphs. In the graph on the left, the curves represent the variation of precision ('y' axis) and recall ('x' axis) scores for all the images in the ground truth when processed by each algorithm. In the graph on the right, histograms measure the same data while also including F-measures, the weighted average of precision and recall values. Both graphs indicate that, according to the evaluation set of the new ground truth, the Group's algorithm perform significantly better than all state-of-the-art saliency-detection algorithms.

selected images) and the definition of outputs (the targets), the Group finally possessed a problem that linear algebra and numerical computing could potentially solve.

Of course, establishing the terms of a problem with a new ground truth was not enough: In order to propose an actual algorithm, the Group also had to design and code lists of instructions that could effectively transform input data into output targets. In order to design and code these lists of instructions, the Group randomly selected 250 images out of the ground truth to form a training set. After formal analysis of the relationships between the inputs and the targets of this training set, the Group extracted several numerical features that expressed – though not completely – these input-target relationships.[13] The whole process of extracting and verifying numerical features and parameters from the training set and implementing them sequentially in the Matlab programming language took almost a month. But at the end of this process, the Group possessed a list of Matlab instructions that were able to transform the input values of the training set into values relatively close to those of the targets.

By the end of March 2014, the Group used the remainder of its ground truth database to evaluate the algorithm and compare it with available saliency-detection algorithms in terms of precision and recall measures (see Figure 7). The results of this confrontation being satisfactory, the features and performances of the Group's algorithm were summarized in a draft paper and submitted to an important European Conference on image processing.

The finalization of Matlab lists of instructions capable of solving the newly defined problem of saliency followed the problematization process in which the Group was engaged. The theoretical reformulation of saliency, the selection of specific images on

Flickr, the coding of a web application, the creation of a Matlab database, the processing of the users' coordinates, all of these practices were required in order to design the ground truth that allowed the creation and evaluation of the algorithm. Of course, the work required for the construction of the ground truth was not sufficient: Critical mathematical choices also needed to be articulated and expressed in machine-readable format. Yet by providing the training set to extract the numerical features of the algorithm and by providing the evaluation set to measure the algorithm's performances, the ground truth – and the process that led to its realization – contributed to the completion of the algorithm.

This characteristic of computational models, being bound to manually gathered and processed data, is not limited to the field of digital image processing. In climatology, the tedious collection, standardization and compilation of weather data in order to produce accurate ground truths – 'data images' – of the Earth's climate are crucial for the both the parametrization and evaluation of General Circulation Models (Edwards, 2013). In the case of machine-learning algorithms for handwriting recognition or spam filtering, 'test data' is important for setting the learning parameters of these algorithms, as well as for evaluating their performances (Burrell, 2016: 5). In computational finance, detailed analysis of former financial transactions as well as the authoritative literature of neoclassical financial economics constitute crucial empirical bases for the shaping and evaluation of 'execution' and 'proprietary trading' algorithms (MacKenzie, 2014: 17–31).[14] Thus behind many algorithms lies a ground truth database that has been used to extract relevant numerical features and evaluate the accuracy of the automated transformations of inputs into targets. Consequently, as soon as such algorithms – 'in the wild', outside of their production sites – automatically process some new data, their respective ground truths are invoked and, to a certain extent, reproduced. Studying the performative effects of such algorithms in the light of the collective processes that constituted the targets that these algorithms try to retrieve could be in turn a stimulating research agenda. This idea that ground truths – and the problematization processes they result from – could be interesting for the study of algorithms will be further developed in the discussion part of this paper.

Almost accepted (yet rejected)

June 19, 2014: The reviewers rejected the Group's paper. The Group was greatly disappointed to see several months of meticulous work unrewarded by a publication, one that they had hoped would launch new research lines and generate many citations. But the feeling was also one of incomprehension and surprise in view of the reasons provided by the three reviewers.

Along with doubts about the usefulness of incorporating face information within saliency detection, the reviewers agreed on one seemingly key deficiency of the Group's paper: The performance comparisons of the computational model were made only with respect to the Group's new ground truth.

> **Assigned Reviewer 1:**
>
> The method does not show that the proposed method also performs better than other state-of-the-art methods on public benchmark ground truths. … The experiment evaluation in this paper

> is conducted only on the self-collected face images. More evaluation datasets will be more convincing. … More experiments need to be done to demonstrate the proposed method.
>
> **Assigned Reviewer 2:**
>
> The experiments are tested only on the ground truth created by the authors. … It would be more insightful if experiments on other ground truths were carried out, and results on face images and non-face images were reported, respectively. This way one can more thoroughly evaluate the usefulness of a face-importance map.
>
> **Assigned Reviewer 3:**
>
> The discussion is still too subjective and not sufficient to support its scientific insights. Evaluation on existing datasets would be important in this sense.

The reviewers found the technical aspects of the paper to be sound. But they questioned whether the new best saliency-detection model – as the Group presents it in the paper – could be evaluated with only the ground truth used to create it. Indeed, why not confront this new model with the already-available ground truths for saliency-detection? If the model were really 'more efficient' than the already published ones, it should also be more efficient on the ground truths used to shape and evaluate the performances of the previously published saliency-detection models. In other words, since the Group presented its model as *commensurable* with former models, the Group should have – according to the reviewers – more thoroughly compared its performance with those others. But why did the Group stop halfway through its evaluation efforts and compare its model only with respect to the new ground truth?

> **Discussion with BJ, Lab's cafeteria, 19 June 2014:**
>
> **FJ**: The committee didn't like that we created our own ground truth?
> **BJ**: No. I mean, it's just that we tested on this one but we did not test on the other ones.
> **FJ**: They wanted you to test on already existing ground truths?
> **BJ**: Yes.
> **FJ**: But why didn't you do that?
> **BJ**: Well, that's the problem: Why did we not test it on the others? We have a reason. Our model is about face segmentation and multiple features. But in the other datasets, most of them do not have more than ten face images. … In the saliency area, most people do not work on face detection and multiple features. They work on images where there is a car or a bird in the center. You always have a bird or something like this. So it just makes no sense to test our model on these datasets. They just don't cover what our model does. … That's the thing: If you do classical improvement, you are ensured that you will present something [at important conferences]. But if you have new things, then somehow people just misunderstand the concept.

It would not have been technically difficult for the Group to confront its model with the previous ground truths: They are available on the Internet and such performance evaluations require roughly the same Matlab scripts as those used to produce the results shown

in Figure 7. The main reason the Group did not do such comparisons is that the earlier models deriving from earlier ground truths would have obtained better performance results: since the Group's model was *not* designed to solve the saliency problem as defined by the previous ground truths, it would have been outperformed by these ground truths' 'native' models.

The rejection of the Group's paper shows again how image-processing algorithms are bound to their ground truths. An algorithm deriving from a ground truth constituted by images whose targets are centered contrastive objects will manage to retrieve these targets. But when tested on a ground truth constituted by images whose targets are multiple decentered objects and faces, the same algorithm may well produce statistically poor results. Similarly, another algorithm deriving from a ground truth constituted of images whose targets are multiple decentered objects and faces will manage to retrieve these targets. But when tested on a ground truth constituted of images whose targets are centered contrastive objects, it may well produce statistically poor results. The algorithms operate in different categories, and their limits lie in the ground truths used to define their range of actions. As BJ suggests, in a dramatic way, to a certain extent we get the algorithms of our ground truths. Algorithms can be presented as statistically more efficient than others only when they derive from the same – or very similar – ground truths. As soon as two algorithms derive from two ground truths with different targets, they can only be presented as different. Qualitative evaluations of the different ground truths in terms of methodology, data selection, statistical rigor or industrial potential can be conducted, but the two computational models themselves are irreducibly different and not commensurable. From the point of view of this study – which may differ from the point of view of the reviewers – the Group's mistake may have been to mix up quantitative improvement of performances with qualitative refinement of ground truths.

Interestingly, one year after this rejection episode, the Group submitted another paper, this time to a smaller conference in image processing. The objects of this paper were rigorously the same as those of the paper that was previously rejected: the same ground truth and the same computational model. Yet instead of highlighting the statistical performances of its model, the Group emphasized its ground truth and the fact that it allows the inclusion of face segmentation within saliency detection. In this second paper that won the 'Best Short Paper Award' of the conference, the computational model was presented as one example of the potential of the new ground truth.

Discussion

This case study – offered as a contribution to bringing algorithms 'down to earth' (Bogost, 2015) and knowing them better (Seaver, 2013) – accounted for a four-month project in saliency detection run by a group of young computer scientists at a European technical institute. For the Group, since industrial applications of saliency-detection algorithms should mimic humans who – according to recent papers in cognitive science – consider faces to be salient features of images, saliency-detection algorithms should also take faces into account. Moreover, as suggested by recent studies in cognitive science collected by the Group, saliency could be seen as a continuous problem, rather than a binary one. Two other elements further contributed to this initial reformulation of the saliency

problem: The Group's previous design of computational models for image segmentation and face detection, and affordable crowdsourcing services that can facilitate social pooling.

Such a reformulation of the saliency problem required the construction of a ground truth database. Indeed, without operational targets of segmented faces and multiple objects, the Group could neither extract numerical features nor evaluate any computational model that might detect them. A new ground truth database gathering a new set of 'natural' images – the inputs – and their manually labelled counterparts – the output 'targets' – thus appeared as a prerequisite for the completion of the Group's desired saliency-detection algorithm. The actual construction of this ground truth involved the selection of a dataset and the shaping and processing of a crowdsourcing experiment. Even though these mundane practices are usually not accounted for in computer science papers and presentations, they are crucial, allowing the constitution of a training set and an evaluation set. The training set was used to extract numerical features that partially expressed the relationships between the inputs and the targets. Once these features were implemented in the Matlab programming language, the whole list of instructions – the algorithm – could be evaluated thanks to the evaluation set. Even though the Group's algorithm could not be reduced to its ground truth, the very existence of the algorithm depended on it.

Reviewers of the Group's academic paper on the algorithm, for an important conference on image processing, were critical. According to them, since the algorithm was presented as more efficient than already published ones, the Group should have conducted comparisons on the available ground truths. But since the available ground truths did not contain the targets that the Group's model was designed to detect, confronting the model with these ground truths would have produced no meaningful results.

This case study suggests two complementary ways of seeing algorithms. First, an algorithm can be considered to be a set of instructions designed to solve a given problem computationally (Cormen et al., 2009). At the end of the Group's project, once the numerical features were extracted from the training set and translated into machine-readable language, several Matlab files with thousands of instructions were solving a given problem. In that sense, the study of such sets of instructions at a theoretical level is fully relevant: How should numbers and machine-readable languages be used to propose a solution to a given problem in the most efficient way?

At the same time, however, the problem an algorithm is designed to solve is the result of a problematization process: a succession of collective practices that aim at defining the *terms* of a problem to be solved. In my case study, the Group first reformulated the saliency problem as face-related and continuous. This first step of the Group's problematization process included mundane and problematic practices such as the critique of previous research results and the inclusion of some of the Lab's recent projects. The second step of the Group's problematization process demanded the constitution of a ground truth that could operationalize the newly formulated problem of saliency. This second step also included mundane and problematic practices, such as the collection of a dataset on Flickr, the organization of a database, the design of a crowdsourcing experiment, and the processing of the results. Only at the very end of this process – once the laboriously constructed targets have been associated with the laboriously constructed dataset to form

the final ground truth database – was the Group able to code and evaluate the set of Matlab instructions capable of transforming inputs into outputs by means of numerical computing techniques. In short, in order to design a numerical method that could solve the new saliency problem, the Group first had to define step by step the boundaries of this new problem.

Thus a first perspective might consider the Group's algorithm as a set of instructions designed to computationally solve a new problem in the best possible way. This *axiomatic* way of considering the Group's algorithm would in turn put the emphasis on the mathematical choices and programming procedures the Group used in order to transform the input values of the new ground truth into their corresponding output values. Did the Group extract relevant numerical features for such task? Did the Group optimally implement the Matlab instructions? In short, this take on the Group's algorithm would analyze it regarding its computational properties.

A second perspective on the Group's algorithm might consider it as a set of instructions designed to computationally retrieve in the best possible way the outputs designed during a specific problematization process. This *problem-oriented* way of considering the Group's algorithm would put the emphasis on the specific situations and practices that led to the definition of the terms of the problem the algorithm was designed to solve. How was the problem defined? How was the dataset collected? How was the crowdsourcing experiment conducted? In short, this take on the Group's algorithm would analyze it with regard to the construction process of the ground truth from which the algorithm ultimately derived.

While the axiomatic and problem-oriented perspectives on algorithms are complementary and should thus be intimately articulated – specific numerical features being suggested by ground truths (and vice-versa) – they draw attention to different practices. By considering the terms of the problem at hand as given, the axiomatic way of considering algorithms may facilitate the definition of the mathematical and programming procedures that end up transforming input sets of values into output sets of values in the best possible ways; by assuming that the transformation of the inputs into the outputs is desirable and relevant, a step-by-step scheme describing this transformation might be proposed. In the case of computer science, different areas of mathematics with many different rules and theorems can be explored and adapted to automate the passage from given inputs to specified outputs: Linear algebra in the case of image processing, probability theory in the case of data compression, graph theory in the case of data structure, number theory in the case of cryptography, etc. Yet for each specific case, the definition of lists of instructions designed to computationally solve a problem will go along with the acceptance of the problem's terms. Acceptance of the terms of a problem is precisely what *enables* the definition of mathematical procedures and their translation into machine-readable languages that, in the end, effectively transform inputs into outputs.

If the problem-oriented perspective on algorithms may not directly facilitate the completion of lists of instructions, it may contribute to 'grounding' discussions about the role of algorithms in public life. Considering algorithms as retrieving entities may put the emphasis in the referential databases that define what algorithms try to retrieve and reproduce. What ground truth defined the terms of the problem this algorithm tries to solve? How was this ground-truth database constituted? When, and by whom?

By pointing at moments and locations where outputs-to-be-retrieved were – or are being – constituted within ground-truths databases, this analytical look at algorithms may suggest new ways of interacting with algorithms and those who design them. We saw that computer scientists rely in part on well-designed referential repositories in order to propose innovative algorithms. And while computer scientists use their intuitions, values, affects, intelligence, and technical skills to effectively define problems, gather orientated datasets and shape targets, some of these problematization processes may benefit from a broader inclusion of individuals who have invested years of efforts in similar topics (cognitive psychologists, linguists, political scientists, artists, citizen, activists, etc.).[15] If we – computer scientists, artists, citizens, activists, consumers, social scientists – need algorithms to engage with the world, many algorithms – as performative retrieving entities – also need problematization practices and ground truths to come into existence. Next to the important questions of *how* to retrieve by means of numerical procedures lies an equally important question of *what* to retrieve by means of numerical procedures. If many algorithms derive from ground-truth databases, working collectively upon the latter may provide refreshing takes on the former.

Acknowledgements

I am immensely grateful to the members of the computer science laboratory who let me follow their activities and more generally helped me with the research. I also want to express my deepest gratitude to Dominique Vinck, Geoffrey Bowker, Alexandre Camus, Aubrey Slaughter, Roderick Crooks, Myles Jeffrey, the anonymous reviewers, and the editors of this journal for their support, help, and insightful suggestions.

Funding

This work was supported by SNF grant Doc.ch (SHS) POLAP1 148948.

Notes

1. Ziewitz (2016) provides a successful account of this depiction of algorithms as both powerful and inscrutable, setting up a real drama. This 'algorithmic drama' (p. 5) comes in two series of studies: the first series of studies presents algorithms as powerful value-laden actors (Beer, 2009; Bucher, 2012; Gillespie, 2014; Hallinan and Striphas, 2016; Introna and Nissenbaum, 2000; Introna and Wood, 2004; Kraemer et al., 2011; Kushner, 2013; Steiner, 2012); the second, symmetric, series of studies presents them as opaque and inscrutable (Anderson, 2011; Diakopoulos, 2015; Graham, 2005). As Ziewitz (2016) puts it '[i]nterestingly, there is a certain recursiveness in this drama: opacity of operations tends to be read as another sign of influence and power' (p. 6).
2. In this paper, I do not discuss the problematic relationships between formal mathematical expressions and their machine-readable counterparts. The term 'algorithm' is highly equivocal, especially among computer scientists: It sometimes refers to formal mathematical expressions, other times to lines of code able to instruct electric pulses. The term 'computational model' is generally used by computer scientists to include these two aspects that are not strictly equivalent and that maintain complex links. For further discussions, see the instructive disputes in De Millo et al. (1979) and Dijkstra (1978), further analyzed in Fetzer (1988) and MacKenzie (1993, 1995).

3. A digital signal is represented by *n* dimensions depending on the independent variables used to describe the signal. A sampled digital sound is for example described as a one-dimensional signal whose dependent variables – amplitudes – vary according to time (*t*); a digital image is usually described as a two-dimensional signal whose dependent variables – intensities – vary on two axes (*x, y*), whereas an audio-visual content will be described as a three-dimensional signal with independent variables (*x, y, t*). For an accessible introduction to signal processing, see Vetterli et al. (2014).
4. Another important neurobiological model of selective-attention method was proposed in Wolfe et al. (1989). This model later inspired competing low-level feature computational models (Tsotsos, 1989; Tsotsos et al., 1995).
5. See Elazary and Itti (2008); Itti and Koch (2001); Itti et al. (1998, 2000); Zhao and Koch (2011).
6. More generally, precision and recall measures are the two pillars of 'information retrieval'. For an introduction to this field, see Manning et al. (2008).
7. Different high-level detection algorithms can nonetheless be assembled as modules in a single program that could for example detect faces *and* cars *and* dogs, etc.
8. At that time, only two saliency-detection algorithms were published, in Itti et al. (1998) and Ma and Zhang (2003). But the ground truths used for the design and evaluation of these algorithms were similar to those used in laboratory cognitive science. The images of these ground truths were, for example, sets of dots disrupted by a vertical dash. Consequently, if these first two saliency-detection algorithms could process natural images, no evaluations of their performances on such images could be conducted.
9. Ground truths created by computer science laboratories are made available online in the name of reproducible research (Vandewalle et al., 2009). The counterpart to this free access is the proper citation of the papers in which these ground truths were first presented.
10. Other proposed algorithms included Achanta et al. (2009); Chang et al. (2011); Goferman et al. (2012); Shen and Wu (2012); Wang et al. (2010).
11. As Irani (2015) notes, large-scale microlabour is not new: Online chatroom moderators, 'Selectric' typewriting secretaries, and even telegraph 'messenger boys' had participated in what is now called the 'information economy'. Yet one undeniable novelty of crowdsourcing is that it allows 'the distribution, collection, and processing of data work at high speeds and large scales' (Irani, 2015: 226). This acceleration and volume growth of data work participates in creating emerging forms of the precariat. For a study of the demographics of crowd workers, see Ross et al. (2010).
12. Matlab is a privately held mathematical software for numerical computing built around its own interpreted high-level programming language. Because of its agility for problems of linear algebra – all integers being considered as scalars – Matlab is widely used for research and industrial purposes in computer science, electrical engineering and economics. Yet, as Matlab works mainly with an interpreted programming language, its programs have to be translated into machine-readable binary code by an *interpreter* in order to actually interact with the data. This additional step makes it less efficient for processing heavy matrices than, for example, programs directly written in C or C++. For a brief history of Matlab, see (Haigh, 2008).
13. The numerical features that were extracted from the training set were related, among others, to 'spatial compactness', 'contrast-based filtering', 'high-dimensional Gaussian filters' and 'element uniqueness'.
14. For a historical study of Automatic Trading Desk, one of the first high-frequency trading firms, see MacKenzie (2017).
15. Several studies have tentatively suggested the need for cooperative design of datasets and ground truths (e.g. Bechmann, 2017; Torralba and Efros, 2011; Vandewalle et al., 2009).

References

Achanta R, Hemami S, Estrada F and Susstrunk S (2009) Frequency-tuned salient region detection. In: *Proceedings of the 2009 IEEE conference on computer vision and pattern recognition*, Miami, FL, 20–25 June. New York: IEEE, pp. 1597–1604.

Ananny M (2016) Toward an ethics of algorithms convening, observation, probability, and timeliness. *Science, Technology & Human Values* 41(1): 93–117.

Anderson CW (2011) Deliberative, agonistic, and algorithmic audiences: Journalism's vision of its public in an age of audience transparency. *International Journal of Communication* 5: 529–547.

Bechmann A (2017) Keeping it real: From faces and features to social values in deep learning algorithms on social media images. In: *Proceedings of the 50th Hawaii international conference on system sciences*, Waikoloa, HI, 4–7 September. New York: IEEE, pp. 1793–1801.

Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11(6): 985–1002.

Bogost I (2015) The cathedral of computation. *The Atlantic*, 15 January. Available at: https://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/ (accessed 11 September 2017).

Borji A (2012) Boosting bottom-up and top-down visual features for saliency estimation. In: *Proceedings of the 2012 IEEE conference on computer vision and pattern recognition*, Providence, RI, 16–21 June. New York: IEEE, pp. 438–445.

Bucher T (2012) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society* 14(7): 1164–1180.

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1): 1–12.

Callon M (1986) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. In: Law J (ed.) *Power, Action and Belief: A New Sociology of Knowledge?* New York: Routledge, pp. 196–223.

Chang KY, Liu TL, Chen HT and Lai SH (2011) Fusing generic objectness and visual saliency for salient object detection. In: *Proceedings of the 2011 IEEE international conference on computer vision*, Barcelona, 6–13 November. New York: IEEE, pp. 914–921.

Chen LQ, Xie X, Fan X, Ma WY, Zhang HJ and Zhou HQ (2003) A visual attention model for adapting images on small displays. *Multimedia Systems* 9(4): 353–364.

Cormen TH, Leiserson CE, Rivest RL and Stein C (2009) *Introduction to Algorithms*, 3rd edn. Cambridge, MA: The MIT Press.

Crawford K (2016) Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology & Human Values* 41(1): 77–92.

De Millo RA, Lipton RJ and Perlis AJ (1979) Social processes and proofs of theorems and programs. *Communications of the ACM* 22(5): 271–280.

Diakopoulos N (2015) Algorithmic accountability: *Journalistic investigation of computational power structures*. *Digital Journalism* 3(3): 398–415.

Dijkstra EW (1978) On a political pamphlet from the middle ages. *ACM SIGSOFT Software Engineering Notes* 3(2): 14–16.

Edwards PN (2013) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Elazary L and Itti L (2008) Interesting objects are visually salient. *Journal of Vision* 8(3): 1–15.

Estellés-Arolas E and González-Ladrón-de-Guevara F (2012) Towards an integrated crowdsourcing definition. *Journal of Information Science* 38(2): 189–200.

Fetzer JH (1988) Program verification: The very idea. *Communications of the ACM* 31(9): 1048–1063.

Fujimura JH (1987) Constructing 'do-able' problems in cancer research: Articulating alignment. *Social Studies of Science* 17(2): 257–293.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies*. Cambridge, MA: The MIT Press, pp. 167–194.

Goferman S, Zelnik-Manor L and Ayellet T (2012) Context-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(10): 1915–1926.

Graham SDN (2005) Software-sorted geographies. *Progress in Human Geography* 29(5): 562–580.

Haigh T (2008) Cleve Moler: Mathematical software pioneer and creator of Matlab. *IEEE Annals of the History of Computing* 30(1): 87–91.

Hallinan B and Striphas T (2016) Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society* 18(1): 117–137.

Heinke D and Humphreys DG (2005) Computational models of visual selective attention: A review. In: Houghton G (ed.) *Connectionist Models in Cognitive Psychology*. London: Psychology Press, 273–312.

Hjelmås E and Low BK (2001) Face detection: A survey. *Computer Vision and Image Understanding* 83(3): 236–274.

Introna L (2016) Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology & Human Values* 41(1): 17–49.

Introna L and Nissenbaum H (2000) Shaping the web: Why the politics of search engines matters. *The Information Society* 16(3): 169–185.

Introna L and Wood D (2004) Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society* 2(2–3): 177–198.

Irani L (2015) Difference and dependence among digital workers: The case of Amazon Mechanical Turk. *South Atlantic Quarterly* 114(1): 225–234.

Itti L (2000) *Models of bottom-up and top-down visual attention*. PhD Thesis, California Institute of Technology, Pasadena, CA.

Itti L and Koch C (2001) Computational modelling of visual attention. *Nature Reviews Neuroscience* 2(3): 194–203.

Itti L, Koch C and Braun J (2000) Revisiting spatial vision: Toward a unifying model. *Journal of the Optical Society of America. Optics, Image Science, and Vision* 17(11): 1899–1917.

Itti L, Koch C and Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(11): 1254–1259.

Jackson SJ, Gillespie T and Payette S (2014) The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing. In: *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing*, Baltimore, MD, 15–19 February. New York: ACM, pp. 588–602.

Jiang B, Zhang L, Lu H, Yang C and Yang MH (2013) Saliency detection via Absorbing Markov Chain. In: *IEEE international conference on computer vision*, Sydney, Australia, 1–8 December. New York: IEEE, pp. 1665–1672.

Judd T, Durand F and Torralba A (2012) A benchmark of computational models of saliency to predict human fixations. Report, MIT, 13 January. Available at: http://dspace.mit.edu/handle/1721.1/68590 (accessed 12 February 2015).

Karthikeyan S, Jagadeesh V and Manjunath BS (2013) Learning top down scene context for visual attention modeling in natural images. In: *Proceedings of the 2013 international conference on image processing*, Melbourne, VIC, Australia, 15–18 September. New York: IEEE, pp. 211–215.

Knobel C and Bowker GC (2011) Values in design. *Communications of the ACM* 54(7): 26–28.

Knorr-Cetina KD (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon Press.

Koch C and Ullman S (1985) Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology* 4(4): 219–227.

Kraemer F, Overveld K and Peterson M (2011) Is there an ethics of algorithms? *Ethics and Information Technology* 13(3): 251–260.

Kushner S (2013) The freelance translation machine: Algorithmic culture and the invisible industry. *New Media & Society* 15(8): 1241–1258.

Latour B (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Latour B (2010) An attempt at a 'compositionist manifesto'. *New Literary History* 41(3): 471–490.

Latour B and Woolgar S (1986) *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Liu T, Sun J, Zheng NN, et al. (2007) Learning to detect a salient object. In: *Proceedings of the 2007 IEEE conference on computer vision and pattern recognition*, Minneapolis, MN, 17–22 June. New York: IEEE, pp. 1–8.

Lowe DG (1999) Object recognition from local scale-invariant features. In: *Proceedings of the Seventh IEEE international conference on computer vision*, Kerkyra, 20–27 September. New York: IEEE, pp. 1150–1157.

Lynch M (1985) *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. London: Routledge & Kegan Paul.

Ma YF and Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: *Proceedings of the eleventh ACM international conference on multimedia*, Berkeley, CA, 2–8 November. New York: ACM, pp. 374–381.

MacKenzie D (1993) Negotiating arithmetic, constructing proof: The sociology of mathematics and information technology. *Social Studies of Science* 23(1): 37–65.

MacKenzie D (1995) The automation of proof: A historical and sociological exploration. *IEEE Annals of the History of Computing* 17(3): 7–29.

MacKenzie D (2014) A sociology of algorithms: High-frequency trading and the shaping of markets (Unpublished paper). Available at: http://www.sps.ed.ac.uk/__data/assets/pdf_file/0004/156298/Algorithms25.pdf (accessed 21 March 2017).

MacKenzie D (2017) A material political economy: Automated trading desk and price prediction in high-frequency trading. *Social Studies of Science* 47(2): 172–194.

Manning CD, Raghavan P and Schütze H (2008) *Introduction to Information Retrieval*. New York: Cambridge University Press.

Movahedi V and Elder JH (2010) Design and perceptual validation of performance measures for salient object segmentation. In: *Proceedings of the 2010 IEEE computer society conference on computer vision and pattern recognition workshops*, San Francisco, CA, 13–18 June. New York: IEEE, pp. 49–56.

Muniesa F (2011) Is a stock exchange a computer solution? Explicitness, algorithms and the Arizona stock exchange. *International Journal of Actor-Network Theory and Technological Innovation* 3(1): 1–15.

Neyland D (2016) Bearing account-able witness to the ethical algorithmic system. *Science, Technology, & Human Values* 41(1): 50–76.

Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1): 62–66.

Ross J, Irani L, Six Silberman M, Zaldivar A and Tomlinson B (2010) Who are the crowdworkers? Shifting demographics in Mechanical Turk. In: *Proceedings of the CHI '10 extended abstracts on human factors in computing systems*, Atlanta, GA, 10–15 April. New York: ACM, pp. 2863–2872.

Santella A, Agrawala M, DeCarlo D, Salesin D and Cohen M (2006) Gaze-based interaction for semi-automatic photo cropping. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, Montréal, QC, Canada, 22–27 April. New York: ACM, pp. 771–780.

Seaver N (2013) Knowing algorithms. *Media in Transition* 8: 1–12.

Seitz F and Einspruch NG (1998) *Electronic Genie: The Tangled History of Silicon*. Chicago, IL: University of Illinois Press.

Shen X and Wu Y (2012) A unified approach to salient object detection via low rank matrix recovery. In: *Proceedings of the 2012 IEEE conference on computer vision and pattern recognition*, Providence, RI, 16–21 June. New York: IEEE, pp. 853–860.

Simondon G (2017 [1958]) *On the Mode of Existence of Technical Objects* (trans. C Malaspina and J Rogove). Minneapolis, MN: Univocal Publishing.

Steiner C (2012) *Automate This: How Algorithms Came to Rule Our World*. New York: Portfolio Hardcover.

Suchman L (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press.

Torralba A and Efros AA (2011) Unbiased look at dataset bias. In: *Proceedings of the 2011 IEEE conference on computer vision and pattern recognition*, Colorado Springs, CO, 20–25 June. New York: IEEE, pp. 1521–1528.

Tsotsos JK (1989) The complexity of perceptual search tasks. In: *Proceedings of the 11th international joint conference on artificial intelligence*, Detroit, MI, 20–25 August. New York: ACM, pp. 1571–1577.

Tsotsos JK, Culhane SM, Kei Wai WY, Lai Y, Davis N and Nuflo F (1995) Modeling visual attention via selective tuning. *Artificial Intelligence* 78(1–2): 507–545.

Vandewalle P, Kovačević J and Vetterli M (2009) Reproducible research in signal processing. *IEEE Signal Processing Magazine* 26(3): 37–47.

Vetterli M, Kovačević J and Goyal VK (2014) *Foundations of Signal Processing*. Cambridge: Cambridge University Press.

Vinck D (2011) Taking intermediary objects and equipping work into account in the study of engineering practices. *Engineering Studies* 3(1): 25–44.

Wang W, Wang Y, Huang Q and Gao W (2010) Measuring visual saliency by site entropy rate. In: *Proceedings of the 2010 IEEE conference on computer vision and pattern recognition*, San Francisco, CA, 13–18 June. New York: IEEE, pp. 2368–2375.

Wang Z and Li B (2008) A two-stage approach to saliency detection in images. In: *Proceedings of the IEEE international conference on acoustics, speech and signal processing*, Las Vegas, NV, 31 March–4 April. New York: IEEE, pp. 965–968.

Wolfe JM, Cave KR and Franzel SL (1989) Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance* 15(3): 419–433.

Zarsky T (2016) The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology & Human Values* 41(1): 118–132.

Zhang C and Zhang Z (2010) A survey of recent advances in face detection. Report, Microsoft Research. Available at: https://www.microsoft.com/en-us/research/publication/a-survey-of-recent-advances-in-face-detection/ (accessed 24 June 2017).

Zhao Q and Koch C (2011) Learning a saliency map using fixated locations in natural scenes. *Journal of Vision* 11(3): 9.

Ziewitz M (2016) Governing algorithms: Myth, mess, and methods. *Science, Technology & Human Values* 41(1): 3–16.

Author biography

Florian Jaton is PhD student in Social Sciences at the University of Lausanne, Switzerland. He is part of the Institute of Social Sciences, the STS Lab and the Laboratory of Digital Humanities of the University of Lausanne (LaDHUL). As well as examining referential repositories for algorithmic procedures, his dissertation deals with computer programming practices and the shaping of mathematical formulas. With Dominique Vinck, he recently edited a special issue for the *Revue d'Anthropologie des Connaissances* entitled 'What data make humanities do (and vice-versa): Unfolding frictions in database projects.'