Prettiness Algorithms: Smartphones and the Manipulation of Taste

Leonardo Impett
University of Cambridge

1. Neural Models of Visual Culture[1]

In 2016, as a graduate student, I got involved in organising a conference called *Ways of Machine Seeing* at Darwin College, Cambridge[2]. The conference's main idea - sadly not my own - was to encourage people to read across two major works on seeing.

The first, John Berger's *Ways of Seeing*(1972), is seminal text (and BBC documentary) on how vision is culturally and socially situated – a crucial problem for art history since the beginning of the last century[3]. Berger's focus was on the historical and socioeconomic relations surrounding pictures, both as they were originally made (e.g. displays of wealth in still-lives) and as they are seen now (e.g. how different economic classes go to galleries). A contemporary review praised his "imaginative and often innovative use of Marxian method"[4]; Berger built on previous Marxist theorists such as Walter Benjamin, and proposed radical new ideas of his own, including the term "male gaze". The second, David Marr's *Vision* (1982), lays out many of the fundamental assumptions of computer vision - partly through an analysis of human vision. The two texts are a several generations old; but both have survived through their central place in the pedagogy of their respective fields.

What this juxtaposition heavily implied was the idea that computer vision models are themselves *ways of seeing* they contain just as much cultural, racial, social, and historical baggage as human vision – though undoubtedly encoded in a different way. This notion allows us to move on from the idea that computer vision or machine learning models are simply *biased*: which is true, but unhelpfully implies the possibility of an unbiased model. An unbiased model would be theoretically possible but practically useless, since we are forced to expose computers to a particular cultural or historical viewpoint whenever we teach them what any man-made thing looks like. Algorithms, explains Louise Amoore, "must necessarily discriminate to have any traction in the world"[5].

To be clear, there is no doubt that algorithms (in computer vision and elsewhere) are often discriminatory. Algorithms used by U.S. courts to determine bail risks are hugely discriminatory against black defendants[6]; commercial face detection systems perform 10-20% worse on darker female faces than lighter male faces[7]; and object detection systems recognise everyday items like

---

[1] This is an edited version of an article which appeared in *EthicAI=Artificial Intelligence & Ethics*, republished by kind permission of the Goethe-Institute Bulgaria.

[2] The conference was organised by Alan Blackwell and Anne Alexander; it has since spawned various other events, networks, and publications, including a recent special issue: Azar, Mitra, Geoff Cox, and Leonardo Impett. "Introduction: ways of machine seeing." *AI & SOCIETY* (2021): 1-12.

[3] E.g. Heinrich Wölfflin's 'History of Seeing'; see Davis, Whitney. "Succession and Recursion in Heinrich Wölfflin's Principles of Art History." *The Journal of Aestheticsand Art Criticism* 73, no. 2 (2015): 157-164.

[4] Wallach, Alan. "John Berger, Ways of Seeing" (book review), in *Artforum* February 1976: 44-45.

[5] Amoore, Louise. *Cloud ethics*. Duke University Press, 2020: 8

[6] Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine bias." *ProPublica,* May 23 (2016): 139-159.

[7] Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." In *Conference on fairness, accountability and transparency*, pp. 77-91. PMLR, 2018.

soap and foodstuffs more easily in the global north than in the global south[8]. But *bias* flattens a complex web of power relations (class, gender, geography) and their visual symptoms to a single percentage difference. Following Berger's Marxist insight, *Ways of Machine Seeing*suggested another route; closer to literary scholar Ted Underwood's recent suggestion that "[t]o understand why neural language models are dangerous (and fascinating), we need to approach them as models of culture"[9].

This approach would have been completely alien to me as a student of computer vision, where techniques either seemed to reflect intrinsic physical reality (in the case of, say, multiple view geometry) or universal features of human vision (e.g. in the similarity between sparse coding of natural images and primary visual cortex receptive fields). But I became involved in the *Ways of Machine Seeing* conference partly because of a short research internship at Microsoft Research in Cairo a year earlier, where I worked on a computer vision problem which I will call *prettiness estimation*.

Why invent a new name for an existing field? Firstly, because computer scientists use a wide and inconsistent range of metaphors to describe the same task: aesthetic quality assessment, aesthetic image assessment, automatic visual aesthetics, photo aesthetic ranking, neural image assessment, aesthetic quality inference… And secondly, because the existing terms are often misleading. Most refer to "aesthetics" – but *prettiness estimation* has very little to do with the rich intellectual history of aesthetics. Scientists ask a set of human annotators for a rating of a digital photo out of 5 stars[10], or "how beautiful is this picture" (a sliding scale from 1 to 5, where 4 is "Professional")[11]; or simply their opinion on "photo quality"[12]. Others use amateur photography websites where online users rank each other's photographs, including Photo.net[13] and DPChallenge.com (a 2012 snapshot of which forms the Aesthetic Visual Analysis dataset, or AVA, perhaps the most widely-used benchmark)[14]. "How pretty is this photo?" is what they're really asking – hence my proposed formulation.

2. Whose prettiness?

On hearing of the existence of this family of algorithms in computer science, colleagues in the humanities most commonly raise two objections: either that the problem is useless (why would anybody need such an algorithm?), and that it is impossible. The first is the easiest to dismiss. Imagine you have a folder of images on your computer or smartphone, or perhaps a large set of photographs taken at a wedding, or during a holiday. Modern user interfaces often display a 'preview image' for such a folder or event; so we have to automatically choose a single image to represent the whole set. Perhaps some photographs are blurry, and others might be 'pocket-dials', taken by accident. We should at least be able to eliminate the *least pretty* images, and

---

[8] De Vries, Terrance, Ishan Misra, Changhan Wang, and Laurens Van der Maaten. "Does object recognition work for everyone?." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 52-59. 2019.

[9] Underwood, Ted. *Mapping the latent spaces of culture.*2021 http://dx.doi.org/10.17613/faaa-1r21

[10] Kong, Shu, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. "Photo aesthetics ranking network with attributes and content adaptation." In *European conference on computer vision*, pp. 662-679. Springer, Cham, 2016.

[11] Schifanella, Rossano, Miriam Redi, and Luca Maria Aiello. "An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures." In *Proceedings of the International AAAI Conferenceon Web and Social Media*, vol. 9, no. 1, pp. 397-406. 2015.

[12] Luo, Wei, Xiaogang Wang, and Xiaoou Tang. "Content-based photo quality assessment." In *2011 International Conference on Computer Vision*, pp. 2206-2213. IEEE, 2011.

[13] Datta, Ritendra, Dhiraj Joshi, Jia Li, and James Z. Wang. "Studying aesthetics in photographic images using a computational approach." In *European Conference on Computer Vision*, pp. 288-301. Springer, Berlin, Heidelberg, 2006.

[14] Murray, Naila, Luca Marchesotti, and Florent Perronnin. "AVA: A large-scale database for aesthetic visual analysis." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2408-2415. IEEE, 2012.

thus choose a reasonable photograph. As we shall see, this is not the only use of prettiness estimators - but it was the use-case we were considering back in 2015.

What about the second objection: how could a machine possibly reproduce deeply subjective judgements around beauty? We might assume everyone involved in creating these datasets has a deeply individual set of preferences; this turns out not to be the case. On the DPChallenge website, photo rating scores go from 1 to 10; the average standard deviation of scores for individual photographs is less than 1.4[15]. The authors of the Aesthetics and Attributes Database (AADB) find that "98.45% batches have significant agreement among raters" - and that therefore "the annotations are reliable for scientific research"[16].

Opinions on photo prettiness are consistent enough in these datasets, then, but are they reproducible algorithmically? Overall performance on the task is constantly improving, like any other in computer vision - but a 2019 model trained on DPChallenge scores gives a Pearson linear correlation coefficient (a measure of how well predictions agree with average user scores, where 1.0 is the highest) of 0.756[17]. To put this into context, the average individual user rating has a Pearson correlation of roughly 0.45 with the overall average[18]. This leads to the paradoxical conclusion that machines have reached 'superhuman' performance on an intrinsically subjective task: in the sense that algorithms are able to reproduce the average prettiness of an image far more reliably than we seem to[19].

Not only are prettiness scores in these datasets surprisingly consistent and reproducible; it appears that even their inconsistencies are predictable. In a 2016 paper[20], former colleagues of mine from the Image and Visual Representation Lab in Lausanne describe an algorithm that not only accurately predicts the average aesthetic score of DPChallenge images, but also the shape of the histogram of scores. Their model can differentiate, in other words, between images whose prettiness (or otherwise) is generally accepted, those rarer images that invoke some kind of controversy or difference of opinion.

We might suppose that the fact that these algorithms are able to crack prettiness so convincingly points to the predictability of our own taste in images. Our own taste - but who are 'we'? Clearly not all eight billion or so humans alive today, as one of the first prettiness estimation papers admitted in 2006: "Ideally, the data should have been collected from a random sample of human subjects under controlled setup, but resource constraints prevented us from doing so"[21].

It turns out that we can point to a 'we'. Machine vision datasets are often 'crowdsourced' through platforms like Amazon Mechanical Turk; making it difficult to understand whose

---

[15] Kim, Won-Hee, Jun-Ho Choi, and Jong-Seok Lee. "Subjectivity in aesthetic quality assessment of digital photographs: Analysis of user comments." In *Proceedings of the 23rd ACM international conference on Multimedia* pp. 983-986. 2015.

[16] Kong, Shu, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. "Photo aesthetics ranking network with attributes and content adaptation." In *European Conference on Computer Vision*, pp. 662-679. Springer, Cham, 2016.

[17] Hosu, Vlad, Bastian Goldlucke, and Dietmar Saupe. "Effective aesthetics prediction with multi-level spatially pooled features." In *Proceedings of the IEEE/CVF Conferenceon Computer Vision and Pattern Recognition*, pp. 9375-9383. 2019.

[18] Code by the author available at doi.org/10.6084/m9.figshare.19196321 ; note that the code makes a small simplification in measuring 1-versus-all rather than 1-versus-rest, given the large number of scores.

[19] Although this is not quite a fair test, since in prettiness estimation datasets individual users are generally being asked to transcribe their own opinion of each image, rather than their estimate of an overall average opinion.

[20] Jin, B., Segovia, M.V.O. and Süsstrunk, S., 2016, September. Image aesthetic predictors based on weighted CNNs. In 2016 *IEEE International Conference on Image Processing (ICIP)* (pp. 2291-2295). Ieee.

[21] Datta, Ritendra, Dhiraj Joshi, Jia Li, and James Z. Wang. "Studying aesthetics in photographic images using a computational approach." In *European conference on computer vision*, pp. 288-301. Springer, Berlin, Heidelberg, 2006.

judgements are being captured. Not so for the the most commonly-used dataset for prettiness estimation, the online photography competition website DPChallenge - which allows members to give their location, biographical summary, age, and even a list of cameras owned. And unlike most computer vision datasets, the selection of users in DPChallenge operates on two levels: since its users are both creating and scoring the images.

At time of writing[22], of the 10 users who have received the most votes on DPChallenge, 6 list their location as within the US (one each in Pennsylvania, Wisconsin, California, Arizona, New Jersey, Massachusetts); 2 as Canada; and 2 the UK. Eight give their age: all between 53 and 75. Five list iPhones alongside their digital cameras.

This is in no way a criticism of the DPChallenge community - it is clearly not *intended* to be a representative cross-section of the global population. Users of DPChallenge are naturally more likely to have the time and disposable income to pursue the hobby of digital photography. "Prosumers"[23], the paper presenting AVA (the dataset based on DPChallenge) calls them, in the sense that they create and consume content; but this is also the industry's name for the market segment of the most expensive amateur cameras. We don't need to be die-hard Bourdieuans to suggest that DPChallenge might constitute a social field, which attracts participants selectively (along geographical, economic, racial, professional, class lines), has them compete for forms of cultural capital (peer voting, winning challenges), and shapes their tastes. We might also hypothesise that the apparent predictability of the dataset's taste is, at least in part, down to the preselection of agents ("sample bias") and the convergent dynamics of this field.

If the taste-system DPChallenge is so predictable, what are its distinguishing characteristics? Challenge-winning photographs[24] often feature extremes of colour: either dramatically-coloured skies, captured in high-dynamic-range (e.g. over Copenhagen[25]; an Icelandic mountain[26]; or Dutch windmills[27]) or in black-and-white (of telephones[28], stairs[29], footpaths[30]). A very large number are landscapes or 'still lifes'. They frequently include domestic or wild animals; only rarely do they include people.

If we take seriously the proposal that trained neural networks are models of dataset culture, one way to see the visual logic of DPChallenge through the lens of an algorithm that's been trained on it. Lu et al, in presenting a new model trained on images from both DPChallenge and Photo.net[31], show the 15 images in the dataset that their algorithm ranks as prettiest. 13 are landscapes; the other 2 are still lifes. All are either monochrome or highly saturated, and none show any people. We might hypothesise that, in this visual logic, Komar and Melamid's 1995 landscape work *USA's Most Wanted Painting* would do rather well. Its conditions of production are

---

[22] https://www.dpchallenge.com/top_10.php?view=most_votes_given - note that the AVA dataset is a "fixed" snapshot taken before 2012. A snapshot which is roughly contemporary is given at https://web.archive.org/web/20120518214949/https://www.dpchallenge.com/top_10.php?view=most_votes_given (in this list, one of the top 10 is based in South Africa). I don't want to give the impression that users are only from North America or the UK – a surprising number turn out to be based in Iceland.

[23] Murray, Naila, Luca Marchesotti, and Florent Perronnin. "AVA: A large-scale database for aesthetic visual analysis." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2408-2415. IEEE, 2012.

[24] All previous challenge-winning photographs are visible at https://www.dpchallenge.com/challenge_archive.php

[25] https://www.dpchallenge.com/image.php?IMAGE_ID=923702

[26] https://www.dpchallenge.com/image.php?IMAGE_ID=933011

[27] https://www.dpchallenge.com/image.php?IMAGE_ID=928097

[28] https://www.dpchallenge.com/image.php?IMAGE_ID=922769

[29] https://www.dpchallenge.com/image.php?IMAGE_ID=921577

[30] https://www.dpchallenge.com/image.php?IMAGE_ID=919866

[31] Lu, Xin, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z. Wang. "Rating image aesthetics using deep learning." *IEEE Transactions on Multimedia* 17, no. 11 (2015): 2021-2034.

somewhat analogous: the artists commissioned a market research firm to gather data on customer preferences about colour, size, iconography etc, and designed a painting based on the survey outcomes.

3. Enhance!

Why is it important to understand - or at least to highlight - the cultural situatedness of prettiness estimation datasets and networks? Because their use, it turns out, goes far beyond preview image selection. Several research papers have already suggested incorporating aesthetic scores into image search ranking algorithms[32,33]. Others suggest using prettiness estimators as the basis for automatic image cropping (e.g. where a square preview of a rectangular image must be generated): keeping only the 'best' part of the image[34,35,36]. Although its model was trained on saliency rather than prettiness, Twitter's image-cropping algorithm generated controversy in October 2020 when it was shown empirically to favour the inclusion of white people over black people[37]. Through various mechanisms, then, prettiness estimation algorithms have the potential to severely influence visibility and invisibility in digital visual culture.

Through a similar logic, other aspects of the image can be manipulated in order to increase the measured prettiness of a digital image: its colour, saturation, brightness levels, and so on. Automatically enhancing images with computer vision has become one of the principal weapons in the smartphone camera arms-race of the past decade - and though there are various techniques, almost all require a training dataset of 'good' images (i.e. something like DPChallenge). Apple's algorithm, *Deep Fusion*, was marketed as a major feature of the new iPhone 11 in 2019. Google's system is instead part of Google Photos. Although we don't know which dataset the Google algorithm is trained on[38], its enhancements[39] follow the visual logic of DPChallenge: highly-saturated HDR images or monochrome geometricism.

Neural image enhancement fundamentally changes the relationship between public visual taste and neural models thereof. Once neural image enhancement algorithms are an everyday part of smartphone photography, they start to play a role in *defining* taste. Neural networks for prettiness estimation, therefore, are influencing the phenomena they aim to model. Before the invention of

---

[32] San Pedro, Jose, Tom Yeh, and Nuria Oliver. "Leveraging user comments for aesthetic aware image search reranking." In *Proceedings of the 21st internationalconference on World Wide Web*, pp. 439-448. 2012.
[33] Redi, Miriam, Frank Z. Liu, and Neil O'Hare. "Bridging the aesthetic gap: The wild beauty of web imagery." In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pp. 242-250. 2017.
[34] Yan, Jianzhou, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. "Learning the change for automatic image cropping." In *Proceedings of the IEEE conference oncomputer vision and pattern recognition*, pp. 971-978. 2013.
[35] Kao, Yueying, Ran He, and Kaiqi Huang. "Automatic image cropping with aesthetic map and gradient energy map." In *2017 IEEE International Conference on Acoustics, Speech and Ssignal Processing (ICASSP)*, pp. 1982-1986. IEEE, 2017.
[36] Wang, Wenguan, Jianbing Shen, and Haibin Ling. "A deep network solution for attention and aesthetics aware photo cropping." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, no. 7 (2018): 1531-1544.
[37] Yee, Kyra, Uthaipon Tantipongpipat, and Shubhanshu Mishra. "Image cropping on twitter: Fairness metrics, their limitations, and the importance of representation, design, and agency." *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (2021): 1-24.
[38] One research paper from Google (an experiment with Street View) explicitly chose not to use DPChallenge, instead working with images from professional photographers directly - see: Fang, Hui, and Meng Zhang. "Creatism: A deep-learning photographer capable of creating professional work." *arXiv preprint* arXiv:1707.03491 (2017).
[39] As its enhancements are often to do with the saturation or desaturation of colour, example images are not suitable for a black-and-white publication; instead, two example enhancements by the author are available at doi.org/10.6084/m9.figshare.19336634

the World Wide Web, sociologist Anthony Giddens described the *double hermeneutic*[40], in which sociological models affect the behaviour of the people they describe. Prettiness estimation algorithms in image enhancement, auto-cropping, and search engines form a double hermeneutic: between neural models of human behaviour and the behaviour itself.

This is a crucial difference between the cultural biases in relatively 'objective' tasks (object detection, person recognition, etc), and those in highly personal tasks like prettiness estimation. At the start of this paper, we saw a study on how Facebook's object detection algorithm performs significantly worse on images of everyday objects taken in low-income countries;[41] presumably because it had been trained on images from higher-income countries. As potentially problematic as this is, there is no suggestion this will lead to changes in the phenomenon being modelled.

We have a feedback loop, then, between public taste in photographs and the neural models intended to model it. As new datasets for prettiness estimation are created, newer tendencies in taste are incorporated in the neural models; whilst some 'online' machine learning systems might be updating their behaviour in real time based on changing user behaviour (i.e. changing taste in images). The loop is biased in at least two ways: firstly, because the initial models created to model the phenomenon are based on very particular subsections of the global population (whether or not they actually use DPChallenge); and secondly, because future datasets will necessarily be similarly skewed towards photographers (smartphone or digital camera owners) and internet users (i.e. those generating quantitative training data on prettiness, possibly unwittingly through the use of social media platforms, search engines etc).

What do we know about the dynamics of this feedback loop? We know that it necessarily has the tendency to homogenise visual taste - and that its centre of gravity is the taste-system we have explored above. Images get made to look like *other pretty pictures* - this is the logic of pattern recognition. We don't know how strong the feedback loop is, and it may be that global, distributed visual taste is barely influenced by the enhancements made by smartphones. But the contemporary importance of smartphones as tools of image-creation (compared, say, to pocket digital cameras); the ubiquity of image enhancement software in newer models; and the association of particular algorithms (Apple's *Deep Fusion*) with expensive hardware (iPhones) might make us suspect otherwise. Because enhancements are often performed silently and automatically at the time of capture, image enhancement algorithms rope us in as collaborators: their pretty pictures are our pretty pictures.

As anyone who has placed a microphone near a loudspeaker knows, feedback loops are not always stable. They can implode just as easily as they explode. There are no direct historical precedents to global cultural feedback systems of this kind – between a distributed system of visual taste, and a centralised algorithm trying to model and predict it. However, many of the the basic methods of image enhancement used today (the manipulation of colour palettes, false depth of field, artificial sharpening) are not so far removed from the colour-filters which made Instagram successful in the early 2010s. Their commercial propositions are in a sense similar: to make images from smartphones look like they came from "real" cameras (in Instagram's case, film cameras; in contemporary photo augmentation, DSLRs). Instagram's homogenisation of

---

[40] Giddens, Anthony. *Social theory and modern sociology*. Stanford University Press, 1987. The 'philosopher of New Labour', Giddens shares little politically with Berger; but the sociological phenomenon he identifies is of relevance nonetheless

[41] De Vries, Terrance, Ishan Misra, Changhan Wang, and Laurens Van der Maaten. "Does object recognition work for everyone?." In *Proceedings of the IEEE/CVF Conferenceon Computer Vision and Pattern Recognition Workshops*, pp. 52-59. 2019.

digital visual culture even led to changes in commercial architecture[42], but it also brought the '#nofilter' trend of 2014 (a contradictory, and only very partial, rejection of Instagram's algorithmic aesthetic). Instagram's filter-based algorithms were much simpler, and it had no 'online learning' feedback loop between data and model; but it is, perhaps, an indication that homogenising algorithms in online visual culture have the potential (even the tendency) to collapse in on themselves.

[42] Newton, Casey. "Instagram is Pushing Restaurants to be Kitschy, Colorful, and Irresistable to Photographers". *The Verge*, July 2020
https://www.theverge.com/2017/7/20/16000552/instagram-restaurant-interior-design-photo-friendly-media-noche