Notes On A (Dis)continuous Surface

Murad Khan

'Differentiated through that which is porous – the skin – a surface perceptive to touch, the body is dissected, fixed "and woven out of a thousand details, anecdotes and stories".' [1]

From content recommendation and social-media feed curation to financial risk assessment and medical diagnoses, machine-learning models have become a pervasive part of our everyday infrastructure. While automated data-processing instruments have long been part of our lives, machine learning provides an accelerated paradigm within which patterns can be unearthed and made actionable across large pools of historical data. Given that these technologies are being deployed in a variety of public and private systems, ethical questions are increasingly being raised when they seem to fail, with particular concern directed at the role that these technologies play in further entrenching racial biases and practices of discrimination. Whether it be failing to recognise darker-skinned subjects,[2] amplifying negative racial stereotypes[3] or denying access to credit, forms of pattern-based learning appear to consistently exacerbate existing racial inequalities and modes of discrimination. With these models increasingly supporting human decision-making in key areas, it is crucial that we understand how racial representation functions within machine-learning systems, asking both how race is understood, and what can be achieved by encoding this understanding.

Differential Visibilities

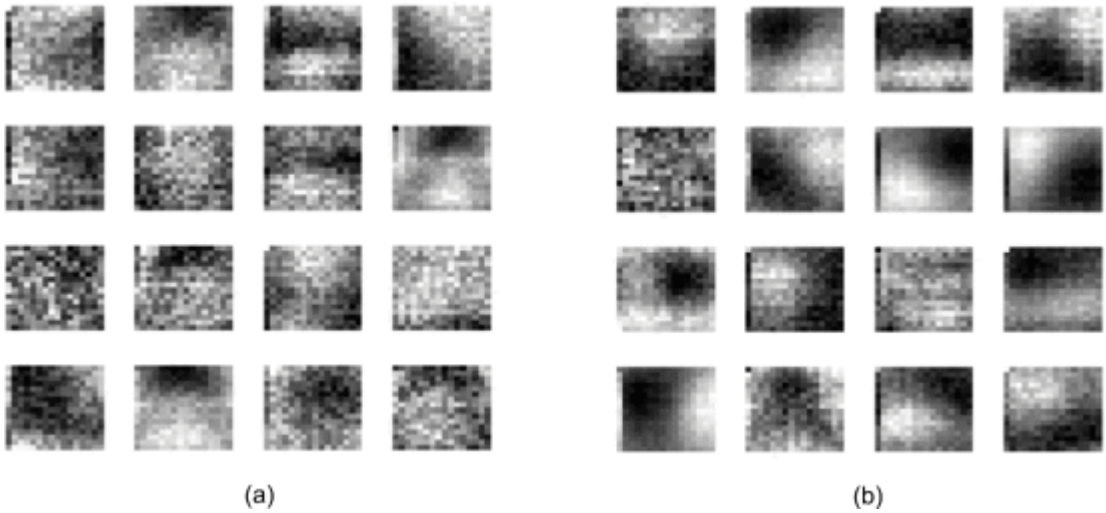


Figure 1. Discriminative race feature representation by multiple layer Convolution Neural Networks (CNN). (a): supervised CNN filters (b): CNN with transfer learning filters [4]

'I am given no chance. I am overdetermined from without. I am the slave not of the 'idea' that others have of me but of my own appearance … I am *fixed*.'[5]

Frantz Fanon's description identifies his own skin as a site of fixity. In an instance of 'epidermalisation', the porous surface enveloping his body enfolds him within the tonal weave of a racial-corporal schema, apprehending him as Black before human and defining the possibilities afforded to him in accordance with the colour of his skin. This schema, which is 'cultural and discursive' rather than solely genetic,[6] is produced and reproduced across morphological designations, stitching a racialised subject out of 'a thousand details, anecdotes, stories',[7] constituting them historically within the limited and specular frame of race-centric discourse. Crucially, such a schema seeks to align the exterior expressions of the body with internal traits corresponding to behaviour, character and cognitive capacity that can be generalised over members of the given racial group. Doing so composes race beyond the remits of the individual body, forming it in concert with the fictive hierarchies that guarantee the colonial arrangement, naturalising racial difference as a twinned condition of the body and mind. To this extent, race is more than just a schema of visual understanding. It forms a perceptive tissue that brings together forms of social

organisation through a psychic operation that safeguards the conditions of the human for certain groups over others, forming the fragmented racial body into a knowable object whenever it is invoked: a legible surface upon which all manner of racial truths may be etched and read in service of maintaining extant social relations. To this degree, it is imperative to outline the ways in which race is figured by a similar series of epidermal abstractions within machine-learning systems, mobilised as a site for perception and identification as well as probabilistic prediction.

Abstraction, Recognition and Prediction

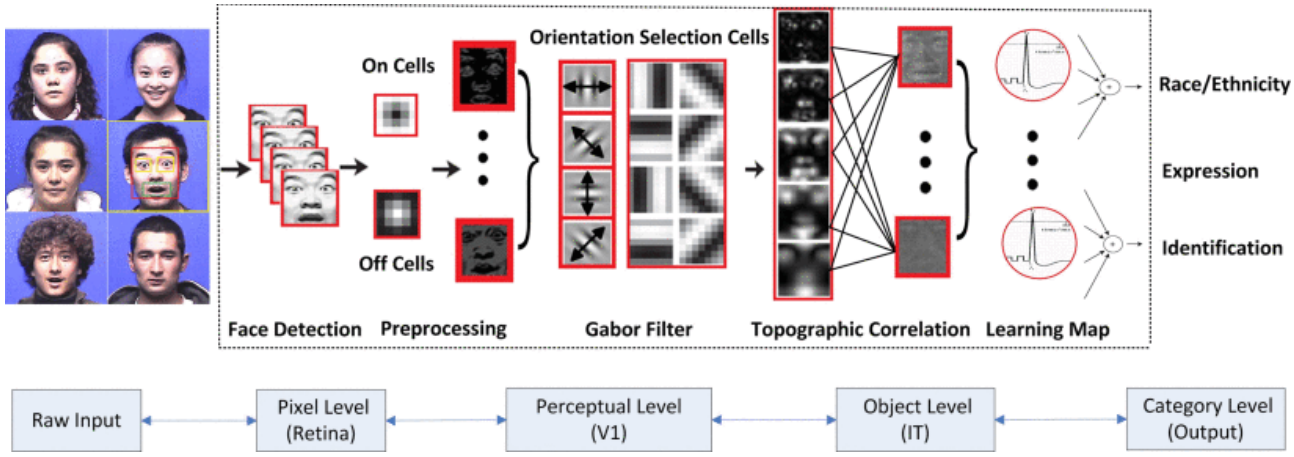


Figure 2. Fu, Siyao, Haibo He, and Zeng-Guang Hou. "Race classification from face: A survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, no. 12 (2014): 2483-2509.

Whilst forms of biometric identification technology have been in use since the 1990s, it is only in the past five years that computational and graphics processing power has improved to such a degree that machine learning can regularly be used to solve problems of face detection and recognition. State-of-t-e-art software now utilises Deep Convolutional Neural Networks (DCNNs), training the learning model on large datasets of faces for authentication, detection and identification scenarios. This is typically done by mapping pixel regions in an input image, wherein facial landmarks (nodal points) such as the distance between the eyes, the tip of the nose, or the corners of the mouth are mapped, extracted and used to detect each unique face. The number of nodal points mapped for each model varies depending on the algorithm used, with some generating an embedding of up to 128 measurements in order to properly map the image to a set of numerical representations. Once these landmarks have been identified and the model trained enough times on these representations, it will have scaled in complexity, moving from an array of indiscernible lines and edges, through to blobs, facial features and eventually to a coherent understanding of a 'face', or a set of values equating to different pixel regions across the image.

Racial representation comes into the equation in supervised learning scenarios, in which the model is provided with labelled images to better classify different types of faces based on these learned patterns of pixels. These labels are key to understanding the racialised nature of facial recognition, as the model learns features corresponding to a given taxonomy of racial classifications, sorting patterns it discovers into these pre-defined spaces of representation, and gauging their proximity (similarity) to one another in order to make a judgement on which racial class an individual face falls into. However, since DCNNs are dependent upon the datasets used to train them, we regularly see instances of failure if the set of faces for a certain racial class is lacking in its training data. Often, this is played out across darker-skinned subjects, causing failure rates to increase once the model encounters them in real-world applications. Subjects either fail to be recognised, or are mis-recognised within the given categories of racial representation. Such failures are exceedingly common, ranging from exam proctoring software barring students from

taking tests[8] to passport applications being rejected.[9]

That these technologies consistently fail when faced with racialised populations is a well-documented issue, making it all the more pernicious that these technologies are continually implemented in public-facing infrastructure.[10] However, those proposing greater diversity in data-representation as a solution to these issues tend to miss the nuances of the problem, failing to recognise that, implemented 'accurately' or otherwise, these racial classifications are going to be put to work in improving predictive policing, surveillance infrastructure and drone targeting systems that render differential levels of harm to racialised populations. For instance, IBM's attempt to create its 'Diversity in Faces' dataset to alleviate racial bias is a prime example of the damage that can be done when large companies latch onto the idea of being more 'diverse' only to reproduce historical understandings about the 'reality' of racial representation. In their search for a diverse and 'racially accurate' dataset that extended beyond the brute classifications of skin colour, not only did researchers from IBM make worrying recourse to craniofacial measurements as an objective indicator of racial grouping,[11] but they did so whilst simultaneously selling custom implementations of their facial recognition software to law-enforcement agencies.[12] Such pseudo-scientific practices have also spilled over into the realm of prediction and 'affective computing', where emotional analysis is carried out on facial expressions.[13] As expected of a system using race-centric data, analysis of the facial expressions of Black men consistently scored them as angrier than White men, replicating social biases.[14] Frank Pasquale summarises the inevitability of bias within such a system, emphasising that 'If a database of aggression is developed from observation of a particular subset of the population, the resulting AI may be far better at finding "suspect behavior" in that subset rather than others.'[15] Thus, by mimicking the long history of pseudo-sciences such as physiognomy and phrenology that tied racialised facial representation to forms of criminality and deviance, such software merely rehashes historic schemas of racial perception under the guise of insightful and objective computational analysis, making them actionable once more.

While expression analysis demonstrates one clearly racialised form of machine prediction, there are other instances in which the learning system may not be presented with race as a defined variable in its input data, but still picks up on cues that implicate race as a latent force within an assemblage of other variables. This associative tendency exacerbates what is referred to as the problem of 'algorithmic bias', denoting the way in which socio-technical apparatuses that leverage statistical (probability-based) models to guide decision-making frequently make predictions based upon implicitly racialised data, amplifying patterns of social bias. Safiya Noble argues that these practices enact similar forms of exclusion and discrimination to 'redlining' practices in the United States. The computation of probabilities, whether for medical diagnoses, credit allocation or even search engine results, depends upon pattern-based abstractions extending a series of equivalencies and probabilities from the physiological designations of the racialised body, proxied for by a wide range of class conditions that reflect and foster structural inequalities, such as access to housing, education history, employment opportunities, life expectancy and so on. Ramon Amaro provides a useful articulation of these discriminatory logics, positing that in the realm of human difference, machine learning has become 'a projection of an already racialised imaginary enacted through technological solution – an imaginary that already understands the black, brown, criminalised, gendered and otherwise Othered human as the principle site of exclusion, quantification, and social organisation.'[16] As such, machine learning can be seen to replay the Fanonian problematics of corporeal representation and psychic differentiation within the sphere of predictive computation, contaminated by the legacies and motivations of the colonial arrangement.

Given these manifestations of race within machine learning, both at the level of visual recognition and within historical data distributions, we can see that the problem of race is best encapsulated not by the question of non-recognition, but of recognition within a discursive environment that has asserted race as a coherent metric for the classification of people as well as a meaningful predictor of future behaviour. Much

as Fanon suggests, racialised subjects are 'overdetermined from without', subject to the legacies and injustices consonant with racial identification and their rearticulation within contemporary technical infrastructure. Contemporary applications of machine-learning[17] In doing so, patterns of probability reach across bodies to form the recurrent possibility of an object both legible and computable, contiguous with the racialised exterior and interior features of an individual. Coerced into an extensive causal surface, the dynamisms of living, breathing individuals are pulled together by the epidermal logic described by Fanon.

---

[1] Simone Browne, 'Digital Epidermalization: Race, Identity and Biometrics', *Critical Sociology,* 36 (1) (February 1, 2010), p. 133.

[2] Joy Buolamwini, Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. *Proceedings of Machine Learning Research,* 81 (2018), pp. 1–15.

[3] See Safiya Noble, *Algorithms of Oppression* (New York: New York University Press, 2018).

[4] Siyao Fu, Haibo He, Zeng-Guang Hou, 'Learning Race from Face: A Survey', *IEEE Transactions on Patter Analysis and Machine Intelligence,* 36 (12) (December 2014). (Fig. 8)

[5] Frantz Fanon, *Black Skin, White Masks* (London: Pluto Press, 1986), p. 87.

[6] Stuart Hall, *The Fact of Blackness: Frantz Fanon and Visual Representation*, ed. Alan Read (London: ICA, 1996), p. 16.

[7] Fanon, *Black Skin, White Masks*, p. 84.

[8] https://venturebeat.com/2020/09/29/examsofts-remot...

[9] https://www.newscientist.com/article/2219284-uk-la...

[10] For instance, in the case of facial recognition for the Home Office's automated passport-photo processing service, a Freedom of Information request revealed that tests had been carried out showing a poor result on darker skinned faces, yet the service was deemed 'sufficient enough to deploy'. See https://www.whatdotheyknow.com/request/skin_colour...

[11] *Diversity in Faces*, https://arxiv.org/abs/1901.10436,

[12] It is worth noting that, in the wake of global Black Lives Matter protests sparked by the deaths of George Floyd, Breonna Taylor and countless others at the hands of the police, IBM chose to announce a moratorium on the sale of facial recognition technology, and to open a dialogue on 'whether and how facial recognition technology should be employed by domestic law enforcement agencies'. Whilst this garnered much applause from 'AI Ethics' advocates, the more cynical among us may note that their announcement only stated that they would no longer offer 'general purpose IBM facial recognition or analysis software' for sale. Whether the software would remain available for custom implementations, such as in police body camera offerings, as they advertise elsewhere on their website, is unclear.

[13] Amazon's Rekognition software, for instance, provides a confidence score for facial emotion. See https://docs.aws.amazon.com/rekognition/latest/dg/...

[14] Lauren Rhue, 'Racial Influence on Automated Perceptions of Emotions' (November 9, 2018). Available at SSRN: https://ssrn.com/abstract=3281765 or http://dx.doi.org/10.2139/ssrn.3281765

[15] https://reallifemag.com/more-than-a-feeling/

[16] Ramon Amaro, *AI and the Empirical Reality of a Racialised Future in AI: More Than Human* (London: Barbican, 2019), p. 126.

[17] Hall, *The Fact of Blackness*, p. 20.

Murad Khan

Murad Khan is a Ph.D researcher at University College London and a Visiting Practitioner at UAL: Central Saint Martins. Directed through an emerging philosophy of noise and adversarial machine learning, his

research seeks to read Frantz Fanon as a philosopher of information, bridging his work on the psychic operations of racialisation with Gilbert Simondon's philosophy of individuation in an exploration of psychopathology and the limits of reason.