Criminal Machine Learning

Responses to Critiques on Machine Learning of Criminality Perceptions (Addendum of arXiv:1611.04135)

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

In November 2016 we submitted to arXiv our paper "Automated Inference on Criminality Using Face Images". It generated a great deal of discussions in the Internet and some media outlets. Our work is only intended for pure academic discussions; how it has become a media consumption is a total surprise to us.

Although in agreement with our critics on the need and importance of policing AI research for the general good of the society, we are deeply baffled by the ways some of them mispresented our work, in particular the motive and objective of our research.

1. Name calling

It should be abundantly clear, for anyone who reads our paper with a neutral mind setting, that our only motive is to know if machine learning has the potential of acquiring humanlike social perceptions of faces, despite the complexity and subtlety of such perceptions that are functions of both the observed and the observer. Our inquiry is to push the envelope and extend the research on automated face recognition from the biometric dimension (e.g., determining the race, gender, age, facial expression, etc.) to the sociopsychological dimension. We are merely interested in the distinct possibility of teaching machines to pass the Turing test on the task of duplicating humans in their first impressions (e.g., personality traits, mannerism, demeanor, etc.) of a stranger. The face perception of criminality was expediently (unfortunately to us in hindsight) chosen as an easy test case, at least in our intuition as explained in our paper:

> *"For validating the hypothesis on the correlations between the innate traits and social behaviors of a person and the physical characteristics of that persons face, it would be hard pushed to find a more convincing experiment than examining the success rates of discriminating between criminals and non-criminals with modern automatic classifiers. These two populations should be among the easiest to differentiate, if social attributes and facial features are correlated, because being a criminal requires a host of abnormal (outlier) personal traits. If the classification rate turns out low, then the validity of face-induced social inference can be safely negated."*

By a magical stretch of imagination, few of our critics intertwine the above passage into some of our honest observations and morph them into the following deduction of, they insist, ours:

> *"Those with more curved upper lips and eyes closer together are of a lower social order, prone to (as Wu and Zhang put it) "a host of abnormal (outlier) personal traits" ultimately leading to a legal diagnosis of "criminality" with high probability."*

We agree that the pungent word criminality should be put in quotation marks; a caveat about the possible biases in the input data should be issued. Taking a court conviction at its face value, i.e., as the "ground truth" for machine learning, was indeed a serious oversight on our part. However, throughout our paper we maintain a sober neutrality on whatever we might find; in the introduction, we declare

> *"In this paper we intend not to nor are we qualified to discuss or debate on societal stereotypes, rather we want to satisfy our curiosity in the accuracy of fully automated inference on criminality. At the onset of this study our gut feeling is that modern tools of machine learning and computer vision will refute the validity of physiognomy, although the outcomes turn out otherwise."*

Nowhere in our paper advocated the use of our method as a tool of law enforcement, nor did our discussions advance from correlation to causality. But still we got interpreted copiously by some with an insinuation of racism. This is not the way of academic exchanges we are used to.

https://arxiv.org/pdf/1611.04135.pdf 2017

Criminal Machine Learning

Responses to Critiques on Machine Learning of Criminality Perceptions
(Addendum of arXiv:1611.04135)

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

In November 2016 we submitted to arXiv our paper "Automated Inference on Criminality Using Face Images". It generated a great deal of discussions in the Internet and some media outlets. Our work is only intended for pure academic discussions; how it has become a media consumption is a total surprise to us.

Although in agreement with our critics on the need and importance of policing AI research for the general good of the society, we are deeply baffled by the ways some of them mispresented our work, in particular the motive and objective of our research.

1. Name calling

It should be abundantly clear, for anyone who reads our paper with a neutral mind setting, that our only motive is to know if machine learning has the potential of acquiring humanlike social perceptions of faces, despite the complexity and subtlety of such perceptions that are functions of both the observed and the observer. Our inquiry is to push the envelope and extend the research on automated face recognition from the biometric dimension (e.g., determining the race, gender, age, facial expression, etc.) to the sociopsychological dimension. We are merely interested in the distinct possibility of teaching machines to pass the Turing test on the task of duplicating humans in their first impressions (e.g., personality traits, mannerism, demeanor, etc.) of a stranger. The face perception of criminality was expediently (unfortunately to us in hindsight) chosen as an easy test case, at least in our intuition as explained in our paper:

> *"For validating the hypothesis on the correlations between the innate traits and social behaviors of a person and the physical characteristics of that persons face, it would be hard pushed to find a more convincing experiment than examining the success rates of discriminating between criminals and non-criminals with modern automatic classifiers. These two populations should be among the easiest to differentiate, if social attributes and facial features are correlated, because being a criminal requires a host of abnormal (outlier) personal traits. If the classification rate turns out low, then the validity of face-induced social inference can be safely negated."*

By a magical stretch of imagination, few of our critics intertwine the above passage into some of our honest observations and morph them into the following deduction of, they insist, ours:

> *"Those with more curved upper lips and eyes closer together are of a lower social order, prone to (as Wu and Zhang put it) "a host of abnormal (outlier) personal traits" ultimately leading to a legal diagnosis of "criminality" with high probability."*

We agree that the pungent word criminality should be put in quotation marks; a caveat about the possible biases in the input data should be issued. Taking a court conviction at its face value, i.e., as the "ground truth" for machine learning, was indeed a serious oversight on our part. However, throughout our paper we maintain a sober neutrality on whatever we might find; in the introduction, we declare

> *"In this paper we intend not to nor are we qualified to discuss or debate on societal stereotypes, rather we want to satisfy our curiosity in the accuracy of fully automated inference on criminality. At the onset of this study our gut feeling is that modern tools of machine learning and computer vision will refute the validity of physiognomy, although the outcomes turn out otherwise."*

Nowhere in our paper advocated the use of our method as a tool of law enforcement, nor did our discussions advance from correlation to causality. But still we got interpreted copiously by some with an insinuation of racism. This is not the way of academic exchanges we are used to.

https://arxiv.org/pdf/1611.04135.pdf 2017