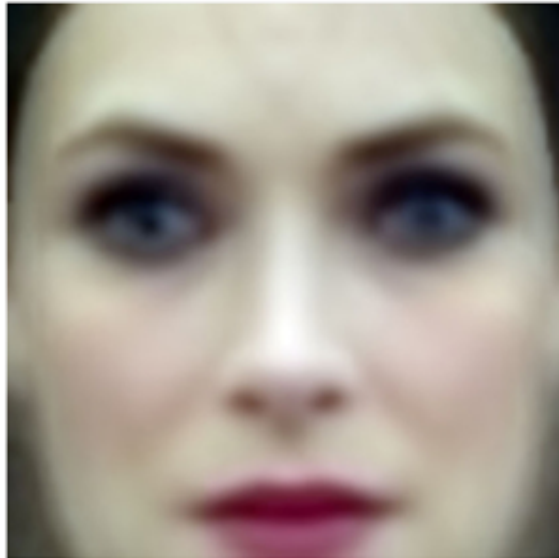


ESSAYS & REVIEWS

# *Invisible Images (Your Pictures Are Looking at You)*

By [TREVOR PAGLEN](#) DECEMBER 8, 2016



# I.

“Winona”  
Eigenface (Colorized),  
Labelled Faces in the Wild  
Dataset  
2016

OUR eyes are fleshy things, and for most of human history our visual culture has also been made of fleshy things. The history of images is a history of pigments and dyes, oils, acrylics, silver nitrate and gelatin—materials that one could use to paint a cave, a church, or a canvas. One could use them to make a photograph, or to print pictures on the pages of a magazine. The advent of screen-based media in the latter half of the 20th century wasn’t so different: cathode ray tubes and liquid crystal displays emitted light at frequencies our eyes perceive as color, and densities we perceive as shape.

We’ve gotten pretty good at understanding the vagaries of human vision; the serpentine ways in which images infiltrate and influence culture, their tenuous relationships to everyday life and truth, the means by which they’re harnessed to serve—and resist—power. The theoretical concepts we use to analyze classical visual culture are robust: representation, meaning, spectacle, semiosis, mimesis, and all the rest. For centuries these concepts have helped us to navigate the workings of classical visual culture.

But over the last decade or so, something dramatic has happened. Visual culture has changed form. It has become detached from human eyes and has largely become invisible. Human visual culture has become a special case of vision, an exception to the rule. The overwhelming majority of images are now made by machines for other machines, with humans rarely in the loop. The advent of machine-to-machine seeing has been barely noticed at large, and poorly understood by those of us who’ve begun to notice the tectonic

shift invisibly taking place before our very eyes.

The landscape of invisible images and machine vision is becoming evermore active. Its continued expansion is starting to have profound effects on human life, eclipsing even the rise of mass culture in the mid 20th century. Images have begun to intervene in everyday life, their functions changing from representation and mediation, to activations, operations, and enforcement. Invisible images are actively watching us, poking and prodding, guiding our movements, inflicting pain and inducing pleasure. But all of this is hard to see.

Cultural theorists have long suspected there was something different about digital images than the visual media of yesteryear, but have had trouble putting their finger on it. In the 1990s, for example, there was much to do about the fact that digital images lack an “original.” More recently, the proliferation of images on social media and its implications for inter-subjectivity has been a topic of much discussion among cultural theorists and critics. But these concerns still fail to articulate exactly what’s at stake.

One problem is that these concerns still assume that humans are looking at images, and that the relationship between human viewers and images is the most important moment to analyze—but it’s exactly this assumption of a human subject that I want to question.

What’s truly revolutionary about the advent of digital images is the fact that they are fundamentally machine-readable: they can only be seen by humans in special circumstances and for short periods of time. A photograph shot on a phone creates a machine-readable file that does not reflect light in such a way as to be perceptible to a human eye. A secondary application, like a software-based photo viewer paired with a liquid crystal display and backlight may create something that a human can look at, but the image only appears to human eyes temporarily before reverting back to its immaterial machine form when the phone is put away or the display is turned off. However, the image doesn’t need to be turned into human-readable form in order for a machine to do something with it. This is fundamentally different than a roll of

undeveloped film. Although film, too, must be coaxed by a chemical process into a form visible by human eyes, the undeveloped film negative isn't readable by a human or machine.

The fact that digital images are fundamentally machine-readable regardless of a human subject has enormous implications. It allows for the automation of vision on an enormous scale and, along with it, the exercise of power on dramatically larger and smaller scales than have ever been possible.



Lake Tenaya  
Maximally Stable External  
Regions; Hough Transform  
2016

## II.

Our built environments are filled with examples of machine-to-machine seeing apparatuses: Automatic License Plate Readers (ALPR) mounted on police cars, buildings, bridges, highways, and fleets of private vehicles snap photos of every car entering their frames. ALPR operators like the company Vigilant Solutions collect the locations of every car their cameras see, use Optical Character Recognition (OCR) to store license plate numbers, and create databases used by police, insurance companies, and the like.[footnote: James Bridle's "How Britain Exported Next-Generation Surveillance" is an

excellent introduction to APLR.] In the consumer sphere, outfits like Euclid Analytics and Real Eyes, among many others, install cameras in malls and department stores to track the motion of people through these spaces with software designed to identify who is looking at what for how long, and to track facial expressions to discern the mood and emotional state of the humans they're observing. Advertisements, too, have begun to watch and record people. And in the industrial sector, companies like Microscan provide full-fledged imaging systems designed to flag defects in workmanship or materials, and to oversee packaging, shipping, logistics, and transportation for automotive, pharmaceutical, electronics, and [packaging industries](#). All of these systems are only possible because digital images are machine-readable and do not require a human in the analytic loop.

This invisible visual culture isn't just confined to industrial operations, law enforcement, and "smart" cities, but extends far into what we'd otherwise—and somewhat naively—think of as human-to-human visual culture. I'm referring here to the trillions of images that humans share on digital platforms—ones that at first glance seem to be made by humans for other humans.

On its surface, a platform like Facebook seems analogous to the musty glue-bound photo albums of postwar America. We "share" pictures on the Internet and see how many people "like" them and redistribute them. In the old days, people carried around pictures of their children in wallets and purses, showed them to friends and acquaintances, and set up slideshows of family vacations. What could be more human than a desire to show off one's children? Interfaces designed for digital image-sharing largely parrot these forms, creating "albums" for selfies, baby pictures, cats, and travel photos.

But the analogy is deeply misleading, because something completely different happens when you share a picture on Facebook than when you bore your neighbors with projected slide shows. When you put an image on Facebook or other social media, you're feeding an array of immensely powerful artificial intelligence systems information about how to identify people and how to recognize places and objects, habits and preferences, race, class, and gender

identifications, economic statuses, and much more.

Regardless of whether a human subject actually sees any of the 2 billion photographs uploaded daily to [Facebook-controlled platforms](#), the photographs on social media are scrutinized by neural networks with a degree of attention that would make even the most steadfast art historian blush. Facebook's "DeepFace" algorithm, developed in 2014 and deployed in 2015, produces three-dimensional abstractions of individuals' faces and uses a neural network that achieves over 97 percent accuracy at identifying individuals— a percentage comparable to what a human can achieve, ignoring for a second that no human can recall the [faces of billions of people](#).

There are many others: Facebook's "DeepMask" and Google's TensorFlow identify people, places, objects, locations, emotions, gestures, faces, genders, economic statuses, relationships, and much more.

In aggregate, AI systems have appropriated human visual culture and transformed it into a massive, flexible training set. The more images Facebook and Google's AI systems ingest, the more accurate they become, and the more influence they have on everyday life. The trillions of images we've been trained to treat as human-to-human culture are the foundation for increasingly autonomous ways of seeing that bear little resemblance to the visual culture of the past.

"Goldfish"

Linear Classifier, ImageNet

Dataset

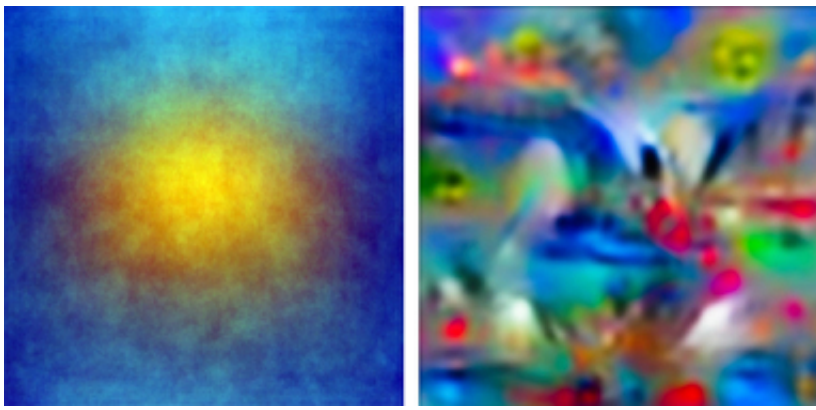
2016

? "Fire Boat"

Synthetic High Activation,

ImageNet Dataset

2016



### III.

If we take a peek into the internal workings of machine-vision systems, we find a menagerie of abstractions that seem completely alien to human perception. The machine-machine landscape is not one of representations so much as activations and operations. It's constituted by active, performative relations much more than classically representational ones. But that isn't to say that there isn't a formal underpinning to how computer vision systems work.

All computer vision systems produce mathematical abstractions from the images they're analyzing, and the qualities of those abstractions are guided by the kind of metadata the algorithm is trying to read. Facial recognition, for instance, typically involves any number of techniques, depending on the application, the desired efficiency, and the available training sets. The Eigenface technique, to take an older example, analyzes someone's face and subtracts from that the features it has in common with other faces, leaving a unique facial "fingerprint" or facial "archetype." To recognize a particular person, the algorithm looks for the fingerprint of a given person's face.

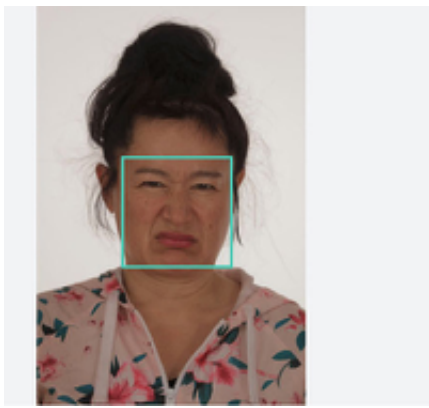
Convolutional Neural Networks (CNN), popularly called "deep learning" networks, are built out of dozens or even hundreds of internal software layers that can pass information back and forth. The earliest layers of the software pick apart a given image into component shapes, gradients, luminosities, and

corners. Those individual components are convolved into synthetic shapes. Deeper in the CNN, the synthetic images are compared to other images the network has been trained to recognize, activating software “neurons” when the network finds similarities.

We might think of these synthetic activations and other “hallucinated” structures inside convolutional neural networks as being analogous to the archetypes of some sort of Jungian collective unconscious of artificial intelligence—a tempting, although misleading, metaphor. Neural networks cannot invent their own classes; they’re only able to relate images they ingest to images that they’ve been trained on. And their training sets reveal the historical, geographical, racial, and socio-economic positions of their trainers. Feed an image of Manet’s “Olympia” painting to a CNN trained on the industry-standard “Imagenet” training set, and the CNN is quite sure that it’s looking at a “burrito.” It goes without saying that the “burrito” object class is fairly specific to a youngish person in the San Francisco Bay Area, where the modern “mission style” burrito was invented. Spend a little bit of time with neural networks, and you realize that anyone holding something in their hand is likely to be identified as someone “holding a cellphone,” or “holding a Wii controller.” On a more serious note, engineers at Google decided to deactivate the “gorilla” class after it became clear that its algorithms trained on predominantly white faces and tended to classify African Americans as apes.

The point here is that if we want to understand the invisible world of machine-machine visual culture, we need to unlearn how to see like humans. We need to learn how to see a parallel universe composed of activations, keypoints, eigenfaces, feature transforms, classifiers, training sets, and the like. But it’s not just as simple as learning a different vocabulary. Formal concepts contain epistemological assumptions, which in turn have ethical consequences. The theoretical concepts we use to analyze visual culture are profoundly misleading when applied to the machinic landscape, producing distortions, vast blind spots, and wild misinterpretations.





```
Detection Result:
JSON:
[
  {
    "faceRectangle": {
      "left": 252,
      "top": 457,
      "width": 332,
      "height": 332
    },
    "scores": {
      "anger": 0.07382989,
      "contempt": 0.04350459,
      "disgust": 0.7045257,
      "fear": 0.0005751435,
      "happiness": 0.00111284887,
      "neutral": 0.04504608,
      "sadness": 0.127928317,
      "surprise": 0.00347741228
    }
  }
]
```

(Research Image)

“Disgust”

Custom Hito Steyerl

Emotion Training Set

## VI.

There is a temptation to criticize algorithmic image operations on the basis that they’re often “wrong”—that “Olympia” becomes a burrito, and that African Americans are labelled as non-humans. These critiques are easy, but misguided. They implicitly suggest that the problem is simply one of accuracy, to be solved by better training data. Eradicate bias from the training data, the logic goes, and algorithmic operations will be decidedly less racist than human-human interactions. Program the algorithms to see everyone equally and the humans they so lovingly oversee shall be equal. I am not convinced.

Ideology’s ultimate trick has always been to present itself as objective truth, to present historical conditions as eternal, and to present political formations as natural. Because image operations function on an invisible plane and are not dependent on a human seeing-subject (and are therefore not as obviously ideological as giant paintings of Napoleon) they are harder to recognize for what they are: immensely powerful levers of social regulation that serve specific race and class interests while presenting themselves as objective.

The invisible world of images isn’t simply an alternative taxonomy of visibility. It is an active, cunning, exercise of power, one ideally suited to molecular police and market operations—one designed to insert its tendrils

into ever-smaller slices of everyday life.

Take the case of Vigilant Solutions. In January 2016, Vigilant Solutions, the company that boasts of having a database of billions of vehicle locations captured by ALPR systems, signed contracts with a handful of local Texas governments. According to documents obtained by the Electronic Frontier Foundation, the deal went like this: Vigilant Solutions provided police with a suite of ALPR systems for their police cars and access to Vigilant's larger database. In return, the local government provided Vigilant with records of outstanding arrest warrants and overdue court fees. A list of "flagged" license plates associated with outstanding fines are fed into mobile ALPR systems. When a mobile ALPR system on a police car spots a flagged license plate, the cop pulls the driver over and gives them two options: they can pay the outstanding fine on the spot with a credit card (plus at 25 percent "service fee" that goes directly to Vigilant), or they can be arrested. In addition to their 25 percent surcharge, Vigilant keeps a record of every license plate reading that the local police take, adding information to their massive databases in order to be capitalized in other ways. The political operations here are clear. Municipalities are incentivized to balance their budgets on the backs of their most vulnerable populations, to transform their police into tax-collectors, and to effectively sell police surveillance data to private companies. Despite the "objectivity" of the overall system, it unambiguously serves powerful government and corporate interests at the expense of vulnerable populations and civic life.

As governments seek out new sources of revenue in an era of downsizing, and as capital searches out new domains of everyday life to bring into its sphere, the ability to use automated imaging and sensing to extract wealth from smaller and smaller slices of everyday life is irresistible. It's easy to imagine, for example, an AI algorithm on Facebook noticing an underage woman drinking beer in a photograph from a party. That information is sent to the woman's auto insurance provider, who subscribes to a Facebook program designed to provide this kind of data to credit agencies, health insurers, advertisers, tax officials, and the police. Her auto insurance premium is

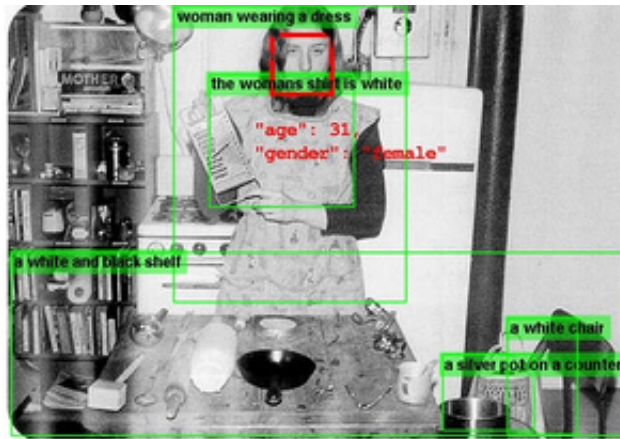
adjusted accordingly. A second algorithm combs through her past looking for similar misbehavior that the parent company might profit from. In the classical world of human-human visual culture, the photograph responsible for so much trouble would have been consigned to a shoebox to collect dust and be forgotten. In the machine-machine visual landscape the photograph never goes away. It becomes an active participant in the modulations of her life, with long-term consequences.

Smaller and smaller moments of human life are being transformed into capital, whether it's the ability to automatically scan thousands of cars for outstanding court fees, or a moment of recklessness captured from a photograph uploaded to the Internet. Your health insurance will be modulated by the baby pictures your parents uploaded of you without your consent. The level of police scrutiny you receive will be guided by your "pattern of life" signature.

The relationship between images and power in the machine-machine landscape is different than in the human visual landscape. The former comes from the enactment of two seemingly paradoxical operations. The first move is the individualization and differentiation of the people, places, and everyday lives of the landscapes under its purview—it creates a specific metadata signature of every single person based on race, class, the places they live, the products they consume, their habits, interests, "likes," friends, and so on. The second move is to reify those categories, removing any ambiguities in their interpretation so that individualized metadata profiles can be operationalized to collect municipal fees, adjust insurance rates, conduct targeted advertising, prioritize police surveillance, and so on. The overall effect is a society that amplifies diversity (or rather a diversity of metadata signatures) but does so precisely because the differentiations in metadata signatures create inroads for the capitalization and policing of everyday life.

Machine-machine systems are extraordinary intimate instruments of power that operate through an aesthetics and ideology of objectivity, but the categories they employ are designed to reify the forms of power that those

systems are set up to serve. As such, the machine-machine landscape forms a kind of hyper-ideology that is especially pernicious precisely because it makes claims to objectivity and equality.



V.

(Research Images)

Magritte, Rosler, Opie

Dense Captioning, Age,

Gender, Adult Content

Detection

Cultural producers have developed very good tactics and strategies for making interventions into human-human visual culture in order to challenge inequality, racism, and injustice.

Counter-hegemonic visual strategies and tactics employed by artists and cultural producers in the human-human sphere often capitalize on the ambiguity of human-human visual culture to produce forms of counter-culture—to make claims, to assert rights, and to expand the field of represented peoples and positions in visual culture. Martha Rosler's influential artwork "Semiotics of the Kitchen," for example, transformed the patriarchal image of the kitchen as a representation of masculinist order into a kind of prison; Emory Douglas's images of African American resistance and solidarity created a visual landscape of self-empowerment; Catherine Opie's images of queerness developed an alternate vocabulary of gender and power. All of these strategies, and many more, rely on the fact that the relationship between meaning and representation is elastic. But this idea of ambiguity, a cornerstone of semiotic theory from Saussure through Derrida, simply ceases

to exist on the plane of quantified machine-machine seeing. There's no obvious way to intervene in machine-machine systems using visual strategies developed from human-human culture.

Faced with this impasse, some artists and cultural workers are attempting to challenge machine vision systems by creating forms of seeing that are legible to humans but illegible to machines. Artist Adam Harvey, in particular, has developed makeup schemes to thwart facial recognition algorithms, clothing to suppress heat signatures, and pockets designed to prevent cellphones from continually broadcasting their location to sensors in the surrounding landscape. Julian Oliver often takes the opposite tack, developing hyper-predatory machines intended to show the extent to which we are surrounded by sensing machines, and the kinds of intimate information they're collecting all the time. These are noteworthy projects that help humans learn about the existence of ubiquitous sensing. But these tactics cannot be generalized.

In the long run, developing visual strategies to defeat machine vision algorithms is a losing strategy. Entire branches of computer vision research are dedicated to creating "adversarial" images designed to thwart automated recognition systems. These adversarial images simply get incorporated into training sets used to teach algorithms how to overcome them. What's more, in order to truly hide from machine vision systems, the tactics deployed today must be able to resist not only algorithms deployed at present, but algorithms that will be deployed in the future. To hide one's face from Facebook, one would not only have to develop a tactic to thwart the "DeepFace" algorithm of today, but also a facial recognition system from the future.

An effective resistance to the totalizing police and market powers exercised through machine vision won't be mounted through ad hoc technology. In the long run, there's no technical "fix" for the exacerbation of the political and economic inequalities that invisible visual culture is primed to encourage. To mediate against the optimizations and predations of a machinic landscape, one must create deliberate inefficiencies and spheres of life removed from market and political predations—"safe houses" in the invisible digital sphere. It

is in inefficiency, experimentation, self-expression, and often law-breaking that freedom and political self-representation can be found.

We no longer look at images—images look at us. They no longer simply represent things, but actively intervene in everyday life. We must begin to understand these changes if we are to challenge the exceptional forms of power flowing through the invisible visual culture that we find ourselves enmeshed within.