Lev Manovich

Automating Aesthetics: Artificial Intelligence and Image Culture

In the original vision of artificial intelligence (AI) in 1950s, the goal was to teach computer to perform a range of cognitive tasks. They included playing chess, solving mathematical problems, understanding written and spoken language, recognizing content of images, and so on. Today, AI (especially in the form of supervised machine learning) has become a key instrument of modern economies employed to make them more efficient and secure: making decisions on consumer loans, filtering job applications, detecting fraud, and so on.

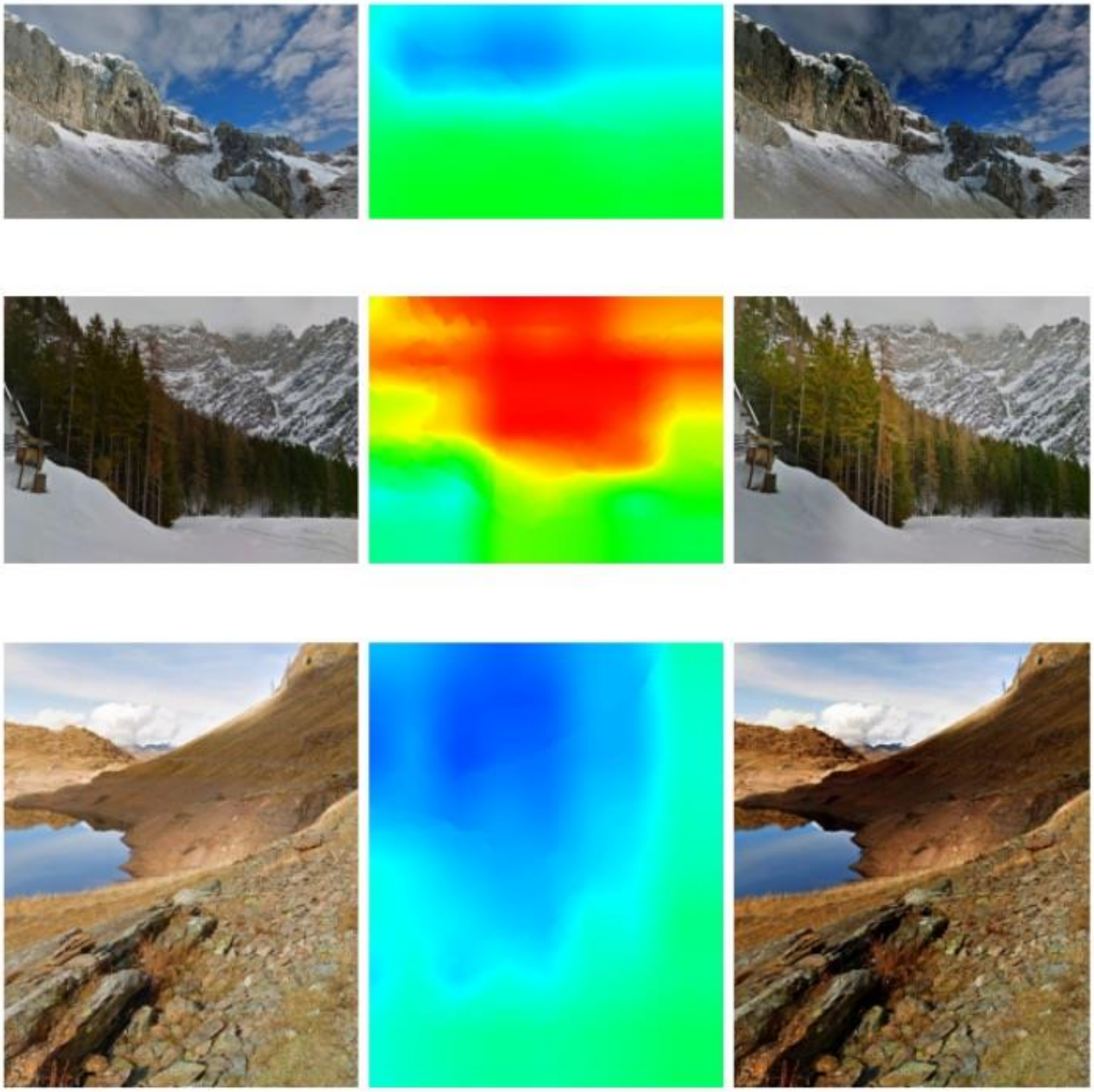
What has been less obvious is that AI now plays an equally important role in our cultural lives, increasingly automating the realm of the aesthetic. Consider, for example, image culture. *Instagram* Explore screen recommends images and videos based on what we liked in the past. *Artsy.net* recommends the artworks similar to the one you are currently viewing on the site.[1] All image apps can automatically modify captured photos according to the norms of "good photography." Other apps "beatify" selfies. Still other apps automatically edit your raw video to create short films in the range of styles.[2] The App The Roll from EyeEm automatically rates aesthetic quality of you photos (**image 1**). EyeEm system also learned the styles of different photo curators using only 20 photos, and then selected similar images from EyeEm large collection.[3] When you upload your photos to EyeEm marketplace for sale, the service automatically assigns keywords.[4] Google designed a system that mimics the skills of a professional photographer such as selecting a photo suitable for editing, cropping, applying filters, and so on (**image 2**).[5] And outside of image culture, the uses of AI range from music recommendations in Spotify, iTunes and other music services to generation of characters and environments in video games and creation of fashion styles.[6]

Does such automation leads to *decrease in cultural diversity over time*? For example, does automatic edits being applied to user photos leads to standardization of “photo imagination”? As opposed to guessing or just following out often un-grounded intuitions, can we use AI methods and large samples of cultural data to measure quantitatively diversity and variability in contemporary culture, and track how they are changing over time?

---

1 https://www.artsy.net/.
2 https://petapixel.com/2016/05/31/trained-algorithm-predict-makes-beautiful-photo/.
3 https://devblogs.nvidia.com/parallelforall/personalized-aesthetics-machine-learning/.
4 https://www.digitaltrends.com/photography/eyeem-adds-auto-keywords/.
5 https://www.theverge.com/2017/7/14/15973712/google-ai-research-street-view-panorama-photo-editing.
6 https://www.engadget.com/2017/08/24/amazon-ai-fashion-brands/.

1. Samples of photographs used to train EyeEm aesthetics ranking system. Source: Appu Shaji, “Personalized Aesthetics: Recording the Visual Mind using Machine Learning,” https://devblogs.nvidia.com/parallelforall/personalized-aesthetics-machine-learning/.



2. Neural network learned to automatically apply "dramatic mask" filter to new photos after being trained on 15,000 professional photographs landscapes. Left: original photos. Center: filter mask. Right: filtered photos. Source: Hui Fang and Meng Zhang. "Creatism: A deep-learning photographer capable of creating professional work," 2017, https://arxiv.org/abs/1707.03491.

While today AI is already automating aesthetic choices, recommending what we should watch, listen to, or wear, and also playing a big role in some areas of aesthetic *production* such as consumer photography (many features of contemporary mobile phone cameras use AI), in the future it will play a larger part in professional cultural production. Its use to design fashion, logos, music, TV commercials and work in other genres of culture industry is already growing. But for now, human experts typically make final decisions or do actual production based on ideas and media generated by computers. The well-known example of *Game of Thrones* is a case in point. The computer suggested plot ideas but the actual writing and the show development was done by humans. So we only will be able to talk about real *AI-driven culture* than computers start creating media products from beginning to end. In this future will not decide if these products should be shown to audiences. They will just trust that AI knows best. We are not here yet. For example, in 2016 IBM Watson created first "AI-made movie trailer."[7] However, the computer only selected many shots from the movie suitable to include in the trailer, and a human editor still had to do the editing.

Today automation of cultural production typically uses the contemporary form of AI called "supervised machine learning." A computer is fed many examples of similar objects such as movie trailers in a particular genre, and it gradually learns the principles of this genre. In fact, AI acts as an art theorist, an art historian or a film scholar who is also repeatedly studies many works in some area of culture to find their common patterns.

However, this is a crucial difference. A human theorist or a historian comes up with explicit principles which we can understand. For example, the standard textbook used in universities in film studies classes - *The Classical Hollywood Cinema* – provides answers to questions such as "How does the typical Hollywood film use the techniques and storytelling forms of the film medium?"[8] But typically the result of training AI on many cultural examples is a black box. Given new examples, it can classify them correctly – for example, it can decide if a particular film does have belong to "the classical Hollywood cinema," or not. But we don't know how a computer came up with this decision. Similarly, it can distinguish between works of different artists or film directors. And it can also generate new objects in the same style. But often we don't know what exactly the computer learned.

This is one of the key issues around cultural uses of AI. Are the results of machine learning *interpretable*, or is it only a black box, efficient in production but useless for human understanding of the domain? Will the expanding use of machine learning to create new cultural objects will also make for us explicit the patterns in the already existing objects? And even if it does, will it be in the form that would be understandable for people without degrees in computer science? And will the companies deploying machine learning to generate movies, ads, designs, images, music and so on expose what they have learned?

Luckily, the use of AI and other computer methods to study culture has also been pursued in the academy where all research is published and accessible, and code and datasets used in studies are also made available to everybody. Since the middle of 2000s researchers in computer science and social sciences have been increasingly analyzing quantitatively patterns in contemporary culture using progressively larger samples. Some of this work focuses on quantifying characteristics of content and aesthetics of hundreds of thousands or millions of artifacts. Others create models that predict which artifacts will be popular,

[7] https://thenextweb.com/insider/2016/09/01/watch-ibm-watson-creates-first-entirely-ai-made-movie-trailer-really-eerie/.
[8] http://www.davidbordwell.net/essays/classical.php.

and how their characteristics affect their popularity (or "memorability," "interestingness," "beauty," or "creativity."[9])

Much of this work uses large samples of content shared on social networks and the data about people behaviors on these networks. The researchers published hundreds of thousands of papers analyze computationally characteristics of images, video and text posts and behavior on all most popular social networks and media sharing services such as Weibo, Facebook, Flickr, YouTube, Pinterest, Tumblr and Instagram.

Let's look at the research about Instagram, for example. To locate only quantitative papers, I add a word "dataset" to a name of the network and then search on Google Scholar. The search for "Instagram dataset" returned 9,210 journal articles and conference papers (July 15, 2017). Here are some of these papers from 2014-2017 that will give you an idea about what questions are being asked in this research. One paper analyzed most popular Instagram subjects and also types of users in terms what subjects they post together.[10] Another paper used a sample of 4.1 million Instagram photos to quantify the effect of using filters on numbers of views and comments.[11] In yet another paper, researchers analyzed temporal and demographic trends in Instagram using 5.5 million Instagram photos with faces they collected. They have also tested three alternative hypothesis about the reasons behind posting selfies in each of 117 countries in their dataset. And a recent paper (2017) analyzed clothing and fashion styles in 44 world cities using 100 million Instagram photos.[12]

What about using computational methods to analyze patterns in historical culture using the available digitized archives? The particularly interesting examples of such research are "Toward Automated Discovery of Artistic Influence,"[13] "Measuring the Evolution of Contemporary Western Popular Music,"[14] "Shot durations, shot classes, and the increased

---

[9] For example, see Redi, M.; O'Hare, N.; Schifanella, R.; Trevisiol, M.; and Jaimes, A. 2014. "6 seconds of sound and vision: Creativity in micro- videos," https://arxiv.org/abs/1411.4080; Rossano Schifanella, Miriam Redi, Luca Aiello, 2015, "An Image is Worth More than a Thousand Favorites: Surfacing the Hidden Beauty of Flickr Pictures," 2014, https://arxiv.org/abs/1505.03358.

[10] Yuheng Hu, Lydia Manikonda, and Subbarao Kambhampati, "What We Instagram: A First Analysis of Instagram Photo Content and User Types," 2014, https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8118/8087.

[11] Saeideh Bakhshi, David A. Shamma, Lyndon Kennedy, Eric Gilbert, "Why We Filter Our Photos and How It Impacts Engagement," 2015, http://comp.social.gatech.edu/papers/icwsm15.why.bakhshi.pdf.

[12] Kevin Matzen, Kavita Bala, Noah Snavely, StreetStyle: Exploring world-wide clothing styles from millions of photos, June 6, 2017, https://arxiv.org/abs/1706.01869.

[13] Babak Saleh et al., "Toward Automated Discovery of Artistic Influence," 2015, https://arxiv.org/abs/1408.3218.

[14] Joan Serrà et al., "Measuring the Evolution of Contemporary Western Popular Music," 2012, https://www.nature.com/articles/srep00521.

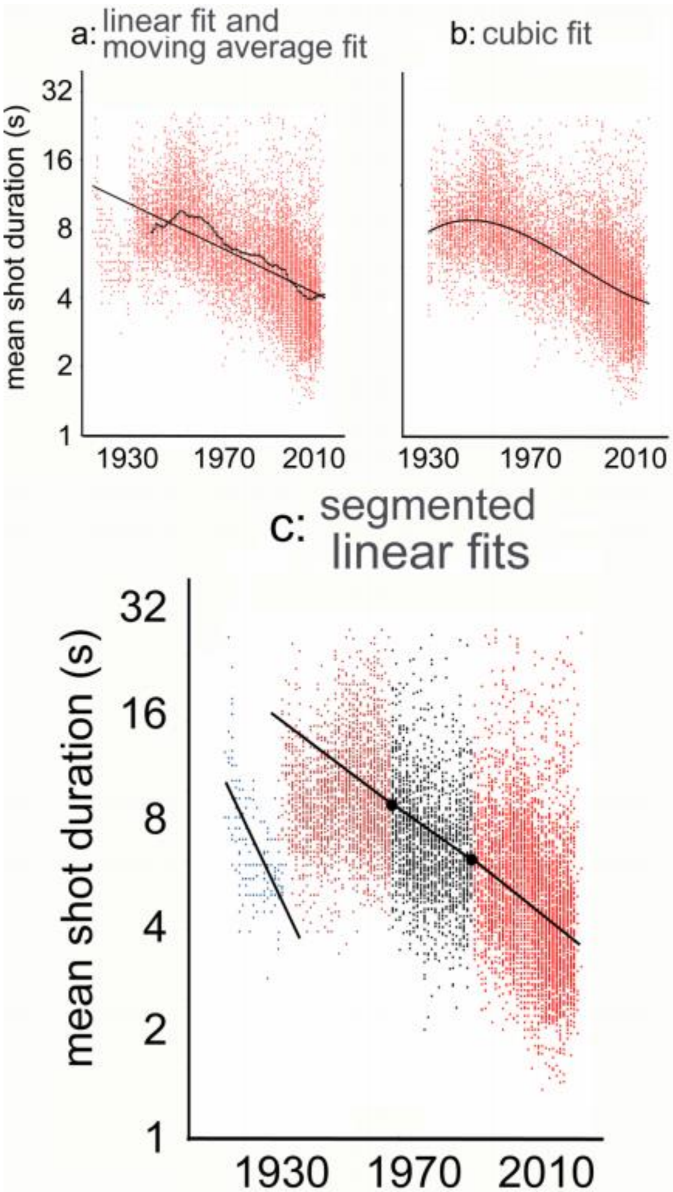
pace of popular movies,"[15] and "OmniArt: Multi-task Deep Learning for Artistic Data Analysis."[16]

The first paper presents a mathematical model for automatic discovery of influence between artists. The model was tested using 1710 images of paintings by 66 well-known artists. While some of the discovered influences are the same ones often described by art historians, the model also suggested other visual influences between artists, thus generating new information. The second paper investigates changes in popular music using a dataset of 464,411 songs from 1955 to 2010. The dataset included "a variety of popular genres, including rock, pop, hip hop, metal, or electronic." The authors concluded that over time, there was "the restriction of pitch transitions" and "the homogenization of the timbral palette" – in other words, some of the musical variability has decreased. The third paper analyzes gradual changes in average shot duration across 9400 English-language film from 1912-2013 (**image 3**). The forth paper used machine learning to predict artist name, period and material of artworks in a dataset of 432,217 art images.

These and many other papers contain valuable and original insights that would be impossible to arrive without computational methods by only using "armchair" theorizing, or small group ethnographic observations. But because most of these studies use a statistical approach, they also have a common limitation. One of the meaning of "statistics" is that it is a summary of characteristics of a collection of information. And any summary by default is going to omit some details, because it is smaller than the original information. Therefore, if we use statistical approach to summarize a collection of cultural artifacts, or a dataset of information about cultural behaviors (for instance, sharing, liking, or commenting on particular images in social networks) to find patterns or propose relationships, they will not apply to everything in the dataset. For example, the paper about effects of Instagram image filters mentioned above found that "filtered photos are 21% more likely to be viewed and 45% more likely to be commented" by analyzing millions of photos. But certainly many filtered photos in this sample were not viewed more frequently or received more comments, because statistical predictions only work some of the time. (While non-qualitative studies of cultural genres and periods such as *Classical Hollywood Cinema* are also summaries proposing artistic techniques and conventions common to a particular set of films, they also analyze particularly important films that can't be reduced to these conventions. We can also recall Roland Barthes's *S/Z* which did not claim to describe all semiotic codes in Balzac's short story, nor to claim that they form a system.[17])

---

[15] J. E. Cutting, J. E. and A. Candan, A. "Shot durations, shot classes, and the increased pace of popular movies," *Projections: The journal for movies and mind,* 2015*, 9*(2), 40-62, http://people.psych.cornell.edu/~jec7/pubs/cutting&candanProj15.pdf.
[16] Gjorgji Strezoski and Marcel Worring. "OmniArt: Multi-task Deep Learning for Artistic Data Analysis." 2017, https://arxiv.org/abs/1708.00684.
[17] https://monoskop.org/images/d/d6/Barthes_Roland_S-Z_2002.pdf.



3. Graphs showing decreasing shot duration in 9400 English-language film from 1912-2012. The average shot duration in each film is represented by a dot. The graphs show different possible fit lines for this data. Source: Cutting, J. E. & Candan, A. (2015). [Shot durations, shot classes, and the increased pace of popular movies.]() *Projections: The journal for movies and mind, 9*(2), 40-62.

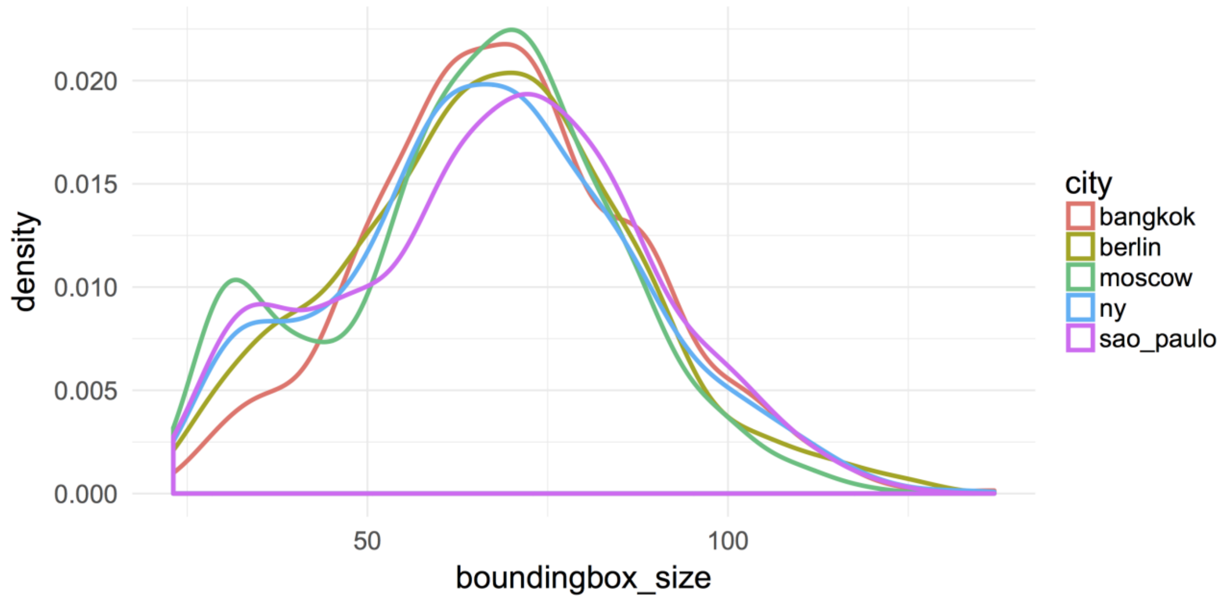
For science having a statistical model which only works sometimes is a problem, because science assumes that such a model should accurately capture characteristics of a phenomena. But we can use computational methods to study culture without such assumptions, and this is a part of the alternative research paradigm that I call Cultural

Analytics.[18] In this paradigm, we do not want to "explain" most or even some of the data using a simple mathematical model, and treat the rest as "error" or "noise" just because our mathematical model cannot account for it. And we do not want to assume that cultural variation is a deviation from a mean. We also do not want to assume that large proportions of works in particular medium of genre follow a single pattern or only a few patterns such as "hero's journey," "golden ratio" or "binary oppositions," or that every culture goes through the same few stages of development as it was claimed by some art historians in the nineteenth century.

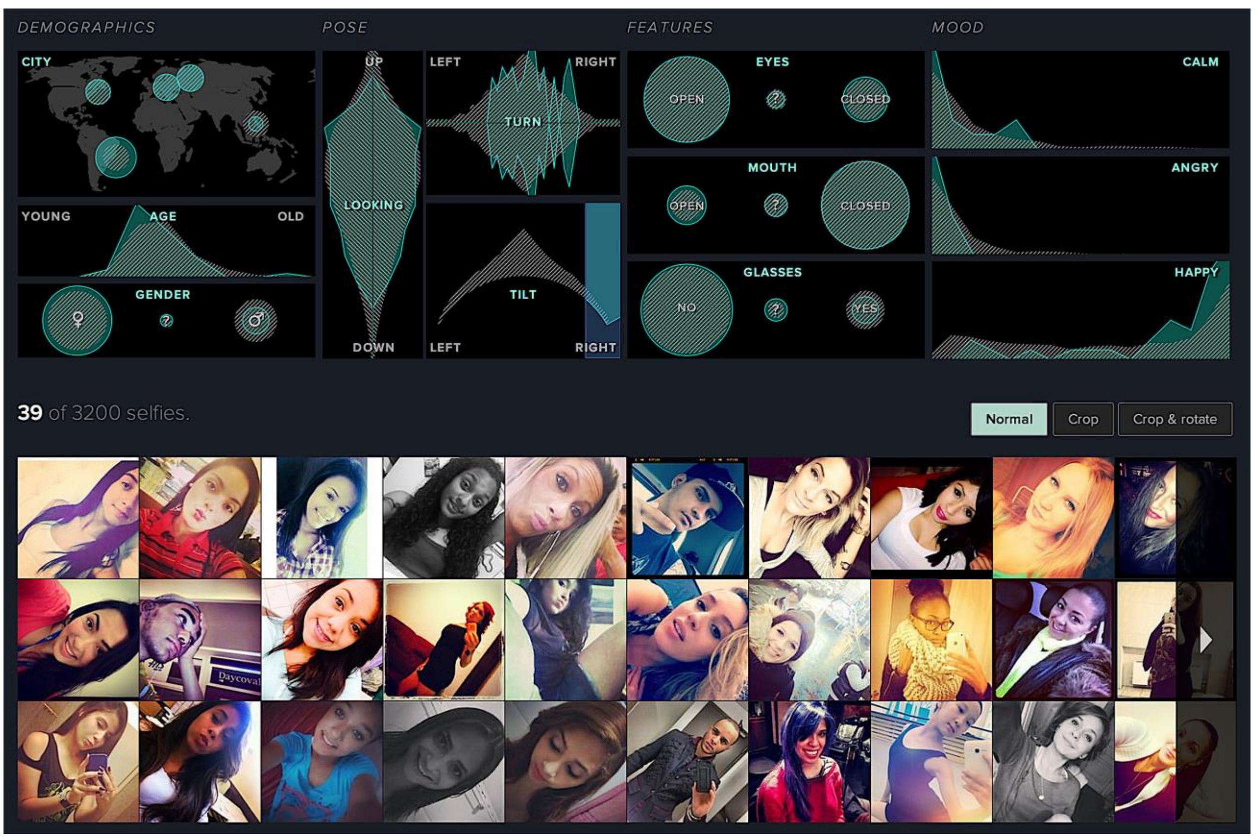
I believe that *we can study cultural diversity without assuming that it is caused by variations from some types or structures*. Does this mean that we are only interested in the differences and that we want to avoid any kind of reduction at all cost? To postulate existence of cultural patterns is to accept that we are doing at least some reduction in analyzing data. Without this, we cannot compare anything, unless we are dealing with extreme minimalism or seriality, where the artist makes everything else equal and only varies a single variable, like Sol LeWitt.

We can define Cultural Analytics as the quantitative study of cultural patterns on different scales. But then we need to immediately qualify this statement. While we want to discover repeating patterns in cultural data, we should always remember that they only account for *some* aspects of the artifacts and their reception.

[18] Lev Manovich, "Cultural Analytics: Visualizing Cultural Patterns in the Era of "More Media" (*Domus*, 2009), http://manovich.net/index.php/projects/cultural-analytics-visualizing-cultural-patterns.



4. Distributions of the face sizes in 3200 Instagram selfie photos shared in five cities, 12/2013. 640 photos from each city are used. Data and additional details: selfiecity.net.



5. Screenshot from selfiecity.net, showing a selection of Instagram selfie photos with similar face title angles. This similarity in face angles makes more clear the differences between faces on all other dimensions.

Unless it is a 100% copy of another cultural artifact, or produced mechanically or algorithmically to be identical with others, every expression and interaction is unique. In some cases, this uniqueness is not important in analysis, and in other cases it is. For example, characteristics of faces we extracted automatically from a dataset of Instagram self-portraits revealed interesting differences in how people represent themselves in this medium in particular cities and periods we analyzed (**image 4**).[19] But the reason we do not get tired looking at endless faces and bodies when we browse Instagram is that each of them is unique. What fascinates us is not repeating patterns but unique details and their combinations (**image 5**).

The ultimate goal of Cultural Analytics should be to map and understand in detail the *diversity* of contemporary professional and user-generated cultural artifacts created globally—i.e. to focus on what is different between numerous artifacts and not only on what they share. In the nineteenth and twentieth century the lack of appropriate technologies to store, organize, and compare large cultural datasets was contributing to the popularity of reductive cultural theories. Today we can use a single computer to map and visualize thousands of differences between tens of millions of objects. We do not have an excuse any more to only focus what cultural artifacts or behaviors share, which is what we do then we categorize them, or perceive them as instances of general types. So while we may have to start with extracting patters first just to draw our initial maps of contemporary cultural production and dynamics given its scale, eventually they may recede in the background or even completely dissolve, as we focus only on the differences between individual objects. This is to me is the ultimate promise behind using AI methods to study culture.

[19] http://selfiecity.net/, 2013-2015.