Large image datasets: A pyrrhic win for computer vision?

Anonymous submission

Paper ID

Abstract

In this paper we investigate problematic practices and consequences of large scale vision datasets. We examine broad issues such as the question of consent and justice as well as specific concerns such as the inclusion of verifiably pornographic images in datasets. Taking the ImageNet-ILSVRC-2012 dataset as an example, we perform a cross-sectional model-based quantitative census covering factors such as age, gender, NSFW content scoring, class-wise accuracy, human-cardinality-analysis, and the semanticity of the image class information in order to statistically investigate the extent and subtleties of ethical transgressions. We then use the census to help hand-curate a look-up-table of images in the ImageNet-ILSVRC-2012 dataset that fall into the categories of verifiably pornographic: shot in a non-consensual setting (up-skirt), beach voyeuristic, and exposed private parts. We survey the landscape of harm and threats both society broadly and individuals face due to uncritical and ill-considered dataset curation practices. We then propose possible courses of correction and critique the pros and cons of these. We have duly open-sourced all of the code and the census meta-datasets generated in this endeavor for the computer vision community to build on. By unveiling the severity of the threats, our hope is to motivate the constitution of mandatory Institutional Review Boards (IRB) for large scale dataset curation processes.

1. Introduction

Born from World War II and the haunting and despicable practices of Nazi era experimentation [4] the *1947 Nuremberg code* [84] and the subsequent *1964 Helsinki declaration* [30], helped to establish the doctrine of **Informed Consent** which builds on the fundamental notions of human dignity and agency to control dissemination of information about oneself. This has shepherded data collection endeavors in the medical and psychological sciences concerning human subjects, including photographic data [8, 56], for the past several decades. A less stringent version of informed consent, *broad consent*, proposed in 45 CFR 46.116(d) of the *Revised Common Rule* [24], has been recently introduced that still affords the basic safeguards towards protecting one's identity in large scale databases. However, in the age of *Big Data*, the fundamentals of informed consent, privacy, or agency of the individual have gradually been eroded. Institutions, academia, and industry alike, amass millions of images of people without consent and often for unstated purposes under the guise of anonymization, a claim that is both ephemeral [57, 68] and vacuous [30]. As can be seen in Table 1, several tens of millions of images of people are found in peer-reviewed literature. These images are obtained without consent or awareness of the individuals or IRB approval for collection. In *Section 5-B* of [79], for instance, the authors nonchalantly state "*As many images on the web contain pictures of people, a large fraction (23%) of the 79 million images in our dataset have people in them*". With this background, we now focus on one of the most celebrated and canonical large scale image datasets: the *ImageNet* dataset.

1.1. ImageNet: A brief overview

The emergence of the ImageNet dataset [21] is widely considered a pivotal moment[2] in the Deep Learning revolution that transformed Computer Vision (CV), and Artificial Intelligence (AI) in general. Prior to ImageNet, computer vision and image processing researchers trained image classification models on small dataset such as CalTech101 (9k images), PASCAL-VOC (30k images), LabelMe (37k images), and the SUN (131k images) dataset (see slide-37 in [51]). ImageNet, with over 14 million images spread across 21,841 synsets, replete with 1,034,908 bounding box annotations, brought in an aspect of scale that was previously missing. A subset of 1.2 million images across 1000 classes was carved out from this dataset to form the ImageNet-1k dataset (popularly called `ILSVRC-2012`) which formed the basis for the *Task-1: classification* challenge in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). This soon became widely touted as the *Computer Vision Olympics*[3]. The vastness of this dataset allowed a Convolutional Neural

[2] "The data that transformed AI research—and possibly the world": https://bit.ly/2VRxx3L

[3] https://engineering.missouri.edu/2014/01/team-takes-top-rankings-in-computer-vision-olympics/

Dataset	Number of images (in millions)	Number of categories (in thousands)	Number of consensual images
JFT-300M ([41])	300+	18	0
Open Images ([50])	9	20	0
Tiny-Images ([79])	79	76	0
Tencent-ML ([89])	18	11	0
ImageNet-(21K,11k[1],1k) ([70])	(14, 12, 1)	(22, 11, 1)	0
Places ([93])	11	0.4	0

Table 1: Large scale image datasets containing people's images

Network (CNN) with 60 million parameters [49] trained by the *SuperVision* team from University of Toronto to usher in the rebirth of the CNN-era (see [3]), which is now widely dubbed the *AlexNet moment* in AI.

Although ImageNet was created over a decade ago, it remains one of the most influential and powerful image databases available today. Its power and magnitude is matched by its unprecedented societal impact. Although an *a posteriori* audit might seem redundant a decade after its creation, ImageNet's continued significance and the culture it has fostered for other large scale datasets warrants an ongoing critical dialogue.

From the questionable ways images were sourced, to troublesome labeling of people in images, to the downstream effects of training AI models using such images, ImageNet and large scale vision datasets (LSVD) in general *constitute a Pyrrhic win* for computer vision. We argue, this win has come at the expense of harm to *minoritized groups* and further aided the gradual erosion of privacy, consent, and agency of both the individual and the collective.

The rest of this paper is structured as follows. In section 2, we cover related work that has explored the ethical dimensions that arise with LSVD. In section 3, we describe the landscape of both the immediate and long term threats individuals and society as a whole encounter due to ill-considered LSVD curation. In Section 4, we propose a set of solutions which might assuage some of the concerns raised in section 3. In Section 5, we present a template quantitative auditing procedure using the ILSVRC2012 dataset as an example and describe the data assets we have curated for the computer vision community to build on. We conclude with broad reflections on LSVDs, society, ethics, and justice.

2. Background and related work

The very declaration of a taxonomy brings some things into existence while rendering others invisible [9]. A gender classification system that conforms to essentialist binaries, for example, operationalizes gender in a cis-centric way resulting in exclusion of non-binary and transgender people [48]. Categories simplify and freeze nuanced and complex narratives, obscuring political and moral reasoning behind a category. Over time, messy and contingent histories hidden behind a category are forgotten and trivialized [75]. With the adoption of taxonomy sources, image datasets inherit seemingly invisible yet profoundly consequential shortcomings. The dataset creation process, its implication for ML systems, and subsequently, the societal impact of these systems has attracted a substantial body of critique. We categorize such body of work into two groups that compliment one another. While the first group can be seen as concerned with the broad downstream effects, the other concentrates mainly on the dataset creation process itself.

2.1. Broad critiques:

The absence of critical engagement with canonical datasets disproportionately negatively impacts women, racial and ethnic minorities, and vulnerable individuals and communities at the margins of society [7]. For example, image search results both exaggerate stereotypes and systematically under-represent women in search results for occupations [47]; object detection systems designed to detect pedestrians display higher error rates for recognition of demographic groups with dark skin tones [87]; and gender classification systems show disparities in image classification accuracy where lighter-skin males are classified with the highest accuracy while darker-skin females suffer the most misclassification [14]. Gender classification systems that lean on binary and cis-genderist constructs operationalize gender in a trans-exclusive way resulting in tangible harm to trans people [48]. With a persistent trend where minoritized and vulnerable individuals and communities often disproportionately suffer the negative outcomes of ML systems, [25] have called for a shift in rethinking ethics not just as a fairness metric to mitigate the narrow concept of bias but as practice that results in justice for the most negatively impacted. Similarly, [46] contend that perspectives that acknowledge existing inequality and aim to redistribute power are pertinent as opposed to fairness-based perspectives. Such understanding of *ethics as justice* then requires a focus beyond 'bias' and fairnesss' in LSVDs and requires questioning of how images are sourced, labelled, and what it means for models to be trained on them. One of the most thorough investigation in this regard comes from [20]. In this recent work, Crawford and Paglen present an in-depth critical examination of ImageNet including the dark and troubling results of classifying people as if they are objects. Offensive and derogatory labels that perpetuate historical and current prejudices are assigned to people's actual images. The authors emphasise that not only are images that were scraped across the web appropriated as data for computer vision tasks, but also the very act of assigning labels to people based on physical features raises fundamental concerns around reviving long-discredited pseudo-scientific ideologies of physiognomy [90].

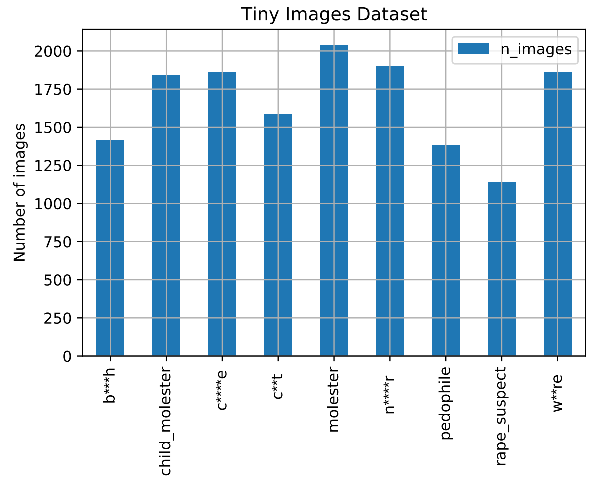
2.2. Critiques of the curation phase:

Within the dataset creation process, *taxonomy sources* pass on their limitations and taken for granted assumptions. The adoption of underlying structures present a challenge where — without critical examination of the architecture — ethically dubious taxonomies are inherited. This has been one of the main challenges for ImageNet given that the dataset is built on the backbone of WordNet's structure. Acknowledging some of the problems, the authors from the ImageNet team did recently attempt to address [91] the stagnant concept vocabulary of WordNet. They admitted that only 158 out of the 2,832 existing synsets should remain in the person sub-tree[4]. Nonetheless, some serious problems remain untouched. This motivates us to address in greater depth the overbearing presence of the *WordNet effect* on image datasets.

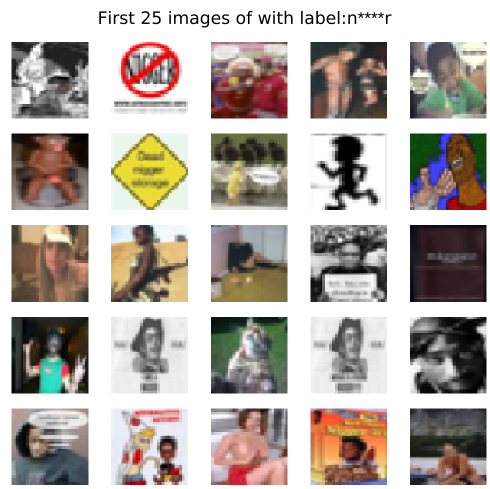
2.3. The WordNet Effect

ImageNet is not the only large scale vision dataset that has inherited the shortcomings of the WordNet taxonomy. The *80 million Tiny Images* dataset [79] which grandfathered the CIFAR-10/100 datasets also used the same path. Unlike ImageNet, this dataset has never been audited or scrutinized and some of the sordid results from inclusion of *ethnophaulisms* in its label space are displayed in Figure 1. The figure demonstrates both the number of images in a subset of the *offensive* classes (sub-figure(a)) and the exemplar images (sub-figure(b)) that show the images in the noun-class labelled `n****r`[5], a fact that serves as a stark reminder that a great deal of work remains to be done by the ML community at large.

And finally, the *labeling and validation* of the curation process also presents ethical challenges. Recent works such as [37] has explored the intentionally hidden labour, which they have termed as *Ghost Work*, behind such tasks. Image labeling and validation requires the use of crowd-sourced platforms such as MTurk, often contributing to the exploitation of underpaid and undervalued *gig workers*. Within the topic of image labeling but with a different dimension and focus, recent work such as [80] and [6] has focused on the shortcomings of human-annotation procedures used during the ImageNet dataset curation. These shortcomings, the authors point out, include single label per-image procedure that causes problems given that real-world images often contain multiple objects, and inaccuracies due to "overly restrictive label proposals".



(a) Class-wise counts of the offensive classes



(b) Samples from the class labelled `n****r`

Figure 1: Results from the *80 Million Tiny Images* dataset

[4] In order to prune all the nodes. They also took into account the *imageability* of the synsets and the skewed representation in the images pertaining to the *Image retrieval* phase

[5] Due to its offensiveness, we have censored this word here, however, it remains uncensored on the website at the time of writing.

3. The threat landscape

In this section, we survey the landscape of harm and threats, both immediate and long term, that emerge with dataset curation practices in the absence of careful ethical considerations and anticipation for negative societal consequences. Our goal here is bring awareness to the ML and AI community regarding the severity of the threats and to motivate a sense of urgency to act on these. We hope this will result in practices such as the mandatory constitution of Institutional Review Boards (IRB) for large scale dataset curation processes.

1: The rise of reverse image search engines, loss of privacy, and the blackmailing threat: Large image datasets, when built without careful consideration of societal implications, pose a threat to the welfare and well-being of individuals. Most often, vulnerable people and marginalised populations pay a disproportionately high price. Reverse image search engines[6] that allow face search such as [2] have gotten remarkably and worryingly efficient in the past year. For a small fee, anyone can use their portal or their API[7] to run an automated process to uncover the "real-world" identities of the *humans of ImageNet* dataset. For example, in societies where sex work is socially condemned or legally criminalized, re-identification of a sex worker through image search, for example, bears a real danger for the individual victim. Harmful discourse such as *revenge porn*, are part of a broader continuum of image-based sexual abuse [52]. To further emphasize this specific point, many of the images in classes such as `maillot`, `brassiere`, and `bikini` contain images of beach voyeurism and other non-consensual cases of digital image gathering (covered in detail in Section-5). We were able to (unfortunately) easily map the victims, most of whom are women, in the pictures to "real-world" identities of people belonging to a myriad of backgrounds including teachers, medical professionals, and academic professors using reverse image search engines such as [63]. Paying heed to the possibility of the *Streisand effect*[8], we took the decision not to divulge any further quantitative or qualitative details on the extent or the location of such images in the dataset besides alerting the curators of the dataset(s) and making a passionate plea to the community not to underestimate the severity of this particular threat vector.

2: The emergence of even larger and more opaque datasets: The attempt to build computer vision has been gradual and can be traced as far back as 1966 to Papert's *The Summer Vision Project* [60], if not earlier. However, ImageNet, with its vast amounts of data, has not only erected a canonical landmark in the history of AI, it has also paved the way for even bigger, more powerful, and suspiciously opaque datasets. The lack of scrutiny of the ImageNet dataset by the wider computer vision community has only served to embolden institutions, both academic and commercial, to build far bigger datasets without scrutiny (see Table 1). Various highly cited and celebrated papers in recent years [10, 16, 41, 77], for example, have used the *unspoken unicorn* amongst large scale vision datasets, that is, the JFT-300M dataset [?][9]. This dataset is inscrutable and operates in the dark, to the extent that there has not even been official communication as to what *JFT-300M* stands for. All that the ML community knows is it purportedly boasts more than 300M images spread across 18k categories. The open source variant(s) of this, the *Open Images V4-5-6* [50] contains a subset of 30.1M images covering 20k categories (and also has an extension dataset with 478k crowd-sourced images across more than 6000 categories). While parsing through some of the images, we found **verifiably**[10] non-consensual images of children that were siphoned off of *flickr* hinting towards the prevalence of similar issues for JFT-300M from which this was sourced. Besides the other large datasets in Table 1, we have cases such as the *CelebA-HQ* dataset, which is actually a *heavily processed* dataset whose grey-box curation process only appears in Appendix-C of [45] where no clarification is provided on this *"frequency based visual quality metric"* used to sort the images based on *quality*. Benchmarking any downstream algorithm of such an opaque, biased and a (semi-)synthetic dataset will only result in controversial scenarios such as [53], where the authors had to hurriedly incorporate addendums admitting biased results. Hence, it is important to reemphasize that the existence and use of such datasets bears direct and indirect impact on people, given that decision making on social outcomes increasingly leans on ubiquitously integrated AI systems trained and validated on such dataset. Yet, despite such profound consequences, critical questions such as where the data comes from or whether the images were obtained consensually are hardly considered part of the LSVD curation process.

The more nuanced and perhaps indirect impact of ImageNet is the **culture** that it has cultivated within the broader AI community; a culture where the appropriation of images of real people as raw material free for the taking has come be to perceived as *the norm*. Such norm and lack of scrutiny has played a role towards the creation of monstrous and secretive datasets without much resistance, prompting further questions such as 'what other secretive datasets currently exist

[6] For example, PimEyes: https://bit.ly/3bSKcZQ

[7] Please refer to the supplementary material for the screenshots

[8] The Streisand effect *"is a social phenomenon that occurs when an attempt to hide, remove, or censor information has the unintended consequence of further publicizing that information, often via the Internet"* [86]

[9] We have decided to purposefully leave the '?' in place and plan to revisit it only after the dataset's creator(s) publish the details of it's curation

[10] See https://bit.ly/2y1sC7i. We performed verification with the uploader of the image via the Flickr link shared.

hidden and guarded under the guise of proprietary assets?' Current work that has sprung out of secretive datasets, such as Clearview AI [40] [11], points to a deeply worrying and insidious threat not only to vulnerable groups but also to the very meaning of privacy as we know it [44].

3: The Creative Commons fallacy: In May 2007 the iconic case of *Chang versus Virgin mobile: The school girl, the billboard, and virgin* [17] unraveled in front of the world, leading to widespread debate on the uneasy relationship between personal privacy, consent, and image copyright, initiating a substantial corpus of academic debate (see [15, 18, 19, 39]). A Creative Commons license addresses only copyright issues – not privacy rights or consent to use images for training. Yet, many of the efforts beyond ImageNet, including the Open Images dataset [50], have been built on top of the *Creative commons* loophole that large scale dataset curation agencies interpret as a *free for all, consent-included* green flag. This, we argue, is fundamentally fallacious as is evinced in the views presented in [54] by the Creative commons organization that reads: *"CC licenses were designed to address a specific constraint, which they do very well: unlocking restrictive copyright. But copyright is not a good tool to protect individual privacy, to address research ethics in AI development, or to regulate the use of surveillance tools employed online."*. Datasets culpable of this *CC-BY heist* such as *MS-Celeb-1M* and *IBM's Diversity in Faces* have now been deleted in response to the investigations (See [28] for a survey) lending further support to the Creative Commons fallacy.

4: Blood diamond effect in models trained on this dataset: Akin to the *ivory carving-illegal poaching* and *diamond jewelry art-blood diamond* nexuses, we posit that there is a similar moral conundrum at play here that effects all downstream applications entailing models trained using a *tainted* dataset. Often, these transgressions may be rather subtle. In this regard, we pick an examplar field of application that on the surface appears to be a low risk application area: *Neural generative art*. Neural generative art created using tools such as BigGAN [11] and Art-breeder [1] that in turn use pre-trained deep-learning models trained on ethically dubious datasets, bear the downstream burden[12] of the problematic residues from non-consensual image siphoning, thus running afoul of the Wittgensteinian edict of *ethics and aesthetics being one and the same.* [29]. We also note that there is a privacy-leakage facet to this *downstream burden*. In the context of face recognition, works such as [74] have demonstrated that CNNs with high predictive power unwittingly accommodate accurate extraction of subsets of the facial images that they were trained on, thus abetting dataset leakage.

5: Perpetuation of unjust and harmful stereotypes: Finally, zooming out and taking a broad perspective allows us to see that the very practice of embarking on a classification, taxonomization, and labeling task endows the classifier with the power to decide what is a legitimate, normal, or correct way of being, acting, and behaving in the social world [9]. For any given society, what comes to be perceived as *normal* or *acceptable* is often dictated by dominant ideologies. Systems of classification, which operate within power asymmetrical social hierarchy, necessarily embed and amplify historical and cultural prejudices, injustices, and biases [75]. In western societies, how "desirable", "positive", and "normal" characteristics and ways of being are constructed and maintained in a way that align with the dominant narrative, giving advantage to those that fit the status quo. Groups and individuals on the margins, on the other hand, are often perceived as the "outlier" and the "deviant". Image classification and labelling practices, without the necessary precautions and awareness of these problematic histories, pick up these stereotypes and prejudices and perpetuate them [31, 58, 59]. AI systems trained on such data amplify and normalize these stereotypes, inflicting unprecedented harm on those that are already on the margins of society. While the ImageNet team did initiate strong efforts towards course-correction [92], the Tiny Images dataset still contains harmful slurs and offensive labels. And worse, we remain in the dark regarding the secretive and opaque LSVDs.

4. Candidate solutions: The path ahead

Decades of work within the fields of Science and Technology Studies (STS) and the Social Sciences show that there is no single straightforward solution to most of the wider social and ethical challenges that we have discussed [5, 25, 76]. These challenges are deeply rooted in social and cultural structures and form part of the fundamental social fabric. Feeding AI systems on the world's beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy [5]. These challenges and tensions will exist as long as humanity continues to operate. Given the breadth of the challenges that we have faced, any attempt for a quick fix risks concealing the problem and providing a false sense of solution. The idea of a complete removal of biases, for example, might in reality be simply hiding them out of sight [36]. Furthermore, many of the challenges (bias, discrimination, injustice) vary with context, history, and place, and are concepts that continually shift and change constituting a moving target [7]. The pursuit of panacea in this context, therefore,

[11] Clearview AI is a US based privately owned technology company that provides facial recognition service to various customers including North American law enforcement agencies. With more than 3 billion photos scraped from the web, the company operated in the dark until its services to law enforcement was reported in late 2019

[12] Please refer to the supplementary material where we demonstrate one such real-world experiment entailing unethically generated neural art replete with responses obtained from human critiques as to what they felt about the imagery being displayed.

Dataset audit card - ImageNet

Census audit statistics

- 83436 images with $101070 - 132201$ persons (Models: DEX ([69]), InsightFace ([38]))
- Mean-age (male): 33.24 (Female):25.58 (RetinaFace [23], ArcFace [22])
- Confirmed misogynistic images: 62. Number of classes with infants: 30
- ($\mu_c^{(A)}$ and $\sigma_c^{(A)}$: Mean and standard-deviation of the gender-estimate of images in class c estimated by algorithm (A).)

Metrics: Class-level mean count ($\eta_c^{(A)}$), mean gender skewness ($\xi_c^{(A)}$) and mean-age ($\alpha_c^{(A)}$):

$$\eta_c^{(A)} = \frac{1}{N_c}\sum_{i=1}^{N_c} I[\phi_i], \alpha_c^{(A)} = \frac{1}{N_c}\sum_{i=1}^{N_c} I[\phi_i] a_i^{(A)} \ and$$

$$\xi_c^{(A)} = \frac{1}{N_c}\sum_{i=1}^{N_c} I[\phi_i]\left(\frac{g_i^{(A)}-\mu_c^{(A)}}{\sigma_c^{(A)}}\right)^3$$

$$\phi_i = \begin{cases} 1 & \text{if face present} \\ 0 & \text{otherwise} \end{cases} \text{ in } i^{th} \text{ image.}$$

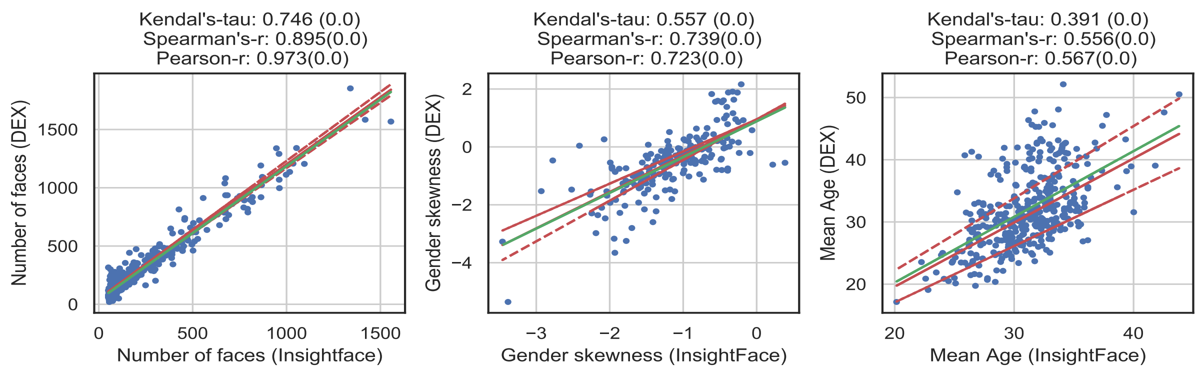


Figure 2: Class-wise cross-categorical scatter-plots across the cardinality, age and gender scores

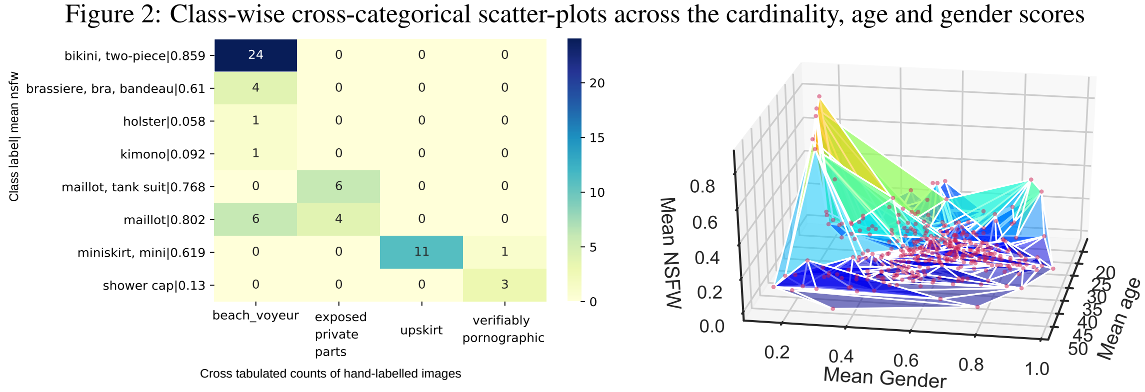


Figure 3: Statistics and locationing of the hand-labelled images

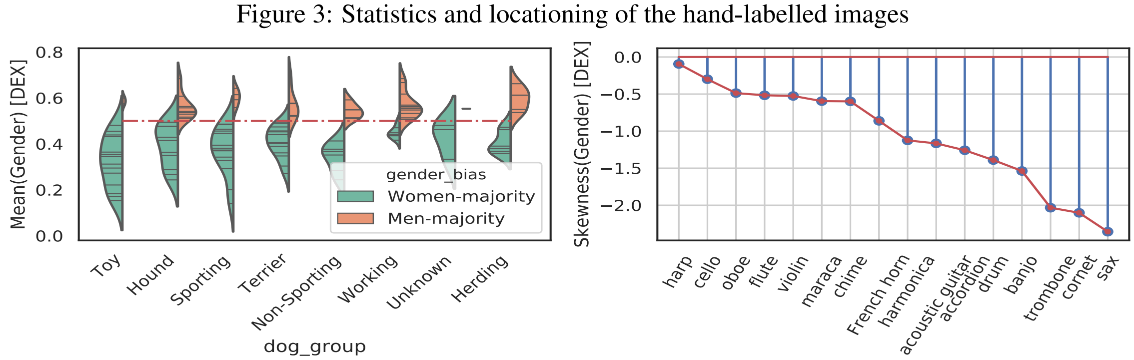


Figure 4: Known *human co-occurrence* based gender-bias analysis

Figure 5: Dataset audit card for the ImageNet dataset

is not only unattainable but also misguided. Having said that, there are remedies that can be applied to overcome the specific harms that we have discussed in this paper, which eventually potentially play constituent roles in improving the wider and bigger social and structural issues in the long run.

4.1. Remove, replace, and open strategy

In [92], the authors concluded that within the *person sub-tree* of the ImageNet dataset, 1593 of the 2832 people categories were *potentially offensive* labels and planned to *"remove all of these from ImageNet."*. We strongly advocate a similar path for the offensive noun classes in the Tiny Images dataset that we have identified in section 2.1, as well as images that fall into the categories of verifiably pornographic, shot in a non-consensual setting (up-skirt), beach voyeuristic, and exposed genitalia in the ImageNet-ILSVRC-2012 dataset. In cases where the image category is retained but the images are not, the option of *replacement with consensually shot financially compensated images* arises. It is possible that some of the people in these images might come forward to consent and contribute their images in exchange for fair financial compensation, credit, or out of sheer altruism [12]. We re-emphasize that our consternation focuses on the non-consensual aspect of the images and not on the category-class and the ensuing content of the images in it. This solution, however, brings forth further questions: does this make image datasets accessible only to those who can afford it? Will we end up with pool of images with a predominantly financially disadvantaged participants?

Science is self-correcting so long as it is accessible and open to critical engagement and this is what we have done given what we know of these LSVDs. The secretive and opaque LSVDs thread a dangerous territory, given that they directly or indirectly impact society. We strongly contend that making them open and accessible is a crucial first step towards an ethical scientific endeavour.

4.2. Differentially private obfuscation of the faces

This path entails harnessing techniques such as DP-Blur [32] with quantifiable privacy guarantees to obfuscate the identity of the humans in the image. The *Inclusive images challenge* [73], for example, already incorporated blurring during dataset curation[13] and addressed the downstream effects surrounding change in predictive power of the models trained on the blurred versions of the dataset curated. We believe that replication of this template that also clearly included avenues for recourse in case of an erroneously non-blurred image being sighted by a researcher will be a step in the right direction for the community at large.

[13] https://www.kaggle.com/c/inclusive-images-challenge

4.3. Synthetic-to-real and Dataset distillation

The basic idea here is to utilize (or augment) synthetic images in lieu of real images during model training. Approaches include using hand-drawn sketch images (*ImageNet-Sketch* [82]), using GAN generated images [26] and techniques such as *Dataset distillation* [83],where a dataset or a subset of a dataset is distilled down to a few representative ***synthetic*** samples. This is a nascent field with some promising results emerging in unsupervised domain adaptation across visual domains [61] and universal digit classification [64].

4.4. Ethics-reinforced filtering during the curation

The specific ethical transgressions that emerged during our longitudinal analysis of ImageNet could have been prevented if there were explicit instructions provided to the *MTurkers* during the dataset curation phase to enable filtering of these images at the source (See Fig.9 in [67] for example). We hope ethics checks become an integral part of the User-Interface deployed during the humans-in-the-loop validation phase for future dataset curation endeavors.

4.5. Dataset audit cards

Much along the lines of *model cards* [55] and *datasheet for datasets* [35], we propose dissemination of *dataset audit cards*. This allows large scale image dataset curators to publish the goals, curation procedures, known shortcomings and caveats alongside their dataset dissemination.
In Figure 5, we have curated an example dataset audit card for the ImageNet dataset using the quantitative analyses carried out in Section 5

5. Quantitative dataset auditing: ImageNet as a template

We performed a cross-categorical quantitative analysis of ImageNet to assess the extent of the ethical transgressions and the feasibility of model-annotation based approaches. This resulted in an *ImageNet census*, entailing both image-level as well as class-level analysis across the 57 different metrics (see supplementary section) covering Count, Age and Gender (CAG), NSFW-scoring, semanticity of class labels and accuracy of classification using pre-trained models. We have distilled the important revelations of this census as a *dataset audit card* presented in Figure 5. This audit also entailed a human-in-the-loop based hybrid-approach that the pre-trained-model annotations (along the lines of [27, 92]) to segment the large dataset into smaller sub-sets and hand-label the smaller subsets to generate two lists covering 62 misogynistic images and 30 image-classes with co-occuring children. We used the `DEX` [69] and the `InsightFace` [38] pre-trained models[14] to generate the cardinality, gender

[14] While harnessing these pre-trained gender classification models, we would like to **strongly emphasize** that the specific models and the *problems*

file_name	shape	file_contents
df_insightface_stats.csv	(1000, 30)	24 classwise statistical parameters obtained by running the `InsightFace` model ([38]) on the ImageNet dataset
df_audit_age_gender_dex.csv	(1000, 12)	11 classwise (ordered by the wordnet-id) statistical parameters obtained from the json files (of the DEX paper) [69]
df_nsfw.csv	(1000, 5)	The mean and std of the NSFW scores of the train and val images arranged per-class. (Unnamed: 0: WordNetID of the class)
df_acc_classwise_resnet50.csv	(1000, 7)	Classwise accuracy metrics (& the image level preds) obtained by running the ResNet50 model on ImageNet train and Val sets
df_acc_classwise_NasNet_mobile.csv	(1000, 7)	Classwise accuracy metrics (& the image level preds) obtained by running the NasNet model on ImageNet train and Val sets
df_imagenet_names_umap.csv	(1000, 5)	DF with 2D UMAP embeddings of the Glove vectors of the classes of the ImageNet dataset
df_census_imagenet_61.csv	(1000, 61)	The MAIN census dataframe covering class-wise metrics across 61 parameters, all of which are explained in df_census_columns_interpretation.csv
df_census_columns_interpretation.csv	(61, 2)	The interpretations of the 61 metrics of the census dataframe above!
df_hand_survey.csv	(61, 3)	Dataframe contaimning the details of the 61 images unearthed via hand survey (Do not pay heed to 61. it is a mere coincidence)
df_classes_tiny_images_3.csv	(75846, 3)	Dataframe containing the class_ind, class_name (wordnet noun) and n_images
df_dog_analysis.csv	(7, 4)	Dataframe containing breed, gender_ratio and survey result from the paper Breed differences in canine aggression'

Table 2: Meta datasets curated during the audit processes

skewness, and age-distribution results captured in Figure 2. This resulted in discovery of **83,436** images with persons, encompassing **101,070 to 132,201** individuals, thus constituting $8-10\%$ of the dataset. Further, we munged together gender, age, class semanticity[15] and NSFW content flagging information from the pre-trained `NSFW-MobileNet-v2` model [34] to help perform a guided search of misogynistic consent-violating transgressions. This resulted in discovery of 62 images[16] across four categories: *beach-voyeur-photography, exposed-private-parts, verifiably pornographic and upskirt* in the following classes: *445-Bikini, 638 -maillot, 639-tank suit, 655-miniskirt* and *459-brassiere* (see Figure 3). Lastly, we harnessed literature from areas spanning from dog-ownership bias ([42],[66]) to engendering of musical instruments ([88], [13]) to generate analysis of subtle forms of *human co-occurrence*-based gender bias in Figure 4. Captured in Table 2 are the details of the *csv* formatted data assets curated for the community to build on. The CAG statistics are covered in df_insightface_stats.csv and df_audit_age_gender_dex.csv. Similarly, we have also curated NSFW scoring (df_nsfw.csv), Accuracy (df_acc_classwise_resnet50/_NasNet_mobile.csv and Semanticity (df_imagenet_names_umap.csv) datasets as well. df_census_imagenet_61.csv contains the 61 cumulative paramaters for each of the 1000 classes (with their column interpretations in df_census_columns_interpretation.csv). We have duly open-sourced these meta-datasets and 14 tutorial-styled Jupyter notebooks (spanning both ImageNet and Tiny-Images datasets) for community access[17].

6. Conclusion and discussion

We have sought to draw the attention of the machine learning community towards the societal and ethical implications of large scale datasets, such as the problem of non-consensual images and the oft-hidden problems of categorizing people. ImageNet has been championed as one of the most incredible breakthroughs in computer vision, and AI in general. We indeed celebrate ImageNet's achievement and recognize the creators' efforts to grapple with some ethical questions. Nonetheless, ImageNet as well as other large image datasets remain troublesome. In hindsight, perhaps the ideal time to have raised ethical concerns regarding LSVD curation would have been in 1966 at the birth of *The Summer Vision Project* [60]. The right time after that was when the creators of ImageNet embarked on the project to "map out the entire world of objects". Nonetheless, these are crucial conversations that the computer vision community needs to engage with **now** given the rapid democratization of imaging scraping tools ([71, 72, 81]) and *dataset-zoos* ([43, 65, 78]). The continued silence will only serve to cause more harm than good in the future. In this regard, we have outlined a few solutions, including *audit cards*, that can be considered to ameliorate some of the concerns raised. We have also curated meta-datasets and open-sourced the code to carry out quantitative auditing using the ILSVRC2012 dataset as a template. However, we posit that the deeper problems are rooted in the wider structural traditions, incentives, and discourse of a field that treats ethical issues as an afterthought.A field where *in the wild* is often a euphemism for *without consent*. We are up against a system that has veritably mastered *ethics shopping*, *ethics bluewashing*, *ethics lobbying*, *ethics dumping*, and *ethics shirking* [33].

Within such an ingrained tradition, even the most thoughtful scholar can find it challenging to pursue work outside the frame of the "tradition". Subsequently, radical ethics that challenge deeply ingrained traditions need to be incentivised and rewarded in order to bring about a shift in culture that centres justice and the welfare of disproportionately impacted communities. We urge the machine learning community to pay close attention to the direct and indirect impact of our work on society, especially on vulnerable groups. Awareness of historical antecedents, contextual, and political dimensions of current work is imperative is this regard. We hope this work contributes to raising awareness and adds to a continued discussion of ethics and justice in machine learning.

that they were intended to solve, when taken in isolation, stand on ethically dubious grounds themselves. In this regard, we strongly concur with previous work such as [85] that *gender classification* based on appearance of a person in a digital image is both **scientifically flawed** and is a technology that bears a high risk of systemic abuse.

[15] Obtained using *GloVe embeddings* [62] on the labels

[16] Listed in `df_hand_survey.csv`

[17] Link: https://rb.gy/zccdps

References

[1] Collaborative art tool for discovering images. https://ganbreeder.app, 2019. [Online; accessed 9-8-2019]. 5

[2] Face search • pimeyes. https://pimeyes.com/en/, May 2020. (Accessed on 05/04/2020). 4

[3] Md Zahangir Alom, Tarek M Taha, Christopher Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C Van Esesn, Abdul A S Awwal, and Vijayan K Asari. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018. 2

[4] Emily Bazelon. Nazi anatomy history: The origins of conservatives' anti-abortion claims that rape can't cause pregnancy. http://www.slate.com/articles/life/history/2013/11/nazi_anatomy_history_the_origins_of_conservatives_anti_abortion_claims_that.html, Nov 2013. (Accessed on 06/16/2020). 1

[5] Ruha Benjamin. *Race after technology: Abolitionist tools for the new jim code*. John Wiley & Sons, 2019. 5

[6] Lucas Beyer, Olivier J. Hénaff, Alexander Kolesnikov, Xiaohua Zhai, and Aäron van den Oord. Are we done with imagenet?, 2020. 3

[7] Abeba Birhane and Fred Cummins. Algorithmic injustices: Towards a relational ethics. *arXiv preprint arXiv:1912.07376*, 2019. 2, 5

[8] Colin Blain, Margaret Mackay, and Judith Tanner. Informed consent the global picture. *British Journal of Perioperative Nursing (United Kingdom)*, 12(11):402–407, 2002. 1

[9] Geoffrey C Bowker and Susan Leigh Star. *Sorting things out: Classification and its consequences*. MIT press, 2000. 2, 5

[10] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018. 4

[11] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018. 5

[12] Alexander L Brown, Jonathan Meer, and J Forrest Williams. Why do people volunteer? an experimental analysis of preferences for time donations. *Management Science*, 65(4):1455–1468, 2019. 7

[13] Claudia Bullerjahn, Katharina Heller, and Jan Hoffmann. How masculine is a flute? a replication study on gender stereotypes and preferences for musical instruments among young children. In *Proceedings of the 14th International Conference on Music Perception and Cognition*, pages 5–9, 2016. 8

[14] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91, 2018. 2

[15] Emma Carroll and Jessica Coates. The school girl, the billboard, and virgin: The virgin mobile case and the use of creative commons licensed photographs by commercial entities. *Knowledge policy for the 21st century. A legal perspective*, pages 181–204, 2011. 5

[16] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017. 4

[17] Creative Commons. Chang v. virgin mobile - creative commons. https://wiki.creativecommons.org/wiki/Chang_v._Virgin_Mobile, Jun 2013. (Accessed on 06/03/2020). 5

[18] Susan Corbett. Creative commons licences: A symptom or a cause? *Available at SSRN 2028726*, 2009. 5

[19] Susan Corbett. Creative commons licences, the copyright regime and the online community: Is there a fatal disconnect? *The Modern Law Review*, 74(4):503–531, 2011. 5

[20] Kate Crawford and Trevor Paglen. Excavating ai. https://www.excavating.ai/, Sep 2019. (Accessed on 04/30/2020). 3

[21] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1

[22] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, 2019. 6

[23] Jiankang Deng, Jia Guo, Zhou Yuxiang, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. In *arxiv*, 2019. 6

[24] Executive departments and agencies of the federal government of the United States. ecfr — code of federal regulations. https://www.ecfr.gov/cgi-bin/text-idx?SID=d387165baf23de2b80af8ea39e2addad&mc=true&node=se45.1.46_1116&rgn=div8, Jun 2020. (Accessed on 06/02/2020). 1

[25] Catherine D'Ignazio and Lauren F Klein. *Data feminism*. MIT Press, 2020. 2, 5

[26] Fabio Henrique Kiyoiti dos Santos Tanaka and Claus Aranha. Data augmentation using gans. *Proceedings of Machine Learning Research XXX*, 1:16, 2019. 7

[27] Chris Dulhanty and Alexander Wong. Auditing imagenet: Towards a model-driven framework for annotating demographic attributes of large-scale image datasets. *arXiv preprint arXiv:1905.01347*, 2019. 7

[28] Chris Dulhanty and Alexander Wong. Investigating the impact of inclusion in face recognition training data on individual face identification, 2020. 5

[29] Robert Eaglestone. One and the same? ethics, aesthetics, and truth. *Poetics Today*, 25(4):595–608, 2004. 5

[30] Editorial. Time to discuss consent in digital-data studies. https://www.nature.com/articles/d41586-019-02322-z, July 2019. (Accessed on 06/02/2020). 1

[31] Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018. 5

[32] Liyue Fan. Image pixelization with differential privacy. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 148–162. Springer, 2018. 7

[33] Luciano Floridi. Translating principles into practices of digital ethics: five risks of being unethical. *Philosophy & Technology*, 32(2):185–193, 2019. 8

[34] Bedapudi Praneeth Gant Laborde. Nsfw detection machine learning model. https://github.com/GantMan/nsfw_model, Jan 2019. (Accessed on 05/31/2020). 8

[35] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *arXiv preprint arXiv:1803.09010*, 2018. 7

[36] Hila Gonen and Yoav Goldberg. Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 609–614, 2019. 5

[37] Mary L Gray and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Eamon Dolan Books, 2019. 3

[38] Jia Guo and Jiankang Deng. deepinsight/insightface: Face analysis project on mxnet. https://github.com/deepinsight/insightface, May 2020. (Accessed on 05/31/2020). 6, 7, 8

[39] Herkko Hietanen. Creative commons olympics: How big media is learning to license from amateur authors. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, 2:50, 2011. 5

[40] Kashmir Hill. *The Secretive Company That Might End Privacy as We Know It*, 2020. 5

[41] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 2, 4

[42] Elizabeth C Hirschman. Consumers and their animal companions. *Journal of consumer research*, 20(4):616–632, 1994. 8

[43] Google Inc. Dataset search. https://datasetsearch.research.google.com/, Sep 2018. (Accessed on 06/17/2020). 8

[44] Khari Johnson. Aclu sues facial recognition startup clearview ai for privacy and safety violations | venturebeat, May 2020. (Accessed on 06/02/2020). 5

[45] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 4

[46] Maximilian Kasy and Rediet Abebe. Fairness, equality, and power in algorithmic decision making. Technical report, Working paper, 2020. 2

[47] Matthew Kay, Cynthia Matuszek, and Sean A Munson. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3819–3828, 2015. 2

[48] Os Keyes. The misgendering machines: Trans/hci implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–22, 2018. 2

[49] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 2

[50] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Tom Duerig, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *arXiv preprint arXiv:1811.00982*, 2018. 2, 4, 5

[51] Fei Fei Li and Jia Deng. Where have we been? where are we going? http://image-net.org/challenges/talks_2017/imagenet_ilsvrc2017_v1.0.pdf, Sep 2017. (Accessed on 05/01/2020). 1

[52] Clare McGlynn, Erika Rackley, and Ruth Houghton. Beyond revenge porn: The continuum of image-based sexual abuse. *Feminist Legal Studies*, 25(1):25–46, 2017. 4

[53] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models, 2020. 4

[54] Ryan Merkley. Use and fair use: Statement on shared images in facial recognition ai - creative commons, Mar 2019. (Accessed on 06/03/2020). 5

[55] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. *arXiv preprint arXiv:1810.03993*, 2018. 7

[56] S Naidoo. Informed consent for photography in dental practice: communication. *South African Dental Journal*, 64(9):404–406, 2009. 1

[57] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE, 2008. 1

[58] Safiya Umoja Noble. *Algorithms of oppression: How search engines reinforce racism*. nyu Press, 2018. 5

[59] Cathy O'neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016. 5

[60] Seymour A Papert. The summer vision project. *AIM-100*, 1966. 4, 8

[61] Xingchao Peng, Ben Usman, Neela Kaushik, Dequan Wang, Judy Hoffman, and Kate Saenko. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2021–2026, 2018. 7

[62] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014. 8

[63] PimEyes. Face search • pimeyes. https://pimeyes.com/en/, Jun 2020. (Accessed on 06/03/2020). 4

[64] Vinay Uday Prabhu, Sanghyun Han, Dian Ang Yap, Mihail Douhaniaris, Preethi Seshadri, and John Whaley. Fonts-2-handwriting: A seed-augment-train framework for universal digit classification. *arXiv preprint arXiv:1905.08633*, 2019. 7

[65] PyTorch. torchvision.datasets — pytorch 1.5.0 documentation. https://pytorch.org/docs/stable/torchvision/datasets.html, Jun 2020. (Accessed on 06/17/2020). 8

[66] Michael Ramirez. "my dog's just like me": Dog ownership as a gender display. *Symbolic Interaction*, 29(3):373–391, 2006. 8

[67] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? *arXiv preprint arXiv:1902.10811*, 2019. 7

[68] Luc Rocher, Julien M Hendrickx, and Yves-Alexandre

De Montjoye. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10(1):1–9, 2019. 1

[69] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018. 6, 7, 8

[70] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015. 2

[71] Anantha Natarajan S. Imagescraper · pypi. https://pypi.org/project/ImageScraper/, May 2015. (Accessed on 06/17/2020). 8

[72] Anubhav Sachan. bingscraper · pypi. https://pypi.org/project/bingscraper/, July 2018. (Accessed on 06/17/2020). 8

[73] Shreya Shankar, Yoni Halpern, Eric Breck, James Atwood, Jimbo Wilson, and D Sculley. No classification without representation: Assessing geodiversity issues in open data sets for the developing world. *arXiv preprint arXiv:1711.08536*, 2017. 7

[74] Congzheng Song, Thomas Ristenpart, and Vitaly Shmatikov. Machine learning models that remember too much. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 587–601, 2017. 5

[75] Susan Leigh Star and Geoffrey C Bowker. Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. *Ethics and Information Technology*, 9(4):273–280, 2007. 2, 5

[76] Lucy Suchman. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press, 2007. 5

[77] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, pages 843–852, 2017. 4

[78] TensorFlow. Tensorflow datasets. https://www.tensorflow.org/datasets, Jun 2020. (Accessed on 06/17/2020). 8

[79] Antonio Torralba, Rob Fergus, and William T Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 30(11):1958–1970, 2008. 1, 2, 3

[80] Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Andrew Ilyas, and Aleksander Madry. From imagenet to image classification: Contextualizing progress on benchmarks, 2020. 3

[81] Amol Umrale. imagebot · pypi. https://pypi.org/project/imagebot/, July 2015. (Accessed on 06/17/2020). 8

[82] Haohan Wang, Songwei Ge, Eric P. Xing, and Zachary C. Lipton. Learning robust global representations by penalizing local predictive power, 2019. 7

[83] Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A Efros. Dataset distillation. *arXiv preprint arXiv:1811.10959*, 2018. 7

[84] Paul Weindling. The origins of informed consent: the international scientific commission on medical war crimes, and the nuremberg code. *Bulletin of the History of Medicine*, pages 37–71, 2001. 1

[85] Sarah Myers West, Meredith Whittaker, and Kate Crawford. Discriminating systems. https://ainowinstitute.org/discriminatingsystems.html, 2019. 8

[86] Wikipedia. Streisand effect - wikipedia. https://en.wikipedia.org/wiki/Streisand_effect, April 2020. (Accessed on 04/29/2020). 4

[87] Benjamin Wilson, Judy Hoffman, and Jamie Morgenstern. Predictive inequity in object detection. *arXiv preprint arXiv:1902.11097*, 2019. 2

[88] Elizabeth R Wrape, Alexandra L Dittloff, and Jennifer L Callahan. Gender and musical instrument stereotypes in middle school children: Have trends changed? *Update: Applications of Research in Music Education*, 34(3):40–47, 2016. 8

[89] Baoyuan Wu, Weidong Chen, Yanbo Fan, Yong Zhang, Jinlong Hou, Jie Liu, and Tong Zhang. Tencent ml-images: A large-scale multi-label image database for visual representation learning. *IEEE Access*, 7:172683–172693, 2019. 2

[90] Blaise Aguera y Arcas, Margaret Mitchell, and Alexander Todorov. Physiognomy's new clothes. *Medium (6 May 2017), online:< https://medium. com/@ blaisea/physiognomys-new-clothesf2d4b59fdd6a*, 2017. 3

[91] Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, and Olga Russakovsky. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 547–558, 2020. 3

[92] Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, and Olga Russakovsky. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 547–558, 2020. 5, 7

[93] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(6):1452–1464, 2017. 2