A Diverse and Interpretable Benchmark for Viti- and Vini-cultural Visual Understanding

Shengli Hu

Dataminr Inc.
sh2264@cornell.edu

Abstract

We present four new datasets for viticultural and vinicultural visual understanding: iVineyard, iCellar, iGrapevine, and VinePathology. We designed, gathered data for, cleaned, and provided numerical and natural language annotations for these datasets in collaboration with domain experts with the aim of (1) accelerating AI adoption in the realms of viticulture and oenology; (2) improving data efficiency and interpretability with data collection, task formulation, and annotation processes informed by domain expertise; (3) benchmarking the performance of representation learning algorithms on a suite of challenging downstream viti- and vinicultural tasks that go beyond standard species classification. We provide analyses of qualitative and quantitative results of downstream tasks including fine-grained visual categorization, fine-grained image retrieval, image geo-localization, and object discovery, thus shedding light on the strengths and weaknesses of feature representations across a diverse set of tasks that are of scientific importance to viticulturists and oenologists.

Fine-grained image classification as a sub-field of computer vision has enjoyed tremendous growth in the past decade, from the proliferation of datasets of various domains, to methodological advancement that enables greater generalizability and flexibility. Large-scale fine-grained classification datasets of dogs, cars, birds, natural species, etc. have been created from media repositories and become standard experimental resources for computer vision researchers to benchmark the progress of classification models over time. While it might appear impressive to enable the computers to identify particular cultivar of grape variety or plant pathology present in an image, recent studies have reminded us of the fact that when working with domain experts or scientists, correct identification of species is not terribly impressive or informative since most likely it's something they already know of, and such AI classification algorithms could help speed up their work at best. What experts are more interested in are downstream tasks. There are far more questions that domain experts would like to ask of these large media repositories in addition to "What species is in this photo?" For instance, given an natural image of a grapevine, besides "what grape variety is in this photo?", a viticulturist may be more interested in knowing, "How old is this grapevine?" or "Is this vine under water stress?" Similarly, an oenologist may want to know, "Is this grape cluster ready for harvest? What is the likely sugar level in this cluster?" Domain experts can certainly answer these questions themselves for a few images. The problem lies in scalability when it come to the ever-growing repositories in face of the limited time and attention of domain experts.

Therefore, we designed four fine-grained image analysis datasets with an overarching ontology and natural language explanations for viticulture and viniculture/oenology in collaboration with domain experts with the aim of benchmarking the performance of representation learning algorithms on multi-task learning of a suite of challenging tasks of practical relevance that go beyond standard fine-grained species classification.

Viticulture and viniculture, intertwined with the $\$417.85$ billion wine industry, represent one of the most ancient and fundamental agricultural product at a global scale. Vines and wines (the final product) constitute a fascinating part of human history, touching on various other disciplines such as geology, geography, chemistry, biology, economics, politics, etc. They allure mere mortals with fragrance and exuberance, inspire the imagination of artists and poets, and attract the powerful and wealthy with esteemed prestige and exclusivity. They also stage a diverse set of interesting challenges from a computer vision perspective: long-tailed data distribution with various zero-shot or low-shot opportunities, occlusions, multi-view and multi-scale inter-class variations, etc.

To summarize our main contributions and provide a brief preview of results:

- We introduce diverse and interpretable benchmarks for viticultural and vinicultural visual understanding informed by domain experts. It includes four curated fine-grained datasets covering various aspects of viticulture and viniculture. These datasets were annotated with an overarching ontology and natural language explanations that embody domain knowledge.
- We motivate and introduce a unique set of visual understanding tasks and provide baseline experiments with qualitative ad quantitative results including fine-grained visual categorization, fine-grained image retrieval, image geo-localization, and object discovery.

Figure 1: Photo-mosaic collages of three fine-grained datasets (from top to bottom, left to right): VinePathology, iGrapevine, iVineyard, iCellar.

Related Work

Fine-grained visual categorization (FGVC) flourished as a sub-field of computer vision in the last decade with a diverse range of FGVC datasets covering domains such as dogs (Khosla et al. 2011), birds (Wah et al. 2011; Berg et al. 2014), flowers (Nilsback and Zisserman 2008), cars (Yang et al. 2015), food (Kaur et al. 2019), retail products (Wei et al. 2019), fashion (Zou et al. 2019; Jia et al. 2019), apples (Thapa et al. 2020), castles (Anderson et al. 2021), natural species (Van Horn et al. 2021), and artworks (Zhang et al. 2019; Conde and Turgutlu 2021), among others. Distinct from most existing FGVC datasets, our sets of viticultural and vinicultural tasks take on a wide range of tasks going beyond species classification for broader applications grounded in real-world use cases, informed by domain experts. More importantly, we included with natural language explanations for justification of resulting annotations, to boost data and labeling efficiency exemplified in ALICE (Liang, Zou, and Yu 2020).

Grapevines are essentially agricultural products, and to improve agriculture is to improve the food supply chain that impacts each and every living being in the world. Despite the tremendous potential impact AI could unlock in this realm, AI for agriculture has been one of the topics with the least amount of research work, according to a recent survey of AI for Social Good (Shi, Wang, and Fang 2020). Reasons cited for such relative literature sparsity include lack of established data collection pipelines or frameworks in agriculture, difficulty in data collection processes, etc., which exacerbated agricultural data mining and direct application of AI techniques. Existing studies largely revolve around growing activities (such as crop disease diagnosis (Quinn, Leyton-Brown, and Mwebaze 2011), yield prediction (You et al. 2017), and crop planning (Von Lücken and Brunelli 2008)), environmental factors (such as drought prediction (Kersting et al. 2012)), and agricultural markets (such as price forecasting (Ma et al. 2019) and market design (Newman et al. 2018)).

For two out of the four viti- and vini-cultural datasets, the tasks directly associated with them could be framed as image geolocalization problems. Image retrieval-based methods for geolocalization have been explored extensively where a query image is matched to the most similar reference images in a large image database. Various low-level image features such as BOW (Schindler, Brown, and Szeliski 2007; Gronat et al. 2013), color or texton histograms (Hays and Efros 2008; Kalogerakis et al. 2009; Hays and Efros 2015), descriptors like SIFT, SURF (Hakeem et al. 2006; Zamir and Shah 2010; Gronat et al. 2013; Arandjelović and Zisserman 2014), point and line features (Ramalingam, Bouaziz, and Sturm 2011; Li, Morariu, and Davis 2014), building patterns (Torii et al. 2013), keypoints (Chen et al. 2011), GIST (Zamir and Shah 2014; Zemene et al. 2018), etc., have been exploited to perform matching, so were feature and geometric correspondence (Li, Snavely, and Huttenlocher 2010; Bansal and Daniilidis 2014; Gopalan 2015; Zemene et al. 2018), representation learning (Gopalan 2015; Zemene et al. 2018; Liu, Li, and Dai 2019), segmentation (Ramalingam et al. 2010; Baatz et al. 2012), discriminative learning (Cao and Snavely 2013), feature voting (Liu et al. 2020), feature reweighting (Kim, Dunn, and Frahm 2017), pose estimation (Ramalingam, Bouaziz, and Sturm 2011), etc. Classification-based image geolocalization provides a more memory and disk efficient alternative to retrieval-based so-

lutions, by treating the task as a classification problem that divides the map into multiple discrete classes. Gronat et al. (2013) leverages geotags to train classifiers per location. Hongsuck Seo et al. (2018) extends (Weyand, Kostrikov, and Philbin 2016) by enhancing the resolution of geoclasses into which convolutional neural networks classify with combinatorial partitioning. We provide baseline results based on SOTA methods with respect to image retrieval, geolocalization, and object discovery.

Datasets

Table 1 summarizes the four datasets we collected in terms of the number of images and classes, and Figure 1 showcases some photo-mosaic collages (e.g., an image of a vineyard/cellar/vine made up by making a collage of various vineyard/cellar/vine images) based on these datasets. The datasets will be released online[1].

Dataset	Class	# Classes	Images
iCellar	Winery cellars	773	108,157
iVineyard	Vineyards	327	41,166
iGrapevine	Grapevines	153	13,550
VinePathology	Vine diseases, pests, weeds, other hazards	42	5,894

Table 1: Datasets for fine-grained viti- and vini-cultural visual understanding.

Tasks

Each of the four datasets — iCellar, iVineyard, iGrapevine, and VinePathology — centers around one central task in vein of species classification, for which we obtained groundtruth labels. For instance, the basic task of iCellar is fine-grained classification of winery cellars around the globe, whereas that of iVineyard is multi-class vineyard classification. Besides respective central tasks, each dataset is also associated with a series of tasks which are identified as possible rationales of the category of the basic task, being of greater interest to domain experts. For instance, in order to correctly identify which vineyard is depicted in an image from iVineyard dataset, visual clues with respect to the type of landscape, trellising system, vineyard quality, soil type, etc. could help inform the identity of the vineyard in question. We summarize in bullet points below for each datasets the correspondingly proposed tasks informed by domain experts. For example, given an image of a grapevine in terms of leaves or clusters, besides training a fine-grained visual classification algorithm to recognize the grape variety, viticulturists are more interested in automatic assessment of vine age, vine health, potential pathology, level of vine stress, and other environmental information that could be derived from the grapevine image. We include the number of classes in the parentheses following each task.

- iCellar — Winery/cellar identification (773):
  - Vessel classification (19): barrel age and type: Barrique, Pièce, Fuder, Stück, Puncheon, Hogshead, Tonneau, Botte; concrete egg, cement vats, stainless steel tanks, etc.;
  - Scale of production prediction (3): small, medium, large.
- iVineyard — vineyard identification (327):
  - Vineyard quality classification (4): none, village level, premier cru, grand cru;
  - Macro-, meso-, and micro-climate classification (9);
  - Trellising and training system classification (26) :Gobelet, Vertical Shoot Positioning, Geneva Double Curtain, Lyre, etc.;
  - Landscape classification (15): steep slope, plateau, flatland, mountain floor, riverbank, fluvial fan, lakeside, etc.;
  - Context identification (8): conditions such as drought, rain, hail, snow, etc.;
  - Soil identification (21): limestone, shale, slate, gravel, granite, schist, clay, chalk, clay, loam, sandstone, etc.
  - Soil condition classification (4): water stress, nutrient deficiency, irrigated, dry farm, etc.
- iGrapevine — grape variety identification (153):
  - ripeness identification (4): phenolic ripeness and physiological ripeness prediction.
- VinePathology (42):
  - Disease identification (24): bacterial, viral, fungal, etc.
  - Pest identification (14): leafroller, phylloxera, nematode, lanternfly, etc.
  - Weed identification (4).

Data Collection, Preprocessing, and Annotation

We collected our datasets by gathering images from major search engines and social media by varying fine-grained search queries with query expansion. Due to the noisy nature of resulting images, we trained an image classifier to filter out unwanted images. By analyzing the composition of images of the initial crawl, we identified 9 major categories of images: images of vineyards/cellars/vines/grapes in a natural scene, images of wine bottles, images of wine labels, marketing images, graphs, images of people, maps, tasting events, winemaking processes. We trained a ResNet-50 (He et al. 2016) based multi-label image type classifier to distinguish between these classes with 3,000 images cleaned for each class. With a resulting F1 at 0.95 (precision at 0.97, recall at 0.92), we applied the classifier to our initial noisy dataset to filter out images of bottles, labels, people, maps, tasting events, winemaking processes, graphs, and stock images. After further deduplicating efforts with tools including *digikam*, *findimagedupes*, *geeqie*, *fdupes*, the remaining images add up to what Table 1 shows.

For each image in four datasets, groundtruth labels for respective central task were obtained by way of the data collection and cleaning process. Labels of auxiliary tasks such

[1] https://github.com/ai-somm/viti-vini-culture.

(a) Canary Islands (b) Finger Lakes (c) Gevrey-Chambertin (d) Santorini

Figure 2: Distinctive visual patches of vineyards in Canary Islands, Finger Lakes, Gevrey-Chambertin, and Santorini.

as vessel classification or trellising classification were provided by domain experts when applicable in the following way: we applied the respective classifier trained for the associated central task to all images in the dataset and identified hard examples with cross high entropy levels to elicit natural language explanations from domain experts. For instance, for an image of a vineyard in Lodi (in central California) misclassified as a vineyard in Yecla (in central Spain), a domain expert wrote: "The Zinfandel vines with signs of leafrolls are indicative of Lodi rather than Yecla, despite both being known for gnarly old vines on barren flatlands. Lodi is more humid, at lower elevation, with sandier soils than Yecla." In total, we collected natural language explanations for $10,635$ images, which were parsed to result in $35,150$ labels for auxiliary tasks.

Baseline experiments

We run experiments to provide baseline performances on three visual understanding tasks using at least one of the datasets proposed here. We train ResNet-50 (He et al. 2016) for two related classification tasks and one task related to clustering: (1) vineyard/cellar/grapevine/pathology recognition, 773-/327-/153-/42-class classification respectively; (2) vineyard/cellar geolocalization as 18-class country classification; and (3) vineyard object discovery where the goal is to surface distinctive patches that uniquely identify each vineyard. We also train models for image retrieval where given a query image, the goal is to provide a ranking of images in terms of visual similarity. In all of our experiments we use a 80/10/10 train/val/test split, ResNet-50 (He et al. 2016) as the backbone pretrained on ImageNet, Dropout rate at 0.1-0.5, learning rate at 0-10^{-3} with a loglinear warm up during initial 5-10 epochs and at 10^{-5}-10^{-3}.

Results

We tabulate the quantitative results in Table 2 for recognition tasks (mean top-K accuracy) and retrieval tasks.

Qualitative results of object discovery are visualized in Figure 2. Interestingly, mountain skylines, bush vines, craters, and dark volcanic ash of Canary Islands separate itself apart from Santorini that also boosts volcanic soil but reflects differently vine training system — short stone walls around bush vines ("kouloura") and the iconic white houses in the distance. Finger Lakes region, on the other hand, showcases lusher and upright vines with presence of water bodies, whereas Gevrey-Chambertin is recognized by ancient stone walls ("clos"), narrowly-winding country roads, and unique local architecture.

Task/ACC	Top-1	Top-5	Task	Top-1	Top-5
Vineyard	41.5	53.6	V-Country	62.2	73.1
Cellar	28.7	39.2	C-Country	38.6	42.3
Grapevine	34.6	44.2	V-Retrieval	11.2	19.4
Pathology	62.7	79.3	C-Retrieval	6.7	12.8

Table 2: Top-1 and Top-5 accuracies (ACC) for vineyard/cellar/grapevine/pathology classification, vineyard/cellar country prediction, and vineyard/cellar retrieval. V-/C- refers to vineyard-/cellar-.

Conclusion and Future Research

We present a diverse set of images for viti- and vini-cultural visual understanding with interpretable explanations. We present baseline experimental results of fine-grained image classification, image retrieval, and object discovery. The diversity and interpretability of these datasets could unlock a variety of further explorations. For instance, multi-task and multimodal active learning methods could potential boost training efficient while keeping annotation costs low, especially when domain experts are involved. Due to the long-tailed distribution of such fine-grained datasets, few-shot, zero-shot, and generalized zero-shot learning methods based on meta learning could accelerate learning new tasks, an important and practical factor for when new tasks or domains surface as the scientific fields such as viticulture evolve. Lastly, even though image features produced by standard supervised methods still largely outperform those produced by self-supervised approaches, there exists evidences (Van Horn et al. 2021) that for particular tasks such as scene classification, self-supervised learning showcases promising potential. Therefore, future works benchmarking self-supervised learning on these viti- and vini-cultural datasets might prove fruitful.

References

Anderson, C.; Teuscher, A.; Anderson, E.; Larsen, A.; Shirley, J.; and Farrell, R. 2021. Have Fun Storming the Castle(s)! In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3703–3712.

Arandjelović, R.; and Zisserman, A. 2014. DisLocation: Scalable descriptor distinctiveness for location recognition. In *Asian Conference on Computer Vision*, 188–204. Springer.

Baatz, G.; Saurer, O.; Köser, K.; and Pollefeys, M. 2012. Large scale visual geo-localization of images in mountainous terrain. In *European conference on computer vision*, 517–530. Springer.

Bansal, M.; and Daniilidis, K. 2014. Geometric urban geolocalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3978–3985.

Berg, T.; Liu, J.; Woo Lee, S.; Alexander, M. L.; Jacobs, D. W.; and Belhumeur, P. N. 2014. Birdsnap: Large-scale fine-grained visual categorization of birds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011–2018.

Cao, S.; and Snavely, N. 2013. Graph-based discriminative learning for location recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition*, 700–707.

Chen, D. M.; Baatz, G.; Köser, K.; Tsai, S. S.; Vedantham, R.; Pylvänäinen, T.; Roimela, K.; Chen, X.; Bach, J.; Pollefeys, M.; et al. 2011. City-scale landmark identification on mobile devices. In *CVPR 2011*, 737–744. IEEE.

Conde, M. V.; and Turgutlu, K. 2021. CLIP-Art: Contrastive Pre-Training for Fine-Grained Art Classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3956–3960.

Gopalan, R. 2015. Hierarchical sparse coding with geometric prior for visual geo-location. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2432–2439.

Gronat, P.; Obozinski, G.; Sivic, J.; and Pajdla, T. 2013. Learning and calibrating per-location classifiers for visual place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 907–914.

Hakeem, A.; Vezzani, R.; Shah, M.; and Cucchiara, R. 2006. Estimating geospatial trajectory of a moving camera. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 2, 82–87. IEEE.

Hays, J.; and Efros, A. A. 2008. IM2GPS: estimating geographic information from a single image. In *2008 ieee conference on computer vision and pattern recognition*, 1–8. IEEE.

Hays, J.; and Efros, A. A. 2015. Large-scale image geolocalization. In *Multimodal location estimation of videos and images*, 41–62. Springer.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hongsuck Seo, P.; Weyand, T.; Sim, J.; and Han, B. 2018. Cplanet: Enhancing image geolocalization by combinatorial partitioning of maps. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 536–551.

Jia, M.; Shi, M.; Sirotenko, M.; Cui, Y.; Hariharan, B.; Cardie, C.; and Belongie, S. 2019. The fashionpedia ontology and fashion segmentation dataset. *Cornell University*.

Kalogerakis, E.; Vesselova, O.; Hays, J.; Efros, A. A.; and Hertzmann, A. 2009. Image sequence geolocation with human travel priors. In *2009 IEEE 12th international conference on computer vision*, 253–260. IEEE.

Kaur, P.; ; Sikka, K.; Wang, W.; Belongie, s.; and Divakaran, A. 2019. FoodX-251: A Dataset for Fine-grained Food Classification. *arXiv preprint arXiv:1907.06167*.

Kersting, K.; Xu, Z.; Wahabzada, M.; Bauckhage, C.; Thurau, C.; Roemer, C.; Ballvora, A.; Rascher, U.; Leon, J.; and Pluemer, L. 2012. Pre-symptomatic prediction of plant drought stress using dirichlet-aggregation regression on hyperspectral images. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Khosla, A.; Jayadevaprakash, N.; Yao, B.; and Li, F.-F. 2011. Novel dataset for fine-grained image categorization: Stanford dogs. In *Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC)*, volume 2. Citeseer.

Kim, H. J.; Dunn, E.; and Frahm, J.-M. 2017. Learned contextual feature reweighting for image geo-localization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3251–3260. IEEE.

Li, A.; Morariu, V. I.; and Davis, L. S. 2014. Planar structure matching under projective uncertainty for geolocation. In *European Conference on Computer Vision*, 265–280. Springer.

Li, Y.; Snavely, N.; and Huttenlocher, D. P. 2010. Location recognition using prioritized feature matching. In *European conference on computer vision*, 791–804. Springer.

Liang, W.; Zou, J.; and Yu, Z. 2020. ALICE: Active Learning with Contrastive Natural Language Explanations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4380–4391.

Liu, D.; Cui, Y.; Guo, X.; Ding, W.; Yang, B.; and Chen, Y. 2020. Visual Localization for Autonomous Driving: Mapping the Accurate Location in the City Maze. *arXiv preprint arXiv:2008.05678*.

Liu, L.; Li, H.; and Dai, Y. 2019. Stochastic Attraction-Repulsion Embedding for Large Scale Image Localization. In *Proceedings of the IEEE International Conference on Computer Vision*, 2570–2579.

Ma, W.; Nowocin, K.; Marathe, N.; and Chen, G. H. 2019. An interpretable produce price forecasting system for small and marginal farmers in india using collaborative filtering and adaptive nearest neighbors. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development*, 1–11.

Newman, N.; Bergquist, L. F.; Immorlica, N.; Leyton-Brown, K.; Lucier, B.; McIntosh, C.; Quinn, J.; and Ssekibuule, R. 2018. Designing and evolving an electronic agricultural marketplace in Uganda. In *Proceedings of the 1st*

ACM SIGCAS Conference on Computing and Sustainable Societies, 1–11.

Nilsback, M.-E.; and Zisserman, A. 2008. Automated flower classification over a large number of classes. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 722–729. IEEE.

Quinn, J.; Leyton-Brown, K.; and Mwebaze, E. 2011. Modeling and monitoring crop disease in developing countries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25.

Ramalingam, S.; Bouaziz, S.; and Sturm, P. 2011. Pose estimation using both points and lines for geo-localization. In *2011 IEEE International Conference on Robotics and Automation*, 4716–4723. IEEE.

Ramalingam, S.; Bouaziz, S.; Sturm, P.; and Brand, M. 2010. Skyline2gps: Localization in urban canyons using omni-skylines. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3816–3823. IEEE.

Schindler, G.; Brown, M.; and Szeliski, R. 2007. City-scale location recognition. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–7. IEEE.

Shi, Z. R.; Wang, C.; and Fang, F. 2020. Artificial intelligence for social good: A survey. *arXiv preprint arXiv:2001.01818*.

Thapa, R.; Zhang, K.; Snavely, N.; Belongie, S.; and Khan, A. 2020. The Plant Pathology Challenge 2020 data set to classify foliar disease of apples. *Applications in Plant Sciences*, 8(9): e11390.

Torii, A.; Sivic, J.; Pajdla, T.; and Okutomi, M. 2013. Visual place recognition with repetitive structures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 883–890.

Van Horn, G.; Cole, E.; Beery, S.; Wilber, K.; Belongie, S.; and Mac Aodha, O. 2021. Benchmarking Representation Learning for Natural World Image Collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12884–12893.

Von Lücken, C.; and Brunelli, R. 2008. Crops Selection for Optimal Soil Planning using Multiobjective Evolutionary Algorithms. In *AAAI*, 1751–1756.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The caltech-ucsd birds-200-2011 dataset.

Wei, X.-S.; Cui, Q.; Yang, L.; Wang, P.; and Liu, L. 2019. RPC: A Large-Scale Retail Product Checkout Dataset. *arXiv e-prints*, arXiv–1901.

Weyand, T.; Kostrikov, I.; and Philbin, J. 2016. Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision*, 37–55. Springer.

Yang, L.; Luo, P.; Change Loy, C.; and Tang, X. 2015. A large-scale car dataset for fine-grained categorization and verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3973–3981.

You, J.; Li, X.; Low, M.; Lobell, D.; and Ermon, S. 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In *Thirty-First AAAI conference on artificial intelligence*.

Zamir, A. R.; and Shah, M. 2010. Accurate image localization based on google maps street view. In *European Conference on Computer Vision*, 255–268. Springer.

Zamir, A. R.; and Shah, M. 2014. Image geo-localization based on multiplenearest neighbor feature matching using generalized graphs. *IEEE transactions on pattern analysis and machine intelligence*, 36(8): 1546–1558.

Zemene, E.; Tesfaye, Y. T.; Idrees, H.; Prati, A.; Pelillo, M.; and Shah, M. 2018. Large-scale image geo-localization using dominant sets. *IEEE transactions on pattern analysis and machine intelligence*, 41(1): 148–161.

Zhang, C.; Kaeser-Chen, C.; Vesom, G.; Choi, J.; Kessler, M.; and Belongie, S. 2019. The iMet collection 2019 challenge dataset. *arXiv preprint arXiv:1906.00901*.

Zou, X.; Kong, X.; Wong, W.; Wang, C.; Liu, Y.; and Cao, Y. 2019. Fashionai: A hierarchical dataset for fashion understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.