A Drone Video Clip Dataset and its Applications in Automated Cinematography

Amirsaman Ashtari [iD], Raehyuk Jung [iD], Mingxiao Li [iD], and Junyong Noh [iD]

✦

AUTHOR STATEMENT

We release our dataset and codes used for our experiments and data collection at our dataset Google Drive link. Please contact the authors to obtain the dataset link and fill out the agreement form to use the data. See Section 6 for details. We prepared an instruction to use each of our codes in its related folder as a "read me" text file. In general, video cut detection evaluation is based on existing methods (ffprobe[1], PySceneDetect[2], Hecate[3], and ORB-SLAM2[4]), and their codes with clear instructions are available to the public. We used existing methods such as DenseNet121 [3], R(2+1)D [4], and Places365-CNN [5] for video classification. All of these codes are also available to the public with clear instructions. We provide our code for automatic drone video collection filtering with detailed instructions and comments. Using this code, one can re-produce all results reported in Section 6 (Figure 9 to 14) of our paper.

The dataset annotations is licensed under the CC BY-NC-SA 4.0[5] (i.e., an annotation-only license). See Section 6 for details. Our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset. For example, YouTube-8M dataset[6] also shares YouTube video links. In addition, video link sharing is provided in these websites as a built-in feature through the video link sharing button, which means that one can easily share the link of a video in these video platforms, and the individuals who uploaded their videos to these websites are aware that their video can be seen and might be shared using these video sharing platforms.

DATASHEETS OF OUR DATASET

For our DVCD18K dataset documentation, we used the popular documentation framework "Datasheets for Datasets" [2].

- *A. Ashtari, R. Jung, and J. Noh (corresponding author) are with the KAIST, Daejeon 34141, South Korea. E-mail: {a.s.ashtari, raeomi93, junyongnoh}@kaist.ac.kr.*
- *M. Li is with the David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada. E-mail: m398li@uwaterloo.ca. M. Li contributed to this work during her internship at KAIST.*

1 MOTIVATION

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description. Drones became popular video capturing tools. Drone videos in the wild are first captured and then edited by humans to contain aesthetically pleasing camera motions and scene. Therefore, edited drone videos have extremely useful information for cinematography and for applications such as camera path planning to capture aesthetically pleasing shots. To design intelligent camera path planners, learning drone camera motions from these edited videos is essential. However, first, this requires to filter drone clips and extract their camera motions out of these edited videos that commonly contain both drone and non-drone content. Moreover, existing video search engines return the whole edited video as a semantic search result and cannot return only drone clips inside an edited video. To address this problem, we proposed the first approach that can automatically retrieve drone clips from an unlabeled video collection using high-level search queries, such as "drone clips captured outdoor in daytime from rural places". The retrieved clips also contain camera motions, camera view, and 3D reconstruction of a scene that can help develop intelligent camera path planners. To train our approach, we needed numerous examples of edited drone videos. To this end, we introduced the first large-scale dataset composed of edited drone videos. This dataset is also used for training and validating our drone video filtering algorithm.

2 COMPOSITION

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description. Our dataset consists of 991 source videos, and each of them is composed of a number of clips (in total over 18000 video clips). We manually identify each clip of all the videos of our dataset and annotate each of them with various levels of information. In the first step, for each clip, we provide manual ground truth annotations on video editing (e.g., cut time and transition), as well as whether it is captured by a drone (drone[7]/non-

---

7. drone clips without any video editing effects that affect the interpretation of the camera paths (i.e., fast forward and slow motion are allowed).

TABLE 1
Percentage of each class for drone video clips.

Type	Class	Percentage (%)
Time: day/night/between	day	72.80
	night	3.47
	between	23.73
Scene type: urban/rural/mixed	urban	23.43
	rural	63.13
	mixed	13.44
Location: indoor/outdoor	indoor	0.60
	outdoor	99.40
Logo: yes/no	logo	28.17
	non-logo	71.83
Text: yes/no	text	9.91
	non-text	90.09

drone[8]/droneE[9]/strobe[10]). In the second step, we manually annotate each drone clip with time information (day/night/between), scene type (rural/urban/mixed), shot location (indoor/outdoor), logo, and text presence. We also provide automatic annotations: camera path by SLAM, 3D scene reconstruction, and social platform metadata (number of likes, comments, and views).

How many instances are there in total (of each type, if appropriate)? Our DVCD18K dataset contains a total of 18,551 drone video clips issued from 991 videos, and the total duration is approximately 44 hours. 85% duration of the clips belong to the drone category, 3% to droneE, 10% to nondrone and 2% strobe. Table 1 shows the percentage of instances of our manual annotations for the drone category. In Section 3.4 of the paper, we also provide some statistics of our dataset as follow:

- distribution of the video duration,
- distribution of the number of clips per video with respect to video duration,
- distribution of the duration of all the identified drone and non-drone video clips,
- distribution of the number of views, likes, and comments.

Among 991 videos in our dataset, we also report which percentage of videos are captured in different categories as follows:

- 26.94% videos captured in different cities such as New York, Los Angeles, Milan, etc.,
- 63.17% videos captured in different countries such as Australia, India, France, etc.,
- 13.02% videos captured in different regions such as Mediterranean Islands, Alps Mountain, etc.,
- 9.59% videos captured in different nature scenes such as iceberg, rainforest, etc.,
- 8.78% videos are captured in different landmarks such as shipwrecks, Unesco cultural sites, stadiums, etc.,
- 1.1% videos captured from different transportation vehicles such as motorcycles, sailboats, balloons, etc. ,
- 15.24% videos captured as the First Person View (FPV) drone shots.

8. clips not captured by drones (e.g., hand-held videos, videos of people talking about drones, animations, and drawings).

9. drone clips with editing effects that affect the interpretation of the camera paths (e.g., rewinding and video shakes).

10. series of extremely short videos/images.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable). While our dataset covers a great diversity in terms of location, appearance, camera motion, scene category, quality, and shot type, the dataset is a sample of *edited* drone videos, captured and edited by human. The dataset might not be representative of all *edited* drone videos, as the distribution of latter is highly diverse and not known exactly. Instead, the motivation was to resemble a variety of *edited* drone videos. In addition, when we trained our models for five tasks (drone/non-drone, logo presence, time information, scene type, and location classification) using our dataset, the trained models were generalized well on totally unseen top 100 Youtube videos with the search query "flying drone". Our drone video filtering algorithm gained high precision for unseen Youtube videos (see Section 6.3 in our paper). This suggests that our dataset can generalize well on unseen *edited* drone videos considering its variety and large scale.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description. Instances contain the following components:

- source video links (specified in a CSV file),
- manual annotations of the source videos such as video cut time and transition, clip category, time information, scene type, location, and text/logo presence (as a CSV file)
- number of views, likes, and comments (as a CSV file)
- drone camera paths extracted via ORB-SLAM2 (as an TXT file),
- drone camera paths extracted via PhotoScan (as an XML file),
- 3D reconstruction of scenes via PhotoScan, including both 3D point clouds and textured meshes (specified in PLY files). In addition, we stored PSZ[11] files of 3D reconstructions.

Is there a label or target associated with each instance? If so, please provide a description. In the first step, for each video clip, we provide manual ground truth annotations on video editing such as cut time and transition, as well as whether it is captured by a drone (drone/non-drone/droneE/strobe). In the second step, we manually annotate each drone clip with time information (day/night/between), scene type (rural/urban/mixed), shot location (indoor/outdoor), logo, and text presence. We also provide automatic annotations such as camera path by SLAM, 3D scene reconstruction, and social platform metadata.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text. No, all of the relevant information has been provided.

11. One can load a PSZ file in PhotoScan to export information in other formats or further proceed to obtain a better 3D reconstruction result.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? *If so, please describe how these relationships are made explicit.* Individual instances (i.e., *edited* drone videos) are independent of each other. Each instance is labeled with its associated annotations based on the video content. Using our annotations, our drone video clips can be explicitly categorized in various groups based on their clip category (drone/non-drone), time information (day/night/between), scene type (rural/urban/mixed), shot location (indoor/outdoor), logo, and text presence.

Are there recommended data splits (e.g., training, development/validation, testing)? *If so, please provide a description of these splits, explaining the rationale behind them.* Among the 991 source videos, we randomly allocate 693 videos for training, 149 videos for validation, and 149 videos for testing. The duration of the training, validation, and testing sets is 31.27 hours (70.7% of the whole dataset), 6.94 hours (15.7%), and 6.04 hours (13.6%), respectively. Because of the big size of our dataset (44 hours) and great variety of the video clips, we did not recommend data splits because it might be hard to develop a logical evaluation criteria for data split recommendation.

Are there any errors, sources of noise, or redundancies in the dataset? *If so, please provide a description.* Through the agreement of multiple independent human annotators, we verified all the manual annotation results to avoid (or at least reduce) errors in the reported labels such as video editing (cut time and transition), time information (day/night/between), scene type (rural/urban/mixed), shot location (indoor/outdoor), logo, and text presence. See annotation validation in Section 3 of our paper for more details.

For the automatic annotations, the quality of the annotation is limited to the performance of ORB-SLAM2 and PhotoScan methods to extract drone camera paths and make 3D reconstructions of scenes, respectively. Sometimes, ORB-SLAM2 (or PhotoScan) might fail to track drone camera paths (or make 3D reconstructions) due to occlusion, aggressive camera motion, or poor visual correspondences between consecutive video frames.

We use the default camera calibration file of ORB-SLAM2 to extract camera paths because predicting camera intrinsic parameters from videos in the wild is still a progressing research field, and existing methods are still exposed to some degree of errors [1]. Therefore, it is hard to extract precise estimations of intrinsic camera parameters where there is no explicit data about cameras, which is the case in most of the videos uploaded on the Internet. However, we noticed when we use the default camera calibration file, ORB-SLAM2 works robustly and the general shapes of returned paths are preserved in most of the cases. As a confirmation, we used the paths extracted by ORB-SLAM2 with the default calibration file to filter circular and linear backward/forward camera motions in Section 6. We showed that we gained perfect precision in filtering out circular and linear paths in Section 6.3 and Figure 14, and the results were satisfactory.

In addition, because we decomposed each edited drone video to its constituent clips based on scene-cuts, each clip is just representing one type of a scene and has visually similar content, which makes it easy to track the camera motion using ORB-SLAM2. Moreover, in our dataset statistics reported in Section 3.4, we showed that the duration of drone clips in edited videos by humans are mostly short (around 5 seconds). ORB-SLAM2 usually fails to track cameras when the clip is long and scene changes a lot, which makes it hard to find visually corresponding points between consecutive frames.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? *If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.* Our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset. We start gathering dataset from 2019, and till then none of our dataset links have been changed or deactivated. Authors might update the links in a case that the dataset links are changed, using the meta-data of our dataset links. To the best of our knowledge, there is no restrictions associated with any of the external resources in sharing video links of their content because video sharing functionality is provided by the YouTube and AIRVŪZ websites.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? *If so, please provide a description.* No, the dataset does not contain any confidential data.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? *If so, please describe why.* No, the dataset does not contain any offensive data.

Does the dataset relate to people? *If not, you may skip the remaining questions in this section.* While our dataset does not relate to specific individuals, our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset.

Does the dataset identify any subpopulations (e.g., by age, gender)? *If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.* No, the dataset does not identify any subpopulations.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? *If so, please describe how.* No, it is not possible to identify individuals, both directly and indirectly.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic

origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description. No, the dataset does not contain sensitive data in any way.

3 COLLECTION PROCESS

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how. The manual annotations were obrained from human annotators. The automatic annotations were obtained using ORB-SLAM2 and PhotoScan methods which their code/software are available to the public. We validated our manual annotations through the agreement of multiple independent human annotators. See annotation validation in Section 3.2 of our paper.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated? The annotators were provided with clear instructions about the labelling. In order to remove any potential confusion, we gave the annotators descriptive requirements with representative examples. To have a common annotation environment, the annotators were told to use the same video player (online player on website). The manual annotations were validated through the agreement of multiple independent human annotators. See annotation validation in Sec. 3.2 of our paper. The automatic annotations were obtained employing ORB-SLAM2 and PhotoScan methods which their code/software are available to the public for free.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)? The dataset is a sample of *edited* drone videos, captured and edited by human. We randomly select 991 source videos from different drone video categories such as nature, cities, and etc.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)? All authors of the paper, in addition to students/interns, were involved in the data collection process. We paid annotators a different amount of money based on a mutual agreement and on hourly basis.

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the time-frame in which the data associated with the instances was created. The dataset was gradually created till June 2022. All of our manual annotations are based on video content and consequently independent of our dataset creation time. Such annotations include video editing (cut time and transition), clip category (drone/non-drone), time information

(day/night/between), scene type (rural/urban/mixed), shot location (indoor/outdoor), logo, and text presence.

Among our automatic annotations, drone camera path and 3D scene reconstruction are also independent of our dataset creation time. However, social platform meta-data such as number of likes, comments, and views of videos might change on a daily basis due to new visitors of video platforms. We updated the social platform meta-data in August 2021.

Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation. No ethical review process was required because we did not require any study participants.

Does the dataset relate to people? If not, you may skip the remainder of the questions in this section. While our dataset does not relate to specific individuals, our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset.

Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself. No, our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset. In addition, video link sharing is provided in these websites as a built-in feature through the video link sharing button, which means that one can easily share the link of a video in these video platforms, and the individuals who uploaded their videos to these websites are aware that their video can be seen and might be shared using these video sharing platforms.

Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented. No, our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate). Not applicable.

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation. Not applicable.

4 PREPROCESSING/CLEANING/LABELING

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tag-

ging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remainder of the questions in this section. No, we just validate our manual annotations through the agreement of multiple independent human annotators. See annotation validation in Sec. 3 of our paper.

5 USES

Has the dataset been used for any tasks already? If so, please provide a description. The dataset is novel and has not been used by others. The dataset has only been used by the authors for three applications, explained as follow:

1) **Drone video classification:** We trained deep learning networks and provided new baseline results (Section 5 in the paper), such as classification of drone vs. non-drone, scene type (rural/mixed/urban), time information (day/between/night), location (indoor/outdoor), and logo vs. non-logo presence.
2) **Drone video semantic search:** We also presented a distillation method which automatically filters video collections to return true drone shots with its associated information (e.g., time, path shape, and scene type). It enables a user to search for drone video clips with high-level semantic search query, such as "drone video shot outdoor in daytime with an orbiting motion" (Section 6 in the paper).
3) **Video cut detection:** We evaluated existing methods of video cut detection (Section 4 in the paper): ORB-SLAM2, Yahoo Hecate, PySceneDetect, and ffprobe.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point. Not applicable. The dataset has only been used by the authors and our code is available to the public in our dataset link.

What (other) tasks could the dataset be used for? Because we introduced the first drone video dataset composed of *edited* drone videos, it brings new challenges and applications as follows (see our future work described in Section 7 for more details):

1) **drone path planning** using our camera path information,
2) **popularity prediction of drone videos** using social meta-data annotation,
3) **music recommendation for drone videos** using audio-visual data of our *edited* drone videos,
4) **recommendation of video transition effects** using our cut transition annotations,
5) **improving drone video classification** using confidence score provided in our dataset to detect drone/non-drone video clips,
6) **making new drone video clip datasets** by extracting drone clips from unstructured video collections using our "scaling to large distilled drone video clips" experiment described in Section 6 of the paper.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms? No.

Are there tasks for which the dataset should not be used? If so, please provide a description. No.

Any other comments? For the suggested future work such as music recommendation or drone path planning, we recommend users to leverage our drone video clips with high number of views or likes because these drone clips are more probable to have aesthetically pleasing audio-visual data or drone camera paths.

6 DISTRIBUTION

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description. Use of the dataset is free to all researchers after signing a data use agreement which stipulates, among other items, that (1) the user will not share the data, and (2) any publication that makes use of the data will also make the relevant code available.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)? The dataset link will be distributed and maintained through the Visual Media Lab (VML) website[12] on a cloud sharing service such as Microsoft OneDrive. Users will have access to our dataset link after contacting authors to sign a data use agreement, as described above. The dataset might also be additionally distributed though figshare[13] or Harvard Dataverse[14] research data repositories, which will ensure the long term data availability. DOI will be assigned upon sharing.

When will the dataset be distributed? The dataset is planned to release to the public after going through the publication process. Users will have access to our dataset link after contacting authors to sign a data use agreement, as described above.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions. The dataset annotations is licensed under the CC BY-NC-SA 4.0[15] (i.e., an annotation-only license). We only provide video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making this dataset. Source videos are owned by the video sharing platforms, i.e., YouTube and AIRVŪZ websites.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any

---

12. https://vml.kaist.ac.kr
13. https://figshare.com/about
14. https://dataverse.harvard.edu/
15. https://creativecommons.org/licenses/by-nc-sa/4.0/

fees associated with these restrictions. There is no restrictions associated with any of the external resources in sharing video links of their content because video sharing functionality is provided by the YouTube and AIRVŪZ websites..

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation. No regulatory restrictions apply apart from the licensed use and agreement form mentioned above.

7 MAINTENANCE

Who is supporting/hosting/maintaining the dataset? Dataset hosting will be supported by the dataset authors and other members of Visual Media Lab (VML) at KAIST on our lab website[16]. In addition, the dataset hosting might be outsourced to the dataset hosting providers such as figshare or Harvard Dataverse research data repositories.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)? Contact information:

- Amirsaman Ashtari (a.s.ashtari@kaist.ac.kr)
- Raehyuk Jung (raeomi93@kaist.ac.kr)
- Mingxiao Li (m398li@uwaterloo.ca)
- Junyong Noh (junyongnoh@kaist.ac.kr)

Is there an erratum? If so, please provide a link or other access point. No, this is the first version of the dataset.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)? If the dataset is updated in future, the older version will still be supported and kept around for consistency. The development of the dataset is planned to continue, and contributions from users are also welcomed. The changes to the dataset will be announced through the associated Google Group, which will be created upon data publication.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced. No, while our dataset does not relate to specific individuals, our dataset provides video links to video sharing platforms such as YouTube and AIRVŪZ, in which people made their videos available to the public, prior to making our dataset.

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to users. Yes, we plan to support versioning of the dataset so that all the versions are available to potential users.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description. New datasets and data samples might be uploaded to the main hosting after discussion with the dataset managers over the email. We will check the validity of a proposed dataset by a contributor. To this end, we will randomly select samples from the proposed dataset and check the validity of its annotations. If the proposed dataset is valid, we will add it to our main dataset and cite the contributor.

REFERENCES

[1] O. Bogdan, V. Eckstein, F. Rameau, and J.-C. Bazin. Deepcalib: a deep learning approach for automatic intrinsic calibration of wide field-of-view cameras. In *Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production*, 2018. 3

[2] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford. Datasheets for datasets. *arXiv preprint arXiv:1803.09010*, 2018. 1

[3] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *CVPR*, 2017. 1

[4] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri. A closer look at spatiotemporal convolutions for action recognition. In *CVPR*, 2018. 1

[5] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *TPAMI*, 2017. 1

---

16. https://vml.kaist.ac.kr