

Dataset name	Dataset version	Origin date	Origin	Description of dataset	Description of dataset collection process	Downloaded outlet	Is outlet licensed?	Is dataset publicly available?	Where license was found	License location	License content	Hash code	Size	Format	Additional notes
CIFAR-10	N/A	2009	https://www.cs.toronto.edu/~kriz/cifar.html	The CIFAR-10 dataset consists of 60000 32x32 colour images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images	The CIFAR-10 and CIFAR-100 are labeled subsets of the 80 million tiny images dataset. They were collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton.	N/A	N/A	Yes	Present on the official dataset website	https://www.cs.toronto.edu/~kriz/cifar.html	<license> <name>cifar paper citation</name> <hash>651A4DCDA5635BF26914F7B21966D57</hash> </license>	MD5: 70270a685842c9e89bb428ec5976c926 (Matlab version); MD5: c58f30108f718f92721af3b95e74349a (Python version); MD5: c32a1d4ab5d03f1284b67883e8d87530 (binary version)	175 MB (Matlab version); 163MB (Python version); 162 MB (binary version)	tar.gz	This dataset is a subset of another dataset called 80 Million Tiny Images
ImageNet	Large Scale Visual Recognition Challenge (ILSVRC) 2012-2017 image classification and localization dataset	2008	https://image-net.org/challenges/LSVRC/2012/2012-downloads.php#images	ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images [...] The data is available for free to researchers for non-commercial use.	We collect candidate images from the Internet by querying several image search engines	N/A	N/A	Yes	Present on the official dataset website, on a separate page	https://image-net.org/download.php	<license> <name>imagenet Terms of Access</name> <hash>55EB32BC75A822E4522317F4545A426B</hash> </license>	MD5: 1d675b47d978889d74fa0da5fadfb00e (train [tasks 1 & 2]); MD5: ccaf1013018ac1037901578038d370da (train [task 3]); MD5: 29b22e2961454d5413dabcf34fc5622 (validation [all tasks]); MD5: e1b8681fff3d63731c599df9b4b6fc02 (test [all tasks])	138GB (train [tasks 1 & 2]); 728MB (train [task 3]); 6.3GB (validation [all tasks]); 13GB (test [all tasks])	.tar	The dataset creators do not state which websites they queried for image data
Cityscapes	N/A	2015	https://www.cityscapes-dataset.com/downloads/	We introduce Cityscapes, a benchmark suite and large-scale dataset to train and test approaches for pixel-level and instance-level semantic labeling. Cityscapes is comprised of a large, diverse set of stereo video sequences recorded in streets from 50 different cities. 5000 of these images have high quality pixel-level annotations; 20 000 additional images have coarse annotations to enable methods that leverage large volumes of weakly-labeled data.	Several hundreds of thousands of frames were acquired from a moving vehicle during the span of several months, covering spring, summer, and fall in 50 cities, primarily in Germany but also in neighboring countries	N/A	N/A	Yes	Present on the official dataset website, on a separate page	https://www.cityscapes-dataset.com/license/	<license> <name>cityscapes dataset license</name> <hash>B3763863E568C02E41E49F7C25209A71</hash> </license>	See Cityscapes Dataset Hash Codes and Sizes	See Cityscapes Dataset Hash Codes and Sizes	.zip	This dataset's data was collected from recorded driving
FFHQ	2017	2017	https://github.com/NVlabs/ffhq-dataset	The dataset consists of 70,000 high-quality PNG images at 1024x1024 resolution and contains considerable variation in terms of age, ethnicity and image background. It also has good coverage of accessories such as eyeglasses, sunglasses, hats, etc	The images were crawled from Flickr [...] Only images under permissive licenses were collected.	N/A	N/A	Yes	Present on the official dataset website, on a separate page	https://github.com/NVlabs/ffhq-dataset/blob/master/LICENSE.txt	<license> <name>ffhq dataset license</name> <hash>7D5D62F93E45FD46363BE7B6CC85648</hash> </name> <name>Creative Commons BY 2.0</name> <hash>A8449650007DAA61ABA0775FBD2D8F97</hash> </name> <name>Creative Commons BY-NC 2.0</name> <hash>2FD66DE7BF107DDACAEF3579B0EDA3D</hash> </name> <name>Public Domain Mark 1.0</name> <hash>43A331AD5C6CE6C2BD9472767B3A02A0</hash> </name> <name>Public Domain CC0 1.0</name> <hash>E588186959B0B3DA645B3DE816AE1C8D</hash> </name> <name>U.S. Government Works license</name> <hash>7815688D3ED7FE7DC97109B3D8FB3395</hash> </name> <name>Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International</name> <hash>3617FFB033C5C83A254CACDCE923FFD0</hash> </license>	N/A	1.28 TB - 255 MB (varying sizes)	.json, .png, .tifrecords, .zip	This dataset's images were collected
VGGFace2	N/A	2016	N/A	The dataset contains 3.31 million images of 9131 subjects, with an average of 362.6 images for each subject. Images are downloaded from Google Image Search and have large variations in pose, age, illumination, ethnicity and profession (e.g. actors, athletes, politicians). The dataset was collected with three goals in mind: (i) to have both a large number of identities and also a large number of images for each identity; (ii) to cover a large range of pose, age and ethnicity; and (iii) to minimise the label noise.	The Images were downloaded from Google Image Search [...] We query in Google Image Search and download 1000 images for each subject. To obtain images with large pose and age variations, we then append the keyword 'sideview' and 'very young' to each name and download 200 images for each. This results in 1400 images for each identity	N/A	N/A	unknown	N/A	N/A	N/A	N/A	N/A	N/A	Neither this dataset nor its license could be located

Dataset name	Dataset version	Origin date	Origin	Description of dataset	Description of dataset collection process	Downloaded outlet	Is outlet licensed?	Is dataset publicly available?	Where license was found	License location	License content	Hash code	Size	Format	Additional notes
MS COCO	N/A	2013	https://cocodataset.org/#download	The MS COCO (Microsoft Common Objects in Context) dataset is a large-scale object detection, segmentation, key-point detection, and captioning dataset. The dataset consists of 328K images.	We collected images from Flickr which tends to have fewer iconic images. Flickr contains photos uploaded by amateur photographers with searchable metadata and keywords.	N/A	N/A	Yes	Present on the official dataset website, on a separate page	https://cocodataset.org/#termsofuse	<license> <name>Creative Commons Attribution 4.0 International</name> <hash>0574E199E3FE568B7FCE2D9B286A06DE</hash> <name>ms coco dataset copyright</name> <hash>2F398401E0138D79FF92E B5A9AC4CADA</hash> <name>Flickr Terms of Use for ms coco dataset</name> <hash>05BC1A173F5B055D04AAE D18B3B166D4</hash> </license>	N/A	13GB (2014 train); 6GB (2014 validation); 6GB (2014 test); 12GB (2015 test); 18GB (2017 train); 1GB (2017 validation); 6GB (2017 test); 19GB (2017 unlabelled)	.zip	This dataset's images were collected from Flickr

Name	MD5 checksum	Size
gtFine_trainvaltest.zip	4237c19de34c8a376e9ba46b495d6f66	241 MB
gtCoarse.zip	1c7b95c84b1d36cc59a9194d8e5b989f	1.3 GB
leftImg8bit_trainvaltest.zip	0a6e97e94b616a514066c9e2adb0c97f	11 GB
leftImg8bit_trainextra.zip	9167a331a158ce3e8989e166c95d56d4	44 GB
rightImg8bit_trainvaltest.zip	4173f20bc5bece70da010e3260fae7a6	11 GB
rightImg8bit_trainextra.zip	9621f0f5b066795c24b9d2991c15e0fe	44 GB
disparity_trainvaltest.zip	2c8272766993c983321f34c73daada5c	3.5 GB
disparity_trainextra.zip	d74998385c3452e91d54b561b8f5c8eb	15 GB
camera_trainvaltest.zip	ee0838cd24c40cf252937817ae017ed0	2 MB
camera_trainextra.zip	5410c2490ae9ce72a92305e053a15de1	8 MB
vehicle_trainvaltest.zip	dac47d38c2085165a51afa89b4ddcc0b	2 MB
vehicle_trainextra.zip	2011af9bdfa345eb97e2d5c4e66a5d5d	7 MB
leftImg8bit_demoVideo.zip	21ae5fb1a3393b3c3b57503dd585ef09	6.6 GB
all_demoVideo.zip	fdcd6584c18ca04418f342c68739c8ff	59 GB
leftImg8bit_blurred.zip	37d49f6d65ea1b747a163d5579c48f40	11 GB
leftImg8bit_sequence_trainvaltest.zip	4348961b135d856c1777f7f1098f7266	324 GB
timestamp_sequence.zip	1c71aca93dd34ad2ea8c4546bf02098a	40 MB
vehicle_sequence.zip	7ed2b1b019c3e2c91af17cb70b1e1bc8	56 MB
rightImg8bit_sequence_trainvaltest.zip	243124abb2e3a2189fa70aa6fabbb296	322 GB
disparity_sequence_trainvaltest.zip	0a10cc2c5a2a16f1ca8c7ec9a208ff21	106 GB
leftImg8bit_allFrames_frankfurt.zip	80e5191a7dbcae9837687d93b99c4b10	241 GB
timestamp_allFrames_frankfurt.zip	76136b7c82c67c2a72bb64a9bfb326a3	29 MB
vehicle_allFrames_frankfurt.zip	856ca1c15e92b20625ec5bbebba70ca4	38 MB
leftImg16bit_trainvaltest.zip	71accfbd463774918712ff171954b3b4	37 GB
rightImg8bit_allFrames_frankfurt.zip	99e9a91729b9e747b215f282db8fcc19	242 GB
leftImg16bit_trainextra.zip	976b896e6827ff228bdd9dc526171d39	149 GB
rightImg16bit_trainvaltest.zip	3cab37116734fb2bb8a6cbe9de9fb437	37 GB
rightImg16bit_trainextra.zip	73cf17baee3c76ce5279783841e00111	148 GB

Dataset/data source name	Is there another data source?	License	License range
CIFAR-10	Yes	Request to cite paper (2009-09-23)	2008-2009
80 Million Tiny Images	Yes	N/A	2005-2006
Google Images	No	Terms of Service (2003-03-29); Terms of Service (2005-11-05)	2005-2006
AltaVista	No	Terms of Use (2005-01-13); copyright (2005-02-12)	2005-2006
Ask	No	Terms of Service (2010-07-30)	2005-2006
Flickr	No	Yahoo! Terms & Conditions of Use (2006-12-31); Flickr.com Terms of Use (2005-12-01)	2005-2006
Cydral	No	N/A	2005-2006
Picsearch	No	Terms of Use (2005-05-24)	2005-2006
Webshots	No	copyright (2005-01-03); Terms of Use (2008-12-03)	2005-2006
ImageNet	Yes, but unknown	Terms of Access (2009-05-16)	2007-2008
Cityscapes	No	License agreement (2017-02-25)	2014-2015
FFHQ	Yes	Custom (2019-12-22); Creative Commons BY 2.0 ; Creative Commons BY-NC 2.0 ; Public Domain Mark 1.0 ; Public Domain CC0 1.0 ; U.S. Government Works ; Creative Commons BY-NC-SA 4.0	2016-2017
Flickr	No	Yahoo! Terms of Service (2017-12-12); Oath Terms of Service (2018-12-12)	2016-2017

Dataset/data source name	Is there another data source?	License	License range
VGGFace2	Yes	License (2017-11-13); Creative Commons Attribution 4.0 International	2015-2016
Google Images	No	Terms of Service (2014-04-30)	2015-2016
MS COCO	Yes	CC Attribution 4.0 International ; Copyright (2017-08-30)	2012-2013
Flickr	No	Yahoo! Terms of Service (2012-01-01)	2012-2013

[illegible]

[illegible]