

## Dataset Description

The VICR dataset is supplied as a single comma separated values file in `Datasets/VICR.csv`. It contains a single header and a row for each image-caption pair. The header lists the following columns:

- `image` – the url of the image
- `caption` – the caption for the image
- `subset` – either train, val, or test
- `num_ratings` – the number of ratings for the pair
- `ratings` – the first rating, marking the start of the list of ratings
- `rating_2` – the second rating
- `rating_3` – the third rating
- `rating_4` – the fourth rating, if available
- `rating_5` – the fifth rating, if available
- `rating_6` – the sixth rating, if available
- `rating_7` – the seventh rating, if available

The dataset is a ragged array, with not every row containing entries in every column (since not every image-caption pair received the same number of ratings). The headings for `rating_2` - `rating_7` were added to allow convenient loading of the dataset with Pandas.

## Dataset Statistics

There are 15,646 image-caption pairs (rows) in the VICR dataset. These have between 3 and 7 ratings each, with a total of 68,217 ratings. There are 9,990 unique images, with 80% of the images being paired with a single caption, and 10% of the images being paired with two captions. The remaining 10% of the images are paired with between 3 and 10 captions. The captions are 10.9 words long on average, with the shortest caption having two words, and the longest having 30 words.

## Qualitative Examples

`VICR_visualize_25.html` contains a sampling of 25 image-caption pairs of varying quality. `VICR_visualize_dataset_order.html` contains a similar presentation of the entire dataset. `VICR_visualize_sorted.html` contains the entire dataset sorted by average rating.

## Dataset Availability / Maintenance

Upon successful completion of the anonymized review process, we will host and manage the dataset and code on our own servers.

## Dataset License

The VICR dataset is licensed under the Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license, which allows for non-commercial use with attribution. Derivative works must share the same license. The full text of the license can be found at

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

The dataset uses assets from Flickr8k-Expert, which is licensed under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license. To view a copy of this license, visit

<http://creativecommons.org/licenses/by/3.0/>

The dataset uses assets from MS-COCO, which is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. To view a copy of this license, visit

<https://creativecommons.org/licenses/by/4.0/legalcode>

## Code License

The code for the caption rating game and for the evaluation is licensed under the MIT license:

Copyright 2023

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

## Author Statement

The authors bear all responsibility in case of violation of rights.

## Code Instructions

**Predictor Model:** The code for the predictor model is in Code/Evaluation/predictor. The two main files are

- `vilbert_embeddings.py` – extracts ViLBERT embeddings given an input csv file and saves them in a custom serialized format
- `vicr.py` – trains and/or evaluates models on pre-extracted ViLBERT embeddings

`vilbert_embeddings.py` requires ViLBERT to run, which may prove challenging to install. We have supplied pre-extracted embeddings in the `Code/Evaluation/predictor/embeddings` folder for all image-caption pairs in our dataset. `vicr.py` only requires Python 3.8, Numpy 1.20, Tensorflow 2.4, and Scipy 1.10. To test the supplied model and evaluate Kendall's tau on the test set, run the following command:

```
python vicr.py models/model-VICR.h5 embeddings/VICR-test-vilbert.emb -kt :test -vb
```

To train a new model, you can specify its architecture on the command line using custom syntax. An example is the following, which creates a dense network with a layer with 512 neurons followed by a layer with 128 neurons and a single output neuron, using relu as the activation functions in the hidden layers and utilizing dropout between the layers. Epochs, batch size, and learning rate are also specified.

```
python vicr.py models/new-model.h5 embeddings/VICR-test-vilbert.emb -t :auto -v :auto
-n "d(512)[relu]->drop(0.6)->d(128)[relu]->drop(0.4)->d(1)" -e 1000 -b 200 -l 0.00001
-vb
```

For more information, consult the documentation by typing:

```
python vicr.py -h
```

**Inter-Rater-Agreement:** The two python scripts in `Code/Evaluation/inter-rater-agreement` were used to obtain Kendall's tau, W, and Fleiss' kappa statistics for the paper.

**Rating Game:** The source code for the rating game can be found in `Code/Game`. It is implemented in `node.js`.

## Rater Instructions

The instructions given to the raters who played our game in order to collect data for the VICR dataset are supplied in `instructions.html`