

# NICE :

## New Frontiers for Zero-shot Image Captioning Evaluation

Sponsors



SEOUL NATIONAL UNIVERSITY

### 01 Overview

The purpose of this project is to challenge the computer vision community to develop robust image captioning models that advance the state-of-the-art both in terms of accuracy and fairness (i.e. mitigating societal biases). Both of these issues must be addressed fully before image captioning technology can be reliably deployed in a large-scale setting.

The challenge and workshop will focus on testing the true limits of image captioning models under the zero-shot image captioning setting. It aims to challenge the models by providing a new evaluation dataset that includes a larger variety of visual concepts from many domains as well as various image types. To accomplish this task, the models need to broadly understand language-vision relations and also learn how to combine language components for a new concept of image.

#### Organizers

Organizing Committee :

- > **Seung Hwan Kim** (LG AI Research) / Co-chair
- > **Kyoung Mu Lee** (Seoul National University) / Co-chair
- > **Bohyung Han** (Seoul National University)
- > **Alessandra Sala** (Shutterstock)

Technical Committee :

- > **Sihaeng Lee** (LG AI Research)
- > **Taehoon Kim** (LG AI Research)
- > **Pyunghwan Ahn** (LG AI Research)
- > **Sangyun Kim** (LG AI Research)
- > **Mark Marsden** (Shutterstock)

#### Important Dates

- > Evaluation Server Open (Validation Data) : **February 1**
- > Test Data Release & Evaluation Server Open (Test Data) : **April 3**
- > Challenge Submission Deadline : **April 30**
- > Challenge Result Notification : **May 15**
- > Full Submission Deadline (Challenge Winners) : **May 30**
- > Workshop @ CVPR 2023 Venue : **June 18 morning**

### 02 Challenge

The challenge aims to evaluate zero-shot image captioning ability of models on a challenging dataset newly provided by Shutterstock. We provide separate sets of validation and test data, and evaluate models using a few widely used metrics to assess generated captions.

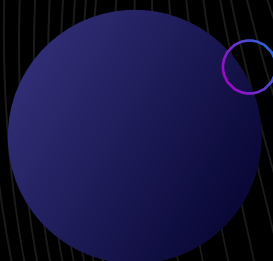
The challenge is divided into two tracks:

**Track 1** Student track (The 1st author of the submission must be a student and affiliation of every author must be an academic institution)

**Track 2** No limit

The winner of each track will be given prize money of \$5,000 and the runner-up will get \$2,500.

The prize winners of the challenge will be requested to give a short talk on their approach at the workshop. They will also be requested to submit tech report and make their implementation public for a certain period. For detailed license statements, please look through the website (<http://nice.lgresearch.ai/>) before submission.



Please visit [nice.lgresearch.ai](http://nice.lgresearch.ai/) for more detail.

# 03 Dataset

In this project, Shutterstock provides a new open-sourced evaluation dataset. This new dataset consists of roughly 26k high-quality images with associated curated metadata and it covers more than 20 general categories and a wide breadth of concepts. With this dataset, we expect the researchers in this community to take a longitudinal evaluation across a variety of metrics to comparatively assess performance of different zero-shot image captioning models.

The dataset is divided into 5k validation data (NICE-val-5k) and 21k test data (NICE-test-21k).

Each image in this collection is licensed and was approved via a human review system which controls for image quality and legal compliance. Images are 500px on the longest side with a single ground truth caption provided. All captions and keywords are in English and were moderated for hate speech, slurs, and expletives. This dataset contains a very broad set of concepts and scenarios (indoor and outdoor, with or without people, with or without animals).



A group of young friends having a barbecue outdoors next to a tent



St. Sebastian's Church, Ramsau, near Berchtesgaden, Bavaria, Germany



3D concept of online banking on mobile phone showing services currencies and graphs



View from sandy beach of picturesque sunset horizon over sea surf lapping on shore

# 04 Workshop

The half-day workshop titled "New Frontiers for Zero-shot Image Captioning Evaluation" will take place in the morning of June 18 (Day 1 of CVPR workshop, ID 292) as a part of Vision + Language track.

- > **Workshop ID** : 292 "New Frontiers for Zero-shot Image Captioning Evaluation"
- > **Schedule** : June 18, morning (final schedule will be announced on the website(<http://nice.lgresearch.ai>))

Opening Talk	Kyoung Mu Lee (Seoul National University)	
Organizer Talks	Alessandra Sala (Shutterstock)	Seung Hwan Kim (LG AI Research)
Challenge Winner Talks	Winner & Runner-up of each track	
Keynote Speech	Honglak Lee (LG AI Research / University of Michigan)	
Invited Talks	Noa Garcia (Osaka University)	Cordelia Schmid (INRIA / Google Research)
	Jack Hessel (Allen Institute for AI)	Hamid Palangi (Microsoft Research / University of Washington)

## Contact

Taehoon Kim      taehoon.kim@lgresearch.ai  
Research Scientist. LG AI Research

Pyunghwan Ahn      p.ahn@lgresearch.ai  
Research Scientist. LG AI Research

Please visit [nice.lgresearch.ai](http://nice.lgresearch.ai) for more detail.

