Im2Text: Describing Images Using 1 Million Captioned Photographs

Vicente Ordonez (presenter), Girish Kulkarni, Tamara L. Berg
Stony Brook University

One of the many stone bridges in town that carry the gravel carriage roads.

A stone bridge over a peaceful river.

An old bridge over dirty green water.

Computer Vision

sky
trees
water
building
bridge

Our Goal
Harness the Web!

SBU Captioned Photo Dataset
1 million captioned images!

Matching using Global Image Features (GIST + Color)

Transfer Caption(s)
e.g. “The water is clear enough to see fish swimming around in it.”

- Smallest house in Paris between red (on right) and beige (on left).
- Bridge to temple in Hoan Kiem lake.
- A walk around the lake near our house with Abby.
- Hangzhou bridge in West lake.
- The Daintree river by boat.
<table>
<thead>
<tr>
<th>Bridge/Location</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron bridge over the Duck river.</td>
<td></td>
</tr>
<tr>
<td>Bridge over Cacapon river.</td>
<td></td>
</tr>
<tr>
<td>The Daintree river by boat.</td>
<td></td>
</tr>
<tr>
<td>The bridge over the lake on Suzhou Street.</td>
<td></td>
</tr>
</tbody>
</table>

Use High Level Content to Rerank (Objects, Stuff, People, Scenes, Captions)

Transfer Caption(s)

e.g. “The bridge over the lake on Suzhou Street.”
Amazing colours in the sky at sunset with the orange of the cloud and the blue of the sky behind.

A female Mallard duck in the lake at Luukki Espoo.

Cat in sink.

Fresh fruit and vegetables at the market in Port Louis Mauritius.

The cat in the window.

The boat ended up a kilometre from the water in the middle of the airstrip.