

# 4-shot examples

## Caption Examples (on COCO Caption dataset)

### 1. Hallucinations.



**openflamingo:** A person is sitting on a couch.

**ours:** A living room with a couch and a wooden coffee table.

**gt:** A living room has a couch and a rustic chest for a coffee table.

A PHOTO OF A LIVING ROOM WITH COUCHES AND A TABLE

The couch is tidy but empty of people.

A living room with a wooden crate coffee table

A white couch sitting in a living room next to a wooden table.



**openflamingo:** A Samsung Galaxy S3 and a HTC One X

**ours:** Two mobile phones side by side, one with a black screen and the other with a green Android logo

**gt:** A new android phone is next to another android phone.

a couple of cellphones are on a table

Two different kind of cell phones sitting on a table

Two phones, one showing the time and one turned off.  
There are two cell phone's one is on with a robot.

## 2. Details.



**openflamingo:** A clock in Prague.

**ours:** The Prague Astronomical Clock is a medieval astronomical clock located in Prague



**gt:** An old ornate clock face set in a brick building.  
 People standing in front of a stone building.  
 A large church tower with a small crowd of people standing underneath it.  
 A building with different clocks on the face of it  
 A very intricately designed old tower clock, with people coming through the arched doorway at the bottom of the photograph



**openflamingo:** A graffiti mural on a wall.

**ours:** A woman walking by a storefront with a black and white street art mural.

**gt:** A woman walks past street posters and graffiti.  
 A woman walks past a boarded up building decorated with black and white graphic art.  
 A woman walking down a sidewalk in front of graffiti.  
 A mural on a city wall with a women walking down the sidewalk.  
 a woman walking down a dusty street in front of a mural

# VQA Examples (on OK-VQA dataset)

## 1. Correctness.



**question:** What type of tool is she using for her hair?

**openflamingo:** comb

**ours:** brush

**gt:** ['brush', 'brush', 'brush', 'brush', 'brush', 'brush', 'brush', 'brush', 'brush', 'brush']

## 2. Usual representation



**question:** What place does this vehicle stop?

**openflamingo:** MRT

**ours:** station

gt: ['station', 'station', 'station', 'station', 'station', 'station', 'station', 'station',  
'depot', 'depot']

### 3. Redundant content.





**question:** Why are these men dressed like this?

**openflamingo:** they are in the military

**ours:** military

gt: ['military', 'military', 'military', 'military', 'military', 'military', 'military',  
'military', 'soldiers', 'soldiers']