

This document provides more detail on the findings laid out in the [Model Evaluation Document](#). While the [Model Evaluation Document](#) summarized testing on several object detection models, this document only focuses on the results of further testing done using our chosen model – the Faster R-CNN model.

The [Model Evaluation Document](#) tested the Faster R-CNN model on a randomized set of images from the Library's [Free to Use and Reuse](#) collection. However, many of the images and collections we wanted to incorporate into our collage tool vary significantly from these initial, randomized images, particularly in terms of quality.

We needed to test not only how the model would work with black and white or sepia toned images; hazy, blurry or otherwise unclear images; and images with limited lighting, but *if* the model's performance would remain despite such characteristics.

To do this, we curated a list of 46 images from collections of interest, keeping in mind the following categories:

- Nature
- Landmark
- People
- Backdrop
- Vehicle

For each category, the amount of images selected were divided as follows:

- At least three images in the “Landmark” category.
- At least three images in the “People” category, paying attention to the color and focus of the image as well as the skin tone of the image’s subject.
- At least two images in the “Nature” (or “Animal”) category.
- At least two images in the Backdrop category (which included miscellaneous items like street lights or fire hydrants).
- At least two images in the “Vehicle” category.

Images could account for more than one category.

In the ‘Observation’ column of Table 1, we list observations about the Faster R-CNN model’s interaction with our sample collection set. In the ‘Description’ column of Table 1, we list out examples of each observation found in the ‘Observations’ column.

The observations in Table 1 will be used to create an assessment of our current iteration of the collage tool and what needs to be changed to better work with the capabilities of the Faster R-CNN model.

For a more generalized overview of what the Faster R-CNN model was able to capture from our sample collection, view the ‘sample-collections’ sheet in our [ifp-collections](#) spreadsheet.

When referencing an item from our sample collection, we will use the column letter and row number. For instance, C2.

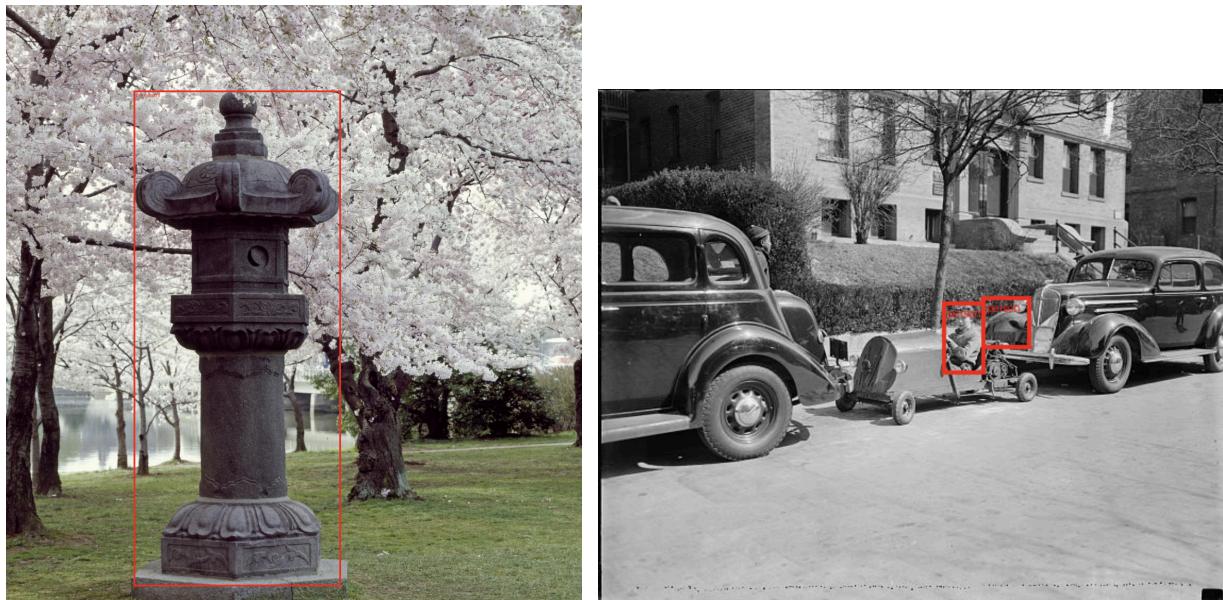
Observation	Description
Privileging of humanoid silhouettes	<p>The model captures humanoid silhouettes the most consistently and with the largest/highest bounding box threshold.</p> <p>In instances where the model would theoretically capture a building-like structure (or landmark), it will not if a humanoid silhouette is close by. C24 has a similar structure to C25, with very similar lighting conditions and background noise. However, the model captured the structure in C25, but not C24 (where a person in the background was captured alone).</p> <p>It is worth noting that C25 is labeled as a “fire hydrant” by the model due to its shape. As such, it’s not read as a background item/building like C24.</p> <p>Along these lines, the model is able to capture statues or illustrations in the shape of humans (C38).</p> <p>See “Illustrations” for more information.</p>
Sizing and Shape	<p>It seems that the positionality of the subject or object in an image impacts whether the model can capture said subject or object.</p> <p>The cat in C2 appears in the model’s vocabulary, and the model is able to capture many other animals in its vocabulary, as seen in C8 (horses), C42 (dog) and C43 (sheep). C43 was a surprising capture since the image is damaged and washed out.</p> <p>However, C2 was not able to be captured, and this may be due to the costume that is placed on the cat that may “obscure” the silhouette of the cat.</p> <p>Initially, we thought that the cat was not able to be captured because of how large and central to the portrait it was, but the model</p>

	<p>was able to capture another close-up portrait (C19).</p>
Shadows [Color/Shading]	<p>Skin tone does not appear to affect the efficacy of the model. People across skin tones were able to be captured.</p> <p>This also applies whether the image is in color, black and white, or sepia-toned.</p> <p>Shape recognition seems to play a bigger role in object detection. Take C11, as an example. Though the two people in the distance are not clearly shown, the model is able to pick up one of them. It seems that the issue here is the person not captured is not clearly “defined” in the image.</p> <p>This explains images like C35. The image has ducks, which should be captured because “bird” is in the model’s vocabulary. However, unlike the near perfect capture of the ducks in C21, the lack of clarity in the image (and by extension, the shape of the duck), means that the ducks in C35 are not captured. The ducks in C35 blend with the background (no defined lines).</p>
Inconsistent capture	<p>The model often inconsistently captures objects and subjects in its vocabulary.</p> <p>There does not seem to be a pattern to or reason for this, including shape (mis)recognition (as seen with C1).</p> <p>For examples, view the following images:</p> <ul style="list-style-type: none"> • C4: Students, chairs, and desks not all captured, even where there is presumably no hindrance to recognition. • C17: Not all subjects captured, even without hindrance to recognition. • C39: Not all subjects captured, even without hindrance to recognition. • C40: Not all subjects captured, even without hindrance to recognition.
Clustering/Grouping	<p>The effects of clustering or grouping of subjects or objects on object detection is indeterminable. There are objects or subjects</p>

	<p>captured even when grouped together and others that are not – even when grouped together.</p> <p>Take C41 as an example. There are two objects with wheels captured. One is a cart and the other is a bicycle. The model is able to detect the bicycle, even though it is grouped with subjects and another object (a cart). However, the cart is not able to be captured despite being more forward and clearer in the image, even though grouped with subjects (ruling out the “privileging of human silhouettes” observation previously noted).</p>
Illustrations	<p>The model does not capture anything from illustrations unless it is a humanoid silhouette. See C10 and C12.</p>

Table 1: Assessment of Faster R-CNN model applied to sample collection.

In general, the model will capture something *if* it can relate it to a structure in its training model, as seen with C31, where a part of a car is labeled as a person and C25, where a building structure is labeled as a “fire hydrant.” If something “distorts” the expected shape of an object or subject (see C1), then the model will not recognize the object or subject, even if they are in the model’s vocabulary.



From left to right: structure not ordinarily captured mislabeled as “fire hydrant” (C25) and part of car mislabeled as “person” (C31).

Citations, from left to right: [1] Harris & Ewing, photographer. (1937) Washington youngster solves parking problem. Washington, D.C., March 30. Transportation and no parking worries. Nelm Clark, 16-year old Washington youngster, solved this problem by combining a lawn mower motor with a set of motor cycle gears to make this unusual midget auto. Costing \$60 to build the contraption weighs only 150 pounds--the weight is its main feature--and if you run out of gas you easily push it or tuck it under your arm and walk home, 3/30. United States Washington D.C. District of Columbia Washington D.C, 1937. [Photograph] Retrieved from the Library of Congress, <https://www.loc.gov/item/2016871444/>. [2] Highsmith, C. M., photographer. Cherry trees along the Tidal Basin with Japanese Lantern placed in the park in . Washington, D.C. United States District of Columbia Washington D.C. Washington D.C, None. [Between 1980 and 2006] [Photograph] Retrieved from the Library of Congress, <https://www.loc.gov/item/2011633233/>.



“Brünnhilde” cat not captured due to headdress and clothing.

Citation: Weidhaas, A. E., photographer. (1936) Brünnhilde. , 1936. [Photograph] Retrieved from the Library of Congress, <https://www.loc.gov/item/2017645524/>.

Given the shortcomings of the model, our initial project deliverable has shifted. We had hoped that users would be able to manipulate items in the “Landmark” category. However, the model is not able to capture buildings or any landscape. As such, anything within this category will be presented as an uneditable background upon which users can place capturable objects and subjects.