Comparison of Cloud-Based Image Processing Algorithms
CIS4010 Cloud Computing Team 3 Project Report

Christian Cornelis, Mohammadamin Sheikhtaheri, and Samuel Tracz
School of Computer Science, University of Guelph, Guelph, Ontario, Canada
{ccorneli, msheikht, traczs}@uoguelph.ca

Keywords: Machine Learning, Image Processing, Computer Vision, Azure, AWS, GCP, Microsoft, Amazon, Google Cloud Platform, Google Vision, Azure Computer Vision, Amazon Rekognition

Abstract: Three pre-trained image processing tools from different cloud platforms were used to compare their labelling-abilities and the limitations they may have. This test was meant to discover and compare edge-cases between the labelling abilities of three major cloud-based image processing APIs: Google Vision, Amazon Rekognition, and Azure Computer Vision.

1 INTRODUCTION

Image recognition is the creation of an artificial intelligence (AI) tool that recognizes and identifies an image. This tool is presented with heaps of images that are already identified and categorized in order to learn and recognize similar images. For example, an AI algorithm would be shown thousands of images of shoes, and would then learn what images of shoes should contain. If you show an image of an egg, the artificial intelligence would compare all the pixels in the egg image to all the images of shoes it's seen and would find no or few matches to shoes, so it would not identify it as a shoe.

Adversarial examples are inputs to a neural network that result in an incorrect output from the network. Neural networks could be trained to recognize images, but sometimes just a change in gradient of a picture is enough for the AI to not recognize what it is. The worst-case scenario would be that the image recognition tool would be confident that the image is something that in reality, it is not. This can have negative real world effects, such as a self-driving car seeing an image of a stop sign, but it is confident that it might be a person instead, and just continues without stopping.

We have selected 3 AI tools that have already been given thousands, if not millions, of pictures in order to avoid having to train them ourselves. This allowed us to focus on what limitations the current pre-trained tools have. Our project focused on determining how well each of the three cloud platforms have trained their image processing tools, and where each of them fail the most. Images have been selected and analyzed by all 3 of our programs that utilized Python software development kits (SDKs) to interact with the 3 cloud services chosen. The three of us then used our judgement to see how accurate each of the three results were, and how confident each tool was in it's labeling. We attempted to find images that at least one tool labelled incorrectly.

1.1 Project Statement

Compare the labelling systems implemented on 3 major pre-trained image processing APIs - Google's Vision API, Amazon's Rekognition API, and Azure's Computer Vision API.

2 OBJECTIVES

This project had a variety of objectives:

- Find limitations that one API seems to have that the others seem to handle appropriately.
- Learn more about Google Cloud Platform due to the fact that it is the least popular of the 3 platforms we are covering, and we did not have the chance to do so through the other aspects of this course.
- Find images that the Google Vision, Azure Computer Vision, or Amazon Rekognition mislabel.

- Compare and contrast differentiating labels between the three services, and attempt to speculate on why these labels may have been computed.

We aim to explore the question ”How can these algorithms label images that humans may have trouble identifying in the first place”?

3 CLOUD ASPECTS

We have 3 created 3 Python scripts that label an image using the 3 different cloud image processing APIs. Each label made by any API was reported accordingly, along with the confidence of each label.

3.1 Azure Computer Vision

As prerequisites for the Azure program, Python must be installed, and a subscription key for Computer Vision must be obtained. The computer vision key provides an API endpoint, along with a key in which you can set to environment variables. The Python libraries needed for this are: Requests - an HTTP library to call on the Azure Computer Vision API, Matplotlib - to display the picture with the label underneath, and pillow - a Python imaging library.

If you do not have an Azure account, you can create a free account here. Once you have an Azure account, you will have to have a subscription key for Computer Vision to call on the API. You can get a free trial key from the Try Cognitive Services site, or you can follow the create a cognitive account tutorial for a paid version.

The credentials given by the cognitive service portal can be set as environment variables. Set `COMPUTER_VISION_SUBSCRIPTION_KEY` and `COMPUTER_VISION_ENDPOINT` to whatever is given to you by Azure.

In this project, the Tag Image method was chosen as the API call, but you can go to the Cognitive Services API documentation here to choose which part of the image processing you want to call.

Within the code, you specify what image you want to open by giving it a path, and then make a POST request which contains the API call, headers (which contains the subscription key), parameters, and the image data. You will then get a JSON response, if successful, containing the analyzed data of the image.

This setup was made by following the Azure Vision Quickstart (Farley, 2020)

3.2 Amazon Rekognition

As prerequisites for running the Rekognition script, you also need Python (one of the many languages that is supported by the API), and the security credentials needed to connect to an AWS account. You can set the security credentials in the ”credentials” file created when installing the AWS CLI. You must also make sure the region is correct in the ”config” file. You can find both of these files in the ”.aws” folder in your computer. A storage bucket was used to store all of the pictures that were processed by the script. The script simply iterated through all of the images in the bucket, processed them, and outputted all of the labels that it caught, along with their highest confidence percentages.

Since Amazon Rekognition is part of the Boto3 library, the user will need to install the library simply by using `pip install boto3`. Connecting to the cloud resources needed via this Python SDK was very easy, and the functions provided were quite simple to understand and use. You can follow the steps to Get Started With Amazon Rekognition. Amazon’s pre-trained algorithm already exists on the cloud, and when using the SDK, we are making use of that algorithm in order to find edge cases.

This setup was made by following the Rekognition Quickstart (Amazon, 2020)

3.3 Google Vision

Using Google’s Vision API required little setup prior to using the Python SDK to gather labels that are assigned to processed images. Setup was performed by following the Vision Quickstart Guide (Google, 2020). To use Google Cloud Platform, an account must be created, or an existing Google Account can be linked for use with GCP. The Google Vision API requires that a billing account also be created as part of the setup process. For the purpose of this project, a Google Cloud Student account was used, which automatically creates a billing account with student credits of $50 US applied to the account. as well as a billing account within the GCP account. This billing account is crucial, as authentication errors are encountered when using the Vision API if billing is not tied to the project that is being used to connect to Google Vision. A project account must be set up after a billing account is created, which can be done directly via the Google Cloud Dashboard. Once the project account is created and is connected successfully to the billing account, the Vision API must simply be enabled for the project. Through the Google Cloud Dashboard, this can be performed by navigating to APIs, and then searching for 'Vision API', and clicking 'Enable'. A Service Account is needed prior to downloading credentials for use by the SDK. From work in another course, a Service Account was already linked to the account used to communicate with the Vision API, so this step was not necessary; however, the steps outlined in the Service Account setup documentation can be followed to create a service account. Once a Service Account is created, the credentials can then be downloaded in the form of a JSON file for the project, which will be used to authenticate whilst using the SDK.

To use the service account credentials downloaded properly, the environment variable `GOOGLE_APPLICATION_CREDENTIALS` must be set on the operating system. The Python SDK will automatically use these credentials when making calls to the Vision API.

The Vision API Python SDK can be downloaded via pip using `pip install google-cloud vision`. Once this has been done, and the aforementioned environment variable is set, the script created for processing all of our resources is set up correctly.

Connecting to this API via the SDK was extremely-easy due to the clear documentation on the matter provided by Google. Google provides code snippets for connecting to the API, which was essentially all of the code necessary to script the labelling of the images we found. Simple, minor tweaks were necessary to format the output of results and ensure that all of our files were being incrementally labelled.

4 MAJOR ACCOMPLISHMENTS AND CONTRIBUTIONS

We feel that the major contributions of this project are the analysis of the false-positive labelling that we encountered, and the ease-of-testing that we found whilst undertaking this project.

4.1 False-Positive Labelling Analysis

We were able to compile a small dataset of images that produce false-positive labelling with high confidence in at least one of the three services investigated in this project. We will break down these false-positives we encountered into the following categories: close-up images, images containing multiple, vastly-different objects, and distorted/odd images. Although our sample size was small for this project, it was interesting to see how the responses from each service differed between each other for the same image. All labels produced by the respective APIs can be found in APPENDIX.

Figure 1: Close-up of a US $100 bill. (Reshetnyak, 2016)

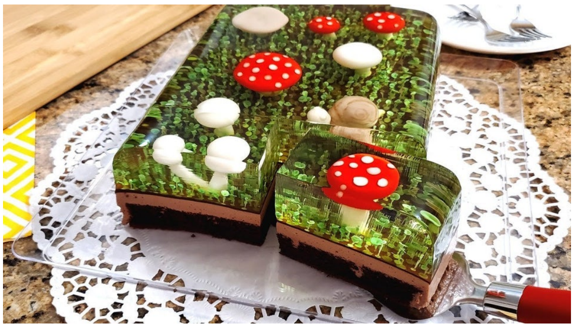
4.1.1 Close-up Images

Close-up images seemed to cause false-positives of an interesting calibre. One image in particular that caused interesting results was that of an American $100 bill, specifically a close-up of Benjamin Franklin's face (Figure 1). Interestingly, each of the three algorithms was not able to label this as anything related to currency. Instead, it determined that the image was that of a drawing (in the case of AWS Rekognition), or an image of an animal (in the case of Azure Computer Vision and Google Vision). We anticipate that this was due to the style in which the portrait of Benjamin Franklin is portrayed on the bill - particularly, we anticipate that the curvature of the lines around the nose and eye, and the portrait itself, are what caused the high-confidence in false-positive labelling in each of the algorithms.

Another example of a close-up image that yielded interesting results was that of a cake (Figure 2). We anticipate that due to the colour scheme of the cake, as well as the close-up nature of the photo, Azure's Computer Vision algorithm was unable to determine that this was a cake, even though it is abundantly clear to the human-eye. Both Rekognition and Google Vision were able to determine, with at least 80% confidence, that this image contained a cake, along with other labels such as 'Dessert', and 'Food'.

The above mentioned images, as well as others tested as part of this project, appear to indicate that close-up images result in each of the tested APIs attempting to focus more on the contours, colour-schemes, and background objects in an image when assigning a label. Unspecific labels were more common with images of this category when labels were assigned via all three tested APIs, indicating that these algorithms aim to assign more vague labels to complex images that may not be of use if a human were to sift through these images in a large dataset. This

Figure 2: Close-up of a cake. (u/St0pX, 2020)



could lead to issues for potential customers if these pre-trained services are being used on data that fits this category.

4.1.2 Images with Multiple, Vastly-Different Objects

This category of images yielded results that tended to vary drastically in detail between the various APIs tested.

In one image of bird on a sidewalk (Figure 3), Amazon's Rekognition API was the only of the three tested to label the image as that of both an animal, and, more specifically, as a bird. The Computer Vision and Vision APIs both did not detect any animals or living creatures in the image, but, they did detect different facets of a typical road and sidewalk, such as asphalt, grass, and walkway.

One specific image that was interesting when testing this category was that of a hotdog in a beer (Figure 4). The Rekognition and Vision APIs only assigned labels to the image in association with food and beverages (both alcoholic and non-alcoholic), while the Computer Vision API assigned many more labels to the image pertaining to what it actually contained. Some of these labels included 'wheat beer', 'beer cocktail', and 'lager', amongst alcoholic beverage titles such as 'daquiri', 'Tom Collins', and others.

Through the specific images mentioned above, as well as other images that were tested, it appears that Microsoft's Computer Vision API appears to label images containing multiple, vastly-different objects with the most false-positive labels, as well as the least amount of labels in some cases. This was especially true when the picture contained multiple objects in a nature setting, indicating that wildlife and nature photos may not be Azure Computer Vision's strong-suit. The Rekognition and Vision API's did quite-well in these areas when assigning labels, although in some cases they also were unsuccessful in assigning labels to plants, wildlife, or "nature".

Figure 3: An image of a bird on a sidewalk. (Hendrycks et al., 2020)

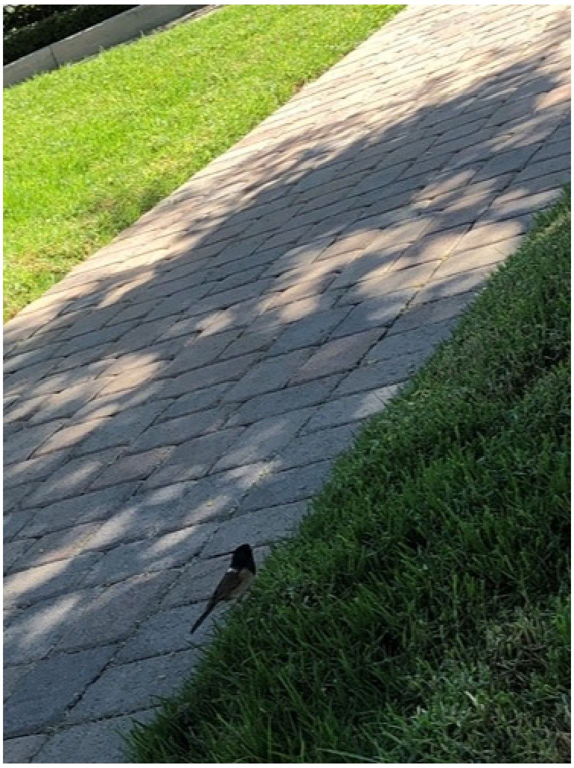
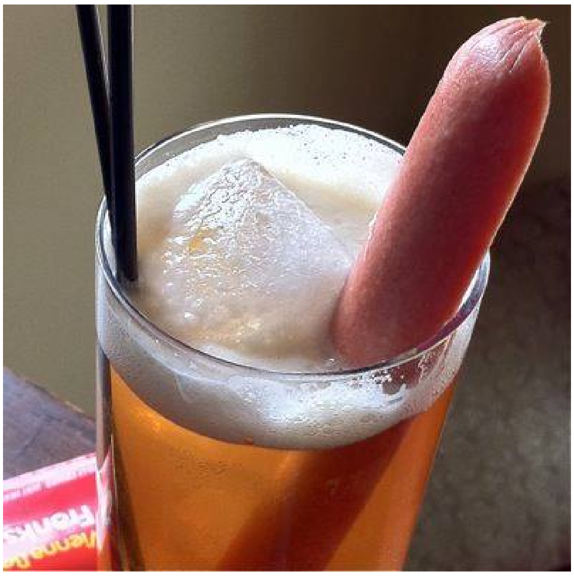


Figure 4: An image of a hotdog in a beer. (CBS, 2014)



4.1.3 Distorted or Odd Images

This category yielded some of the most interesting results that we perceived in this project. The first image that was tested was that of multiple corn cobs stacked in an alternating pattern in rows, as can be seen in Fig-

Figure 5: An image of stacked corn cobs (Hendrycks et al., 2020)

ure 5. Both the Azure Computer Vision and Computer Vision services were over 80% confident in assigning the label of 'orange' to this image, which, while appropriate to some degree, does not provide much context if these services are being used to associate image contents with possible search terms. The Rekognition service yielded somewhat-more appropriate labels, such as 'plant', 'food'. and 'fruit', while also providing the completely-incorrect label of 'pineapple' with 63% confidence. This was an interesting false-positive in that none of the tested services were able to assign a label that was pertaining to what the image actually contained.

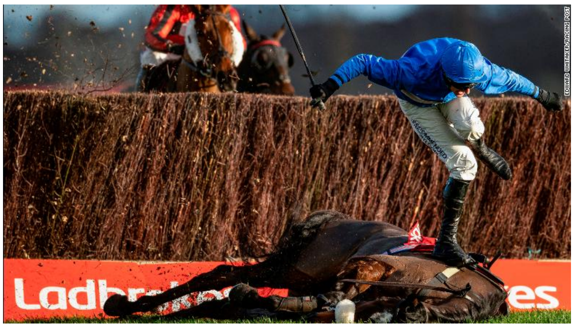
A second image of particular interest within this category was that of an equestrian accident containing a horse and a rider tumbling in a race (Figure 6). The Computer Vision and Vision services only assigned labels that did not pertain to the main objects within the image; yielding 'outdoor', 'text' and 'jumping; and 'stunt performer', 'soil', 'tree', 'competition event'. and 'games' respectively. Rekognition, on the other hand, assigned copious amounts of labels to the images, some that were of more context to the image such as 'animal', 'horse', and 'equestrian'.

This category made it abundantly clear that images containing objects in unusual forms, such as that of in an accident, in an unnatural context, or with interesting effects on them, tend to yield in high-confidence labelling that is incorrect in all of the APIs tested.

4.2 Ease-of-Testing

Through this project, we determined that testing the capabilities of each of these three services requires very little setup. Aside from gaining authentication credentials, and in the case of Google Vision, enabling the API itself, there were almost no prerequisites to testing these services aside from having an account on any of the three cloud providers that were

Figure 6: An image of a jockey and their horse falling (Whittaker, 2019). Both were safe after this fall!



utilized for this project. This means that future testing of the limitations and behavior of each of these APIs is a potential research topic for most anyone with the ability to set up an account on their preferred platform, which could lead to interesting results if vast amounts of time were to be spent on testing the false-positive labelling abilities of these platforms in their current state. This may in-turn lead to a deeper-understanding of how image-processing algorithms' deterministic abilities change over time, as these abilities could be measured as the algorithms undergo further training, thus, enhancing their precision when assigning labels to images.

5 FUTURE WORK

To improve the quality of our results, more time could have been spent attempting to find images that resulted in false-positives in any of the three image processing services tested in this project. This would increase the confidence in our analyses, and perhaps allow us to gain a deeper understanding of these "black box" algorithms that provide little-to-no reasoning for the results that they generate. In doing so, this may also reveal more categories of images that these algorithms do not label properly, or at all.

If work were to continue, the logical next steps for improvement would be to expand the scope of the project to investigate more than just the labelling capabilities of these services. All three of these services also provide object-detection results when they're used, which would be an interesting area to focus on alongside the labelling capabilities. Some of the tools used also provide the functionality of adding captions to photos, which was not utilized in this project's current state in order to compare the three services as equally as possible. However, by using these three abilities (labelling, captioning, and object detection) where applicable, a much more complex analysis of these three services could be performed, perhaps identifying a broader category of edge-cases where all three services seem to fail, which would be extremely valuable for potential customers of these services to understand.

6 TEAM MEMBER CONTRIBUTIONS

6.1 Christian Cornelis

- Setup of repo for all code and structuring of data for ease-of-use by all APIs
- Setup of GCP Student Account and associated billing account
- Google Vision API Python Script
- Analysis of findings
- Data aggregation for analysis
- Report
- Google Vision demo
- Presentation

6.2 Samuel Tracz

- Setup of Azure Cognitive Service
- Azure Computer Vision API Python Script
- Analysis of findings
- Report
- Azure Computer Vision demo
- Presentation

6.3 Mohammadamin Sheikhtaheri

- Setup of Amazon Rekognition service
- AWS Amazon Rekognition API Python Script
- Data collection
- Analysis of findings
- Report
- Rekognition demo
- Presentation

REFERENCES

Amazon (2020). Getting started with amazon rekognition. https://docs.aws.amazon.com/rekognition/latest/dg/labels-detect-labels-image.html/.

CBS (2014). The future is here: A minor league team will reportedly sell this hot-dog flavored cocktail. https://twitter.com/CBSSports/status/451714785264951298.

Farley, P. (2020). Quickstart: Analyze a local image - rest, python - azure cognitive services. https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/quickstarts/python-disk.

Google (2020). Quickstart: Setup the vision api — cloud vision api — google cloud. https://cloud.google.com/vision/docs/setup#windows.

Hendrycks, D., Zhao, K., Basart, S., Steinhardt, J., and Song, D. (2020). hendrycks/natural-adv-examples. https://github.com/hendrycks/natural-adv-examples.

Reshetnyak, V. (2016). Full frame shot of eye. www.pexels.com/photo/full-frame-shot-of-eye-251287/.

u/St0pX (2020). r/pics - forest floor mushroom jelly cake. https://www.reddit.com/r/pics/comments/fyh9vc/forest_floor_mushroom_jelly_cake/.

Whittaker, E. (2019). Stunning shots from horse racing's photographer of the year. https://edition.cnn.com/2019/12/05/sport/gallery/horse-racing-photos-ed-whitaker-winning-post-spt-intl/index.html.

APPENDIX

Appendix 1: Table of responses when running the various images we found through each respective image processing API. The images referenced can be found in our repo in the `resources` folder.

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
bill_eye.jpg	cat (98.66% confidence) animal (98.01% confidence) sketch (96.79% confidence) drawing (96.03% confidence) reptile (94.35% confidence)	Drawing (98.00% Confidence) Parents: Art Art (98.00% Confidence) Sketch (93.73% Confidence) Parents: Drawing Art	Elephant (96.93% confidence) Skin (96.50% confidence) Close-up (94.74% confidence) Eye (94.48% confidence) Wrinkle (93.47% confidence) Elephants and Mammoths (91.34% confidence) Nose (84.20% confidence) Organ (83.25% confidence)
bird_on_sidewalk.jpg	grass (99.99% confidence) outdoor (99.99% confidence) ground (98.48% confidence) plant (91.05% confidence) sidewalk (82.69% confidence) walkway (16.62% confidence) curb (9.10% confidence)	Path (99.68% Confidence) Walkway (98.80% Confidence) Parents: Path Bird (98.63% Confidence) Parents: Animal Animal (98.63% Confidence) Pavement (91.29% Confidence) Parents: Path Sidewalk (91.29% Confidence) Parents: Path Roof (73.85% Confidence) Slate (68.36% Confidence) Grass (59.50% Confidence) Parents: Plant Plant (59.50% Confidence) Cobblestone (56.07% Confidence) Parents: Path Walkway Sidewalk	Grass (95.82% confidence) Lawn (89.51% confidence) Grass family (80.71% confidence) Shadow (79.75% confidence) Asphalt (78.37% confidence) Road surface (73.28% confidence) Sidewalk (69.33% confidence) Path (69.17% confidence) Tree (68.69% confidence) Walkway (67.54% confidence)
cake.jpg			

Image	**Azure Computer Vision Labels**	**AWS Rekognition Labels**	**Google Vision Labels**
camera.jpg	table (99.89% confidence) Christmas tree (88.91% confidence) plate (87.73% confidence) decoration (85.09% confidence) slice (61.28% confidence) Christmas (60.48% confidence) sliced (25.45% confidence)	Plant (92.10% Confidence) Dessert (85.70% Confidence) Parents: Food Cake (85.70% Confidence) Parents: Dessert Food Food (85.70% Confidence) Birthday Cake (85.70% Confidence) Parents: Dessert Cake Food Candle (84.60% Confidence) Fork (76.45% Confidence) Parents: Cutlery Cutlery (76.45% Confidence) Egg (75.32% Confidence) Parents: Food Fungus (62.48% Confidence) Torte (60.34% Confidence) Parents: Dessert Cake Food Agaric (59.45% Confidence) Parents: Plant Fungus Mushroom Mushroom (59.45% Confidence) Parents: Plant Fungus	Food (82.71% confidence) Cake (81.78% confidence) Cuisine (78.64% confidence) Dessert (75.25% confidence) Chocolate cake (73.45% confidence) Petit four (69.41% confidence) Dish (67.61% confidence) Baked goods (61.26% confidence) Recipe (60.01% confidence) Torte (54.55% confidence)

Image	**Azure Computer Vision Labels**	**AWS Rekognition Labels**	**Google Vision Labels**
cartoon_face.jpg	text (94.09% confidence) map (87.90% confidence)	Electronics (97.28% Confidence) Camera (93.76% Confidence) Parents: Electronics GPS (84.33% Confidence) Parents: Electronics Monitor (79.70% Confidence) Parents: Electronics Screen Screen (79.70% Confidence) Parents: Electronics Display (79.70% Confidence) Parents: Electronics Screen Mobile Phone (71.10% Confidence) Parents: Phone Electronics Cell Phone (71.10% Confidence) Parents: Phone Electronics Phone (71.10% Confidence) Parents: Electronics Video Camera (64.88% Confidence) Parents: Camera Electronics Digital Camera (56.19% Confidence) Parents: Camera Electronics	Cameras & optics (95.94% confidence) Camera (94.76% confidence) Electronics (94.69% confidence) Technology (85.39% confidence) Digital camera (82.84% confidence) Electronic device (81.64% confidence) Screen (78.17% confidence) Gps navigation device (77.23% confidence) Camera accessory (72.89% confidence) Automotive navigation system (68.18% confidence)

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
	text (95.71% confidence) cartoon (94.50% confidence) drawing (90.10% confidence) child art (81.85% confidence) painting (80.36% confidence) sketch (74.93% confidence)	Head (97.14% Confidence) Hair (93.30% Confidence) Black Hair (93.30% Confidence) Parents: Hair Art (93.16% Confidence) Person (75.50% Confidence) Human (75.50% Confidence) Label (57.33% Confidence) Parents: Text Text (57.33% Confidence)	Face (97.17% confidence) Cartoon (95.86% confidence) Eyebrow (88.58% confidence) Head (87.62% confidence) Forehead (87.55% confidence) Cheek (83.80% confidence) Chin (82.75% confidence) Black hair (71.52% confidence) Jaw (63.02% confidence) Smile (54.04% confidence)
corn.png	orange (92.49% confidence) outdoor (89.15% confidence) outdoor object (33.00% confidence) several (12.55% confidence) colored (10.15% confidence)	Plant (97.12% Confidence) Food (88.52% Confidence) Fruit (88.52% Confidence) Parents: Food Plant Pineapple (63.04% Confidence) Parents: Food Plant Fruit	Orange (81.77% confidence)
gryphon_statue_group_picture.png	outdoor (99.95% confidence) sky (99.95% confidence) tree (99.73% confidence) clothing (94.84% confidence) person (92.82% confidence) footwear (86.99% confidence) man (83.30% confidence) sculpture (80.19% confidence) jeans (76.41% confidence) woman (68.05% confidence) statue (59.69% confidence) stone (15.05% confidence)	Human (99.85% Confidence) Person (99.85% Confidence) Shoe (98.69% Confidence) Parents: Footwear Clothing Clothing (98.69% Confidence) Apparel (98.69% Confidence) Footwear (98.69% Confidence) Parents: Clothing Pants (96.40% Confidence) Parents:	Statue (96.80% confidence) Sculpture (92.89% confidence) Monument (86.52% confidence) Tourism (79.41% confidence) Leisure (75.98% confidence) Vacation (73.13% confidence) Art (70.89% confidence) Tree (68.69% confidence) Recreation (59.43% confidence) Park (54.11% confidence)

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
		Clothing Tree (91.95% Confidence) Parents: Plant Plant (91.95% Confidence) Abies (90.69% Confidence) Parents: Tree Plant Fir (90.69% Confidence) Parents: Tree Plant Mammal (89.00% Confidence) Parents: Animal Elephant (89.00% Confidence) Parents: Animal Mammal Wildlife Animal (89.00% Confidence) Wildlife (89.00% Confidence) Parents: Animal Inflatable (82.15% Confidence) Conifer (78.15% Confidence) Parents: Tree Plant Jeans (70.45% Confidence) Parents: Clothing Pants Denim (70.45% Confidence) Parents: Clothing Pants	

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
		Building (68.54% Confidence) Architecture (68.54% Confidence) Parents: Building Art (59.08% Confidence) Sculpture (59.08% Confidence) Parents: Art	
horse_rider_accident.jpg	outdoor (97.55% confidence) text (97.41% confidence) jumping (54.40% confidence)	Human (99.04% Confidence) Person (99.04% Confidence) Nature (92.15% Confidence) Outdoors (91.15% Confidence) People (87.51% Confidence) Parents: Person Clothing (84.76% Confidence) Apparel (84.76% Confidence) Soil (75.01% Confidence) Symbol (73.56% Confidence) Arrow (73.56% Confidence) Parents: Symbol Countryside (73.06% Confidence) Parents: Outdoors Nature Land (72.80% Confidence) Parents: Outdoors Nature Animal (71.50% Confidence) Face (68.55% Confidence) Parents: Person Mammal (64.11% Confidence) Parents: Animal	Stunt performer (74.96% confidence) Soil (69.82% confidence) Tree (68.69% confidence) Competition event (56.94% confidence) Games (52.12% confidence)

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
		Horse (62.67% Confidence) Parents: Mammal Animal Field (62.65% Confidence) Water (61.41% Confidence) Hunting (59.16% Confidence) Parents: Person Team (57.77% Confidence) Parents: People Person Team Sport (57.77% Confidence) Parents: Team Sport People Person Polo (57.77% Confidence) Parents: Person Equestrian People Team Sport Team Sport Animal Mammal Horse Equestrian (57.77% Confidence) Parents: Mammal Horse Animal Person Sports (57.77% Confidence) Parents:	

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
		Person Sport (57.77% Confidence) Parents: Person Rural (56.16% Confidence) Parents: Outdoors Nature Countryside Ice (55.82% Confidence) Parents: Outdoors Nature Leisure Activities (55.38% Confidence)	
hotdog_beer.jpg	cup (99.99% confidence) soft drink (99.82% confidence) cocktail (99.06% confidence) table (98.03% confidence) juice (97.47% confidence) food (96.42% confidence) beverage (93.22% confidence) coffee (91.75% confidence) drink (88.86% confidence) glass (86.08% confidence) beer (84.97% confidence) cocktail garnish (81.31% confidence) classic cocktail (77.05% confidence) non-alcoholic beverage (75.97% confidence) sour (74.41% confidence) beer glass (73.95% confidence) paradise (73.30% confidence) pint glass (70.75% confidence) fizz (70.65% confidence) alcoholic beverage (69.91% confidence) zombie (67.45% confidence) wheat beer (66.52% confidence)	Beverage (97.34% Confidence) Drink (97.34% Confidence) Alcohol (97.04% Confidence) Parents: Beverage Beer (96.07% Confidence) Parents: Alcohol Beverage	Drink (96.27% confidence) Food (92.74% confidence) Non-alcoholic beverage (91.19% confidence) Batida (86.17% confidence) Alcoholic beverage (78.83% confidence) Beer cocktail (76.27% confidence) Juice (73.50% confidence) Distilled beverage (72.66% confidence) Bay breeze (72.14% confidence) Ingredient (70.07% confidence)

Image	Azure Computer Vision Labels	AWS Rekognition Labels	Google Vision Labels
	pint (65.26% confidence) fuzzy navel (64.91% confidence) punch (64.65% confidence) orange drink (64.35% confidence) long island iced tea (62.81% confidence) highball glass (62.10% confidence) aguas frescas (59.91% confidence) planter's punch (57.85% confidence) beer cocktail (57.73% confidence) highball (57.68% confidence) harvey wallbanger (56.82% confidence) drinking straw (56.36% confidence) bellini (56.26% confidence) tom collins (55.80% confidence) mai tai (55.20% confidence) painkiller (54.22% confidence) agua de valencia (53.79% confidence) daiquiri (53.73% confidence) bloody mary (53.28% confidence) apple cider (52.12% confidence) bia hi (52.02% confidence) lager (51.17% confidence) aparitif (51.10% confidence) rum swizzle (51.02% confidence) greyhound (50.87% confidence) screwdriver (50.83% confidence) alcohol (35.10% confidence) fruit drink (32.71% confidence)		
jellyfish.jpg	reef (99.68% confidence) marine invertebrates (97.88% confidence) aquarium (96.78% confidence) fish (86.56% confidence) fungus (80.98% confidence) organism (74.35% confidence) animal (66.84% confidence) underwater (65.58% confidence)	Fungus (97.32% Confidence) Sea Life (94.69% Confidence) Parents: Animal Animal (94.69% Confidence) Invertebrate (92.89% Confidence) Parents: Animal	Nature (95.35% confidence) Mushroom (92.61% confidence) Fungus (91.03% confidence) Organism (87.41% confidence) Agaricaceae (87.15% confidence) Agaricomycetes (83.81% confidence) Edible mushroom (82.00% confidence) Water (80.73% confidence)

Image	**Azure Computer Vision Labels**	**AWS Rekognition Labels**	**Google Vision Labels**
	invertebrate (62.87% confidence) marine biology (61.25% confidence) mushroom (57.52% confidence)	Jellyfish (57.43% Confidence) Parents: Invertebrate Sea Life Animal	Wildlife (78.65% confidence) Terrestrial animal (74.44% confidence)
shiny_legs.png	nail (89.54% confidence) finger (70.70% confidence) vein (68.53% confidence) toe (67.62% confidence) hand (62.32% confidence) limb (58.74% confidence) feet (28.97% confidence)	Skin (99.96% Confidence) Heel (99.13% Confidence) Person (85.01% Confidence) Human (85.01% Confidence)	Skin (95.96% confidence) Leg (89.54% confidence) Tan (89.44% confidence) Thigh (88.99% confidence) Finger (87.41% confidence) Nail (86.14% confidence) Joint (85.65% confidence) Human leg (84.89% confidence) Knee (81.03% confidence) Close-up (78.43% confidence)