Visual AI Webinar

Using Image Insights to Drive Modeling Decisions

July 21, 2020
1. Goals of the Webinar
2. Case Studies
   a. Case 1 - Simple Example
   b. Case 2 - Activation Maps
   c. Case 3 - Embeddings
   d. Case 4 - Baseline Model
   e. Case 5 - Layer Levels
3. Questions
4. Catered Lunch
   a. For those that provided address for delivery
Goals of the Session

1. You are able to use Image Insights in DataRobot to *identify* scenarios where your data or models could be improved

2. You have an idea of what *action* to take next to improve them

3. You will stay *awake* for the next 45 minutes (stretch goal)
Case 1: Grocery Items

Your grocery store chain wants to know which produce is running out of stock throughout the day when it happens. You want to use Visual AI with a system of in-store cameras to accomplish this.
You work with agricultural growers that need to monitor weeds that are threatening their crops. You want to use Visual AI to build systems that can estimate the prevalence of weeds in a plot of land. You might even want to build a sensor that you can drag along the ground that will spray pesticide if it identifies a weed.
**Observation**

My model’s confusion matrix identifies a few commonly confused classes

My model appears to be overfitting to features of the image unrelated to the class

**Action**

Examine the activation maps of the model on those classes to understand what the model is using to make a prediction

Crop your photos appropriately if possible. Consider training a model to identify the photos that contain the unrelated feature and removing them from the training set.
Your department needs to quickly process car insurance claims and sort them by how much manual review they’ll need. You’re interested in using Visual AI to quantify damage, so you want to run a pilot to see how well Visual AI does at classifying damaged cars from undamaged ones.
Observation

My embedding clearly separates my images into two classes.

There are differences in images from each class unrelated to the problem (such as how the image was taken).

Action

Examine the images close to the boundary of the two clusters and at either extreme to gather possible explanations.

Resample your images to account for this imbalance, or crop them appropriately if possible.
Case 4b: Counting Skittles

Your manufacturing plant assembles bags of hundreds of small items, like pieces of candy, chemical pellets, or screws. You currently weigh the bags to estimate quantity, but you want Visual AI to count the number of items before they’re bagged as an extra quality control measure.
Observation

The Baseline Image Model does nearly as well as the more advanced models.

There's a part of the image (or the whole image) whose brightness correlates with the class.

Action

Examine the activation maps of the models to understand what parts of the image they are using to make a prediction.

Decide whether this is a blessing or a data quality problem, depending on your use case.
You're working with a government utility that wants to survey hard-to-reach rural areas to map out the precise location of power lines and phone lines that are suspected to be damaged. You want to use Visual AI on photos taken from cameras aboard low-flying aerial vehicles to identify these locations.
A Closer Look at a Convolutional Neural Network

Pretrained Featurizer

Low-level features → Medium-level features → High-level features → Highest-level features → Concatenated feature vector (usually 1000+ numbers)

Traditional DR modeler learns from features
Observation

The activation maps don’t show what the model *should* be using to make predictions.

Adjusting the layer levels still didn’t seem to produce the scores I would have expected.

Action

Tune the model to use the layer level of features you expect to be the most useful.

Try a more powerful network architecture capable of extracting more signal from the features.
Goals of the Session - Recap

1. You are able to use Image Insights in DataRobot to *identify* scenarios where your data or models could be improved
   - The confusion matrix reveals a pair of commonly confused classes
   - Activation maps are not showing what you expected
   - Image Embeddings reveal separation based on features that shouldn’t be correlated with the target
   - The Baseline Model scores better than you expect

2. You have an idea of what *action* to take next to improve them
   - Resample your training data
   - Crop / Preprocess your training data more
   - Tune with alternate settings for layer levels

3. If you are reading this, you are currently *awake*
Team Members

Ivan Pyzow
Deep Learning Engineer

Jack Jablonski
AI Success Manager

Rajiv Shah
Data Scientist
Questions?
Engage, learn, and accelerate your AI/ML journey
Connect with peers to find solutions to AI challenges
Explore helpful content to take your AI to the next level
Build your brand as an AI expert & thought leader
Join your peers today at community.datarobot.com

Questions: aisuccess-webinars@datarobot.com
Appendix
Image Activation Maps

- **Layer**
  - **Input**
    - (224, 224, 3)
    - (224, 224, 3)
    - (111, 111, 64)
  - **Output**
    - (224, 224, 3)
    - (111, 111, 64)
    - (55, 55, 64)

- **Input Image Size**: (224, 224, 3)
- **Convolutional Layer conv1**: 64 3x3
- **Pooling Layer pool1**: (111, 111, 64)
- **Fire Module fire2**: (55, 55, 64)
- **Fire Module fire3**: (55, 55, 128)
- **Pooling Layer pool3**: (55, 55, 128)
- **Fire Module fire4**: (27, 27, 128)
- **Fire Module fire5**: (27, 27, 256)
- **Pooling Layer pool5**: (27, 27, 256)
- **Fire Module fire6**: (13, 13, 256)
- **Fire Module fire7**: (13, 13, 384)
- **Fire Module fire8**: (13, 13, 384)
- **Fire Module fire9**: (13, 13, 512)

- **Global Average Pooling (GAP)**
  - (55, 55, 128) → 128
  - (27, 27, 256) → 256
  - (13, 13, 384) → 384
  - (13, 13, 512) → 512

- **Low-level features (128)**
- **Medium-level features (256)**
- **High-level features (384)**
- **Highest-level features (512)**
Image Activation Maps

Fully featurized images

Measure feature importance (SHAP or internal Feature Impact)

Ground truth

Weighted sum of layer activations and estimated feature importances

Postprocessing: smoothing, clipping
Image Embeddings

Layer:
- input image
- conv1 64x3x3
- pool1
- fire2
- fire3
- pool3
- fire4
- fire5
- pool5
- fire6
- fire7
- fire8
- fire9

Input:
- (224, 224, 3)
- (224, 224, 3)
- (111, 111, 64)
- (55, 55, 64)
- (55, 55, 128)
- (55, 55, 128)
- (27, 27, 128)
- (27, 27, 256)
- (27, 27, 256)
- (13, 13, 256)
- (13, 13, 384)
- (13, 13, 384)
- (13, 13, 512)

Output:
- (224, 224, 3)
- (224, 224, 3)
- (111, 111, 64)
- (55, 55, 64)
- (55, 55, 128)
- (55, 55, 128)
- (27, 27, 128)
- (27, 27, 256)
- (27, 27, 256)
- (13, 13, 256)
- (13, 13, 384)
- (13, 13, 384)
- (13, 13, 512)

Highest-level features (512) → TriMap dimensionality reduction (512D → 2D) → (X, Y) coordinates

Global Average Pooling
(13, 13, 512) → 512
Your teams need to survey large campuses and estimate the condition of the pavement and the concrete buildings quickly. You have a drone that can collect close-up photos, but you’d like Visual AI to assess those tens of thousands of photos instead of you, manually.

Case 4a: Concrete Cracks