RDS Deep Learning Series:
Visual Search and more

AI Ruevolution II

RGG Data Science

Outline

Image Recognition and Visual Search

Product tagging - Product cataloging

Image Embedding - Visual similarity

Hierarchy Automation

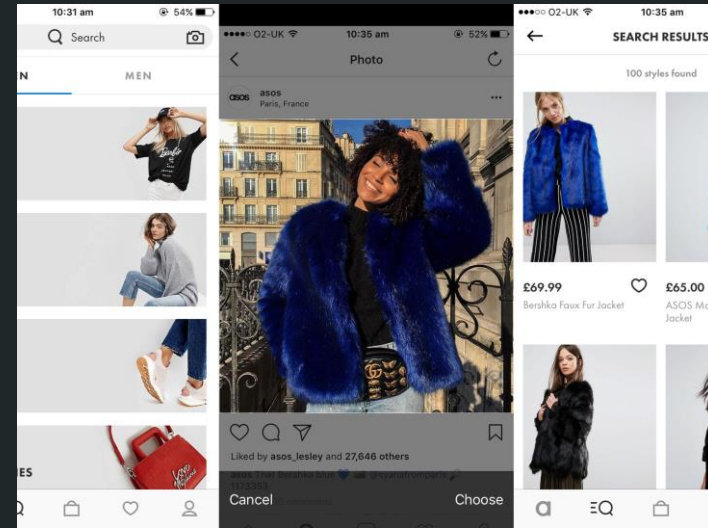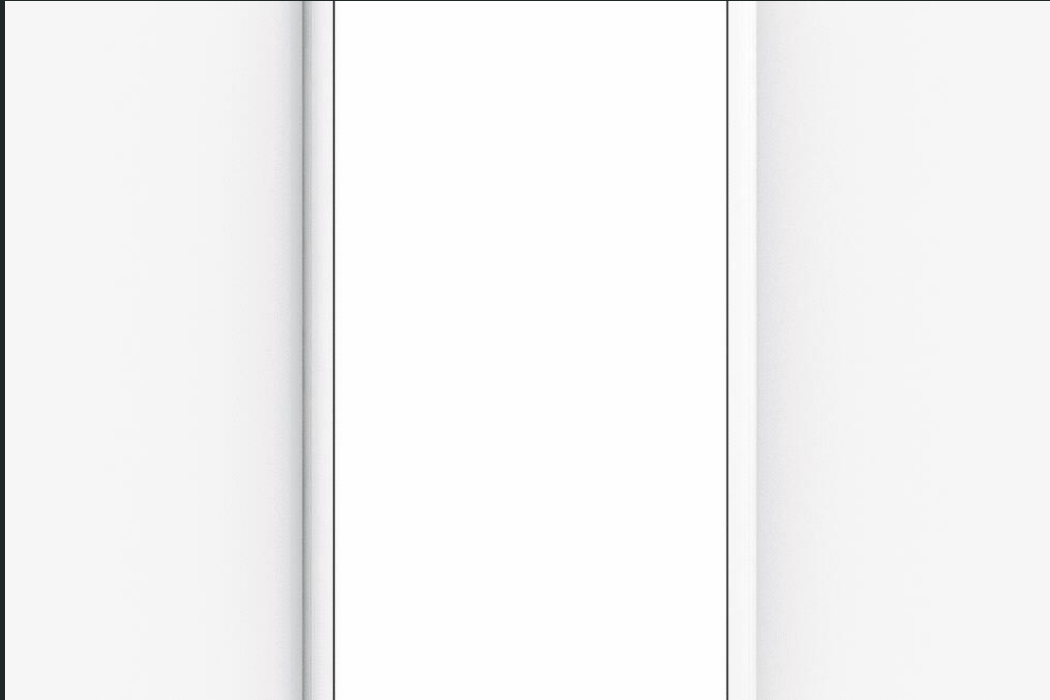Visual Search - Street to Shop

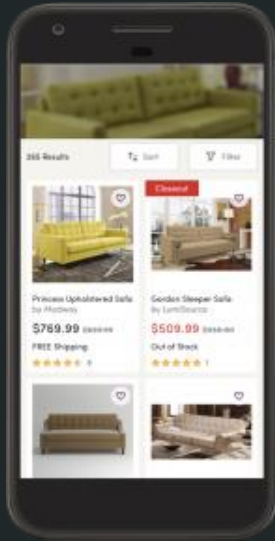Shop the look

Existing Work

Street to Shop

Pinterest

ASOS

Existing Work

Shop the look

Wayfair

Image Models Overview

Finds image groups from the image

Get vector and/or tags from the Image

Find similar products to the reference image group

Object Detection

Image Group Detector

Image Embedding

Generates Vectors

Image Tagging

Generates Tags

Real time Similarity Lookup

Produces image groups

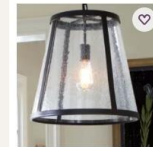*Image groups:*
Dress
Glasses
Shoes
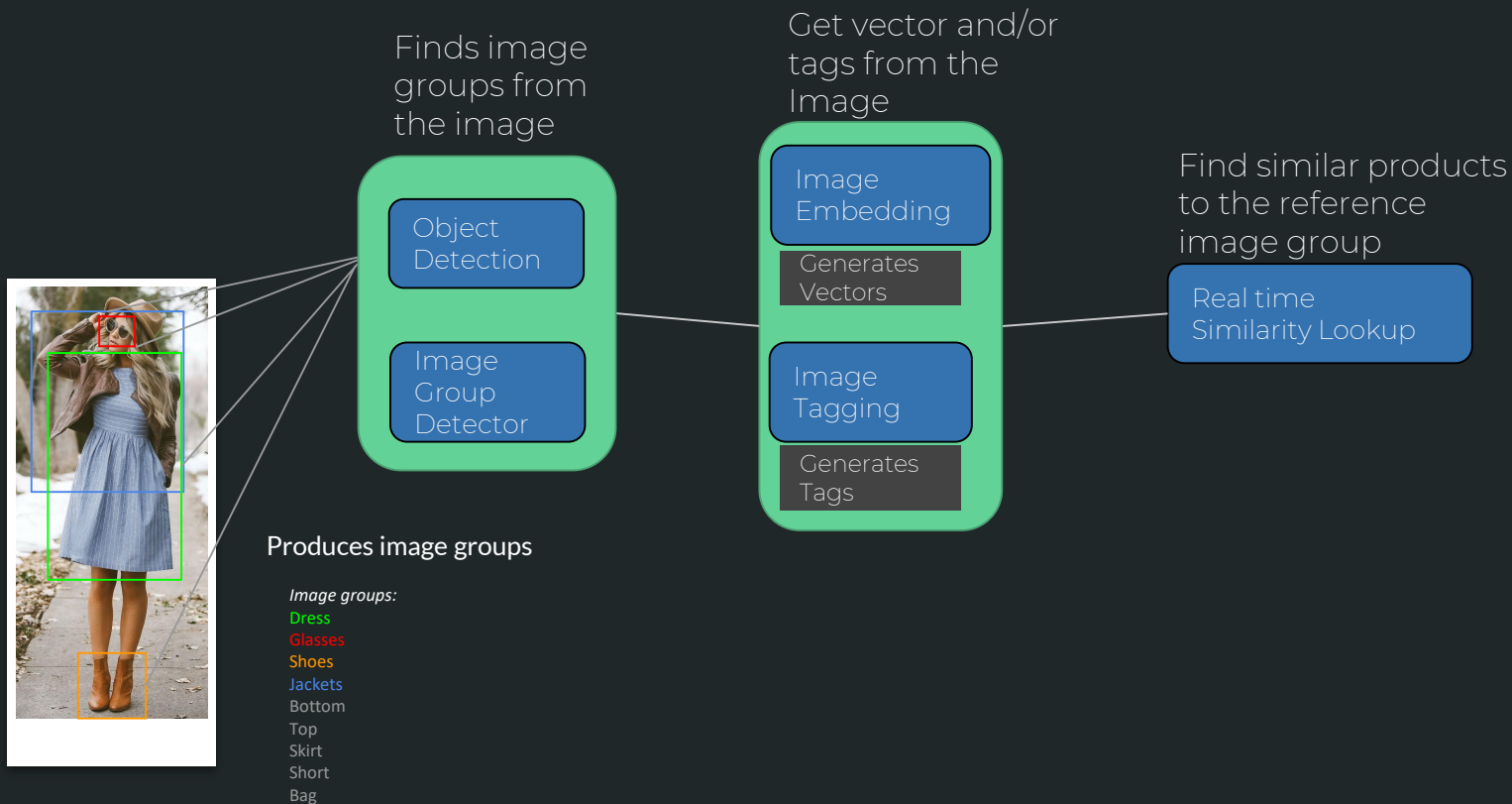Jackets
Bottom
Top
Skirt
Short
Bag

Image Groups

Mapping multiple rows from Hierarchy into new Classes

Will correspond to correct Class level in the hierarchy

	confidence	dept_tag	hierarchy	name_tag	name_tokens	rds_tag	style_id	style_nm	tag
59761	1.000000	None	Women_Clothing_Coats & Jackets_Leather	jacket	jacket	jacket	1411605097	Bagatelle Drape Jacket	jacket_women
84255	0.999993	shoes	Women_Shoes_Sandals_Wedges	shoes	sandal	None	1311668137	Donald Pliner Fifi Leather Sandal	shoes_women
55955	1.000000	jewelry	Women_Jewelry_Necklaces_Necklaces	jewelry	necklace	None	6030737500	Rachel Reinhardt Silver Labradorite & Chalcedo...	jewelry_women
10608	0.997483	jewelry	Women_Jewelry_Rings_Rings	jewelry	ring	None	6030505798	Rivka Friedman 18K Clad Crystal Ring	jewelry_women
7604	1.000000	jewelry	Women_Jewelry_Rings_Rings	jewelry	ring	None	6030030275	Diana M 14K Rose Gold 1.08 ct. tw. Diamond & T...	jewelry_women
33325	0.999780	None	Women_Clothing_Swimwear_Cover-Ups	dress	dress	inner	1411441223	Red Carter Cover-Up Dress	dress_women
106397	0.790925	shoes	Women_Shoes_Pumps_Pumps	shoes	pump	None	1311251028	Vince Camuto Tannen Suede Pump	shoes_women
107187	0.999721	None	Men_Clothing_Shirts_Dress	top	shirt	None	1010722682	Robert Graham Elm Canyon Classic Fit Woven Shirt	top_men
102979	0.999984	None	Men_Clothing_Sport Coats_Tailored	suit	blazer	None	1011586071	Ted Baker Mini Design Wool-Blend Blazer	jacket_men
118055	1.000000	None	Women_Clothing_Shorts_Shorts	short	short	shorts	1411535531	Juicy Couture Keep It Juicy Venice Track Short	short_women
86248	1.000000	jewelry	Women_Jewelry_Bracelets_Bracelets	jewelry	cuff	None	6030201116	Melinda Maria Plated CZ Cuff	jewelry_women

Image groups

Image Group Detection Model

Goal: given an image classify it into one of the following categories:

Image Groups: (dress, top, bottom, jacket, skirt, scarf, shoes, hat, bag, formal set, glasses, watch, jewelry)

Model: ResNet-50 trained on Rue Images w/ data augmentation. Accuracy: 97%

Input Image:

Output:

Predicted Image Group: Dress

Image Group Model

ResNet-50: Deep Residual Network (w/ 50 layers)

(1) Features skip (residual) connections for easier training

(2) Fewer parameters and higher classification accuracy on ImageNet

ImageNet accuracy:

method	top-5 err. (**test**)
VGG [41] (ILSVRC'14)	7.32
GoogLeNet [44] (ILSVRC'14)	6.66
VGG [41] (v5)	6.8
PReLU-net [13]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

Table 5. Error rates (%) of **ensembles**. The top-5 error is on the test set of ImageNet and reported by the test server.

skip connections:
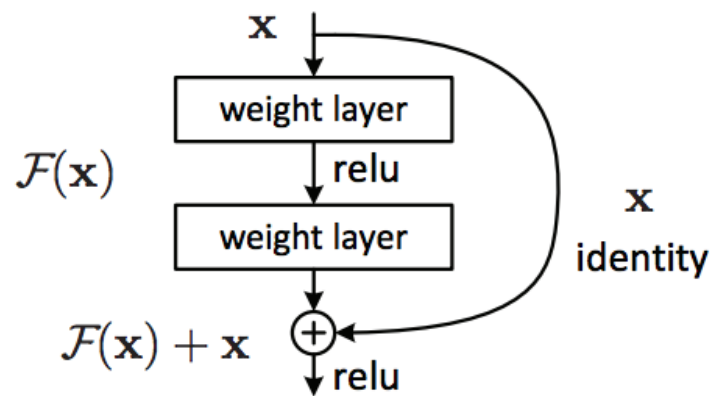
Object Detection

Classification & Regression:

Where is the top located in the image?

Model output:
class label + location
top location_data

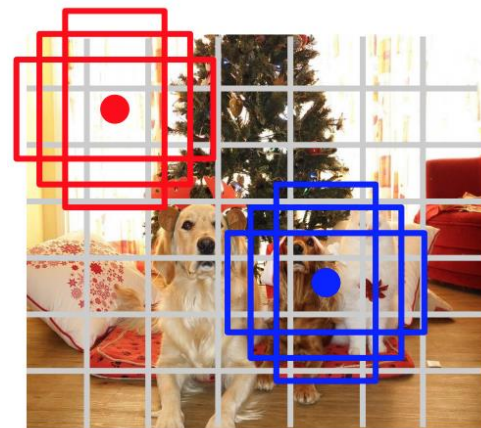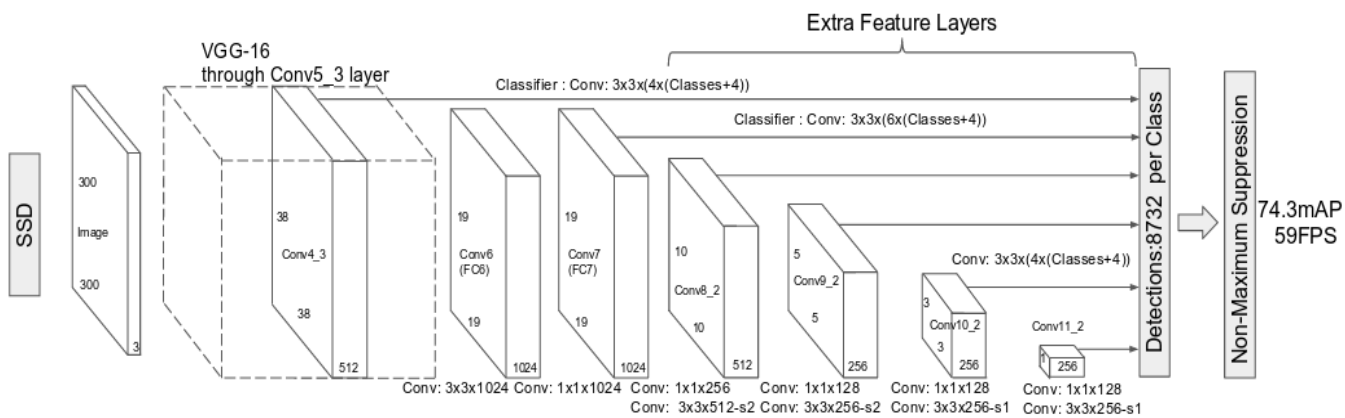**A vector for every object:**

Multiple fashion items per image!
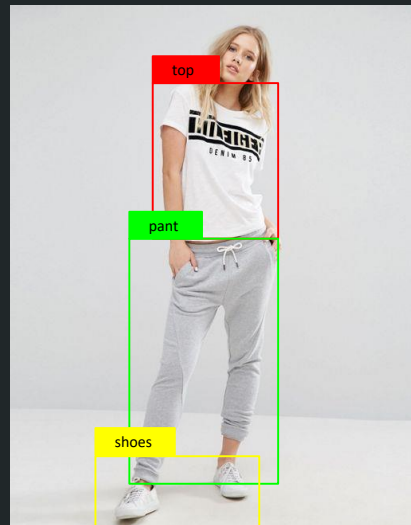
Model output:
top   location_data
pant   location_data
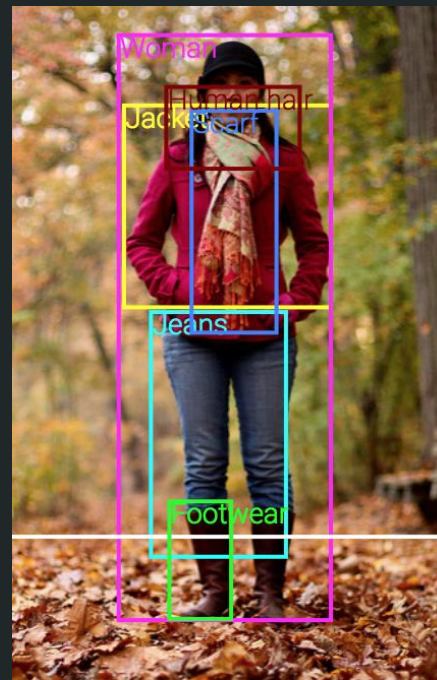*shoes   location_data*

Object Detection Model

SSD: Single-Shot Multi-Box Detector

(1) divide input image into a grid

(2) introduce (many) prior boxes

(3) detect objects for each prior box
(anchors)

(4) output class confidence + bounding box coordinates
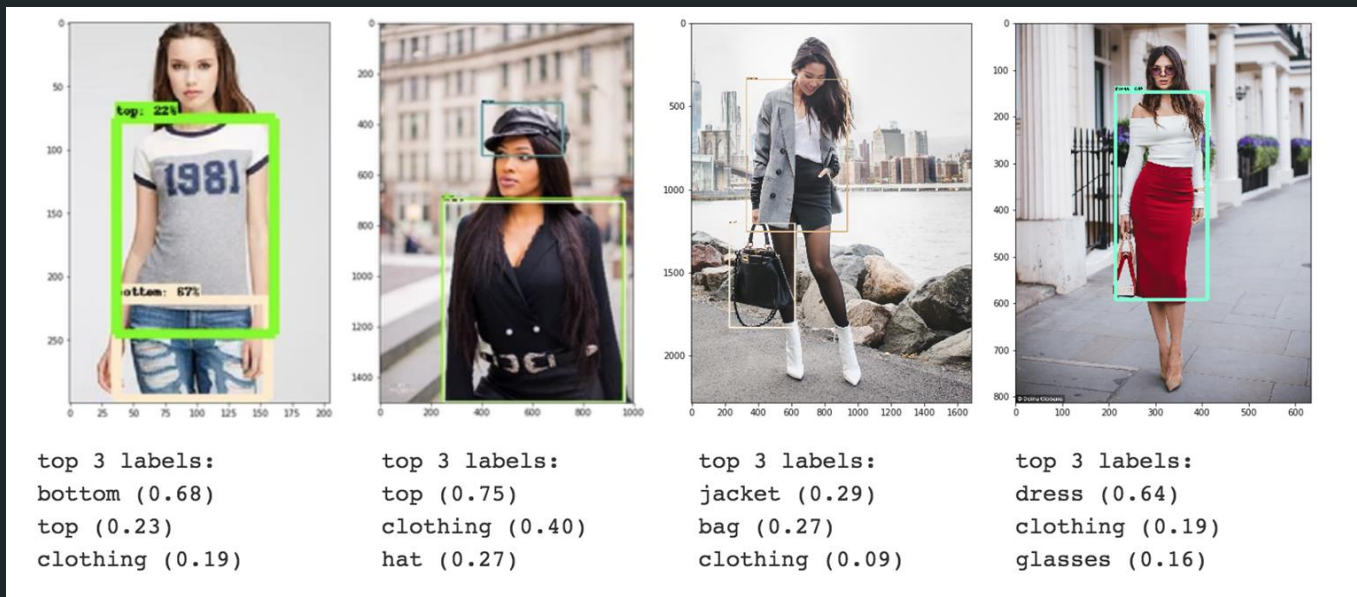
Open Images Dataset

(600K fashion images)
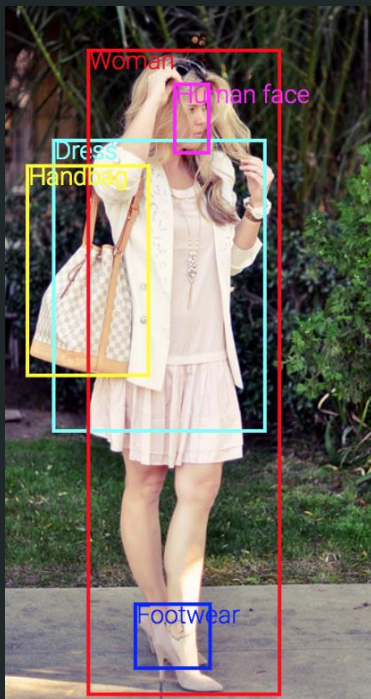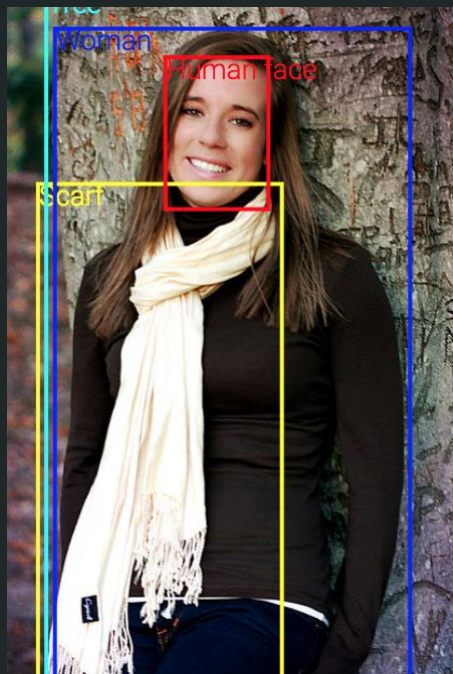
❏ Multiple bounding boxes per image ❏ Realistic images (background, occlusions, etc)

Object Detection Results

Model: TensorFlow object detection API Using **SSD** with Open Image dataset (92K images)

Tagging Models

Model: ResNet-50 trained with Vue.AI tags

NLP based tags for many groups without VueAI

One model for each image group

Challenge: Need more and better QAed tag labels
 for accuracy

Reference Image

```
Image Group Detected: "dress_women"

Tags Detected: {
    "dresslength": "Thigh",
    "dresstype": "ShiftSack",
    "neckline": "Crewneck",
    "patternuw": "Floral",
    "sleevelength": "Sleeveless"
}
```

Tagging Model - with vue.ai

Goal: tag category-specific attributes of the image based on the data labeled by vue.AI: (dress, top, bottom, skirt, shoes, bag)

Challenges: low tagging accuracy, some categories (jackets, glasses, jewelry, accessories) not covered, lack of labeled data for pants, bags and shoes.

Model: ResNet-50 trained on Rue Images w/ data augmentation. Accuracy: 50-90%

Input Image:

Output:
Predicted Image Group: Dress
Dress Type: Column
Dress Length: Ankle
Neckline: Plunge Neck
Pattern: Solid
Sleeve Length: Sleeveless

Tagging Model - with NLP

Goal: fill in the gaps - categories that are not tagged by VueAI (or tagged with the very low accuracy)

Approach: we have style names and product descriptions available for every product, we can extract these attributes w/ RegEx, TF-IDF, Word Embeddings, NER

Frame shape: *butterfly*

Frame color: *gold*

Lens color: *light burgundy gradient*

These frames flatter those with an oval, square, or triangle shaped face Plastic lenses with 100% UV protection

Lens: 63mm wide Bridge: 14mm wide Arms: 125mm long Case included; case size and color may vary Made in Italy

Knit khaki with **cropped sleeves** and faux leather belt
Jacket is approximately 22" in length
Notched collar
Three **button front closure** with distressed gold buttons
Princess seams
Contrasting stitching
Belted accent around the waist is approximately 2.25" wide
Cropped sleeves are approximately 18" in length
Fully lined interior with tan fabric

Model: ResNet-50 trained on Rue Images w/ data augmentation. Accuracy: 50-90%

Getting Similarity from CNN models

(1) Continuous space similarity:

    (a) extract image embeddings (vectors) from fully connected layer

    (b) compute distances between vectors and retrieve nearest neighbors

(2) Discrete space similarity: extract image labels at the output

    (a) extract categorical outputs (labels) from the final softmax layer

   (b) Compute exact matches and retrieved top-K similar candidates

(1) Extract output from this layer as an embedding vector

(2) Extract labels
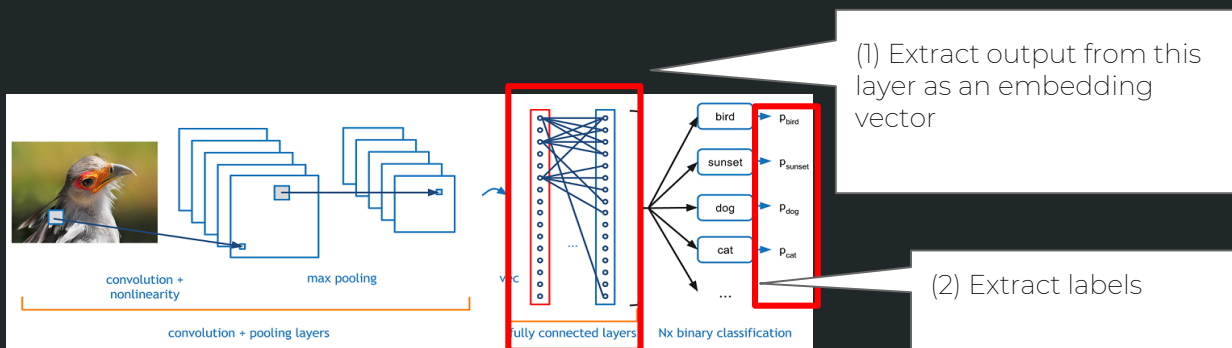
Real-time Similarity Lookup

Goal: Dynamic lookup, low latency

Solution: Approximate Nearest Neighbors (ANN) based on Locality Sensitive Hashing (LSH) (Takes 5ms to lookup 100,000 vectors)

(1) Randomly partition data space into regions (random projection LSH)
(2) Recurse / construct a tree until each leaf has at most K data points (nearest neighbors)
(3) Create many trees (a forest) that use slightly different random partitions
(4) Union root-to-leaf node traversals from multiple trees (to increase the chance of finding true K nearest neighbors)
(5) Use priority queue to merge results from multiple trees (where priority is a distance between the query point and the previously computed union of candidate points)

Demo

http://vsearch.ruetracking.com:3000/
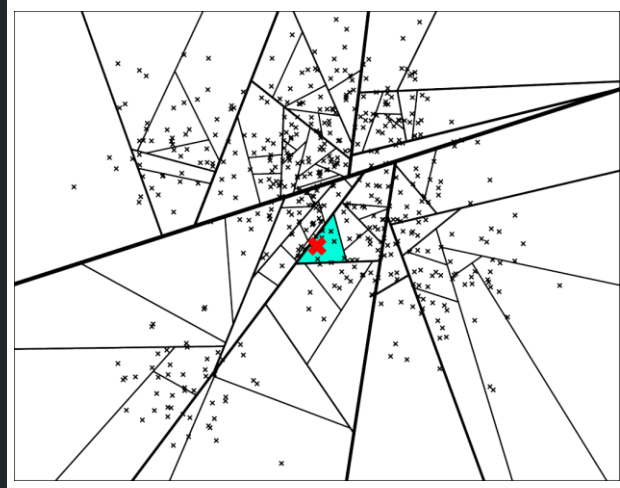
Applications

Codename: ARIA

Rue's First Ever AI

A - Automated
R - Realtime
I - Intelligent
A - Assistants

Visual Similarity

Visual Similarity

Current NLP based Model

Visual Similarity

Similar Images with non-white constant background

Reference Image

Hierarchy Automation

Existing hierarchy issues

Which one is it?

Hierarchy Automation

Automatically classify products into a new hierarchy (with Image and/or text)

Input Image:

Output from Image Group Model

Image Group: Dress

Output from Dress Tagging Model

Dress Type: a-line

Automatically mapped hierarchy

DIVISION: Women
DEPARTMENT: Clothing
CLASS: Dress
SUBCLASS: Day Dresses
CATEGORY: ALine

Hierarchy Automation

CatNav Now

All Women

Clothing

Dresses & Skirts

CatNav Could be

All Women

Clothing

Dresses

    A-line Dresses

    Shift Dresses

    Maxi Dresses

Attributes as filters

Tags Detected: {
    "dresslength": "Thigh",
    "dresstype": "ShiftSack",
    "neckline": "Crewneck",
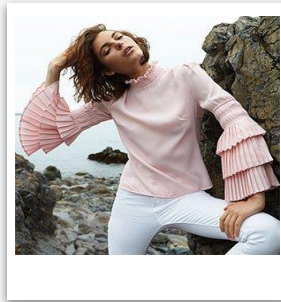   "patternuw": "Floral",
    "sleevelength": "Sleeveless"
}
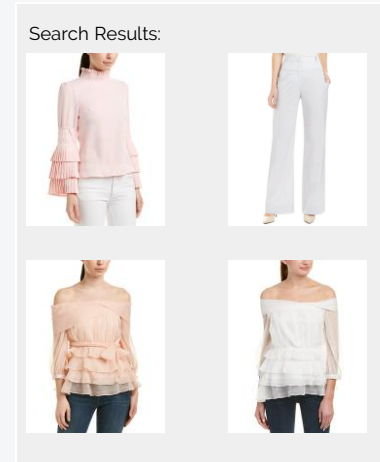
* Just an example

Visual Search

Search products by an Image

Users upload a picture and we provide recommendations based on the picture
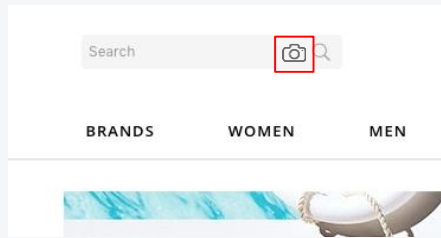
Search by Image button on Search bar

User uploads Image

Search Results based on the uploaded Image

Visual Navigation/Discovery

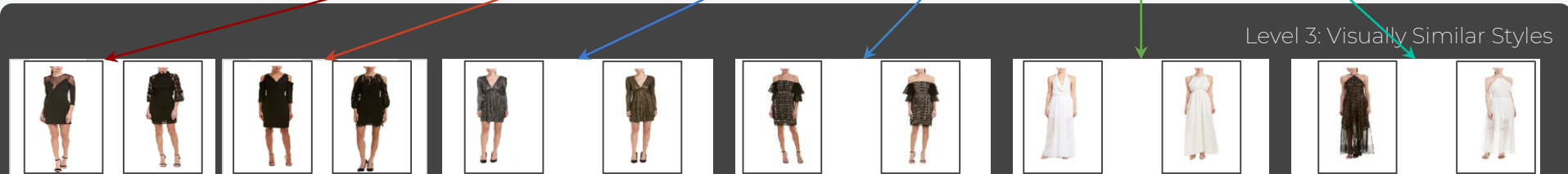Navigating our catalog through product images

Main Application in Voila

Powers Image based Streaming

Level 1: Distinct Styles

When user clicks on one of the above

Level 2:
Text based
Similar Styles

When user clicks on one of the above

Level 3: Visually Similar Styles

Shop the Look

Automatically detect objects in product shots and link to actual products

Women's Fedora
C Brand
Price: 30$

Similar products to these

Women's Round Sunglasses
A Brand
Price: 50$

Similar products to these

Some brand Leather Jacket
Some brand
Price: 100$

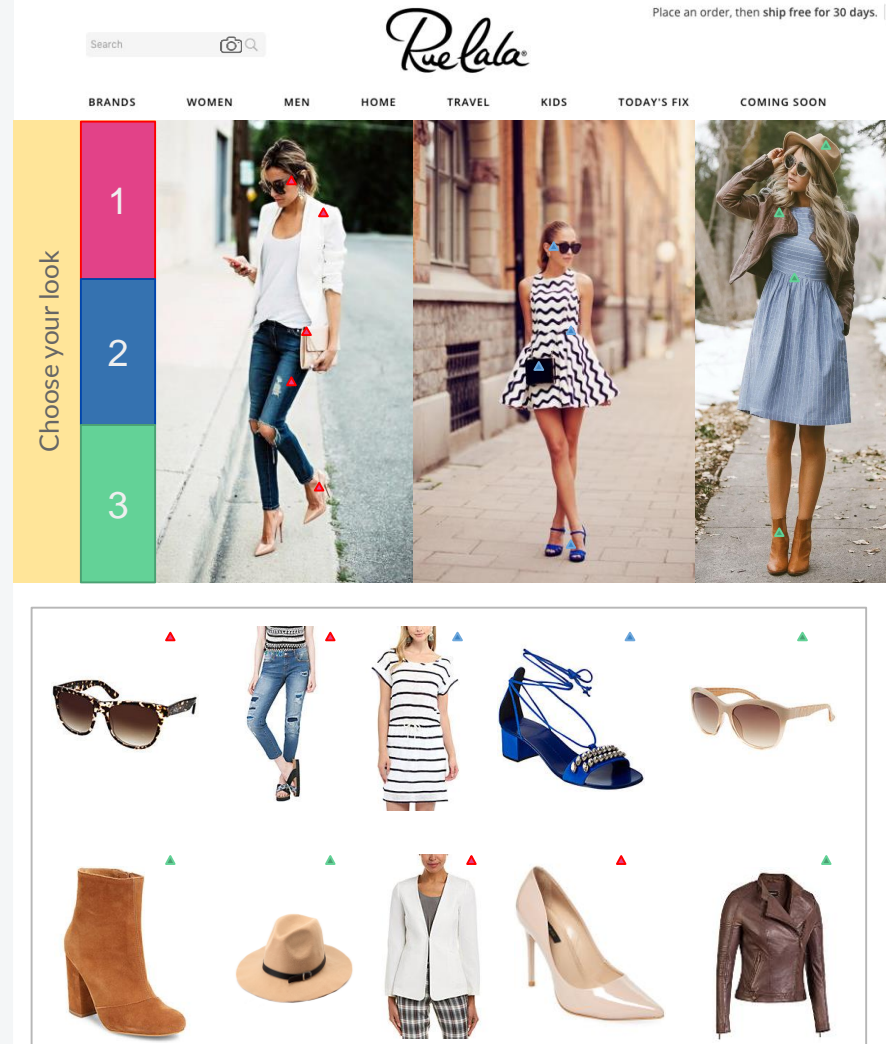**Similar products to these**

X Brand A-line Dress
X Brand
Price: 80$

Similar products to these

Leather Ankle Shoes
Z Brand
Price: 150$

Similar products to these

Shop the Look BTQ

Now Imagine that in a PLP or an exclusive interactive boutique

Also Integrate with Store Manager to simplify Boutique creation process and adding products to Boutiques
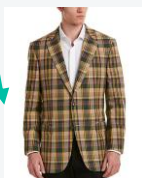
Instead of …
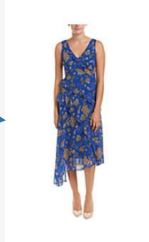
Shop the Look
Interactive Boutique

Boutique Main Impact Shot

Actual products in the Image

Similar products to the one before but with distinct styles or colors

Thank you!