Using Deep Learning to rank and tag millions of hotel images

15/11/2018 - PyParis 2018

Christopher Lennan (Senior Data Scientist)
@chris_lennan

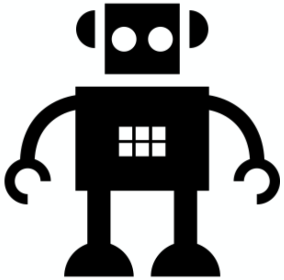
Tanuj Jain (Data Scientist)
@tjainn

#idealoTech

Agenda

idealo

1. idealo.de
2. Business Motivation
3. Models and Training
4. Image Tagging
5. Image Aesthetics
6. Summary



Some Key Facts

idealo

18 More than 18 years experience

Germany's 4th largest eCommerce website

Active in 6 different countries (DE, AT, ES, IT, FR, UK)

700 “idealos” from 40 nations

16 million users/month [1]

50.000 shops

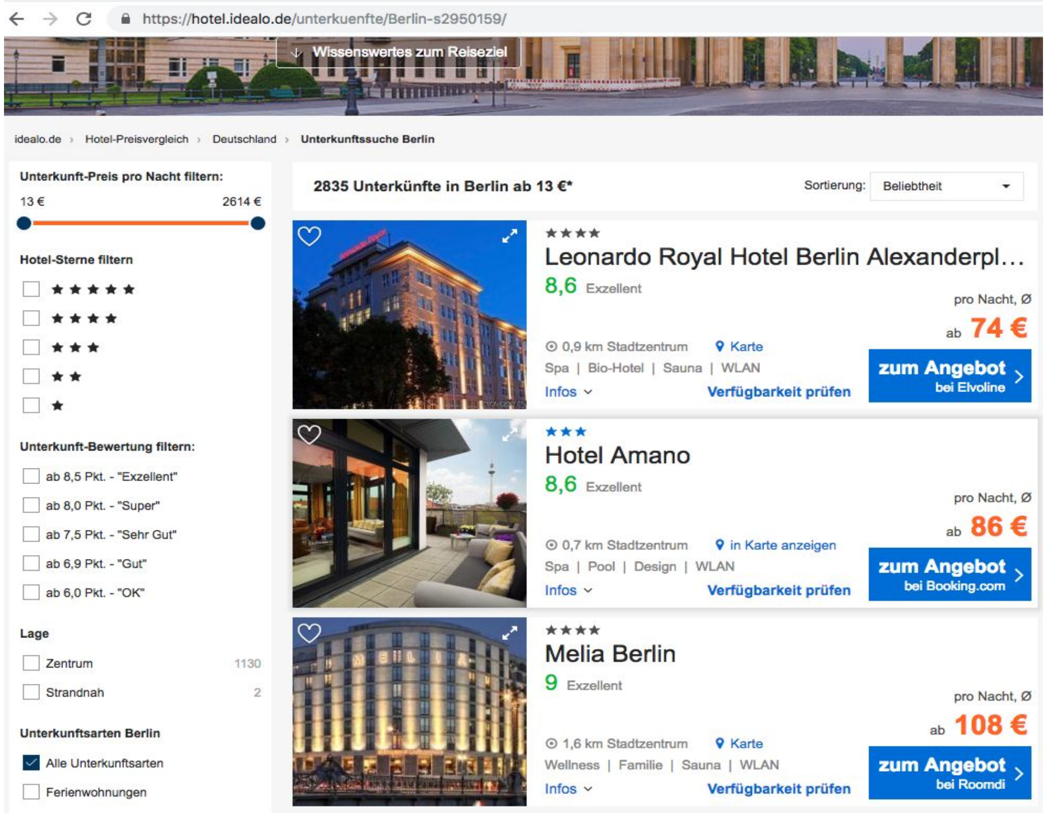
Over 330 million offers for 2 million products

Tüv certified comparison portal [2]

Motivation

idealo hotel price comparison

hotel.idealo.de



- 2.306.658 accommodations
- 308.519.299 images
- ~ 133 images per accommodation

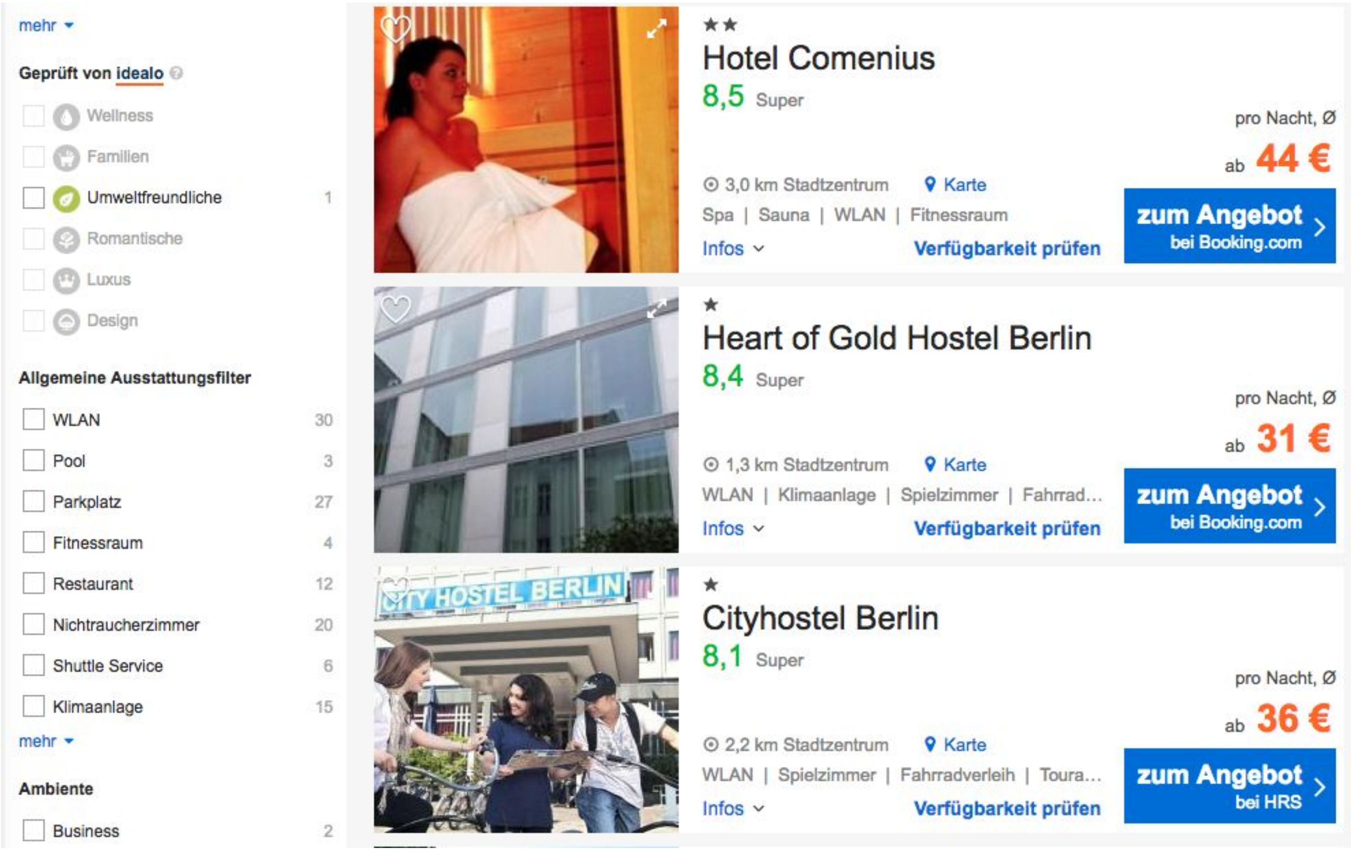
Importance of Photography for Hotels

idealo

"... after price, photography is the most important factor for travelers and prospects scanning OTA sites.."

".. Photography plays a role of 60% in the decision to book with a particular hotel .."

".. study published today by **TripAdvisor**, it would seem like **photos have the greatest impact** driving engagement from travelers **researching on hotel and B&B pages** .."


mehr
Geprüft von idealo
Wellness
Familien
Umweltfreundliche 1
Romantische
Luxus
Design
Allgemeine Ausstattungsfilter
WLAN 30
Pool 3
Parkplatz 27
Fitnessraum 4
Restaurant 12
Nichtraucherzimmer 20
Shuttle Service 6
Klimaanlage 15
mehr
Ambiente
Business 2
★★
Hotel Comenius
8,5 Super
pro Nacht, Ø
ab 44 €
3,0 km Stadtzentrum
Karte
Spa | Sauna | WLAN | Fitnessraum
Infos
Verfügbarkeit prüfen
zum Angebot bei Booking.com
★
Heart of Gold Hostel Berlin
8,4 Super
pro Nacht, Ø
ab 31 €
1,3 km Stadtzentrum
Karte
WLAN | Klimaanlage | Spielzimmer | Fahrrad...
Infos
Verfügbarkeit prüfen
zum Angebot bei Booking.com
CITY HOSTEL BERLIN
★
Cityhostel Berlin
8,1 Super
pro Nacht, Ø
ab 36 €
2,2 km Stadtzentrum
Karte
WLAN | Spielzimmer | Fahrradverleih | Toura...
Infos
Verfügbarkeit prüfen
zum Angebot bei HRS

idealo

★★

Hotel Comenius

8,5 Super

pro Nacht, Ø

ab 44 €

3,0 km Stadtzentrum   Karte

Spa | Sauna | WLAN | Fitnessraum

Infos   Verfügbarkeit prüfen

zum Angebot bei Booking.com

idealo

★★

Hotel Comenius

8,5 Super

pro Nacht, Ø

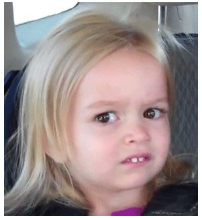
ab 44 €

3,0 km Stadtzentrum | Karte

Spa | Sauna | WLAN | Fitnessraum

Infos | Verfügbarkeit prüfen

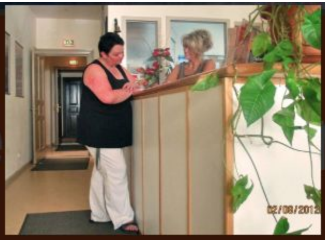
zum Angebot
bei Booking.com



1

2

3



4

5

6

7

8

9

10

11

12

13

Image Aesthetics

Current image placement

Position: 1

Position: 19

Image Aesthetics

Current image placement

Position: 3

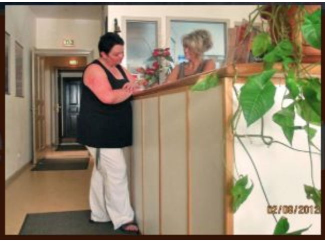
Position: 17

Beautiful images should appear earlier in the gallery

1

2

3



4

5

6

7

8

9

10

11

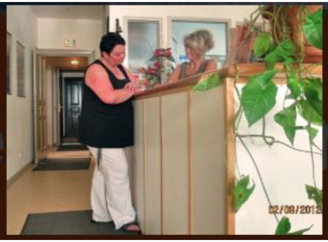
12

13

1

2

3



4

5

6

7

8

9

10

11

12

13

Ensure different areas get depicted

Bedroom

1

2

3

Bathroom

4

Restaurant

5

Facade

6

Fitness Studio

7

Kitchen

8

Understanding Image Content

Two part problem

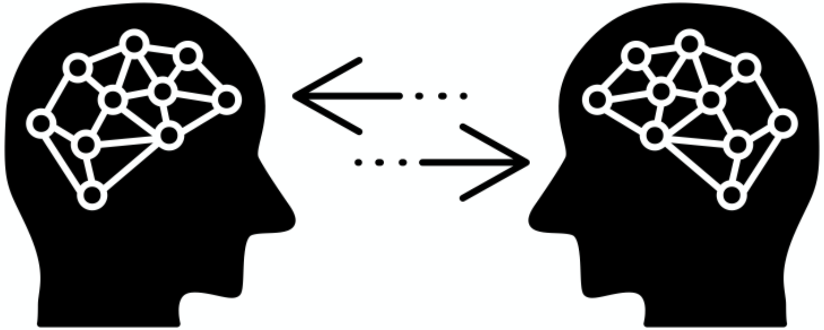
1. Tag the image with the hotel property area
2. Predict aesthetic quality

Models & Training

Transfer Learning

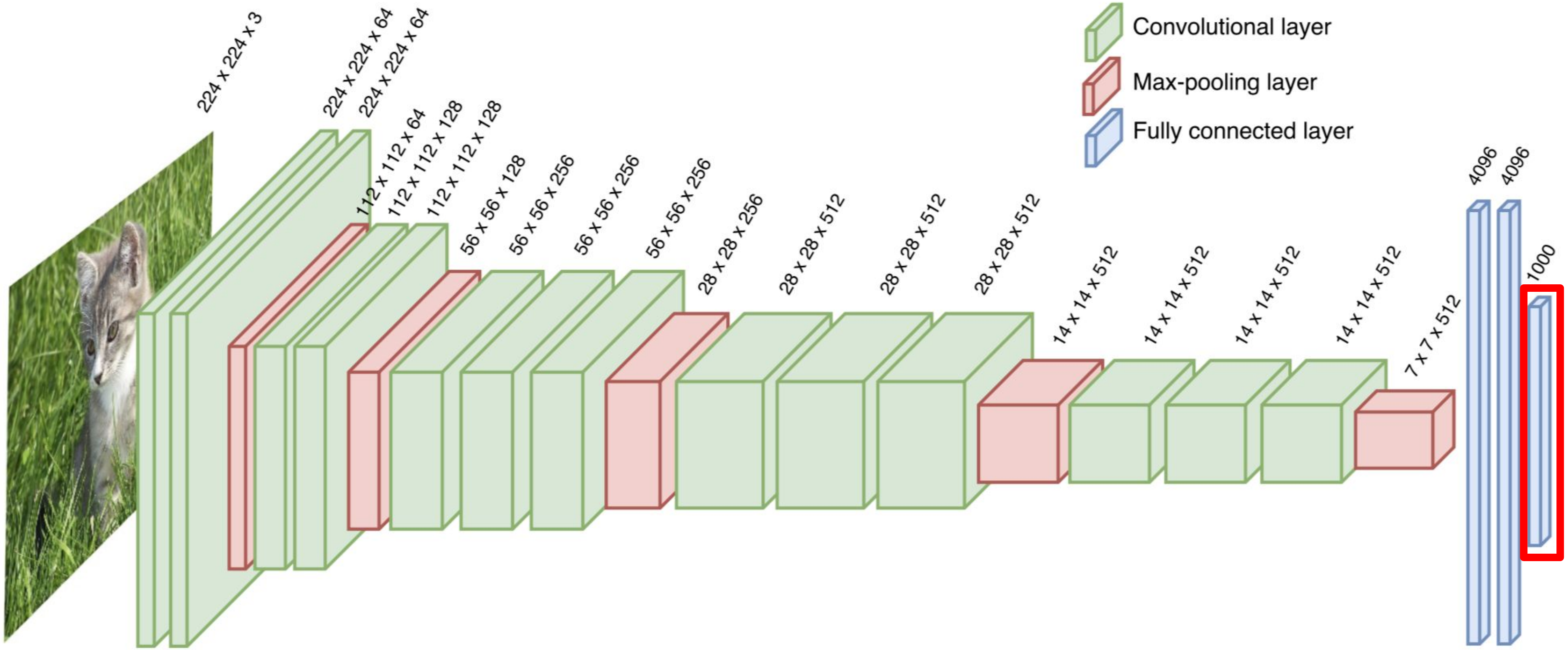
- Use pre-trained CNN that was trained on millions of images

  (e.g. MobileNet or VGG16)
- Replace top layers so that the output fits with classification task
- Train existing and new layer weights



Transfer Learning

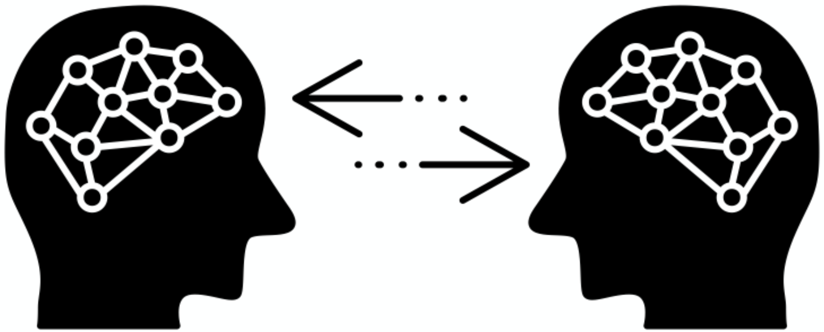
CNN architecture (VGG16)



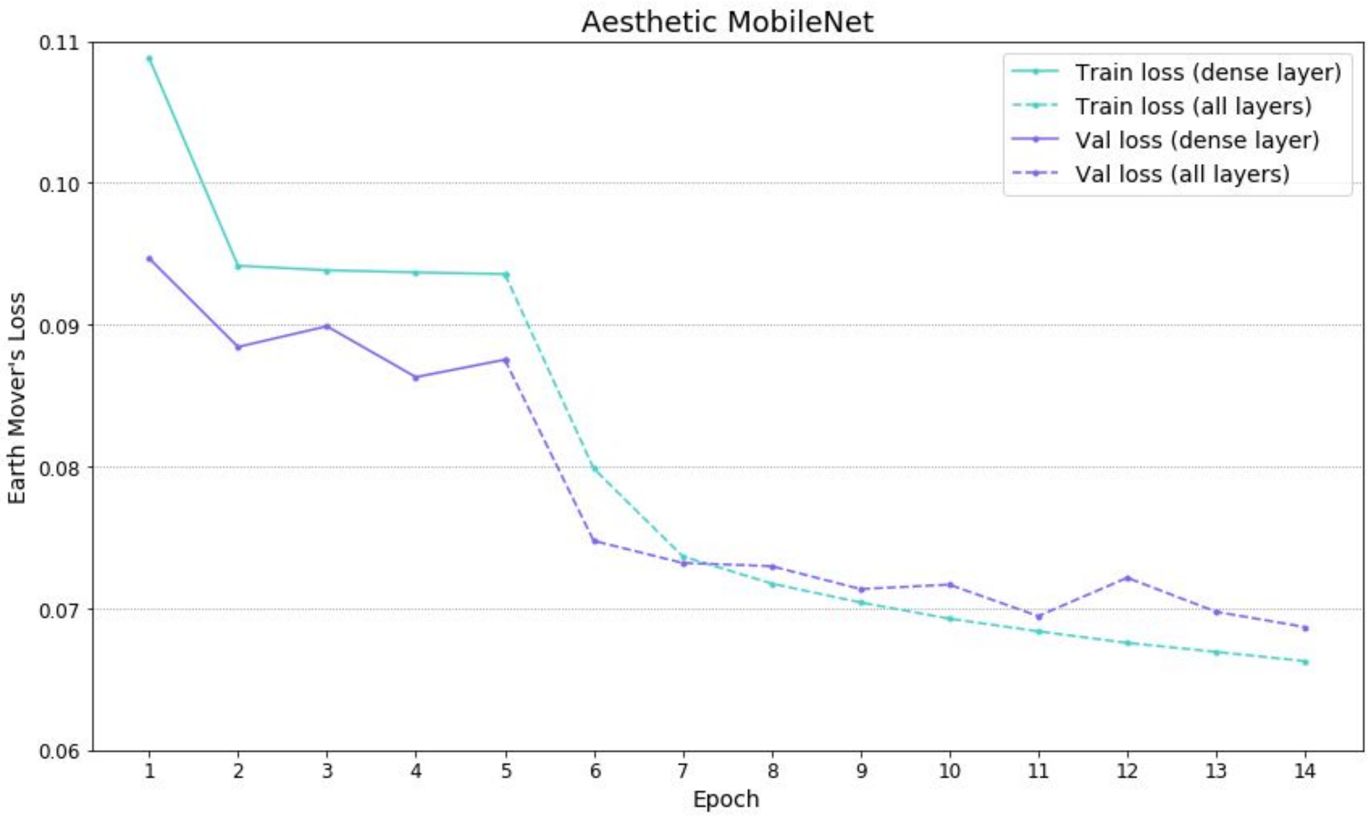
Training regime

1. Only train the newly added dense layers with high learning rate
2. Then train all layers with low learning rate

Goal: Do not juggle around the pre-trained convolutional weights too much



Training regime



Loss functions

Cross-entropy loss (CEL)

- CEL generally used for “one-class” ground truth classifications (e.g. image tagging)
- CEL ignores inter-class relationships between score buckets

$\hat{\mathbf{y}}$ $\qquad$ $\mathbf{y}$

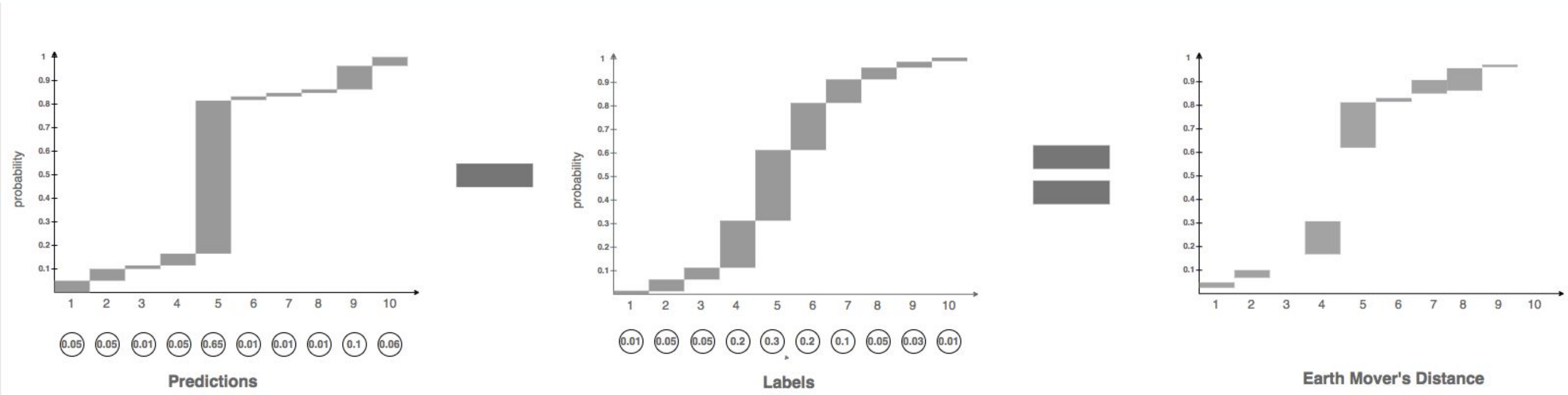
$$\begin{bmatrix} 0.1 \\ 0.5 \\ 0.4 \end{bmatrix} \quad D(\hat{\mathbf{y}}, \mathbf{y}) = -\sum_j y_j \ln \hat{y}_j \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

source: https://ssq.github.io/2017/02/06/Udacity%20MLND%20Notebook/

Loss functions

Earth Mover’s Distance (EMD)

- For ordered classes, classification settings can outperform regressions
- Training on datasets with intrinsic ordering can benefit from EMD loss objective



GPU training workflow

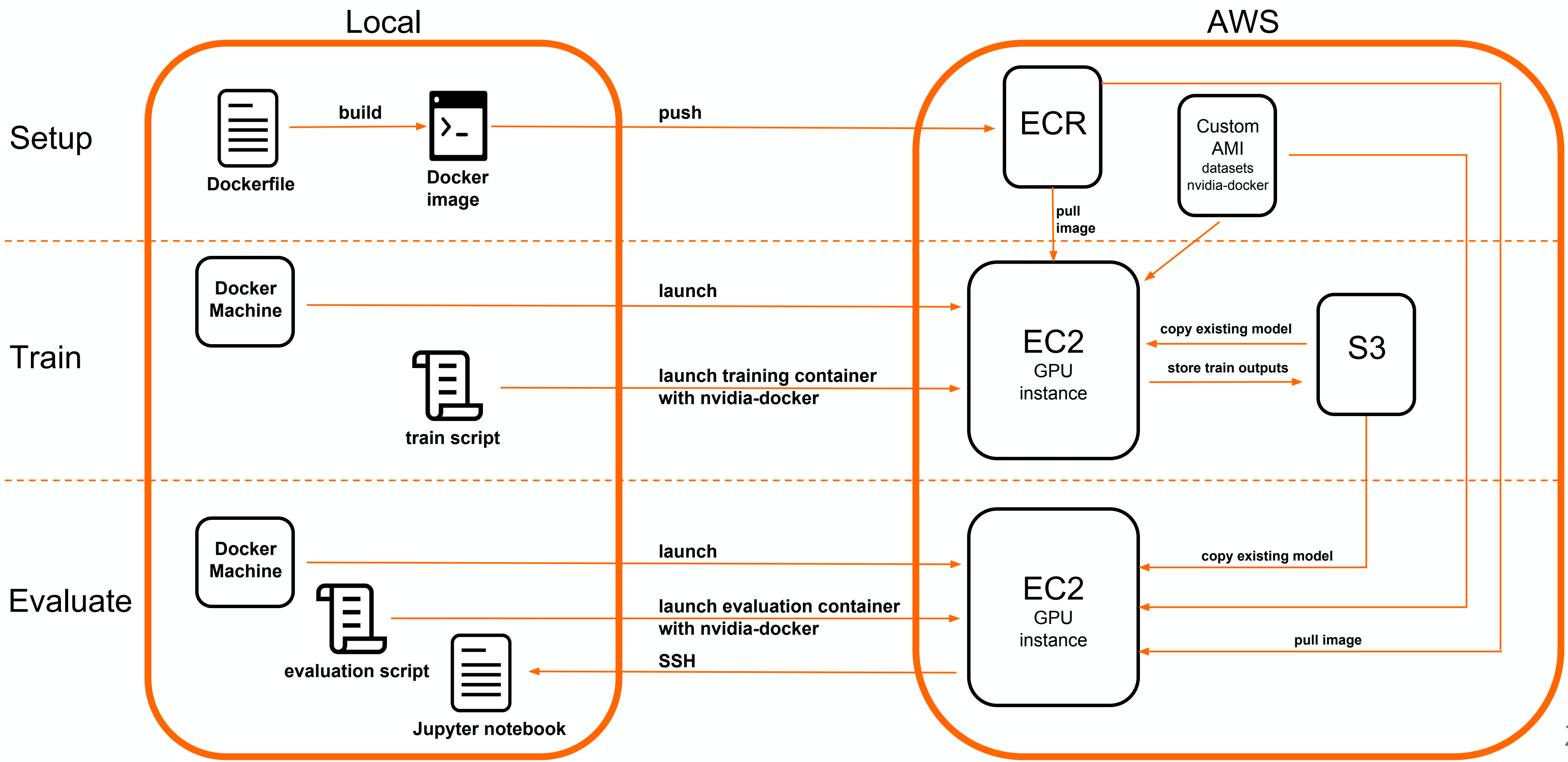


Image Tagging

Tagging Problem

- Given an image, tag it as belonging to a **single** class

- Multiclass classification model with classes:
  - Bedroom
  - Bathroom
  - Foyer
  - Restaurant
  - Swimming Pool
  - Kitchen
  - View of Exterior (Facade)
  - Reception

Multiple Datasets

Will go over them one-by-one and see:

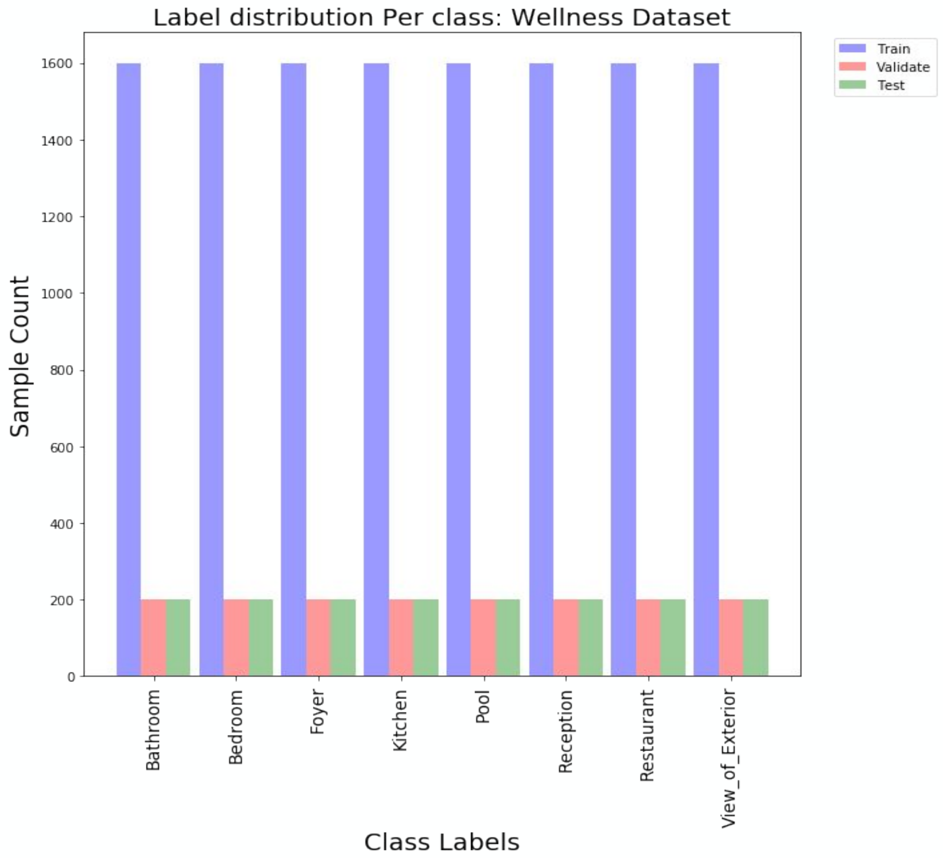
- Dataset properties
- Results
- Issues

Wellness Dataset

- Idealo in-house **pre-labelled** images
- Mostly pictures of 2 or 3 stars properties

Wellness Dataset

- Balanced: Equal sample count in all categories for all sets



Wellness Dataset: Metrics

idealo

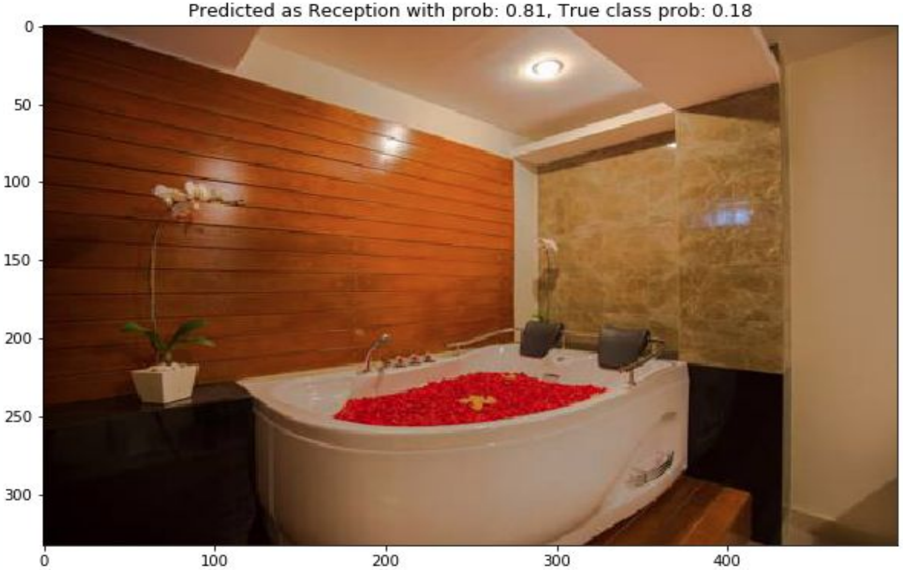
Top-1- accuracy: 86%

	precision	recall	f1-score	support
0	0.97	0.89	0.92	200
1	0.85	0.90	0.87	200
2	0.74	0.74	0.74	200
3	0.85	0.94	0.89	200
4	0.89	0.98	0.94	200
5	0.76	0.86	0.81	200
6	0.91	0.80	0.85	200
7	0.98	0.81	0.89	200
avg / total	0.87	0.86	0.86	1600

	Bathroom	Bedroom	Foyer	Kitchen	Pool	Reception	Restaurant	View_of_Exterior
Bathroom	177	4	2	8	0	9	0	0
Bedroom	1	180	17	1	0	1	0	0
Foyer	3	24	149	2	0	19	2	1
Kitchen	0	1	2	187	0	8	2	0
Pool	0	0	1	0	196	0	3	0
Reception	0	1	8	16	0	172	3	0
Restaurant	2	2	12	5	7	10	160	2
View_of_Exterior	0	1	10	0	16	6	5	162

Wellness Dataset: Wrong Predictions

True Class of these images: BATHROOM, Predicted as: RECEPTION



Rectangular structure = Reception with high probability
→ BIAS!

Wellness Dataset: Wrong Predictions

True Class of these images: BATHROOM

Wrong true label of images
→ NOISE in the dataset!

Correcting Bias

- **Augmentation** operations, same for every class:
  - Random cropping
  - Rotation
  - Horizontal flipping

- **Data enrichment**:
  - External data from google images

Augmented Wellness + Google Dataset: Metrics

Top-1- accuracy: 88%

	precision	recall	f1-score	support
0	0.97	0.93	0.95	244
1	0.87	0.93	0.90	241
2	0.88	0.67	0.76	258
3	0.93	0.89	0.91	277
4	0.97	0.93	0.95	283
5	0.71	0.94	0.81	250
6	0.91	0.82	0.86	268
7	0.87	0.94	0.90	281
avg / total	0.89	0.88	0.88	2102

	Bathroom	Bedroom	Foyer	Kitchen	Pool	Reception	Restaurant	View_of_Exterior
Bathroom	227	2	0	9	0	6	0	0
Bedroom	1	225	5	1	0	7	0	2
Foyer	4	22	173	1	0	40	8	10
Kitchen	2	0	2	246	0	21	6	0
Pool	0	0	1	0	264	0	4	14
Reception	0	3	6	5	0	234	2	0
Restaurant	1	4	5	2	3	19	220	14
View_of_Exterior	0	2	4	0	5	2	3	265

Gotta Clean!

Cleaning Dataset

idealo

- Hand-cleaned each category:
  - Deleted pictures that do not belong in its category
  - Removed duplicates (presence of duplicates can give us wrong metrics)
  - Added more images from external sources for classes with a small number of images left after cleaning

Cleaned Data: Metrics

idealo

Top-1- accuracy: 91%

	precision	recall	f1-score	support
0	0.98	0.95	0.97	170
1	0.95	0.95	0.95	176
2	0.86	0.75	0.80	159
3	0.86	0.97	0.92	189
4	0.93	1.00	0.96	185
5	0.77	0.79	0.78	105
6	0.91	0.94	0.92	167
7	1.00	0.85	0.92	145
avg / total	0.91	0.91	0.91	1296

	Bathroom	Bedroom	Foyer	Kitchen	Pool	Reception	Restaurant	View_of_Exterior
Bathroom	162	0	0	7	0	1	0	0
Bedroom	2	168	1	4	0	1	0	0
Foyer	1	5	119	6	0	16	12	0
Kitchen	0	0	1	184	0	2	2	0
Pool	0	0	0	0	185	0	0	0
Reception	0	2	11	9	0	83	0	0
Restaurant	0	1	2	3	2	2	157	0
View_of_Exterior	0	0	4	0	13	3	2	123

Cleaned Dataset: Results

- Bathroom vs. Reception confusion has almost vanished!
- View_of_exterior vs Pool confusion has reduced
- Foyer performance:
  - Most misclassifications of Foyer gets assigned to Reception
  - This is human problem as well!

Foyer or Reception?

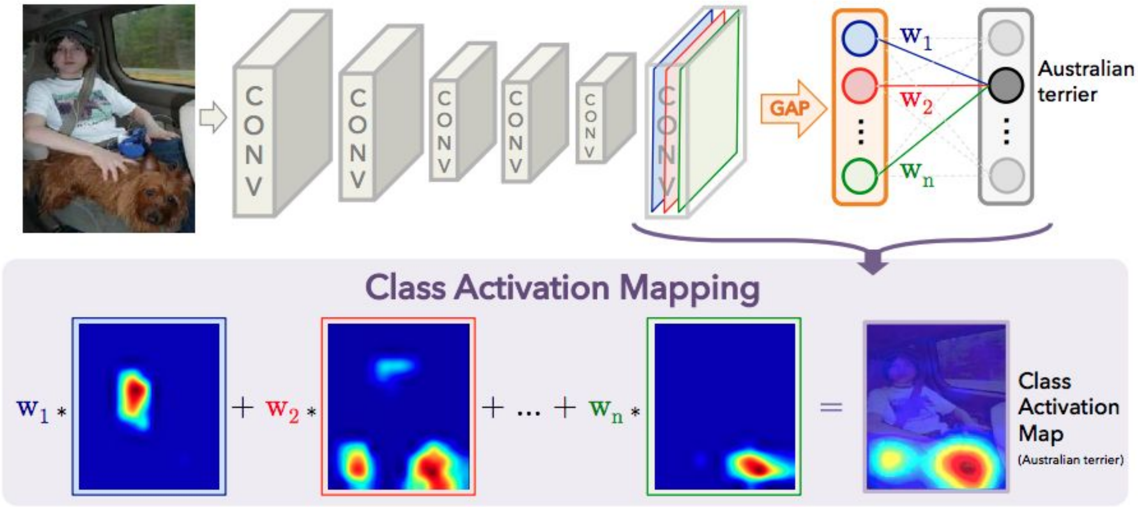
Learnings so far

- The model can only be as good as the data (cleaning)
- Foyer is a hard category to predict

Understanding Model Decisions

Understanding Decisions: Class Activation Maps

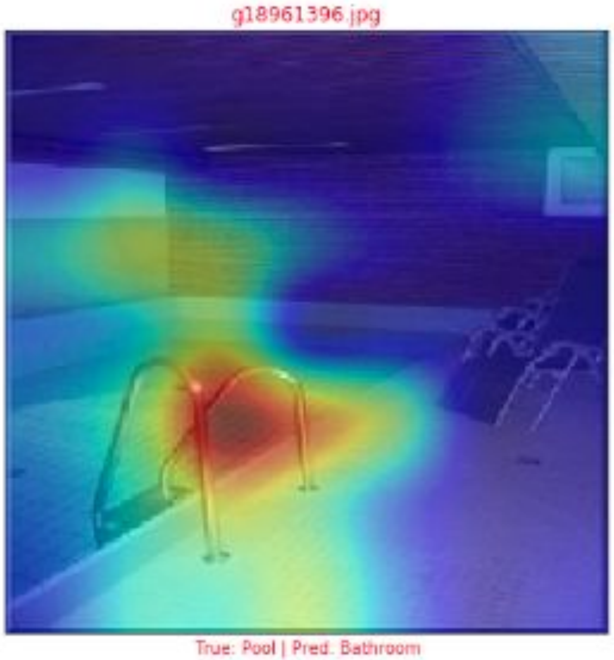
- Use the penultimate Global Average Pooling Layer (GAP) to get class activation map
- Highlights discriminative region that lead to a classification



Insights With CAM

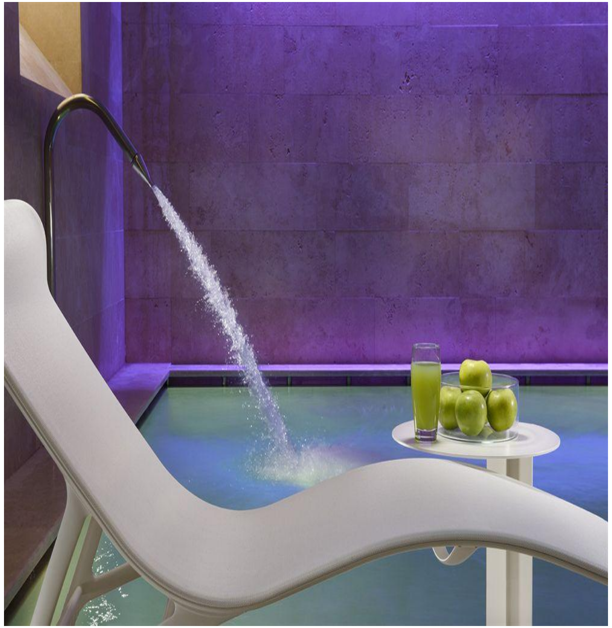
Swimming Pool misclassified as Bathroom

CAM

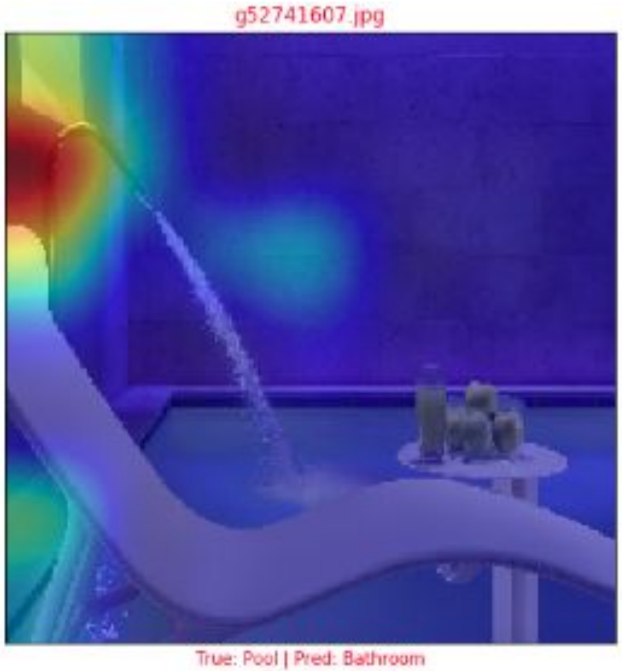


Insights With CAM

Swimming Pool misclassified as Bathroom



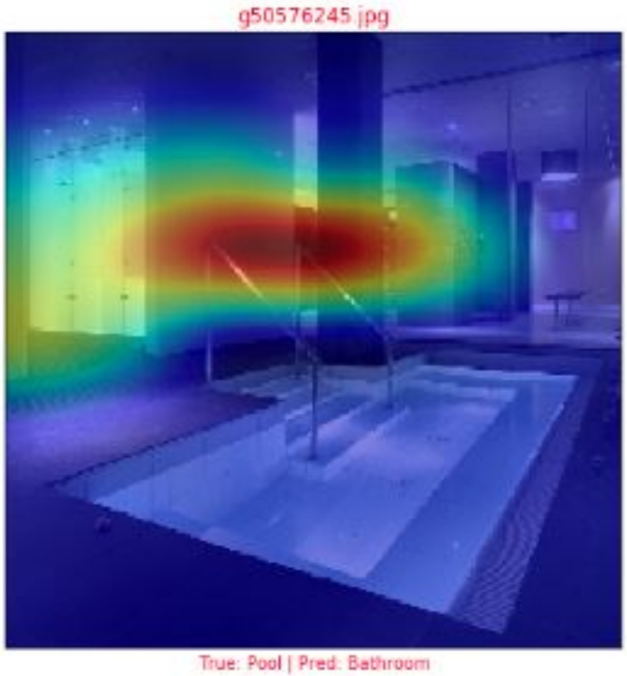
CAM



Insights With CAM

Swimming Pool misclassified as Bathroom

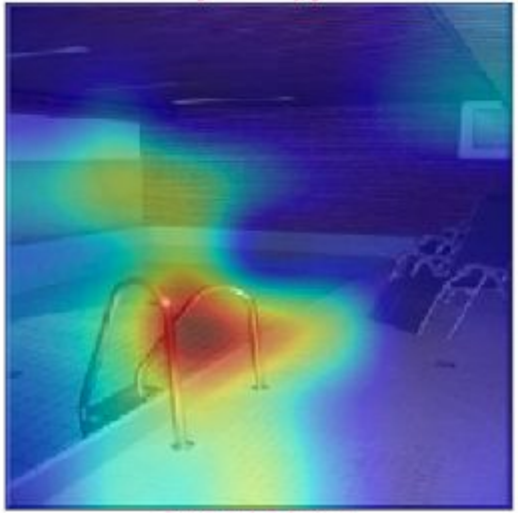
CAM



Insights With CAM

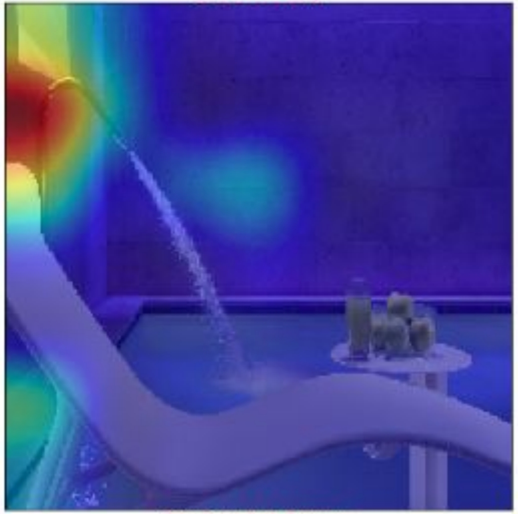
Swimming Pool misclassified as Bathroom

g18961396.jpg



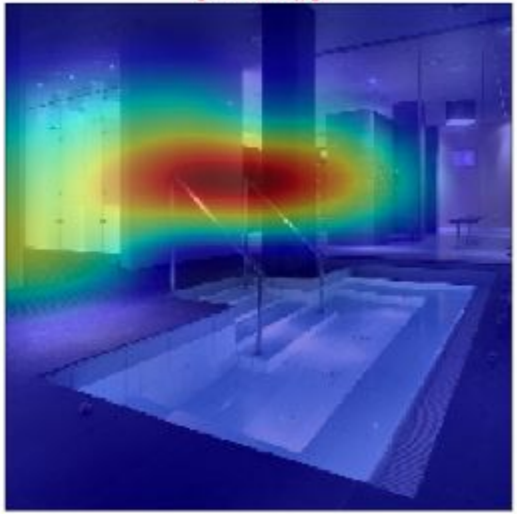
True: Pool | Pred: Bathroom

g52741607.jpg



True: Pool | Pred: Bathroom

g50576245.jpg



True: Pool | Pred: Bathroom

Using rails to misidentify Pool as Bathroom.

Insights With CAM

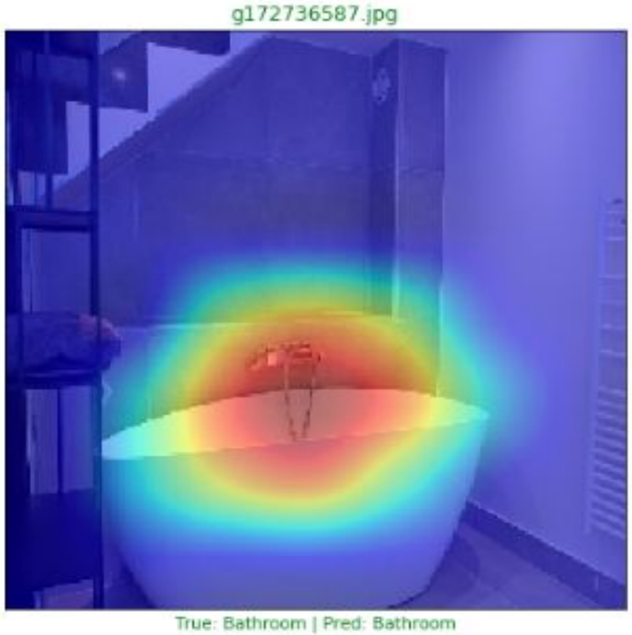
Bathroom correct classification

CAM

Insights With CAM

Bathroom correct classification

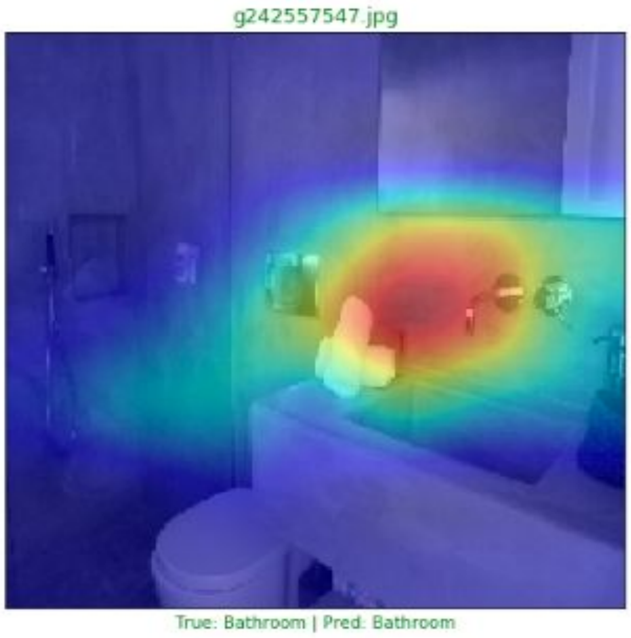
CAM



Insights With CAM

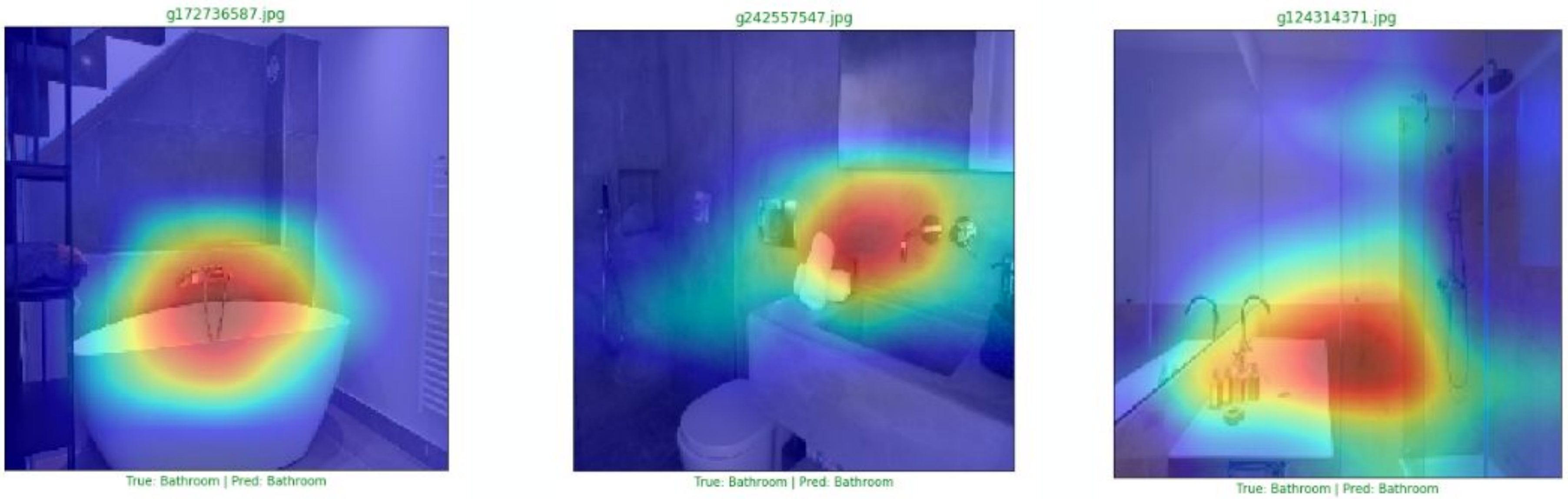
Bathroom correct classification

CAM



Insights With CAM

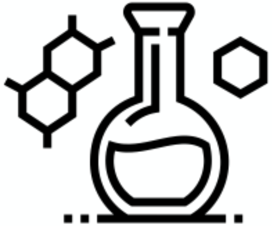
Bathroom correct classification



Using faucets to correctly identify Bathroom.

Learnings so far

- Attribution techniques like CAM lend interpretability
- CAM can drive data collection in specific directions



Tagging Next Steps

1. Add still more data
    a. Explore manual tagging options for training (Example: Amazon Mechanical Turk)
2. Add more classes
    a. Fitness Studio
    b. Conference Room
    c. Other

Image Aesthetics

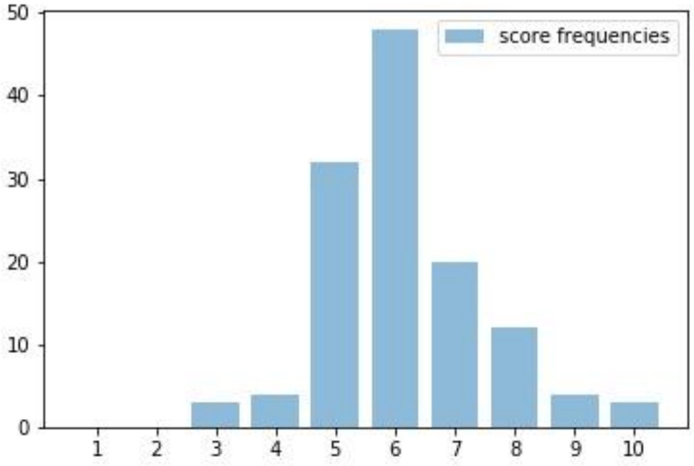
Ground Truth Labels

For the NIMA model we need “true” probability distribution over all classes for each image:

- AVA dataset: we have frequencies over all classes for each image

  → normalize frequencies to get “true” probability distribution

(6.151 / 1.334)

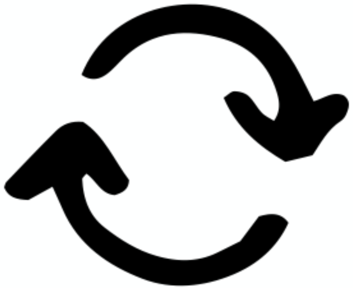


Iterations

idealo

We have gone through two iterations of the aesthetic model:

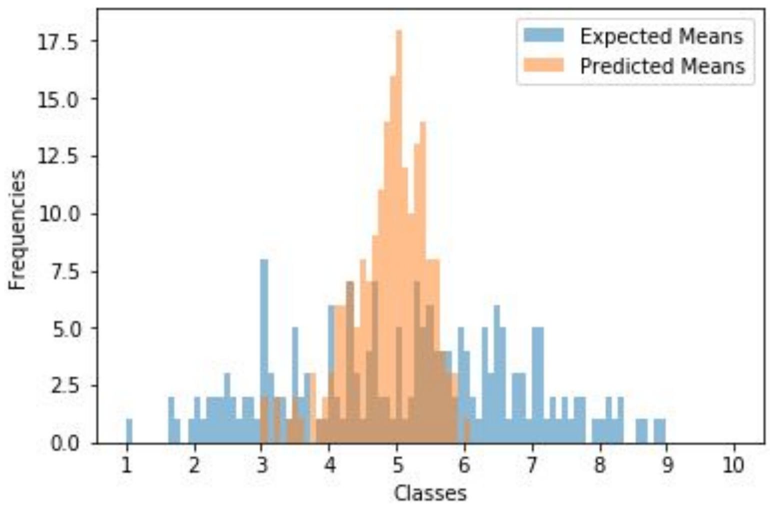
- First iteration - Train on AVA Dataset
- Second iteration - Fine-tune first iteration model on in-house labelled data

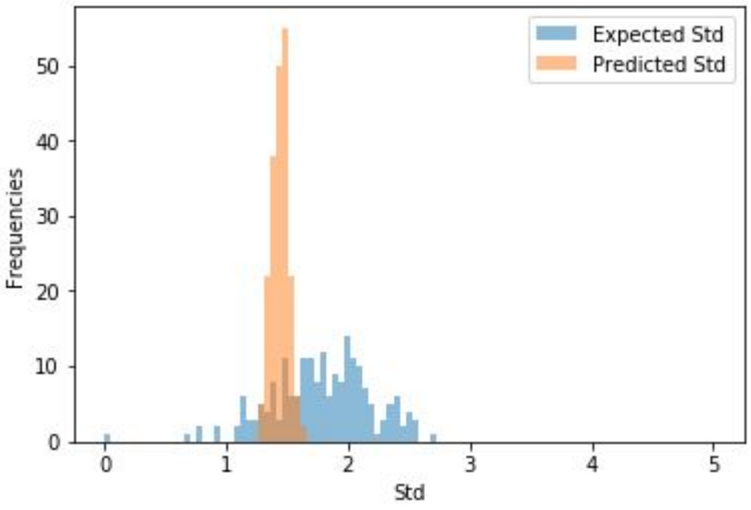


Results - first iteration

Aesthetic model - MobileNet

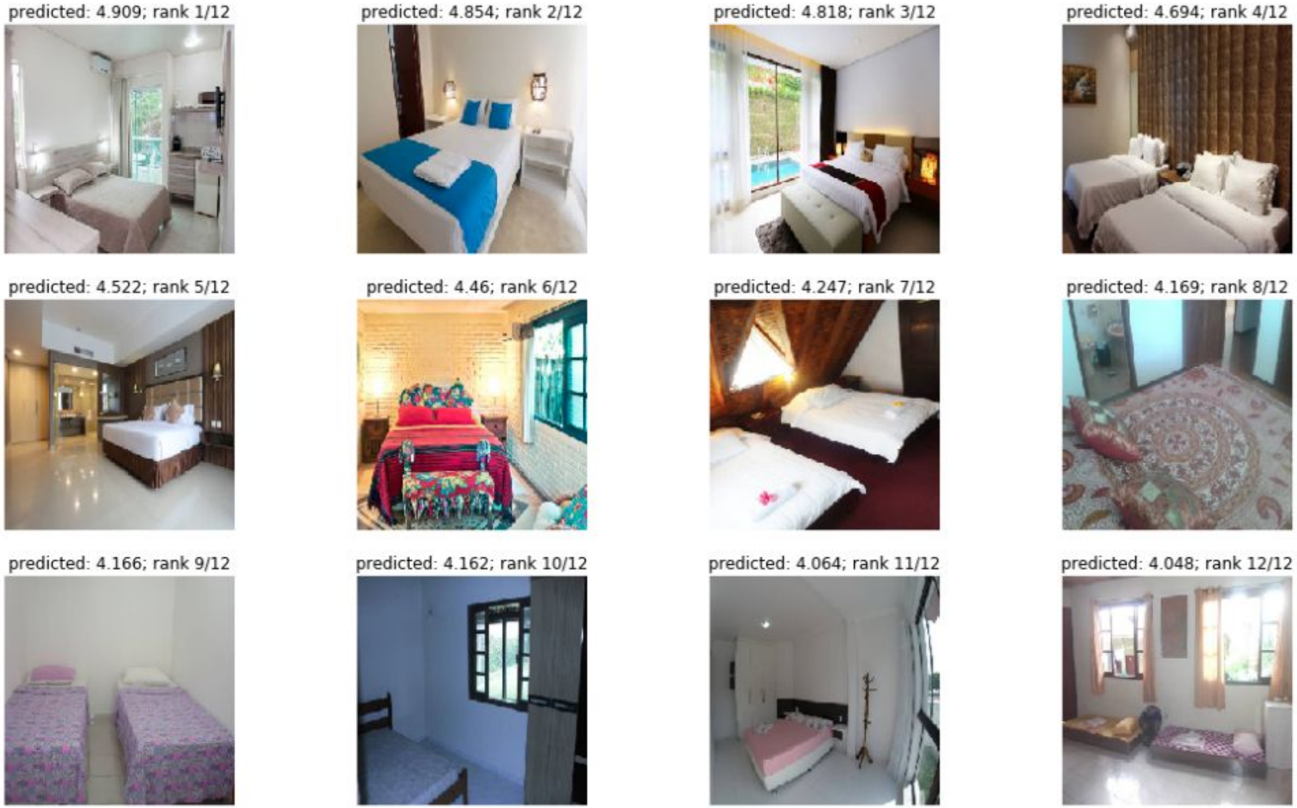
Linear correlation coefficient (LCC): 0.5987
Spearman's correlation coefficient (SCRR): 0.6072
Earth Mover's Distance: 0.2018
Accuracy (threshold at 5): 0.74





Examples - first iteration

Aesthetic model



Examples - first iteration

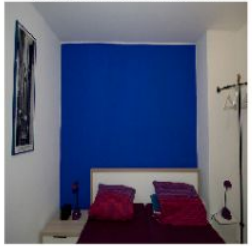
Aesthetic model

Examples - first iteration

Aesthetic model

predicted: 5.386; rank 1/27

predicted: 5.294; rank 2/27



predicted: 5.279; rank 3/27

predicted: 5.186; rank 4/27

predicted: 5.182; rank 5/27

predicted: 5.174; rank 6/27

predicted: 5.171; rank 7/27

predicted: 5.163; rank 8/27

predicted: 5.111; rank 9/27

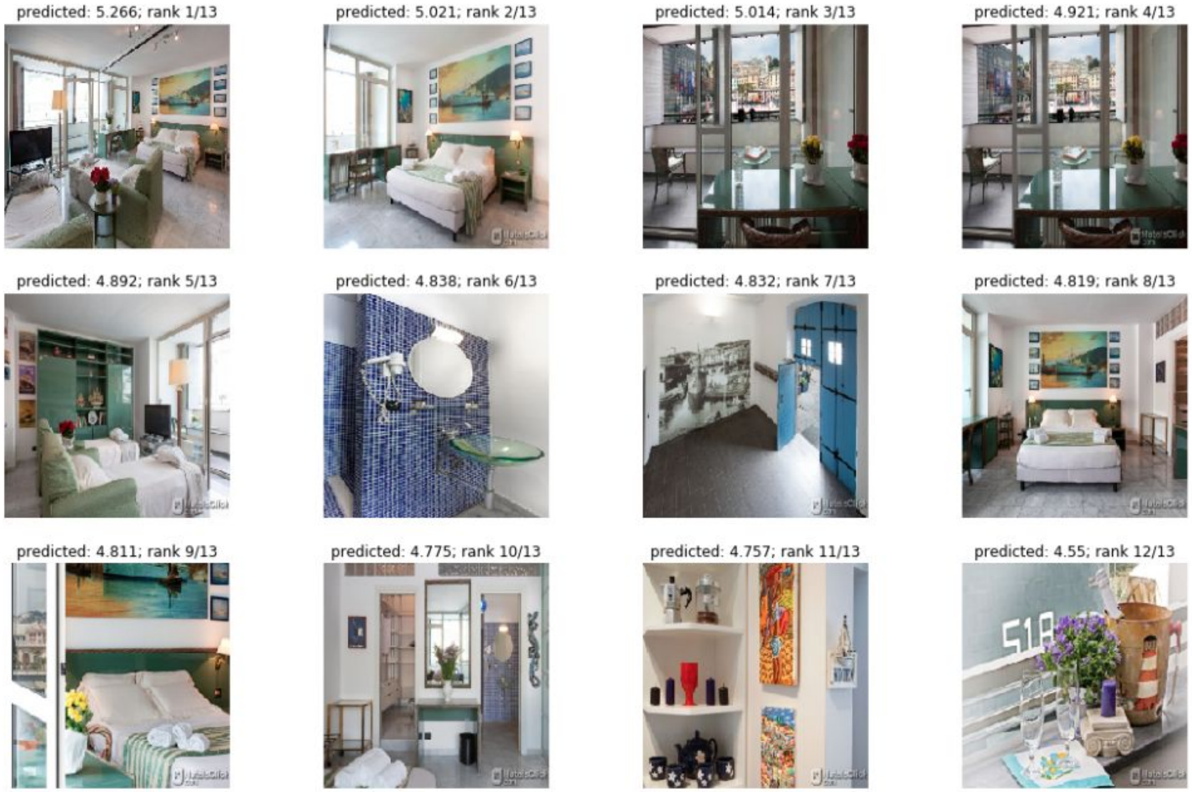
predicted: 5.102; rank 10/27

predicted: 5.095; rank 11/27

predicted: 5.075; rank 12/27

Examples - first iteration

Aesthetic model



Examples - first iteration

Aesthetic model

predicted: 5.115; rank 1/50

predicted: 5.103; rank 2/50

predicted: 5.095; rank 3/50

predicted: 5.062; rank 4/50

predicted: 5.043; rank 5/50

predicted: 4.95; rank 6/50

predicted: 4.943; rank 7/50

predicted: 4.888; rank 8/50

predicted: 4.888; rank 9/50

predicted: 4.842; rank 10/50

predicted: 4.835; rank 11/50

predicted: 4.831; rank 12/50

Results - second iteration

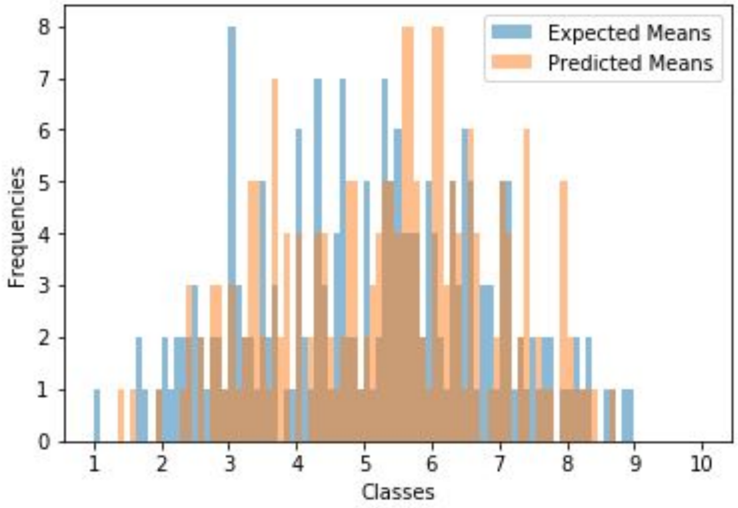
idealo

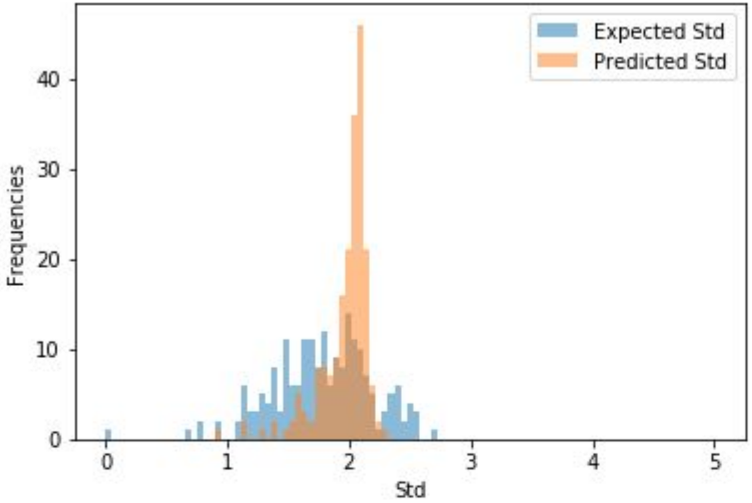
- We built a simple labeling application
- http://image-aesthetic-labelling-app-nima.apps.eu.idealo.com/
- ~ 12 people from idealo Reise and Data Science labeled
    - 1000 hotel images for aesthetics
- We fine-tuned the aesthetic model with 800 training images
- Built aesthetic test dataset with 200 images

Results - second iteration

Aesthetic model - MobileNet

Linear correlation coefficient (LCC): 0.7986
Spearman's correlation coefficient (SCRR): 0.7743
Earth Mover's Distance: 0.1236
Accuracy (threshold at 5): 0.85





Examples - second iteration

Aesthetic model

Examples - second iteration

Aesthetic model

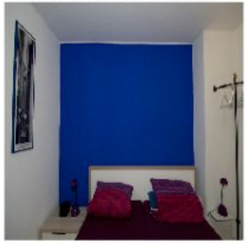
Examples - second iteration

Aesthetic model

predicted: 6.39; rank 1/27

predicted: 5.431; rank 2/27

predicted: 5.232; rank 3/27



predicted: 5.159; rank 4/27

predicted: 5.1; rank 5/27

predicted: 5.07; rank 6/27

predicted: 5.025; rank 7/27

predicted: 4.893; rank 8/27

predicted: 4.855; rank 9/27

predicted: 4.773; rank 10/27

predicted: 4.769; rank 11/27

predicted: 4.741; rank 12/27

Examples - second iteration

Aesthetic model

predicted: 7.613; rank 1/13

predicted: 6.538; rank 2/13

predicted: 6.459; rank 3/13

predicted: 6.248; rank 4/13

predicted: 6.227; rank 5/13

predicted: 6.043; rank 6/13

predicted: 5.884; rank 7/13

predicted: 5.421; rank 8/13

predicted: 4.995; rank 9/13

predicted: 4.588; rank 10/13

predicted: 4.553; rank 11/13

predicted: 4.382; rank 12/13

Examples - second iteration
Aesthetic model

predicted: 7.281; rank 1/50

predicted: 6.734; rank 2/50

predicted: 6.546; rank 3/50

predicted: 6.447; rank 4/50

predicted: 6.053; rank 5/50

predicted: 5.851; rank 6/50

predicted: 5.845; rank 7/50

predicted: 5.815; rank 8/50

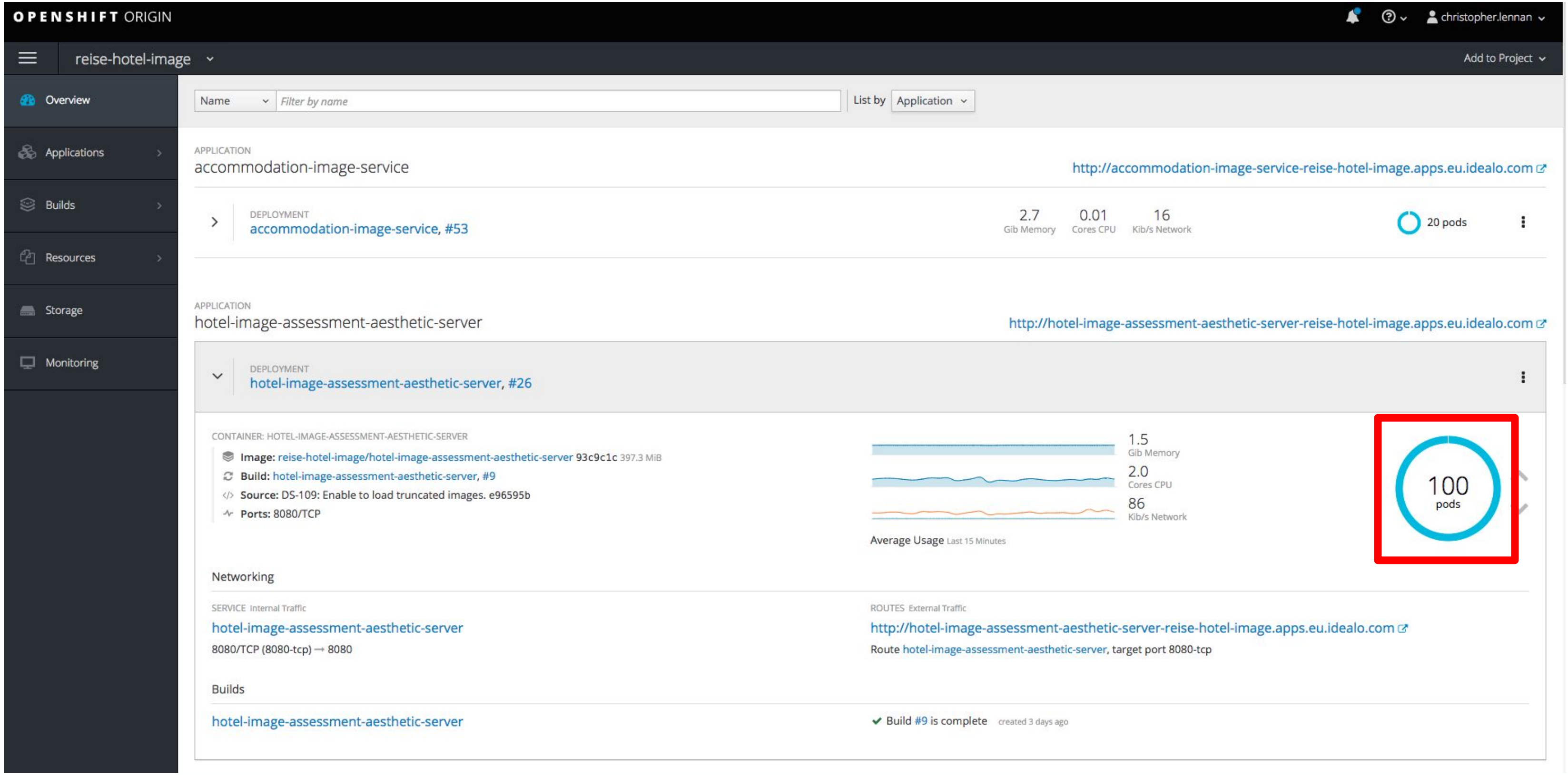
predicted: 5.634; rank 9/50

predicted: 5.616; rank 10/50

predicted: 5.535; rank 11/50

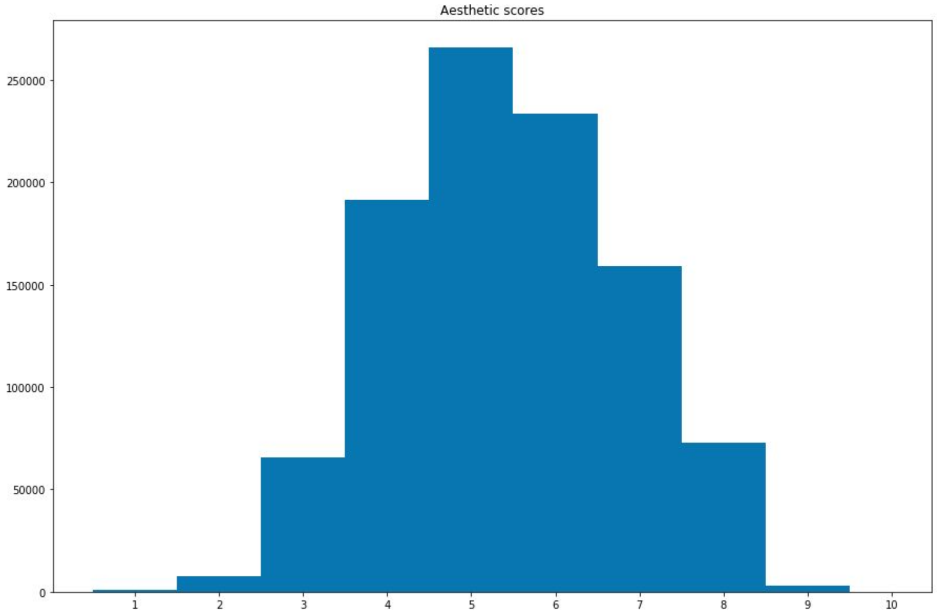
predicted: 5.519; rank 12/50

Production
Aesthetic model



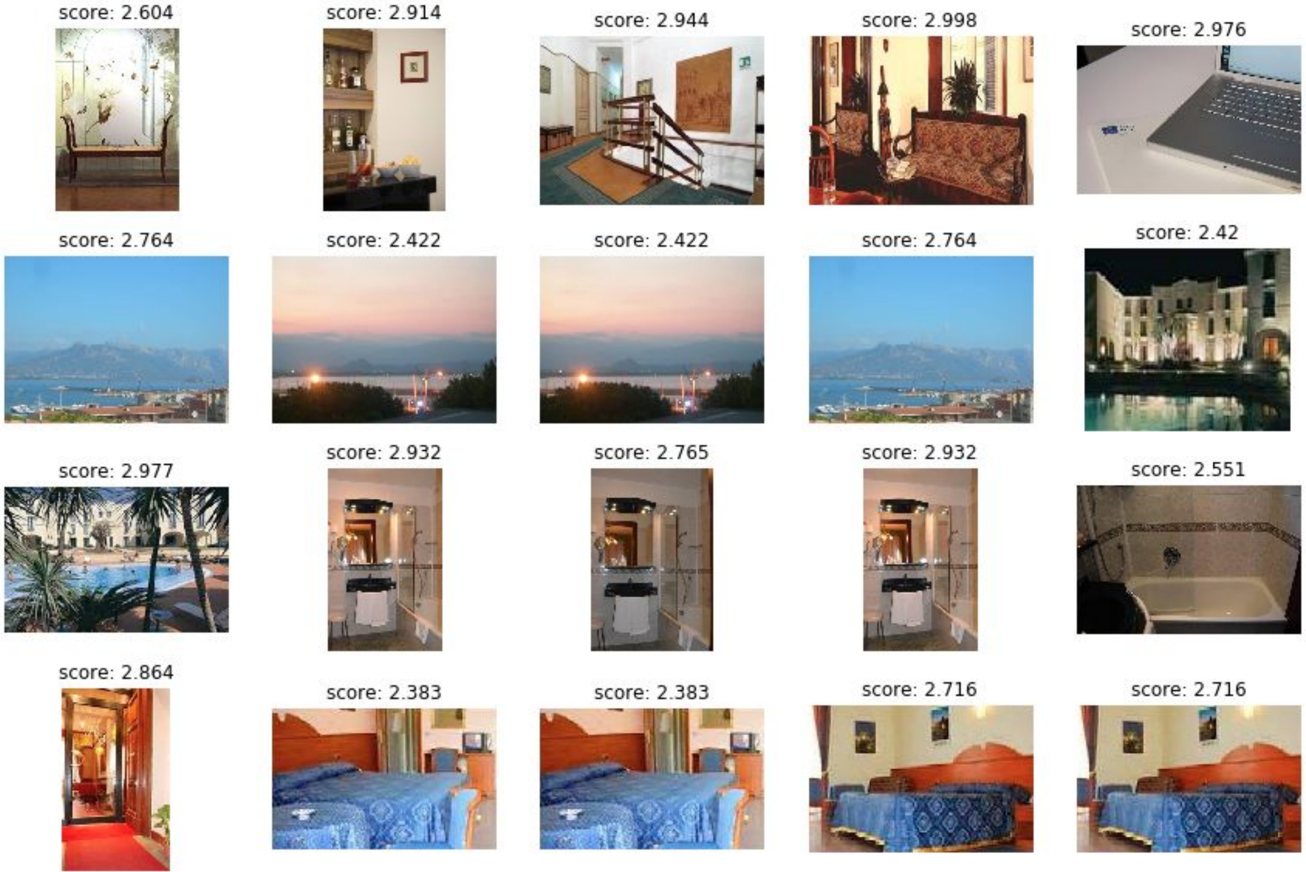
Production
Aesthetic model

- To date we have scored ~280 million images
- Distribution of scores (sample of 1 million scores):



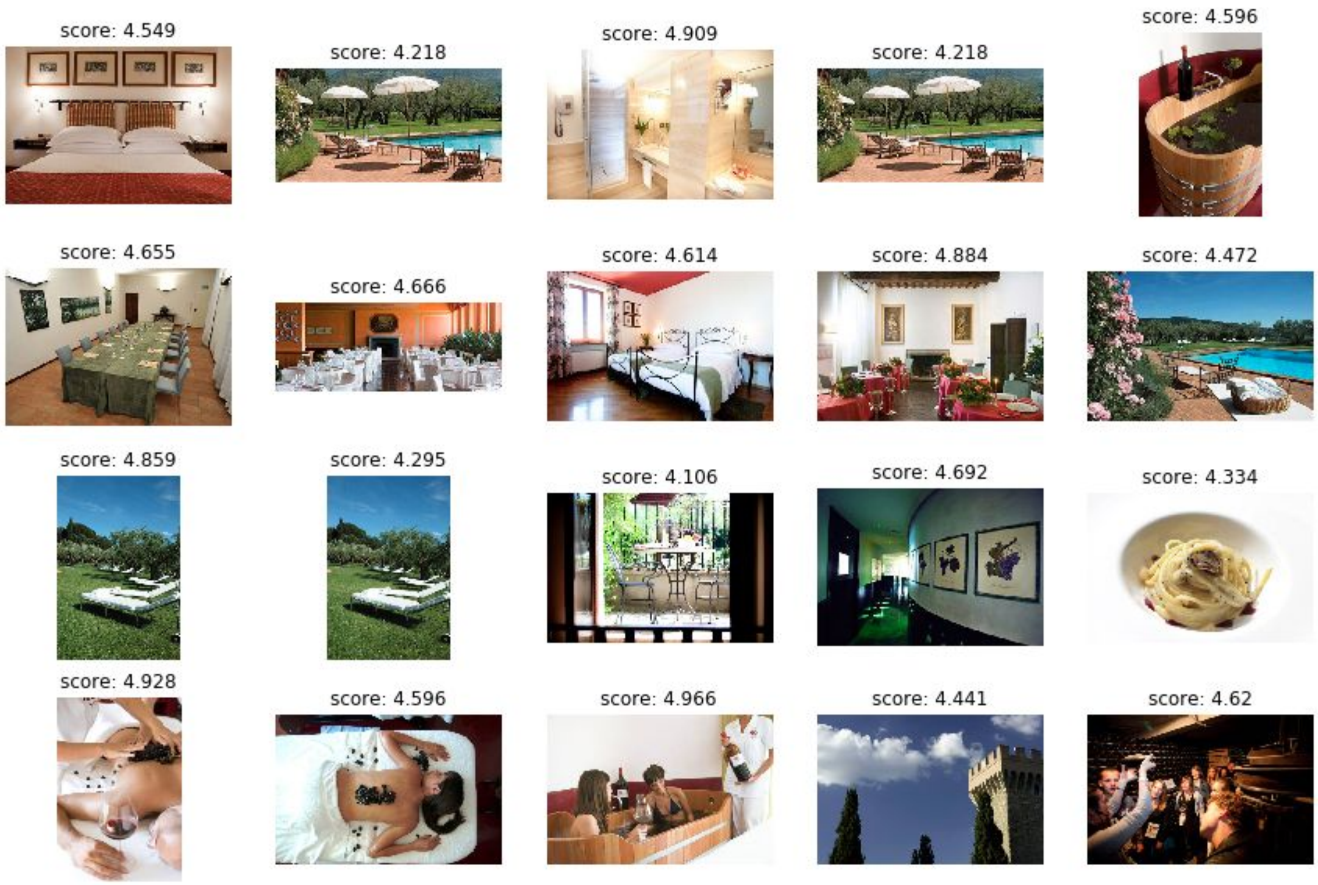
Production - Low Scores

Aesthetic model



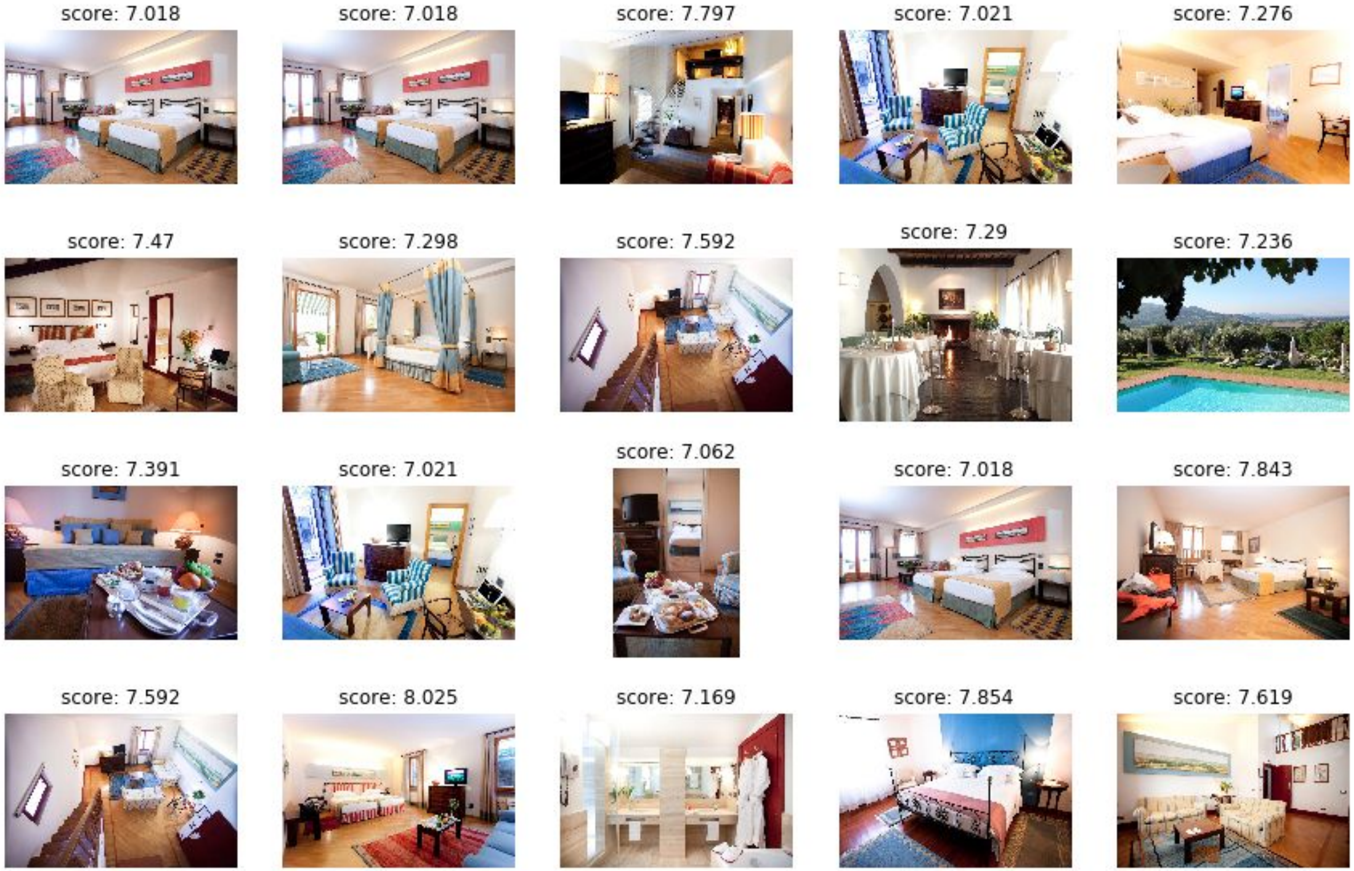
Production - Medium Scores

Aesthetic model



Production - High Scores

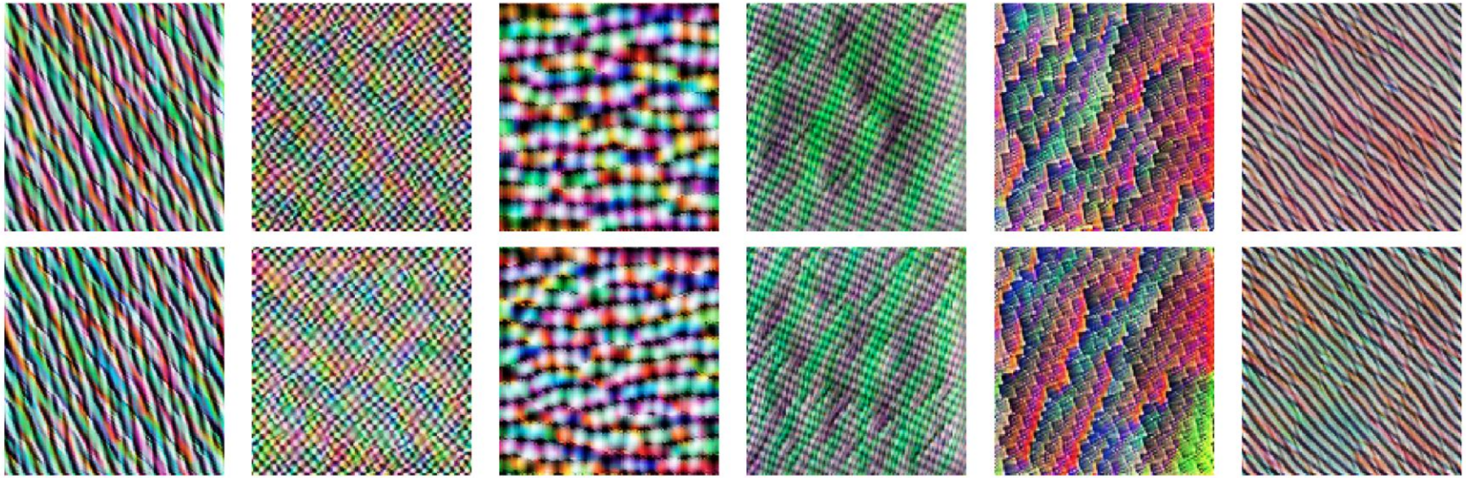
Aesthetic model



Understanding Model Decisions

Convolutional Filter Visualisations

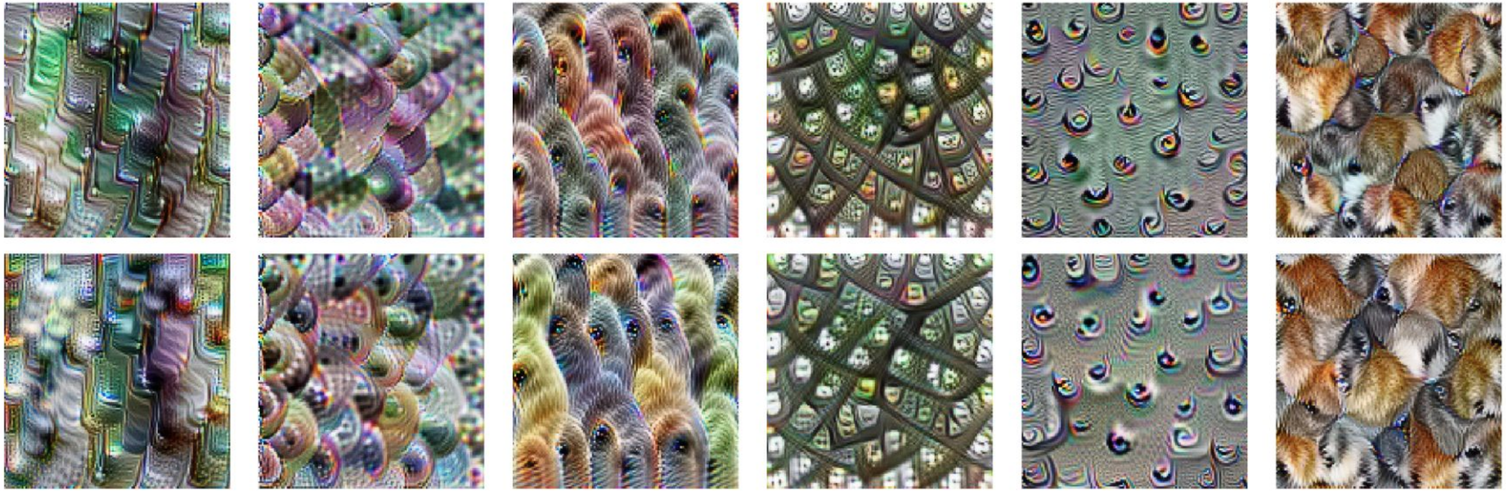
Layer 23 MobileNet original



MobileNet Aesthetic

Convolutional Filter Visualisations

Layer 51 MobileNet original

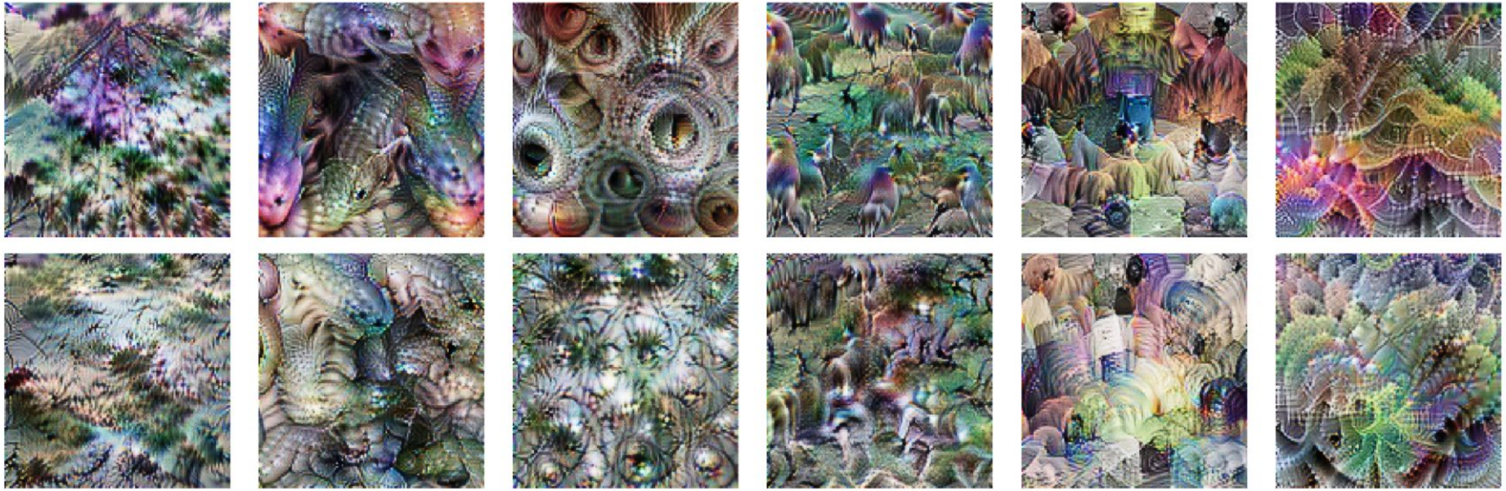


MobileNet Aesthetic

Convolutional Filter Visualisations

Layer 79

MobileNet original

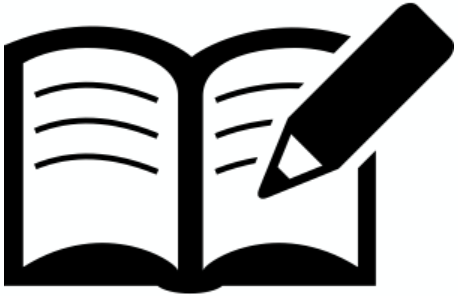


MobileNet Aesthetic

Aesthetic Learnings

idealo

- **Hotel specific labeled data is key** - Aesthetic model improved markedly from 800 additional training samples
- NIMA only requires **few samples** to achieve **good results** (EMD loss)
- Labeled hotel images also important for **test set** (model evaluation)
- Training on GPU significantly improved training time **(~30 fold)**



Aesthetics Next Steps

- Continue labeling images for aesthetic classifier
- Introduce new desirable biases in labeling (e.g. low technical quality == low aesthetics)
- Improve prediction speed of models (e.g. lighter CNN architectures)

Summary

Summary

- Transfer learning allowed us to train image tagging and aesthetic classifiers with a few thousand domain specific samples
- Showed the importance of having noise-free data for quality predictions
- Use of attribution & visualization techniques helps understand model decisions and improve them

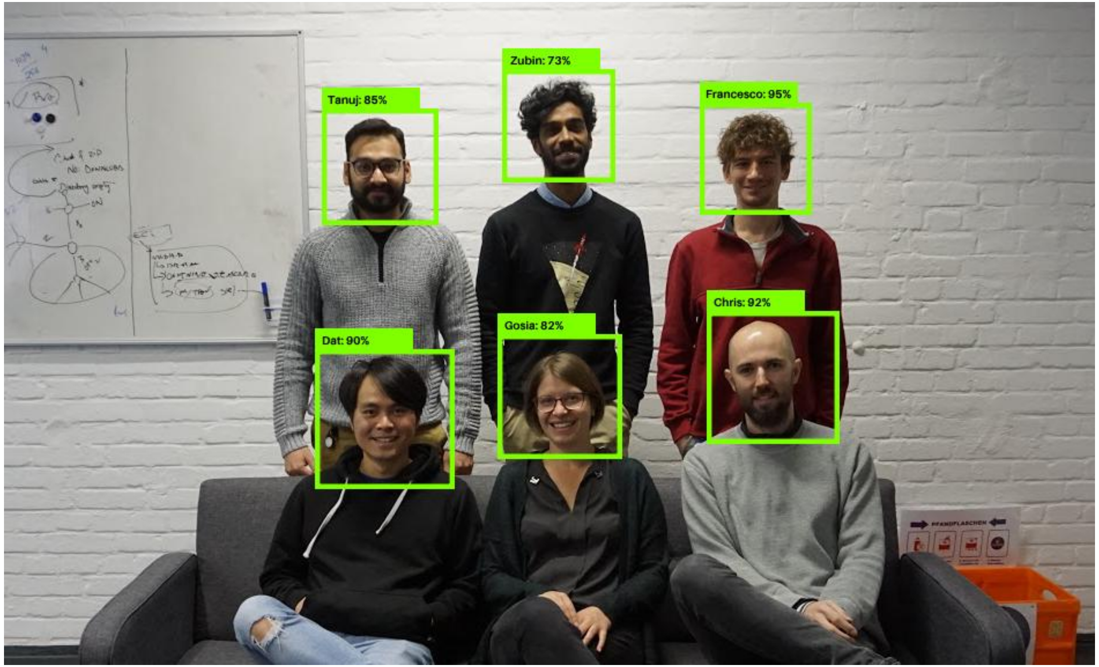
Check us out! #idealoTech

idealo

https://github.com/idealo

https://medium.com/idealo-tech-blog

We're hiring!

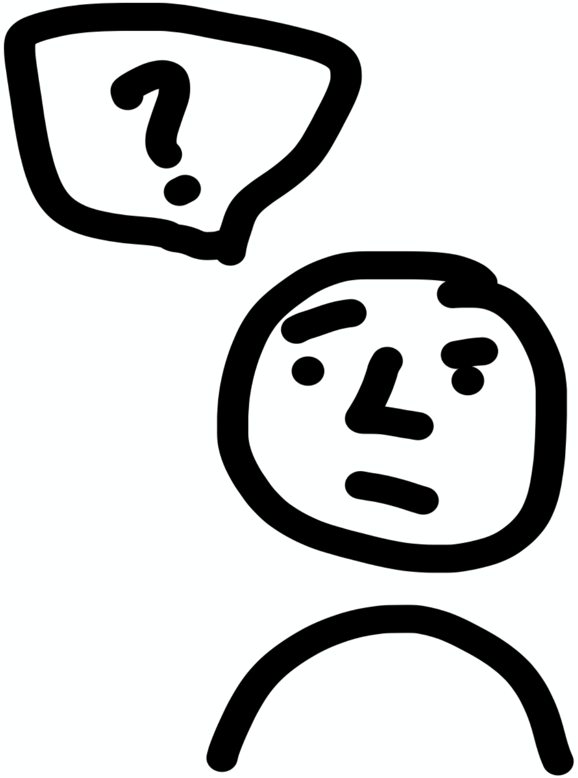


Data Engineers, DevOps Engineers across different teams

Check out our job postings: jobs.idealo.de

Tanuj Jain
tanuj.jain@idealo.de
@tjainn

Christopher Lennan
christopher.lennan@idealo.de
@chris_lennan



THE END