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

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

- 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
- 50,000 shops
- Over 330 million offers for 2 million products
- Tüv certified comparison portal
Motivation
# idealo hotel price comparison

[hotel.idealo.de](https://hotel.idealo.de/unterkuenfte/berlin-s2950198/)

- 2,306,658 accommodations
- 308,519,299 images
- ~133 images per accommodation
Importance of Photography for Hotels

“.. 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 ..”
Hotel Comenius
8,5 Super
Ø 3,0 km Stadtzentrum | Karte
Spa | Sauna | WLAN | Fitnessraum
Infos ▼ Verfügbarkeit prüfen
pro Nacht, Ø
ab 44 €

Heart of Gold Hostel Berlin
8,4 Super
Ø 1,3 km Stadtzentrum | Karte
WLAN | Klimaanlage | Spielzimmer | Fahrradverleih
Infos ▼ Verfügbarkeit prüfen
pro Nacht, Ø
ab 31 €

Cityhostel Berlin
8,1 Super
Ø 2,2 km Stadtzentrum | Karte
WLAN | Spielzimmer | Fahrradverleih | Touristverbindungen
Infos ▼ Verfügbarkeit prüfen
pro Nacht, Ø
ab 36 €
Hotel Comenius

8,5 Super

Ø 3,0 km Stadtzentrum  Karte
Spa | Sauna | WLAN | Fitnessraum

Infos ▽ Verfügbarmkeit prüfen

pro Nacht, Ø
ab 44 €

zum Angebot
bei Booking.com
Hotel Comenius
8,5 Super

Ø 3,0 km Stadtzentrum  Karte
Spa | Sauna | WLAN | Fitnessraum

Infos  Verfügbarkeit prüfen

pro Nacht, Ø
ab 44 €

zum Angebot bei Booking.com
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
Ensure different areas get depicted
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

Aesthetic MobileNet

- Train loss (dense layer)
- Train loss (all layers)
- Val loss (dense layer)
- Val loss (all layers)
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

\[
D(\hat{y}, y) = - \sum_{j} y_j \ln \hat{y}_j
\]
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

Setup
- Dockerfile
- Docker image
- Docker Machine
- build
- push
- ECR
- pull image
- Custom AMI datasets
- nvidia-docker

Train
- train script
- launch
- EC2 GPU instance
- launch training container with nvidia-docker
- S3
- copy existing model
- store train outputs

Evaluate
- evaluation script
- Docker Machine
- launch
- EC2 GPU instance
- launch evaluation container with nvidia-docker
- SSH
- Jupyter notebook
- pull image
- copy existing model
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

Top-1- accuracy: 86%

<table>
<thead>
<tr>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.97</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>1</td>
<td>0.85</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>2</td>
<td>0.74</td>
<td>0.74</td>
<td>0.74</td>
</tr>
<tr>
<td>3</td>
<td>0.85</td>
<td>0.94</td>
<td>0.89</td>
</tr>
<tr>
<td>4</td>
<td>0.89</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>5</td>
<td>0.76</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>6</td>
<td>0.91</td>
<td>0.80</td>
<td>0.85</td>
</tr>
<tr>
<td>7</td>
<td>0.98</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>avg / total</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bathroom</th>
<th>Bedroom</th>
<th>Foyer</th>
<th>Kitchen</th>
<th>Pool</th>
<th>Reception</th>
<th>Restaurant</th>
<th>View_of_Exterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathroom</td>
<td>177</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Bedroom</td>
<td>1</td>
<td>180</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Foyer</td>
<td>3</td>
<td>24</td>
<td>149</td>
<td>2</td>
<td>0</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>Kitchen</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>187</td>
<td>0</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Pool</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>196</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Reception</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>16</td>
<td>0</td>
<td>172</td>
<td>3</td>
</tr>
<tr>
<td>Restaurant</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>160</td>
</tr>
<tr>
<td>View_of_Exterior</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>16</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>
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%

<table>
<thead>
<tr>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.97</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>1</td>
<td>0.87</td>
<td>0.93</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>0.67</td>
<td>0.76</td>
</tr>
<tr>
<td>3</td>
<td>0.93</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>4</td>
<td>0.97</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>0.71</td>
<td>0.94</td>
<td>0.81</td>
</tr>
<tr>
<td>6</td>
<td>0.91</td>
<td>0.82</td>
<td>0.86</td>
</tr>
<tr>
<td>7</td>
<td>0.87</td>
<td>0.94</td>
<td>0.90</td>
</tr>
</tbody>
</table>

avg / total | 0.89   | 0.88     | 0.88    | 2102    |

<table>
<thead>
<tr>
<th></th>
<th>Bathroom</th>
<th>Bedroom</th>
<th>Foyer</th>
<th>Kitchen</th>
<th>Pool</th>
<th>Reception</th>
<th>Restaurant</th>
<th>View_of_Exterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathroom</td>
<td>227</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bedroom</td>
<td>1</td>
<td>225</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Foyer</td>
<td>4</td>
<td>22</td>
<td>173</td>
<td>1</td>
<td>0</td>
<td>40</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Kitchen</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>245</td>
<td>0</td>
<td>21</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Pool</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>264</td>
<td>0</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Reception</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>234</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Restaurant</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>19</td>
<td>220</td>
<td>14</td>
</tr>
<tr>
<td>View_of_Exterior</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>265</td>
</tr>
</tbody>
</table>
Gotta Clean!
Cleaning Dataset

- 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

Top-1- accuracy: 91%

<table>
<thead>
<tr>
<th>precision</th>
<th>recall</th>
<th>f1-score</th>
<th>support</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.98</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>1</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>0.86</td>
<td>0.75</td>
<td>0.80</td>
</tr>
<tr>
<td>3</td>
<td>0.86</td>
<td>0.97</td>
<td>0.92</td>
</tr>
<tr>
<td>4</td>
<td>0.93</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>5</td>
<td>0.77</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>6</td>
<td>0.91</td>
<td>0.94</td>
<td>0.92</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>0.85</td>
<td>0.92</td>
</tr>
<tr>
<td>avg / total</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Bathroom</th>
<th>Bedroom</th>
<th>Foyer</th>
<th>Kitchen</th>
<th>Pool</th>
<th>Reception</th>
<th>Restaurant</th>
<th>View_of_Exterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathroom</td>
<td>162</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bedroom</td>
<td>2</td>
<td>168</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Foyer</td>
<td>1</td>
<td>5</td>
<td>119</td>
<td>6</td>
<td>0</td>
<td>16</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Kitchen</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>184</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pool</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>185</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Reception</td>
<td>0</td>
<td>2</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>83</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Restaurant</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>157</td>
<td>0</td>
</tr>
<tr>
<td>View_of_Exterior</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>3</td>
<td>2</td>
<td>123</td>
</tr>
</tbody>
</table>
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
Insights With CAM

Swimming Pool misclassified as Bathroom
Insights With CAM

Swimming Pool misclassified as Bathroom
Insights With CAM

Swimming Pool misclassified as Bathroom

Using rails to misidentify Pool as Bathroom.
Insights With CAM

Bathroom correct classification
Insights With CAM

Bathroom correct classification

CAM
Insights With CAM

Bathroom correct classification
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
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
Examples - first iteration

Aesthetic model
Examples - first iteration

Aesthetic model
Results - second iteration

- We built a simple labeling application
- ~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
Examples - second iteration
Aesthetic model

predicted: 7.28; rank 1/50
predicted: 6.73; rank 2/50
predicted: 6.54; rank 3/50
predicted: 6.44; rank 4/50

predicted: 6.03; rank 5/50
predicted: 5.85; rank 6/50
predicted: 5.84; rank 7/50
predicted: 5.81; rank 8/50

predicted: 5.63; rank 9/50
predicted: 5.62; rank 10/50
predicted: 5.53; rank 11/50
predicted: 5.52; rank 12/50
To date we have scored ~280 million images

Distribution of scores (sample of 1 million scores):
Production - Low Scores
Aesthetic model

Production - Low Scores
Aesthetic model
Production - Medium Scores

Aesthetic model
Production - High Scores
Aesthetic model

score: 7.018  score: 7.018  score: 7.797  score: 7.021  score: 7.276

score: 7.47  score: 7.298  score: 7.592  score: 7.29  score: 7.236

score: 7.391  score: 7.021  score: 7.062  score: 7.018  score: 7.843

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

- **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)
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

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
THE END