

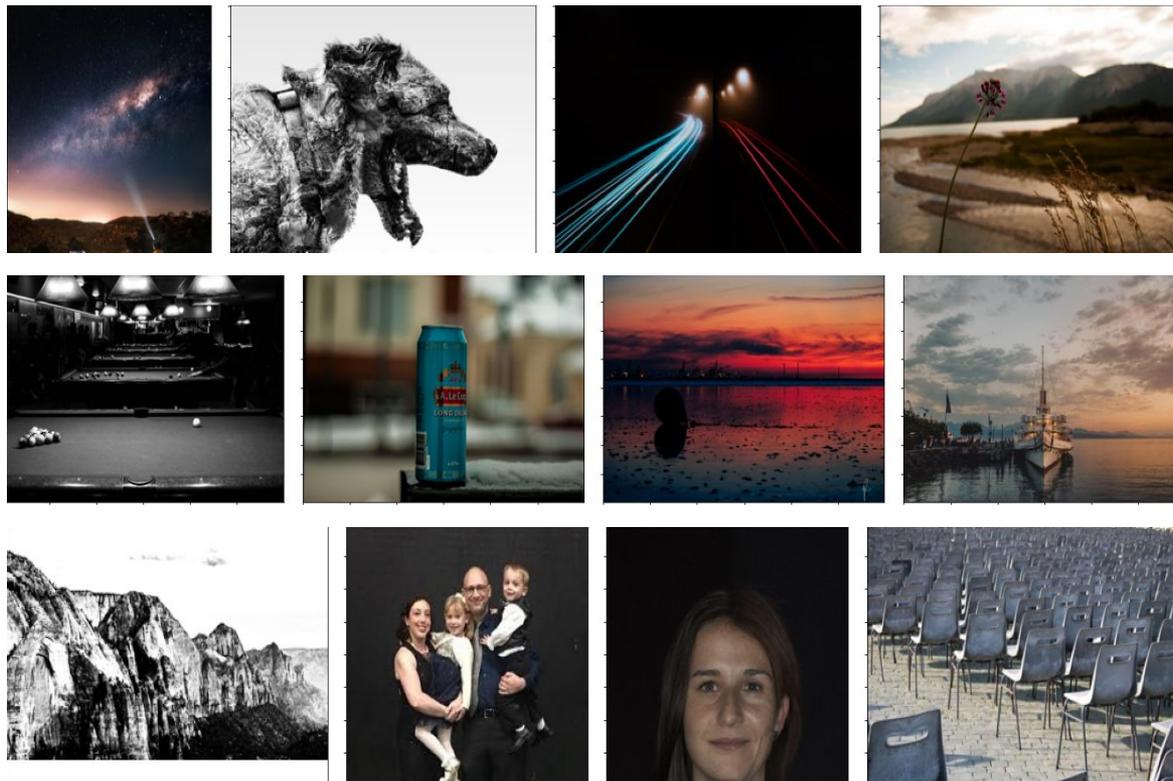
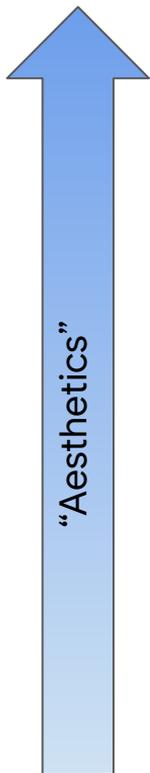


Understanding Aesthetics with Language: A Photo Critique Dataset for Aesthetic Assessment

Daniel Vera Nieto, Luigi Celona and Clara Fernández Labrador

<https://mediatechnologycenter.github.io/aestheval/>

Motivation





Definition

aes·thet·ic

/es'THedik/

adjective

1. Concerned with beauty or the appreciation of beauty.
“The pictures give great aesthetic pleasure”

NOT easy to define

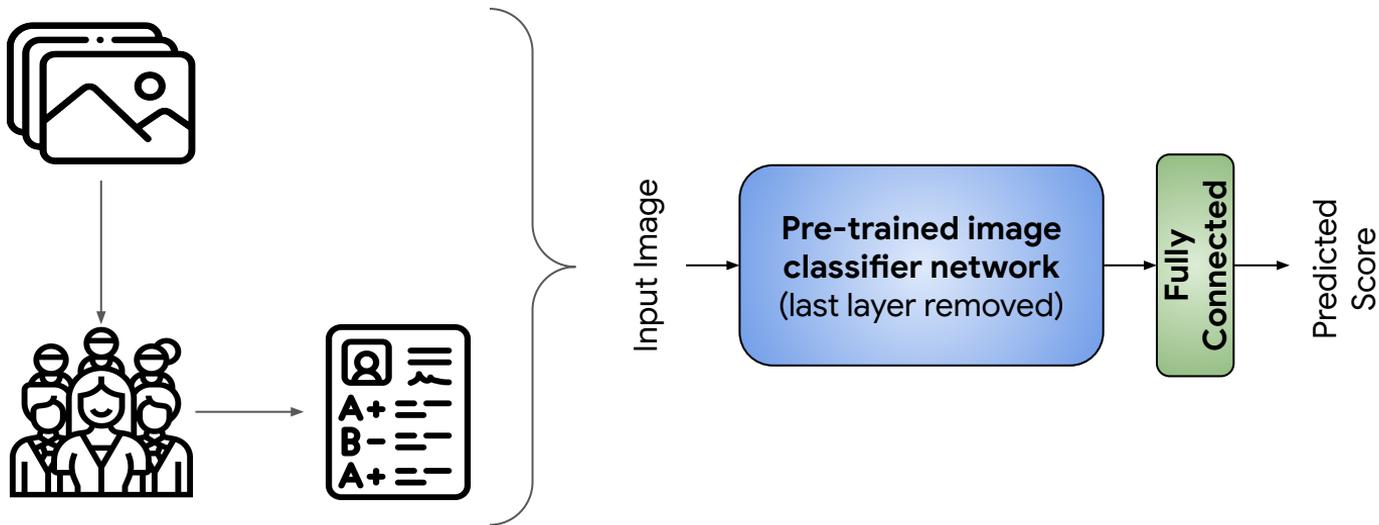
Subjectivity

It all depends on the viewer's:

- Preferences
- Experiences
- Photography knowledge
- ...

Prior work

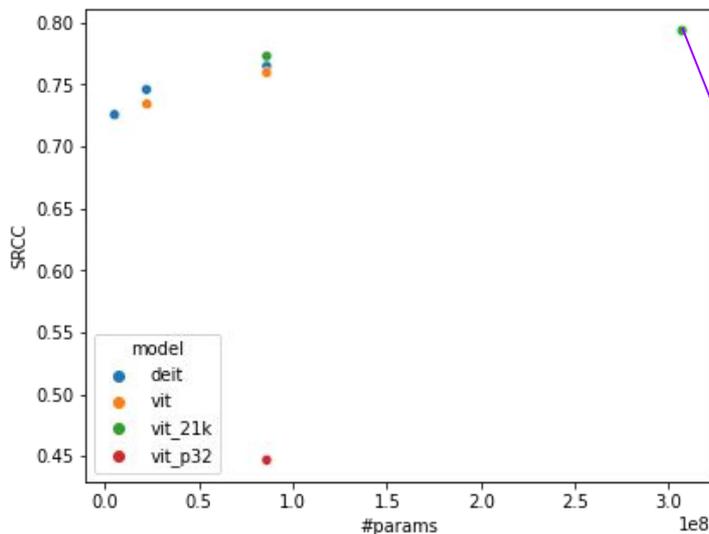
AVA Dataset [1]



[1] AVA: A large-scale database for aesthetic visual analysis. Naila Murray; Luca Marchesotti; Florent Perronnin (2012)

Following the traditional approach

Selection of the best model in terms of tradeoff between performance and number of parameters



Comparison of our baseline with state-of-the-art methods on the AVA dataset

| Model | SRCC | LCC | Accuracy (%) |
|---------------------------|--------------|--------------|--------------|
| Murray <i>et al.</i> [28] | – | – | 66.70 |
| Lu <i>et al.</i> [24] | – | – | 74.46 |
| Ma <i>et al.</i> [26] | – | – | 81.70 |
| Kong <i>et al.</i> [20] | 0.558 | – | 77.33 |
| Talebi <i>et al.</i> [37] | 0.612 | 0.636 | 81.51 |
| Chen <i>et al.</i> [6] | 0.649 | 0.671 | 83.20 |
| Xu <i>et al.</i> [40] | 0.724 | 0.725 | 80.90 |
| Ke <i>et al.</i> [19] | 0.726 | 0.738 | 81.15 |
| Celona <i>et al.</i> [4] | 0.731 | 0.732 | 80.75 |
| Hosu <i>et al.</i> [14] | <u>0.756</u> | <u>0.757</u> | 81.72 |
| ViT-L/16 - 21k | 0.793 | 0.793 | <u>82.85</u> |

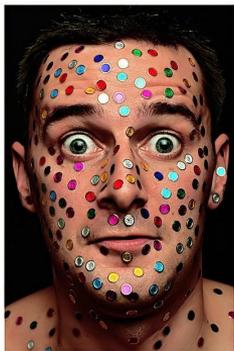
Challenges

- Summarizing the aesthetic judgment in a **single value** limits the representation of visual aesthetics.
- Aesthetic scores are highly **dependent on the voting procedure** (i.e., voting scale, number of stimuli, questions and adjectives in the voting scale)
- Lack of **interpretability** (Why an image has a score of 6? Why not 7?)



Can we use comments on the photos for
aesthetic assessment?

Existing datasets



SCORE: 6.1

AVA-Comments - Joint Image and Text Representation for Aesthetics Analysis

#funny...I like it...

#Ha! As a post-challenge trick, try erasing all evidence of his face. I wonder whether the spots alone would provide enough contours to tell it was a face.

#LOL.. good one. Like the colors of the picture and also the spots :o) #The dot on the adam's apple is a nice touch, heh.

#Bood? Correct me if I'm wrong. I love your self portraits. This one is awesome. Love the bulging eyes :P

#very nice, like that look. well lighted and cropped.

Photo Critique Captioning Dataset (PCCD) - Aesthetic Critiques Generation for Photos

"subject_of_photo": "In this case I feel as if the subject (the child) [...]"

"use_of_camera": "Your camera settings are spot on.[...]"

"description": "His action and characteristics after he had made a mess and was caught."

"depth_of_field": "Personally I don't think fake depth of field is ever a good idea[...]"

"focus": "Your camera has locked onto the child and [...]"

"color_lighting": "I feel as if the lighting is a bit flat in this image. I can see [...]"

"general_impression": "Hi Jon. It's great to see a photographer experimenting in post production and [...]"

"composition": "In this shot you have chosen to shoot above looking [...]"



SCORE: 7.3

Existing datasets



SCORE: 6.1

AVA-Comments - Joint Image and Text Representation for Aesthetics Analysis

#funny
#Ha! A
provide
#LOL..
touch,
#Bood
#very nice, like that look. well lighted and cropped.

ner the spots alone would
the adam's apple is a nice
ve the bulging eyes :P

Many comments are **not very informative** of the aesthetics of the image

Photo Critique Captioning Dataset (PCCD) - Aesthetic Critiques Generation for Photos

"subject_of_photo": "In this case I feel as if the subject (the child) [...]"
"use_of
"descri
"depth
"focus"
"color_
"general_impression": "In fact, it's great to see a photographer experimenting in post production and [...]",
"composition": "In this shot you have chosen to shoot above looking [...]"

Critiques are very informative about aesthetics, but there are **only 4236 images**



SCORE: 7.3

Introducing r/Photocritique Dataset (RPCD)

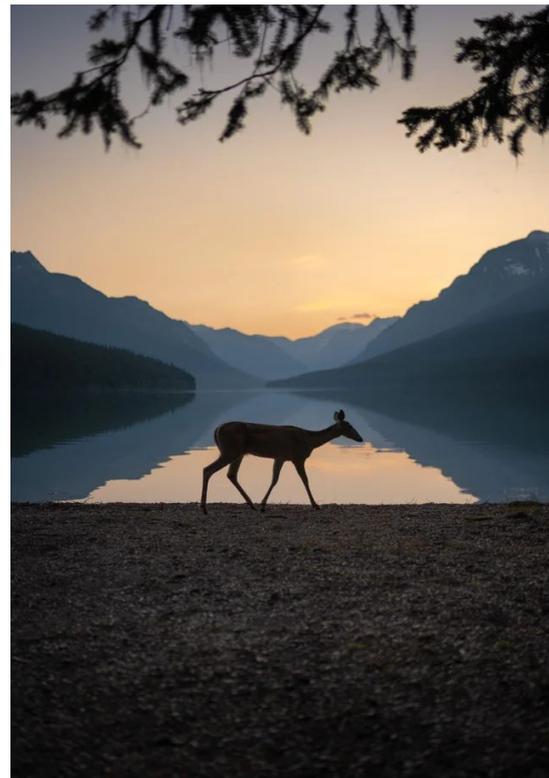
[r/photocritique](#) is a subreddit with a huge community where people submit their photos seeking critique from the rest of the users.

A user submits a photo, usually providing a comment of the intention or technical details of the photo:

“The framing on this was very intentional. I wanted to get a centered photo with the deer coming across, and thought the branches coming down from above would add some interest. Struggling on whether I should have kept it darker and made it more of a true silhouette. Curious what your tastes gravitate towards, and would love some critiques to think over. Thanks, guys!”

Then, the rest of the users comment on the photo regarding both artistic and technical aspects:

“The shot is great. Personally, I would have lightened the shadows, but that is just my taste. I am not afraid of noise in my photos, I don't think it is as annoying as others do. The photo is although already light enough to see every detail you need to see, and also the small contours on the deer's fur is an interesting eyecatcher. This is perhaps a once in a life time shot, so be proud of your work!”



Benefits

- **New source of data**
- **Rich description** of aesthetic aspects of the images
- **Free-form text** annotation of data, not limited to a single value
- **~17x images** than PCCD, **~3x avg.** words per comment than AVA-comments

Assessing Aesthetics from critiques

Sentiment Score

We propose to use the mean of the predicted sentiment polarity [2] of the comments as Aesthetic Score

$$s_k = \frac{\sum_{l=0}^2 p_l l}{2}, \quad s_i = \frac{1}{K} \sum_{k=0}^K s_k$$

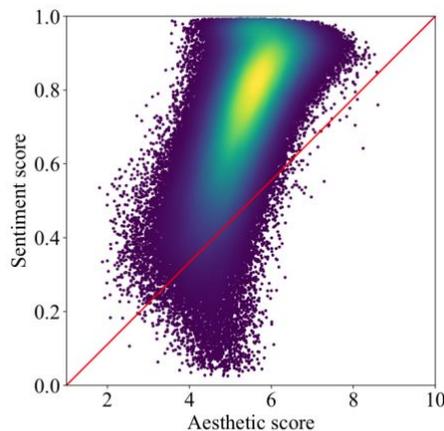
s_k : score for critique k

s_i : score for image i

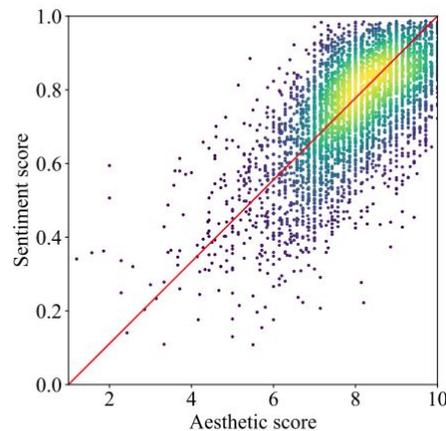
l : label for negative, neutral or positive sentiment respectively (0,1,2)

p_l : probability associated to each label

Correlation between our proposed Sentiment Score vs ground truth Aesthetic Score on AVA & PCCD datasets.



(a) AVA



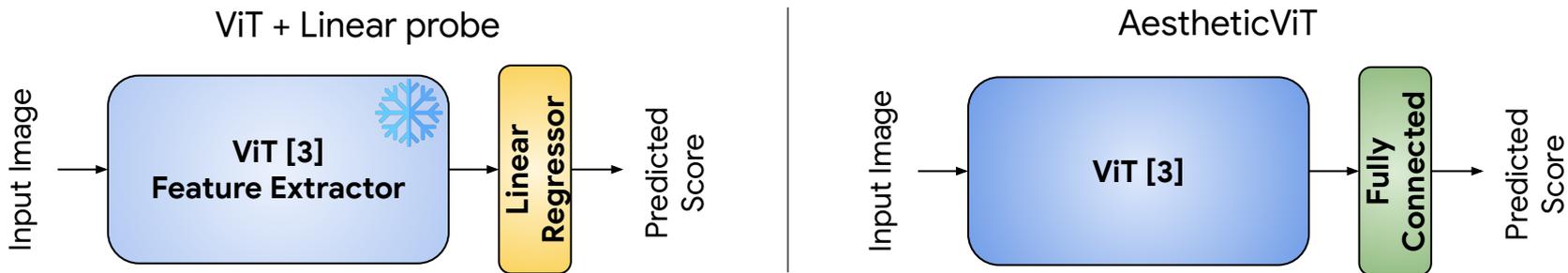
(b) PCCD

[2] [TweetEval: Unified benchmark and comparative evaluation for tweet classification](#)

Experiments

Table 2: Sentiment score baseline on the three considered datasets.

| Method | AVA | | | PCCD | | | RPCD | | |
|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | SRCC | LCC | Acc. (%) | SRCC | LCC | Acc. (%) | SRCC | LCC | Acc. (%) |
| NIMA [31] | 0.253 | 0.259 | 90.20 | 0.066 | 0.070 | 93.87 | 0.120 | 0.116 | 63.25 |
| ViT + Linear probe* | 0.570 | 0.570 | 76.43 | 0.156 | 0.165 | 93.04 | 0.172 | 0.173 | 64.58 |
| AestheticViT* | 0.544 | 0.550 | 90.54 | 0.228 | 0.262 | 93.86 | 0.250 | 0.253 | 65.27 |



[3] [An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale](#)

Experiments

Table 2: Sentiment score baseline on the three considered datasets.

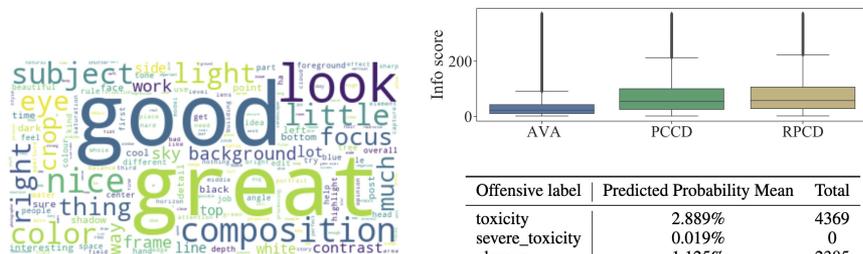
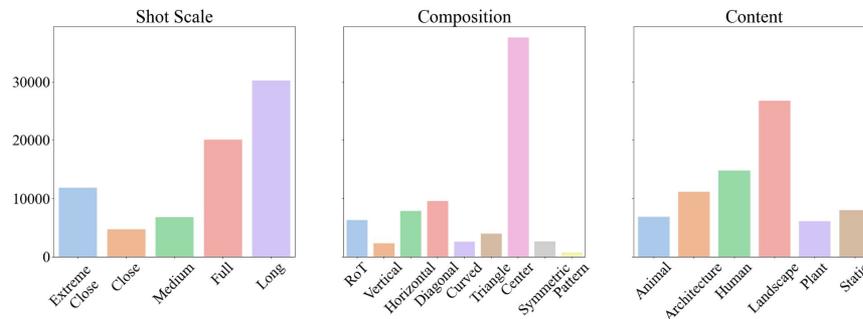
| Method | AVA | | | PCCD | | | RPCD | | |
|---------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | SRCC | LCC | Acc. (%) | SRCC | LCC | Acc. (%) | SRCC | LCC | Acc. (%) |
| NIMA [31] | 0.253 | 0.259 | 90.20 | 0.066 | 0.070 | 93.87 | 0.120 | 0.116 | 63.25 |
| ViT + Linear probe* | 0.570 | 0.570 | 76.43 | 0.156 | 0.165 | 93.04 | 0.172 | 0.173 | 64.58 |
| AestheticViT* | 0.544 | 0.550 | 90.54 | 0.228 | 0.262 | 93.86 | 0.250 | 0.253 | 65.27 |

Difficult for the model to predict from the images the more abstract aesthetic score!

Other annotated aspects

Apart from the sentiment score we **have automatically estimated** other aspects of the images and comments in RPCD

- **For images:** semantic content, shot scale type, composition rule
- **For comments:** main topics, informativeness, offensive contents



| Offensive label | Predicted Probability Mean | Total |
|-----------------|----------------------------|-------|
| toxicity | 2.889% | 4369 |
| severe_toxicity | 0.019% | 0 |
| obscene | 1.125% | 2385 |
| identity_attack | 0.374% | 336 |
| insult | 0.672% | 742 |
| threat | 0.418% | 287 |
| sexual_explicit | 0.259% | 439 |

Limitations

- **Limited number** of comments per image
- Better methods to use the comments as labels instead of sentiment labels automatically computed
- **Dataset bias:** Western cultural and geographical context





Understanding Aesthetics with Language: A Photo Critique Dataset for Aesthetic Assessment

Daniel Vera Nieto, Luigi Celona and Clara Fernández Labrador

<https://mediatechnologycenter.github.io/aestheval/>