WED-AM-166

PIQ23: An Image Quality Assessment Dataset for Portraits

Nicolas Chahine, Ana-Stefania Calarasanu, Davide Garcia-Civiero, Theo Cayla, Sira Ferradans, Jean Ponce

DXOMARK FOCUSED ON EXCELLENCE

Inria

ENS | PSL

cnrs

NYU

Outline

- Motivation
- PIQ23
- Advanced statistical analysis
- SEM-HyperIQA
- Summary

Motivation - Camera quality assessment

TID2013

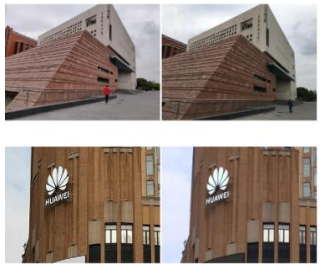
LIVE in the wild

Traditional BIQA

- Aesthetics
- Includes any content
- Global score
- Evaluated by casual users

CQA

- Specific attributes
- Predefined scenarios (charts)
- Sub-scores
- Usually evaluated by experts

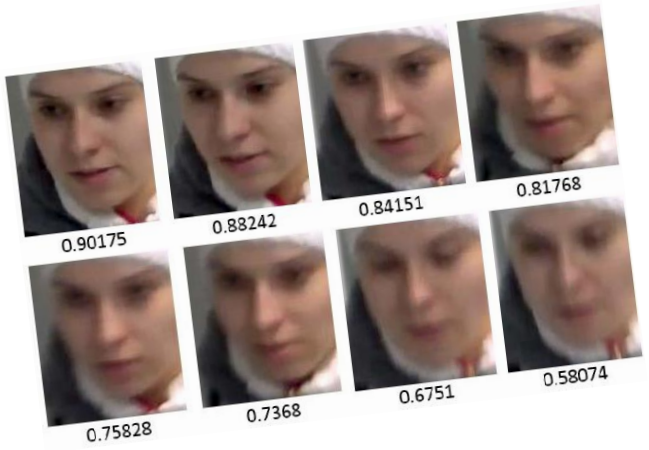


PCB2021

CARCO

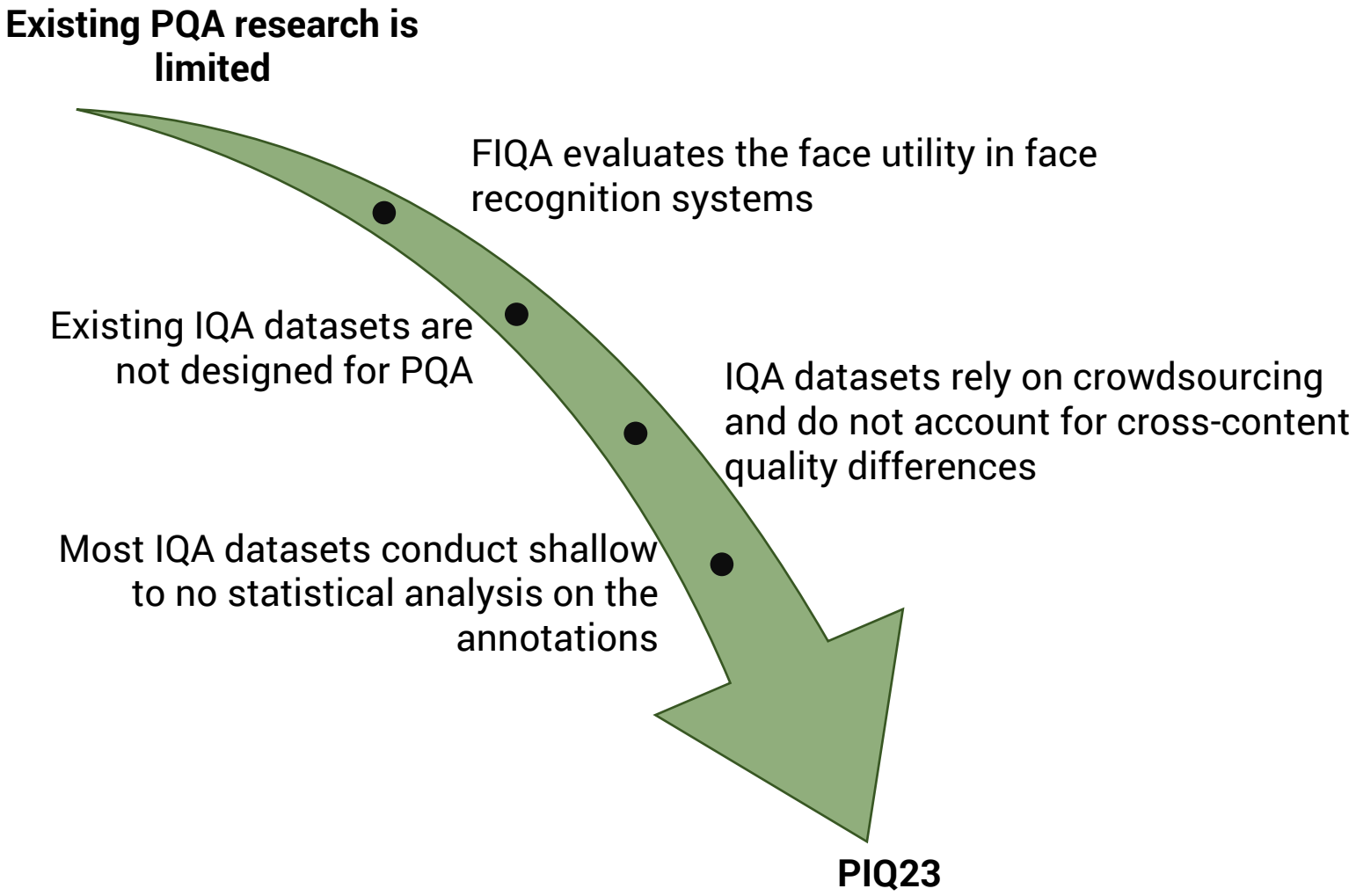
SPAQ

Motivation - Portrait quality assessment



Face IQA	PQA
Wide research area	Only limited research
Pre-processing step	Standalone metric
Utility for face recognition systems	Quality and aesthetics of portraits

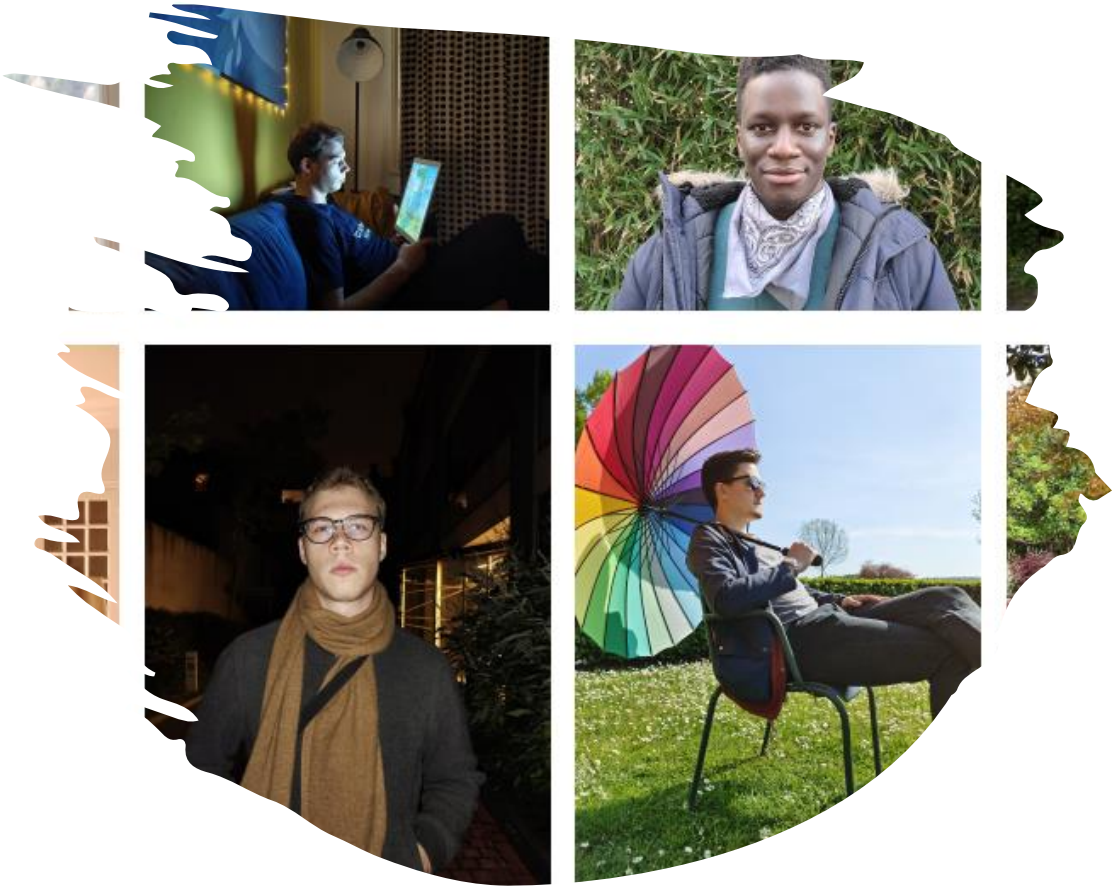
Motivation - Limitations



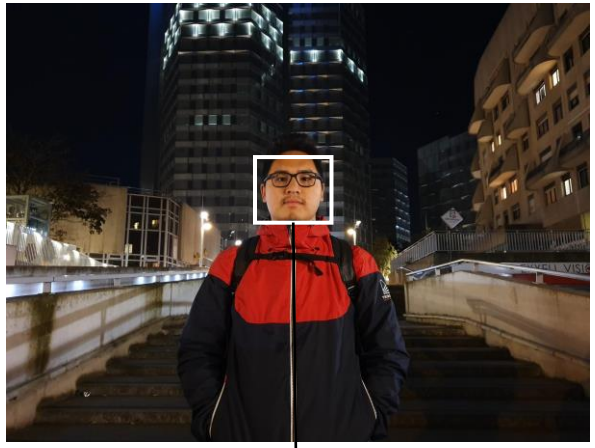
PIQ23 - Characteristics

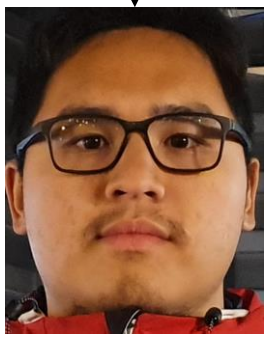
5,116 authentically distorted portrait images

- 50 pre-defined scenarios (of around 100 images each) covering four lighting conditions: outdoor, indoor, lowlight and night.
- 100+ smartphone devices from 14+ different brands, covering over a decade of smartphone camera technology.
- Individuals of different ethnicities and genders.
- Three quality attributes:
  - Face details preservation;
  - Target exposure on the face;
  - Overall portrait quality;
  - Global color quality (dropped).
- Annotated by 30+ image quality experts using pairwise comparisons, following specific guidelines.
- Annotated by pairwise comparisons and scaled to JOD units.
- Around 600k comparisons in total, averaging 4000 per scene for each attribute (1666+ hours of annotations).
- Complies with GDPR where individuals have explicitly granted permission for their image rights.



PIQ23 - Perceptual analysis





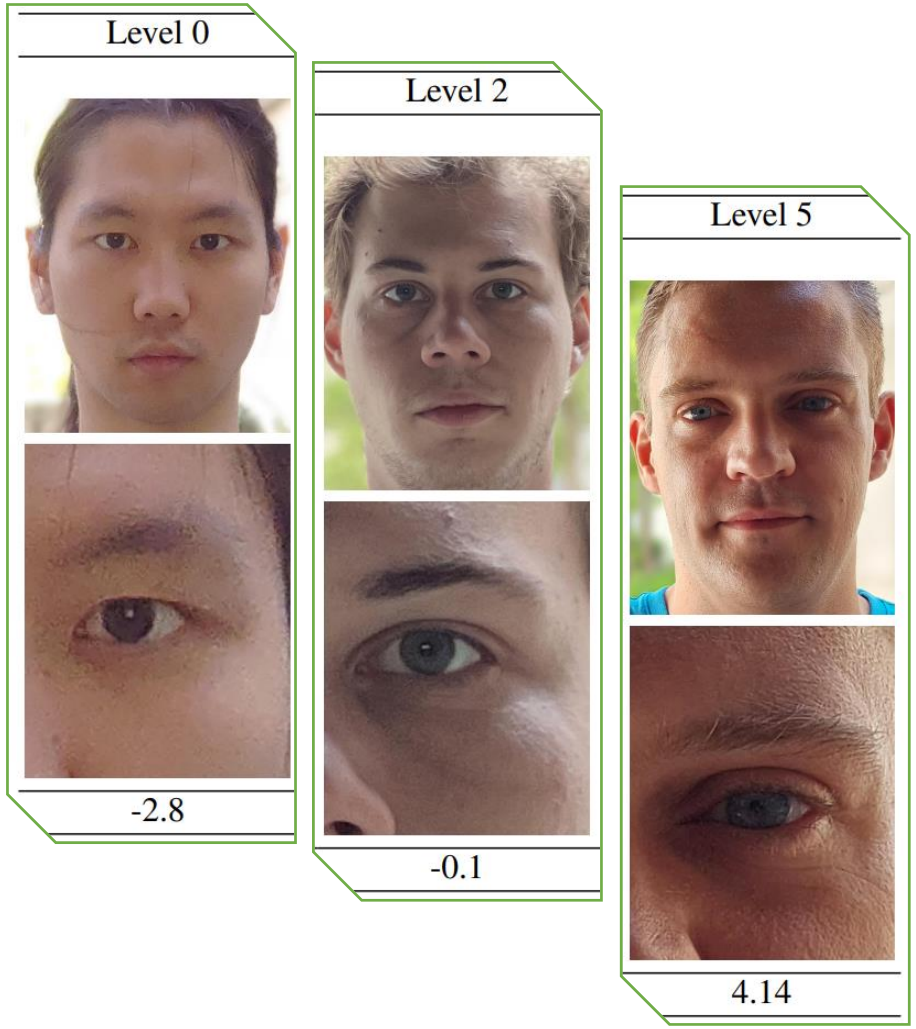
Each attribute was analyzed on a different ROI

Upscaled Face for Details and Exposure

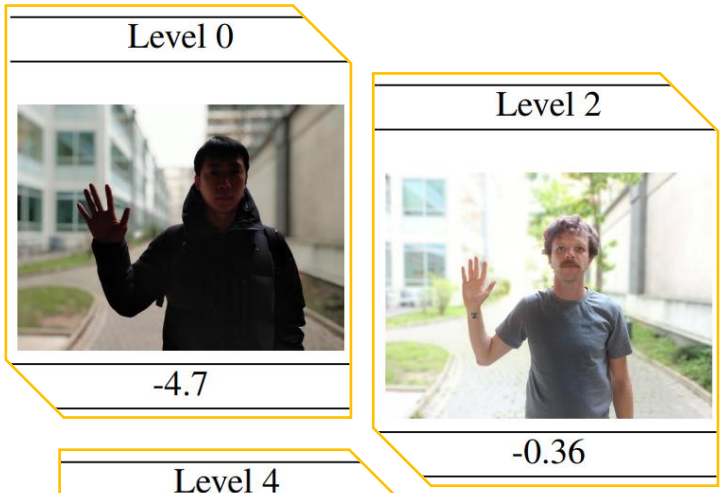
Rescaled image for Overall quality

PIQ23 - Perceptual analysis

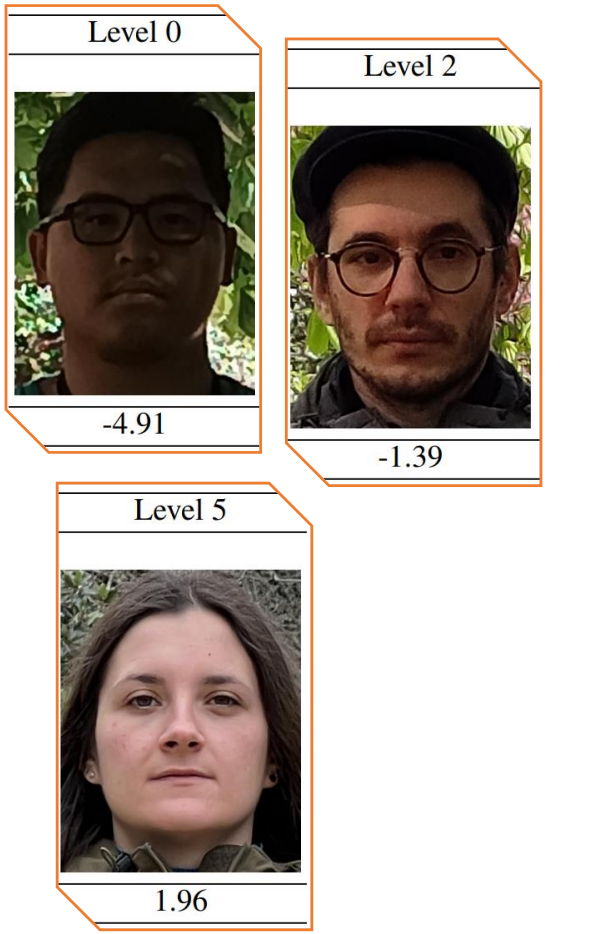
Face details preservation



Overall portrait quality



Target exposure on the face



Advanced statistical analysis - Motivation

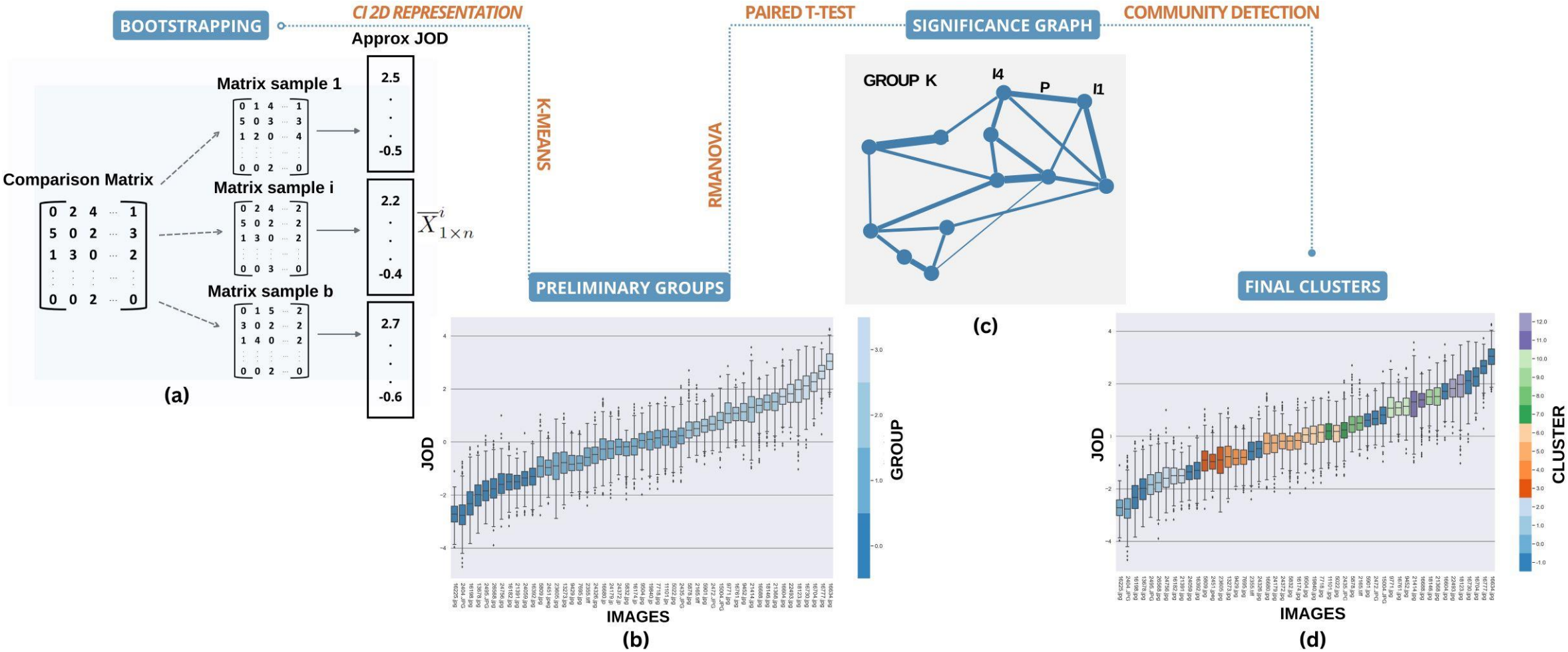
Develop a statistical analysis approach on PWC experiments

- Image quality annotations are noisy:
  - Intra-annotator variance;
  - Inter-annotator variance;
  - Sampling and annotator bias;
  - Scaling errors.

- Confidence intervals are not enough to evaluate the annotation uncertainty:
  - How to interpret the overlap of confidence intervals?
  - How to account for the score covariance due to the JOD scaling algorithm?

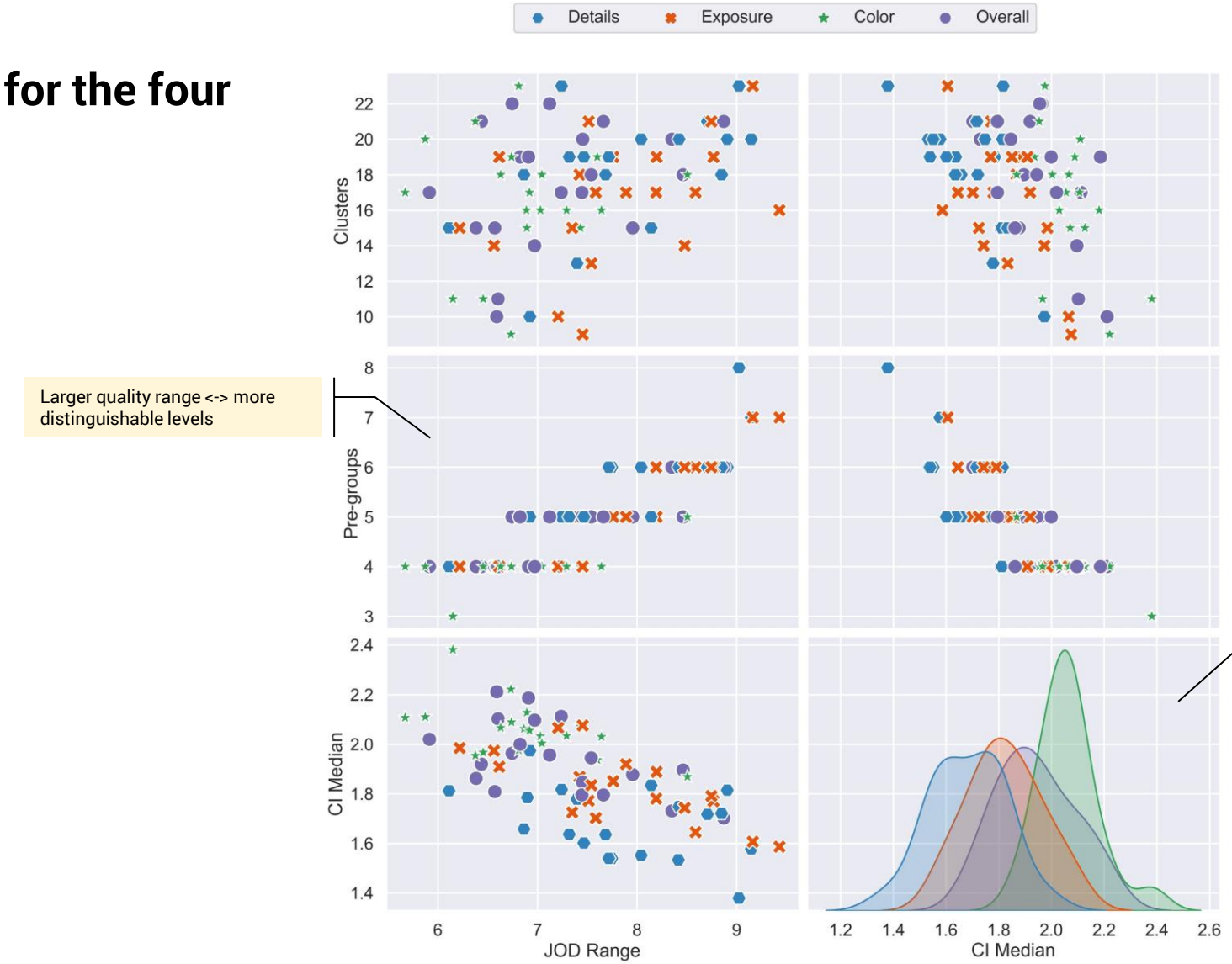
- We have developed a new statistical analysis method that arranges the images into groups of significantly different quality levels, thus precisely quantifying the JOD scale.

Advanced statistical analysis - Method



Advanced statistical analysis - Results

Results on 20 scenes for the four annotated attributes



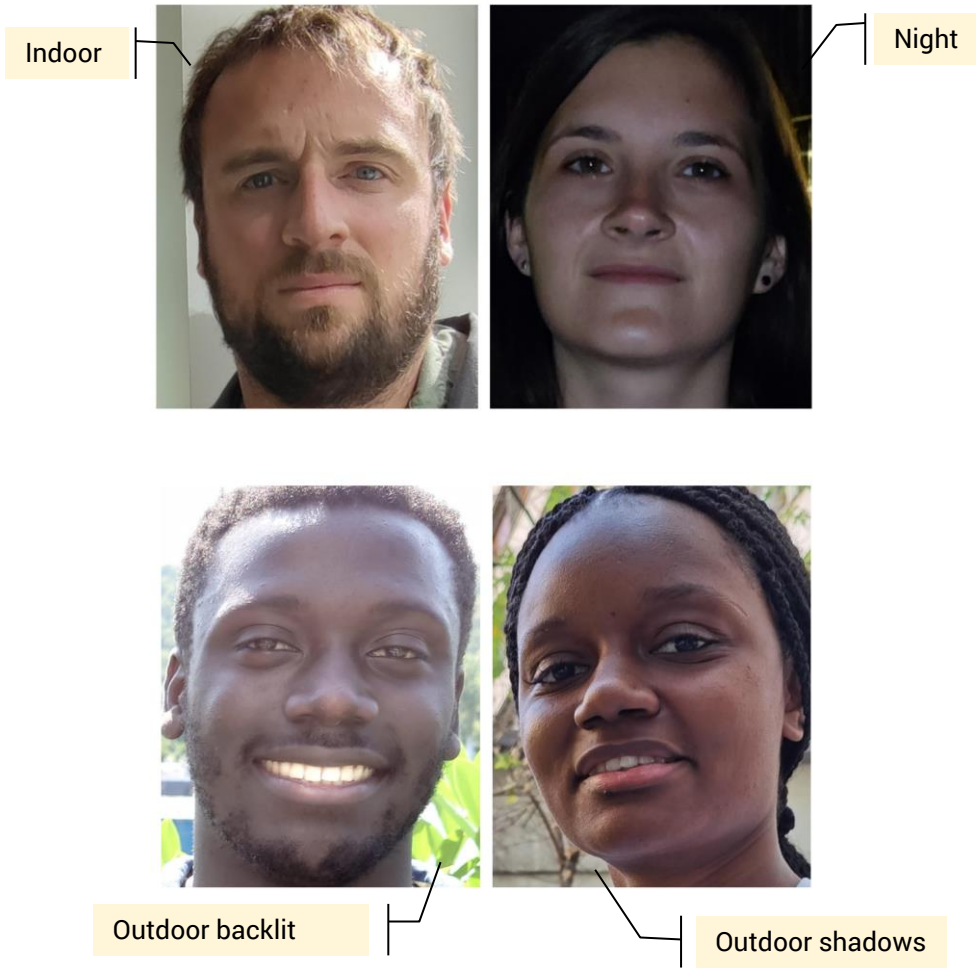
Blind IQA on PIQ23

How to train deep learning models on PIQ23?

Domain shift challenge

- Each scene has its own JOD scale due to the nature of psychometric scaling which is relative to the experiment's settings
- The quality is content dependent. For example, evaluating the target exposure of night conditions is different from sunny conditions

Target exposure quality with similar JOD on different lighting conditions



Blind IQA on PIQ23

How to train deep learning models on PIQ23?

Domain shift challenge

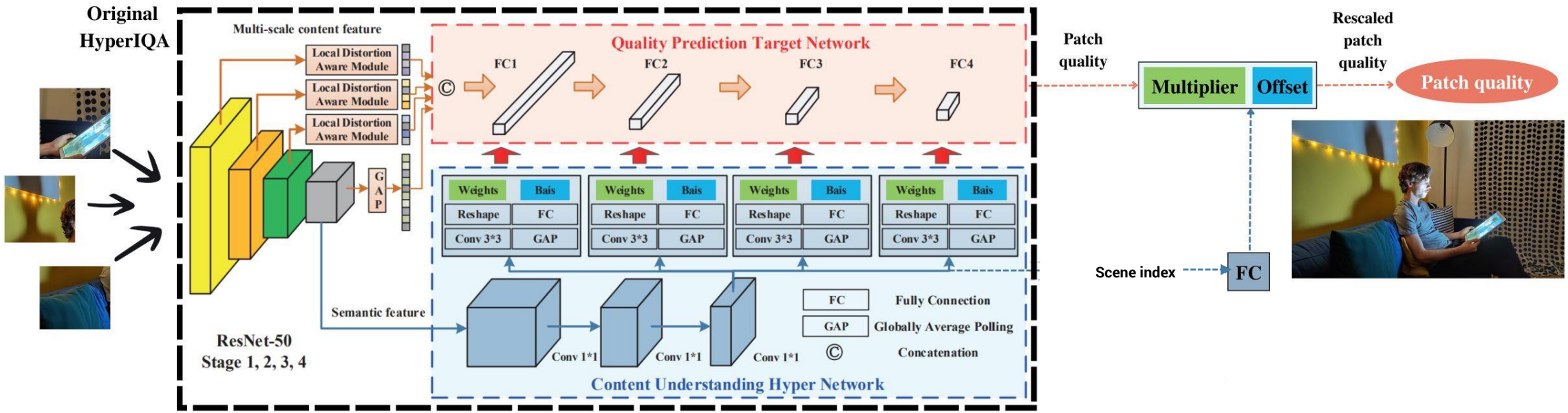
- Each scene has its own JOD scale due to the nature of psychometric scaling which is relative to the experiment's settings
- The quality is content dependent. For example, evaluating the target exposure of night conditions is different from sunny conditions

Solution

- S1: Rescaling the quality for each scene
- S2: Semantic aware quality predictions

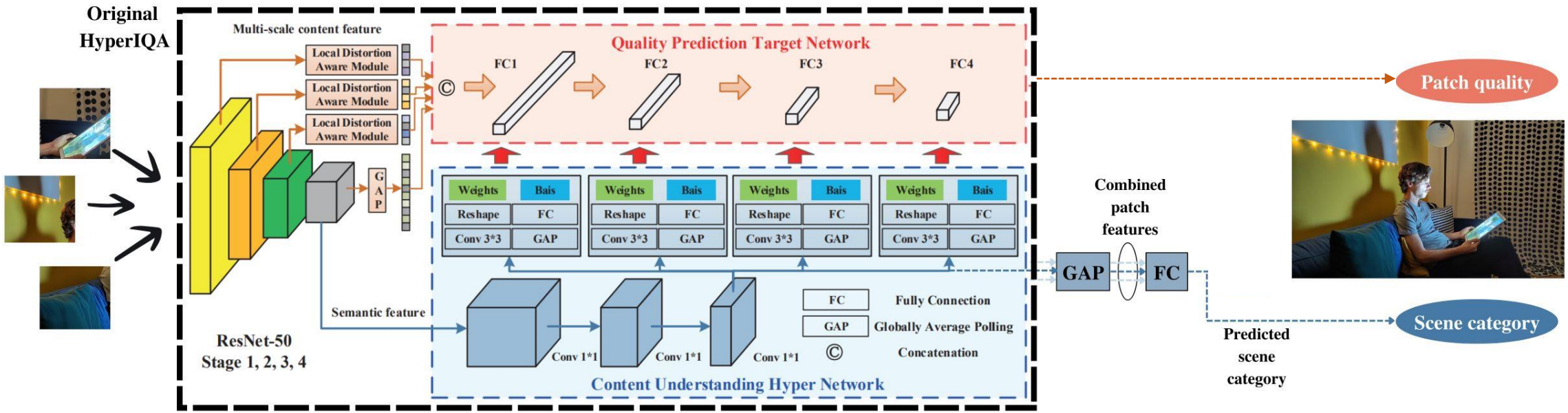
Blind IQA on PIQ23

SEM-HyperIQA-SO (S1)



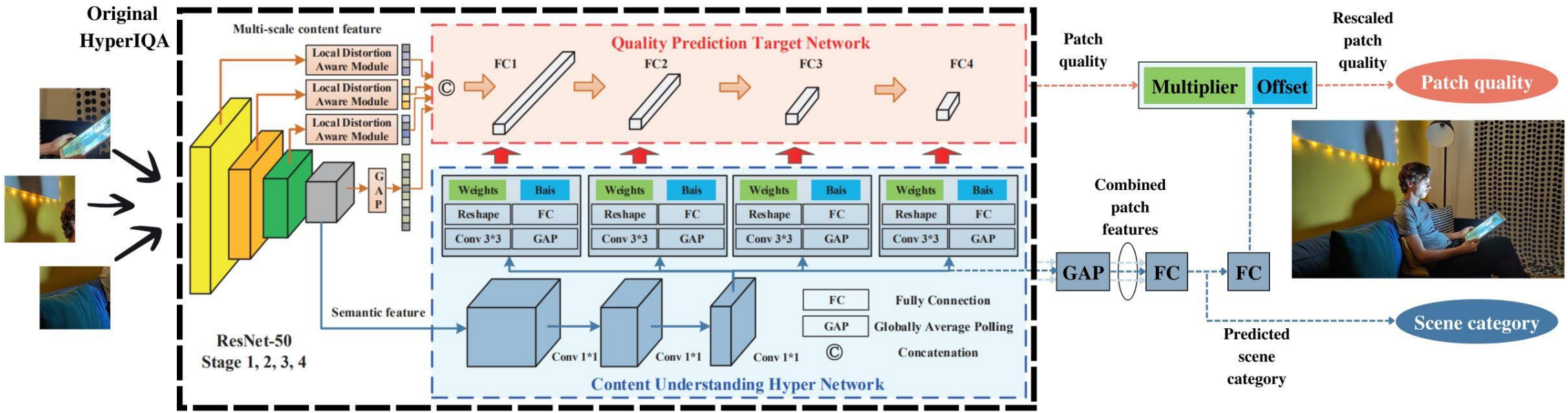
Blind IQA on PIQ23

SEM-HyperIQA-CO (S2)



Blind IQA on PIQ23

SEM-HyperIQA (S1 + S2)



Blind IQA on PIQ23 - Benchmarks

SROCC on PIQ23

	#	Method	Details	Exposure	Overall
	1	BRISQUE [36]	0.323	0.307	0.192
	2	NIQE [37]	0.378	0.265	0.298
	3	ILNIQE [64]	0.353	0.312	0.214
	4	DB-CNN [67]	0.628 ±0.07	0.635 ±0.06	0.555 ±0.07
	5	HyperIQA [52]	0.649 ±0.08	0.706 ±0.04	0.611 ±0.06
	6	MUSIQ [25]	<u>0.671</u> ±0.07	**0.725** ±0.04	0.589 ±0.07
S1+S2	7	SEM-HyperIQA	<u>0.671</u> ±0.07	0.71 ±0.04	<u>0.621</u> ±0.06
S1	8	SEM-HyperIQA-SO	**0.722** ±0.06	<u>0.721</u> ±0.06	**0.642** ±0.08
S2	9	SEM-HyperIQA-CO	0.664 ±0.07	0.71 ±0.06	0.621 ±0.07

Traditional approaches underperform

For Details and Exposure, the scene semantics cannot be easily extracted from the input (face ROI), thus only S1 applies, which explains the good performance of SEM-HyperIQA-SO. Our proposition outperforms MUSIQ for Details and is on par for Exposure.

For Overall, the scene semantics can be easily extracted (the input is a full image), thus S1 + S2 applies, since HyperIQA already extracts semantic information, SEM-HyperIQA and SEM-HyperIQA-CO have similar performances. The S1 solution boosts the results since we feed the index as input making the task easier.

Summary

- In this paper, we have introduced PIQ23, a portrait-specific image quality assessment dataset of 5k images on 50 scenes acquired by 100+ smartphones and annotated by 30+ experts.

- We have developed an advanced statistical analysis method to quantify the results of an IQA experiment.

- With an extensive benchmarking on PIQ23, we have demonstrated how a simplified approach to domain shift can enhance the results of IQA models.

THANK YOU !

24-26, quai Alphonse le Gallo - 92100 Boulogne-Billancourt - France
www.dxomark.com