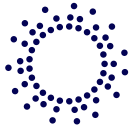
Junyong You, Norwegian Research Centre (NORCE), Bergen, Norway
Yuan Lin, Kristiania University College, Bergen, Norway
Jari Korhonen, University of Aberdeen, Aberdeen, UK

Half of An Image Is Enough for Quality Assessment

Motivation



- How does a deep net for IQA work?
- Do IQA models need all image regions?
- Any relationship between IQA and semantic meanings?

Motivation and approach

- How does a deep net for IQA work?
- Do IQA models need all image regions?
- Any relationship between IQA and semantic meanings?

- Positional masking on TRIQ model for region importance for IQA
  - J. You, and J. Korhonen, “Transformer for image quality assessment,” ICIP’21
- Semantic measures

Can traditional XAI models work on IQA?

- LIME [1]
- Grad-CAM [2]

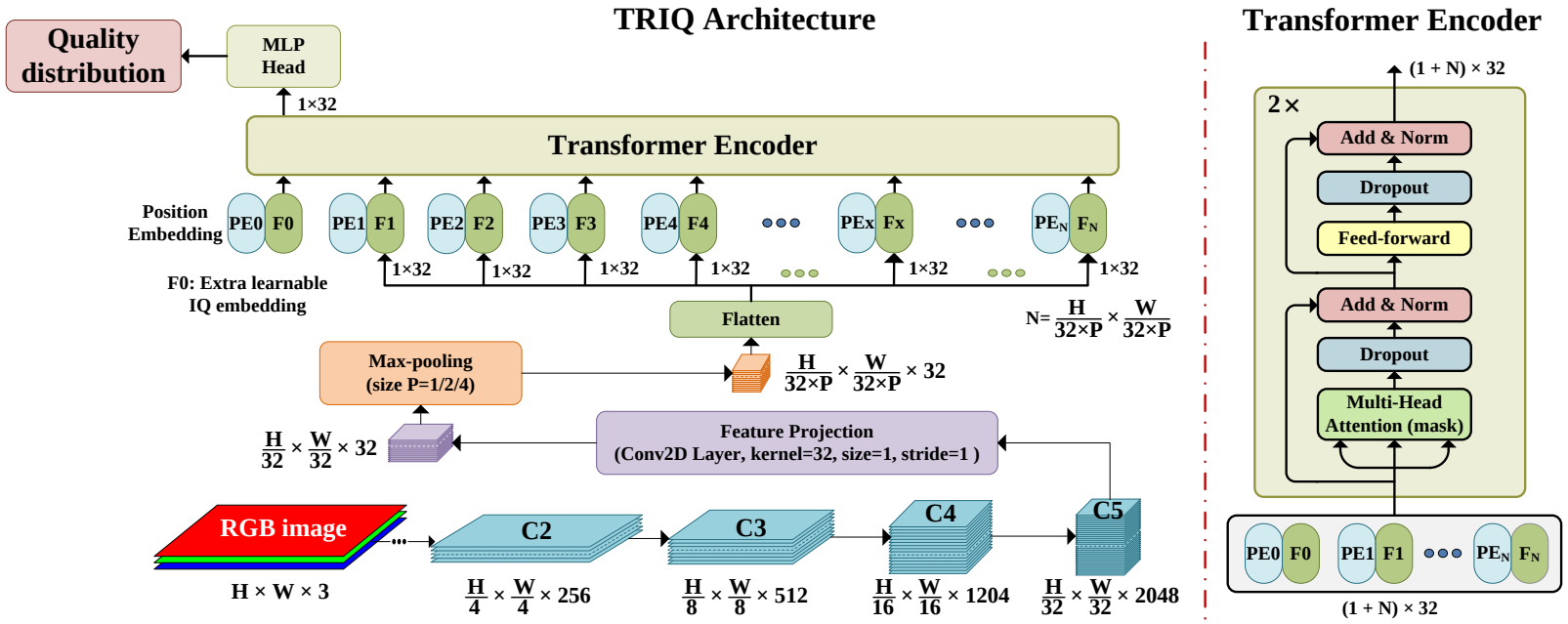
Original mages LIME maps Grad-CAM maps

Derived from the TRIQ model

[1] M.T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. ACM SIGKDD Knowl. Discov. Data Min. Aug. 2016, San Francisco, CA, USA.
[2] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2017, Venice, Italy.

Positionally masked TRIQ

- Insert a masking operation in transformer encoder
  - Adding a very high negative scalar value (-1e9) to the dot product of query and key at the masked positions in MHA
  - The Softmax produces outputs (attention weights) close to 0 at the masked positions

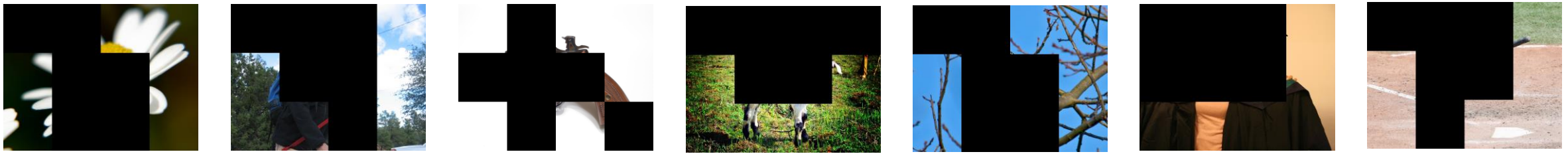


Region importance in TRIQ for IQA

- Feature matrix: 24x32x2048, divided into 3x4 grids, each block contains 8x8 feature vectors (2048 features)
- Manually masking *n* blocks, *n*=1,2,…,11, in TRIQ
- Correlation between masked TRIQ and original TRIQ indicates region importance of a block
- Loop *n* through all possible combinations in C_{12}^{n}
- Average correlations of individual regions over all combinations

Region importance in TRIQ for IQA

- Divide the 12 regions into two groups based on correlations
- High correlation → low importance (trivial regions)
- Low correlation → high importance (important regions)

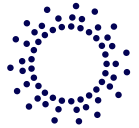


Trivial regions

Important regions

- KonIQ dataset: 0.97 (trivial regions) VS 0.67 (important regions)
- SPAQ dataset: 0.96 (trivial regions) VS 0.68 (important regions)

Semantic and explainable image measures



- **Saliency**: degrees of regions in an image where viewers fixate with high priority
- **Spatial frequency**: characteristic of patterns that are periodic across position in space
  - Important regions often contain objects with mid-range of spatial frequencies
  - In accordance with CSF
- **Objectness**: likelihood of a group of pixels in an image to be a (foreground) object

- High matching degree between the measures and important regions

Experiment I: Is half image enough for IQA?

- Zeroing important regions and trivial regions respectively, and then run IQA models
- Correlation on predicted quality scores between zeroed images and original images

Models	KonIQ-10k					SPAQ				
	PLCC on test sets	Important regions		Trivial regions		PLCC on test sets	Important regions		Trivial regions	
		Predicted	MOS	Predicted	MOS		Predicted	MOS	Predicted	MOS
TRIQ	0.922	0.680	0.671	0.972	0.902	0.916	0.695	0.675	0.969	0.907
AIHIQnet	0.929	0.816	0.804	0.959	0.896	0.928	0.810	0.801	0.958	0.912
Koncept	0.916	0.817	0.813	0.960	0.901	0.831	0.802	0.788	0.957	0.811
DBCNN	0.856	0.804	0.796	0.963	0.887	0.894	0.808	0.797	0.960	0.868
Swin-IQA	0.956	0.835	0.811	0.974	0.921	0.933	0.804	0.782	0.963	0.902

("PLCC on test sets" indicate the performance of individual models, "Important regions" means zeroing the important regions, "Trivial re-gions" zeroing the trivial regions, "Predicted" is the PLCC between model predicted quality values on the original images and the zeroed images on the test sets; and "MOS" indicates the PLCC between the model predicted quality on zeroed images and the MOS values on original images on the test sets)

Experiment II: Matching degree of semantics

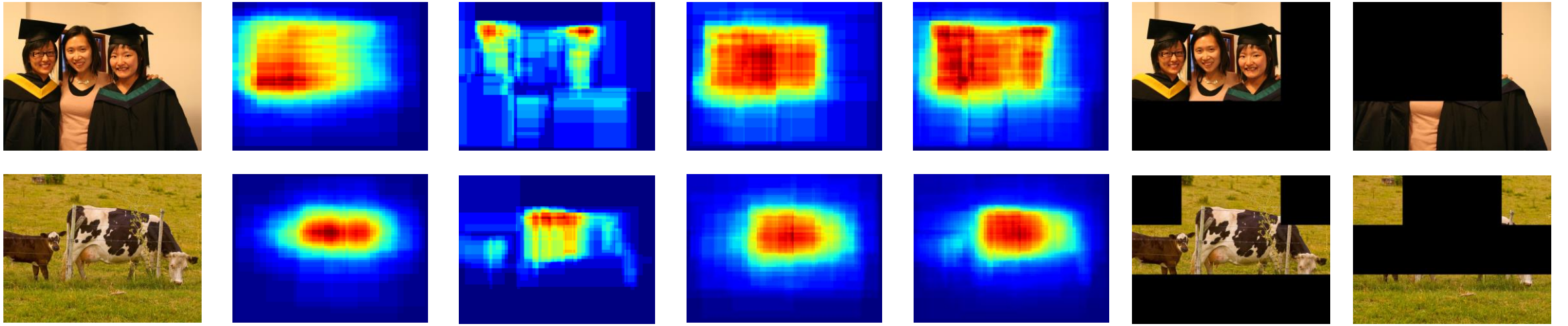


Image | saliency | spatial frequency | objectness | average | important regions | trivial regions

- A quantitative analysis of matching degrees also performed

Experiment II: Matching degree of semantics

- Quantitative analysis of matching degrees between semantic measures and important/trivial regions

Measures	T=1	T=2	T=3	T=4	T=5
Saliency	100	96	69.2	52.4	37.8
Frequency	100	99	70.1	60.1	40.8
Objectness	100	100	77.3	66.6	44.7
Averaged	100	100	79.5	65.8	45.6

Taking an example of T=4, the result shows that for 66.6% images in the two datasets, at least 5 regions wither higher values of objectness measure belong to the 6 important regions for IQA, whilst other 5 or 6 regions with lower objectness levels are trivial for IQA.

Conclusion

- Half of an image might be enough for a CNN-based IQA model

- Three semantic measures (saliency, spatial frequency, objectness) show high accordance with IQA perception