Re-IQA: Unsupervised Learning for Image Quality Assessment in the Wild

IEEE/CVF CVPR'2023 Poster ID # TUE-PM-164
Vancouver, BC, Canada, June 18-22, 2023



Avinab Saha*
avinab.saha@utexas.edu

Sandeep Mishra*
sandy.mishra@utexas.edu

Alan C. Bovik
bovik@ece.utexas.edu

Laboratory of Image and Video Engineering, The University of Texas at Austin

Overview of Re-IQA

- Perceptual *Image* ***Q****uality* ***A****ssessment* ***(IQA)*** affects ***billions of internet and social media users daily***
- We propose a ***Mixture of Experts*** approach to independently train two encoders to learn image features relating to
  - ***High Level Image Content*** (Content Aware Encoder)
  - ***Low Level Technical Image Quality*** (Quality Aware Encoder)
- Encoders are trained in an ***Unsupervised setting***
- We call this framework to train the encoders ***Re-IQA***
- For ***IQA in-the-Wild***, complementary ***low*** & ***high*** level image representations are used to ***train a regressor*** to map ***image representations*** to ground truth **M**ean **O**pinion **S**cores (MOS)

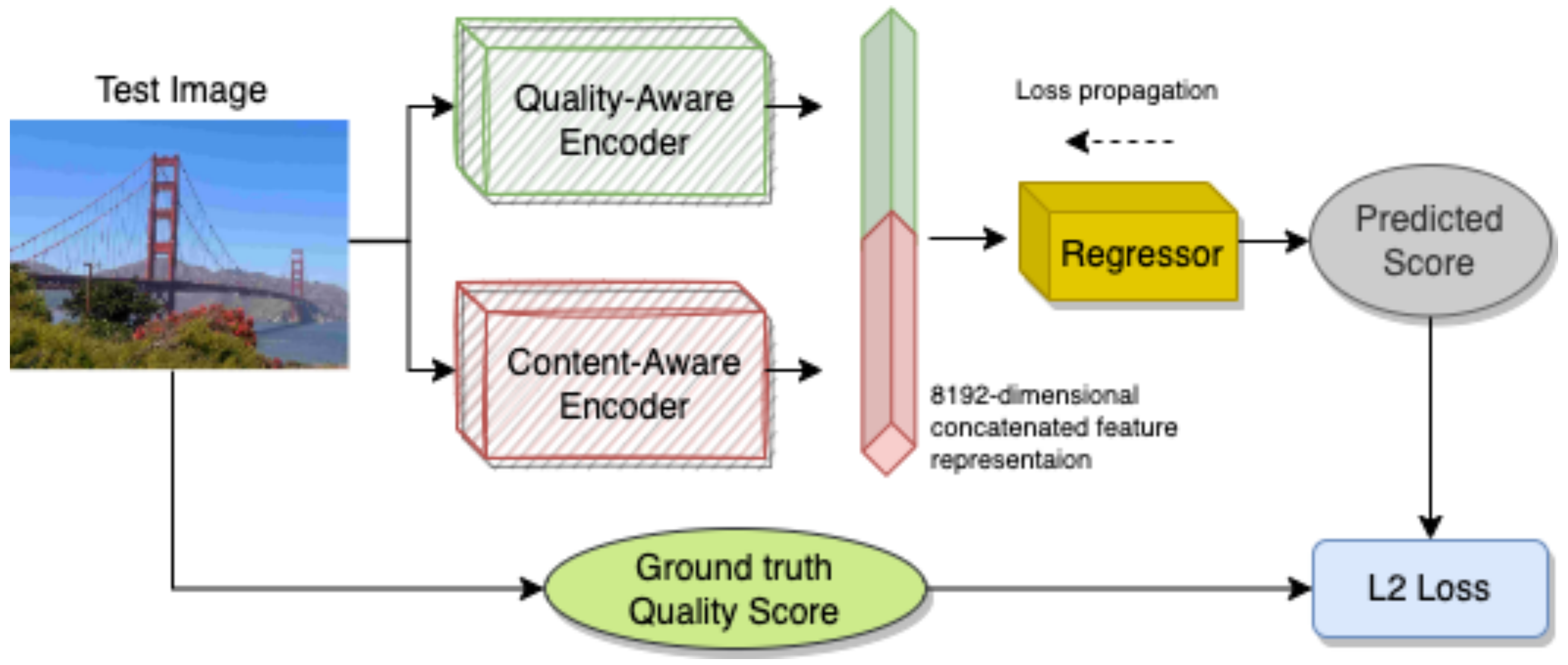


Fig 1 : Content-Aware and Quality-Aware encoders are frozen while the regressor learns to map image representations to quality predictions

No Reference IQA : Challenges

- No-Reference IQA for ***Images in the Wild*** presents challenges due to the ***complex interplay*** among the various kinds of ***distortions***

- Due intricate nature of human visual system, ***image content affects quality perception***

- Image distortion perception is highly ***content dependent***, and is heavily affected by content related ***perceptual processes like masking***

(a) JPEG Compressed : 1 (b) JPEG Compressed : 2 (c) Motion Blur - Camera Shake (d) Overlaid Film Grain/ Noise

Fig 2 : Exemplar Synthetically and “In-the-wild” distorted pictures

No Reference IQA : Challenges

- Also well-known, perceived quality does not ***correlate well with image metadata*** like
    - Image resolution, file size, color profile, compression ratio etc
- Unlike Full-Reference IQA, that has access to the pristine source image, No-Reference IQA (NR-IQA) ***lacks both the information about source image & applied distortions***.

(a) JPEG Compressed : 1 (b) JPEG Compressed : 2 (c) Motion Blur - Camera Shake (d) Overlaid Film Grain/ Noise

Fig 2 : Exemplar Synthetically and "In-the-wild" distorted pictures

Our Method

- Our work draws inspiration from ***Momentum Contrastive Learning*** methods' success in image classification task
- We engineer ***Re-IQA*** to learn ***content*** and ***quality-aware*** image representations for NR-IQA on ***real, authentically distorted*** pictures
- ***Mixture of Experts*** approach is used to train two encoders to learn image features relating to
    - **Expert 1** : ***High Level Image Content*** (Content Aware Encoder)
    - **Expert 2** : ***Low Level Technical Image Quality*** (Quality Aware Encoder)
- The representations from both encoders are utilized to train a regressor that maps ***image representations*** to ground truth ***M***ean ***O***pinion ***S***cores (MOS)

Key Contributions

- ***Unsupervised representation learning*** framework for ***low-level image quality*** that are ***complementary*** to ***high-level image-content*** representations
    - ***Mixture of Content and Quality Features*** achieve competitive image quality predictions compared to existing SoTA methods
- Proposed a novel ***Image Augmentation*** and ***Intra-Pair Image Swapping*** scheme to enable learning of **low-level image quality** representations
    - ***Dynamic*** nature of Image Augmentation prevents learning of ***discrete distortion classes*** enforcing learning of ***perceptually relevant*** image-quality features

Re-IQA : Content Aware

- ***Unsupervised pre-trained MoCo-v2**** Resnet-50 backbone trained on ImageNet database is used as the ***Content-Aware Module***
- ***High Level Working*** of MoCo-v2 :
    - Two augmented crops of same image are labelled as ***positive pairs***
    - Crops from different images are labelled as ***negative pairs***
    - Positive & negative pairs are used in the ***InfoNCE loss*** to train the networks
- ***Issue :*** Two augmentations of even the same image crop have varying Image quality
    - MoCo-V2 framework needs to ***modified*** for learning ***quality-aware*** representations

*https://arxiv.org/abs/2003.04297

Re-IQA : Quality Aware

- MoCo-v2 framework is modified using our proposed ***Image Augmentation*** and ***Intra-Pair Image Swapping*** scheme
- We also use the ***3 hypothesis*** inspired by our knowledge of distortion perception in the HVS
  - ***H1***: For two overlapping crops from the same image
    - higher overlap => more similar quality features
  - ***H2***: crops with different content => dissimilar quality features
  - ***H3***: same crop, different distortion => dissimilar quality features
- ***Augmentation bank*** comprises of ***25 distortion methods***, each realized at ***5 levels*** of severity

Re-IQA : Quality Aware (Contd.)

- Using n unique distortion augmentations from the bank on ***two overlapping crops*** (c_1, c_2) of the training image (i^k), we define a chunk of images as :

$$chunk^{k_{c1}} = [i^{k_{c1}}, {i_1}^{k_{c1}}, {i_2}^{k_{c1}}, \dots, {i_n}^{k_{c1}}]$$
$$chunk^{k_{c2}} = [i^{k_{c2}}, {i_1}^{k_{c2}}, {i_2}^{k_{c2}}, \dots, {i_n}^{k_{c2}}]$$

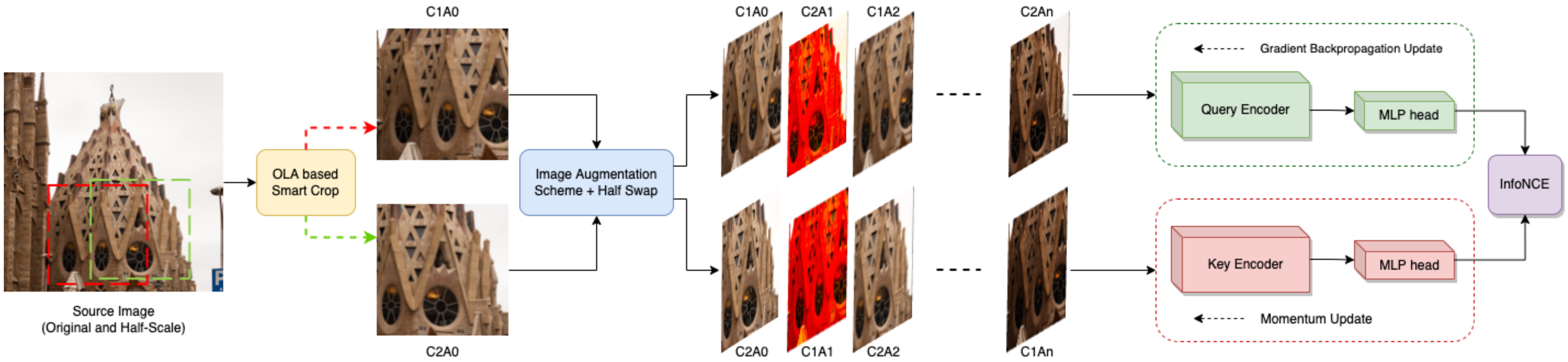
Similar Quality as obtained as per H1

- The ***Intra-Pair Image Swapping*** scheme then generates the following pairs

$$\begin{array}{cccccccc} i^{k_{c1}} & {i_1}^{k_{c1}} & \dots & {i_m}^{k_{c1}} & {i_{m+1}}^{k_{c1}} & \dots & {i_{n-1}}^{k_{c1}} & {i_n}^{k_{c1}} \\ \updownarrow & \updownarrow & & \updownarrow & \updownarrow & & \updownarrow & \updownarrow \\ i^{k_{c2}} & {i_1}^{k_{c2}} & \dots & {i_m}^{k_{c2}} & {i_{m+1}}^{k_{c2}} & \dots & {i_{n-1}}^{k_{c2}} & {i_n}^{k_{c2}} \end{array} \longrightarrow \begin{array}{cccccccc} i^{k_{c1}} & {i_1}^{k_{c2}} & \dots & {i_m}^{k_{c1}} & {i_{m+1}}^{k_{c2}} & \dots & {i_{n-1}}^{k_{c1}} & {i_n}^{k_{c2}} \\ \updownarrow & \updownarrow & & \updownarrow & \updownarrow & & \updownarrow & \updownarrow \\ i^{k_{c2}} & {i_1}^{k_{c1}} & \dots & {i_m}^{k_{c2}} & {i_{m+1}}^{k_{c1}} & \dots & {i_{n-1}}^{k_{c2}} & {i_n}^{k_{c1}} \end{array} \longrightarrow \begin{array}{l} [i_m^{k_{c1}}, i_m^{k_{c2}}] \mapsto similar-quality \\ [i_m^{k_{c1}}, i_l^{k_{c2}}] \mapsto different-quality \\ [i_m^{k_{c1}}, i_l^{k_{c1}}] \mapsto different-quality \\ [i_m^{k_{c1}}, i_l^{j_{c2}}] \mapsto different-quality \end{array}$$

Re-IQA : Quality Aware (Contd.)

- We train the ***Query*** encoder by back-propagating the ***InfoNCE*** loss of the batch calculated from the output features obtained using the paired inputs
- The weights of the ***Key*** encoder are updated using the momentum update method



Re-IQA : Quality Aware (Contd.)

- We conducted extensive ablation studies to select the hyper-parameters :
  - ***Number of Augmentations*** (n_{aug}) used to generated a chunk
  - ***Patch Size :*** Size of Crops used in training Re-IQA Quality-Aware module
  - ***Overlapping Area Bound*** between crops from a same image
- Based on our results, the following configurations were chosen :
  - n_{aug} : 11 , Patch Size : 160 , OLA Bound : 10-30%
- Performance comparison among various configurations can be found in ***Table 1 (Main Paper)***

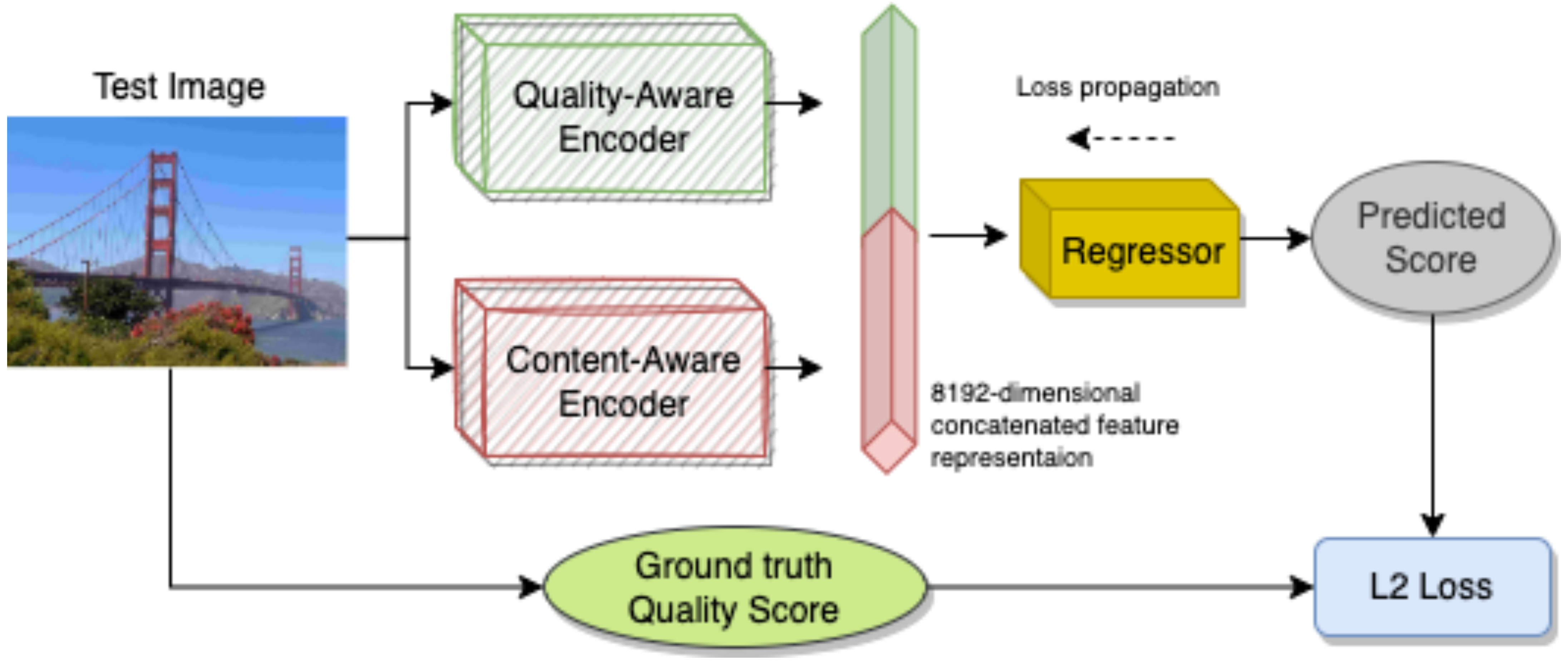
Training Dataset

- For Content-aware model: ImageNet-1K (~ 1.28 million images)
- For Quality-aware model we use a combination of authentic and synthetic distorted images from the following databases:
  - KADIS[1] : We use the 140,000 pristine images in the dataset
  - AVA[2] : 225,000 authentically distorted images
  - COCO[3] : 330,000 authentically distorted images
  - CERTH-Blur[4] : 2450 authentically distorted images
  - VOC[5] : 33,000 authentically distorted images

[1] Hanhe L, et al. Kadid-10k: A large-scale artificially distorted iqa database. IEEE QoMEX 2019
[2] Naila M, et al. Ava: A large-scale database for aesthetic visual analysis.CVPR 2012
[3] Tsung-Yi L, et al. Microsoft coco: Common objects in context. ECCV 2014
[4] Eftichia M, et al. No-reference blur assessment in natural images using fourier transform and spatial pyramids.ICIP 2014
[5] Everingham M, et al. The pascal visual object classes (voc) challenge. IJCV 2010

IQA Regression

- Image is passed through the ***two frozen pre-trained encoders*** to generate image representations
- The concatenated image representations are fed to a **Linear Regressor** to predict a quality score
- The predicted quality score is compared with the ground truth human opinion score (**MOS**) to train the Regressor head



Evaluation Datasets

- We evaluate our model on both UGC ***(In-the-Wild)*** and Synthetic datasets
- ***UGC-IQA Datasets:***
    - KonIQ[1] (10,000), SPAQ[2] (11,000), CLIVE[3] (1162), FLIVE[4] (40,000)
- ***Synthetic-IQA Datasets:***
    - LIVE-IQA[5] (779), CSIQ-IQA[6] (866), TID-2013[7] (3,000), KADID[8] (10,125)

[1] Hosu, V., et al., KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. IEEE TIP 2020
[2] Fang, Y., et al., Perceptual quality assessment of smartphone photography. CVPR 2020
[3] Ghadiyaram, D. et al., Massive online crowdsourced study of subjective and objective picture quality. IEEE TIP 2015
[4] Ying Z., et al., From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. CVPR 2020
[5] Sheikh, H.R., et al., A statistical evaluation of recent full reference image quality assessment algorithms. IEEE TIP 2006
[6] Larson, E.C. et al., Most apparent distortion: full-reference IQA and the role of strategy. Journal of Electronic Imaging 2020
[7] Ponomarenko, N., et al. , Color image database TID2013: Peculiarities and preliminary results. IEEE EUVIP 2013.
[8] Lin, H., et al., KADID-10k: A large-scale artificially distorted IQA database. IEEE QoMEX 2019

Objective NR-IQA Results

- Checkout ***Table 2 (Main Paper)*** for comparison of our proposed method with various SoTA algorithms

- Our method performs at par with MUSIQ[1], which is built on Transformers

- Our model ***performs better*** than ***SoTA methods*** on ***most datasets***, and competitively similar on the rest.

Method	FLIVE (SRCC ↑)	SPAQ (SRCC ↑)
HyperIQA[2]	0.535	0.916
CONTRIQUE[3]	0.580	0.914
MUSIQ (Transformer based)	**0.646**	**0.917**
Re-IQA (Content + Quality Experts)	**0.645**	**0.918**

Table 1: Comparison of SRCC scores of Re-IQA against MUSIQ (Transformer based approach), Hyper-IQA and CONTRIQUE (CNN based approach) on UGC-IQA datasets

[1] Ke, J., et al.. Musiq: Multi-scale image quality transformer. IEEE/CVF ICCV 2021
[2] Su, S., et al. Blindly assess image quality in the wild guided by a self-adaptive hyper network. CVPR 2020
[3] Madhusudana, P.C., et al. Image quality assessment using contrastive learning. IEEE TIP 2020

Objective NR-IQA Results

- Our method performs ***better than most*** when evaluated on ***Synthetic datasets***

- For some datasets, ***Quality-only expert performs better than Mixture-of-experts***

Method	LIVE-IQA (SRCC ↑)	CSIQ-IQA (SRCC ↑)
HyperIQA	0.962	0.923
CONTRIQUE	0.960	0.942
Re-IQA (Quality Expert only)	**0.971**	**0.944**
Re-IQA (Content + Quality Experts)	**0.970**	**0.947**

Table 2: Comparison of SRCC scores of Re-IQA (Content+Quality Experts), Re-IQA (Quality Expert only), Hyper-IQA and CONTRIQUE (CNN based approach) on Synthetic IQA datasets

Cross Database Generalization

- ***Cross-database*** generalization isCchallenging NR-IQA problem
  - Common phenomenon that model performance ***degrades*** when ***trained and tested*** on ***different datasets***
- ***SRCC Comparison*** of cross database generalization of Re-IQA with SoTA NR-IQA methods shown below

Training Database	Testing Database	NR-IQA Algorithms			
		PQR	HyperIQA	CONTRIQUE	Re-IQA
CLIVE	**KonIQ**	0.757	**0.772**	0.676	**0.769**
KonIQ	**CLIVE**	0.770	**0.785**	0.731	**0.791**
LIVE-IQA	**CSIQ-IQA**	0.719	0.744	**0.823**	**0.808**
CSIQ-IQA	**LIVE-IQA**	0.922	**0.926**	0.925	**0.929**

- Re-IQA has ***superior cross-database generalizability***!

Conclusion

- Developed a holistic approach to ***NR-IQA*** by individually targeting the impact of ***content*** and ***distortion*** on the ***overall image quality score***
- ***Re-engineered*** the MoCo-v2 framework for learning ***quality-aware*** representations to include our proposed ***Image Augmentation***, ***OLA-based smart cropping***, and ***Intra-Pair Swapping*** scheme
- Results show ***Re-IQA consistently achieves SoTA performance*** on eight popular NR-IQA databases
- Lastly, our method is ***flexible to encoder architecture designs*** and can be extended to other CNN and Transformer based models.