Multi-Scale Features and Parallel Transformers Based Image Quality Assessment

Presented by:
Abhisek Keshari

Guide:
Dr. Vinit Jakhetiya

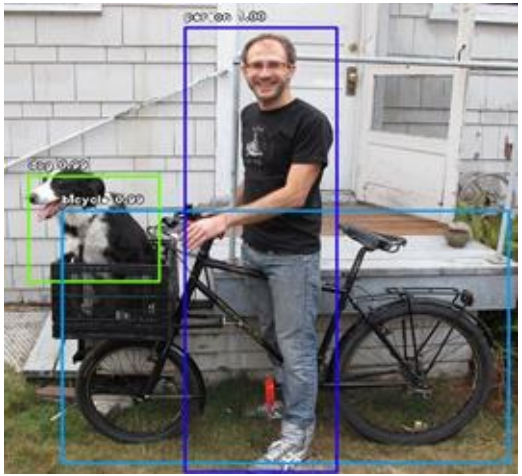
Outline

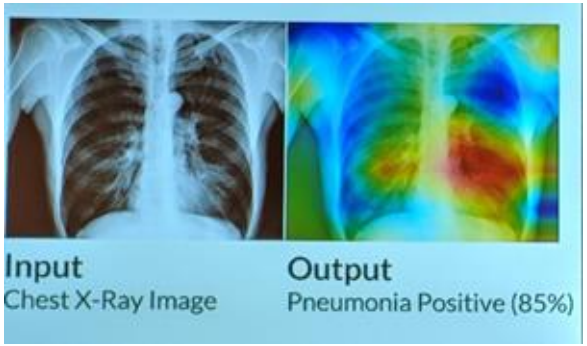
- Introduction
- Image Quality Assessment
  - Distortions
  - Type of Distortions
  - Mean Opinion score & Other Metrics
- Datasets Overview
- Problem Statement
- Proposed Solution
- Results
- Conclusion

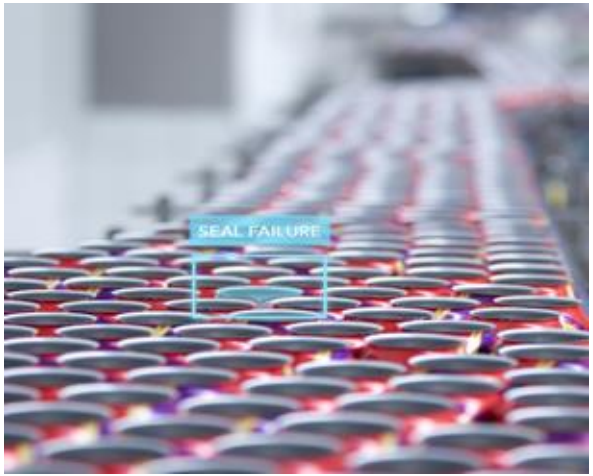
Introduction

Computer Vision: Unveiling Visual Insights

- Emulates human sight to interpret visual data.
- Methods decode raw pixels into content comprehension.
- **Applications :** Identifies objects, Health care, Quality control.







Introduction

Image Processing: Enhancing Visual Information

- Manipulation, analysis, and interpretation of digital images.
- Basic adjustments (like resizing) to advanced tasks (object detection).
- Aims to extract valuable data and improve visual quality.

Image after segmentation

Image after segmentation and morphological processing

Introduction

Image Quality Assessment (IQA)

- Quantitative representation of human perception quality.
- Utilizes a blend of human judgement and objective metrics.
- Ensures a balance between perception and technical accuracy.

(a) Reference Image | (b) Distorted Image 1 MOS: 0.431 | (c) Distorted Image 2 MOS: 0.521 | (d) Distorted Image 3 MOS: 0.702 | (e) Distorted Image 4 MOS: 0.449 | (f) Distorted Image 5 MOS: 0.320

Figure 1. Example reference and distorted Images from PIPAL Dataset. [17]

IQA Metrics for Algorithm Evaluation

- Used to assess algorithm performance in computer vision.
- Applicable in applications like image compression, transmission, and processing.
- Measures effectiveness in preserving visual quality.

Types of IQA

- Full-Reference IQA
- No-Reference IQA
- Reduced-Reference IQA

Distortions

- **Image Distortions** : Alterations impacting an image's visual quality and content, whether intentional or accidental.
- **Essential in IQA :** Distortions play a vital role in Image Quality Assessment (IQA) by assessing the effects of processes on image fidelity.
- **Process Impact :** Evaluation of distortions informs how transformations impact an image, guiding decisions for processing and enhancing image quality.

Types of Distortions

- **Traditional :** Gaussian blur, Motion blur, Image compression.
- **Super-Resolution :** Interpolation method, SR with kernel mismatch.
- **Denoising :** Mean filtering, Deep-learning-based methods.
- **Mixture Restoration :** SR of noisy images, SR after denoising.
- **GANs based :** Noise and Artifacts, Loss of Context, Visual Artifacts.

Gaussian blurring (a) DMOS = 59.0 | JPEG-2000 compression (b) DMOS = 67.1 | White noise (c) DMOS = 74.6 | JPEG-2000 compression (d) DMOS = 82.7

Metrics

Mean Opinion Score:

- Quantifies perceived image quality.
- Averages human observer ratings.

Advantages:

- Bridges technical and human perception.
- Enhances image processing methods.

Usage:

- Evaluates algorithms, compression, etc.
- Measures perceptual aspects.

(a) Reference Image (b) Distorted Image 1 MOS: 0.431 (c) Distorted Image 2 MOS: 0.521 (d) Distorted Image 3 MOS: 0.702 (e) Distorted Image 4 MOS: 0.449 (f) Distorted Image 5 MOS: 0.320

Figure 1. Example reference and distorted Images from PIPAL Dataset. [17]

Metrics

Peak Signal-to-Noise Ratio (PSNR):

- The ratio between the highest signal power (original image) and noise power (difference between original and distorted images).

$$
\begin{aligned}
PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\
&= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\
&= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE).
\end{aligned}
$$

Structural Similarity Index (SSIM):

- It assesses structural similarity between original and distorted images, considering luminance, contrast, and structure factors. It aims to align better with human perception compared to PSNR.

$$
\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}
$$

with:

- μ_x the average of x;
- μ_y the average of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;
- $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator;
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$);
- $k_1 = 0.01$ and $k_2 = 0.03$ by default.

Dataset Overview

Four benchmark image quality datasets are utilized in our experiments:

1. LIVE (2006)
2. TID2013 (2013)
3. KADID-10k (2019)
4. PIPAL (2020)

A tabulated summary of the datasets used for the performance comparison.

Database	Reference Images	Distorted Images	Distortion Types	Ratings	Rating Type	Distortion Type	Environment
LIVE [41]	29	779	5	25k	MOS	traditional	lab
TID2013 [35]	25	3000	25	524k	MOS	traditional	lab
KADID-10k [26]	81	10.1k	25	30.4k	MOS	traditional	crowdsourcing
PIPAL [17]	250	29k	40	1.13m	MOS	trad. + algo outputs	crowdsourcing

PIPAL
(Perceptual Image Processing ALgorithms IQA Dataset)

- PIPAL training set includes:
  - 200 reference images
  - 40 distortion types
  - 23,000 distortion images
  - Over one million human ratings
  - GAN-based algorithms outputs introduced as a new GAN-based distortion type
- The Elo rating system is used to assign Mean Opinion Scores (MOS) for the ratings.

Source : https://paperswithcode.com/dataset/pipal-perceptual-iqa-dataset

PIPAL
(Perceptual Image Processing ALgorithms IQA Dataset)

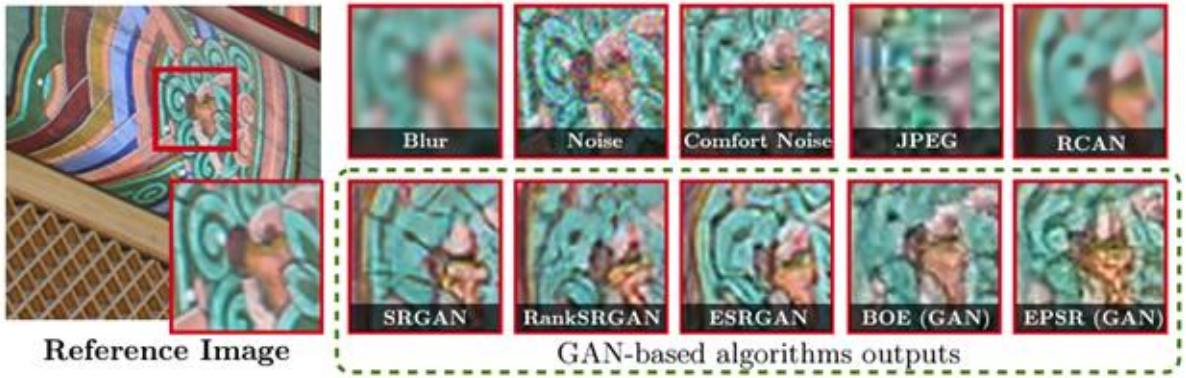


Fig. 1: Visualizing different distortions. Unlike the distortions in the upper row, which do not follow the natural image distribution. The GAN-based outputs are actually similar to natural images. However, their details are wrong

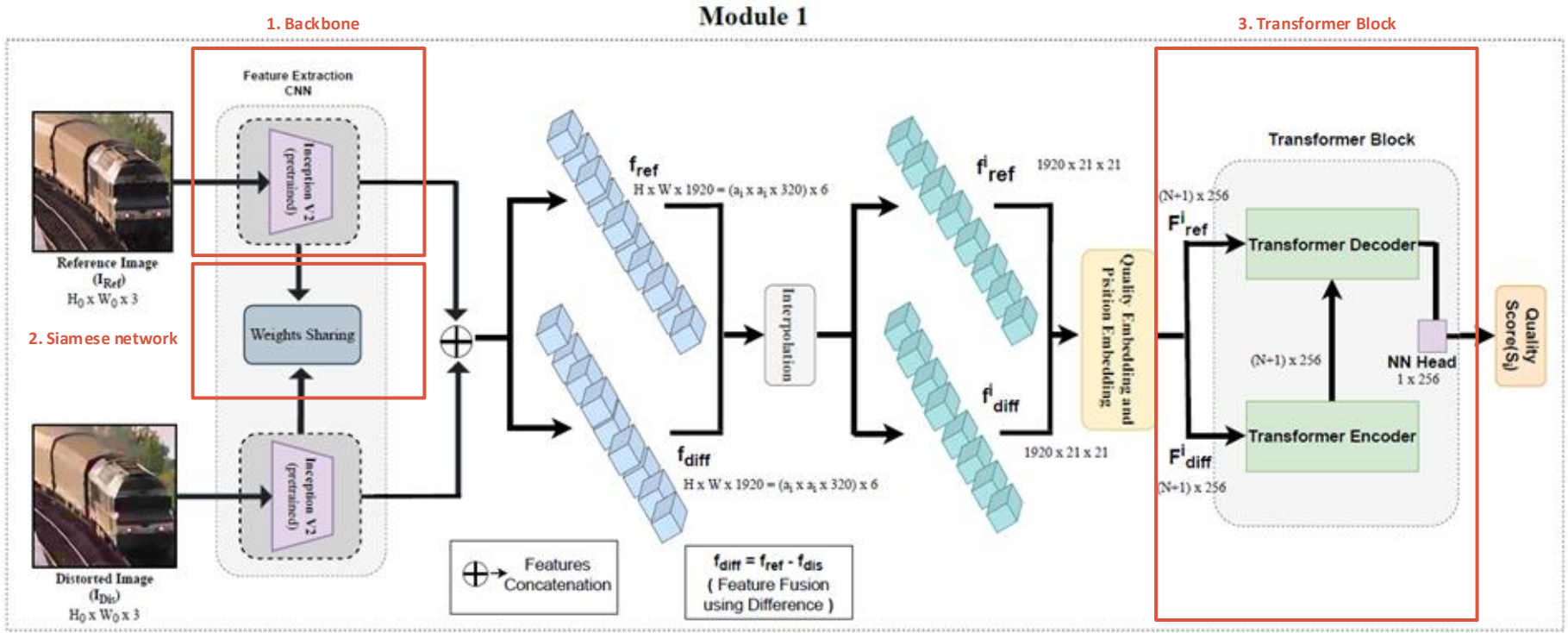
Source : https://arxiv.org/abs/2007.12142

Problem Statement

Perceptual Image Quality Assessment Challenge:

- **Objective** : Create a metric to predict the Mean opinion score (MOS).
- **Evaluation Metrics** : Predicted MOS value for the validation set would be compared with the true MOS value using:
  - Pearson linear correlation coefficient (PLCC)
  - Spearman rank-order correlation coefficients (SROCC)
- **Pre-training Allowance**: Pre-training with non-IQA datasets like ImageNet is permitted within the competition guidelines.
- **Dataset** : One must only use PIPAL dataset.
- **Disqualification Criteria**: Non-fully-referenced methods and models using extra labelled IQA datasets will be disqualified from final ranking.

Proposed Solution

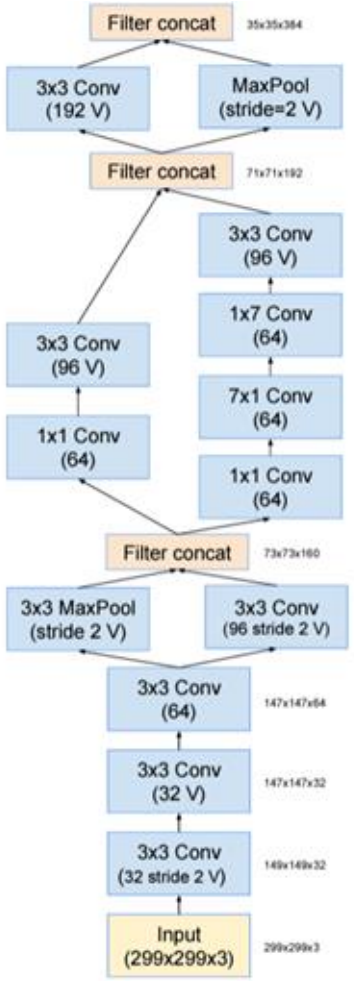


Source : https://arxiv.org/abs/2204.09779

Backbone

Inception-ResNet-v2

- This model is used as backbone because of its Top 1% accuracy in image classification.
- Following schema represents stem of the pure Inception-v4 and Inception-ResNet-v2 networks.
- Pretrained on ImageNet database.

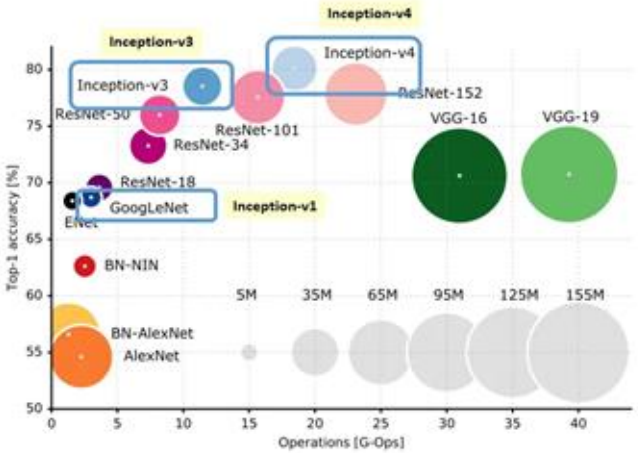


Source : https://arxiv.org/abs/1602.07261

Backbone

Inception-ResNet-v2 :

- Schema for Inception-ResNet-v1 and Inception ResNet-v2 networks.This schema applies to both networks but the underlying components differ.

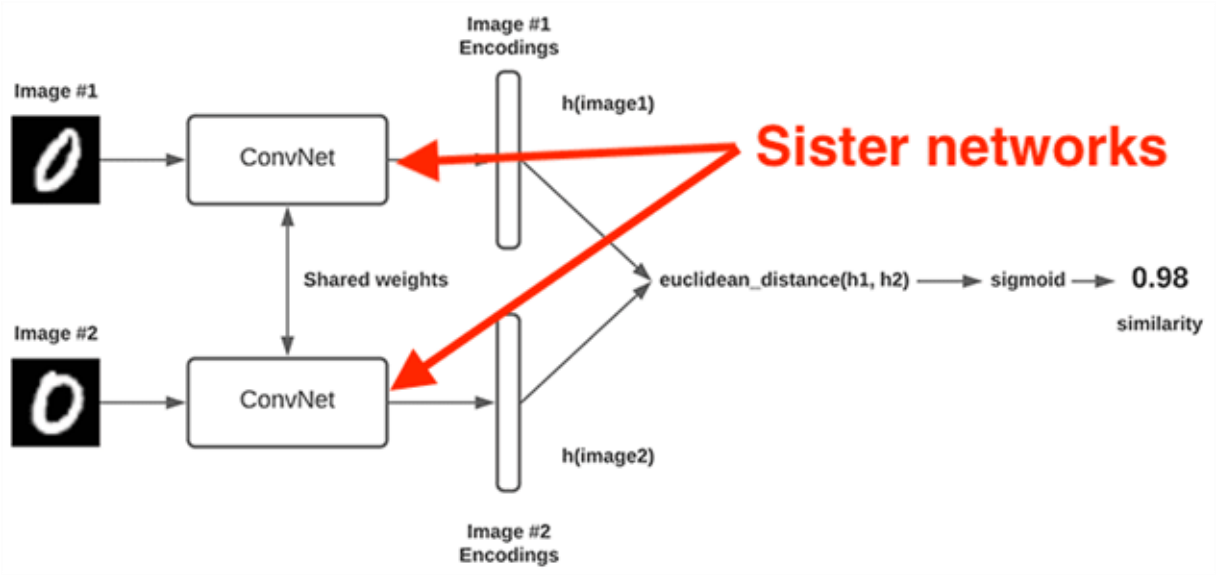


Top-1 Accuracy against Number of Operations (Size is the number of parameters)

Siamese Network

- **Twin CNN Structure :** Utilizes a combination of two shallow CNNs with a few hidden layers each, designed with flexibility in mind.
- **Weight and Bias Sharing :** Employs shared parameters between the CNNs, ensuring identical weights and biases for both networks. A single set of weights is trained and applied to both.
- **Loss Function Approach :** Implements either triplet or contrastive loss functions, contributing to effective learning of shared features within the twin CNN framework.

Siamese Network



Transformer Block

Attention is All You Need Transformer :

- **No Sequential Processing :** Dispenses with sequential processing, employing self-attention to establish global dependencies.
- **Parallel Computations :** Allows parallelization of computations, enhancing efficiency and scalability.
- **Enhanced Performance :** Revolutionizes tasks like language translation and image analysis, outperforming traditional sequential models.

Source : https://arxiv.org/abs/1706.03762

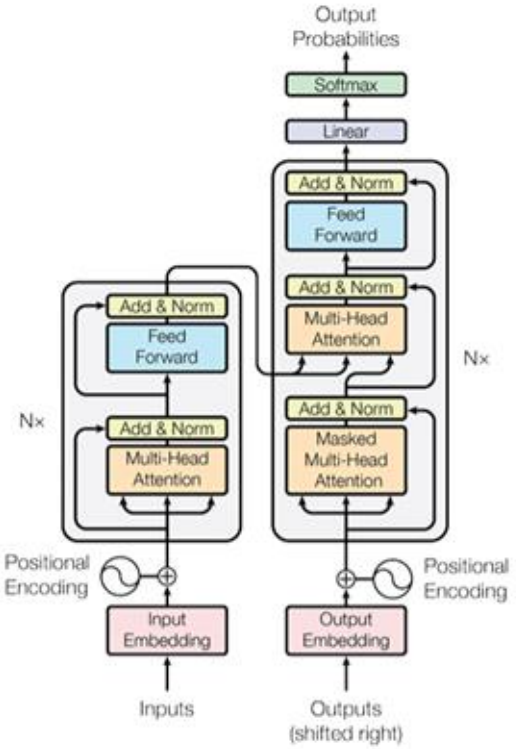
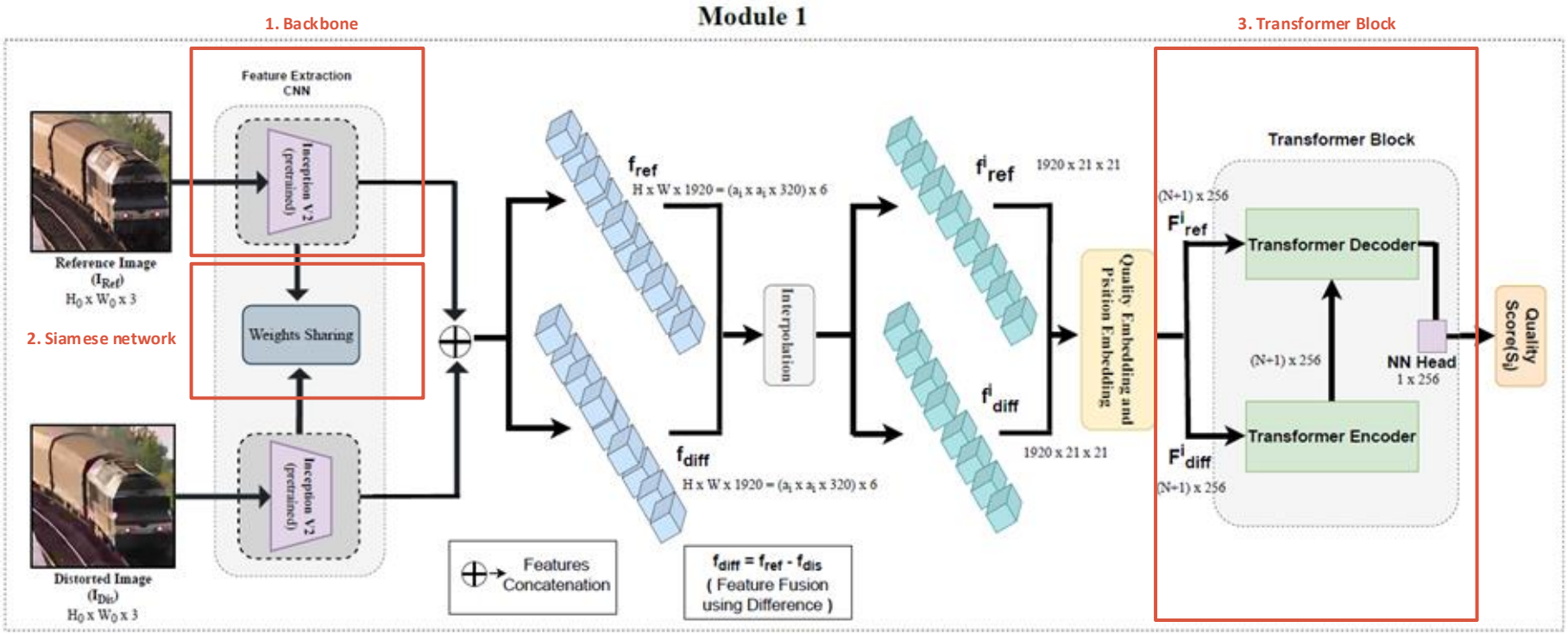


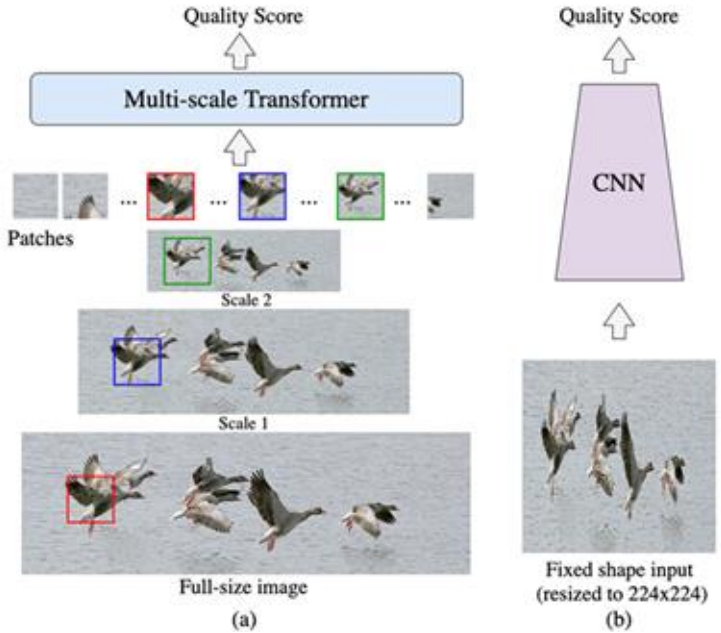
Figure 1: The Transformer - model architecture.

Workflow Diagram



Proposed Solution

- **Baseline MUSIQ :** In CNN-based models (b), images need to be resized or cropped to a fixed shape for batch training. However, such preprocessing can alter image aspect ratio and composition, thus impacting image quality.
- Patch-based MUSIQ model (a) can process the full-size image and extract multi-scale features, which aligns with the human visual system.



Source : https://arxiv.org/abs/2108.05997

Proposed Solution

Algorithm 1 MultiScale Transformer based IQA

Input: A pair of reference R_{img} and distored D_{img} image
Output: A predicted IQA score

Denotes feature extraction as FE,
enc_inp_emb ={x_{ij}, where i∈ {1 ... BatchSize}, j∈ {1 ... SequenceLength}, x_{ij}=1 },
dec_inp_emb ={x_{ij}, where i∈ {1 ... BatchSize}, j∈ {1 ... SequenceLength}, x_{ij}=1 },
Denotes Transformer block as TB

for $j \leftarrow 1$ **to** 4 **do**
  $f_{ref_j}, f_{diff_j} := FE(R_{img}, D_{img}, Scale=j)$
  $f^i_{ref_j} := Interpolate(f_{ref_j})$
  $f^i_{diff_j} := Interpolate(f_{diff_j})$
  $S_j := TB(f^i_{ref_j}, enc\_inp\_emb, f^i_{diff_j}, dec\_inp\_emb)$
end for
Final Score := $Avg(S_1,S_2,S_3,S_4)$

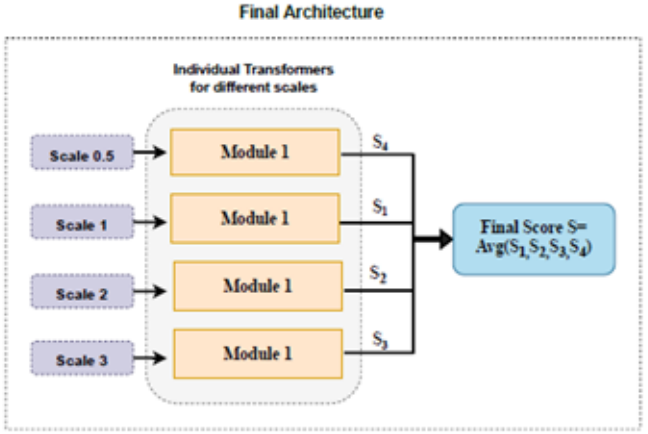


Figure 3. Workflow Diagram of the proposed overall model.

$$FinalQualityScore(S) = \frac{\sum_{i=1}^{4} S_i}{4}$$

Results

Table 2. Performence comparison over LIVE [41] and TID2013 [35] Datasets. [51]

Method	LIVE			TID2013		
	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC
PSNR	0.865	0.873	0.68	0.677	0.687	0.496
SSIM [45]	0.937	0.948	0.796	0.777	0.727	0.545
MS-SSIM [47]	0.94	0.951	0.805	0.83	0.786	0.605
VSI [53]	0.948	0.952	0.806	0.9	0.897	0.718
MAD [25]	0.968	0.967	0.842	0.827	0.781	0.604
VIF [40]	0.96	0.964	0.828	0.771	0.677	0.518
FSIMc [55]	0.961	0.965	0.836	0.877	0.851	0.667
NLPD [24]	0.932	0.937	0.778	0.839	0.8	0.625
GMSD [48]	0.957	0.96	0.827	0.855	0.804	0.634
WaDIQaM [6]		0.947	0.791	0.834	0.831	0.631
PieAPP [36]	0.908	0.919	0.75	0.859	0.876	0.683
LPIPS [56]	0.934	0.932	0.765	0.749	0.67	0.497
DISTS [12]	0.954	0.954	0.811	0.855	0.83	0.639
SWD [15]	-	-	-	-	0.819	0.634
IQT [9]	-	0.97	0.849	0.943	0.899	0.717
IQT-C [9]	-	0.917	0.737	-	0.804	0.607
MSFPT-1	0.962	0.976	0.874	**0.955**	**0.949**	**0.807**
MSFPT-2	0.958	0.964	0.846	0.872	0.857	0.673
MSFPT-3	0.944	0.955	0.824	0.853	0.828	0.635
MSFPT-0.5	0.963	0.976	**0.875**	0.831	0.796	0.598
MSFPT-avg	**0.972**	**0.977**	0.874	0.929	0.92	0.752

Table 3. Performence comparison over KADID Dataset. [26]

Method	KADID		
	PLCC	SRCC	KRCC
SSIM [45]	0.723	0.724	0.537
MS-SSIM [47]	0.801	0.802	0.609
IWSSIM [46]	0.846	0.850	0.666
MDSI [33]	0.873	0.872	0.682
VSI [53]	0.878	0.879	0.691
FSIM [55]	0.851	0.854	0.665
GMSD [48]	0.847	0.847	0.664
SFF [7]	0.862	0.862	0.675
SCQI [2]	0.853	0.854	0.662
ADD-GSIM [19]	0.817	0.818	0.621
SR-SIM [52]	0.834	0.839	0.652
MSFPT-1	0.822	0.846	0.653
MSFPT-2	0.796	0.799	0.613
MSFPT-3	0.667	0.674	0.495
MSFPT-0.5	0.857	0.857	0.672
MSFPT-avg	**0.888**	**0.883**	**0.700**

Results

Table 6. Performance comparison over Validation Dataset of NTIRE-2022 FR [18]

Model Name	Main Score	SRCC	PLCC
MSFPT-avg (our)	1.598	0.81	0.788
PSNR	0.503	0.234	0.269
NQM [10]	0.666	0.302	0.364
UQI [44]	0.966	0.461	0.505
SSIM [45]	0.696	0.319	0.377
MS-SSIM [47]	0.457	0.338	0.119
RFSIM [54]	0.539	0.254	0.285
GSM [29]	0.829	0.379	0.45
SRSIM [52]	1.155	0.529	0.626
FSIM [55]	1.005	0.452	0.553
VSI [53]	0.905	0.411	0.493
NIQE [32]	0.141	0.012	0.129
MA [30]	0.196	0.099	0.097
PI [4]	0.198	0.064	0.134
Brisque [31]	0.06	0.008	0.052
LPIPS-Alex [56]	1.175	0.569	0.606
LPIPS-VGG [56]	1.162	0.551	0.611
DISTS [12]	1.243	0.608	0.634

Table 7. Performance comparison over Testing Dataset of NTIRE-2022 FR [18]

Model Name	Main Score	SRCC	PLCC
MSFPT-avg (our)	1.45	0.738	0.713
PSNR	0.526	0.249	0.277
NQM [10]	0.76	0.364	0.395
UQI [44]	0.87	0.42	0.45
SSIM [45]	0.753	0.361	0.391
MS-SSIM [47]	0.532	0.369	0.163
RFSIM [54]	0.632	0.304	0.328
GSM [29]	0.874	0.409	0.465
SRSIM [52]	1.209	0.573	0.636
FSIM [55]	1.075	0.504	0.571
VSI [53]	0.975	0.458	0.517
NIQE [32]	0.166	0.034	0.132
MA [30]	0.287	0.14	0.147
PI [4]	0.249	0.104	0.145
Brisque [31]	0.14	0.071	0.069
LPIPS-Alex [56]	1.137	0.566	0.571
LPIPS-VGG [56]	1.228	0.595	0.633
DISTS [12]	1.342	0.655	0.687

Ablation Study

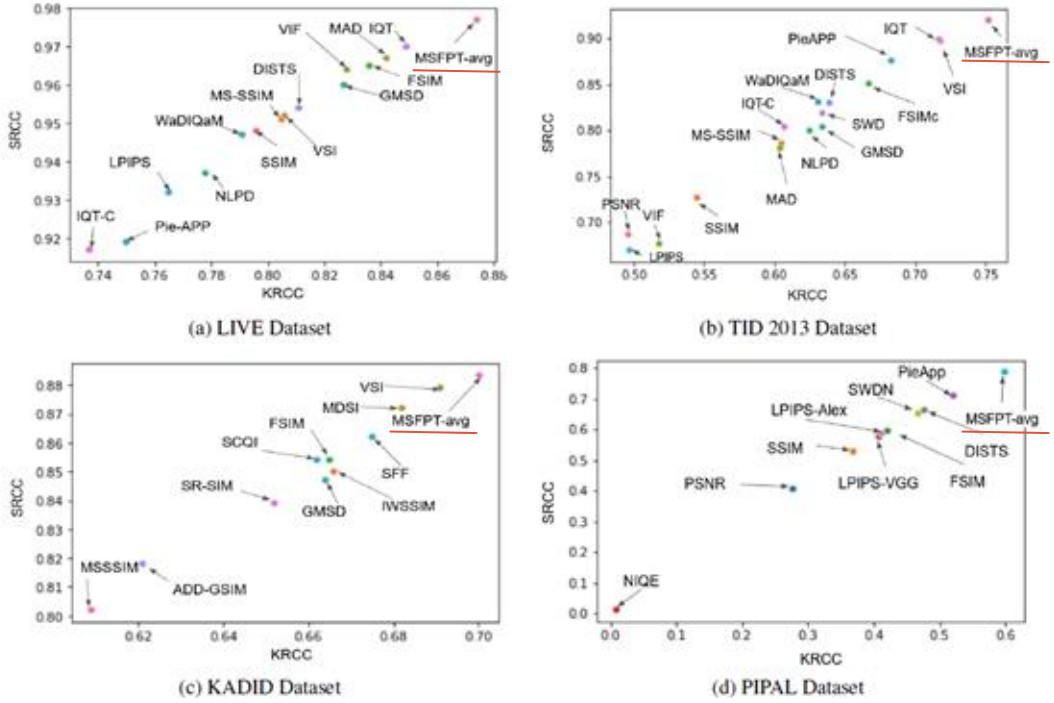
Table 4. Ablation study with respect to the different scales.

Model Name	Validation		
	Main Score	PLCC	SRCC
MSFPT-1	1.552	0.784	0.768
MSFPT-2	1.522	0.773	0.749
MSFPT-3	1.47	0.749	0.721
MSFPT-0.5	-	-	-
MSFPT-avg	1.598	0.810	0.788
Model Name	**Testing**		
	Main Score	**PLCC**	**SRCC**
MSFPT-avg	1.450	0.738	0.713
MSFPT-1	1.254	0.637	0.617
MSFPT + Bert + Scale1	1.383	0.699	0.684
MSFPT + Bert + Scale2	1.361	0.698	0.663
MSFPT + Bert + Scale3	1.182	0.621	0.561
MSFPT + Bert + Avg. of 1,2,3	1.44	0.73	0.71

Table 5. Performance comparison of the proposed algorithm in NTIRE IQA challenge, Testing phase.

Team Name	Main score	PLCC	SRCC
Anynomus1	1.651	0.826	0.822
Anynomus2	1.642	0.827	0.815
Anynomus3	1.64	0.823	0.817
Anynomus4	1.541	0.775	0.766
Anynomus5	1.538	0.772	0.765
Anynomus6	1.501	0.763	0.737
Pico Zen(ours)	1.45	0.738	0.713
Anynomus8	1.403	0.703	0.701

Quantitative Comparison



(a) LIVE Dataset (b) TID 2013 Dataset

(c) KADID Dataset (d) PIPAL Dataset

Figure 4. Quantitative comparison of IQA methods. (a) LIVE Dataset, (b)TID 2013 Dataset, (c)KADID Dataset, (d) PIPAL Dataset.

Conclusion

- **Algorithm :** Presented a full-reference image quality assessment algorithm integrating parallel transformers and multi-scale CNN features.
- **Transformer Network :** Utilized encoders and decoders within transformers for quality prediction, enhancing accuracy.
- **Experimental Validation :** Conducted comprehensive experiments to showcase the effectiveness of the parallel transformers and multi-scale features combination.

Conclusion

- **Performance :** Demonstrated the algorithm's superiority over alternative network combinations, highlighting its enhanced performance.
- **Outperforming State-of-the-Art :** Evaluated against current image quality assessment methods, the proposed approach outperforms in terms of assessment accuracy.

Questions?

Thank You