Multi-dimension Attention Network for No-Reference Image Quality Assessment (MANIQA)

It aims to evaluate the quality of images without needing a reference image.

This is particularly challenging for images distorted by Generative Adversarial Networks (GANs), as these distortions can create textures that look realistic to humans but don't align with the prior knowledge of deep-learning models used for Image Quality Assessment (IQA).

Key Points of MANIQA

1. Challenge with Existing NR-IQA Methods:

- Traditional No-Reference Image Quality Assessment (NR-IQA) methods struggle to accurately predict the quality of images that have GAN-based distortions.

- GAN-based distortions can produce unreal textures that are visually pleasing but deviate from what existing models recognize as high quality.

2. Proposed Solution: MANIQA:

- MANIQA utilizes the Vision Transformer (ViT) to extract features from images.

- **Transposed Attention Block (TAB):** Enhances global interactions within the image, helping the model understand the overall structure and content.

- **Scale Swin Transformer Block (SSTB)**: Focuses on local interactions, improving the model's ability to assess fine details.

- **Dual Branch Structure:** Used for patch-weighted quality prediction, allowing the model to evaluate different parts of the image with varying importance.

3. Performance and Validation:

- MANIQA outperforms state-of-the-art methods on standard datasets.

- It ranked first in the NTIRE 2022 Perceptual Image Quality Assessment Challenge Track

- The method shows impressive results on various datasets, including PIPAL, LIVE, CSIQ, TID2013, and KADID-10K.

4. Addressing GAN-based Distortion:

- MANIQA is particularly effective in handling GAN-based distortions, overcoming the limitations of existing NR-IQA methods.

- The architecture of MANIQA includes specific components designed to enhance both global and local feature interactions, making it better suited to evaluate the complex distortions introduced by GANs.

5. Architecture of MANIQA:

- **Feature Extractor:** Uses Vision Transformer (ViT) to extract high-quality features from the image.

- **Transposed Attention Block (TAB):** Enhances global interactions within the image, capturing the overall content and structure.

- **Scale Swin Transformer Block (SSTB)**: Focuses on local interactions, capturing fine details and textures.

- **Dual Branch Structure**: Utilizes patch-weighted quality prediction to evaluate different parts of the image with appropriate weighting.

6. Experimental Validation:

- Experiments on multiple datasets demonstrate that MANIQA consistently outperforms existing methods.

- The ablation study highlights the importance of each component (ViT, TAB, SSTB, dual branch structure) in improving the model's performance.

Conclusion

MANIQA provides a comprehensive and effective solution for No-Reference Image Quality Assessment, especially for images with GAN-based distortions.

Its advanced architecture and the integration of multiple attention mechanisms enable it to deliver superior performance across various datasets, outperforming current state-of-the-art methods.