Life-IQA: Boosting Blind Image Quality Assessment through GCN-enhanced Layer Interaction and MoE-based Feature Decoupling

Supplementary Material

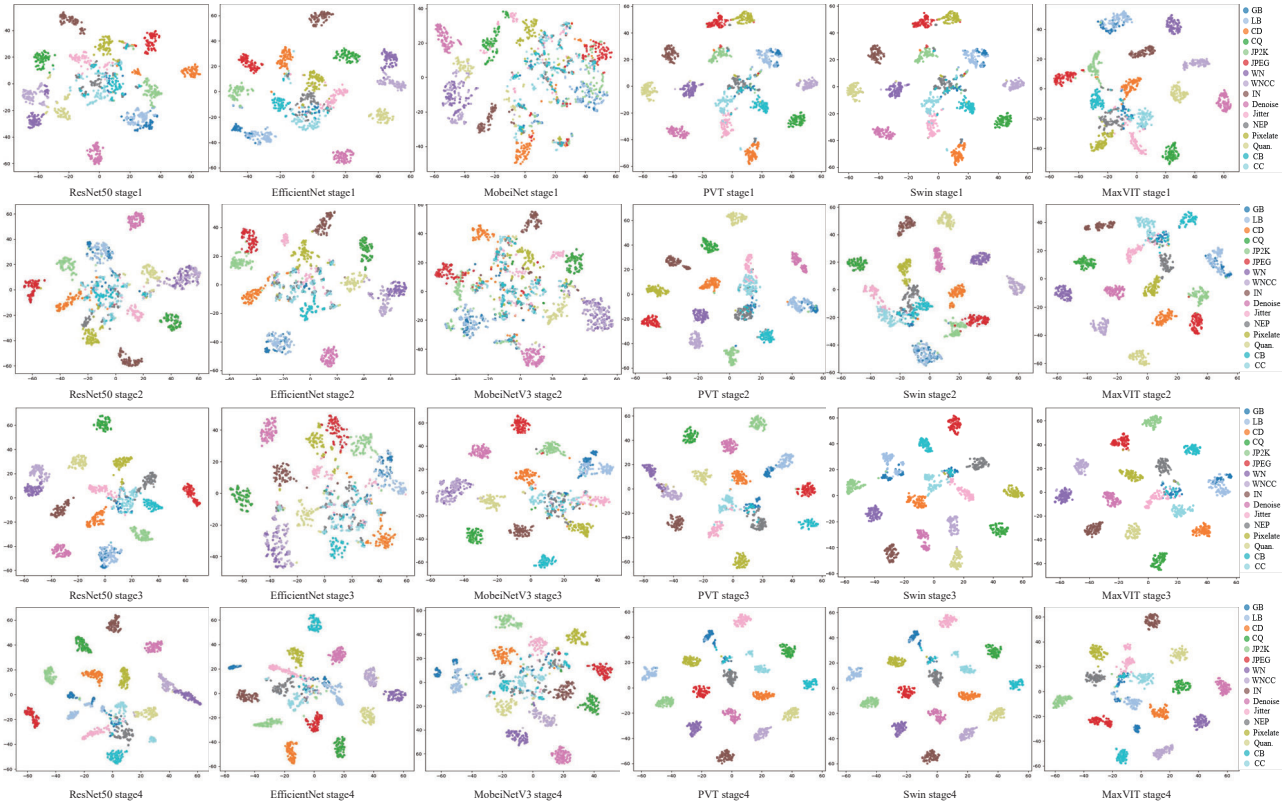

Figure 10. t-SNE visualization of multi-scale features from different pretrained models on the KADID-10k dataset.

Method	CSIQ		KADID-10K	
	SROCC	PLCC	SROCC	PLCC
Stage2⟺Stage1	0.951	0.959	0.915	0.918
Stage3⟺Stage1	0.960	0.965	0.937	0.940
Stage3⟺Stage2	0.962	0.968	0.938	0.938
Stage4⟺Stage1	0.958	0.965	0.930	0.930
Stage4⟺Stage2	0.964	0.968	0.937	0.939
Life-IQA	**0.966**	**0.971**	**0.940**	**0.943**

Table 6. Cross-layer pairing study: deep and shallow interactions for BIQA

6. Cross-layer pairing study

Table 6 varies only the interaction pair in the GCN-enhanced layer interaction, while the decoder depth, heads, embedding size, query count, training recipe, and parameter budget are kept fixed. A consistent pattern emerges:

- **Deep with deep performs best.** Using Stage4 with Stage3 (Life-IQA) achieves 0.966/0.971 on CSIQ and 0.940/0.943 on KADID-10K.
- **Deep with shallow is weaker.** Relative to Stage4 with Stage1, the deep–deep pairing gains +0.008/+0.006 SROCC/PLCC on CSIQ and +0.010/+0.013 on KADID-10K.
- **Deep with mid sits in between.** Compared with Stage4 with Stage2, the deep–deep pairing improves by +0.002/+0.003 on CSIQ and +0.003/+0.004 on KADID-10K; compared with Stage3 with Stage2, the gains are +0.004/+0.003 on CSIQ and +0.002/+0.005 on KADID-10K.
- **Shallow with shallow is the weakest.** The Stage2⇔Stage1 pairing yields 0.951/0.959 on CSIQ and 0.915/0.918 on KADID-10K, underperforming

the deep–deep setting by -0.015/-0.012 (CSIQ) and -0.025/-0.025 (KADID-10K).

The trend holds for both rank correlation (SROCC) and linear correlation (PLCC), indicating that interactions between semantically compatible deep representations yield more reliable quality cues, whereas shallow–shallow coupling provides limited semantic separation and thus the weakest performance.

7. t-SNE analysis of stage-wise features

Figure 10 visualizes features from four stages of several pretrained backbones on KADID-10K (25 distortion types). A clear monotonic trend is observed across all architectures: from stage1 to stage4, clusters become progressively more compact with larger inter-class margins, while cross-class overlap is reduced. Early layers exhibit scattered manifolds with substantial mixing among distortion categories, indicating dominance of low-level textures and color edges. Mid-level layers begin to separate families of artifacts (e.g., blur, noise, compression), yet boundaries remain porous. Deep layers form well-delimited groups with sparse interstitial points, suggesting that higher-level representations encode distortion-aware semantics that are more linearly separable. The pattern holds for both convolutional (ResNet50, EfficientNet, MobileNetV3) and Transformer-style models (PVT, Swin, MaxViT), with the latter generally showing clearer margins at deeper stages. These observations support using deep features for query formation and shallower deep features as keys/values in the proposed interaction module, as deep–deep pairing exploits the stronger class separability of later stages while retaining complementary detail from preceding stages.