Task-Adaptive FFN Editing for Continual Blind Image Quality Assessment

Satish Maurya* and Parimala Kancharla†
School of Computing and Electrical Engineering
Indian Institute of Technology Mandi, Himachal Pradesh 175005, India
Email: *s24012@students.iitmandi.ac.in, †parimala@iitmandi.ac.in

***Abstract*—Blind Image Quality Assessment (BIQA) models trained on one distortion distribution often degrade when exposed to new ones, making sequential adaptation without forgetting a fundamental challenge. While continual learning offers a natural solution, existing methods typically retrain the entire backbone per task, limiting scalability and parameter efficiency. We propose ContEditIQA, a parameter-efficient framework for continual BIQA that selectively edits a pre-trained Vision Transformer (ViT) rather than retraining it. Following a locate-then-edit strategy, a lightweight attention-guided hypernetwork identifies distortion-sensitive Feed-Forward Network (FFN) parameters for each incoming task and restricts updates to those regions, while attention layers remain frozen to preserve globally shared representations. This targeted editing enables robust sequential adaptation without model expansion or memory replay. Experiments across six BIQA benchmarks demonstrate superior knowledge retention and cross-dataset generalization while modifying fewer than 30% of backbone parameters, establishing selective model editing as an effective and scalable paradigm for continual BIQA.**

***Index Terms*—Blind Image Quality Assessment, Continual Learning, Model Editing**

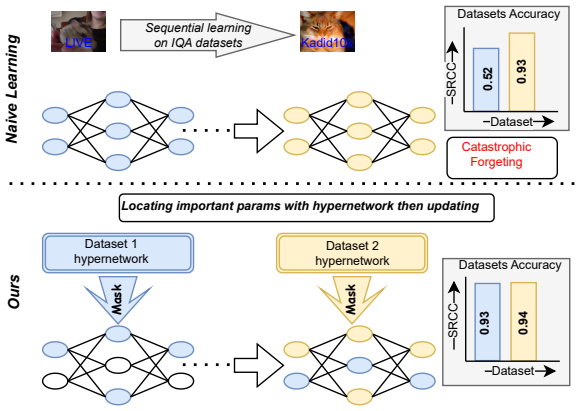

Fig. 1: Naive sequential learning vs. the proposed mask-based continual learning approach for BIQA. Color-coded masks and parameters indicate task-specific updates guided by the hypernetwork, mitigating catastrophic forgetting.

I. Introduction

Blind Image Quality Assessment (BIQA) [1], [2], [3] predicts perceptual quality scores for a given distorted image without needing its reference pristine image. This is essential for streaming services and social media where "perfect" ground truths do not exist. While some datasets like LIVE [4] and CSIQ [5] focused on synthetic distortions, some benchmarks such as BID [6], CLIVE [7], KonIQ-10K [8], and KADID-10K [9] capture complex real-natural degradations or artifacts. Despite these advancements, modern architectures [10], [11], [12] often fail to generalize; they learn dataset-specific characteristics rather than universal quality principles, causing models trained on synthetic noise to fail when applied to real camera artifacts. If we directly fine-tune the model on new distortion, this may result in *catastrophic forgetting*[13],[14] of previously seen distortions.

Continual Learning for BIQA: Previous works[15], [16] address this by balancing plasticity (learning new information) with stability (retaining old knowledge). While traditional BIQA research focuses on maximizing correlation within a single dataset, continual BIQA requires sustained performance across an evolving stream of tasks. [17], [18], [19] formulated continual learning for BIQA and introduced plasticity and stability indices to quantify this trade-off. Early methods like Learning without Forgetting (LwF) [20] used knowledge distillation to prevent drift, and recent research integrated distillation with task-specific heads [17]. [21] improved adaptation by freezing convolution filters and learning only task-specific normalization layers, while [22] introduced channel modulation kernels to encode channel attention and mitigate model divergence across successive tasks through attention distillation. However, these CNN-based methods lack precise adaptation mechanisms, and the global reasoning power of Vision Transformers (ViT) has yet to be fully exploited in this context.

Model editing offers a surgical alternative to full retraining by updating only behavior-specific network components, avoiding the instability of global updates. While widely explored in language models [23], [24], [25], its application in computer vision remains limited [26]. In BIQA, different parameters encode sensitivities to distinct distortions, enabling targeted adaptation while preserving orthogonal knowledge. We propose ContEditIQA, a mask-guided selective editing framework for continual BIQA. Focusing on the FFN layers of a ViT backbone, an attention-gated hypernetwork generates task-specific binary masks to identify which weights require updating, while self-attention modules remain fixed to preserve

global reasoning.

Our framework follows a two-stage *locate–then–edit* strategy. A lightweight hypernetwork first localizes distortion-sensitive FFN parameters for each incoming dataset, after which only the selected parameters are updated to enable targeted adaptation while preserving prior knowledge. This approach achieves fine-grained plasticity without replay buffers or full-network fine-tuning and consistently outperforms existing continual BIQA methods across six benchmarks. Fig 1 compares naive sequential learning with hypernetwork-guided selective editing for mitigating catastrophic forgetting. Our contributions are summarized as follows:

- We propose a selective transformer editing framework for continual BIQA that adapts only task-relevant FFN layers while preserving global attention representations.
- We introduce a hypernetwork-driven *locate–then–edit* mechanism for dynamic parameter localization and editing.
- We achieve a favorable stability–plasticity trade-off while updating only $\leq$29% of backbone parameters.

II. PROPOSED METHOD

A. Continual Learning of Dataset-Specific Heads for BIQA

Continual Blind Image Quality Assessment (BIQA) requires models to adapt to newly emerging distortion distributions without degrading performance on previously learned datasets. A straightforward approach is to sequentially update both the backbone and prediction heads; however, this often induces global parameter drift, leading to overfitting and catastrophic forgetting.

We consider a continual learning setting with a sequence of datasets $\{D_1, D_2, \ldots, D_T\}$, each characterized by distinct distortion types. Our BIQA architecture consists of a backbone feature extractor $f_\phi : \mathbb{R}^{H\times W} \to \mathbb{R}^L$ and task-specific prediction heads $\mathbf{h}_{\psi_t}(\cdot)$. The objective is to sequentially update the shared backbone and dataset-specific heads while preserving previously acquired knowledge.

Rank-Based Loss: We adopt a fidelity-aware ranking loss [27], which better reflects perceptual ordering than regression-based objectives. For an image pair $(x, y) \sim D_t$, the probability that x has higher perceptual quality than y is defined as

$$\hat{p}_t(x, y) = \Phi\left(\frac{h_{\psi_t}(f_\phi(x)) - h_{\psi_t}(f_\phi(y))}{\sqrt{2}}\right)$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Given MOS values $q(x), q(y) \in \mathbb{R}$, the binary label is defined as $p(x, y) = 1$ if $q(x) \geq q(y)$ and 0 otherwise. The fidelity loss is formulated as

$$\ell(x, y; \phi, \psi_t) = 1 - \sqrt{p(x, y)\hat{p}_t(x, y)} - \sqrt{(1 - p(x, y))(1 - \hat{p}_t(x, y))}. \quad (1)$$

The overall training objective is given by

$$\min_{\phi_t, \{\psi_t\}} \ell(x, y; \phi_t, \psi_t) + \lambda \sum_{k=1}^{t-1} \ell(x, y; \phi_{t-1}, \psi_k), \quad (2)$$

where ϕ denotes the backbone parameters and ψ_t corresponds to the head for dataset t. For previous tasks, ground-truth labels are unavailable; therefore, pseudo-labels generated by the previous model are used, denoted as $p_k(x, y)$ for $k = 1, \ldots, t-1$ and $p(x, y)$ in (1) becomes $p_k(x, y)$ and used in the second term of (1) . This formulation enables sequential optimization across datasets while allowing prediction heads to evolve with incoming data. The second term penalizes the model from drifting from previous knowledge, where λ controls the penalty strength.

However, fully fine-tuning the backbone at each task induces global parameter drift and leads to catastrophic forgetting. To mitigate this issue, we introduce a selective backbone editing strategy that updates only a subset of parameters while freezing distortion-invariant representations. This selective editing mechanism enables efficient task adaptation while preserving prior knowledge, forming the foundation of the proposed *ContEditIQA* framework.

B. Selective Transformer Editing via Hypernetwork Localization

Our approach builds on the Vision Transformer (ViT) backbone, but departs from conventional fine-tuning by introducing a selective editing mechanism for continual adaptation. Instead of updating the entire network, we identify specific parameters that are most responsive to particular distortions.

A Transformer block consists of Multi-Head Self-Attention (MHSA) and Feed-Forward Network (FFN) modules. While MHSA primarily captures global structural relationships that remain largely distortion-invariant, FFN layers encode task-dependent contextual representations[28]. This observation motivates a targeted adaptation strategy: preserving attention pathways while selectively modifying FFN parameters to accommodate new datasets.

The FFN sublayer is defined as

$$\text{FFN}(x) = \text{Linear}_2\left(\text{GELU}\left(\text{Linear}_1(x)\right)\right). \quad (3)$$

we keep the other components of backbone frozen while adapting only the FFN components, as illustrated in Fig. 2. This design constrains representational drift while enabling task-specific plasticity.

For finding distortion-aware parameters, we used an attention-guided hypernetwork that generates task-specific masks for editing FFN weights. Rather than updating the whole FFN, the hypernetwork learns which parameters are critical for the incoming task and applies multiplicative modulation to those weights. The backbone f_{ϕ_t} remains fixed except for the masked FFN parameters, allowing controlled adaptation without overwriting prior knowledge. This selective transformer editing strategy forms the core of the proposed *ContEditIQA* framework, enabling continual learning through localized parameter updates while maintaining stable global representations.

C. Hypernetwork-Guided Selective Editing

Let W_1 and W_2 denote the linear projection matrices of the FFN within a Transformer block. Given an image

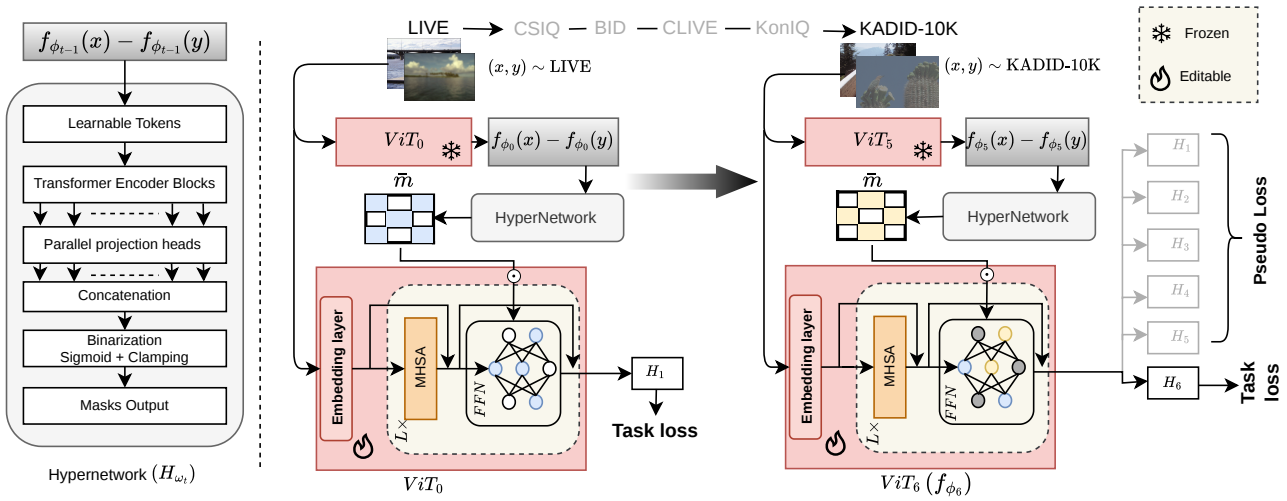

Fig. 2: System diagram of the proposed ContEditIQA framework for continual learning. The framework processes datasets in chronological order as shown above. At each task t, the frozen backbone ViT_{t-1} from the previous task(for the first task (t=1), ViT_0 is initialized with ImageNet-pretrained weights) generates features for two given images x, y then the difference $z_t = f_{\phi_{t-1}}(x) - f_{\phi_{t-1}}(y)$ is fed to the hypernetwork to produce binary masks $\tilde{m}$ for selective parameter localization.The trainable ViT_t applies masked editing to FFN layers via elementwise multiplication. Task-specific heads h_{ψ_1} to h_{ψ_t} compute the combined objective: task loss $\ell(x, y; \phi_{\text{edited}}, \psi_t)$ for the current dataset and pseudo loss $\sum_{k=1}^{t-1} \ell(x, y; \phi_{\text{edited}}, \psi_k)$ from previous heads to prevent catastrophic forgetting. Then, the same hypernetwork is learned via (1) for the next dataset.

Algorithm 1 Bi-level Hypernetwork Optimization

Require: $d_t \in \mathcal{D}_t$, $f_{\phi_{t-1}}$ (Frozen), γ, β (Learning Rates)
Ensure: ω_t (Hypernetwork params), ψ_t (Task Head)

Initialize: ω_t, ψ_t randomly
for Batch $(x, y) \in \mathcal{D}_t$ **do**
  $z_t \leftarrow f_{\phi_{t-1}}(x) - f_{\phi_{t-1}}(y)$
  Randomly initialize $\tilde{m}$
  for $i = 1$ to $Outer$ **do**
    for $i = 1$ to $Inner$ **do**
      $\mathcal{L}_{\text{task}} \leftarrow \ell(x, y; \phi \odot \tilde{m}, \psi_t)$
      $\tilde{m}^* \leftarrow \tilde{m} - \gamma \nabla_{\tilde{m}} \mathcal{L}_{\text{task}}$
    end for
    $\hat{m} \leftarrow H_{\omega_t}(z_t)$ ▷ Hypernet Prediction
    $\mathcal{L}_{\text{align}} \leftarrow \text{BCE}(\hat{m}, \tilde{m}^*)$
    $\omega_t \leftarrow \omega_t - \beta \nabla_{\omega_t} \mathcal{L}_{\text{align}}$ ▷ Update Hypernet
  end for
end for

pair $(x, y) \sim D_t$, we compute the feature difference $z_t = f_{\phi_{t-1}}(x) - f_{\phi_{t-1}}(y)$, which captures distortion-aware variations under the previously learned backbone.

A lightweight hypernetwork H_{ω_t} predicts sparse binary masks indicating which FFN parameters require updating:

$$M_1^t, M_2^t = H_{\omega_t}(z_t), \tag{4}$$

where the edited FFN weights are obtained via element-wise modulation: $\tilde{W}_1^t = W_1 \odot M_1^t$ and $\tilde{W}_2^t = W_2 \odot M_2^t$. The masked

TABLE I: Performance Comparison (mSRCC, mPI, mSI, mPSI)

Method	mSRCC	mPI	mSI	mPSI
LwF-KG[17]	0.815	0.8563	0.9796	0.9180
TSN-IQA[21]	0.84	0.860	0.985	0.923
ContEditIQA-KG	**0.8187**	**0.8382**	**0.9553**	**0.8968**
ContEditIQA-O	**0.8579**	**0.8662**	**0.9859**	**0.9261**

TABLE II: Performance comparison of previous SOTAs after trained on the last dataset of the sequence (KADID-10K).

Method	Test Dataset (SRCC)					
	LIVE	CSIQ	BID	CLIVE	KonIQ	KADID
LwF-KG [20], [17]	0.8996	0.7983	0.8128	0.7742	0.8520	0.7622
TSN-IQA[21]	0.954	0.801	0.829	0.786	0.870	0.830
MH-CMKernel-KG[22]	0.9327	0.8905	0.7437	0.8134	0.7263	0.8243
CCL-IQA[19]	0.9112	-	-	0.8021	0.8653	0.8092
ContEditIQA-KG	**0.9476**	**0.8775**	**0.7697**	**0.7511**	**0.7453**	**0.8214**
ContEditIQA-O	**0.9493**	**0.9087**	**0.7928**	**0.8168**	**0.8214**	**0.8794**

FFN is then defined as

$$\text{FFN}_t(x) = \tilde{W}_2^t \cdot \text{GELU}(\tilde{W}_1^t \cdot x). \tag{5}$$

Once masks are learned via bi-level optimization (Algorithm 1), they remain fixed during backbone adaptation. Given a training batch B_t from dataset D_t, we jointly optimize the

masked backbone parameters ϕ_{edited} and task-specific head ψ_t:

$$\min_{\phi_{\text{edited}},\psi_t} \mathbb{E}_{(x,y)\sim B_t}\left[\ell(x,y;\phi_{\text{edited}},\psi_t)+\sum_{k=1}^{t-1}\ell(x,y;\phi_{\text{edited}},\psi_k)\right], \tag{6}$$

where pseudo-label losses from previous heads enforce functional consistency while adapting to the current task. This locate–then–edit procedure repeats sequentially across datasets, as illustrated in Fig. 2.

D. Inference Strategy

We evaluate the proposed framework under two complementary inference protocols that reflect both practical deployment and ideal task-aware settings.K-means Gating (KG) [17] enables task-agnostic inference by computing a weighted aggregation of predictions from all learned heads. The coefficients a_t are obtained using standard gating procedures to estimate task affinity at test time.Task-Oracle inference assumes knowledge of the test task and directly selects the corresponding head:

$$\hat{q}_{\text{KG}}(x) = \sum_{t=1}^{T} a_t \, h_{\psi_t^*}\big(f_{\phi^{\text{Edited}*}}(x)\big)$$

$$\hat{q}_{\text{Oracle}}(x) = h_{\psi_{\hat{t}}^*}\big(f_{\phi^{\text{Edited}*}}(x)\big)$$

Together, these protocols provide complementary perspectives on real-world usability and upper-bound performance, highlighting the robustness of the proposed continual editing framework across both task-agnostic and task-aware scenarios.

III. Experiments and Results

A. Experimental Setup and Performance Evaluation

We evaluate the proposed method on six widely used BIQA benchmarks: LIVE [4], CSIQ [5], BID [6], LIVE Challenge (CLIVE) [7], KonIQ-10K [8], and KADID-10K [9]. Following [17], the datasets are arranged chronologically to simulate a continual learning setting, with a 70%/10%/20% train/validation/test split. Content independence is ensured for LIVE, CSIQ, and KADID-10K by splitting based on reference images, and 500 samples per dataset are reserved for hypernetwork training and excluded from the main splits.

All experiments are conducted using the Adam optimizer with learning rates of 2×10^{-4} for head training(rehearsal) and 2×10^{-5} during editing. Models are trained for 6 epochs with a batch size of 32 using random 224×224 crops. Results are reported after training on last dataset.

We compare against regularization-based methods EWC [29], SI [30], and MAS [31]; distillation-based LwF [20]; and architectural variants MH-CL and SH-CL. All methods use the same ViT-B backbone [32] and training protocol. Performance is evaluated using SRCC, mSRCC, mean Plasticity Index (mPI), mean Stability Index (mSI), and mean Plasticity–Stability Index (mPSI) [17].

Comparison with Joint Learning: Interestingly, ContEditIQA-O outperforms joint learning (JL) across all datasets (Table II) despite using only 29% of trainable parameters. While joint learning trains on all datasets simultaneously, which typically provides an upper bound on performance, selective FFN editing achieves better generalization. This can be explained by the localized parameter updates reducing cross-task interference: by constraining modifications to distortion-sensitive regions, the model maintains cleaner decision boundaries for each distortion type, whereas unconstrained joint optimization may lead to parameter configurations that compromise across different distortion families.

Parameter Efficiency and Scalability: The proposed method edits only ≤ 25.36M parameters (29% of the 86.57M ViT backbone), achieving superior stability-plasticity trade-off compared to full-parameter methods. The attention-based hypernetwork requires minimal overhead, using only 500 samples per task for learning mask generation. This efficiency is particularly valuable for continual learning scenarios where computational resources and storage are constrained.

Result Analysis: Careful analysis reveals distinct performance characteristics of ViT: it demonstrates robust performance on most datasets while showing marginally lower correlation on KonIQ-10K. CNN methods [17], [21], [22] perform better on KonIQ-10K but lag substantially on KADID-10K. This suggests that ViT's global self-attention may be less effective at capturing localized authentic artifacts while excelling at systematic synthetic distortions. Our use of vanilla ViT-B [32] ensures fair baseline comparison, indicating potential for advanced ViT variants with enhanced local feature modeling.

TABLE III: Ablation study on editing strategies for continual BIQA.

Method	Test Dataset (SRCC)					
	LIVE	CSIQ	BID	CLIVE	KonIQ	KADID
Only-FFN-LwF	0.7228	0.6856	0.6718	0.6704	0.7122	0.7627
Rand-Mask-edit	0.8468	0.7017	0.6014	0.7164	0.7287	0.6908
6-layer-edit	0.9470	0.8787	0.8089	0.8117	0.8029	0.8472
Ours	**0.9493**	**0.9087**	**0.7928**	**0.8168**	**0.8214**	**0.8794**

TABLE IV: Performance comparison after training VIT with different continual methods on the last dataset of the sequence (KADID-10K).

Method	Params (M)	Test Dataset (SRCC)					
		LIVE	CSIQ	BID	CLIVE	KonIQ	KADID
JL	86.57	0.9131	0.7428	0.6799	0.6960	0.7447	0.7922
SH-CL	86.57	0.8577	0.7296	0.7062	0.4548	0.6666	0.8394
MH-CL-O	86.57	0.8618	0.7402	0.7094	0.4216	0.6455	0.8634
EWC-O [29]	86.57	0.8533	0.6309	0.5207	0.4614	0.6167	0.6270
SI-O [30]	86.57	0.8575	0.6882	0.4819	0.4819	0.6909	0.8066
MAS-O [31]	86.57	0.7850	0.7731	0.6831	0.7169	0.7338	0.8053
LwF-O [20]	86.57	0.7947	0.7387	0.6984	0.7204	0.7423	0.8073
ContEditIQA-O	**$\leq$ 25.36**	**0.9493**	**0.9087**	**0.7928**	**0.8168**	**0.8214**	**0.8794**

$\leq$ shows that in FFN layers itself not all params are being edited, mask may be sparse

IV. Conclusion

We proposed ContEditIQA, a selective ViT editing framework for continual BIQA that achieves a strong stability–

plasticity trade-off by restricting parameter updates to distortion-sensitive FFN layers. Experiments across six benchmarks show that targeted editing of fewer than 30% of backbone parameters consistently outperforms full-backbone continual learning methods, demonstrating that surgical model editing is a viable and scalable alternative to retraining for unified BIQA under evolving data conditions.

ACKNOWLEDGMENT

The authors acknowledge financial support from ANRF, India (Anusandhan National Research Foundation), under the PMECRG scheme (Grant No. ANRF/ECRG/2024/004417/ENS; Project No. IITM/ANRF/KP/529/2025).

REFERENCES

[1] M. Song, C. Chen, W. Song, and Y. Fang, "Uni-iqa: A unified approach for mutual promotion of natural and screen content image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 7, pp. 6371–6385, 2025.

[2] X. Min, Y. Gao, Y. Cao, G. Zhai, W. Zhang, H. Sun, and C. Wen Chen, "Exploring rich subjective quality information for image quality assessment in the wild," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 8, pp. 7778–7791, 2025.

[3] C. Wang, J. Lin, and X. Li, "Structural-equation-modeling-based indicator systems for image quality assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 8, pp. 7093–7107, 2025.

[4] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[5] D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[6] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2016.

[7] A. Ciancio, A. L. Targino da Costa, E. A. da Silva, A. Said, R. Samadani, and P. Obrador, "No-reference blur assessment of digital pictures based on multifeature classifiers," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 64–75, 2011.

[8] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.

[9] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqa database," in *Proceedings of the 11th International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–3.

[10] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2018.

[11] Y. Yang, L.-K. Huang, S. Chen, K. Ma, and Y. Wei, "Learning where to edit vision transformers," *Advances in Neural Information Processing Systems*, vol. 37, pp. 134 459–134 487, 2024.

[12] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.

[13] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory." *Psychological review*, vol. 102, no. 3, p. 419, 1995.

[14] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," in *Psychology of learning and motivation*. Elsevier, 1989, vol. 24, pp. 109–165.

[15] H. Zhao, F. Zhu, B. Ni, F. Zhu, G. Meng, and Z. Zhang, "Practical continual forgetting for pre-trained vision models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2026.

[16] F. Lyu, G. Cheng, D. Liu, L. Zhao, Z. Zhang, F. Hu, and L. Wang, "Mitigating catastrophic forgetting in online continual learning with dual-margin contrastive replay," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2025.

[17] W. Zhang, D. Li, C. Ma, G. Zhai, X. Yang, and K. Ma, "Continual learning for blind image quality assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 2864–2878, 2023.

[18] Y. Luo, J. Liu, W. Zhou, and X. Jin, "Multi-attribute continual learning for blind image quality assessment," in *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2025, pp. 1–5.

[19] J. Yang, Z. Wang, B. Huang, and L. Deng, "Continuous learning for blind image quality assessment with contrastive transformer," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[20] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.

[21] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Task-specific normalization for continual learning of blind image quality models," *IEEE Transactions on Image Processing*, vol. 33, pp. 1898–1910, 2024.

[22] H. Li, L. Liao, C. Chen, X. Fan, W. Zuo, and W. Lin, "Continual learning of blind image quality assessment with channel modulation kernel," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[23] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

[24] H. Chen, Y. Yang, N. Zhong, and K. Ma, "Hiding images in diffusion models by editing learned score functions," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 18 663–18 673.

[25] X. Li, S. Li, S. Song, J. Yang, J. Ma, and J. Yu, "Pmet: Precise model editing in a transformer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 17, 2024, pp. 18 564–18 572.

[26] V. Mazzia, A. Pedrani, A. Caciolai, K. Rottmann, and D. Bernardi, "A survey on knowledge editing of neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[27] X. Liu, J. Van De Weijer, and A. D. Bagdanov, "Rankiqa: Learning from rankings for no-reference image quality assessment," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1040–1049.

[28] M. Geva, R. Schuster, J. Berant, and O. Levy, "Transformer feed-forward layers are key-value memories," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 5484–5495.

[29] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[30] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *International Conference on Machine Learning*. PMLR, 2017, pp. 3987–3995.

[31] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, "Memory aware synapses: Learning what (not) to forget," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 139–154.

[32] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.