Breaking Teacher's Mean Regression Bias: Boundary-Aware Contrastive Distillation for Endoscopic Super-Resolution

Wenbin Zuo[1], Hongying Liu[1⋆], Yushuai Cui[1], Fanhua Shang[2], and Liang Wan[1,3]

[1] Medical School, Tianjin University, Tianjin, China
[2] School of Computer Science and Technology, Tianjin University, Tianjin, China
[3] School of Computer Software, Tianjin University, Tianjin, China
{wbzuo,hyliu2009,cuiyushuai02,fhshang,lwan}@tju.edu.cn

Abstract. Endoscopic image super-resolution (SR) is vital for identifying subtle clinical abnormalities. However, given the ill-posed nature of SR, optimization using L_1 or L_2 losses inherently leads to a mean regression problem, resulting in smoothed textures. This issue is particularly severe in endoscopic images due to their highly uniform backgrounds. To enable clinical deployment under limited computational resources, Knowledge Distillation (KD) provides a practical compression solution; however, it significantly exacerbates this mean regression issue. Since the teacher model is already a variance-suppressed product of mean regression, forcing a lightweight student to rigidly mimic it causes the student to passively inherit and amplify this bias. To address this unexplored critical problem, we propose a novel boundary-aware contrastive distillation framework (BACD). By establishing extreme degraded samples as a lower bound to repel blurry artifacts, and utilizing gradient-based teacher-optimized inputs as a safe upper bound to extract latent high-frequency priors, our contrastive loss fundamentally reshapes the optimization space. This method effectively overcomes the mean regression problem exacerbated by KD, enabling the lightweight student network to successfully recover high-fidelity clinical details.

Keywords: Super-Resolution · Knowledge Distillation · Mean Regression · Contrastive Learning · Endoscopy.

1 Introduction

High-fidelity endoscopic imaging is the cornerstone of modern minimally invasive surgery and computer-assisted diagnosis, where the precise delineation of mucosal textures and microvascular patterns is critical for identifying early-stage lesions [1]. However, physical constraints on the size of distal sensors in endoscopes impose a fundamental bottleneck on spatial resolution. While deep

⋆ Corresponding author.

learning-based single image super-resolution (SISR) has emerged as a computational remedy, its deployment in clinical settings is severely compromised by a theoretical pathology inherent to standard optimization objectives: the mean regression problem [2–6].

Mathematically, SISR is a highly ill-posed inverse problem where a single low-resolution (LR) observation x maps to a diverse manifold of plausible high-resolution (HR) solutions y. From a probabilistic perspective, optimizing a neural network via the prevalent L_1 or L_2 loss functions forces the model to learn the conditional central tendency of the posterior distribution $p(y|x)$, which corresponds to the conditional median under L_1 loss and the conditional mean under L_2 loss [5, 7] (In what follows, the discussion focuses on the L_1 case for clarity.). In natural scene images, pronounced edge structures often act as anchors that implicitly limit the variance of the posterior distribution. Endoscopic imagery, by contrast, operates near the opposite extreme: the visual field is largely composed of smooth, weakly textured mucosal surfaces, offering little structural guidance for uncertainty reduction. Under severe downsampling (e.g., $4\times$, $8\times$), distinct HR vascular topologies collapse into visually identical LR patches, causing the posterior $p(y|x)$ to become exceptionally broad and multi-modal. To minimize expected risk over this uncertainty, the network converges to the statistical average of all possible structures—resulting in a deterministic, textureless blur that obliterates diagnostic details [2].

To meet the strict latency requirements of operating rooms, knowledge distillation (KD) is routinely employed to compress heavy SR models into lightweight students [7–12]. Crucially, we identify a critical flaw in this paradigm: standard KD catastrophically exacerbates mean regression through a mechanism we term bias inheritance. The teacher model, having been trained with pixel-wise losses, is itself a variance-suppressed approximation of reality, restricted to a smoothed sub-manifold of the HR space [2]. When a student is forced to mimic this teacher (via feature or output alignment), it is not learning the true data distribution but rather a double regression of an already simplified statistic. The student is thus trapped by dual anchors—the dataset median (via its own L_1 loss) and the teacher's smoothed output (via KD loss)—creating a deep local minimum where high-frequency variance is mathematically suppressed. To the best of our knowledge, no prior work has explicitly addressed the amplification of mean regression bias in the distillation of super-resolution models.

In this paper, we propose **Boundary-Aware Contrastive Distillation (BACD)**, a pioneering framework that abandons the conventional paradigm of point-to-point mimicking. Our core insight is that while the teacher's outputs are heavily smoothed by dataset-level uncertainty, its latent parameter space retains the capacity to synthesize high-frequency clinical details. Instead of blindly trusting the teacher's compromised predictions, BACD reshapes the optimization landscape by probing the teacher to construct explicit statistical boundaries: (1) The lower bound (negative anchor): to mathematically isolate the mean regression bias, we completely strip the input of its structural priors via extreme degradation. Deprived of geometric context, the frozen teacher is forced into ab-

solute epistemic uncertainty. This uniquely establishes a precise representation of the mean regression trap, serving as a highly effective negative anchor; (2) The upper bound (positive anchor): Conversely, extracting the teacher's maximal high-fidelity capability is non-trivial. Simply injecting Ground Truth (GT) signals into the low-resolution input triggers a severe Manifold Mismatch; the teacher's convolutional filters interpret these out-of-distribution (OOD) frequencies as adversarial noise and actively suppress them. To unlock the teacher's true potential, we introduce a novel Gradient-Based Input Optimization strategy By leveraging the teacher's own gradient to inversely optimize the input tensor, subject to strict topological constraints within the valid LR manifold, we safely project the GT prior into a space the teacher comprehends. This prompts the teacher to bypass its uncertainty filters and generate a variance-rich, hallucination-free upper bound. By integrating these bounds into a contrastive InfoNCE objective, BACD exerts a dynamic repulsive gradient. This explicitly penalizes the bias inheritance, dynamically injecting structural variance back into the student network and successfully recovering high-fidelity microvascular details. The main contributions of this work are summarized as follows:

- We analyze the exacerbation of mean regression in knowledge distillation for super-resolution, demonstrating that conventional distillation implicitly drives the student toward variance-suppressed teacher outputs, thereby amplifying dataset-level smoothing bias.
- We introduce a gradient-based input optimization strategy. Instead of directly injecting ground truth signals which causes Manifold Mismatch, we inversely optimize the input to safely extract the teacher's latent high-fidelity capabilities. This uniquely establishes a topologically safe and variance-rich positive anchor (upper bound) for contrastive learning.
- We propose BACD, a novel framework that reshapes the optimization landscape by introducing a push and pull mechanism to counteract mean regression collapse. Extensive experiments on endoscopic datasets demonstrate that our lightweight student network achieves state-of-the-art performance.

2 Methodology

2.1 Theoretical Formulation of Exacerbated Mean Regression

SR is fundamentally an ill-posed inverse problem where a single LR input $x \in \mathcal{X}$ maps to a highly multimodal distribution of plausible HR images, denoted as $P(y|x)$ [2, 5, 6]. When training a Teacher network T_θ using the standard $\mathcal{L}_1$ loss, the optimization objective is defined as:

$$\theta^* = \arg\min_\theta \mathbb{E}_{(x,y)\sim\mathcal{D}} \left[\|T_\theta(x) - y\|_1\right] \tag{1}$$

Statistically, the optimal solution for the $\mathcal{L}_1$ objective is the conditional spatial median of the target distribution: $T^*(x) = \text{Median}(P(y|x))$. In endoscopic imaging, where textures are highly uniform and distinct high-frequency structural

landmarks are sparse. Consequently, the conditional median fundamentally acts as a low-pass filter, forcing $T(x)$ to collapse into a smoothed, variance-suppressed manifold, resulting in the well-known mean regression bias.

In standard KD, the student network S_ϕ is optimized via a joint objective combining the GT and the teacher's knowledge:

$$\mathcal{L}_{KD\_Total} = \lambda_{gt}\|S_\phi(x) - y\|_1 + \lambda_{kd}\|S_\phi(x) - T_\theta(x)\|_1 \tag{2}$$

This formulation explicitly reveals the root cause of the exacerbated mean regression problem. The first term, tracking the GT, already suffers from spatial averaging due to the ill-posed mapping. Critically, the second term ($\mathcal{L}_{kd}$) acts as an aggressive variance regularizer rather than providing structural guidance. Since the teacher's output $T_\theta(x)$ is a deterministic, smoothed scalar field ($\mathrm{Var}(T(x)) \ll \mathrm{Var}(y)$), minimizing the distillation loss explicitly penalizes the student for generating any high-frequency variance that deviates from this smoothed state.

It should be noted that the above discussion is intended as an analytical motivation. Throughout the following sections, the terms "lower bound" and "upper bound" do not refer to the worst or best samples in the data distribution. Instead, they are used in an operational sense to denote low- and high-fidelity teacher-induced response anchors within the frozen teacher's learned parameter and feature space.

2.2 Negative Sample Construction: The Repulsive Lower Bound

To break this bias inheritance, our first step is to mathematically isolate the pure mean regression artifact. We achieve this by intentionally stripping the input x of its latent structural priors. Specifically, we apply a severe degradation transformation $\mathcal{D}(\cdot)$ parameterized by a low-pass Gaussian filter with variance σ^2_{blur} and additive noise $\eta \sim \mathcal{N}(0, \sigma^2_{noise})$:

$$x_{deg} = \mathcal{D}(x) = (x \circledast k_{LP}(\sigma_{blur})) + \eta \tag{3}$$

By feeding x_{deg} into the frozen teacher, we obtain $y^- = T_\theta(x_{deg})$. Deprived of informative geometric context, the teacher is forced into a state of absolute epistemic uncertainty.

2.3 Positive Sample Construction: The Latent Upper Bound

Conversely, constructing a valid positive anchor remains challenging. A natural approach to counteract mean regression is to enhance the LR-to-HR correspondence by injecting additional high-frequency cues. While directly downsampling ground-truth high-frequency content into the LR input appears intuitive, forcing the teacher to assimilate such sharp signals leads to severe manifold mismatch, since the teacher's fixed convolutional filters treat these out-of-distribution frequencies as noise. To safely access the teacher's high-fidelity response space, we introduce a Gradient-Based Input Optimization strategy.

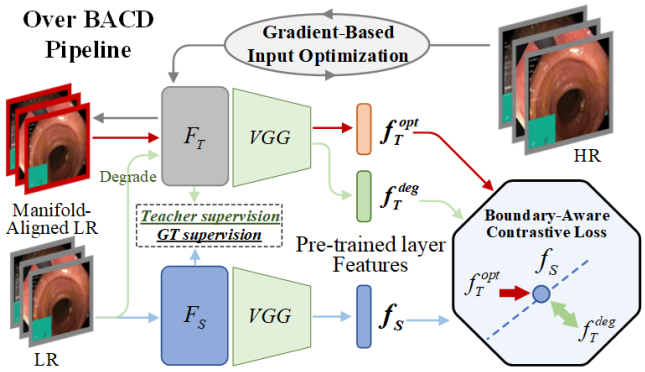

Fig. 1. The overall architecture of BACD.

We initialize a learnable latent state $x_{opt}^{(0)} = x$. Keeping the teacher parameters θ strictly frozen, we inversely optimize the input tensor over K steps to minimize the structural discrepancy with the GT:

$$x_{opt}^{(k+1)} = x_{opt}^{(k)} - \alpha \nabla_{x_{opt}} \left\| T_\theta(x_{opt}^{(k)}) - y \right\|_1, \quad k \in \{0, 1, \ldots, K-1\} \tag{4}$$

where α is the optimization step size. The final optimized input prompts the teacher to generate the positive anchor: $y^+ = T_\theta(x_{opt}^{(K)})$. This gradient-guided formulation holds profound theoretical significance. Instead of modifying the fixed parameter space θ, which would destroy the teacher's intrinsic learned priors, we project the high-frequency structural cues of y directly onto the valid LR input manifold. The optimized state $x_{opt}^{(K)}$ acts as an "idealized probe". It bypasses the teacher's epistemic uncertainty filters and naturally elicits the theoretical maximum structural fidelity the network is capable of synthesizing. Thus, this mechanism successfully unlocks a variance-rich, topologically safe upper bound that is natively comprehensible to the student without OOD hallucinations.

2.4 Boundary-Aware InfoNCE Optimization

To increase the contrastive sample density and enforce spatial consistency, we introduce a geometric augmentation mechanism. For a given input pair (x, y), we define a set of spatial transformations $\mathcal{A} = \{I, F_h, F_v, F_t\}$, corresponding to identity, horizontal flip, vertical flip, and transpose. These augmentations are applied prior to the bound construction, yielding a multi-view batch $x_{aug} = \mathcal{A}(x)$. After obtaining the bounds $T_\theta(x_{deg,aug})$ and $T_\theta(x_{opt,aug})$, we apply the inverse transformations $\mathcal{A}^{-1}$ to spatially align the generated tensors. This yields spatially consistent multi-view positive and negative sets: $\mathcal{P} = \{y_i^+\}_{i=1}^{|\mathcal{A}|}$ and $\mathcal{N} = \{y_i^-\}_{i=1}^{|\mathcal{A}|}$. With the bounds established, we project the student's prediction $\hat{y} = S_\phi(x)$, alongside the sets $\mathcal{P}$ and $\mathcal{N}$, into a deep perceptual manifold using a

pre-trained feature extractor $\Phi_{vgg}(\cdot)$. The features are aggregated via Adaptive Average Pooling and L_2-normalized to obtain compact semantic embeddings: $f_{\hat{y}}$, $\{f_{y_i^+}\}$, and $\{f_{y_i^-}\}$. We formulate our Boundary-Aware Contrastive Distillation (BACD) loss using a multi-positive InfoNCE objective:

$$\mathcal{L}_{BACD} = -\log \frac{\sum_{p\in\mathcal{P}} \exp(\text{sim}(f_{\hat{y}}, f_p)/\tau)}{\sum_{p\in\mathcal{P}} \exp(\text{sim}(f_{\hat{y}}, f_p)/\tau) + \sum_{n\in\mathcal{N}} \exp(\text{sim}(f_{\hat{y}}, f_n)/\tau)} \tag{5}$$

where $\text{sim}(\cdot,\cdot)$ denotes the cosine similarity and τ controls the temperature scaling.

3 Experiments and Results

3.1 Datasets and Implementation

We evaluate the proposed BACD framework on two public endoscopic benchmarks. The Labeled Images subset of HyperKvasir [13] contains 10,662 images spanning 23 clinical categories, exhibiting a pronounced long-tailed distribution that reflects real-world data imbalance and poses challenges for stable representation learning. In addition, we adopt SegSTRONG-C [14], a dataset designed for robust segmentation under three adverse conditions: blood, low illumination, and smoke. We merge its original validation and test sets (900 images per condition each) and re-split them with a 4:1 training-to-testing ratio, yielding 1,440 training and 360 testing images per condition.

Implemented in PyTorch (BasicSR) on twp RTX 4090 GPUs, the Student is trained for 150K iterations using Adam (batch size 16, initial lr 4×10^{-4} with multi-step decay). For BACD, we set the optimization step $K = 3$, learning rate $\alpha = 0.5$, temperature $\tau = 0.07$, and loss weights $\lambda_{kd} = 0.007$, $\lambda_{gt} = 1.0$. Evaluation metrics include PSNR, LPIPS [15], and downstream clinical metrics (ACC, F1).

3.2 Experiments

To evaluate the generalization capability and effectiveness of BACD, we conduct experiments using two representative architectures: the CNN-based RCAN [16] and the Transformer-based SwinIR [17]. For RCAN, the Teacher is configured with 64 channels and 10×20 blocks (15.6M parameters), while the Student is compressed to 10×6 blocks (5.2M parameters). For SwinIR, the Teacher employs 180 channels with 6 blocks (11.9M parameters), and the Student is reduced to 60 channels and 4 blocks (1.2M parameters). BACD is compared against the corresponding pre-trained Teachers and several representative KD baselines, including $\mathcal{L}_1$ logits, FAKD [8], DCKD [18], and MiPKD [12]. All methods are retrained under identical data protocols for fair comparison.

Table 1. Quantitative comparison on HyperKvasir and SegSTRONG-C datasets. Best student results are in bold.

Method	**HyperKvasir**				**SegSTRONG-C**			
	×4		×8		×4		×8	
	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓
CNN-based Architectures (RCAN)								
Teacher	39.20	0.1506	32.91	0.3526	29.12	0.1285	25.45	0.3215
Logits	38.64	0.1541	32.62	0.3329	26.91	0.1512	24.32	0.3056
FAKD	38.71	0.1525	32.68	0.3412	26.95	0.1485	24.38	0.3112
DCKD	38.69	0.1538	32.65	0.3455	26.98	0.1442	24.42	0.3155
MiPKD	38.75	0.1518	32.70	0.3384	27.05	0.1395	24.46	0.3188
Ours	**38.84**	**0.1351**	**32.73**	**0.2834**	**27.18**	**0.1156**	**24.55**	**0.2854**
Transformer-based Architectures (SwinIR)								
Teacher	39.51	0.1434	33.38	0.3352	29.45	0.1224	25.85	0.3125
Logits	38.72	0.1534	32.81	0.3241	27.15	0.1488	24.55	0.2985
FAKD	38.85	0.1485	32.89	0.3315	27.22	0.1425	24.62	0.3021
DCKD	38.81	0.1492	32.86	0.3288	27.28	0.1386	24.68	0.3055
MiPKD	38.90	0.1456	32.85	0.3295	27.35	0.1312	24.75	0.3088
Ours	**39.06**	**0.0816**	**32.96**	**0.2735**	**27.48**	**0.1045**	**24.88**	**0.2745**

Table 2. Downstream segmentation performance on SegSTRONG-C.

Method (SwinIR)	**Scale × 4**		**Scale × 8**	
	Acc (%) ↑	**F1 (%) ↑**	**Acc (%) ↑**	**F1 (%) ↑**
Teacher (Upper Bound)	97.55	95.82	96.64	94.45
Logits (Baseline)	96.21	93.55	95.12	92.25
FAKD	96.52	93.92	95.65	92.84
DCKD	96.68	94.25	95.88	92.15
MiPKD	96.85	94.61	95.05	92.48
Ours (BACD)	**97.28**	**95.34**	**95.92**	**92.86**

Performance Comparison Table 1 reports the quantitative comparisons. Although existing KD methods (FAKD, DCKD, MiPKD) achieve marginal PSNR gains over the baseline, their LPIPS scores consistently degrade toward the teacher, indicating aggravated mean-regression bias and suppressed structural variance. This observation validates our analysis that rigid alignment to an $\mathcal{L}_1$-smoothed teacher traps the student in a smoothing regime. In contrast, BACD simultaneously attains the highest PSNR and significantly improved LPIPS, even surpassing the teacher. As illustrated in Fig. 2, BACD reconstructs sharper microvascular structures with clearer high-frequency details, whereas competing KD methods produce over-smoothed results. By repelling the mean-regression bounds and aligning with optimized positive counterparts, BACD effectively restores structural variance, enabling the compact student to escape the regression ceiling.

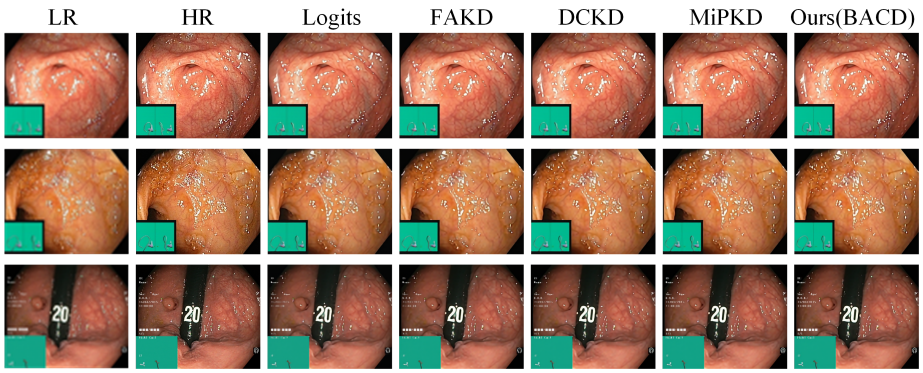

Fig. 2. Visual comparison with state-of-the-art methods.

Table 3. Ablation study on the HyperKvasir dataset (SwinIR × 8).

Model	Neg Anchor	Naive Pos Anchor	Opt Pos Anchor	PSNR ↑	LPIPS ↓
Baseline	No	No	No	32.81	0.3241
Variant A	Yes	No	No	32.86	0.3015
Variant B	Yes	Yes	No	32.83	0.3150
BACD	**Yes**	No	**Yes**	**32.96**	**0.2735**

Downstream Clinical Utility To assess whether variance recovery translates into clinical benefit, we evaluate the reconstructed HR images on downstream lesion segmentation using SegSTRONG-C. As reported in Table 2, conventional KD methods suffer from boundary blurring caused by mean regression, resulting in inferior F1 scores. In contrast, BACD achieves 97.28% Accuracy and 95.34% F1 at ×4, outperforming all baselines. These results indicate that mitigating mean-regression bias effectively restores discriminative lesion boundaries, thereby improving downstream segmentation reliability.

3.3 Ablation Study

We perform ablation experiments on HyperKvasir (SwinIR ×8) to validate the proposed bound design. As shown in Table 3, the baseline exhibits pronounced mean-regression bias. Introducing only the repulsive Negative Anchor (Variant A) alleviates this effect and improves LPIPS to 0.3015. However, directly injecting GT high-frequency signals (Variant B) leads to manifold mismatch, degrading LPIPS to 0.3150. In contrast, BACD employs gradient-based inverse optimization to project GT priors into the teacher manifold, resolving the mismatch and achieving the best PSNR–LPIPS trade-off.

4 Conclusion

In this paper, we propose Boundary-Aware Contrastive Distillation (BACD) to fundamentally solve the bias inheritance problem in endoscopic super-resolution. By establishing a degraded repulsive lower bound and a gradient-optimized, hallucination-free upper bound, BACD creates a dynamic push-pull optimization space that actively prevents manifold mismatch. This liberates highly compressed student networks from mean-regression blur, achieving state-of-the-art perceptual fidelity and significantly improving real-time downstream clinical segmentation.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 62276182), Tianjin Natural Science Foundation (Nos. 24JCYBJC01230, 24JCYBJC01460, 24JCZDJC01190), and Tianjin Municipal Education Commission Research Plan (No. 2024ZX008).

Disclosure of Interests. We have no conflicts of interest to disclose.

References

1. Park, J.Y.: Image-enhanced endoscopy in upper gastrointestinal disease: focusing on texture and color enhancement imaging and red dichromatic imaging. Clinical Endoscopy **58**(2), 163–180 (2025)
2. Xu, T., Li, L., Mi, P., Zheng, X., Chao, F., Ji, R., Shen, Q.: Uncovering the oversmoothing challenge in image super-resolution: Entropy-based quantification and contrastive optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(9), 6199–6215 (2024)
3. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690. IEEE, Piscataway (2017)
4. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Springer, Cham (2018)
5. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp. 694–711. Springer International Publishing, Cham (2016)
6. Freirich, D., Michaeli, T., Meir, R.: A theory of the distortion-perception tradeoff in wasserstein space. In: Advances in Neural Information Processing Systems 34, pp. 25661–25672. Curran Associates, Inc., Red Hook (2021)
7. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6228–6237. IEEE, Piscataway (2018)
8. He, Z., Dai, T., Lu, J., Jiang, Y., Xia, S.T.: Fakd: Feature-affinity based knowledge distillation for efficient image super-resolution. In: 2020 IEEE International Conference on Image Processing (ICIP), pp. 518–522. IEEE, Piscataway (2020)

9. Lee, W., Lee, J., Kim, D., Ham, B.: Learning with privileged information for efficient image super-resolution. In: European Conference on Computer Vision, pp. 465–482. Springer International Publishing, Cham (2020)
10. Wang, Y., Lin, S., Qu, Y., Wu, H., Zhang, Z., Xie, Y., Yao, A.: Towards compact single image super-resolution via contrastive self-distillation. arXiv preprint arXiv:2105.11683 (2021)
11. Zhang, Y., Chen, H., Chen, X., Deng, Y., Xu, C., Wang, Y.: Data-free knowledge distillation for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7852–7861. IEEE, Piscataway (2021)
12. Li, S., Zhang, Y., Li, W., Chen, H., Wang, W., Jing, B., Hu, J.: Knowledge distillation with multi-granularity mixture of priors for image super-resolution. In: 13th International Conference on Learning Representations (ICLR 2025), pp. 27216–27232. ICLR, Singapore (2025)
13. Borgli, H., Thambawita, V., Smedsrud, P.H., Hicks, S., Jha, D., Eskeland, S.L., De Lange, T.: HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Scientific Data **7**(1), 283 (2020)
14. Ding, H., Zhang, Y., Lu, T., Liang, R., Shu, H., Seenivasan, L., Unberath, M.: SegSTRONG-C: Segmenting surgical tools robustly on non-adversarial generated corruptions–An EndoVis' 24 Challenge. arXiv preprint arXiv:2407.11906 (2024)
15. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595. IEEE, Piscataway (2018)
16. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 695–704. IEEE, Piscataway (2018)
17. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844. IEEE, Piscataway (2021)
18. Zhou, Y., Qiao, J., Liao, J., Li, W., Li, S., Xie, J., Lin, S.: Dynamic contrastive knowledge distillation for efficient image restoration. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 10861–10869. AAAI Press, Washington, DC (2025)