Towards Generalized Representations for Low-Light Understanding: When Signal Constancy Meets Semantic Enrichment

Yifan Li[1] Haofeng Huang[1] Wenhan Yang[2] Jiaying Liu[1*]
[1] Wangxuan Institute of Computer Technology, Peking University [2] Peng Cheng Laboratory
liyifan02@stu.pku.edu.cn, yangwh@pcl.ac.cn, {hhf, liujiaying}@pku.edu.cn

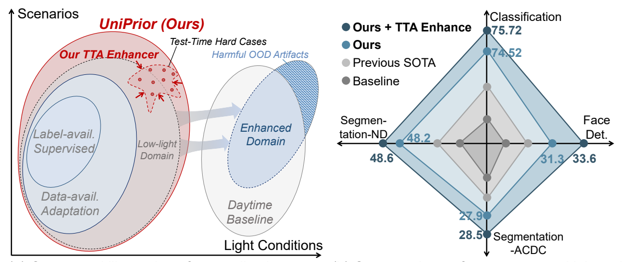
Abstract

Low-light degradation hampers machine understanding at night. Existing methods either overfit labeled data (paired supervision) or specific distributions (unpaired supervision), resulting in poor generalization under unseen degradations. In this paper, we propose ***UniPrior****, a unified prior-based low-light adaptation framework that integrates the general semantic prior embedded in vision foundation models (VFMs) with illumination-invariant priors, to capture both stable and changing semantics under varied low-light degradation without any real low-light training data. In detail, the illumination-invariant prior is used as an auxiliary input, and a parallel decoder reconstructs it as a regularization target, enforcing representation consistency and reducing feature drift. Such signal constancy enables us to build a VFM-aligned semantic space via a contrastive training strategy guided by VFM self-correlation maps, enriching features with high-level cues, thereby improving adaptation to diverse low-light conditions. Beyond high-level features, we also give a joint consideration of such unified prior and low-level signal space through our machine-oriented enhancement scheme. We extend the signal prior to handle overexposure and inject VFM-guided semantic cues into the enhancement process via a CLIP-based loss. This coupling of semantic alignment and pixel correction enables sample-adaptive optimization to improve performance. Extensive experiments on multiple low-light tasks demonstrate our method's superiority and practical utility. Our project is publicly available at https://lyf1212.github.io/UniPrior.*

1. Introduction

Low-light conditions cause a series of visual degradations, including detail loss, reduced visibility, and intensive noises, which severely hinder both human perception and machine understanding. Many works have effectively improved the human visual quality of low-light images, ranging from early manually designed algorithms [13, 31] to recent learning-based models [12, 40]. However, most existing methods are designed for human visual preferences and often introduce artifacts or lose task-relevant features, leading to degraded performance in downstream machine vision tasks, as illustrated by Fig 2(c). Moreover, solely utilizing human visual-oriented enhancement methods for machine perception and freezing the understanding module, will inherently hinder generalization to diverse real-world conditions. Despite remarkable progress in machine perception, low-light degradation remains a fundamental obstacle to deploying robust vision systems in real-world scenarios such as autonomous driving and intelligent surveillance.

*Corresponding author.



(a) Conceptual illustration of generalization gaps (b) Our superior performance on multiple tasks

Figure 1. (a) Existing methods are constrained by the biased light conditions and scenarios. By comparison, our **UniPrior** integrates unified signal/feature-level priors, enabling broader generalization. Building upon this, our TTA Enhancer further improves generalizability to test-time hard cases through sample-adaptive optimization. (b) Our method achieves state-of-the-art performance on multiple low-light understanding tasks by a large margin.

Several approaches have been proposed to mitigate this issue. Supervised training has demonstrated powerful learning capabilities in the era of deep learning across many domains, as denoted by Fig. 2(a). Unfortunately, in the context of low-light image enhancement, the scarcity of large-scale labeled low-light data makes the fully-supervised paradigm prone to overfitting to specific training sets, leading to poor generalization [14, 16, 36]. To overcome the overfitting and label scarcity issues, much literature adopts a domain adaptation paradigm, which transfers knowledge learned from labeled normal-light data by leveraging unlabeled low-light data as a target domain indicator [23, 29, 38], as denoted

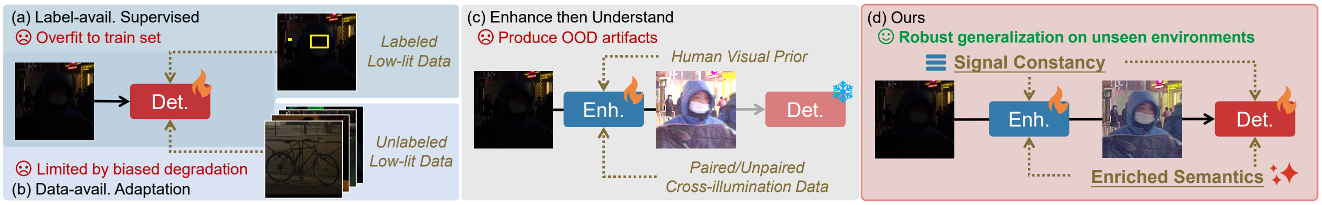


Figure 2. Principle diagram comparison of different learning paradigms for low-light understanding. (a) *Supervised methods with available labels* are limited by data constraints and poor generalization. (b) *Adaptation methods with available data* do not need labels, but may overfit to data-specific degradations or scenarios. (c) *Enhance then understand* fails to guarantee understanding performance with solely signal-level constraints. (d) Unlike previous methods, we propose a **unified prior** framework that integrates signal intrinsics and enriched semantic priors, reducing data/label dependence while ensuring real applicability with superior generalization under low-light conditions.

by Fig. 2(b). However, these methods are constrained by dataset-specific degradation patterns and scene diversity, leading to poor generalization to real-world and extremely low-light conditions.

Beyond modeling low-light distribution from data, some methods turn to capturing intrinsic priors embedded in visual signals, aiming for a principled characterization of visual degradation, *e.g.*, simulated ISP [6], Retinex theory [33], or color invariant edge detection [20], which effectively disentangle semantic content and reduces the impact of biased low-light conditions. However, such physics-informed priors are still vulnerable to complex noise patterns, such as sensor-dependent, non-Gaussian, and scene-specific noise [20, 40]. Accordingly, Sim-MinMax [27] is not limited to real physical imaging processes, but synthesizes more challenging low-light data to enhance discriminative representations through multi-task contrastive learning. Although existing methods leverage prior knowledge to introduce visual inductive bias, they lack generalizable semantic representations and are limited in robustness to varying low-light degradation and dynamic scenarios.

Based on the above analysis, we summarize the key challenges as two-fold: 1) Low-light environments involve complex distortions like low visibility, noise, and motion blur, making exhaustive sampling impractical and hindering generalization, especially in rare scenarios like neon lights or nighttime reflections. 2) Existing methods lack large model priors, limiting semantic consistency and generalization across diverse degradation patterns in real-world.

To address the diverse and complex degradations as well as dataset discrepancies inherent in low-light scenarios, we propose **UniPrior**, a unified prior-integrated framework enriched with the knowledge from large models to achieve low-light adaptation without any real low-light training data. Our approach is characterized by two core ideas: 1) establishing an illumination-invariant signal constancy to stabilize feature distributions, and 2) enriching semantic representations using the general prior embedded in Vision Foundation Models (VFMs) [30, 32] to ensure robust adaptation across diverse low-light environments. The use of large models with prior knowledge improves both semantic stability and discriminability, helping the model better adapt to complex low-light conditions.

In detail, we first establish signal constancy with an illumination-invariant prior and a decoding-based regularization scheme. We then enrich semantic representations by jointly leveraging VFM-driven correlation alignment and spatial context-informed contrastive learning to construct a unified feature space. To bridge high-level perception and low-level correction, we incorporate the unified prior into a test-time adaptive (TTA) machine-oriented enhancer that dynamically characterizes the unseen low-light distribution. Extensive experiments demonstrate that our method achieves strong generalization and state-of-the-art performance across low-light classification, segmentation, and face detection tasks.

Our contributions are summarized as follows:

- We incorporate the illumination-invariant prior as signal constancy to alleviate feature shifts with negligible parameter overhead. To effectively enforce the network to learn compact and illumination-invariant representations, we design a decoding-based regularization scheme, enabling universal generalization under complex and unknown lighting conditions without any real low-light data.
- Beyond signal constancy, we propose a semantic enrichment mechanism that leverages the universal and illumination-insensitive knowledge embedded in VFMs. The proposed mechanism aligns and enhances representations within the VFM embedding space. We design an innovative contrastive regularization scheme based on the VFM-driven spatial intra-feature correlation. The combination of signal intrinsic priors and VFM-enriched semantics improves our generalization synergistically.
- Built on stable signal constancy and enriched semantics, we propose a machine-oriented enhancement scheme. An extended signal prior robust to overexposure and VFM-guided semantic cues are injected into low-level process, linking semantic alignment with pixel correction to reduce domain gap. Our method is naturally illumination-free and distribution-agnostic, enabling per-sample test-time adaptation (TTA) for improved performance.

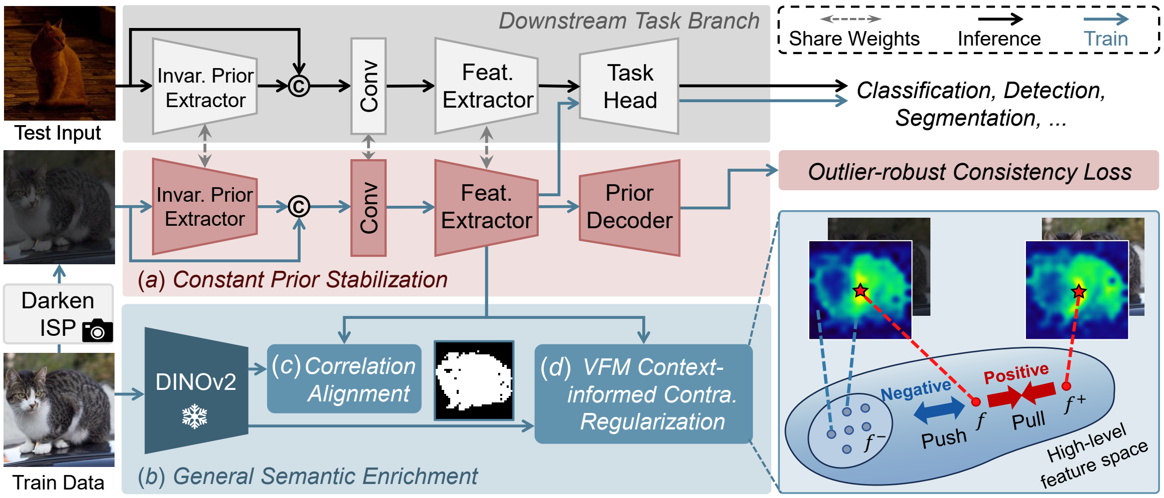


Figure 3. The overall framework of our proposed **UniPrior**. (a) **Constant Prior Stabilization** leverages the illumination-invariant prior as an auxiliary input through residual convolution, while decoding-based regularization enforces consistent representation. (b) **General Semantic Enrichment** fully exploits the general and degradation-robust prior embedded in VFM, including (c) **Correlation Alignment** and (d) **VFM Context-informed Contrastive Regularization**, constructs an aligned semantic space that helps achieve robust representation across varying lighting conditions. These modules effectively integrate signal constancy, semantic enrichment, and semantic alignment, collectively enhancing generalization in a data- and label-free manner.

2. Related Works

Low-light Image Enhancement. This category of methods offers an intuitive solution to achieve low-light adaptation by first restoring low-light inputs to those of normal light. Low-light enhancement methods can be categorized into three paradigms based on the priors they mainly utilize: The first one employs **physical priors** [10, 40, 41, 48, 49] to decouple images into illumination intensity components and illumination-invariant counterparts, and adjusts the former one to obtain normal-lit images. The second one adopts **handcrafted priors** aiming at human visual perception [3, 9, 12, 13, 21, 28, 31] by imposing constraints on appearance attributes, including color, noise, histograms, etc. The third paradigm leverages the learning capability of deep networks to utilize **data priors**, constructing pixel-wise mappings [1, 17, 42, 43, 47] or distribution-wise mappings [18, 19, 24, 37, 45] from under-exposed images to normal-light ones. Although "enhance then understand" poses a straightforward way for low-light machine perception, it only operates in the intermediate pixel space and leaves the downstream model alone, which is impaired by unrealistic artifacts and detail loss as illustrated in Fig. 1(a), hampering accurate understanding.

Low-light Domain Adaptation. Beyond preprocessing with enhancement methods, domain adaptation seeks to transfer knowledge gained from labeled normal-lit data to unlabeled low-light scenes. YOLO-in-the-dark [35] combines restoration and recognition through a "glue layer" to utilize shared features. HLA-Face [38, 39] proposes a joint high-low adaptation framework with contrastive learning. WiiD [23] designs a bi-directional low-level alignment and masked contrastive training strategy to highlight important regions. Morawski *et al* [29] leveraged CLIP [32] to learn content and context cues with prompt tuning. Although effective, these methods are inevitably hindered by biased light conditions and scenarios inherent in training data, failing to elucidate the intrinsic semantics.

Zero-shot Domain Adaptation in Low-light Conditions. This branch has become a promising trend for comprehensive low-light understanding, requiring no real low-light data and mitigating biased overfitting. MAET [6] introduces an orthogonal tangent regularization with a physical ISP simulator, while CIConv [20] uses a color-invariant physical prior for illumination robustness. SimMinMax [27] employs a similarity min-max framework for compact representation via multi-task contrastive learning, and DAI-Net [8] enhances illumination invariance through reflectance decomposition-based regularization. By leveraging illumination-invariants and VFMs, our method better generalizes to complex and unforeseen low-light scenarios.

3. UniPrior: Unifying Priors via Signal Constancy and Semantic Enrichment

3.1. Motivations and Architecture

Low-light signals are often affected by external factors such as poor lighting, noise, or interference, which distort how they appear. However, the objects behind these signals usually have stable and intrinsic semantic natures that remain unchanged across different conditions. Most existing models struggle to capture stable features and handle the wide signal variations caused by complex environments and se-

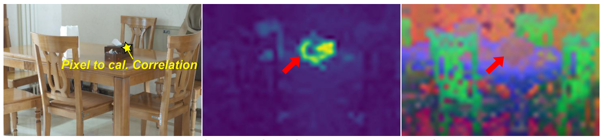


Figure 4. Empirical verification of DINOv2 **correlation map** compared to original features. We calculate the similarity of the anchor feature (marked by the yellow star at the input image). The tissue box is blended and indistinguishable from surrounding irrelevant objects in the original DINOv2 feature, but is highlighted in our proposed DINOv2 correlation map, demonstrating that the correlation map could provide more discriminative cues.

mantics. The rise of illumination-invariant priors [20, 40] and large vision foundation models [30, 32] offers chances to improve generalization under low-light environments: 1) *Illumination-invariant priors* provide stable representations by suppressing signal-level perturbations across diverse lighting conditions; 2) Foundation models embed strong, generalizable *semantic priors* that remain robust across scenes and degradations.

To fully exploit the complementary strengths of these two types of priors, we propose **UniPrior**, a unified prior-integrated framework tailored for low-light understanding without any real low-light training data. The overall architecture is shown in Fig. 3. Specifically, we first construct a signal constancy mechanism through a prior-conditioned architecture and a decoding-based regularization, which stabilizes features and narrows the domain gap. We then enrich high-level semantics by aligning features with vision foundation models, and further inject the learned semantic cues into the low-level signal space via a sample-adaptive machine-oriented enhancement strategy. This cross-level design bridges pixel correction with semantic alignment, naturally enabling a unified prior that supports robust and generalizable adaptation to diverse low-light scenarios.

3.2. Signal Constancy: Illumination Invariant Prior-conditioned Adaptation

We build the signal constancy around the physically-informed Illumination-Invariant Prior [40], proposing the corresponding module and training constraints. As for the *module design*, we utilize the prior as a stable representation that is robust to complex lighting conditions, feeding into the original backbone F with an auxiliary convolution layer. Given an image $I \in \mathbb{R}^{h\times w\times 3}$, we first extract the prior $p_{\text{iip}} = G_{\text{iip}}(I) \in \mathbb{R}^{h\times w\times 6}$. G_{iip} is initialized by the pretrained prior extractor in [40] and is finetuned online within our framework. More details can be found in our supplementary material. We then utilize a simple convolution with kernels $K_{\text{merge}} \in \mathbb{R}^{c_{\text{out}}\times c_{\text{in}}\times k_{\text{h}}\times k_{\text{w}}}$ to merge these two channel-concatenated components, where $c_{\text{in}} = 3+6 = 9$. To seamlessly integrate this new signal without disrupting the initial capability of pretrained models, we adopt a zero-initialization strategy, *i.e.,* the weights corresponding to the prior channels are initialized to zero. The network then progressively learns to extract and utilize meaningful representations. More illustrations and ablations can be found in our supplementary material.

At the *training constraint* end, we propose (1) Prior Consistency Loss for more stable constancy, and (2) Decoding-based Regularization that enforces the learning of illumination-invariant semantic representations.

As for the **prior consistency loss**, we first utilize an inverse ISP-based darkening process [6] to synthesize low-light data, aiming to narrow the gap between the priors of synthesized low-light images $p^l_{\text{iip}} = G_{\text{iip}}(I_{\text{low}})$ and their normal-lit counterparts $p^n_{\text{iip}} = G_{\text{iip}}(I_{\text{normal}})$. We further employ an outlier-robust loss to ensure training stability against artifacts in the synthetic data. Specifically, we disregard the instances that exhibit a high discrepancy (i.e., outliers), preventing them from contributing to the gradient update. Formally, we first define the discrepancy map $d = |G_{\text{iip}}(I_{\text{low}}) - G_{\text{iip}}(I_{\text{normal}})| \in \mathbb{R}^{h\times w}$. After that, we utilize the α-quantile of the discrepancy map as an adaptive threshold d_α, filtering outliers with a large difference:

$$\mathcal{F} = \begin{cases} 0, & d > d_\alpha, \\ 1, & \text{otherwise} \end{cases} . \tag{1}$$

After this, we apply L_2 loss to the remained regions:

$$\mathcal{L}_{\text{iip-consis}} = \left\| \mathcal{F} \odot \left(G_{\text{iip}}(I_{\text{low}}) - G_{\text{iip}}(I_{\text{normal}}) \right) \right\|_2^2 . \tag{2}$$

As for the **decoding-based regularization**, we attach a lightweight auxiliary decoder $\mathcal{D}_{\text{prior}}$ to the backbone's intermediate features $\{f^i\}_{i=1}^N$. This decoder is tasked with a cross-illumination reconstruction: given features from low-light inputs and predict the illumination-invariant prior $p^{\text{normal}}_{\text{iip}}$, which is derived from normal-lit images:

$$\mathcal{L}_{\text{iip-decode}} = \left\| \mathcal{D}_{\text{prior}} \left(\{f^i\}_{i=1}^N \right) - p^{\text{normal}}_{\text{iip}} \right\|_2^2 . \tag{3}$$

The decoder is intentionally small (with $\sim 10\%$ of the backbone's parameters), such that the proposed decoding-based regularization scheme can serve as an information bottleneck, compelling the backbone to distill the most compact, illumination-invariant information.

3.3. Semantic Enrichment: VFM-guided Representation Augmentation

We have established the signal constancy with an illumination-invariant prior and a prior-conditioned architecture, stabilizing feature distribution. For the sake of universal understanding under dynamic low-light scenes, we then enrich such signal constancy with the semantic general prior embedded in VFMs through the proposed two-fold method, dubbed *contrastive augmentation* and *statistical alignment*. The details are as follows:

Contrastive Augmentation. Our motivation is to augment our semantic representations by incorporating the VFM powerful features. Instead of static feature embeddings, we leverage the DINOv2 spatial correlation map to provide more discriminative cues for regularization. As illustrated in Fig. 4, such high-order spatial intra-feature correlation provides more accurate structural semantics, highlighting the semantic-relevant regions more precisely compared to the original feature embedding.

Specifically, given an extracted VFM feature $f_{\text{vfm}} \in \mathbb{R}^{(N_h \cdot N_w) \times D}$ where N_h, N_w denote the patch number along height and width, respectively, we leverage the self-similarity matrix $\mathcal{S} \in \mathbb{R}^{(N_h \cdot N_w) \times (N_h \cdot N_w)}$ to depict the high-order manifold of DINOv2 feature space:

$$\mathcal{S} = \left(\frac{f_{\text{vfm}}}{||f_{\text{vfm}}||_2^2}\right) \cdot \left(\frac{f_{\text{vfm}}}{||f_{\text{vfm}}||_2^2}\right)^\top. \tag{4}$$

We then utilize the self-similarity matrix $\mathcal{S}$ to compute the masks of semantic clusters appointed by certain anchors, which indicate semantic-relevant spatial regions. We filter out low-distance ones with an adaptive percentage threshold $\mathcal{S}_\alpha \in \mathbb{R}^{(N_h \cdot N_w) \times 1}$ for each row in $\mathcal{S}$, obtaining a set of binary masks $\mathcal{M}_s \in \{0, 1\}^{(N_h \cdot N_w) \times (N_h \cdot N_w)}$ which indicates the highlighted regions of each patch position:

$$\mathcal{M}_s = \begin{cases} 0, & \mathcal{S} < \mathcal{S}_\alpha \\ 1, & \text{otherwise} \end{cases}. \tag{5}$$

Finally, we select the anchor region $\mathcal{M} \in \{0, 1\}^{N_h \times N_w}$ based on a given keypoint $p = (n_h, n_w)$:

$$\hat{\mathcal{M}} = \mathcal{M}_s[n_h \cdot N_w + n_w, :], \tag{6}$$

$$\mathcal{M} = \text{reshape}\left(\hat{\mathcal{M}}, (N_h, N_w)\right). \tag{7}$$

After obtaining the semantic relationship mask $\mathcal{M}$, a contrastive feature augmentation mechanism is proposed to enhance discriminability and robustness to degradation at the same time. Treating features inside the mask as semantic-relevant, and those outside ones as semantic-irrelevant, we can construct a positive set to be pulled together, and a negative counterpart to be pushed away. Formally, given an anchor feature f of the synthesized low-light image, we leverage the normal-lit feature at the same position as the positive sample f^+ to maximize similarities, and utilize the features outside the mask as negative samples f^- to maximize the discrepancy in the high-level feature space, as illustrated in Fig. 3. We utilize InfoNCE loss [2] for training:

$$\mathcal{L}_{\text{contra}} = -\log\left(\frac{\sigma(f, f^+)}{\sigma(f, f^+) + \sum_{M[i]=0} \sigma(f, f_i^-)}\right), \tag{8}$$

where $\sigma(x, y) = \exp(x \cdot y / \tau)$, τ denotes the temperature parameter. With this contrastive paradigm, the model can mine regions critical to understanding via VFM-guided structural semantic prior, and minimize feature distances under diverse lighting conditions, improving both robustness and discriminability.

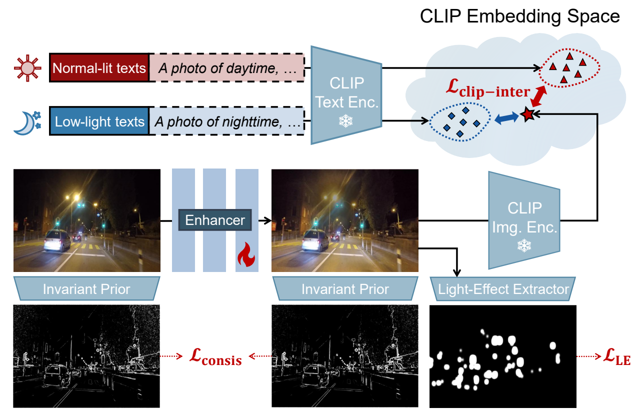


Figure 5. **Machine-oriented low-light enhancement.** We leverage the illumination-invariant prior as a consistent constraint and the light-effect suppression loss for broader generalization. A CLIP-based intermediate loss is proposed to reduce the impact of illumination-relevant attributes, achieving brightened images with preserved and restored semantic cues.

Correlation Alignment. Considering the huge capacity gap between the VFM and our task-specific backbones, forcing the small models to directly mimic the VFM's feature distribution is not only exceptionally difficult but may also introduce unexpected biases, thereby limiting generalization. Therefore, we propose a correlation map-based alignment loss $\mathcal{L}_{\text{align}}$ to construct a unified feature space between the task-oriented backbone and VFM. Given features f, f_{dino} from our backbones and DINOv2, respectively, we first adopt simple `nearest` interpolation to align the resolution, and compute self-correlation map $\mathcal{S}, \mathcal{S}_{\text{dino}}$ as defined in Eq. 4. Then the alignment loss is formulated as:

$$\mathcal{L}_{\text{align}} = \text{CE}(\text{Softmax}(\mathcal{S}), \text{Softmax}(\mathcal{S}_{\text{dino}})), \tag{9}$$

where "CE" denotes cross-entropy loss, "Softmax" indicates the normalized probability along each row of $\mathcal{S}, \mathcal{S}_{\text{dino}}$. Instead of enforcing consistency on static features, we align the VFM embedding space with our smaller backbone using such dynamic self-correlation maps, which focus on contextualized patterns rather than absolute feature values, achieving more generalizable representations.

3.4. Machine-oriented Low-light Enhancement

Although high-level semantics are not fully correlated with low-level signals, empirical evidence suggests that pixel corruption in low-level signal space usually hampers high-level feature extraction. Existing enhancement methods [12, 40], which mainly focus on human visual perception, often overlook the needs of machine perception. Therefore, we propose a semantic-preserved enhancement strategy. The **key challenge** is to obtain richer, high-quality semantic representations instead of degrading them during enhancement. For improving adaptability to complex illumination distortion, we adopt CoLIE [3] as our enhancement backbone, since it learns a unique, per-image enhancement mapping for each low-light input. In pursuit of accurately restored semantics, we extend the signal prior to robustly handle overexposed scenarios, and leverage the rich and robust representations of VFM [32] to design novel, semantics-aware regularizations, as illustrated in Fig. 5.

Overall Enhancement Procedure. Our enhancer is constructed with a four-layer MLP and non-linear activations, which is primarily supervised with four human visual losses proposed by [3], denoted as $\mathcal{L}_{\text{enh}}$. More details can be referred to the supplementary. We further propose three machine-oriented objectives as follows:

- **Light Effect Suppression Loss**. To prevent amplified overexposed light effects, which are detrimental to machine perception, we introduce a light effect suppression loss $\mathcal{L}_{\text{LE}}$. We first extract the illumination-dominated regions $\mathcal{M}_{LE} \in \{0, 1\}^{h \times w}$ following [11], which extends the signal prior to handle overexposure robustly. As illustrated in Fig. 5 (a), this mask effectively captures overexposed areas, such as the reflected light from car lights and streetlights along a nighttime road. We present more visualizations in our supp. The loss is then formulated as the mean value of $\mathcal{M}_{LE}$, which penalizes the total area of the overexposed regions:

$$\mathcal{L}_{\text{LE}} = \frac{1}{hw} \sum_{i=1}^{h} \sum_{j=1}^{w} \mathcal{M}_{i,j}. \tag{10}$$

- **Signal Prior Consistency Loss**. To further maintain semantic integrity, we apply consistency regularization on the illumination-invariant prior:

$$\mathcal{L}_{\text{consis}} = \text{Outlier-Robust-L}_1(p_{\text{iip}}^{\text{enh}}, p_{\text{iip}}^{\text{low}}), \tag{11}$$

  where $p_{\text{iip}}^{\text{enh}}$ and $p_{\text{iip}}^{\text{low}}$ denote the signal prior extracted from the enhanced image and corresponding low-light input. We also utilize the same outlier-robust L_1 in Sec. 3.2.

- **CLIP-based Intermediate Loss**. Inspired by feature whitening [4], we are motivated to erase the illumination attributes of enhanced images, as illustrated in Fig. 5 (c). Since texts naturally embed the attributes of low-light environments, we leverage the powerful cross-modal VFM, CLIP [32] by injecting its semantic cues into pixel correction process. We first collect N_1 daytime texts and N_2 low-light texts, and extract their CLIP embeddings, denoted by $\mathcal{T}_{\text{day}} \in \mathbb{R}^{N_1 \times D}$ and $\mathcal{T}_{\text{night}} \in \mathbb{R}^{N_2 \times D}$, where D is CLIP feature dimension. To reduce the detrimental effect of poor lighting conditions, we create an intermediate manifold that lies in the middle area between normal-lit and low-light CLIP embedding space, where illumination attributes are removed to reduce domain shift. Formally,

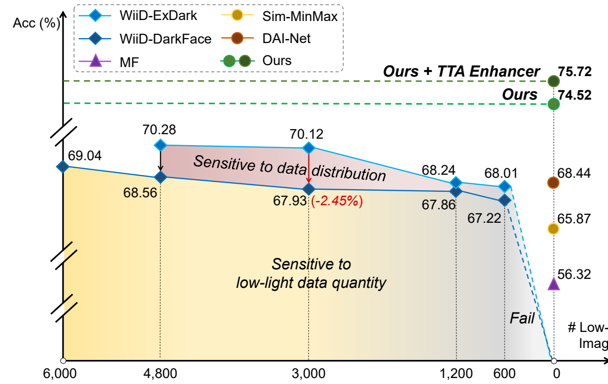


Figure 6. Our **superior performance** and **more generalizable practicality** compared to methods including low-light data-available adaptation (◇), zero-shot adaptation (○), and zero-shot signal enhancement (△), which suffer from the sensitivity of low-light distribution and data quantity.

  we first compute the softmax-normalized cosine similarity $p_{\text{sim}} \in \mathbb{R}^{N_1+N_2}$ between CLIP features of enhanced images $\mathcal{E}_{\text{clip}}(I_{\text{enh}})$ and $\mathcal{T}_{\text{day}}$, $\mathcal{T}_{\text{night}}$:

$$\mathcal{T}_{\text{day-night}} = \text{Concat}(\mathcal{T}_{\text{day}}, \mathcal{T}_{\text{night}}), \tag{12}$$

$$p_{\text{sim}} = \text{Softmax}\left(\frac{\mathcal{E}_{\text{clip}}(I_{\text{enh}}) \cdot \mathcal{T}_{\text{day-night}}}{||\mathcal{E}_{\text{clip}}(I_{\text{enh}})|| \cdot ||\mathcal{T}_{\text{day-night}}||}\right), \tag{13}$$

  where I_{enh} denotes enhanced images. A max-entropy loss is then proposed to construct an intermediate manifold:

$$\mathcal{L}_{\text{clip-inter}} = \sum_{i} p_{\text{sim}}^{i} \cdot \log p_{\text{sim}}^{i}. \tag{14}$$

3.5. Overall Pipeline of Two-Stage Training

Our **UniPrior** can improve the low-light generalization of well-trained daytime models without requiring any real-world low-light data, by leveraging signal constancy and enriched semantics. It consists of two stages:

Stage 1: Offline Adaptation Training. We utilize Dark-ISP [6] to synthesize paired labeled low-light data, and train the high-level adaptation components:

$$\mathcal{L}_{\text{high}} = \lambda_{\text{task}} \cdot \mathcal{L}_{\text{task}} + \lambda_{\text{con}} \cdot \mathcal{L}_{\text{iip-consis}} + \lambda_{\text{dec}} \cdot \mathcal{L}_{\text{iip-decode}} + \lambda_{\text{contra}} \cdot \mathcal{L}_{\text{contra}} + \lambda_{\text{ca}} \cdot \mathcal{L}_{\text{align}},$$

where $\mathcal{L}_{\text{task}}$ denotes task-specific loss.

Stage 2: Test-Time Adaptation Enhancement. With the learned illumination-invariant prior, UniPrior performs sample-adaptive optimization to refine low-level robustness on unseen real data:

$$\mathcal{L}_{\text{low}} = \lambda_{\text{enh}} \cdot \mathcal{L}_{\text{enh}} + \lambda_{\text{entropy}} \cdot \mathcal{L}_{\text{entropy}} + \lambda_{\text{consis}} \cdot \mathcal{L}_{\text{consis}}. \tag{15}$$

4. Experiments

We apply our low-light adaptation framework to tasks including classification, segmentation, and face detection.

Table 1. Top-1 classification accuracy on the CODaN low-light test set [20] and computational complexity.

Method	Top-1 (%)	FLOPs	Params	Time (s)
Baseline				
ResNet-18 [15]	53.32	1.80 G	11.70 M	0.005
Low-light Enhancement				
MF [9]	56.32	-	-	0.062
EnlightenGAN [19]	56.68	12.61 G	8.64 M	0.014
Zero-DCE [12]	56.91	4.01 G	79.42 k	0.010
CoLIE [3]	57.36	662 G	133.12 k	0.629
Zero-DCE++ [21]	57.96	5.17 M	10.56 k	0.002
SCI [28]	58.68	23.79 M	0.258 k	0.001
URetinexNet [42]	58.72	43.8 G	340.11 k	0.088
GEFU [37]	60.92	190.02 G	598 M	0.121
Zero-shot Low-light Domain Adaptation				
MAET [6]	56.48	1.80 G	11.70 M	0.005
CIConv [20]	60.32	1.73 G	11.69 M	0.005
Sim-MinMax [27]	65.87	1.80 G	11.70 M	0.005
DAI-Net [8]	68.44	1.80 G	11.70 M	0.005
Ours	**74.52**	2.04 G	11.72 M	0.005
Ours + TTA Enhancer	**75.72**	8.66 G	11.85 M	2.593

4.1. Low-light Image Classification

Experimental Setting. CoDaN [20] is used for evaluation, which collects 12,500 normal-lit images and 2,500 low-light images within 10 classes for low-illumination adaptation. We adopt ResNet-18 [15] as our baseline.

Quantitative Evaluation. As shown in Tab. 1, enhancement methods concentrate on a plausible human visual experience, however hampering machine perception. Previous zero-shot low-light adaptation methods lack more general prior guidance, leading to suboptimal results. By comparison, our method achieves the best classification accuracy by a large margin, demonstrating our superior capability to handle more complex low-light domain shifts. We further report the computational complexity, network parameters, and processing time for input images of resolution 224×224×3 in Tab. 1. Compared to prior methods, our main model introduces negligible overhead. We also present an optional, TTA-enhanced version ('*Ours + TTA Enhancer*') for further improved performance. In addition, our accuracy achieves 72.12% after 2 epochs training, surpassing current SotA [8] by 3.68%, demonstrating both superior efficiency and effectiveness of our UniPrior.

Justification of Practicality in Real Scenarios. We argue that adaptation methods that require low-light data will inevitably overfit to specific degradation or scenarios. To substantiate this claim, we retrain a data-required adaptation method, WiiD [23], with different low-light datasets (DarkFace [46], 6,000 images & ExDark [26], 4,800 images) and varying amounts of data, observing a significant performance drop, as illustrated in Fig. 6. When the low-light data distribution changes, the accuracy drops by 1.72% under the same data quantity. In addition, when the training data is reduced from 4,800 to 600, the accuracy drops by 2.27% at most. As can be seen from the above experiments, low-light data-required adaptation is sensitive to both the *distribution* and *quantity* of low-light data due to the lack of universal priors for generalizable understanding. However, in real-world applications, it is infeasible to obtain a large amount of low-light data with appropriate degradation and scenes that match dynamic test-time conditions. By comparison, our UniPrior **outperforms all** existing zero-shot adaptation methods with a significant gap (7.28%), and requires strictly no low-light data for training, establishing a new state-of-the-art on low-light understanding with superior performance, robustness, and practical utility.

Table 2. Low-light semantic segmentation and face detection performance on Nighttime Driving (ND) [7], ACDC-Nighttime (ACDC-N) [34] and DarkFace [46].

Method	Segmentation		Face Detection
	ND	ACDC-N	DarkFace
Baseline			
RefineNet [25]	34.3	23.8	-
DSFD [22]	-	-	16.1
Low-light Enhancement			
EnlightenGAN [19]	25.2	16.8	31.3
URetinexNet [42]	28.1	17.9	31.4
SCI [28]	28.6	20.9	34.5
MF [9]	28.8	17.4	41.4
ZeroDCE [12]	29.9	18.7	41.3
GEFU [37]	30.2	19.9	29.4
ZeroDCE++ [21]	32.7	21.0	40.9
CoLIE [3]	33.2	26.8	32.7
Zero-shot Low-light Domain Adaptation			
MAET [6]	28.1	18.2	25.2
CIConv [20]	41.2	27.4	18.4
DAI-Net [8]	41.4	25.8	28.0
Sim-MinMax [27]	44.9	27.6	25.7
Ours	**48.2**	**27.9**	**31.3**
Ours + TTA Enhancer	**48.6**	**28.5**	**33.6**

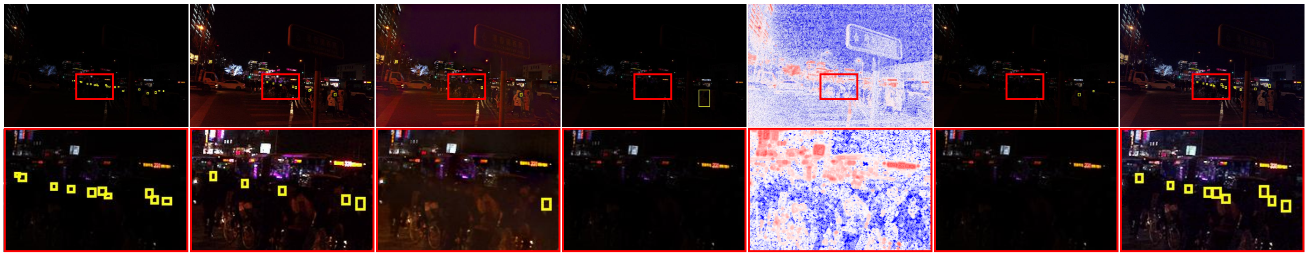
4.2. Low-light Semantic Segmentation

Experimental Setting. We further evaluate on semantic segmentation with RefineNet [25] as a baseline. We use Cityscapes [5] dataset as a daytime training domain, which contains 2,975 images. We utilize Nighttime Driving [7] and ACDC-Nighttime [34] as segmentation benchmarks to evaluate with 50 and 106 low-light images, respectively.

Quantitative Evaluation. Tab. 2 reports performance comparison by mIoU(%). Enhancement methods yield worse results than the baseline due to the nonuniform illumination of the nighttime street scene in these two datasets. Sim-MinMax [27] employs the hardest synthesized low-light images to train high-level models within a contrastive paradigm; however lacks of general priors. In contrast, our framework leverages the most comprehensive signal and semantic priors, achieving the best performance.

4.3. Low-light Face Detection

Experimental Setting. We first conduct our experiments on low-light face detection following HLA-Face [38]. WIDER Face [44] serves as the daytime source domain, which holds 12,880 normal-lit images with human face annotations for training. We evaluate our method on the Dark-



(a) Ground Truth (b) RUAS (c) EnlightenGAN (d) DSFD (e) CIConv (f) DAI-Net (g) Ours

Figure 7. Qualitative comparison of different low-light face detection methods.

Table 3. Ablation study of high-level feature augmentation and low-level signal enhancement on classification on CoDaN [20] and segmentation on Nighttime Driving [7].

Variants	Top-1 ACC (%)	mIOU-ND (%)
High-level Feature Augmentation		
w/o invariant prior	67.24	44.62
w/o zero-init.	62.25	39.31
w/o prior consistency loss	72.23	45.72
L_1 as prior consistency loss	73.63	47.59
w/o prior decoder	70.26	45.23
w/o contrastive aug.	68.89	47.07
w/o correlation align.	69.71	46.67
with naive feature align.	63.26	41.72
Ours	**74.52**	**48.20**
Low-level Signal Enhancement		
w/o machine-oriented designs	71.48	46.44
w/o prior consistency loss	74.11	47.26
w/o CLIP intermediate loss	74.86	48.52
w/o light-effect loss	74.29	48.39
Ours + TTA Enhancer	**75.72**	**48.61**

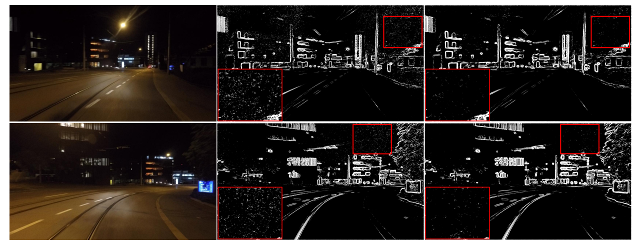
Face [46] test set, which contains 4,000 low-light images with small human faces and nonuniform illumination conditions. We adopt DSFD [22] as the baseline detection model.

Quantitative Evaluation. Quantitative performances are reported by mean Average Precision (mAP) under threshold 0.5 at Tab. 2. We provide a subjective comparison in Fig. 7. Our method outperforms all zero-shot adaptation techniques, establishing a new state-of-the-art with a margin of 5.6% mAP over the prior leading method [8]. Benefiting from our TTA enhancement in Sec. 3.4, our framework is further improved by pre-processing the input signal in pixel space to adapt to unseen low-light distribution.

4.4. Ablation Experiments and Analysis

To rigorously validate the effectiveness of our each design, we conduct a thorough ablation in Tab. 3 and provide further in-depth analysis with visualizations in Fig. 8.

High-level Feature Augmentation. As shown in Tab. 3, removing signal constancy modules or semantic enrichment modules both lead to a significant performance drop, demonstrating their necessity. The decoding-based regularization achieves a significant performance gain (70.26% vs 74.52%), underscoring its crucial contribution. Notably, replacing our correlation alignment with naive feature alignment using $\mathcal{L}_1$ loss causes one of the most substantial degradations, which validates our core insight (Fig. 4) that VFM correlation maps provide more discriminative cues than static feature embeddings, thereby enabling more compact and generalizable semantic representations.



Input Pretrained QuadPrior Ours

Figure 8. Visualization of the pretrained/ours signal prior.

Machine-oriented Enhancement. Our TTA Enhancer contributes to a substantial performance boost (74.52% vs 75.72%) by adaptively correcting pixel values according to the input distribution with the guidance of our proposed machine-oriented regularizations. Compared to our main model, the TTA-version surpasses all existing methods, offering a reasonable cost-complexity trade-off when the maximal performance is prioritized. We provide more visualized comparisons in our supp.

Signal Prior Improvement by Generalizable Training. We visualize the signal priors from our framework and the original QuadPrior [40] in Fig. 8. While QuadPrior is designed to align with human visual preferences, it still retains significant noise artifacts. In contrast, our machine-oriented prior, refined through generalization training, is cleaner and more stable, proving to be a more effective and generalizable signal representation.

5. Conclusion

We propose UniPrior, a unified prior-based low-light adaptation framework that ensures signal constancy with an illumination-invariant prior and enhances semantics using a universal, degradation-insensitive prior embedded in VFMs. A VFM context-informed contrastive training strategy is proposed to enable generalizable adaptation under varying low-light conditions. By integrating unified prior into pixel-level correction, our model reduces the domain gap with rich semantics. Extensive experiments across multiple tasks validate our superiority and wide applicability.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grant 62332010, and in part by the Beijing Major Science and Technology Project under Contract No. Z251100008425023.

References

[1] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Int. Conf. Comput. Vis.*, 2023. 3

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Int. Conf. Machine Learning*, 2020. 5

[3] Tomáš Chobola, Yu Liu, Hanyi Zhang, Julia A. Schnabel, and Tingying Peng. Fast context-based low-light image enhancement via neural implicit representations. In *Eur. Conf. Comput. Vis.*, 2024. 3, 6, 7

[4] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2021. 6

[5] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2016. 7

[6] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. Multitask aet with orthogonal tangent regularity for dark object detection. In *Int. Conf. Comput. Vis.*, 2021. 2, 3, 4, 6, 7

[7] Dengxin Dai and Luc Van Gool. Dark model adaptation: Semantic image segmentation from daytime to nighttime. In *Int. Conf. Intel. Transportation Systems*, 2018. 7, 8

[8] Zhipeng Du, Miaojing Shi, and Jiankang Deng. Boosting object detection with zero-shot day-night domain adaptation. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024. 3, 7, 8

[9] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John W. Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Process.*, 129: 82–96, 2016. 3, 7

[10] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2023. 3

[11] Zhigang Wang Dong Wang Bin Zhao Guanzhou Lan, Yuqi Yang and Xuelong Li. Night-to-day translation via illumination degradation disentanglement. *arXiv:2411.14504*, 2024. 6

[12] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2020. 1, 3, 5, 7

[13] Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.*, 26(2):982–993, 2017. 1, 3

[14] Khurram Azeem Hashmi, Goutham Kallempudi, Didier Stricker, and Muhammad Zeshan Afzal. FeatEnHancer: Enhancing hierarchical features for object detection and beyond under low-light vision. In *Int. Conf. Comput. Vis.*, 2023. 1

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2016. 7

[16] Mingbo Hong, Shen Cheng, Haibin Huang, Haoqiang Fan, Shuaicheng Liu, et al. You only look around: Learning illumination-invariant feature for low-light object detection. In *Annu. Conf. Neural Inf. Process. Systems*, 2024. 1

[17] Hai Jiang, Ao Luo, Songchen Han, Haoqiang Fan, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. In *Siggraph Asia*, 2023. 3

[18] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. LightenDiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *Eur. Conf. Comput. Vis.*, 2024. 3

[19] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.*, 30:2340–2349, 2021. 3, 7

[20] Attila Lengyel, Sourav Garg, Michael Milford, and Jan C. van Gemert. Zero-shot domain adaptation with a physics prior. In *Int. Conf. Comput. Vis.*, 2021. 2, 3, 4, 7, 8

[21] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(8): 4225–4238, 2021. 3, 7

[22] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. DSFD: dual shot face detector. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2019. 7, 8

[23] Yunan Li, Yihao Zhang, Shoude Li, Long Tian, Dou Quan, Chaoneng Li, , and Qiguang Miao. Watching it in dark: A target-aware representation learning framework for high-level vision tasks in low illumination. In *Eur. Conf. Comput. Vis.*, 2024. 1, 3, 7

[24] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Int. Conf. Comput. Vis.*, 2023. 3

[25] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian D. Reid. RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2017. 7

[26] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Comput. Vis. and Image Understanding*, 178:30–42, 2019. 7

[27] Rundong Luo, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Similarity min-max: Zero-shot day-night domain adaptation. In *Int. Conf. Comput. Vis.*, 2023. 2, 3, 7

[28] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022. 3, 7

[29] Igor Morawski, Kai He, Shusil Dangi, and Winston H Hsu. Leveraging content and context cues for low-light image enhancement. *IEEE Trans. Multimedia*, 2025. 1, 3

[30] Maxime Oquab, Timothée Darcet, Theo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Russell Howes, Po-Yao Huang, Hu Xu, Vasu Sharma, Shang-Wen Li, Wojciech Galuba, Mike Rabbat, Mido Assran, Nicolas Ballas, Gabriel Synnaeve, Ishan Misra, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision. *IEEE Trans. on Machine Learning Research*, 2024. 2, 4

[31] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller. Contrast-limited adaptive histogram equalization: Speed and effectiveness. In *Vis. Biomed. Comput.*, 1990. 1, 3

[32] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Int. Conf. Machine Learning*, 2021. 2, 3, 4, 6

[33] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. Retinex processing for automatic image enhancement. *J. Electron. Imaging*, 2004. 2

[34] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In *Int. Conf. Comput. Vis.*, 2021. 7

[35] Yukihiro Sasagawa and Hajime Nagahara. YOLO in the dark - Domain adaptation method for merging multiple models. In *Eur. Conf. Comput. Vis.*, 2020. 3

[36] Fangtong Sun, Congyu Li, Ke Yang, Yuchen Pan, Hanwen Yu, Xichuan Zhang, and Yiying Li. FRBNet: Revisiting Low-Light Vision through Frequency-Domain Radial Basis Network. In *Annu. Conf. Neural Inf. Process. Systems*, 2025. 1

[37] Sen Wang, Shao Zeng, Tianjun Gu, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, and Lizhuang Ma. From enhancement to understanding: Build a generalized bridge for low-light vision via semantically consistent unsupervised fine-tuning. In *Int. Conf. Comput. Vis.*, 2025. 3, 7

[38] Wenjing Wang, Wenhan Yang, and Jiaying Liu. HLA-Face: Joint high-low adaptation for low light face detection. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2021. 1, 3, 7

[39] Wenjing Wang, Xinhao Wang, Wenhan Yang, and Jiaying Liu. Unsupervised face detection in the dark. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):1250–1266, 2023. 3

[40] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2024. 1, 2, 3, 4, 5, 8

[41] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Brit. Mach. Vis. Conf.*, 2018. 3

[42] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022. 3, 7

[43] Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. HVI: A new color space for low-light image enhancement. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2025. 3

[44] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. WIDER FACE: A face detection benchmark. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2016. 7

[45] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Int. Conf. Comput. Vis.*, 2023. 3

[46] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Trans. Image Process.*, 29:5737–5752, 2020. 7, 8

[47] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with A generative diffusion model. In *Int. Conf. Comput. Vis.*, 2023. 3

[48] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM Int. Conf. Multimedia*, 2019. 3

[49] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *Int. J. Comput. Vis.*, 129(4):1013–1037, 2021. 3